DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIEDSCIENCES

THE INVESTIGATION OF THE EFFECTS OF VALUE OF TIME AND RELIABILITY ON THE USE OF PUBLIC TRANSPORTATION

by

Husam J.M. ABEDKHADER

December, 2014

İZMİR

THE INVESTIGATION OF THE EFFECTS OF VALUE OF TIME AND RELIABILITY ON THE USE OF PUBLIC TRANSPORTATION

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of DokuzEylül Universityin
Partial Fulfillment of the Requirements for the Degree of Master ofScience in
Civil Engineering, Transportation Program

by

Husam J.M. ABEDKHADER

December, 2014

İZMİR

M.Sc. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"THE INVESTIGATION OF THE EFFECTS OF VALUE OF TIME AND RELIABILITY ON THE USE OF PUBLIC TRANSPORTATION"** completed by **HUSAM J.M. ABEDKHADER** under supervision of **ASS. PROF. DR. MUSTAFA ÖZUYSAL** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Ass. Prof. Dr. Mustafa ÖZUYSAL
Supervisor

Assoc. Prof. Dr. Serhan TANYEL
(Jury Member)

Ass. Prof. Dr. Umay Uzunoğlu KOÇER
(Jury Member)

Prof. Dr. Ayşe OKUR
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENT

I am using this opportunity to express my special appreciation and thanks to my supervisor, Ass. Prof.Dr. Mustafa ÖZUYSAL for his continuous support during my M.Sc. study and research, for his patience, motivation, enthusiasm and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my M.Sc. thesis.Herein, I would also give my sincere thanks to Assoc. Prof. Dr. Serhan TANYEL. Dr. TANYEL's great personality, strong background in Transportation engineering and his continuing help contributed to the completion of this thesis.Thanks are also given to Prof. Dr. Burak ŞENGÖZ and to Ass. Prof.Dr. Yıldırım ORAL for their support throughout my master study.

I would also like to thank all of my friends who supported me in writing, and incented me to strive towards my goal. Continues thanks also to my classmates for their help and supportduring my study.

A special thanks to great Izmir and their lovely people. It gave me wonderful and unforgettable memories that will stay in my mind till the last breath.

Last but not the least; I would like to thank my great family. Words cannot express how grateful I am tomy parents Jamal and Sana'a, for giving birth to me at the first place and supporting me spiritually throughout my life.

Husam J.M. ABEDKHADER

THE INVESTIGATION OF THE EFFECTS OF VALUE OF TIME AND RELIABILITY ON THE USE OF PUBLIC TRANSPORTATION

ABSTRACT

Reliability measures are typically used to evaluate the service quality of public transport systems. A definition of reliability is based on the analysis of the travel time reliability, waiting time reliability, arrival time reliability and probability of arriving to the stops with delay no more “m” minutes. Studies show that reliability can affect the passengers’ mode choice. Moreover, the unreliability contributes to the low level of public transport competition with the privet vehicles. The importance of this study is to show the effects of value of time (VOT) and reliability measures on the public transport smart card use, which forces the transit operators to make more reliable services.

In this thesis, reliability has been measured as a function of real travel time only. The bus arrival time on each stop has been taken and then reliability has calculated for each bus stop in the study area. After this, stops reliability has been converted to represent the 57 districts in the study area. Becker’s (1965) model has been used to predict value of time, by using SPSS software to model the household travel survey dataset which it had been collected in 2007 by Izmir Municipality. Also, average income and development level for each district had taken a place in this study. Thereafter, the relations between these different variables and public transport smart card use of three months for a specified zone in İzmir/ Turkey have been examined. Calculations show that people of Izmir prefer to take public transport bus or to wait the next service than use other mode. Reliability measures do not affect the consuming of public transport use. In other words, the boarding on public transport services increase even if the reliability measures are low and do not satisfy the passengers' needs (travel time).

Keywords: Reliability measures, value of time, public transport, smart card, logistic regression.

GÜVENİLİRLİK VE ZAMAN DEĞERİNİN TOPLU ULAŞIM KULANIMI ÜZERİNDEKİ ETKİLERİNİN İNCELENMESİ

ÖZ

Güvenilirlik ölçümleri genel olarak toplu taşıma sistemlerinin hizmet kalitesini değerlendirmek için kullanılır. Güvenilirliğin tanımı, yolculuk süresi güvenilirliğine, bekleme süresi güvenilirliğine, varış zamanı güvenilirliğine ve duraklara "m" dakikadan daha az gecikme ile varış olasılığına bağlıdır. Yapılan çalışmalar güvenilirliğin yolcuların tür seçimlerini de etkileyebildiğini göstermektedir. Ayrıca, güvensizlik toplu taşımanın özel araçlarla olan rekabetinin düşük seviyede olmasına katkı sağlamaktadır.

Bu çalışmada, güvenilirlik sadece gerçek yolculuk zamanının bir fonksiyonu olarak ölçülmüştür. Her bir durak için otobüs varış zamanı alınmış ve çalışma sahasındaki her bir durak için güvenilirlik hesaplanmıştır. Daha sonra, durakların güvenilirlikleri çalışma sahasındaki 57 bölgeyi temsil edecek şekilde dönüştürülmüştür. Zaman değerinin tahmini, 2007 yılında İzmir Büyükşehir Belediyesi tarafından toplanan hane halkı ulaşım anketi verilerini modellemek için SPSS yazılımı kullanılarak, Becker (1965) modeli ile yapılmıştır. Ayrıca bu çalışmada her bölge için ortalama gelir ve gelişmişlik seviyesi de yer bulmuştur. Daha sonra, bu farklı değişkenler ve sağlanan İzmir ilindeki belirli bir bölgede 3 aylık toplu taşıma akıllı kart (kentkart) kullanımı arasındaki ilişkiler incelenmiştir.

Hesaplamalar İzmir halkının belediye otobüsüne binmeyi ya da daha sonraki seferi beklemeyi diğer ulaşım modu tercih ettiklerini göstermektedir. Güvenilirlik ölçümleri toplu taşıma kullanımının tüketimini etkilememektedir. Başka bir deyişle, güvenilirlik ölçümleri düşük de olsa ve yolcuların ihtiyaçlarını karşılamasa da toplu taşıma servislerinin kullanımı artmaktadır.

Anahtar kelimeler: Güvenilirlik ölçümleri, zamanın değeri, toplu ulaşım, akıllı kart, lojistik regresyon.

CONTENTS

Page

THESIS EXAMINATION RESULT FORM ii
ACKNOWLEDGEMENTS iii
ABSTRACT iv
ÖZ v
LIST OF FIGURES ix
LIST OF TABLES xi

CHAPTER ONE – INTRODUCTION 1

1.1 Background 1
1.2 Thesis Outline 2

CHAPTER TWO – LITERATURE REVIEW 4

2.1 Definition of Value of Time 4
2.1.1 Empirical Estimates of the Value of Time 7
2.1.2 Value of Time Models 9
2.1.2.1 Becker (1965) 10
2.1.2.2 Oort (1969) 11
2.1.2.3 De-Serpa (1971) 12
2.1.2.4 Small (1982) 15
2.1.2.5 Jara– Diaz (2003) 16
2.2 Discrete Choice Analysis 21
2.2.1 The Behavioral Models 21
2.2.2 The Logit Model 23
2.2.2.1 Maximum Likelihood Estimation 24
2.2.2.2 Limitation of Logit Model 26
2.2.3 Logistic Regression 26
2.3 Definition of Reliability 28
2.3.1 Factors Affecting Reliability 30

2.3.2 Factors Affecting Reliability 31
2.3.2.1 Travel Time 31
2.3.2.2 On-Time Performance 32
2.3.2.3 Headway Regularity 33
2.3.2.4 Waiting Time 34
2.3.2.5 Transfer Time 35
2.3.2.6 Buffer Time 35
2.3.3 Measures of Reliability 36
2.3.3.1 Travel Time Reliability 36
2.3.3.2 Headway Reliability 37
2.3.3.3 Passenger Wait Time Reliability 37
2.4 Public Transport Smart Card 38
2.4.1 Smart Card Data Use in Public Transport Planning 40
2.4.1.1 Strategic-Level Studies 41
2.4.1.2 Tactical-Level Studies 42
2.4.1.3 Operational-Level Studies 43

CHAPTER THREE – MODEL STRUCTURE 45

3.1 Value of Time Model 46
3.1.1 Becker Model (1965) 47
3.2 Logistic Regression Analysis 50
3.2.1 Multiple Logistic Regression Analysis 52
3.3 Reliability Measure Model 53
3.3.1 Travel Time Reliability Model 53

CHAPTER FOUR – STUDY AREA 55

4.1 General Information about İzmir 55
4.2 Study Area 61
4.3 The Datasets 64
4.3.1 Statistical Analysis of the Household Travel Survey Dataset 66

CHAPTER FIVE – MODELING PROCESS 71

5.1 Value of Time Estimation 71
5.1.1 Logistic Regression Using SPSS 71
5.1.1.1 Block 0: Beginning Block 72
5.1.1.2 Block 1: Method = Enter 73
5.1.2 The Value of Time 77
5.2 Average Income and Development Level 83
5.3 Reliability Estimation 86
5.4 Statistical data of Smart Card Use 92
5.5 Correlation 98
5.5.1 Pearson's Product-moment Coefficient 97
5.6 Linear Regression 100
5.6.1 WdScB as Dependent Variable 104
5.6.2 RaWdUScB as Dependent Variable 105
5.6.3 RaWeUScB as Dependent Variable 107
5.6.4 RaTUScB as Dependent Variable 108
5.6.5 Reliability 1 as Dependent Variable 109
5.6.6 Reliability 2 as Dependent Variable110 110

CHAPTER SIX – CONCLUSION 111

REFERENCES 113

LIST OF FIGURES

Page

Figure 2.1 Atypical smart card system flowchart 39
Figure 2.2 Simplified database model for boarding data storage..... 40
Figure 3.1 Modeling process..... 46
Figure 3.2 The logit transformation 51
Figure 4.1 İzmir Metropolitan Municipality 55
Figure 4.2 İzmir Metropolitan Municipality's public transport systems 56
Figure 4.3 The stations of light rail system (METRO) 57
Figure 4.4 The stations of commuter rail system (IZBAN) 58
Figure 4.5 The rail systems (METRO) and (IZBAN) and there stations..... 58
Figure 4.6 The ferry system and the ferry stations 59
Figure 4.7 Bus system network..... 60
Figure 4.8 The study area boundaries 61
Figure 4.9 The 57 districts that have a complete dataset 62
Figure 4.10 The bus lines routes inside the boundaries of the study area 63
Figure 4.11 Overview of respondents 65
Figure 4.12 The percentage of individuals who own autos 66
Figure 4.13 The percentage of individuals who own P.T smart card 66
Figure 4.14The individuals' gender percentage of the household travel survey dataset 67
Figure 4.15 Gender vs. mode choice 67
Figure 4.16 The students percentage in the household travel survey dataset 68
Figure 4.17 Student or non- vs. mode choice..... 68
Figure 4.18 The education level of the household travel survey dataset 69
Figure 4.19 Education level vs. mode choice 69
Figure 4.20The individuals' working percentage of the household travel survey dataset 70
Figure 4.21 Working or not vs. mode choice..... 70
Figure 5.1 The estimated VOT 1 of the study area..... 80
Figure 5.2 The estimated VOT 2 of the study area 81
Figure 5.3 The estimated VOT 3 of the study area 82

Figure 5.4 The average income of the study area 84
Figure 5.5 The development level of the study area 85
Figure 5.6 The conceptual plan of route 121 in Izmir Metropolitan Municipality with its stops location 86
Figure 5.7 The boxed section is chosen for the detailed stop reliability study 87
Figure 5.8 The Reliability 1 measures of the study area 90
Figure 5.9 The Reliability 2 measures of the study area 91
Figure 5.10 The RWdScB for all districts in study area 95
Figure 5.11 The RWeScB for all districts in study area 95
Figure 5.12 The RaWdUScB for all districts in study area 96
Figure 5.13 The RaWeUScB for all districts in study area 96

LIST OF TABLES

Page

Table 2.1 Empirical estimates of the value of time 9
Table 2.2 Value of time from the main time allocation approaches 19
Table 2.3 The definitions shortcuts that used in the time allocation models 20
Table 2.4 Review of studies on smart card systems for strategic public transport ... 41
Table 2.5 Review of studies on smart card systems for tactical transit planning 43
Table 2.6 Review of studies on smart card systems for operational transit planning 44
Table 4.1Sample of public transport boarding types for weekdays (these data have been collected between 07 and 11 May, 2012) 57
Table 4.2 Public transport mode choice according to gender case 67
Table 4.3 Public transport mode choice according to student case 68
Table 4.4 Public transport mode choice according to the education level 69
Table 4.5 Public transport mode choice according to the working case 70
Table 5.1 Case processing summary 72
Table 5.2 Classification table 72
Table 5.3 Variables in the equation table 5.4 variables not in the equation 73
Table 5.4 Variables not in the equation 73
Table 5.5 Omnibus tests of model coefficients 74
Table 5.6 Model summary 74
Table 5.7 Variables in the equation 75
Table 5.8 Classification table 76
Table 5.9 The three model estimated value of time for all districts 79
Table 5.10 Average income and development level for all districts 83
Table 5.11 Reliability calculations for Alsancak 311 district 88
Table 5.12 The reliability values for all districts 89
Table 5.13 The statistical terms of the public transport smart card use 92
Table 5.14 The statistical values of public transport smart card usage 94
Table 5.15Correlations matrix for the study's main measures and the statistical usage of public transport smart card 99
Table 5.16 Correlations matrix of the study main measures 103
Table 5.17 Summary output (Dependent variable is WdScB) 105

Table 5.18 Summary output (Dependent variable is RaWdUScB) 106
Table 5.19 Summary output (Dependent variable is RaWeUScB) 107
Table 5.20 Summary output (Dependent variable is RaTUScB)..... 108
Table 5.21 Summary output (Dependent variable is Reliability 1) 109
Table 5.22 Summary output (Dependent variable is Reliability 2) 110

CHAPTER ONE
INTRODUCTION

1.1 Background

Reliability is one of the most important characteristics of public transportation service quality both from service recipients (passengers) and service providers (operators) points of view. It has a direct impact on mode choice, ridership, and the cost of providing the service. Reliability affects public transport users in one of the two ways: as a delay when picking up the passenger and as a delay when the passenger is on the service. Reliability measures are typically used within performance regimes to evaluate the quality of service of public transport providers. Moreover, the unreliability contributes to the low level of public transport competition with the private vehicles.

Reliability can be defined as the ability of the service to provide a consistent service over a period of time (Polus, 1978). This definition of reliability is based on the analysis of the different types of time: travel time reliability, waiting time reliability, arrival time reliability, and punctuality. In this thesis reliability has been measured as a function of real arrival time on each bus stop in the study area, and then these measures have been converted to represent the 57 districts in the study area. Travel time reliability is defined as the mean over standard deviation of travel time. The inclusion of the mean in addition to the standard deviation as in Polus (1978) is to allow for comparison of the variations in bus routes or sections of a route that have significantly different mean values.

Value of time is an explanatory variable used for the modeling of mode choice and represents the importance of passengers' time loss while they are traveling. The investigations show that the users pay more importance for the time loss occurring out of vehicle especially in waiting time at stops and access to stops. Public transport smart cards that have been used in public transportation systems of many cities in Turkey, provide the ability to predict the departure and arrival times, travel time,

origin-destination and mode choice behavior of the users by tracking them in an individual manner.

In this research the public transport smart card data of three months for a selected analysis zones in Izmir/ Turkey provided by ESHOT General Directorate of İzmir Public Bus System [ESHOT] are analyzed to determine the most effective reliability measure on passengers' behavior. Value of time is also predicated on a selected analysis zone by using the data 2007 İzmir household travel survey with several methods. Then the relationship between the reliability measures, the value of time and public transport use (public transport smart card usage number) are investigated by using linear and nonlinear modeling techniques. For this purpose, multiple regression analysis by using SPSS Statistic has been used.

1.2 Thesis Outline

The remainder of the thesis is organized as follows:

Chapter 2 presents the literature review for this study, including both value of time models and reliability models, in addition to these two main models district choice models has been also discussed in this chapter, logistic regression has beenusedfor modeling district choice analysis. Finally, the public transportation smart card system has been also presented in this chapter.

Chapter 3 shows in detail the model structure for this study. Both the value of time models and reliability measures has beenexplained in a simple way. This chapter illustrates the development of the value of time models through five different models from the earliest Becker's 1965 model until the last one Jara- Diaz's 2003 model. Discrete choice model and reliability measures take a place in this chapter.

Chapter 4 describes the study area and the data collected for this study. The data sources include Geographic Information Services (GIS) data that describes the

transportation environment in Izmir Metropolitan Municipality. Also, the different types of public transport system have been discussed in general in this chapter.

Chapter 5 provides the study's empirical results. These results contain the three different values of time (VOT 1, VOT 2 and VOT 3), the two reliability measures (Reliability 1 and Reliability 2) and also both average income and development level for each district in the study area. These seven variables have been estimated by using linear and nonlinear modeling techniques and then these findings have been discussed and explained clearly by using maps and charts.

Finally, Chapter 6 concludes the dissertation with a summary of the main findings and limitations that must be attached to them.

CHAPTER TWO
LITERATURE REVIEW

This chapter focuses on and illustrates the different literature researches and studies that have a strong relationship with this thesis. It starts with the main time allocation theories from Baker's 1965 model until the model of Jara-Diaz 2003. And then, moves to discrete choice analysis and shows how maximum likelihood and logistic regression technique are used to find the value of time. Thereafter, reliability measures take a place in this chapter in six different measures as it shown below. Finally, this chapter explains how public transport smart cards help in collecting data which these data have significant role in analyze and plan the public transport services.

2.1 Definition of Value of Time

The review of several time allocation studies showed that the value of time can be theoretically defined as a derivation from a budget constrained allocation problem of time cross various activities, such as work, leisure and time required consuming goods and services. Time is a finite resource that can either be used for work or leisure. Normally, value of time is equal to the opportunity cost, or to the foregone wage of leisure time. The value of time in these allocation theories is connected to the hourly wage rate.

In the earliest models, time was considered as a commodity not as an input, this consideration has been changed by Becker (1965) when he introduced the time as an input into the model of consumer choice. The treatment that has been made by Becker (1965) on the time as an input in the production of household goods, such as child rearing and other household chores, it also has nontrivial consequences in the field of transportation research. The demand for travel is derived from not only goods but also services that requiring out-of-home travel, because of that, there are different meanings for the value of time. When individuals deciding what to consume, how much and how frequently, they must also take into account that

consumption requires travel and travel requires time. In the contrast with this theory the goods are the only source of satisfaction for the traditional budget constrained utility maximization theory.

DeSerpa (1971) went further by considering that the decision to consume a specified amount of any commodity requires some minimum amount of time to be allocated to it. Because of this the value of saving time has been divided into the value of time as a resource and the value of time as a commodity (Jara-Diaz, 2003). Shaw (1992) announced that the value of an individual's time in an activity varies between the wage rate at a maximum and zero at a minimum. This theory, though, ignores those goods consumption cannot be realized without incurring travel or communication costs.

In this context, there is a difference between the value and the cost of time. The value of time can be represented as a subjective marginal benefit of time spent in a certain activity, while the opportunity cost of time reflects the net benefit of time spent in the next best alternative activity. Thus, the opportunity cost of time is affected directly by the wage rate, the activity a person is engaged in and also by the list of the alternative activities available. The timing of the decision has to be considered when a decision is taken to change from activity to another in the marginal case (Jara-Diaz, 2003).

The developments which are made by Becker (1965) and DeSerpa (1971) in the consumer theory have significant implications. They show that the value of time not only can vary for the same individual depending on the undertaken particular time for a trip differ among individuals, but also it differs among individuals. Both empirical and theoretical researches indicate that the value of time can be significantly higher or lower than the current wage rate of an individual depending on activities that people are involved in (Jara-Diaz, 2002).

For example, Gronau (1976) says that the average wage rate can only provide an approximation of the value of time, while the marginal wage rate is the most used

way that the value of time depends on it. The problem that arises in this situation is that the marginal wage is not directly observable. Thus, the difference between the observable average wage rate and value of time can be caused by such factors as the marginal disutility of work, the costs incurred through work, the marginal utility of travel, the possibility of engaging in other activities while traveling, institutional barriers to changes in the number of working hours, and other factors. The inclusion of this distinction is that a low or zero value of time is not because individuals have a no observable market wage. Similarly, unemployed or low wage individuals do not necessarily have a low opportunity cost of time (Shaw, 1992).

The value of time has different interpretations; these interpretations depend on how the time allocation problem is modeled. For example, Johnson (1966) keeps time out of utility by assuming travel time only as a cost whereas, DeSerpa (1971) includes time in the utility function and consequently has a commuting value. For that, he defines commuting time as the value of time saved or the difference between the scarcity value which is based on alternative use of time and the commodity value of time. There are different ways to model consumers' utility maximization problem and these ways affect the measurements of the value of time. Time can be included in the utility function or in the constraint function. Time does not have a value per second and cannot directly affect the utility when it is only included in the constraint function because it is treated as cost and acts as a budgetary constraint to consumption. On the other hand, when time is included in the utility function, time gains commodity value and can directly influence consumer's utility.

Reichman (1976) concludes in his study that time saving is important for the travelers as other trip characteristics, such as comfort, scenery, physical and mental efforts. This conclusion may explain why people sometimes may be willing to pay for using toll ways even when they provide relatively small time savings. In other words, when travelers facing with a money-time trade-off, they take into account the total cost of the trip, which is determined as the sum of time and utility loss, rather than as the cost of travel time itself. Thus, people sometimes choose the more pleasant travel options even it is slower, reflecting their lower total time cost, or they

can be willing to pay extra for more comfortable seats, reduced transfers, air conditioning, etc. (Litman, 2007).

Hensher (1976) states that numerous empirical travel mode choices studies fail to separate the pure value of time, which is mode-abstract, from the comfort and convenience features of the trip that affect the composite value of travel time. As a result, the separation of the pure value of travel time from another trip attributes affecting travelers' willingness to pay is extremely difficult.

Eventually, the literature review of several theoretical models showed that the value of time has not one unique definition. Through the review of the literature, it is shown that the value of time hinges on the assumption of a minimum-time constraint that requires individuals to travel more than they would choose to travel without the constraints.

2.1.1 Empirical Estimates of the Value of Time

The review of the studies that concerning on the time allocation problem and the value of time that have been mentioned in the above section have shown that personal travel time unit costs are usually estimated at 25 to 50 percent of the prevailing average wage rate. These estimations depend on multiple factors, such as, demographics of the travelers and the characteristics of the trip. As is well known, the cost of travel time tends to be higher for uncomfortable, unsafe and stressful conditions. The earliest studies on the value of time like Becker (1965), Beesley (1965) and Lisco (1967) produced larger estimates ranging from 30 to 50 percent of the average wage.

Calfee and Winston (1998) in particular, by using a stated-preference approach, they found that the commuters' average willingness to pay to reduce travel time ranging from 14 to 26 percent of the gross wage. And they also found that travel conditions do not affect the value of travel time. The fact that commuters are able to

adjust to congestion through their modal and departure time choices, as well as the choice of residential and workplace location are behind these lower estimations.

Various travel time values have been suggested by the United States Department of Transportation [USDOT] to be used in evaluating transportation projects. These suggestion values vary from 50 to 120 percent of the wage rate, depending on the type of travel (business or personal) and also on the length, but they didn't take qualitative factors such as reliability or travel comfort into consideration (USDOT, 2003).

The estimations for the travel time value have been made according to various qualitative factors, such as type of travel (commercial vs. personal), type of traveler (driver, adult passenger, child passenger), the transportation mode (automobile, bus, bicycle, walk), and travel condition (measured by the level-of-service). For example, Waters (1992) suggests that the valuation of in- vehicle travel time can vary from 50 to 100 percent of the wage for the best (A rating) and the worst (F rating) level-of-service respectively. He also recommends that commercial vehicle travel time should be valued at 120 percent and 170 percent of wage rate for the best and worst LOS rating respectively. Table 2.1 summarizes the empirical estimates of the value of time according to the reviewed travel time studies.

Table 2.1 Empirical estimates of the value of time. (Florida Department of Transportation, Synthesis of research on value of time and value of reliability, 2009.)

Study	Data Used	Value of time estimate
(Becker, 1965)		40% of wage rate
(Beesley, 1965)	Data from the survey of government employees in London, UK	31%-50% of wage rate
(Lisco, 1967)		20%-51% of wage rate
(Miller, 1989)	Survey of multiple route choice models	60% of gross wage (on average)
(Small, 1992)	Values derived from multiple mode choice transportation models	20% to 100% of gross wage; 50% reasonable average
(Waters, 1992)	Travel data from British Columbia, Canada	50%-100% average wage rate for personal travel, depending on LOS; 120% - 170% of average wage rate for commercial travel, depending on LOS
(Waters, 1996)	Travel data from 15 commuting studies in North America	40%-50% of after tax wage rate (mean: 59% of after tax wage rate; median: 42% of wage rate)
(Calfee and Winston, 1998)	Data from National Family Opinion survey, covering commuters from major U.S. metropolitan areas	14%-26% of gross wage; 19% of wage - average estimate
(Small and Yan, 2001)	Data on commute travelers on SR-91 in California	Average VOT is $22.87/hour, or 72% of sample wage rate
(Brownstone and Small, 2003)	Travel data from ETC facilities in HOT lanes on SR-91 and I-15 in Southern California	VOT saved on the morning commute: $20-$40 per hour, or 50%-90% of average wage rate in the sample
(USDOT, 2003)	Estimates are based on multiple sources of data	50%-120% of the wage rate depending on type of travel (personal vs. business); 50% of wage rate for personal local travel 100% of wage rate for commercial local travel
(Small, Winston and Yan, 2005)	Travel from SR-91 in greater Los Angeles area (CA), collected over 10-month period in 1999-2000	Median VOT is $21.46/hour or 93% of average wage rate
(Tseng, Ubbels and Verhoef, 2005)	Data for Dutch commuters who drive to work two or more times per week. Collected in June 2004	Mean VOT for all travelers: 10 Euros/hour (approximately $12.10/hour)
(Litman, 2007)	Results are drawn from multiple travel time studies	25%-50% of prevailing wage (for personal travel)

2.1.2 Value of Time Models

Becker (1965) used time as an input to obtain the final good while, other authors like Oort (1969) and De-Serpa (1971) included all activity duration as a direct source of utility. Evans (1972) assumed in his model that the only source of utility is the time assigned to activities. Later on, the consumption-activity model has gradually settled in the literature. Both Small (1982) and Gronau (1986) adopted this more

general approach as well. Jara-Diaz (2003) showed in his model that there is a need to establish two families of relations between goods consumption and time use. In this section, some of the most important theories will be shown and discussed.

2.1.2.1 Becker (1965)

In 1965, Becker's pioneering work was one of the first economic works that introduce the time dimension to the utility function of consumers. He emphasizes that, when decisions took by individuals about a non-work activities the value of time plays an important role. Consumption is not only the purchases of market goods; time is also needed to consume the market goods. Because of this, he defines commodities as a function of both market goods and time. The individuals are the producers of commodities.

When an individual decides to travel by bus, both purchase bus tickets and the spent time are forming the commodity. Commodity *i* can be written as:

$$Z_i = f_i\,(X_i, T_i) \tag{2.1}$$

f_i is production function needed to produce commodity *i*, X_i is a vector of market goods and T_i is a vector of time inputs.

Individuals maximize their utility function given by:

$$U = U\,(Z_1, \ldots .. Z_m) = U\,(f_1, \ldots .. f_m) = U\,(X_1, \ldots .. X_m; T_1, \ldots .. T_m) \tag{2.2}$$

The subscript m denotes the number of commodities that are being produced by the consumer. The goods constraint can be written as:

$$\sum_{1}^{M} P_i\, X_i \; = I = I_f + Ww \tag{2.3}$$

P_i is a vector giving the unit prices of X_i, I is the total income, I_f is income from other sources, *W*is a vector giving the hours spent at work and w is a vector giving the earnings per unit of *W*.

The commodity price divides into direct part which is the market goods used, and indirect part which is the time used. In Becker's model, the value of time is equal to consumer's wage rate.

$$\frac{\mu}{\lambda} = w \tag{2.4}$$

Where, μ is the marginal utility of time and λ is the marginal utility of income.

2.1.2.2 Oort (1969)

After Becker's theory appeared, Oort (1969) said that the travel time should be included in the utility. This was intuitively attractive, as an exogenous reduction in travel time would increase the leisure or work, which makes it more attractive if travel was burdensome in itself. He defines the individual's utility function as:

$$U = U(W, L, t, G) \tag{2.5}$$

Where, U is the individual's utility, W is the amount of working time, L is the amount of leisure time, tis the exogenous travel time and G is the money income.

All the variables in utility function W, L, t, and G are not independent. The total time available for individual can be written as:

$$T = W + L + t \tag{2.6}$$

Where, T is the total time available.

The money income given by

$$G = Ww \tag{2.7}$$

Where, w is the individual's wage rate.

The individual is assumed to maximize the utility function, subject to these two constraints. After this maximizing the following results are going to be obtained:

$$\frac{\mu}{\lambda} = \frac{\partial U / \partial tv}{\lambda} = w + \frac{\partial U / \partial tw}{\lambda} \tag{2.8}$$

And

$$-\frac{\partial U / \partial tr}{\lambda} = w + \frac{\partial U / \partial tw}{\lambda} - \frac{\partial U / \partial tr}{\lambda} \tag{2.9}$$

Where, $\partial U / \partial tr$ is the total effect on utility exogenous change in travel time.

Equation 2.9 shows that the value of time $\frac{\mu}{\lambda}$ equals the money value of leisure which is equal to the wage rate plus the money value of work.The value of a reduction in travel time in the Equation 3.21 is equal to the value of leisure minus the money value of travel time in utility. If both work and travel do not affect utility directly the value of a reduction in travel time would be equal to the wage rate only.

2.1.2.3 De-Serpa (1971)

By 1971, De-Serpa identified three essential features for his model: first, utility is a function not only of commodities, but also of the time allocated to them; second, the individual's decision is subjected to two resource constraints, a money constraint and a time constraint; and third, the decision to consume a specified amount of any commodity requires that some minimum amount of time to be allocated to it, but the individual may spend more time in that activity if he/she desires. According to these features three different concepts of time value appear: the value of time as a resource, the value of assigning time to an activity, and the value of saving time in a constrained activity. De-Serpa's model provides another approach to treat time and consumption goods as two separate arguments in the utility function. He defines a set of commodity bundles as:

$$X = (X_1, \ldots X_n, T_1, \ldots T_n) \tag{2.10}$$

Where, $X_i$denotes some quantity of the ith consumption good and T_i denotes the amount of time allocated to the ith good. Corresponding to this the utility function of consumer can be written as:

$$U = U(X) \tag{2.11}$$

Individuals want to maximize their utility function subject to the constraints of money income and time given by:

$$I_f = \sum_{i=1}^{n} P_i X_i \tag{2.12}$$

Where, $I_f$is the money income and Pi Xiis the amount of money spent on good i.

$$T = \sum_{i=1}^{n} T_i \tag{2.13}$$

Here T is the total amount of time available to consumers and T_i is the time allocated to the consumption of good i.

By combining all three constraints, the optimization problem of the individual can be written as this Lagrange function:

$$L = U(X_1, \ldots X_n, T_1, \ldots T_n) + \lambda (I_f - \sum_{i=1}^{n} P_i X_i) + \mu(T - \sum_{i=1}^{n} T_i) + \sum_{i=1}^{n} K_i (T_i - a_i X_i) \tag{2.14}$$

λ is strictly positive Lagrange- multiplier representing the marginal utility of money, μ is strictly positive Lagrange- multiplier representing the marginal utility of time and K_i is the Lagrange- multiplier related to each consumption activity, and can be seen as the marginal utility of saving time on activity i.

Due to the fact that this constraint may be non-binding, the value of K_i is either zero or positive. And $\mu, \lambda > 0$. The first order conditions of this problem are given by:

$$K_i (T_i - a_i X_i) = 0 \tag{2.15}$$

That is, either $T_i = a_i X_i$, or $K_i = 0,\ i = 1,\ldots,\ n$. Depending on whether the constraint binds or not for consumption activity myself.

Dividing through Equation 2.15 by λ

$$\frac{\partial U / \partial T_i}{\lambda} = \frac{\mu}{\lambda} - \frac{K_i}{\lambda} \tag{2.16}$$

$\frac{\mu}{\lambda}$ is the marginal rate of substitution between time and money. This can be interpreted as the resource value of time, which is equal to the wage rate. While, $\frac{\partial U / \partial T_i}{\lambda}$ is the marginal of substitution of *Ti* for money. This represents the value of time as a commodity. The value of time as a commodity and the value of time as a resource are equal only if an individual spends more time on that activity than the required amount. An example of this could be a pleasurable activity such as leisure. De-Serpa states that the differences between the value of time as a resource and the value of time as a commodity determine the value of saving time for that activity. The value of saving time consuming can be written as:

$$\frac{\partial U / \partial T_i}{\lambda} - \frac{\mu}{\lambda} = \frac{K_i}{\lambda} \tag{2.17}$$

During leisure activities the value of saving time is equal to zero. For travelling, the constraint in Equation 2.17 will bind since it is a burdensome activity. This will cause a discrepancy between the value of time as a resource and the value of time as a commodity.

De-Serpa's model emphasize that the value of time will be different for each consumption activity. This is the contrast with Becker's model (1965) that simply equaled the value of time to the foregone wages for each consumption activity.

2.1.2.4 Small (1982)

Small builds his model according to Becker and Johnson time allocation models. He included scheduling consideration to both utility function and constraints. The time divided in three types in his model: leisure time, work time, and consumption time. The time when the trip has begun or the arrival time at work, in work trips has been entered which influences utility, travel time, and travel cost. The utility function can be written as:

$$U = U(G, L, W, s) \tag{2.18}$$

Where, *U* is the individual's utility, *G* is the money income, *L* is the leisure time, *W* is the work time and s is the schedule time (a specific time of the day).
Subjected to

$$G + c(s) = I_f + wW \tag{2.19}$$

$$L + t(s) = T - W \tag{2.20}$$

$$F(s, W; w) = 0 \tag{2.21}$$

Where, *c* is the travel cost, I_f is the individual's fixed income, *w* is the wage rate, t is the exogenous travel time, *T* is the total time available and *F* is the function that accounts for the limitations imposed by the institutional setting within which employment opportunities are encountered. The Lagrangian function for these constraints can be written as:

$$L = U + \lambda[I_f + Ww - G - c(s)] + \mu[T - L - t(s) - W] - \nu F(s, W; w) \tag{2.22}$$

Where, λ is the Lagrange multiplier of income restriction, μ is the Lagrange multiplier of time restriction and v is the Lagrange multiplier of schedule restriction. By maximizing over *G, L,* and *W* (for fixed s) yields to this time allocation model

$$\frac{\mu}{\lambda} = w + \frac{\partial F / \partial W}{\lambda} - \nu \frac{\partial F / \partial W}{\lambda} \qquad (2.23)$$

According to this time allocation model which links working hours, schedule time, and the wage rate. The value of time would be affected by the work schedule.

2.1.2.5 Jara – Diaz (2003)

Jara- Diaz started by establishing two technical relations between good consumption and time used. He combined De-Serpa's (1971) with Evans' (1972) constraints to build these to general functions: A (X, T) ≥ 0, the activity possibility function, and G (X, T) ≥ 0, the consumption possibility function, where equality defined the activity possibility frontier for a given X in the former, and the consumption possibility frontier for a given T in the latter. These technical relations can be written as:

$$T_i \geq f_i (X) \qquad (2.24)$$

$$X_i \geq g_i (T) \qquad (2.25)$$

Where, $f_i(X)$ is the minimum time for activity *i* as a function of goods and $g_i(T)$ is the minimum consumption of good i as a function of activities.

Equation 2.24 shows that goods consumption impose minimum levels on activity duration, where Equation 2.25 shows that activities impose minimum levels on goods consumption. These relations can be included as a restriction in a general framework for consumer behavior that includes time, which generate an additional term in the willingness to bay to reduce in constrained activities. Individual's utility model can be written as:

$$U = U (X, T) \qquad (2.26)$$

Where, *U* is the utility function, *X* is the consumption of good *I* and *T* is the time assigned for activity *i*. Subject to

$$w\,T_w - \sum P_i\,X_i \geq 0 \quad (2.27)$$

$$T - \sum T_i = 0 \quad (2.28)$$

$$T_i - f_i(X) \geq 0 \quad (2.29)$$

$$X_i - g_i(T) \quad (2.30)$$

Where, *w* is the wage rate, $T_w$is the time assigned to work and *T* is the total time available. The first order condition for an activity *k* given by:

$$\frac{\partial U}{\partial T_i} - \mu + k_i - \sum_{i=1}^{g} \psi_i \frac{\partial g_i}{\partial T_i} = 0 \quad (2.31)$$

From Equation 2.31 a new expression can be obtained for the value of time $\frac{\mu}{\lambda}$.

$$\frac{ki}{\lambda} = \frac{\mu}{\lambda} - \frac{\partial U / \partial Ti}{\lambda} + \frac{1}{\lambda}\sum_{i=1}^{g} \psi_i \frac{\partial gi}{\partial Ti} \quad (2.32)$$

Where, λ is the Lagrange multiplier of income restriction, μ is the Lagrange multiplier of time restriction, ψ is the Lagrange multiplier of the minimum consumption requirement of good i.

The value of time as a resource $\frac{\mu}{\lambda}$ in Equation 2.32 is affected by the value of time assigned for an activity $i\frac{\partial U / \partial Ti}{\lambda}$, the value of saving time in an activity $i\frac{ki}{\lambda}$, wage rate w, and it is also affected by the new term, the value of total consumption saved when activity i is reduced $\frac{1}{\lambda}\sum_{i=1}^{g} \psi_i \frac{\partial gi}{\partial Ti}$.

Becker (1965) used time as an input to obtain the final good while, other authors like Oort (1969) and De-Serpa (1971) included all activity duration as a direct source

of utility. Evans (1972) assumed in his model that the only source of utility is the time assigned to activities. Later on, the consumption-activity model has gradually settled in the literature. Both Small (1982) and Gronau (1986) adopted this more general approach as well. Jara-Diaz (2003) showed in his model that there is a need to establish two families of relations between goods consumption and time use. In this section, some of the most important theories will be shown and discussed.

There are other microeconomic models dealing with time allocation and the value of time, like De Donnea (1971), Pollack and Wachter (1975), Michael and Becker (1973), or Dalvi (1978). Table 2.2 summarizes what consider the main contributions to the analysis of the value of time. And Table 2.3 presents the definitions of the shortcuts that used in the time allocation models. For this study Becker's (1965) model has been used to predict the value of time. This model will be explained more clearly in the modeling structure chapter (chapter three).

Table 2.2 Value of time from the main time allocation approaches

Author	**Model**	**Value of time**
Becker (1965)	$U = U(Z_1, \ldots .. Z_m)$ $\sum_{1}^{M} P_i X_i = I = I_f + Ww$ $T_c = T - W$ $Z_i = f_i (X_i, T_i)$	$\frac{\mu}{\lambda} = w$
Oort (1969)	$U = U (W, L, t, G)$ $T = W + L + t$ $G = W w$	$\frac{\mu}{\lambda} = \frac{\partial U / \partial tv}{\lambda}$
De-Serpa (1971)	$U = (X_1, \ldots . X_n, T_1, \ldots . T_n)$ $I_f = \sum_{i=1}^{n} P_i X_i$ $T = \sum_{i=1}^{n} T_i$	$\frac{\partial U / \partial T_i}{\lambda} - \frac{\mu}{\lambda} = \frac{K_i}{\lambda}$
Small (1982)	$U = U (G, L, W, s)$ $G + c(s) = I_f + wW$ $L + t (s) = T - W$ $F (s, W; w) = 0$	$\frac{\mu}{\lambda} = w + \frac{\partial F / \partial W}{\lambda} - \nu \frac{\partial F / \partial W}{\lambda}$
Jara- Diaz (2003)	$U = U (X, T)$ $w T_w - \sum P_i X_i \geq 0$ $T - \sum T_i = 0$ $T_i \geq f_i (X)$ $X_i \geq g_i(T)$	$\frac{ki}{\lambda} = \frac{\mu}{\lambda} - \frac{\partial U / \partial Ti}{\lambda} + \frac{1}{\lambda} \sum_{i=1}^{g} \psi_i \frac{\partial gi}{\partial Ti}$

Table 2.3 The definitions shortcuts that used in the time allocation models

Shortcuts	Definitions
U	The individual's utility function.
f_i	a function "production function" needed to produce commodity i.
X_i	A vector of market goods.
T_i	A vector of time inputs.
I	The total income.
W	The amount of working time.
w	The wage rate.
T_c	A vector giving the total time spent on consumption.
T	The total time available.
λ	The Lagrange multiplier of income restriction.
μ	The Lagrange multiplier of time restriction.
L	The amount of leisure time.
t	The exogenous travel time.
G	The money income.
I_f	The money income.
P_i	the price of good i
X	the consumption of good i
a_i	may be interpreted as technologically or institutionally determined minimum amount of time required to consume one unit of Xi
K_i	the Lagrange- multiplier related to each consumption activity, and can be seen as the marginal utility of saving time on activity i.
s	The schedule time "a specific time of the day".
c	The travel cost.
I_f	The individual's fixed income.
F	The function that accounts for the limitations imposed by the institutional setting within which employment opportunities are encountered.
v	The Lagrange multiplier of schedule restriction.
T_i	the time allocated to the consumption of good i.
T_w	The time assigned to work.
Ψ	the Lagrange multiplier of the minimum consumption requirement of good i.

2.2 Discrete Choice Analysis

This section has focused on the logit model and the logistic regression model. A vast amount of value of time studies has been conducted using these two types of discrete choice models. In this thesis the logistic regression has used to find the value of time by using IBM SPSS Statistics software. This way and specially this program give good opportunity to arrange the dataset and to preview the results in a good way with an accurate calculation.

2.2.1 The Behavioral Models

The derivation of discrete choice models based on the maximize of behavior utility by the decision maker. In 1927 Thurstone was one of the first authors who developed the concepts in terms of psychological stimuli. Later on, Marschak (1960) interpreted the stimuli as utility and provided a derivation from utility maximization. According to this assumption he derived the random utility models.

The key aspect of travel demand analysis is modeling travel demand behavior, where aggregate demand is the accumulation of individuals' decision. Travel decisions include different choices, such as choice of destination for a non-work trip, choice of departure time, choice of traffic route and also choice of travel mode. Travel behavior analysis is disaggregated because the models represent the choice behavior of individual travels (Ben-Akiva & Lerman, 1985).

Discrete choice models are the methodology used to understand the behavior process that leads the individual to choose from a choice set. Choice set includes all alternatives that individuals can choose from. Alternatives have to be exclusive; meaning that individuals can choose only one alternative, also alternatives must be exhaustive. Finally, the number of alternatives must be finite (Train, 2009). An individual's choice can be described as a function of both observable and unobservable factors. By following Train (2009)notation, this can be written as:

$$y = h(x, \varepsilon) \tag{2.33}$$

$x$are the observable factors and ε is the error term that accounts for the unobserved factors. The unobserved factors $\varepsilon$are known to the individual, but unknown to the researcher. This means that choices taken by an individual cannot be predicted exactly. Instead, the probabilities of individual alternatives in a choice set can be calculated. The probability that an individual chooses a specific alternative is the probability that the unobserved factors are such that the behavioral process results in that outcome (Train, 2009). This can be written as:

$$P(y \mid x) = \Pr ob(\varepsilon s.t. h(x, \varepsilon) = y) \tag{2.34}$$

ε is the error term that accounts for the unobserved factors, There are different ways of specifying h (x, ε) giving rise to different discrete choice models. Crucial here is the assumption about the error term ε. Assuming that ε is distributed logistically over the population yields the logit model. If ε is assumed to be normally distributed this results in the probit model.

The representative utility function of each available alternative for individual is allocated to calculate choice probability. The representative utility includes only the observable part of utility, this can be written as:

$$V_{ni} = \beta x_n \tag{2.35}$$

β is a vector of parameters and x_n is the corresponding vector of individual-specific variables. These variables could be the cost and other characteristics of the alternative and the socio-economic attributes of each individual (Jara-Diaz, 2000). The actual utility for an individual n choosing alternative i, is given by:

$$U_{ni} = V_{ni} + \varepsilon_{ni} \tag{2.36}$$

Equation 2.36 shows that the actual utility is a combination of both representative utilityV_{ni} and unobservable error termε_{ni}.Because$\varepsilon_{ni}$is unknown to the researcher, it is

only possible to make probability statements about what alternative is going to be chosen by an individual. The error term can be seen as the unmeasured characteristics for an individual relating to an alternative.

Train (2009) shows that the absolute level of utility is irrelevant to both decision maker's behavior and the researcher's model. Also, he stated that scaling utility of different alternatives by a constant does not chance the choice probabilities. The probability of individual n choosing alternative i can be written as:

$$P_{ni} = \Pr ob(U_{ni} > U_{nj} \forall j \neq i) \tag{2.37}$$

$$P_{ni} = \Pr ob(\varepsilon_{ni} - \varepsilon_{nj} < V_{ni} - V_{nj} \forall j \neq i) \tag{2.38}$$

Equation 2.38 shows that the probability of an individual does not depend only on the observed facts, but also it depends on the size of the error terms associated with each individual foe each alternative.

2.2.2 The Logit Model

Logit formula is the most popular and easiest formula to use in discrete choice models. Because the discrete choice probabilities take the closed form and is readily interpretable (Train, 2009). Lace (1959) derived his formula according to choice probability characteristics "the independence from irrelevant alternatives (IIA)". Marshak (1960) said that the model is harmonious with the maximization of utility. Luce and Suppes (1965) cite that Marly established the relation between the Logit formula and the distribution of unobserved utility. McFadden (1974)showed that unobserved utility is distributed extreme value in a light formula for the choice probabilities. In other words, the logit formula for choice probabilities imply that the error terms $\varepsilon_{nj}$for different alternatives are independent identically distributed extreme value. Meaning that the value of other error terms cannot be predicted by knowing the value of other error terms.

If the representative utility function has been specified in a suitable way and the unobserved parts of utility are truly independent, then the logit model is compatible to use. The density of the unobserved part of utility can be written as

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}} \tag{2.39}$$

For the unobserved part of utility the largest model makes it possible to express choice probabilities in a closed form, like the formula shown below

$$P_n(i) = \frac{1}{1+e^{-\mu\beta(x_{in}-x_{jn})}} = \frac{e^{-\mu\beta x_{in}}}{e^{-\mu\beta x_{in}} + e^{-\mu\beta x_{jn}}} \tag{2.40}$$

As Equation 2.39 states, V_{ni} is the representative utility for individual n choosing alternative i. For one individual the sum of all choice probabilities for all alternatives is equal 1.

In the logit model the choice probabilities are functions of the explanatory variable x, and the corresponding parameter vector β. To estimate these parameters the maximum likelihood estimation should be applied (Train, 2009).

2.2.2.1 Maximum Likelihood Estimation

First of all the probability of individual n choosing the alternative was observed to choose should be found. If the parameter y_{ni} indicates whether the alternative was chosen or not, this can be written in an equivalent way as shown below

$$\prod_{i=1}^{j} (P_{ni})^{y_{ni}} \tag{2.41}$$

For the chose alternative$y_{ni}$is 1 and$y_{ni}$is 0 for all other alternatives. Equation 2.41 shows that it is possible to make a special expression for the probability of every individual in the sample choosing case when all alternatives were observed to choose. This is called the likelihood function (β), and is calculated by multiplying choice probabilities for the observed choices over all individuals.

$$L(\beta) = \prod_{n=1}^{N} \prod_{i=1}^{j} (P_{ni})^{yni} \tag{2.42}$$

β are the maximum likelihood estimators that maximize this function. Biorn (2010) states that maximizing the likelihood function is equivalent to maximizing the logarithm of the likelihood function. By taking the logarithm of this expression it yields to the log- likelihood function, denoted as (β).

$$LL(\beta) = \sum_{N=1}^{N} \sum_{i=1}^{J} y_{ni} \ln P_{ni} \tag{2.43}$$

According to Train's (2009) study, the log-likelihood function is globally concave. Thus, it becomes easier to find the vector of parameters β that maximizes the log-likelihood function by finding the first order condition for each parameter.

$$\frac{\partial LL(\beta)}{\partial \beta} = 0 \tag{2.44}$$

Train (2009) shows how this can be written as

$$\frac{\partial LL(\beta)}{\partial \beta} = \sum_{N=1}^{N} \sum_{i=1}^{J} (y_{ni} - P_{ni}) x_{ni} = 0 \tag{2.45}$$

It is possible to manipulate this expression for easier interpretation.

$$\frac{1}{N} \sum_{N=1}^{N} \sum_{i=1}^{J} y_{ni} x_{ni} = \frac{1}{N} \sum_{N=1}^{N} \sum_{i=1}^{J} P_{ni} x_{ni} \tag{2.46}$$

In this equation the left hand side represents the observed average value of the explanatory variable vector x of the sampled decision makers. While the right hand side represents the average of x for the predicted choices of the sampled decision makers Therefore, the maximum likelihood estimates for$\beta$are those that make the predicted average of each explanatory variable equal to the observed average in the sample (Train, 2009).

2.2.2.2 Limitation of Logit Model

Train (2009) describes three topics that can cause problems for the estimation of the logit model. These topics are: taste variation, substitution patterns, and repeated choices over time. These topics can be summarized as follows:

1- The logit model can represent taste variation for observed characteristics of an individual, but cannot represent taste variation for the unobserved characteristics of an individual.

2- The logit model implies proportional substitution across alternatives, given the researcher's specification of representative utility. To capture more flexible forms of substitution, other models are needed.

3- If unobserved factors are independent over time in repeated choice situations, then logit can capture the dynamics of repeated choice, including state dependence. However, logit cannot handle situations where unobserved factors are correlated over time.

According to these limitations biased estimates occurs when the logit model is been used.To avoid this shortage there a need to shift from using the logit model to using the more advanced mixed logit model.

2.2.3 Logistic Regression

Logistic regression, also called a logit model, is a type of probabilistic statistical classification model (Bishop, 2006). It is also used to analyze a dataset in which there are one or more independent variables that define an outcome. The outcome is measured with a binary variable. That is why; logistic regression is used in estimating the parameters of a qualitative response model. The probabilities describing the possible outcomes of a single trial are modeled, as a function of the explanatory variables, using a logistic function. Logistic regression is used to refer

specifically to the problem in which the dependent variable is binary, which it means that the number of available categories is two, while problems with more than two categories are referred to as multinomial logistic regression. Or, if the multiple categories are ordered, as ordered logistic regression.

The relationship between a categorical dependent variable and one or more independent variables is measured by logistic regression; by using probability scores as the predicted values of the dependent variable, these variables are usually continuous. Logistic regression can be seen in many ways to be similar to ordinary regression, and allows us to look at the fit of the model as well as at the significance of the relationships between dependent and independent variables that we are modeling.

Logistic regression usually comes in two cases binomial or multinomial. Binomial or binary logistic regression occurs when the observed outcome for a dependent variable can have only two possible types (for example, dead vs. alive, male vs. female or as in our study car vs. public transit). Multinomial logistic regression deals with situations where the outcome can have three or more possible types. In binary logistic regression, the outcome is usually coded as 0 or 1, as this leads to the most straightforward interpretation (Hosmer et al., 2000). Code 1 means that the observed outcome for the dependent variable is noteworthy possible outcome (such as a success or a case), while code 0 is the contrary outcome (such as a failure or a non-case).Logistic regression is used to predict the odds of being a case based on the values of the independent variables (predictors). The odds are defined as the probability that a particular outcome is a case divided by the probability that it is a non-case.

In this thesis, logistic regression has been used to predict the value of time because it gives an opportunity to analyze the relationships between a dependent variable and independent variables. (SPSS Statistics supports Multinomial Logistic Regression that can be used with more than two groups, but the focus in this study is

on binary logistic regression for two groups). The logistic regression computational method can be explained in these steps:

- Multiple regression uses the least-squares method to find the coefficients for the independent variables in the regression equation, i.e. it computed coefficients that minimized the residuals for all cases.

- Logistic regression uses maximum-likelihood estimation to compute the coefficients for the logistic regression equation. This method finds attempts to find coefficients that match the breakdown of cases on the dependent variable.

- The overall measure of how will the model fits is given by the likelihood value, which is similar to the residual or error sum of squares value for multiple regression. A model that fits the data well will have a small likelihood value. A perfect model would have a likelihood value of zero.

2.3 Definition of Reliability

Reliability of public transport systems is critically important for the most of public transport users because passengers are adversely affected by the consequences associated with unreliability, such as additional waiting time, late or early arrival at destinations and missed connections, which increases their anxiety and discomfort (Bates et al., 2001). Reliability has also been identified, as the ability of a public transport system to adhere the specified schedule, headway and travel time (Chen et al., 2003). Turnquist and Bowman (1981) have noted that the mod choice is affected by reliability. A more comprehensive definition was given by Chapman (1976) covering all aspects of the service as “safety, punctuality, regularity, cleanliness, passenger comfort, documents, security”. Therefore, it may be stated that unreliability in public transport drives away existing and prospective passengers.

Reliability is critically important for passengers and operators alike. For passengers, unreliable services lead individuals to make adjustments in their trips to avoid problems resulting from a service not operating normally. Arrival time variability forces traveler who uses public transport to arrive at their destination late, to avoid this lateness, the traveler has to take the earlier service. This situation sometimes causes the travelers to arrive too early at their destination, hence they have to wait around or make up time. For operators, many difficulties in timetabling and resource planning are caused because of unreliable services. Also, unreliable services are typically more unevenly loaded, causing issues of passenger overloading. Thus, the concept of reliability is of significance not only for public transport users, but also for the service operators and the regulatory regime such as the local authorities.

Public transport has sources of uncertainty, while using a car contributes to reducing the level of control users have over their own situation caused primarily by the reduced flexibility of public transport; car users can manage their trip time to the minute, whereas a public transport users have to follow an existing timetable.

For a public transport user, the journey consists mainly of:

- In vehicle time (IVT)
- Headway/ Waiting time
- Walk time
- The fare paid

In vehicle time (IVT) represents travel time spent in the vehicle and access/ egress to the vehicle. Headway covers the interval between public transport services and is a measure of services frequency. Waiting time occurs either prior to the arrival of the first vehicle or during the course of the journey when the interchange is required. Walk time represents time spent walking to and from the public transport mode.

The variations of in-vehicle travel time are usually caused by either infrastructure or vehicle failure. While, waiting time variation is caused primarily by a previous in-vehicle time variation, but can also be caused by service cancellation. For an

individual using public transport, if a service is running early, he/she faces the real possibility that he/she may miss it given their arrival time at the station. In this case, an individual should wait for the next service, thus increasing their wait time substantially.

2.3.1 Factors Affecting Reliability

There are many different factors that give a share in public transport unreliability: Such as day-to-day and within-day variability in traffic flow and congestion levels cause delays and make predictions of public transport journey time uncertain, excess passenger demand causes longer dwell time at public transport stops; public transport do not always run on schedule, or appear in bunches; operators have insufficient spare capacity to cover for service breakdowns (Sterman & Schofer, 1976; Abkowitz & Tozzi, 1987). Generally, these factors can be classified into the four following groups:

Passenger characteristics: These include variability in passenger volume, passenger volume at stops, passenger route choice, passenger arrival distributions. Turnquist and Bowman (1981) have noted that inter-relationship of factors such as passenger demand, link travel time variability and the service frequency may cause public transport to bunch on a route, thus adversely affecting the service reliability.

Traffic characteristics: These include day-to-day and within-day variation in travel demand, traffic composition, and traffic congestion levels.

Route characteristics: These include number of lanes in each route, length of the route, location of the bus stops (on-road or bus-bays), provision of bus lanes (whether in short stretches or all along the route), number of intersections (signalized and priority), priorities at junctions for buses (sensor/detector system), and on-street parking.

Public transport operational characteristics: Such as scheduling system, staff shortages, fleet availability, fleet maintenance, ticketing system, and variability in

driver behavior and experience. Prashker (1978) has identified mechanical reliability as one of the three main components of reliability; the other two are parking and travel-time reliability.

2.3.2 Reliability Indicators

In the absence of one unique measure to assure all reliability aspects of service, distinctive measures are called for. Based on this, an exigency is created to secure a series of indicators derived from the demands of the various collaborators and the performing status. The appropriate and valid process is to compile an overall group of indicators (Pullen, 1993). In this section, six categories indicators related to different service attributes are described in detail.

2.3.2.1 Travel Time

In consonance with Kaparias et al. (2008), most travel time reliability marks adopt different features of the travel time allocation. Lomax et al. (2003) classified them in three groups as follow: statistical range measures, buffer measures and tardy trip indicators. When dealing with people's perceptions, it appears to be more appealing to separate physical from psychometric performance indicators (Pronello & Camusso, 2012). For travel time reliability, physical indicators depict it as 'it is what it is), while psychometric indicators say it is more (what it is perceived to be'. The following debates physical achievement indicators.

Statistical Range Indicators: This type of measure typically serves as an approximate estimate of the range of trip situations experienced by passengers, calculated on standard deviation statistics. Standard deviation of travel time represents reliability in such way that small values are considered reliable. Percent variation of travel time, statistically known as the coefficient of variation, provides a clearer picture of the trends and performance characteristics than the standard deviation by eliminating route length from the calculation. Moreover, percent variation is dimensionless enabling a comparison between links and routes. The

travel time window is defined as the average travel time plus or minus the standard deviation of travel time, and can provide the passenger with an idea of how much the travel time will vary (Lomax et al., 2003). The variability index is defined as a ratio of peak to off-peak variation in travel conditions, and is calculated as a ratio of the difference in the upper 95% and lower 95% confidence intervals between the peak period and the off-peak period.

Tardy Trip Indicators: Tardy trip measures are extreme values of travel time. The tardy trips are identified by setting unacceptable limit values in the form of additional minutes plus expected time or percentage over expectation. In most cases, these values are arbitrarily set. The Florida reliability measure (FRM) uses a percentage of the average travel time in the peak to estimate the limit of the tolerable travel time range. Travel time exceeding the expectations is termed a tardy trip (Shaw & McLeod, 1998). Extended FRM uses travel rate (travel time per unit distance) instead of travel time, so as to provide a length-neutral way of grading the service performance (Lomax et al., 2003). The misery index examines trip reliability by using the difference between the average travel rates of the worst trips and all trips.

Skew-Width Indicators: Skew and width of travel time distribution measures are based on percentiles (van Lint & van Zuylen, 2005). Skew of travel time distribution is defined as the ratio of the difference between the 90th and 50th percentile and the difference between the 50th and 10th percentile. Width of travel time distribution indicates the distribution compactness. The wider the distribution is, the lower the reliability will be.

2.3.2.2 On-Time Performance

Schedule adherence plays a significant role, for the low frequency services routes, considering the fact that passengers are awaited to schedule their arrivals to coordinate with the scheduled departures to minimize waiting time at stops with a tolerance probability of missing the trips. On time performance is a frequently adopted schedule adherence measure in applied environments, defined as the

percentage of trips that depart up to “m” minutes late and “n” minutes early from the scheduled departure time. The US Transportation Research Board (TRB) presented a service delivery measure survey where zero minutes was the most common earliness threshold and 5 minutes was the most common lateness threshold (Kittelson & Associates et al., 2003).

2.3.2.3 Headway Regularity

In the case of high-frequency service routes, headway related measures gain more importance (Currie et al., 2012). In this case, passengers are more likely to be on the stops disconnectedly and the combined waiting time is shrinked when passengers are uniformly dispersed (Osuna & Newell, 1972). Various indicators are suggested for this situation. Whereas some are determined by comparison with scheduled headway, like service regularity, headway ratio and percentage regularity deviation mean (van Oort & van Nes, 2004), others are determined by headway distribution like for instance standard deviation, coefficient of variance, average waiting time as well as probability-based headway regularity measure (Lin & Ruan, 2009). Also, two indicators are established for distinguished aims. The headway regularity indicator is an identificator of the vehicle bunching problem (Henderson et al., 1991) whereas the irregularity indicator is able to validly point out long differences between vehicles (Golshani, 1983).

On-time performance and headway regularity are indicators derived from schedules. The issue is that an universal benchmarking verge marking the difference between frequent and infrequent services and determining the on-time tolerance interval is nonexistent. Additionally, the reflection of demand-side approach of reliability is not possible. By switching the on time tolerance gap from 5 minutes to 10 minutes, the calculated service conduct rises without drawing the attention of passengers to it. On-time distribution indicator endorsed by Nakanishi (1997) is convenient for the client to obtain a meticulous understanding of the work of specific routes, as well as catering operators with precise causes of defects.

2.3.2.4 Waiting Time

Waiting time at stop is, from the passenger's point of view, the most important constituent of public transit travel and generally mentioned as one of the most significant elements obstructing the use of bus transit. Customarily, waiting time indicators are sorted into two groups, referred to as: mean-variance based and extreme-value based (van Oort & van Nes, 2004).

Mean-Variance based: Excess waiting time (EWT) is explained as the difference between the average waiting time (AWT) and the scheduled waiting time (SWT) (Trompet et al., 2011). For usual services, the SWT is explained as the ordinary time passengers are prompt to wait in the case of an strictly according to plan operating service (Liu & Sinha, 2007). For high frequency utilities, a generally used AWT index is half the progression of consecutive buses, derived from three suppositions: passengers show up at random, passengers catch the first coming bus, and vehicles arrive regularly (Fan & Machemehl, 2009).

Extreme-Value based: passengers tend to be more anxious about acute values in their judgment of service achievement when budgeting their arrival at stops. Budget waiting time is determined as 95th percentile waiting time for regular utilities. It serves as the total waiting time that a traveller is supposed to allocate for a trip to prevent missing awaited services at a stop under certain circumstances. Probable waiting time, explained as the difference between budgeted waiting time and mean waiting time, plays the buffer time that a passenger is expected to schedule for their show up at stops (Furth & Muller, 2006). The notion of extreme-value derived indexes distinguishes the impact on operations from the impact on traveller scheduling. Extreme-value based waiting time depends more on service reliability than mean-variance based AWT does.

2.3.2.5 Transfer Time

Scheduled stops are a way of measuring transfer time (Jang, 2010). Accordingly, the application of data indexes for calculating transfer time reliability; coefficient of variation of transfer delays for instance; is possible (Turnquist & Browman, 1980). Nevertheless, day-to-day influx time fluctuations tend to complicate the analysis (Kittelson & Associates et al., 2003). Transfer waiting time is commonly used as a translation time reliability index. Goverde (1999) derived an expected transfer waiting time model, a function of arrival delays distribution, incorporating the risk and significance of missing connections.

2.3.2.6 Buffer Time

The buffer time points out extra travel time demanded to assure the passengers' on time appearance. Usually, it is explained as the difference between percentile and the ordinary travel time. The scheduling time is determined as the percentile travel time. It shows the over-all time that a rider must allocate for the trip. Buffer time index is explained as the buffer time divided by the ordinary travel time. These indicators relate directly with how travellers decide to go on a trip (Lomax et al., 2003). Uniman (2009) suggested the general type of an basic set of reliability buffer time calculations under the 'percentile-based' and 'slack time' procedure. Reliability buffer time, determined as the difference between the maximum percentile xx, and a midway or minimum percentile yy, is the extra time possibly demanded to be xx-percent positive of being at the terminal on time. Excess Reliability Buffer Time (ERBT) is explained as the difference between the confirmed stages of reliability passengers experience and what they should have experienced had everything gone as scheduled. The ERBT indicator can be used to seize the incident-caused extra unreliability above deriving from recurring elements.

2.3.3 Measures of Reliability

In this section, three types of bus reliability measures are discussed. These are: travel time reliability, headway reliability and passenger wait time reliability. These definitions are not the exclusive definitions to the type of reliability measures. For example, the travel time reliability indicator defined in this study as the mean over standard deviation in journey times of a particular route within a particular time period. In the literature, travel time reliability has also been defined as the inverse of the standard deviation of journey times (e.g. Polus, 1978; Sterman & Schofer, 1976), the percentage of travel on a bus route that takes no longer than the expected travel time plus an acceptable additional time (Strathman et al., 2001), and the percentage of time that a bus's journey time is no more than 10% higher than average (Meyer, 2002).

2.3.3.1 Travel Time Reliability

Travel time reliability can be defined as the variability in bus journey time for a specific bus routes within a specific time interval and at a specific level of service or probability that a trip can be completed within a given time. Travel time reliability is defined as the mean over standard deviation of travel time. The inclusion of the mean in addition to the standard deviation as in Polus (1978) is to allow for comparison of the variations in bus routes or sections of a route that have significantly different mean values. This definition can be written as:

$$RT_i = \frac{\mu_{ti}}{\sigma_{ti}} \tag{2.47}$$

Where, μ mean of the sample x, σ the standard deviation of the sample x and t_{inm} journey time of bus trip n on day (or simulation run) m along route i

With this definition, a bigger value of RT indicates a better reliability. In this study, the values of reliability for all districts have been measured according to this formula.

2.3.3.2 Headway Reliability

Maintaining the scheduled headways and providing regularity of buses helps reduce passenger discomfort by minimizing the average passenger wait time. The bunching of buses results in absence of service during the scheduled time which then results in peaks in passenger demand for subsequent bus. Such peaks in passenger demand may result in 6 buses getting full to its capacity and having to bypass subsequent stops due to inadequate capacity.

A number of different reliability measures have been proposed in the literature (e.g. Chapman, 1976; Turnquist & Browman, 1980; Oort & Nes, 2003; Yin et al., 2003). In this study, the definition of Bowman et al. (1991) has been adopted and defines headway reliability as the standard deviation over mean headway:

$$RH_i = \frac{\sigma_{hi}}{\mu_{hi}} \tag{2.48}$$

Where, h_{inms} headway between bus trip n and n+1 on day m at bus stop s along route i.

Different to the definition of RT, smaller values of RH indicate better headway reliability.

2.3.3.3 Passenger Wait Time Reliability

Passengers value the waiting time for urban transit almost twice the in-vehicle time (Oort & Nes, 2003). High waiting times in a public transport network is highly likely to affect the patronage and is therefore an important factor to focus on.

The concept that bus service reliability should be measured from a passenger's point of view has been taken on by Transport for London who uses an excess wait time (EWT) measure as a key indicator of reliability for high frequency services (e.g.

a service frequency at five or more buses per hour) (TfL, 2006). The EWT assessment includes calculation of the following two elements:

• Average scheduled wait time (SWT) – the time passengers would wait, on average, if the service ran exactly as schedule; and

• Actual wait time (AWT) – the average time that passengers actually waited.

Then EWT is the difference between AWT and SWT and represents the additional waiting time experienced by passengers. For frequent service, SWT can be taken as half of the service frequency and AWT the accumulated passenger wait time at the bus stop. In this study, we adopt EWT as the passenger wait time reliability measure and calculate it for each bus route and at each bus stop:

$$RW_{is} = EWT_{is} = AWT_{is} - SWT_{is} \tag{2.49}$$

Where, w_{ims} total passenger wait time at stop s for service i on day m. It is clear that small values of RW indicate better passenger wait time reliability.

2.4 Public Transport Smart Card

The smart card system is involved in intelligent public transport systems by both operators and users as it provides great advantages in widespread use every day (Trepanier, et al., 2007). In recent years, benefits such as decreased travel times and general convenience to passengers have driven a shift towards smartcard ticketing systems on public transport systems in Izmir and around the world. As well as providing more efficient transport services to passengers, smartcard ticketing systems enable service operators and transit authorities to collect comprehensive data on the travel behavior of passengers. With this information at hand service operators are able to achieve passengers' needs, and allocate resources more efficiently. Figure 2.1 shows how the smart card system gives the help of collecting the data and transmitting it to the main centers to evaluate it by different types of software.

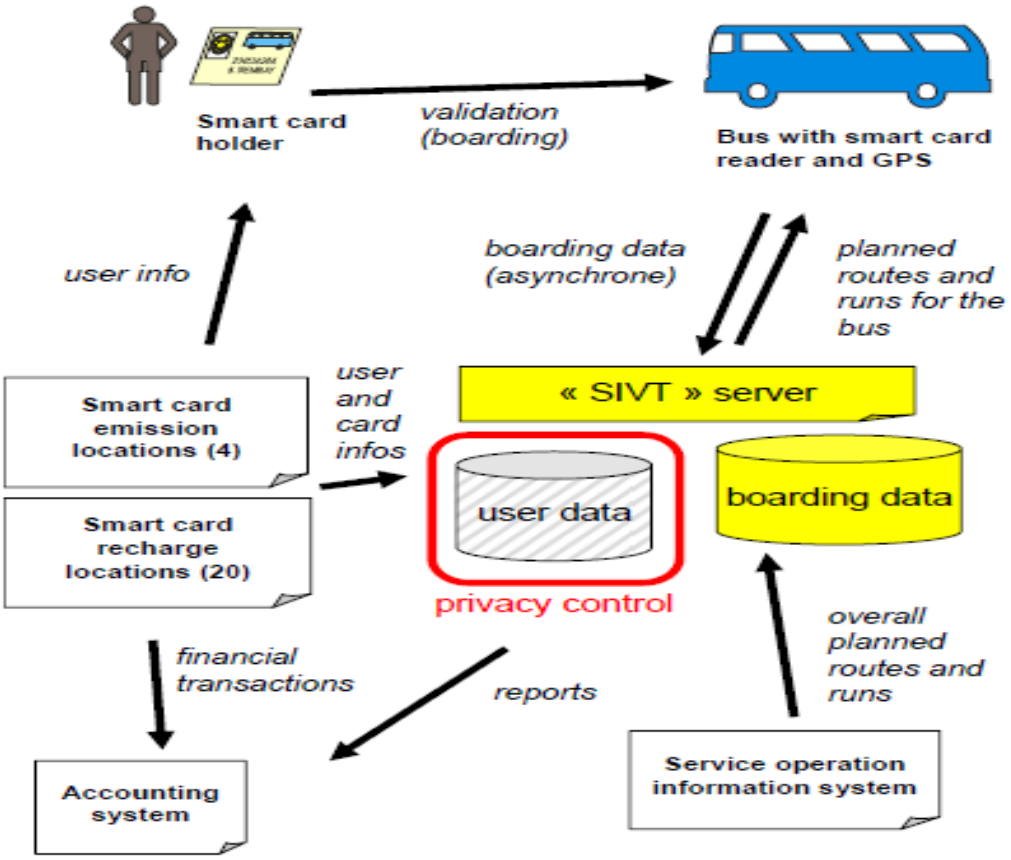


Figure 2.1 A typical smart card system flowchart. (Bruno, A., & Catherine, M. (2007).Mining public transport user behavior from smart card data).

When the passenger boards the public transport service, the smart card's fare gets validated. The validation is done following those steps: the public transport service system contains the planned runs for the day. The Global Positioning System (GPS) reader on the public transport service identifies the stop where the boarding is made. The system validates the run (correct route) at this location. Card ID, service ID, driver ID, stop ID, route direction, route number, cost type (full, student, etc.), transfer case and boarding information such as time and date are recorded by the system.

Figure 2.2 presents the simplified data model that stores boarding data. The most important one is boarding, which contains a record for each smart card validation on the network. There is also a link between boarding and the public transport network geometry (routes and route-stops). This link is to ensure the validity of the time and

location of boarding. Boarding refusals and card recharges are also included in this model.

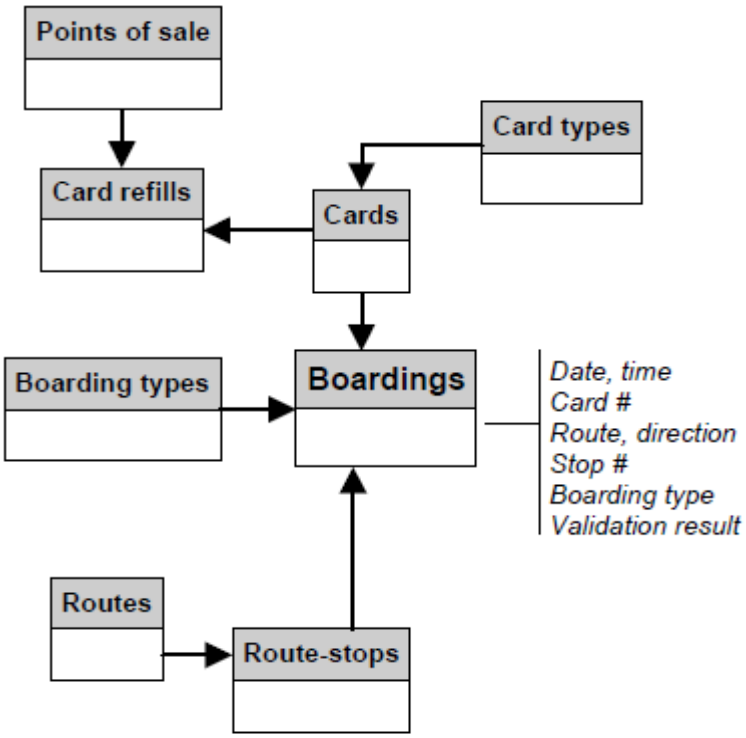


Figure 2.2 Simplified database model for boarding data storage. (Bruno, A., & Catherine, M.(2007).Mining public transport user behavior from smart card data).

2.4.1 Smart Card Data Use in Public Transport Planning

Several studies have been conducted during recent years on the use of smart card data for public transport planning purposes. These studies have been grouped in this section into three categories: Strategic-level studies, tactical-level and operational-level.

2.4.1.1 Strategic-Level Studies

These studies are related to long-term network planning, customer behavior analysis and demand forecasting. Table 2.4 presents various work related to long-term planning.

Table 2.4 Review of studies on smart card systems for strategic public transport planning. (Marie, A., & Catherine, M. (2009), Smartcard data in public transit planning: A review).

Author(s)	Data	Analysis/use	Benefits
Agard et al. (2006)	Boarding date, time, and location. Card type.	Define typical user type andmeasure their trip habits. Analyze use variability according to the day, the week, or season.	Better understand user behavior.
Bagchi& White (2005)	Time, space, and structure. Personal and travel data.	Turnover analysis. Marketing.	Analyze consistency of users travel behavior over time. Produce targeted marketing campaigns to retain users in specific groups.
Blythe (2004)	Route load profiles.	Manage the demand through the network.	Make public transport more attractive.
Chu &Chapleau (2008), Chu et al. (2009)	Boarding date, time, and location. Card type. alighting point. Card type.	Runtime estimation. Itinerary reconstruction. Spatio-temporal portrait on the network. Concept of Driver Assisted Bus Interview (DABI).	Make adjustments to network geometry and schedules. Richer information than that from travel survey. Adapt network to user needs.
Deakin& Kim (2001)	Information from '*Travel Information Services'.*	Provide information in real time on actual and planned conditions of the network.	Allow users to follow an alternative itinerary or to make a well-informed choice before journey.

The large amount of data collected gives a better understanding of travel behavior, because there are more observations in space and time with this system than with any other means of data collection (Agard et al., 2006; Bagchi& White 2005). For every smart card fare collection transaction, the date, time, and card numbers are available, which facilitates the calculation of ridership statistics at a precise temporal scale. In some cases, boarding location is also available, which means that these statistics can be spatially detailed over a network. However, since

there is usually no information available on the card holders themselves, these data lack sociodemographic attributes, and so there is a need to enrich them by traditional means of collection, like the household survey (Trépanier et al., 2009).

2.4.1.2 Tactical-Level Studies

On the tactical side, the focus is on schedule adjustment, longitudinal and individual trip patterns. Service adjustment is the most frequent topic as table 2.5 shows. That is, most transit authorities will provide a similar schedule for all weekdays. However, there can be large variations in ridership during weekdays, and so a different schedule could be posted for each day

Since most systems do not have the alighting point for individual trips, an algorithm like the one proposed by Trépanier et al. (2007) is needed to estimate the most probable alighting point. This is done by looking at the next boarding point in the day or by applying similarities to other trips made with the same card in the historical database. Then, the load profile of the route can be calculated because the boarding and alighting points are available for each transaction. When the boarding stop is not known, it is theoretically possible to estimate the boarding location by looking at the time of boarding compared to the bus schedule, but the application of this method was not found in the literature.

Table 2.5 Review of studies on smart card systems for tactical transit planning.(Marie, A., & Catherine, M. (2009), Smartcard data in public transit planning: A review).

Author(s)	Data used	Analysis/use	Benefits
Bagchi& White (2004)	Trip data and personal data.	Reconstruction of user trips, by identifying bus used and transfer point.	Transport offering adjustment. Improved data quality. Statistics available. Schedule adjustment.
Blythe (2004)	Pattern behavior, profile, and preference.	Customer-loyalty scheme.	Better understanding of customer needs.
Bagchi& White (2005)	Origin and destination.	Construct trips made and examine travel pattern.	Service adjustment.
Chapleau& Chu (2007)	Boarding date, time, and location. Estimated alighting point. Card type.	Analyze the boarding passengers variability on a specific route. Detect transfer coincidences. Analyze transfer activities. Identify linked and no-linked itinerary.	Detection of maximum boarding point and return runs. Schedule coordination between bus and metro.
Hoffman et al. (2009)	Magnetic card data on entry points.	Iterative classification algorithm to obtain more information on transfer journeys.	Better view of transfer journeys. Could be applied on smart card data.
Utsunomiya et al. (2006)	History of use.	Determine frequency and consistency in user patterns.	Service adjustment.

2.4.1.3 Operational -Level Studies

Operational -Level Studies are related to supply and demand indicators, as well as to smart card system operations. Table 2.6 shows that smart card systems can be used to calculate precise performance indicators on a transit network, like schedule adherence, vehicle-kilometers and person-kilometers for each individual run, route, or day (Trépanier et al., 2009). Schedule adherence can be estimated by comparing the boarding times at given stops along the route with the route schedule. In this case, data must be carefully filtered to retain the first transaction at each stop to represent the bus arrival time, because boarding can take several seconds. Indeed, smart card time stamps can also help estimate the average boarding time at different stops, and for different types of route and vehicles.

Table 2.6 Review of studies on smart card systems for operational transit planning.(Marie, A., & Catherine, M. (2009), Smartcard data in public transit planning: A review).

Author(s)	Data used	Analysis/use	Benefits/Potential
Attoh-Okine&Shen (1995)	Personal data, fares.	Payment.	Reduces perception cost.
Chapleau& Chu (2007)	Boarding date, time, and location.	Detect, quantify, and analyze errors and inconsistencies in transaction data.	Improvements can be proposed to system, and corrected data can be obtained.
Deakin& Kim (2001)	History of use and personal data.	Implementation of fare structure permitting time-of-day pricing.	Better flexibility in fare perception.
Morency et al. (2007), Trépanier et al. (2009)	Boarding date, time, and location. Estimated alighting point. Card type.	Analyze transit user behavior with data mining techniques. Follow buses.	Performance indicators. Schedule adherence.
Park & Kim (2008)	Personal data and bus run data.	Describe user characteristics (transfer point, boarding time, bus run by mode, type of user).	Better understanding of user habits.
Reddy et al. (2009)	Entry-only automated fare collection system data. Magnetic fare cards.	Several operational statistics available at individual level.	Reduces costs previously needed to calculate performance indicators.
Trépanier& Vassivière (2008)	Boarding date, time, and location. Estimated alighting point. Card type.	Several operational statistics available at bus run level. Intranet for dissemination.	Service can be adjusted at micro level. Defective equipment can be detected if data are incorrect or missing. Improves use of embedded system.

Smart cards can also help to detect irregularities and errors in the smart card payment system itself, in addition to fulfilling the core purpose of these systems, which is payment management, (Deakin& Kim, 2001). The presence of errors in transaction data can lead to rapidly identifying defective equipment, fraud, or employee errors.

CHAPTER THREE
MODEL STRUCTURE

This chapter contains the detailed models that have been used to find both value of time and reliability measures variables. Becker's (1965) model has been chosen to predict the value of time by using binary logistic regression technique in SPSS Statistic, B values have been taken for the three different cases, and then these values have been entered into Becker's equation to predict the three different value of time representatives (VOT 1, VOT 2 and VOT 3). Value of time (VOT) has been modeled by SPSS Statistics in three different ways; the first one is by modeling cost values and time values separately as explanatory variables (independent) and then dividing the B of time values by B of cost values as it shown in Equation 5.1 (VOT 1). The second way is by modeling cost values over time values as explanatory variables and then taking the invers for the B values (VOT 2) and the last one is by using the model of mode choice using the explanatory variables as time values over cost values directly such like Equation 5.1 and getting the B values as a value of time (VOT 3).

Reliability also has been measured as a function of travel time and for that mission Polus' (1978) model has been used to fined two different reliability measures (Reliability 1 and Reliability 2). Average income and Development level for each district have been included in the calculations. After including these two variables, each district has seven different variables now; VOT 1, VOT 2, VOT 3, Reliability 1, Reliability 2, Average income and Development level. The correlation has been checked for these variables and for the public transport smart card statistical by using SPSS Statistic and then linear regression has also taken a place in this study. Figure 3.1 shows the modeling process.

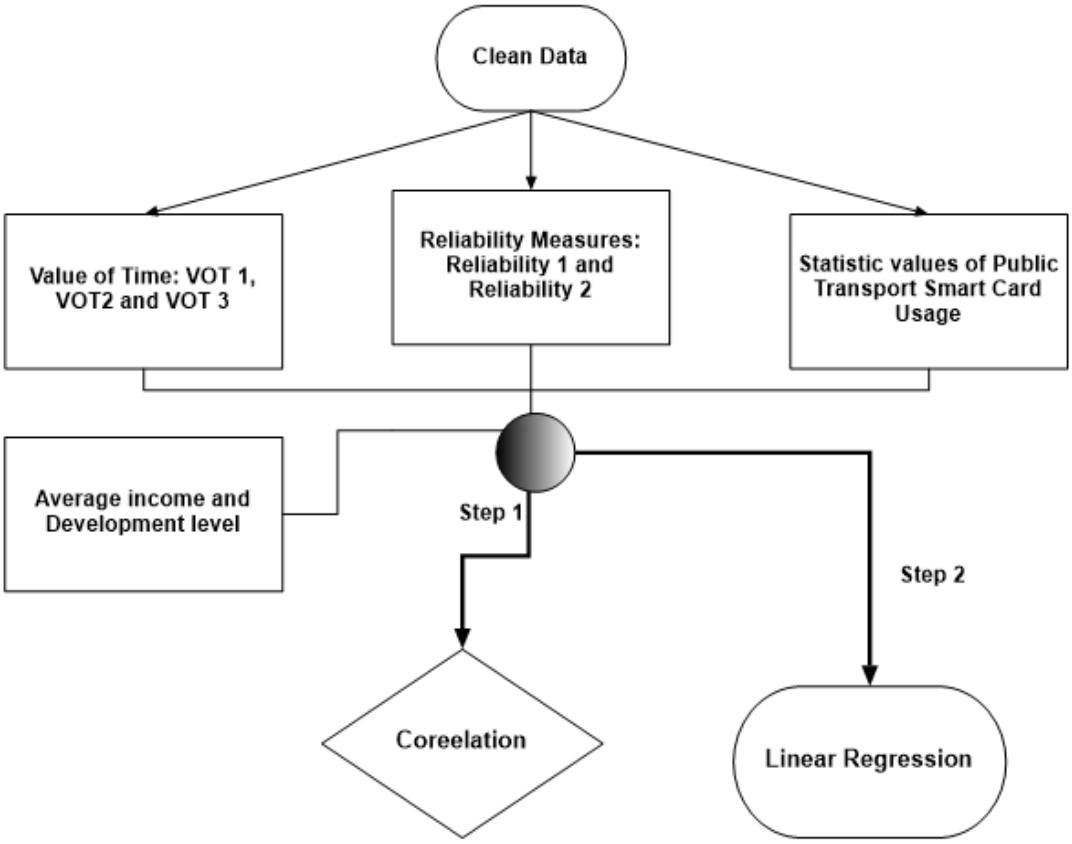


Figure 3.1 Modeling process.

3.1 Value of Time Model

As it was mentioned before, researchers and economists from many different perspectives have started to study value of time a half century ago, including those who interest in studying the home production or the labor market, and also by researchers who interest to understand the travel behavior. After Becker's (1965) theory, it has been recognized that time plays an important role in consumer behavior. In traditional models individuals try to achieve the highest level of satisfaction given the constraint faced.

Becker (1965) used time as an input to obtain the final good. While, other authors like Oort (1969) and De-Serpa (1971) included all activity duration as a direct source

of utility. Evans (1972) assumed in his model that the only source of utility is the time assigned to activities. Later on, the consumption-activity model has gradually settled in the literature. Both Small (1982) and Gronau (1986) adopted this more general approach as well. Jara-Diaz (2003) showed in his model that there is a need to establish two families of relations between goods consumption and time use.

When time is included in consumer theory, there are three important things should be taken into consideration: first, how time enters the utility function; second, there is a need to include a time constraint; and the third thing is the need to identify the relation between time allocation and goods consumption (Jara-Diaz, 2007). Backer's (1965) is one of the most important theories will be shown and discussed. In this study as has been mentioned before, value of time is going to be modeled by using Backer's (1965) model.

3.1.1 Becker (1965)

In 1965, Becker's pioneering work was one of the first economic works that introduce the time dimension to the utility function of consumers. He emphasizes that, when decisions took by individuals about a non-work activities the value of time plays an important role. Consumption is not only the purchases of market goods; time is also needed to consume the market goods. Because of this, he defines commodities as a function of both market goods and time. The individuals are the producers of commodities.

When an individual decides to travel by bus, both purchase bus tickets and the spent time are forming the commodity. Commodity i can be written as:

$$Z_i = f_i (X_i, T_i) \tag{3.1}$$

f_i is a production function needed to produce commodity *i*, X_i is a vector of market goods and T_i is a vector of time inputs.

Individuals maximize their utility function given by:

$$U = U(Z_1, \ldots Z_m) = U(f_1, \ldots f_m) = U(X_1, \ldots X_m; T_1, \ldots T_m) \quad (3.2)$$

The subscript m denotes the number of commodities that are being produced by the consumer. The goods constraint can be written as:

$$\sum_{1}^{M} P_i X_i = I = I_f + Ww \quad (3.3)$$

P_i is a vector giving the unite prices of X_i, I is the total income, I_f is income from other sources, $W$is a vector giving the hours spent at work and w is a vector giving the earnings per unit of W.

The time constraints can be written as:

$$\sum_{1}^{M} T_i = T_c = T - W \quad (3.4)$$

$T_c$is a vector giving the total time spent on consumption and T is a vector giving the total time available.

The production functions in Equation 3.1 can be written in the equivalent form

$$T_i = a_i Z_i \quad (3.5)$$

$$X_i = b_i Z_i \quad (3.6)$$

$a_i$is the amount of time spent per unit of Z_i and $b_i$is the amount of market goods spent per unit of Z_i.

He admits that the goods and time constraints are not independent because of Wentering in both expressions; time can be converted into goods by using less time at consumption and more at work. By substituting for Win Equation 3.3 its equivalent in Equation 3.4 gives the single constraint.

$$\sum_{1}^{M} P_i X_i + \sum_{1}^{M} T_i w = I_f + T w \quad (3.7)$$

By using Equations 3.5 and 3.6, Equation 3.7 can be written as:

$$\sum(P_i\ b_i + a_i\ w)Z_i = I_f + T\,w \qquad (3.8)$$

With, $\pi_i = P_i\ b_i + a_i\ w$ (3.9)

$$S = I_f + Tw \qquad (3.10)$$

The equilibrium condition resulting from maximizing Equation 3.2 subject toEquation 3.8 takes a very simple form, as long as *w* were constant

$$U_i = \frac{\partial U}{\partial Z_i} = \lambda\pi_i \qquad i = 1, \ldots m \qquad (3.11)$$

Where, λ is the marginal utility of income.

Equation 3.9 shows that the full price of commodity i is not only the price of the market good itself, but also the value of time used to consume the market good. Thus, the commodity price divides into direct part is consisting the market goods used, and indirect part consisting of time used. In Becker's model, the value of time is equal to consumer's wage rate.

$$\frac{\mu}{\lambda} = w \qquad (3.12)$$

Where, μ is the marginal utility of time and λ is the marginal utility of income.

Becker's model equalizes the value of time with the individual wage rate. Hence, the wage rate is the ratio between the Lagrange multiplier of time restriction μ and the Lagrange multiplier of income restriction λ. And to find these two Lagrange multiplier values μ and λ logistic regression analysis has been used, a vast amount of value of time studies has been conducted using these two types of discrete choice models. For this operation of analysis IBM SPSS Statistics program has been used to find the B values, which they are the represented values for the Lagrange multiplier μ and λ. This way and specially this program give good opportunity to arrange the dataset and to preview the results in a good way with an accurate calculation.

3.2 Logistic Regression Analysis

The goal is to estimate the regression coefficients in a model as in linear regression, given a sample of (X, Y) pairs. In the case of logistic regression, the X's can be numerical or categorical, but Y's are generally coded as 0 (for those who do not have the event) or 1 (for those who have the event).

A linear relationship between the natural logarithm (ln) of the odds of an event and a numerical independent variable is assumed in the simple logistic model. The form of this relationship is as follows:

$$\mathrm{L} = \ln(0) = \ln(\frac{\mathrm{p}}{1-\mathrm{p}}) = \beta_0 + \beta_1 X + \varepsilon \tag{3.13}$$

Where, Y is binary and represent the event of interest (response), coded as 0/1 for failure/success, p is the proportion of successes, o is the odds of the event, L is the ln(odds of event), X is the independent variable, β_0 and β_1 are the Y-intercept and the slope, respectively, and Ɛ is the random error.

Computation of the estimates of β_0 and β_1 in logistic regression are far more complicated. p is the probability of the event, and then the odds of the event are:

$$Odds = 0 = \frac{p}{1-p} \tag{3.14}$$

L is defined as, L = ln (odds of event Y), sometimes called the "log odds" or logit of Y. L can be written in terms of p, Probability (Y=1), as follows:

$$L = \ln(0) = \ln(\frac{\mathrm{p}}{1-\mathrm{p}}) \tag{3.15}$$

The laws of exponents and logs and some algebra can be used to express p (the proportion of successes or risk of the event) in terms of L:

$$\ln(0) = \ln(\frac{p}{1-p}), \text{ then } e^{L} = 0 = \frac{p}{1-p}, \text{so } p = e^{L} - pe^{L},$$
$$\text{and hence } p = \frac{e^{L}}{e^{L} - pe^{L}} \tag{3.16}$$

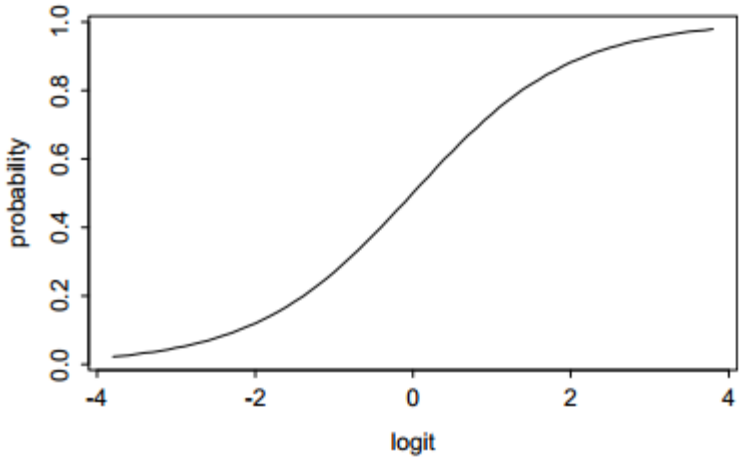


Figure 3.2 The logit transformation.

This is called the logistic function and its graph is as follows. Notice that p, the probability of the event, and increases from 0 to 1. The logistic regression model may be written in terms of p, the risk of event Y, assuming that L is a linear function of X as follows:

$$P(x) = \frac{e^{\beta_0 + \beta_1 X + \varepsilon}}{1 + e^{\beta_0 + \beta_1 X + \varepsilon}} \tag{3.17}$$

As in linear regression, our goal is to estimate the regression coefficients β_0 and β_1, given sample of (X, Y) pairs. Here as in the linear regression framework, X can be numerical or categorical. However, the Y is not numerical as in linear regression; rather is coded 0(for those who do not have the event) or 1(for those who do have the event). Unfortunately, the techniques used in linear regression to estimate the regression coefficients cannot be applied to the logistic regression case.

3.2.1 Multiple Logistic Regression Analysis

In general, the applications that use multiple logistic regression analysis are involving a single binary dependent variable which it is denoted by Y and multiple independent variables which they are denoted by X_1, X_2,…,X_p. The form of the multiple linear regression equation is:

$$L = \ln(0) = \ln(\frac{p}{1-p}) = \beta_0 + \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon \qquad (3.18)$$

Where, Y is binary and represent the event of interest (response),coded as 0/1 for failure/success, p is the proportion of successes, o is the odds of the event, L is the ln(odds of event), X_1, X_2,…,X_p are the independent variables, β_0 is the intercept and β_1, β_2, … , β_p are the slope coefficient (i.e., the expected change in Y relative to one unit change in Xi), and Ɛ is the random error.

Unlike in linear regression, in logistic regression dependent variable, Y, is a binary variable. The independent variables, X_i can be continuous variables or dichotomous (sometimes called indicator) variables. The formulas used to estimate the intercept and slope coefficients are computationally complex.
The general multiple logistic regression models in terms of p is:

$$P(x) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon}} \qquad (3.19)$$

The odds ratio comparing the odds of an event for two people, one with X = (X_{a1}, X_{a2}, … ,X_{ap}) and one with X = (X_{b1}, X_{b2}, … , X_{bp}), is:

$$\widehat{OR} = e^{\bar{\beta}_1 (x_{a1} - x_{b1}) + \bar{\beta}_2 (x_{a2} - x_{b2}) + ... + \bar{\beta}_p (x_{ap} - x_{bp})} \qquad (3.20)$$

3.3 Reliability Measure Model

Polus (1978) defined bus service reliability as "the ability of the service to provide consistent service over a period of time". A more comprehensive definition was given by Chapman (1976) covering all aspects of the service as "safety, punctuality, regularity, cleanliness, passenger comfort, documents, security".

Public transport reliability can be divided into two main groups: the route-based reliability and the stop-based reliability. Lower-level measures (e.g. the stop level) are also applicable at higher-level (i.e. the route or network level) (Kittelson, 2003). Stop-based reliability concentrates on reliability on the stop level, which can be used to find the reliability of specific stop, route or transit network. Whereas, route-based reliability reflects the reliability performance of a bus operation at the route level, which can be employed to assess the reliability of a specific route or a whole transit network.

In this section, three types of bus reliability measures are discussed. These are: travel time reliability, headway reliability and passenger wait time reliability. These definitions are not the exclusive definitions to the type of reliability measures. For example, the travel time reliability indicator defined in this study as the mean over standard deviation in journey times of a particular route within a particular time period. In the literature, travel time reliability has also been defined as the inverse of the standard deviation of journey times (e.g. Polus, 1978; Sterman & Schofer, 1976), the percentage of travel on a bus route that takes no longer than the expected travel time plus an acceptable additional time (Strathman et al, 2001), and the percentage of time that a bus's journey time is no more than 10% higher than average (Meyer, 2002).

3.3.1 Travel Time Reliability Model

Travel time reliability can be defined as the variability in bus journey time for a specific bus routes within a specific time interval and at a specific level of service or

probability that a trip can be completed within a given time. Travel time reliability is defined as the mean over standard deviation of travel time. The inclusion of the mean in addition to the standard deviation as in Polus (1978) is to allow for comparison of the variations in bus routes or sections of a route that have significantly different mean values. This definition can be written as:

$$RT_i = \frac{\mu_{ti}}{\sigma_{ti}} \tag{3.21}$$

Where, μ mean of the sample x, σ the standard deviation of the sample x and t_{inm} journey time of bus trip n on day (or simulation run) m along route i. With this definition, a bigger value of RT indicates a better reliability. In this study, the values of reliability for all districts have been measured according to this formula.

In this study, reliability for all districts has been measured according to the travel time reliability formula. First of all, reliability has been determined for each stop from all bus lines that pass through each stop. These values have been calculated by using the mean and standard deviation values of arrival time at each stop. In other words, the reliability of district is calculated by finding the reliability weighted average of the stops that located inside the district boundaries during the 61 working days.

CHAPTER FOUR
STUDY AREA

4.1 General Information about İzmir

İzmir is the largemetropolis in the western of Anatolia. It is located in the eastern coast of the Aegean Sea. İzmir is the third most populous city in Turkey, with an area of 12,007 km^2 and a total population of 4,061,074. İzmir Metropolitan Municipality is located in 24 zones which has 743 districts (Türkiye İstatistik Kurumu [TÜİK], 2013).

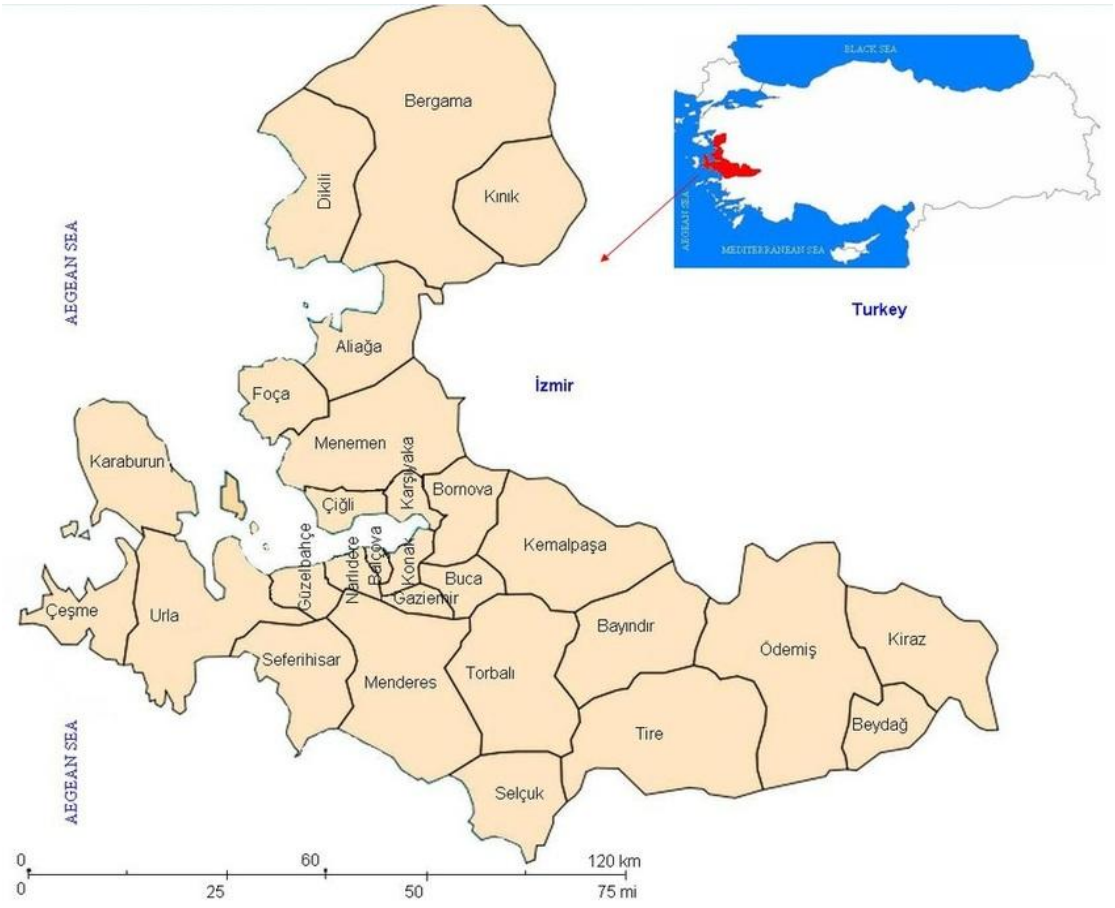


Figure 4.1 İzmir Metropolitan Municipality.

By 2013 the Metropolitan Municipality provided three types of public transport services: wheeled system (bus, minibus and taxi), rail system (commuter rail-IZBAN, light rail system- METRO) and ferry service on the bay sides. All these types of public transport are operated by İzmir Metropolitan Municipality except the minibus and taxi services.

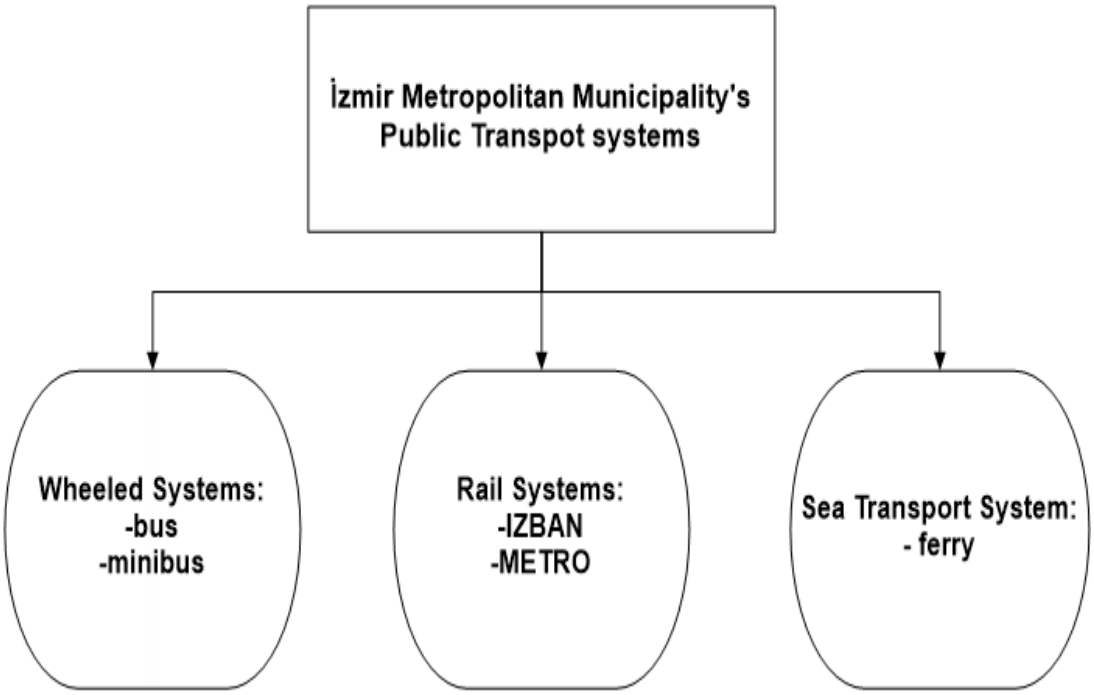


Figure 4.2 İzmir Metropolitan Municipality's public transport systems.

Currently, the İzmir Metropolitan Municipality's public transport system contains: 336 bus lines, 1450 buses and around 6700 bus stations. In addition to bus system, 17 km light rail line length and 17 stations (METRO), 79 km commuter rail line length and 31 stations (IZBAN) and 8 ferry stations on the bay sides with 15 ferry lines and 23 ferries (ESHOT,2013).

In December 2011, a survey showed that the average boarding number of working days were 1.378.729, 1.127.050 for Saturdays and 780.155 for Sundays. The total boarding shares are 77% for buses, about 3% for ferries, 11% for light rail system

(METRO) and 9% for commuter rail system (IZBAN). Based on these data the most widely used type is the bus system.

Table 4.1 Sample of public transport boarding types for weekdays (These data have been collected between 07 and 11 May, 2012)

Public Transport type	**Boarding Type**					**Total**	**Rate**
	Full	**Student**	**Free**	**3-5 Ticket**	**Discount**		
Bus	601210	321015	146867	8057	10627	**1087776**	**75%**
METRO	103066	61608	12252	2561	1373	**180860**	**12%**
IZBAN	98941	38646	10736	1864	1966	**152153**	**10%**
Ferry	27746	5984	2711	1515	274	**38230**	**3%**
Total	**830964**	**427254**	**172566**	**13997**	**14239**	**1459019**	**100%**

Rail Systems: There are two rail systems in İzmir city. The first one has been built in 2000 and it was connecting Bornova and Üçyol. In March 2012 Ege University Campus and Evka 3 stations have been added to the rail line and in June 2014 the rail line is extended to F. Altay station. Today light rail system (METRO) offers a service between F. Altay and Evka 3 stations for 17 Km with 17 stations.

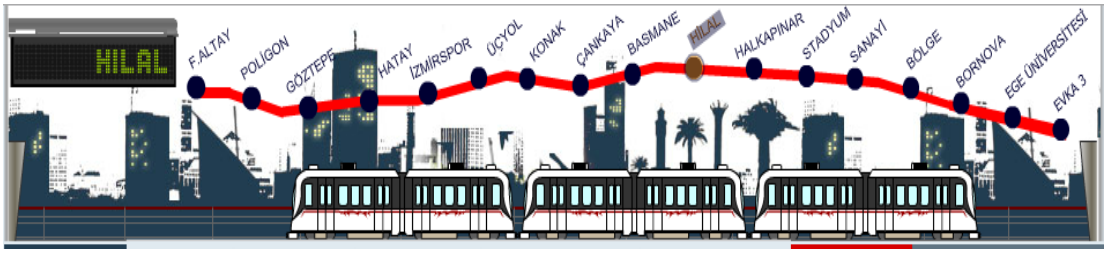


Figure 4.3 The stations of light rail system (METRO).

In 2011 IZBAN came to connect both sides of İzmir city the eastern part with the western part. It extents on a 79 Km with 31 stations and it is connecting Aliağa and Cumaovası zones. With 79 km line length, IZBAN ranked first all over Turkey in terms of length.



Figure 4.4 The stations of Commuter rail system (IZBAN).



Figure 4.5 The rail systems (METRO) and (IZBAN) and there stations.

Both Halkapınar and Hilal stations are the transfer stations for those passengers who want to transfer between the two systems or even to take a bus services as it shown in (Figure 4.5).

Ferry System: This public transport system has 23 ferries divided into 15 lines which cover the both sides of the bay. Today, about 200 trips per day are scheduled all over the 15 ferry lines. Konak, Bostanlı, Karşıyaka, Üçkuyular, Göztepe, Pasaport, Alsancak and Bayraklı are the 8 stations distributed on the bay sides. About 3% of passengers using the ferry system, which is quite low share.

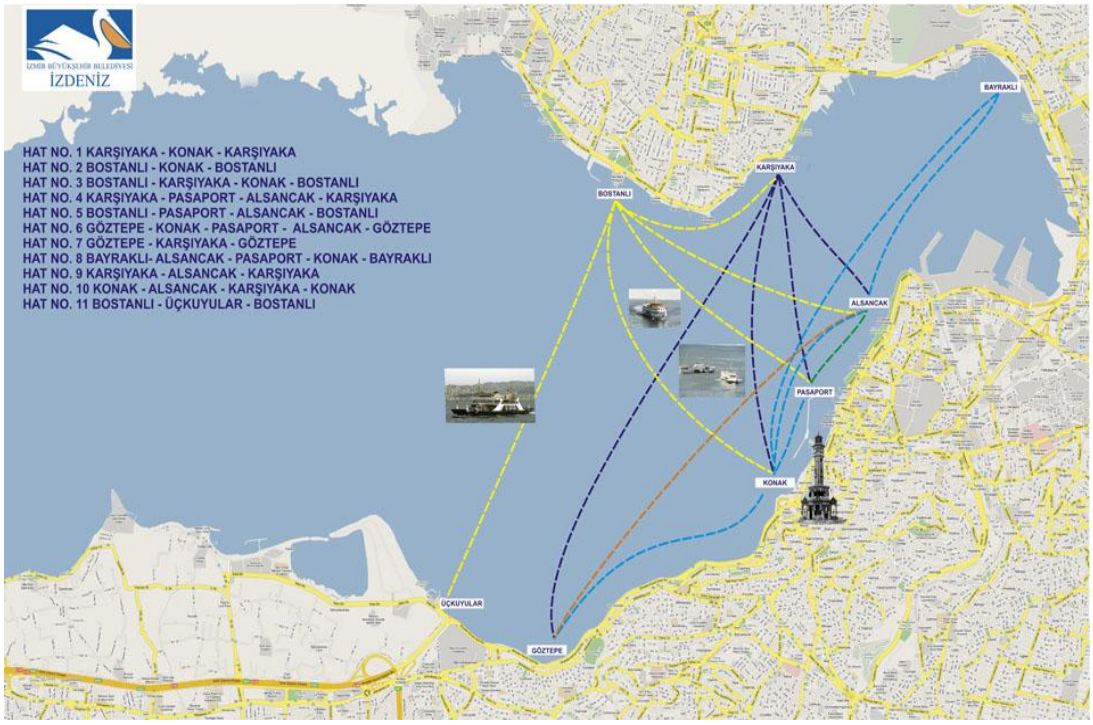


Figure 4.6 The ferry system and the ferry stations.

Bus System: This system is providing service for the most percentage of passengers. About 75% of passengers take the bus services to move all over the city. The bus system is connecting all zones and districts with a wide service network, which makes the system dominant.

Within the boundaries of the metropolitan area, public transport system has a total of 23 units of transfer center. These units give passengers the ability to change through these public transport systems. For instance, Konak and Bornova transfer centers are ranked as the highest passengers' volume with (56,010 trip /day) and (24,868 trip /day) respectively (ESHOT, 2012). Konak transfer center provides connection between bus, ferry and light rail (METRO) systems, which is located in the commercial, administrative and tourist center of Izmir city. Whereas Bornova transfer center is connecting bus system and light rail system (METRO), it is located near Ege University and the Hospital of Ege University which the significant destination points.

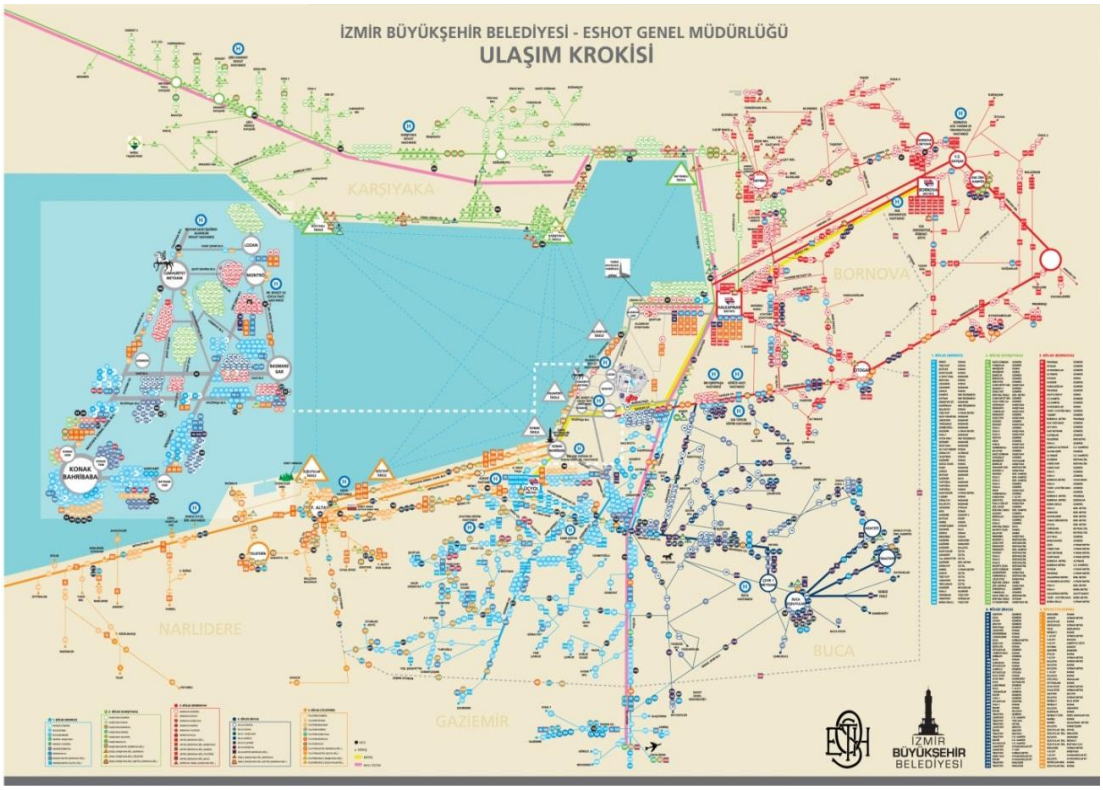


Figure 4.7 Bus system network.

4.2 Study Area

The study area is located in the northern residential part of İzmir Metropolitan. Study area is including 4 zones: Çiğli, Karşıyaka, Bayraklı and Konak. These 4 zones contain more than 72 districts as it shown in figure 4.8, while this study is going to deal with only 57 districts of them because of missing data. The study area contains different types of land use such as industrial, residential, commercial and administrative regions. Also there are different income groups for the residential regions showing continuous improvement which leads to increase the number of population in these regions. Figure 4.8 shows the study area boundaries, while figure 4.9 presents the 57 districts that have a complete data.



Figure 4.8 The study area boundaries.



Figure 4.9 The 57 districts that have a complete dataset.

Inside the boundaries of the study area there are some main streets such as (Anadolu, Altınyol and CemalGürsel) that connect the 4 zones together. In these zones, public transport systems are varying between three systems; bus system which it is the most one in the public transport system, commuter rail system (IZBAN) and ferry system. There are 60 bus lines inside the study area boundaries, 45 of these bus lines their starts and ends inside the study area as it shown in Figure 4.10.

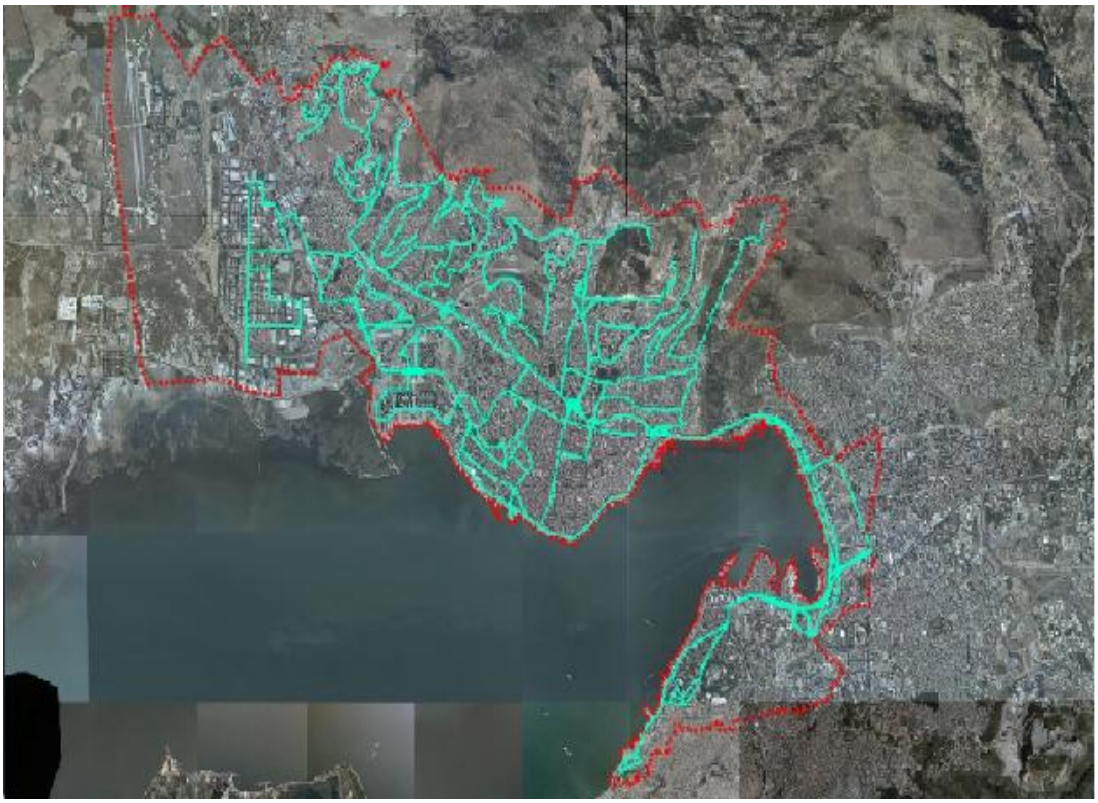


Figure 4.10 The bus lines routes inside the boundaries of the study area.

4.3 The Datasets

In this study two types of dataset have been used: the household travel survey dataset used for value of time analysis which it had been collected in 2007 by İzmir Municipality and the smart card data of public transport system. The household travel survey dataset contains important information for individuals live in each single district such as: gender, student or not, working or not, has an auto or not, has a driving license or not, has a public transport smart card or not, trip kind, trip time, trip cost, which transportation mode they choose to reach their destinations (public transport, private car or walking/ cycling), Figure 4.11 shows some of these information as statistical charts. After cleaning for the value of time analysis, 13.622 respondents remained. These respondents assigned to questionnaires for auto users, public transport users and walk/cycle. Figure 4.10 provides an overview for this dataset.

The other dataset is smart card data used for reliability analysis. This dataset had been conducted by ESHOT between October 1st and December 31st 2012. It was for 92 days, but in this study the weekends and other holidays have been eliminated, so 61 days remain. This elimination has been done to obtain more realistic reliability measures, because individuals give more interest for the reliability measures in working days.

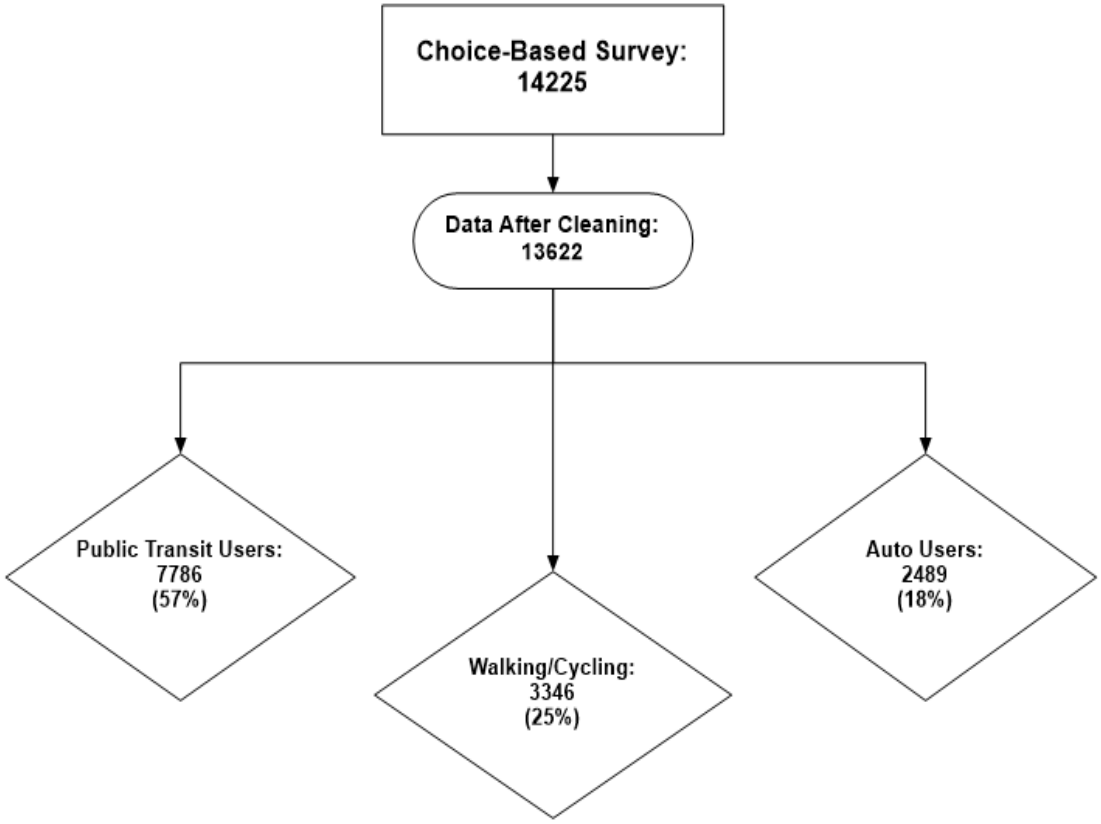


Figure 4.11 Overview of respondents.

Figure 4.11 shows that 7786 respondents were assigned to the public transport users. Out of these, 2489 answered were assigned auto as mode of transport and another 3346 answered were assigned walking/cycling as mode of transport.

4.3.1 Statistical Analysis of the Household Travel Survey Dataset

This section presents some statistical analysis for household travel survey collected in 2007 by İzmir Municipality.

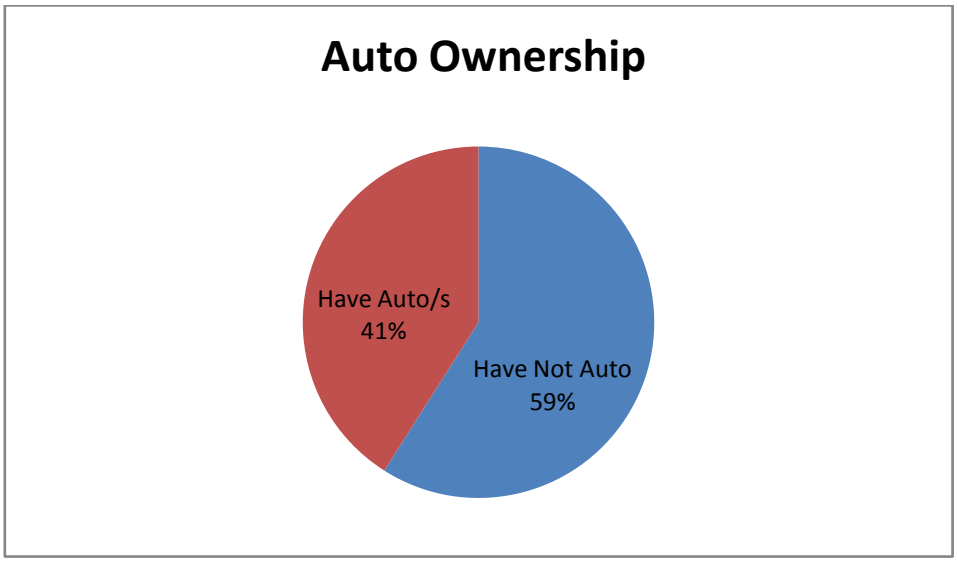


Figure 4.12 The percentage of individuals who own autos.

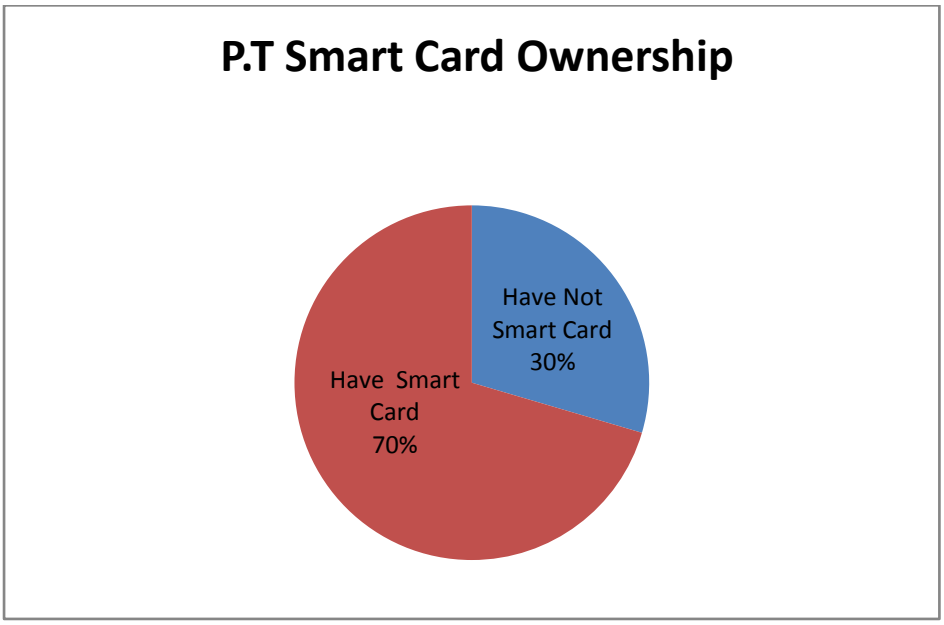


Figure 4.13 The percentage of individuals who own P.T smart card.

The above two figures show that41% of individuals who answered this survey have auto/s, and 70% of them have public transport smart card.

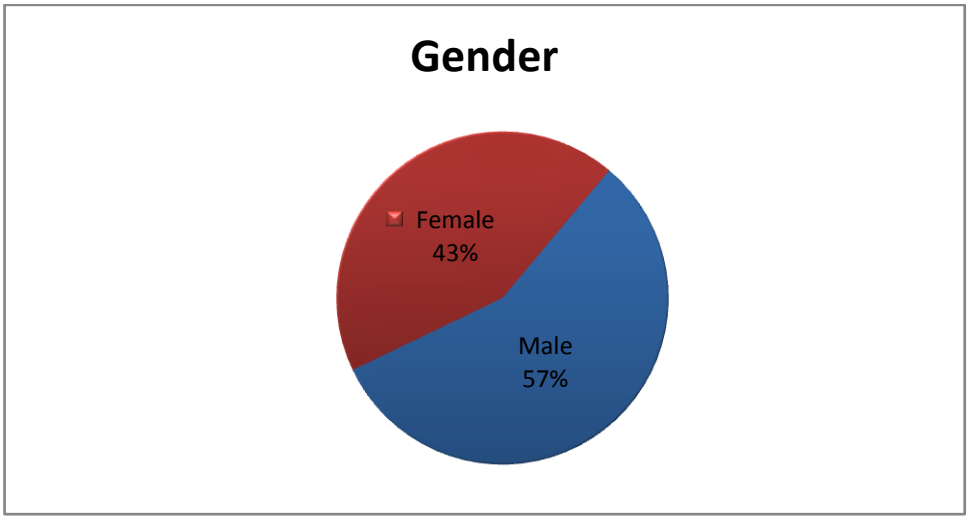


Figure 4.14 The individuals' gender percentage of the household travel survey dataset.

Table 4.2 Public transport mode choice according to gender case.

	Auto	Public Transport	Walking/Cycling
Male	1816	4490	1762
Female	802	3608	1747

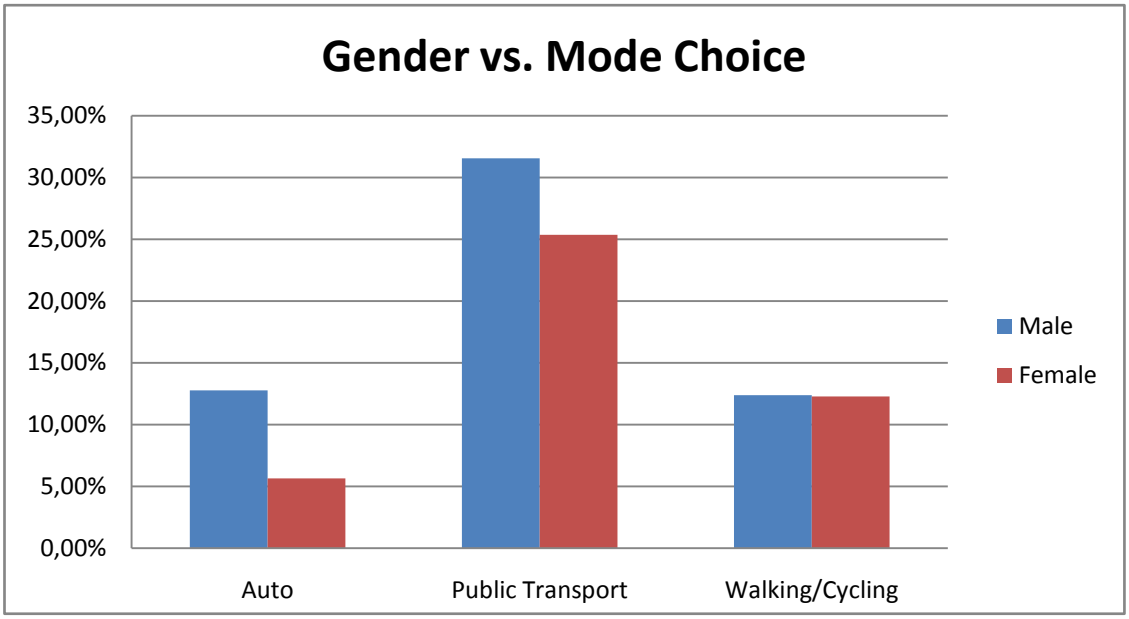


Figure 4.15 Gender vs. mode choice.

From these figures, around of 31% of public transport users are male and 25% of them are female.

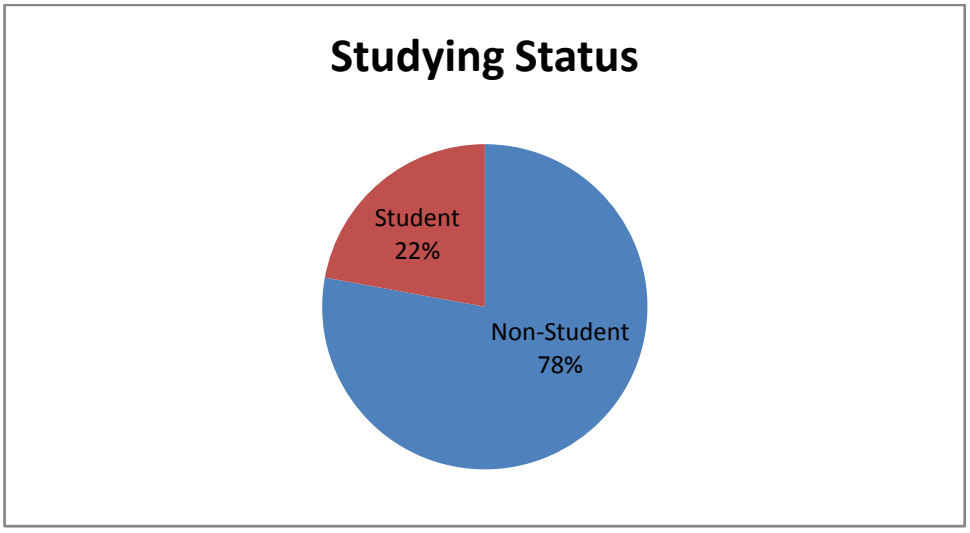


Figure 4.16 The students percentage in the household travel survey dataset.

Table 4.3 Public transport mode choice according to student case.

	Auto	Public Transport	Walking/Cycling
Non-Student	2511	6236	2313
Student	107	1862	1196

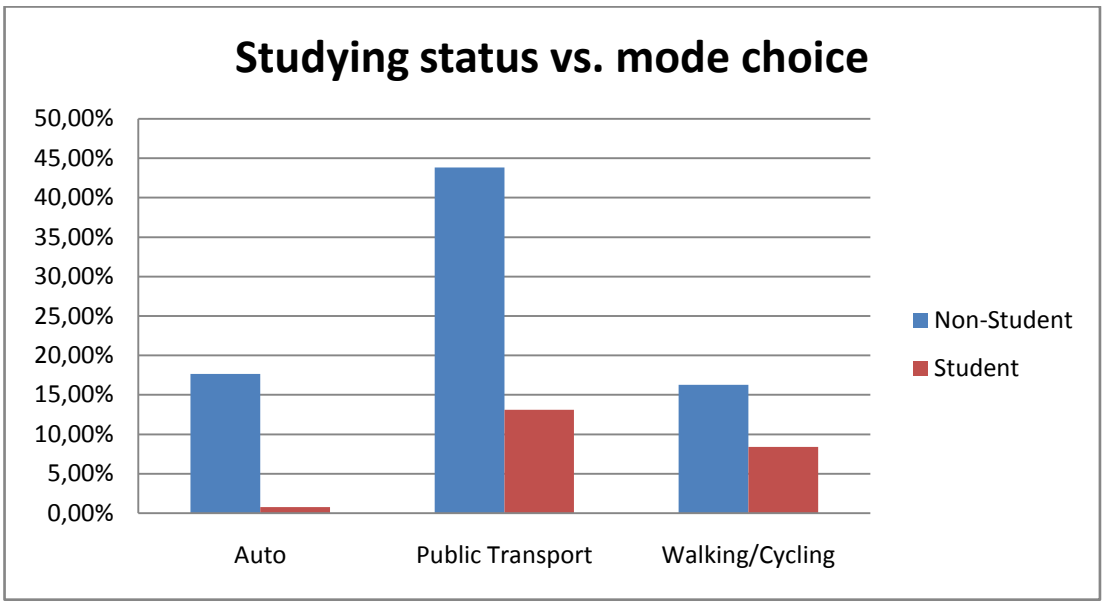


Figure 4.17 Student or non- vs. mode choice.

Students in this survey prefer to us public transport for their daily use, because public transport is cheap comparing to other mode and most of them do not have driving license.

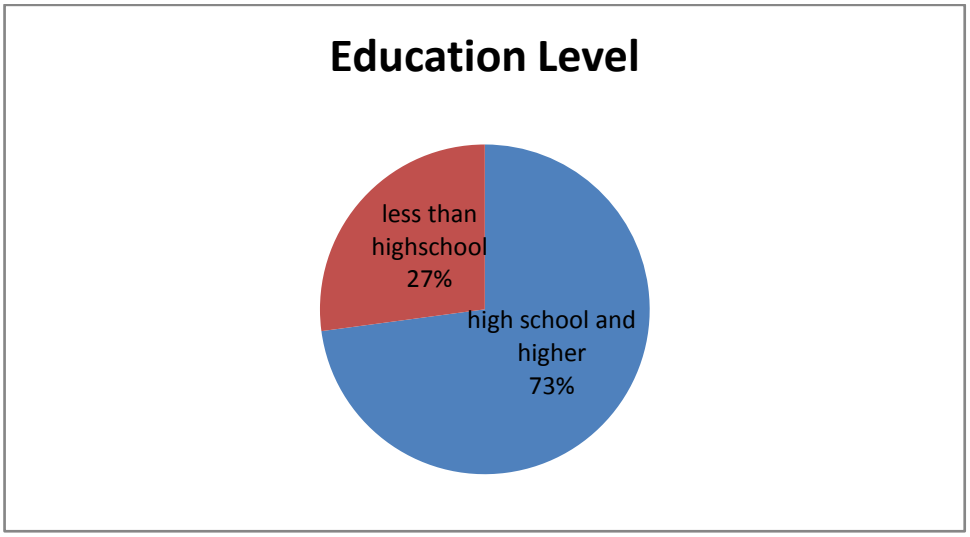


Figure 4.18 The education level of the household travel survey dataset.

Table 4.4 Public transport mode choice according to the education level.

	Auto	Public Transport	Walking/Cycling
Less than High-School	1319	6083	2962
High-School and Higher	1299	2015	547

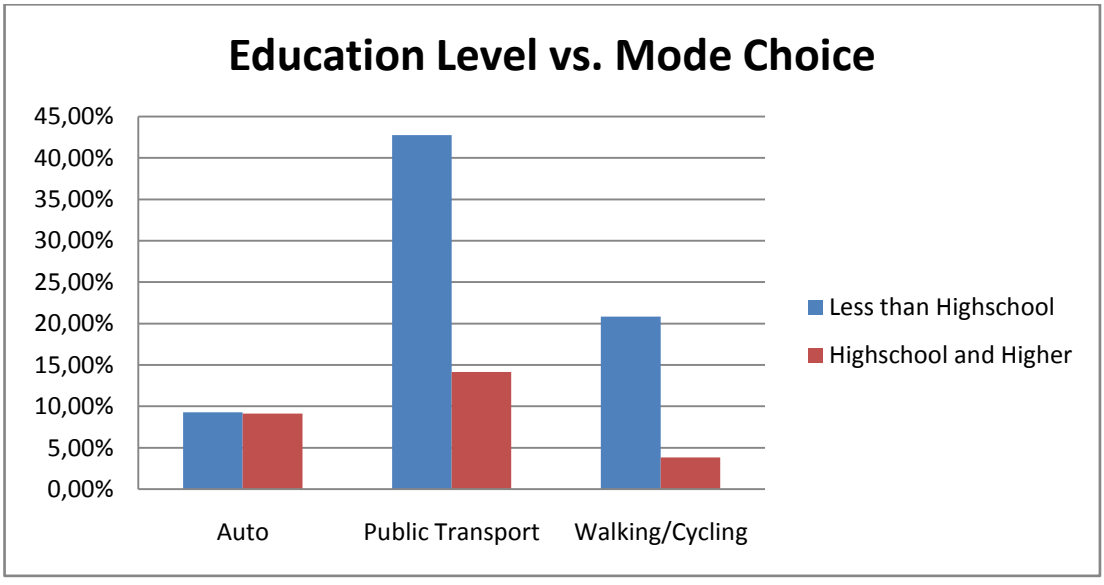


Figure 4.19 Education level vs. mode choice.

Education level effects public transport use as it shown in the above chart, individuals who have low education level use public transport more than other.

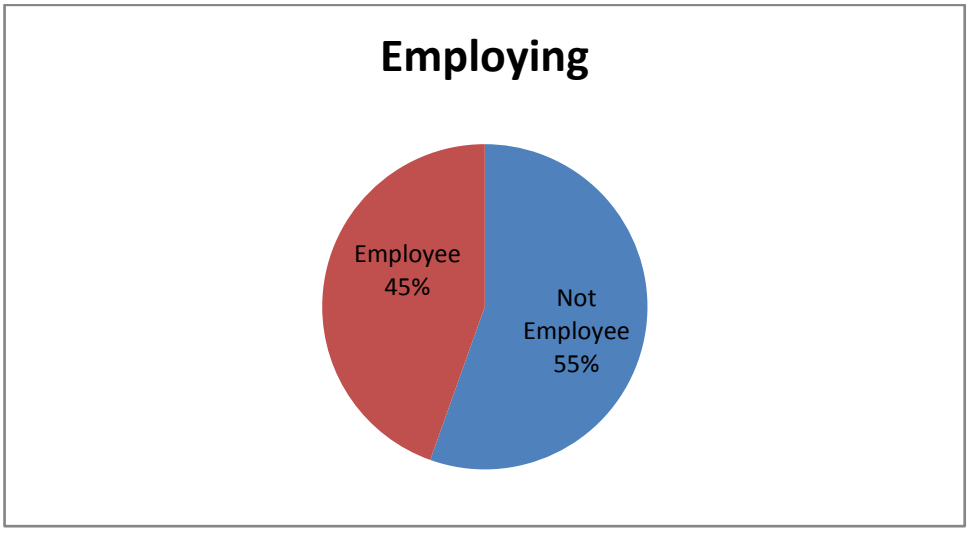


Figure 4.20 The individuals' employing percentage of the household travel survey dataset.

Table 4.5 Public transport mode choice according to the employing case.

	Auto	**Public Transport**	**Walking/Cycling**
Working	1736	3732	887
Not Working	883	4366	2622

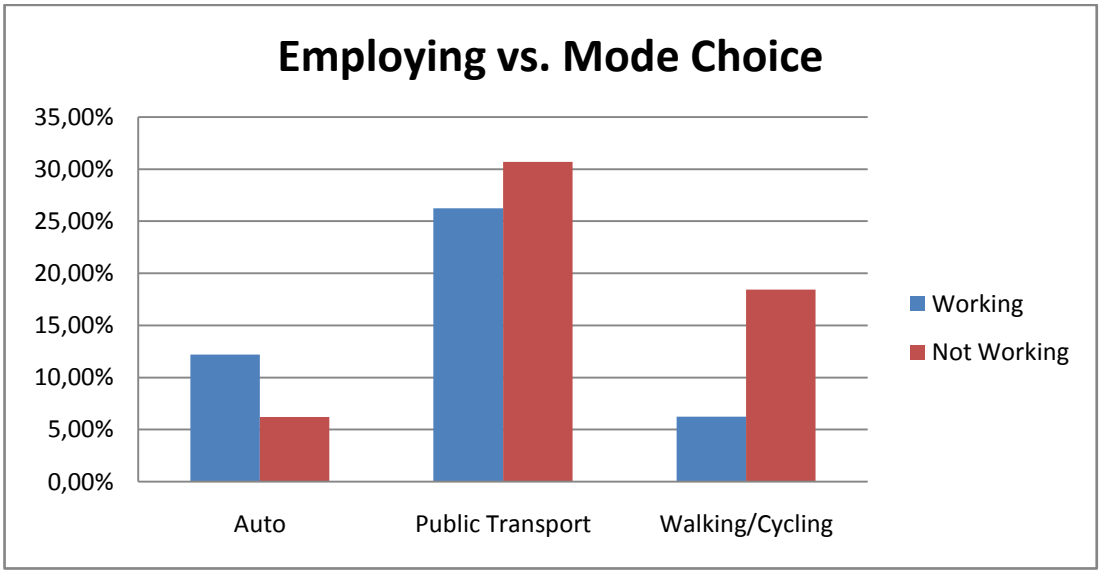


Figure 4.21 Employing vs. mode choice.

From the above chart, around of 26% of the employees in this survey prefer to use public transport mode.

CHAPTER FIVE
MODELING PROCESS

5.1 Value of Time Estimation

According to the time allocation models that had been discussed before, the simplest way to compute the value of time is to equal it with the individual wage rate. Wage rate is the ratio between the Lagrange multiplier of time restriction μ and the Lagrange multiplier of income restriction λ. In this study value of time has been estimated according to the following equation:

$$w = \frac{\mu}{\lambda} \tag{5.1}$$

Where, w is the value of time, μ is the Lagrange multiplier of time restriction and λ is the Lagrange multiplier of income restriction.

5.1.1 Logistic Regression by Using SPSS

To get a clearer sense of logistic regression, (Alsancak 311) district has been taken as an example to analyze it by using IBM SPSS Statisticssoftware, as it has a logistic regression module which is adequate for the study's purposes. Whether individuals claim to take auto or public transport is going to be check here. Let the dependent variable AUTO/PT takes the value 1 if an individual goes somewhere using an auto, and 0, otherwise (if he/she uses public transportation). So, AUTO/PT is a binary variable which means that logistic regression cab be used in the analysis. There are also some independent variables (auto, gender, driving license, public transport card and time/cost). Thegoal is to analyze and interpret the correlation between these variables and AUTO/PT.

The case processing summary Table 5.1 shows that there are 274 cases used in the analysis.

Table 5.1 Case processing summary

UnweightedCases[a]		N	Percent
Selected Cases	Included in Analysis	274	100.0
	Missing Cases	0	0.0
	Total	274	100.0
Unselected Cases		0	0.0
Total		274	100.0

a. If weight is in effect, see classification table for the total number of cases.

The first few tables SPSS program printout are not of major importance. The first one to take note of is the Classification table Table 5.2 in Block 0 Beginning Block.

5.1.1.1 Block 0: Beginning Block

Block 0: Beginning Block. Block 0 presents the results with only the constant included before any coefficients (i.e. those relating to AUTO, GENDER, DRIVING LICENSE, PUBLIC TRANSPORT CARD and TIME/COST) are entered into the equation. Logistic regression compares this model with a model including all the predictors (AUTO, GENDER, DRIVING LICENSE, SMART CARD and TIME/COST) to determine whether the latter model is more appropriate.

Table 5.2 Classification table

Observed			Predicted		
			AUTO_PT		
			AUTO	PUBLIC TRANSPORT	Percentage Correct
Step 0	AUTO_PT	PUBLIC TRANSPORT	0	119	0.0
		AUTO	0	155	100.0
	Overall Percentage				56.60

a. Constant is included in the model.

b. The cut value is .500

Classification Table 5.2 suggests that the base rates of the two decision options (155/274 = 56.6% decided to use car, 43.4% decided to use public transit), and no other information, the best strategy is to predict, for every case, that the passenger will decide to use car.

Table 5.3 Variables in the equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.264	.122	4.702	1	.030	1.303

Variables in the equation Table 5.3 shows that the intercept-only model is ln(odds) = 0.264. If the exponentiation for both sides of this expression are taken it is found that predicted odds [Exp(B)] = 1.303. That is, the predicted odds of deciding to use public transit is 1.303. Since 119 of passengers decided to use public transit and 155 decided to use car, the observed odds are 155/119= 1.303.

Table 5.4 Variables not in the equation

			Score	df	Sig.
Step 0	Variables	AUTO	59.862	1	.000
		GENDER	3.699	1	.054
		DRIVING LICENSE	37.414	1	.000
		SMART CARD	20.368	1	.000
		TIME/COST	51.437	1	.000
	Overall Statistics		111.218	5	.000

The variables not in the equation Table 5.4shows that whether each independent variable improves the model. The answer is positive for all variables, with AUTO, DRIVING LICENSE, TIME/COST and SMART CARD which are slightly better than GENDER, as the all are significant (which is less than or equal to 0.05) and if included would add to the predictive power of the model. If they had not been significant and able to contribute to the prediction, then termination of the analysis would occur at this point.

5.1.1.2 Block 1: Method = Enter

Block 1 Method = Enter. This presents the results when the predictors (AUTO, GENDER, DRIVING LICENSE, SMART CARD and TIME/COST) are included. There are a number of different methods we can select from. The default is ‘enter’, the other options are forms of stepwise regression where variables are entered in

order of significance. Later SPSS prints a classification table which shows how the classification error rate has changed from the original 56.6%. The 79.2% accuracy can be predicted by adding the variables (see Classification Table 5.8 below). The model appears good; it is evaluate model fit and significance as well. SPSS will offer you a variety of statistical tests for model fit and whether each of the independent variables included make a significant contribution to the model.

Table 5.5 Omnibus tests of model coefficients

		Chi-square	df	Sig.
Step 1	Step	144.769	5	.000
	Block	144.769	5	.000
	Model	144.769	5	.000

The overall significance is tested using what SPSS calls the Model Chi square, which is derived from the likelihood of observing the actual data under the assumption that the model that has been fitted is accurate.

At this stage, SPSS has added the independent variables as predictors. Omnibus Tests of Model Coefficients gives a Chi-Square of 144.769 on 5 degrees of freedom, significant beyond 0.000. Thus, the indication is that the model has a poor fit, with the model containing only the constant indicating that the predictors do have a significant effect and create essentially a different model.

Table 5.6 Model summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	230.332[a]	.410	.550

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Model summary Table 5.6 shows that the -2 Log Likelihood statistics is 230.332. This statistic measures how poorly the model predicts the decisions (the smaller the statistic the better the model). Although there is no close analogous statistic in logistic regression to the coefficient of determination R^2 the Model Summary Table 5.6 provides some approximations. Cox and Snell's R-Square attempts to imitate

multiple R-Square based on 'likelihood', but its maximum can be (and usually is) less than 1.0, making it difficult to interpret. Here it is indicating that 40.1% of the variation in the depended variable is explained by the logistic model. The Nagelkerke modification making the range from 0 to 1 is a more reliable measure of the relationship. Nagelkerke's R^2 will normally be higher than the Cox and Snell measure. Nagelkerke's R^2 is part of SPSS output in the 'Model Summary' table and is the most-reported of the R-squared estimates. In this case, it is 0.550, indicating a moderately strong relationship of 55% between the predictors and the prediction.

Table 5.7 Variables in the equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	AUTO	-1.387	.276	25.208	1	.000	.250
	GENDER	.160	.345	.216	1	.642	1.174
	DRIVING LICENSE	-1.511	.377	16.106	1	.000	.221
	SMART CARD	.935	.342	7.460	1	.006	2.547
	TIME/COST	.103	.022	21.213	1	.000	1.108
	Constant	.973	.470	4.276	1	.039	2.645

a. Variable(s) entered on step 1: AUTO, GENDER, DRIVING LICENSE, PUBLIC TRANSPORT CARD and TIME/ COST.

The variables in the equation Table 5.7 has several important elements. This table contains the coefficients for the (fitted) line and other relative information about the coefficients. The following column, labeled B, gives the unstandardized coefficients. These are the values for the logistic regression equation for predicting the dependent variable from the independent variable. The equation of the line found from the output is:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \quad (5.2)$$

Expressed in terms of the variables used in this example, the logistic regression equation is:

$$\ln\left(\frac{p}{1-p}\right) = 0.973 - 1.387 \text{ AUTO} + 0.160 \text{ GENDER} - 1.511 \text{ DRIVING LICENSE} + 0.953 \text{ SMART CARD} + 0.103 \text{ TIME/COST} \quad (5.3)$$

These estimates present the relationship between the independent variables and the dependent variable, where the dependent variable is on the logit scale. These estimates show the amount of increase (or decrease, if the sign of the coefficient is negative) in the predicted log odds of possession of literature = 1 (which means possession) that would be predicted by a 1 unit increase (or decrease) in the independent variable, holding all other predictors constant.

The Wald statistic and associated probabilities provide an index of the significance of each predictor in the equation. The Wald statistic has a chi-square distribution. The simplest way to assess Wald is to take the significance values and if less than .05 reject the null hypothesis as the variable does make a significant contribution. In this case, the driving license contributed significantly to the prediction (p = 0.000) but gender did not (p = 0.642). The Exp(B) column in Table 5.7 presents the extent to which raising the corresponding measure by one unit influences the odds ratio. If the value exceeds 1 then the odds of an outcome occurring increase; if the figure is less than 1, any increase in the predictor leads to a drop in the odds of the outcome occurring.

Table 5.8 Classification table[a]

Observed			Predicted		
			AUTO_PT		Percentage Correct
			PT	AUTO	
Step 1	AUTO_PT	PT	89	30	74.8
		AUTO	27	128	82.6
	Overall Percentage				79.2

a. The cut value is .500

The Classification Table 5.8 shows that 128 / 155 = 82.6% of the subjects where the predicted event (deciding to use auto) was observed. This is known as the sensitivity of prediction, the P(correct | event did occur), that is, the percentage of occurrences correctly predicted. Also this rule allows us to correctly classify 89 / 119 = 74.8% of the subjects where the predicted event was not observed. This is known as the specificity of prediction, the P(correct | event did not occur), that is, the percentage of nonoccurrences correctly predicted.

Value of time is represented by B values for the correlated TIME/COST variables (Sig. is less than or equal to 0.05) for each district. If it is insignificant, value of time considered zero. From Table 5.7 the value of time for (Alsancak 311) district is 0.103. Since the p-value of the variable time/cost is 0.00 (which is less than or equal to 0.05), TIME/COST has an effect on the variable AUTO/PT choice. Since the sign of B value (0.103) is positive, these variables are positively correlated. For a one unit increase in the variables TIME/COST, the log odds of car usage (AUTO/PT) increases by 0.103.

5.1.2 The Value of Time

Value of time (VOT) has been modeled by SPSS Statistics in three different ways; the first one is by modeling cost values and time values separately as explanatory variables (independent) and then dividing the B of time values by B of cost values as it shown in Equation 5.1 (VOT 1). The second way is by modeling cost values over time values as explanatory variables and then taking the invers for the B values (VOT 2) and the last one is by using the model of mode choice using the explanatory variables as time values over cost values directly such like Equation 5.1 and getting the B values as a value of time (VOT 3). According to the prediction results, the zero values mean that value of time for this model has no relationship (uncorrelated) with Auto/PT binary variable; because they have a significant value equal to 0.05 or bigger. Negative values mean that value of time is correlated negatively with Auto/PT and they correlated positively when it has a positive value of time sign. Table 5.9 represents these three values for all districts.

According to the time allocation models that had been discussed before, the simplest way to compute the value of time is to equal it with the individual wage rate. Wage rate is the ratio between the Lagrange multiplier of time restriction μ and the Lagrange multiplier of income restriction λ. In this study value of time has been estimated according to the Equation 5.1.

In this study the value of time has small values (less than 1). As it has been discussed in literature view chapter, the different time allocation theories calculate value of time as a function of individuals' wage rate, but unfortunately, the dataset that has been used in this study does not have the individuals' wage rate. Instead the cost of each trip has been used to estimate the value of time. In other words, value of time in this study is a function of trip cost not a function of wage rate. Besides that, time has been used as minutes, if it was taken as hours the coefficients would be 60 times of these coefficients. Table 5.9 represents the value of time prediction results.

Table 5.9 The three model estimated value of time for all districts

District Code	District Name	VOT 1	VOT 2	VOT 3
24	Ahmet Taner Kışlalı	0	0.03	0.07
25	Atatürk	0	0	0
26	Egekent	0	0.02	0
27	Esentepe	0	0	0
28	Evka-2	0	0	0
29	Evka-5	0	0	0
30	Evka-6	0	0	0
31	Güzeltepe	0	0	0
32	İzkent	0	0	0
34	Yakakent	0	0	0
35	Aydınlıkevler	0	0	0
36	Balatçık	0	0	0
37	Çağdaş	0.11	0.09	0
38	İstasyonaltı	0.04	0.04	0
39	Köyiçi	0	0	0
40	Küçük Çiğli	0.04	0.07	0
41	Maltepe	0	0	0
42	Şirintepe	0	0	0
43	Yeni Mahalle	0	0	0
44	Aksoy	0	0.07	0
45	Alaybey	0.01	0.02	0.13
46	Atakent	0	0	0
47	Bahariye	0.04	0.04	0.03
48	Bahçelievler	0.03	0.05	0.04
49	Bahriye Üçok	0	0.04	0
50	Bostanlı	0.04	0.04	0.03
51	Dedebaşı	0.03	0.04	0
52	Demirköprü	0	0	0.06
53	Donanmacı	0.07	0.07	0.02
54	Fikri Altay	0	0	0
56	Mavişehir	0.1	0.1	0.03
57	Nergiz	0	0.13	0
58	Şemikler	0	0	0
59	Tersane	0.02	0.03	0
60	Tuna	0	0.07	0
61	Yalı	0.03	0	0
62	Cumhuriyet	0.03	0.04	0
63	Emek	0.1	0.06	0
64	Gümüşpala	0.1	0.05	0
65	İmbatlı	0	0	0
66	Onur	0.03	0.06	0
67	Org. Nafiz Gürman	0.16	0.03	0
68	Örnekköy	0.08	0.05	0
69	Postacılar	0.04	0.04	0
70	Soğukkuyu	0	0	0
72	Yamanlar	0	0.05	0
73	Zübeyde Hanım	0.17	0.05	0
78	Adalet	0.07	0.15	0
80	Bayraklı	0	0	0
84	Fuat Edip Baksi	0	0.05	0
88	75. Yıl	0	0.05	0
282	Çınarlı	0	0.04	0.08
288	Halkapınar	0.68	0	0
311	Alsancak	0.17	0.07	0.1
322	Kahramanlar	0.16	0	0
326	Kültür	0.05	0.18	0
327	Mimar Sinan	0.09	0.08	0

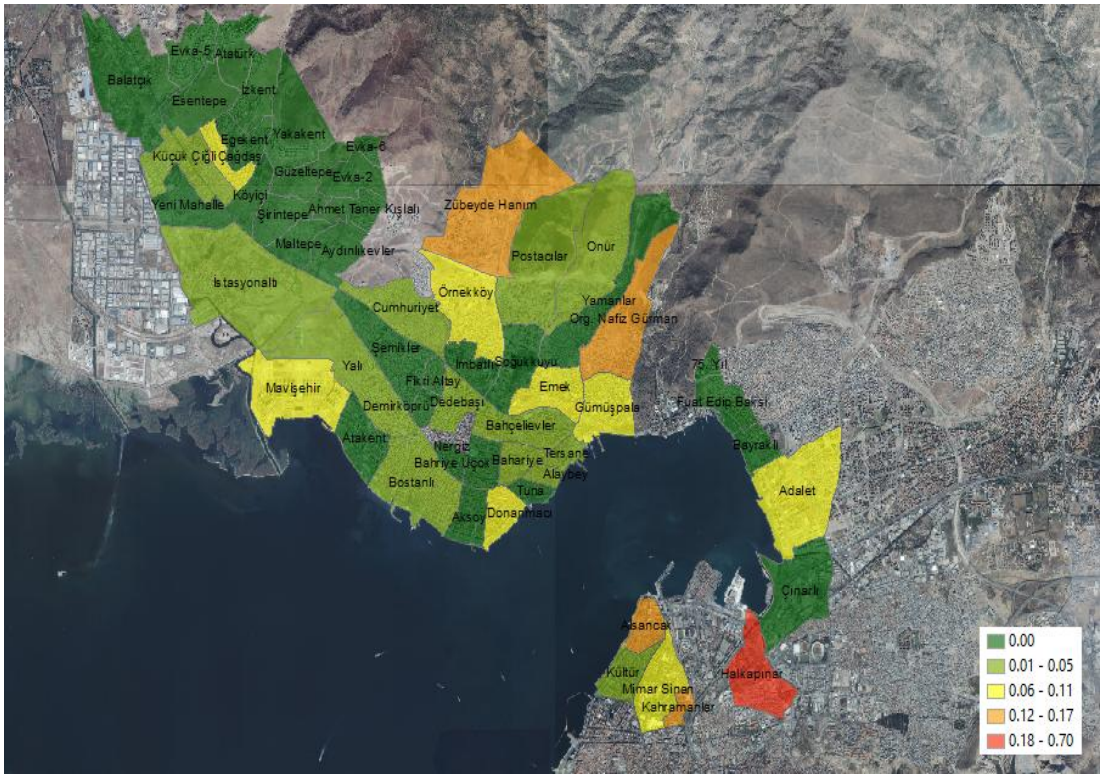


Figure 5.1 The estimated VOT 1 of the study area.

Figure 5.1 representing VOT 1 model shows that Halkapınar district has the highest value of time about 0.68, while the 31 districts that have a dark green color get a zero values. For this zone, it seems a bit suspicious. Right Halkapınar is the transfer area to change between three types of the main transportation systems (METRO, IZBAN and Bus Network) but the survey that made by Izmir Municipality and it has been used in this study to find the value of time, it was for the people that they are living in these districts and Halkapınar district is an industrial area not a residential area. Around 63 users answered that survey in this district, and that many not be enough to determine the behavior of whole district. For instance Alsancak district has a high value of time also and that is logical, because the families there have a high income comparing to other districts and they worth more about time and value of time.

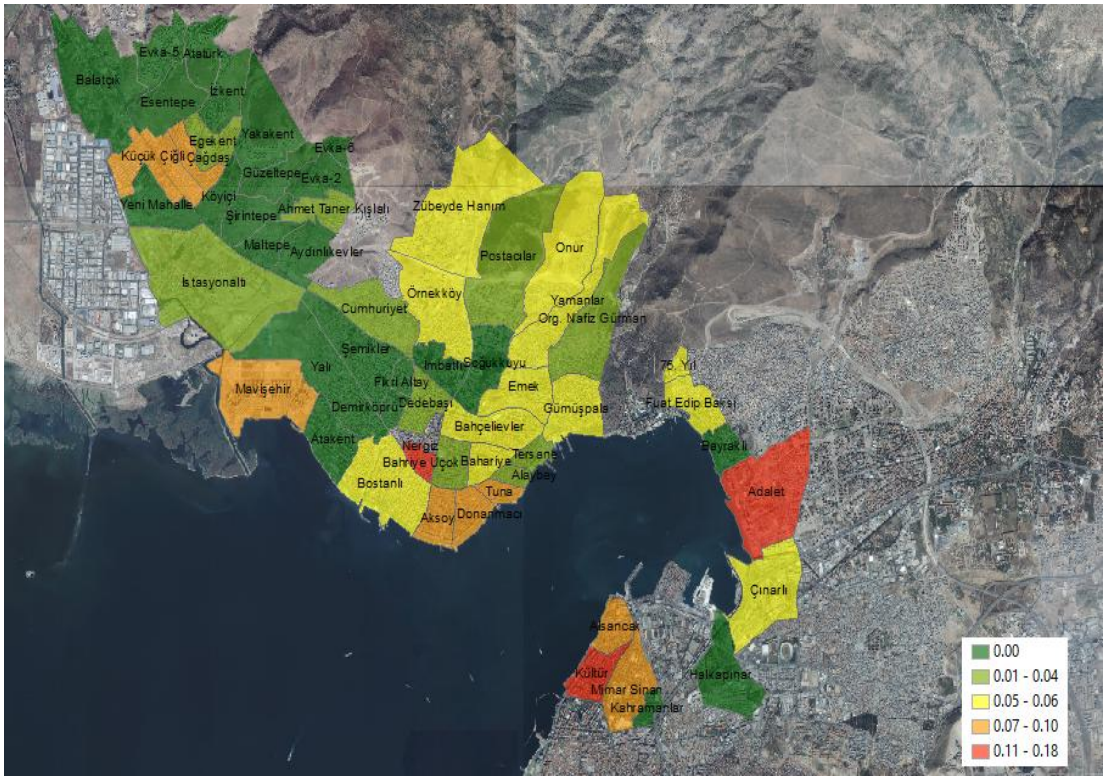


Figure 5.2 The estimated VOT 2 of the study area.

In Figure 5.2, demonstrating the distribution of VOT 2, Kültür, Nergiz and Adalet districts have the highest value of time, while 24 districts have a zero value. Alsancak and other sea side neighborhoods have also high value of time and that seems logical because high income families are living in these districts. Halkapınar district in this model has a zero value.

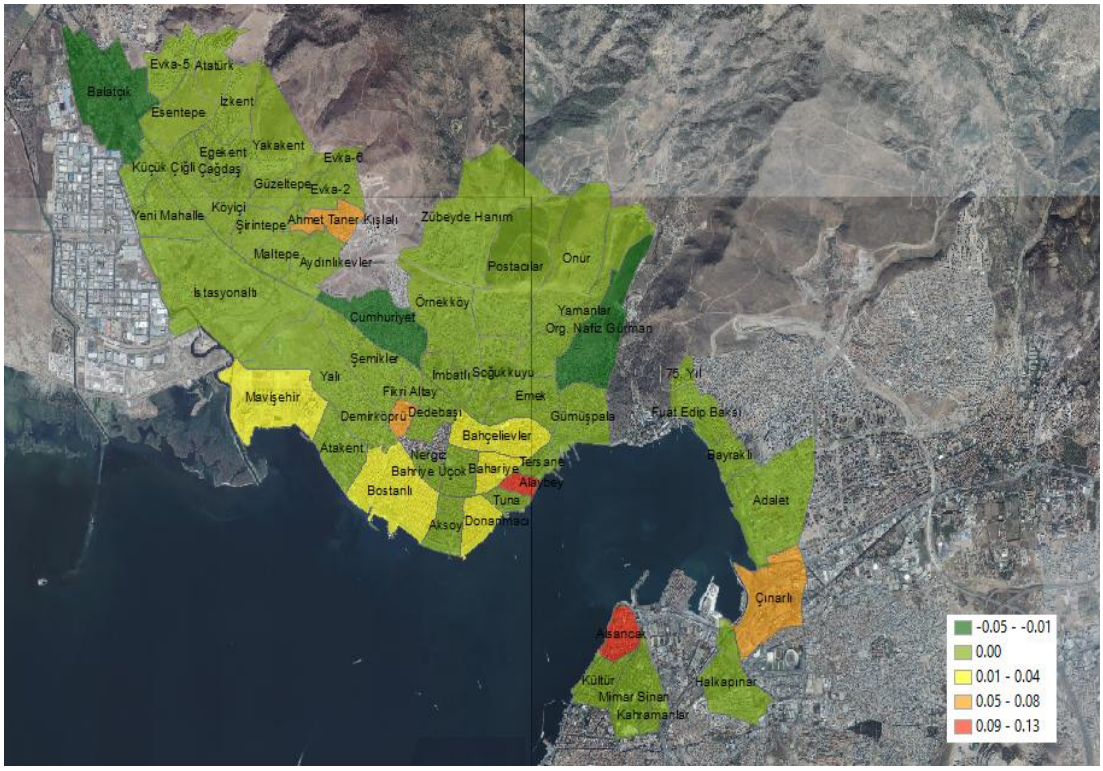


Figure 5.3 The estimated VOT 3 of the study area.

Figure 5.3 represents VOT 3 model. Around 50 districts have a zero value. That means this model is not efficient enough to give accurate values. However, Alsancak and Alaybey and some sea side neighborhoods still have non-zero positive values of time.

The most logical, reasonable and efficient model to represent the value of time for districts is the VOT 2 model. Only 24 districts that have a zero values which is less than VOT1 and VOT3 models, and it gives a logical high values for the districts that located in the sea side.

5.2 Average Income and Development Level

The following table shows the average income and development level for each district inside the boundaries of the study area.

Table 5.10 Average income and development level for all districts

District Code	District Name	Average Income	Development Level
24	Ahmet Taner Kışlalı	3.99	4
25	Atatürk	2.44	5
26	Egekent	3.24	5
27	Esentepe	3.85	5
28	Evka-2	3.13	5
29	Evka-5	3.19	5
30	Evka-6	3.75	5
31	Güzeltepe	2.36	5
32	İzkent	3.3	5
34	Yakakent	2.53	5
35	Aydınlıkevler	3.56	5
36	Balatçık	2.42	5
37	Çağdaş	3.86	5
38	İstasyonaltı	3.71	4
39	Köyiçi	2.61	5
40	Küçük Çiğli	2.14	5
41	Maltepe	2.36	5
42	Şirintepe	2.04	5
43	Yeni Mahalle	3.52	5
44	Aksoy	6.27	1
45	Alaybey	5.46	1
46	Atakent	6.29	1
47	Bahariye	4.9	2
48	Bahçelievler	5.35	2
49	Bahriye Üçok	5.13	2
50	Bostanlı	6.22	1
51	Dedebaşı	4.67	3
52	Demirköprü	4.36	3
53	Donanmacı	5.98	1
54	Fikri Altay	4.26	3
56	Mavişehir	6.96	1
57	Nergiz	4.86	2
58	Şemikler	3.98	3
59	Tersane	5.45	1
60	Tuna	5.49	1
61	Yalı	4.39	3
62	Cumhuriyet	2.89	4
63	Emek	2.42	5
64	Gümüşpala	2.48	5
65	İmbatlı	3.07	4
66	Onur	2.4	5
67	Org. Nafiz Gürman	2.75	5
68	Örnekköy	4.1	4
69	Postacılar	3.27	4
70	Soğukkuyu	3.53	4
72	Yamanlar	2.25	5
73	Zübeyde Hanım	3.41	4
78	Adalet	4.13	3
80	Bayraklı	4.91	2
84	Fuat Edip Baksi	4.16	3
88	75. Yıl	4.84	2
282	Çınarlı	2.91	3
288	Halkapınar	4.09	1
311	Alsancak	5.81	1
322	Kahramanlar	5.33	2
326	Kültür	6.11	1
327	Mimar Sinan	5.82	1

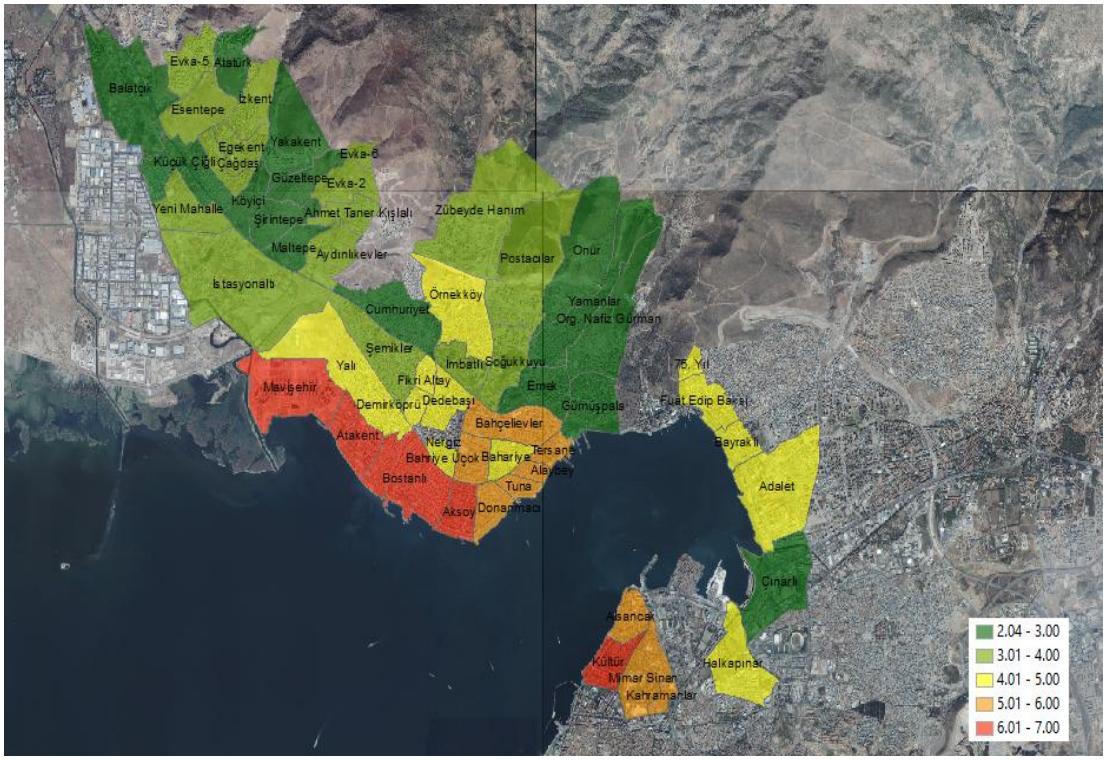


Figure 5.4 The average income of the study area.

The value of average income values in this study are based on the transportation master plan of Izmir city, 2008. As Figure 5.4 illustrates, those districts that located on the sea side have the highest value of average income and that make sense, residential location choice in these zones highly depends on income. For that, both value of time and reliability have to be high for these districts. Although they are crowded, the zones far from the city facilities have been shown in green on this figure and that is reasonable, because they also have low values of time. In other words, individuals in these districts cannot manage their time as they want; they have to be at the station on time otherwise they have to wait the next service or to pay more money to take a private mode.

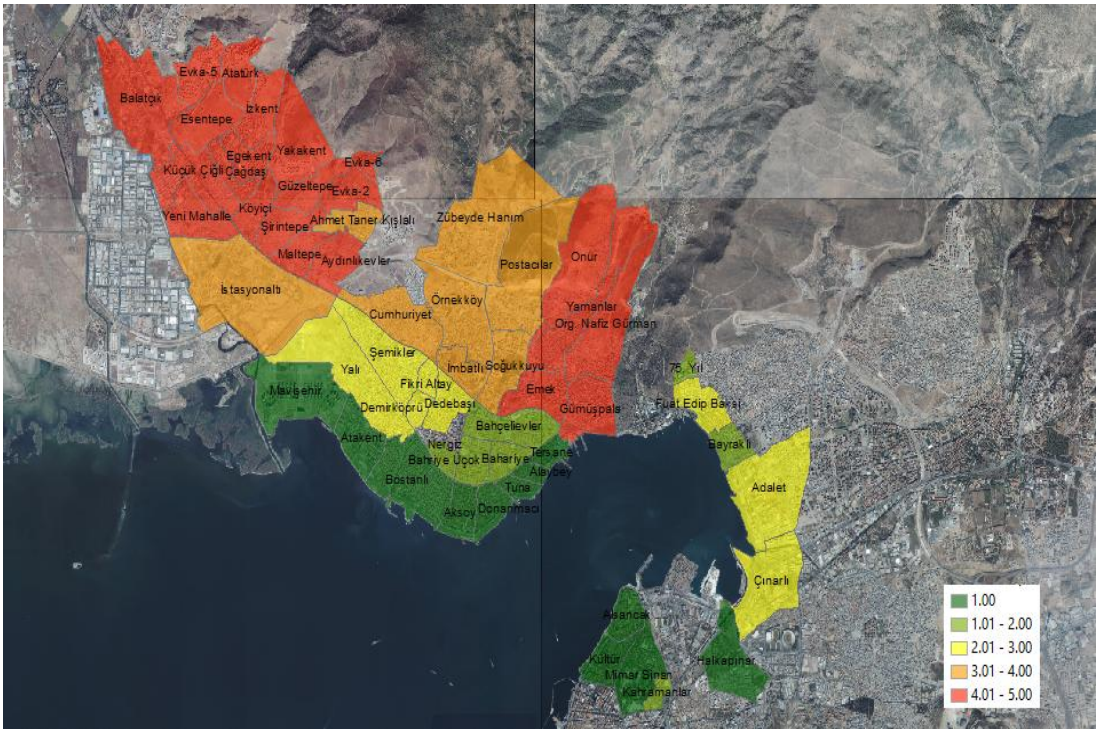


Figure 5.5 The development level of the study area.

The value of development level is based on the statistics provided by TUIK (2012). Here in Figure 5.5, development level is opposite to the average income, which means those districts that have a high value of average income will represent in a low value of development level.

The development level of the study area districts are varying from 1-6, the highest average income takes 1 and the lowest takes 6. As an example, the value of average income of Kültür district is 6.11 while the development level of the same district is 1. And then because of this, it is opposite to average income.

5.3 Reliability Estimation

For an in-depth analysis and evaluation of reliability of bus services in Izmir Metropolitan Municipality, it is demonstrated on one service as an example route. The bus line 121 runs from Bostanlı ferry station to Konak via seaside Figure 5.6 and links up residential areas in the north of the city with administration area, Konak. The bus length for the two ways is 46.10 km and passes 20 stations when it is going from Bostanlı ferry station to Konak and when it comebacks from Konak to Bostanlı ferry station it passes 24 stations. The bus starts working at 06:00 and finishes at 00:00.



Figure 5.6 The conceptual plan of route 121 in Izmir Metropolitan Municipality with its stops location. (ESHOT, Bus line number 121 Bostanlı ferry station to Konak,2012).

In this section, the values of stop reliability have been determined for (Alsancak 311) district for one single day. Bus 121 passes (Alsancak 311) district through 3 stops: “DevletHastanesi 10109” and “Talatpaşa 1011” for the direction Bostanlı - Konak, and “Talatpaşa 1010” for the direction Konak - Bostanlı (Figure 5.7).

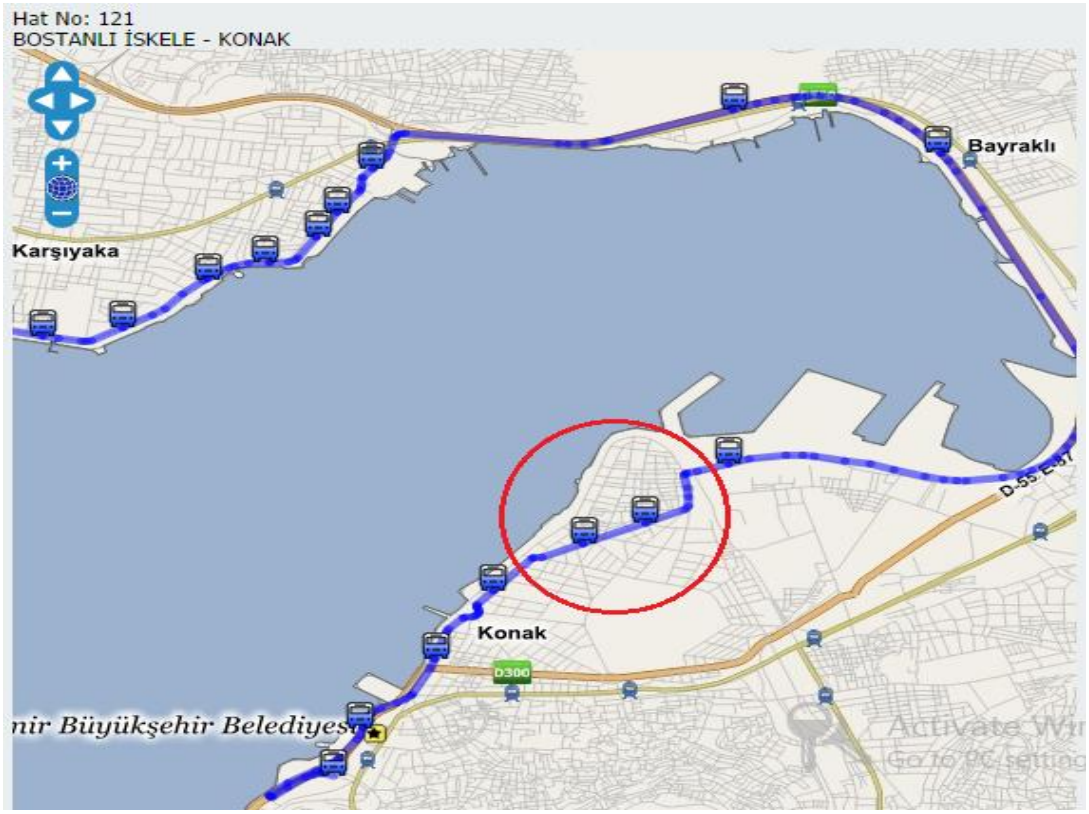


Figure 5.7 The boxed section is chosen for the detailed stop reliability study. (ESHOT, Bus line number 121 Bostanlı ferry station to Konak,2012).

Table 5.11 presents the stop reliability calculation for (Alsancak 311) district; these calculations have been done for a one single day (October 1st, 2012). The reliability has been measured as travel time reliability for each stop by using the following equation:

$$RT_i = \frac{\mu_{ti}}{\sigma_{ti}} \tag{5.4}$$

Where, μ_{ti} is the mean of the sample, σ_{ti} is the standard deviation of the sample and t_{inm} is the journey time of bus trip n on day (or simulation run) m along route i.

Table 5.11 Reliability calculations for (Alsancak 311) district

Start point	District Name	Stop Code	No. of Bus Line Passing The Stop	No. of Boarding on The Stop	Time Mean Value "min."	Standard Deviation	Stop Reliability for One Day
Konak	Alsancak	10109	1	4163	10.76	2.95	3.65
Bostanlı	Alsancak	10110	1	4071	46.56	6.55	7.11
Konak	Alsancak	10111	1	4162	14.47	3.22	4.49

In this study, reliability has been measured in two different ways: the first one is by taking the average value of the stops reliability for each district as district reliability (Reliability 1). And the second way is by taking the weighted average of the stops reliability according to the bus lines that pass the stop (Reliability 2). According to Equation 5.4, a bigger value of RT indicates a better reliability. As it has been said before, the collecting data were for 61 working days during October 1st to December 31st 2012. To analyze and determine the reliability for each stop, Matlab based algorithm has been used. Table 5.12 presents the two travel time reliability and also it includes the normalized value of both Reliability 1 and Reliability 2. The following equation is the normalization formula that used in this study:

$$X_{nor} = \frac{X - \mu_x}{\sigma_x} \tag{5.5}$$

Where, X is the reliability value, μ_x is the average reliability for all districts and σ_x is the reliability standard deviation of all districts.

Table 5.12 The reliability values for all districts

District Code	District Name	Stop Number	Reliability 1	Reliability 2	Nor. Reliability 1	Nor. Reliability 2
24	Ahmet Taner Kışlalı	18	3.12	3.02	-0.14	-0.31
25	Atatürk	17	2.18	2.65	-1.41	-0.81
26	Egekent	24	2.78	2.78	-0.6	-0.64
27	Esentepe	6	3.23	3.29	0.01	0.04
28	Evka-2	25	2.67	2.64	-0.74	-0.82
29	Evka-5	15	3.41	3.71	0.25	0.59
30	Evka-6	8	3	2.95	-0.3	-0.41
31	Güzeltepe	22	2.84	2.84	-0.51	-0.55
32	İzkent	17	2.31	2.36	-1.24	-1.19
34	Yakakent	26	3.11	2.87	-0.16	-0.52
35	Aydınlıkevler	21	2.97	2.95	-0.34	-0.42
36	Balatçık	1	4.38	4.38	1.56	1.47
37	Çağdaş	18	2.62	2.69	-0.81	-0.76
38	İstasyonaltı	39	3.27	5.04	0.06	2.35
39	Köyiçi	14	2.88	2.85	-0.46	-0.54
40	Küçük Çiğli	6	3.5	3.1	0.38	-0.22
41	Maltepe	16	3.32	3.17	0.13	-0.13
42	Şirintepe	27	2.91	2.92	-0.41	-0.45
43	Yeni Mahalle	10	3.47	3.33	0.33	0.09
44	Aksoy	10	2.96	2.96	-0.36	-0.4
45	Alaybey	4	4.5	4.5	1.73	1.63
46	Atakent	15	3.44	3.33	0.29	0.09
47	Bahariye	1	1.51	1.51	-2.3	-2.31
48	Bahçelievler	20	3.02	3.08	-0.27	-0.24
49	Bahriye Üçok	8	2.41	2.72	-1.1	-0.72
50	Bostanlı	47	3.54	3.28	0.43	0.03
51	Dedebaşı	8	3.1	3.14	-0.17	-0.16
52	Demirköprü	3	3.99	3.99	1.03	0.95
53	Donanmacı	5	2.73	2.73	-0.66	-0.7
54	Fikri Altay	8	2.91	3.23	-0.42	-0.04
56	Mavişehir	5	3.49	3.47	0.37	0.27
57	Nergiz	25	3.1	2.88	-0.17	-0.51
58	Şemikler	6	3.28	3.43	0.08	0.22
59	Tersane	15	4.54	4.54	1.77	1.68
60	Tuna	2	5.04	4.88	2.45	2.14
61	Yalı	11	2.87	3	-0.48	-0.35
62	Cumhuriyet	25	3.09	3.19	-0.18	-0.09
63	Emek	36	2.98	3.05	-0.32	-0.29
64	Gümüşpala	27	2.78	2.91	-0.6	-0.46
65	İmbatlı	13	3.05	3.13	-0.23	-0.18
66	Onur	9	3.08	2.9	-0.19	-0.48
67	Org. Nafiz Gürman	30	2.76	2.71	-0.63	-0.73
68	Örnekköy	35	2.92	2.87	-0.41	-0.52
69	Postacılar	21	3.15	2.91	-0.1	-0.46
70	Soğukkuyu	13	2.84	2.81	-0.52	-0.6
72	Yamanlar	19	2.85	2.81	-0.51	-0.6
73	Zübeyde Hanım	20	2.83	2.86	-0.53	-0.53
78	Adalet	26	3.05	3.17	-0.24	-0.13
80	Bayraklı	11	3.41	3.4	0.25	0.18
84	Fuat Edip Baksi	18	3.13	3.09	-0.13	-0.23
88	75. Yıl	7	0	0	0	0
282	Çınarlı	3	2.92	2.88	-0.41	-0.51
288	Halkapınar	12	3.76	3.76	0.73	0.66
311	Alsancak	1	4.01	4.01	1.06	0.98
322	Kahramanlar	3	0	0	0	0
326	Kültür	14	3.04	3.53	-0.25	0.35
327	Mimar Sinan	8	3.65	3.59	0.57	0.43

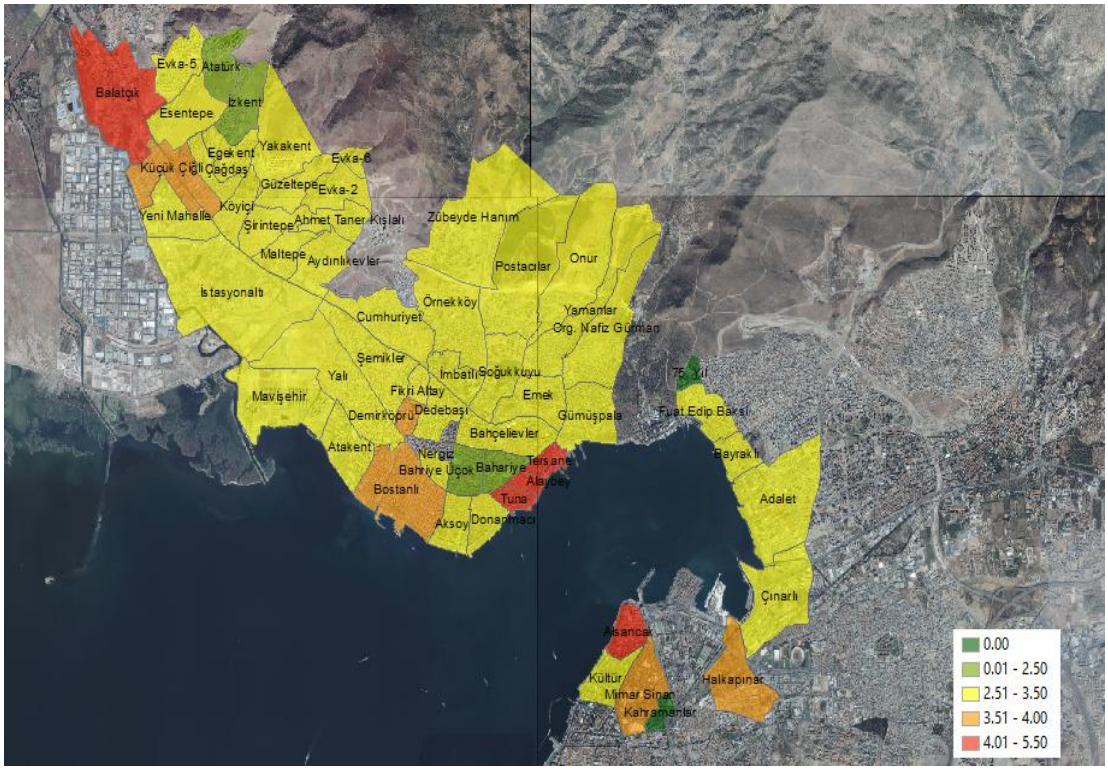


Figure 5.8 The Reliability 1 measures of the study area.

As Figure 5.8 shows, most of districts in the study area have reasonable values of Reliability 1. It means that those districts such as Alsancak and Alaybey districts have the highest reliability values because they are well serviced zones, which they have different modes of public transport. For instance in these two districts there are three modes: ferry station, commuter rail system (IZBAN) passes through these zones and there are more than 1 bus line service these areas. In addition to these three modes there are the sharing taxi system and the normal taxi, all of these modes increase the reliability values for these districts. Reliability has been measured as a function of travel time. In other words, varying of transportation modes help individuals to choose between public transport services which are save time and increase the reliability.

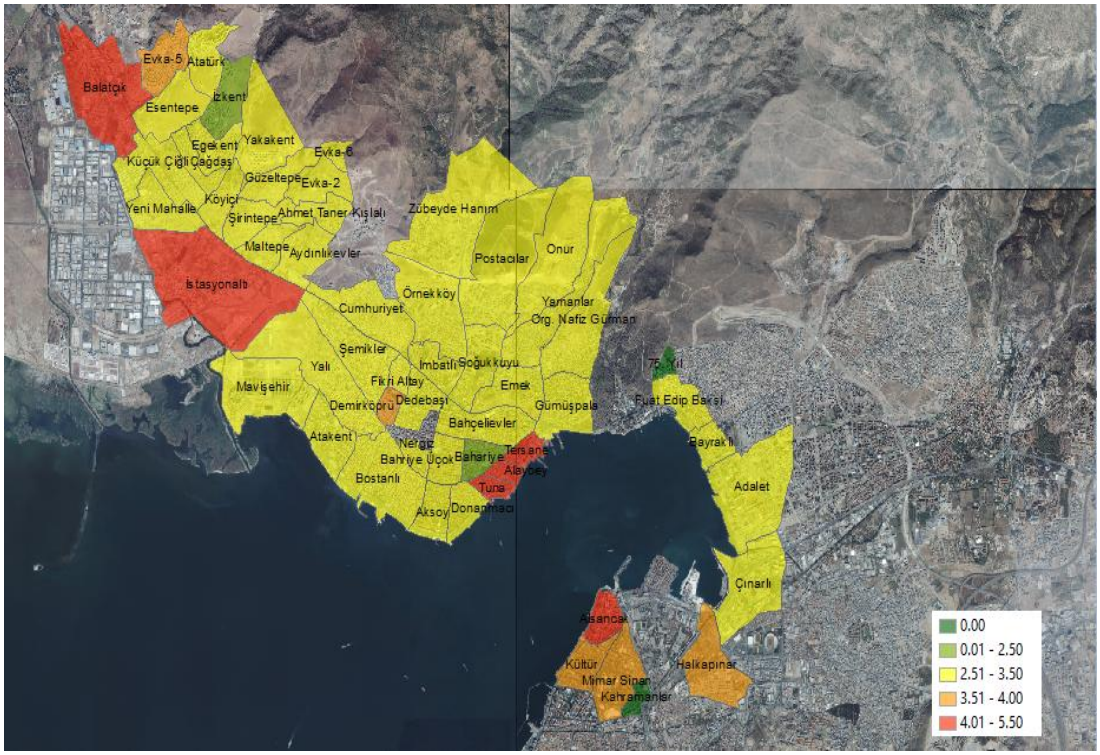


Figure 5.9 The Reliability 2 measures of the study area.

Approximately 41 districts as it seen in Figure 5.8 and 5.9, both reliability measures have reliability values between 2.5 and 3.5. That is because most of these districts have only the bus public transport system beside the taxi and minibus systems. In other words, the domain system that affects the reliability measures is the bus system for these 41 districts.

Both Kahramanlar and 75.Yıl districts have zero reliability values. Actually, the reliability measures are unknown. Because the reliability dataset that has been used in this study does not include any statistic for these two districts.

5.4 Statistical Data of Smart Card Use

Table 5.13 below contains the 17 statistical terms that have been used in this study. As it shown each term has defined clearly and denoted with a suitable shortcut, while Table 5.14 presents these statistical values.

Table 5.13 The statistical terms of the public transport smart card use

Shortcut	Clarification	Definition
ScNo	Smart card usage number	(the amount of smart cards usage for each district during the study duration 61 days)
ScNoWd	Smart card number in working days	(the amount of smart cards usage for each district in **working days** during the study duration 61 days)
ScNoWe	Smart card usage number in weekends	(the amount of smart cards usage for each district in **weekends** during the study duration 61 days)
WdScB	Working days smart card boarding	(the total S.C boarding number for each district in **working days** during the study duration 61 days)
WeScB	weekends smart card boarding	(the total S.C boarding number for each district in **weekends** during the study duration 61 days)
RWdScB	Rate of Working days smart card boarding	(**working days** boarding number over the total number of smart cards for each district)
RWeScB	Rate of weekends smart card boarding	(**weekends** boarding number over the total number of smart cards for each district)
RaScNoWd	Ratio of smart card usage number in working days	(the ratio of **working days** smart cards usage amount to the amount of smart cards for each district during the study duration 61 days)
UScNoWd	Unique smart card usage number in working days	(the unique number of smart cards usage for **working days**)
RaUScNoWd	Ratio of unique smart card usage number in working days	**(UScNoWd/ ScNoWd)** (the ratio of the unique usage number of smart cards for **working days** to the amount of smart cards usage for each district in **working days** during the study duration 61 days)
UScNoWe	Unique smart card usage number in weekends	(the unique usage number of smart cards for **weekends**)
RaUScNoWe	Ratio of unique smart card usage number in weekends	**(UScNoWe/ScNoWe)** (the ratio of the unique usage number of smart cards for **weekends** to the amount of smart cards for each district in **weekends** during the study duration 61 days)

Table 5.13 The statistical terms of the public transport smart card use (continue)

Shortcut	Clarification	Definition
UScNo	Unique smart card usage number	(the unique usage number of smart cards for the whole study duration 61 days)
RaUScNo	Ratio of unique smart card usage number	**(UScNo/ScNo)** (the ratio of the unique usage number of smart cards for the whole duration of study to the amount of smart cards for each district during the study duration 61 days)
RaWdUScB	Ratio of working days unique smart card boarding	**(WdScB /UScNoWd)** (the ratio of the total S.C boarding number for each district in **working days** during the study duration 61 days to the amount of smart cards usage for each district in **working days** during the study duration 61 days)
RaWeUScB	Ratio of weekends unique smart card boarding	**(WeScB /UScNoWe)** (the ratio of the total S.C boarding number for each district in **weekends** during the study duration 61 days to the amount of smart cards usage for each district in **weekends** during the study duration 61 days)
RaTUScB	Ratio of total unique smart card boarding	**(WdScB+WeScB)/ UScNo)** (the ratio of the total S.C boarding for each district of the whole study duration to the unique usage number of smart cards for the whole study duration 61 days)

The most important smart card statistics are the boarding ratios and the boarding rates, which they give a clear vision about the use of public transportation and about the daily use of the different systems of public transport. In other words, these statistics are the most credible predictors that play an important role especially when reliability and value of time are going to be examined. The amount of these boarding ratios and rates are varying from district to district as they shown in the following four figures.

The following figures show the ratios and rates of boarding for the working days and the weekends. It is shown that both ratios and rates of boarding are high for the districts that have low average income values. The districts with red color are in general normal residential districts (not a high class districts), which means the users in these zones are dependent on the public transport services and therefor boarding ratios and rates are high for these districts. While the sea side districts do not have a high boarding ratios and rates, because the users have the ability to choose between public transport services and private car or to take a taxi, in these zones.

Table 5.14 The statistical values of public transport smart card usage

District Code	District Name	ScNo	ScNoWd	ScNoWe	WdScB	WeScB	RWdScB	RWeScB	RaScNoWd	UScNoWd	RaUScNoWd	UScNoWe	RaUScNoWe	UScNo	RaUScNo	RaWdUScB	RaWeUScB	RaTUScB
24	Ahmet Taner Kışlalı	93343	70208	23135	211390	67120	3.011	2.901	0.752	10412	0.148	7234	0.313	17646	0.189	20.303	9.278	15.783
25	Atatürk	96473	70374	26099	228801	80549	3.251	3.086	0.729	10093	0.143	8139	0.312	18232	0.189	22.669	9.897	16.967
26	Egekent	36167	26657	9510	80011	26702	3.002	2.808	0.737	5766	0.216	3854	0.405	9620	0.266	13.876	6.928	11.093
27	Esentepe	11436	8483	2953	24179	8260	2.850	2.797	0.742	1673	0.197	1128	0.382	2801	0.245	14.452	7.323	11.581
28	Evka-2	98650	74464	24186	224907	70926	3.020	2.933	0.755	9283	0.125	6692	0.277	15975	0.162	24.228	10.599	18.518
29	Evka-5	160401	118171	42230	358111	124652	3.030	2.952	0.737	13517	0.114	11145	0.264	24662	0.154	26.493	11.185	19.575
30	Evka-6	7778	6232	1546	18983	4447	3.046	2.876	0.801	1035	0.166	601	0.389	1636	0.210	18.341	7.399	14.322
31	Güzeltepe	19355	14571	4784	46333	14448	3.180	3.020	0.753	2048	0.141	1594	0.333	3642	0.188	22.624	9.064	16.689
32	İzkent	76749	57675	19074	177672	55685	3.081	2.919	0.751	8067	0.140	5968	0.313	14035	0.183	22.025	9.331	16.627
34	Yakakent	121497	88313	33184	286671	103951	3.246	3.133	0.727	10330	0.117	8697	0.262	19027	0.157	27.751	11.953	20.530
35	Aydınlıkevler	46291	34795	11496	101772	31572	2.925	2.746	0.752	5626	0.162	3805	0.331	9431	0.204	18.090	8.298	14.139
36	Balatçık	114869	79532	35337	218727	94598	2.750	2.677	0.692	16256	0.204	12811	0.363	29067	0.253	13.455	7.384	10.779
37	Çağdaş	116974	86856	30118	267645	87118	3.081	2.893	0.743	11942	0.137	9207	0.306	21149	0.181	22.412	9.462	16.774
38	İstasyonaltı	536397	378490	157907	915216	372022	2.418	2.356	0.706	101481	0.268	72168	0.457	173649	0.324	9.019	5.155	7.413
39	Köyiçi	65019	47359	17660	108926	39154	2.300	2.217	0.728	14031	0.296	8856	0.501	22887	0.352	7.763	4.421	6.470
40	Küçük Çiğli	63203	45446	17757	116771	43598	2.569	2.455	0.719	11418	0.251	7827	0.441	19245	0.304	10.227	5.570	8.333
41	Maltepe	63958	48292	15666	121233	37837	2.510	2.415	0.755	13196	0.273	7408	0.473	20604	0.322	9.187	5.108	7.720
42	Şirintepe	20202	14794	5408	43533	14900	2.943	2.755	0.732	2666	0.180	2025	0.374	4691	0.232	16.329	7.358	12.456
43	Yeni Mahalle	96228	67716	28512	179449	72930	2.650	2.558	0.704	15337	0.226	12070	0.423	27407	0.285	11.700	6.042	9.209
44	Aksoy	60351	44334	16017	111461	36948	2.514	2.307	0.735	12536	0.283	7661	0.478	20197	0.335	8.891	4.823	7.348
45	Alaybey	219595	159888	59707	398920	142825	2.495	2.392	0.728	35844	0.224	26557	0.445	62401	0.284	11.129	5.378	8.682
46	Atakent	167725	129224	38501	357598	97763	2.767	2.539	0.770	31944	0.247	17119	0.445	49063	0.293	11.195	5.711	9.281
47	Bahariye	288069	194120	93949	444786	205946	2.291	2.192	0.674	81911	0.422	62014	0.660	143925	0.500	5.430	3.321	4.521
48	Bahçelievler	372417	274334	98083	743685	249930	2.711	2.548	0.737	62363	0.227	41901	0.427	104264	0.280	11.925	5.965	9.530
49	Bahriye Üçok	12794	9895	2899	26174	7404	2.645	2.554	0.773	3189	0.322	1635	0.564	4824	0.377	8.208	4.528	6.961
50	Bostanlı	635891	461291	174600	1101520	398665	2.388	2.283	0.725	113457	0.246	83361	0.477	196818	0.310	9.709	4.782	7.622
51	Dedebaşı	110312	82599	27713	225054	74443	2.725	2.686	0.749	21706	0.263	12720	0.459	34426	0.312	10.368	5.852	8.700
52	Demirköprü	17286	13400	3886	38413	9715	2.867	2.500	0.775	3660	0.273	1843	0.474	5503	0.318	10.495	5.271	8.746
53	Donanmacı	163156	116618	46538	246498	92077	2.114	1.979	0.715	45864	0.393	27596	0.593	73460	0.450	5.375	3.337	4.609
54	Fikri Altay	113172	84768	28404	230677	76311	2.721	2.687	0.749	25350	0.299	13205	0.465	38555	0.341	9.100	5.779	7.962
56	Mavişehir	86931	64877	22054	171082	52871	2.637	2.397	0.746	16998	0.262	11440	0.519	28438	0.327	10.065	4.622	7.875
57	Nergiz	89316	67533	21783	184817	56053	2.737	2.573	0.756	19763	0.293	11066	0.508	30829	0.345	9.352	5.065	7.813
58	Şemikler	469749	331413	138336	858718	347467	2.591	2.512	0.706	69602	0.210	53884	0.390	123486	0.263	12.338	6.448	9.768
59	Tersane	55819	42239	13580	109315	33818	2.588	2.490	0.757	9419	0.223	5938	0.437	15357	0.275	11.606	5.695	9.320
60	Tuna	382444	263368	119076	638398	298214	2.424	2.504	0.689	108425	0.412	80990	0.680	189415	0.495	5.888	3.682	4.945
61	Yalı	48768	34449	14319	75058	28521	2.179	1.992	0.706	12433	0.361	8404	0.587	20837	0.427	6.037	3.394	4.971
62	Cumhuriyet	64861	47656	17205	156278	54821	3.279	3.186	0.735	7625	0.160	5956	0.346	13581	0.209	20.495	9.204	15.544
63	Emek	34617	26142	8475	78095	23988	2.987	2.830	0.755	5224	0.200	3423	0.404	8647	0.250	14.949	7.008	11.806
64	Gümüşpala	78726	60135	18591	188811	55637	3.140	2.993	0.764	10537	0.175	7462	0.401	17999	0.229	17.919	7.456	13.581
65	İmbatlı	7236	5187	2049	9018	3789	1.739	1.849	0.717	2157	0.416	1376	0.672	3533	0.488	4.181	2.754	3.625
66	Onur	148920	112976	35944	380127	114849	3.365	3.195	0.759	13252	0.117	10437	0.290	23689	0.159	28.685	11.004	20.895
67	Org. Nafiz Gürman	110933	83593	27340	270200	86387	3.232	3.160	0.754	10246	0.123	7563	0.277	17809	0.161	26.371	11.422	20.023
68	Örnekköy	75020	56628	18392	170019	52205	3.002	2.838	0.755	10339	0.183	7063	0.384	17402	0.232	16.444	7.391	12.770
69	Postacılar	92324	71992	20332	235578	63005	3.272	3.099	0.780	9371	0.130	6698	0.329	16069	0.174	25.139	9.407	18.581
70	Soğukkuyu	113419	85997	27422	249590	75500	2.902	2.753	0.758	22936	0.267	13075	0.477	36011	0.318	10.882	5.774	9.028
72	Yamanlar	147602	108627	38975	368763	127200	3.395	3.264	0.736	14560	0.134	12000	0.308	26560	0.180	25.327	10.600	18.673
73	Zübeyde Hanım	146643	108966	37677	343714	112556	3.154	2.987	0.743	13009	0.119	10609	0.282	23618	0.161	26.421	10.609	19.319
78	Adalet	341903	255624	86279	694514	228575	2.717	2.649	0.748	64451	0.252	38625	0.448	103076	0.301	10.776	5.918	8.955
80	Bayraklı	389596	266949	122647	671082	304101	2.514	2.479	0.685	68585	0.257	53773	0.438	122358	0.314	9.785	5.655	7.970
84	Fuat Edip Baksi	64160	50319	13841	144560	38543	2.873	2.785	0.784	8696	0.173	5235	0.378	13931	0.217	16.624	7.363	13.144
88	75. Yıl	79248	58000	21248	191946	68053	3.309	3.203	0.732	7038	0.121	5652	0.266	12690	0.160	27.273	12.041	20.488
282	Çınarlı	282088	219803	62285	509034	145902	2.316	2.342	0.779	94447	0.430	42074	0.676	136521	0.484	5.390	3.468	4.797
288	Halkapınar	1279	964	315	1761	617	1.827	1.959	0.754	610	0.633	270	0.857	880	0.688	2.887	2.285	2.702
311	Alsancak	188453	132107	56346	308365	127992	2.334	2.272	0.701	57303	0.434	37345	0.663	94648	0.502	5.381	3.427	4.610
322	Kahramanlar	16373	12669	3704	33741	10037	2.663	2.710	0.774	3437	0.271	1733	0.468	5170	0.316	9.817	5.792	8.468
326	Kültür	590183	400645	189538	822901	381465	2.054	2.013	0.679	158479	0.396	109706	0.579	268185	0.454	5.192	3.477	4.491
327	Mimar Sinan	36607	26678	9929	50262	18330	1.884	1.846	0.729	13307	0.499	6609	0.666	19916	0.544	3.777	2.773	3.444

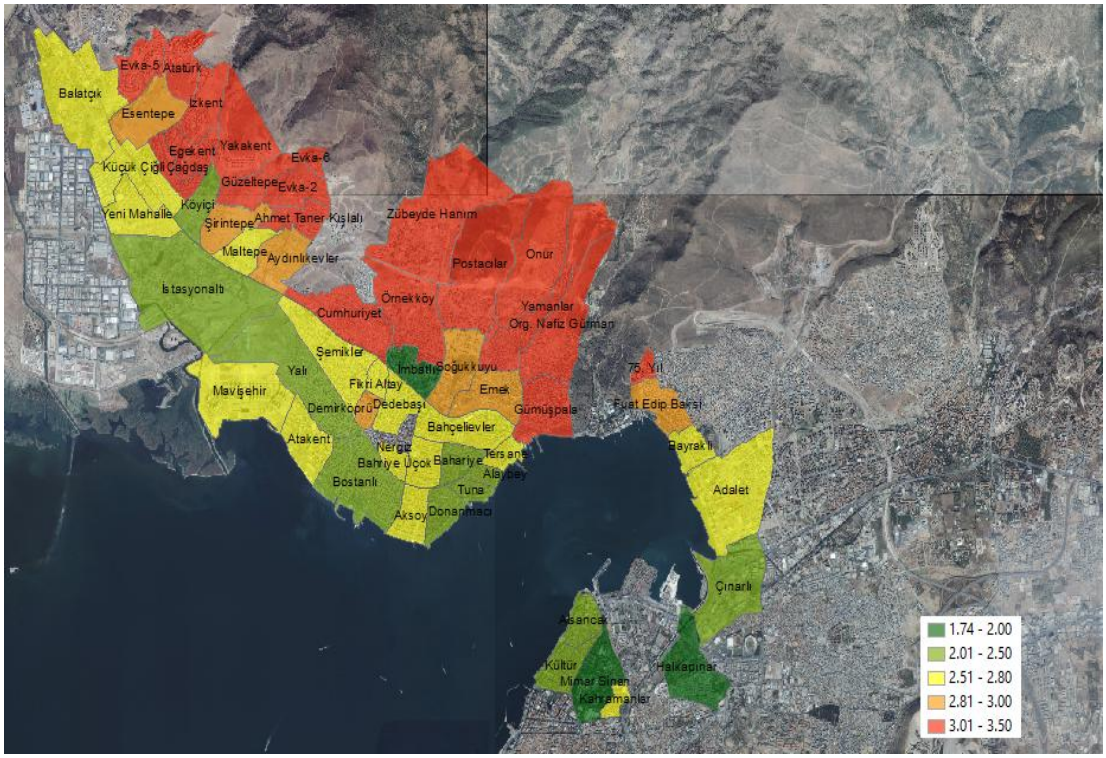


Figure 5.10 The RWdScB for all districts in study area.

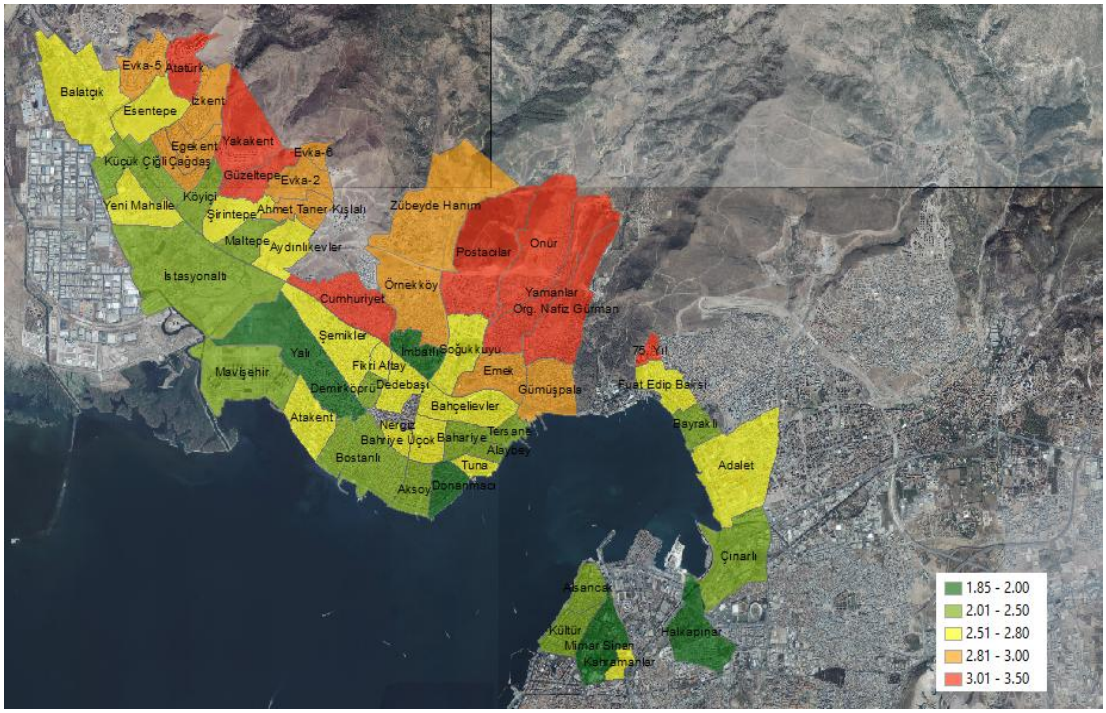


Figure 5.11 The RWeScB for all districts in study area.

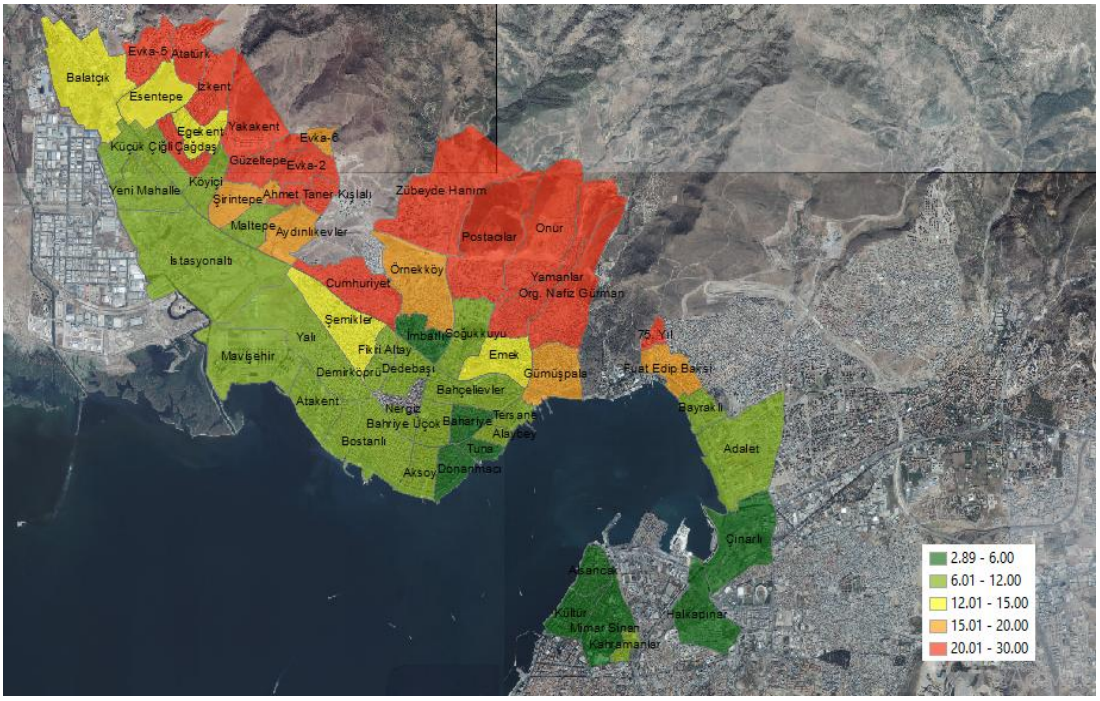


Figure 5.12 The RaWdUScB for all districts in study area.

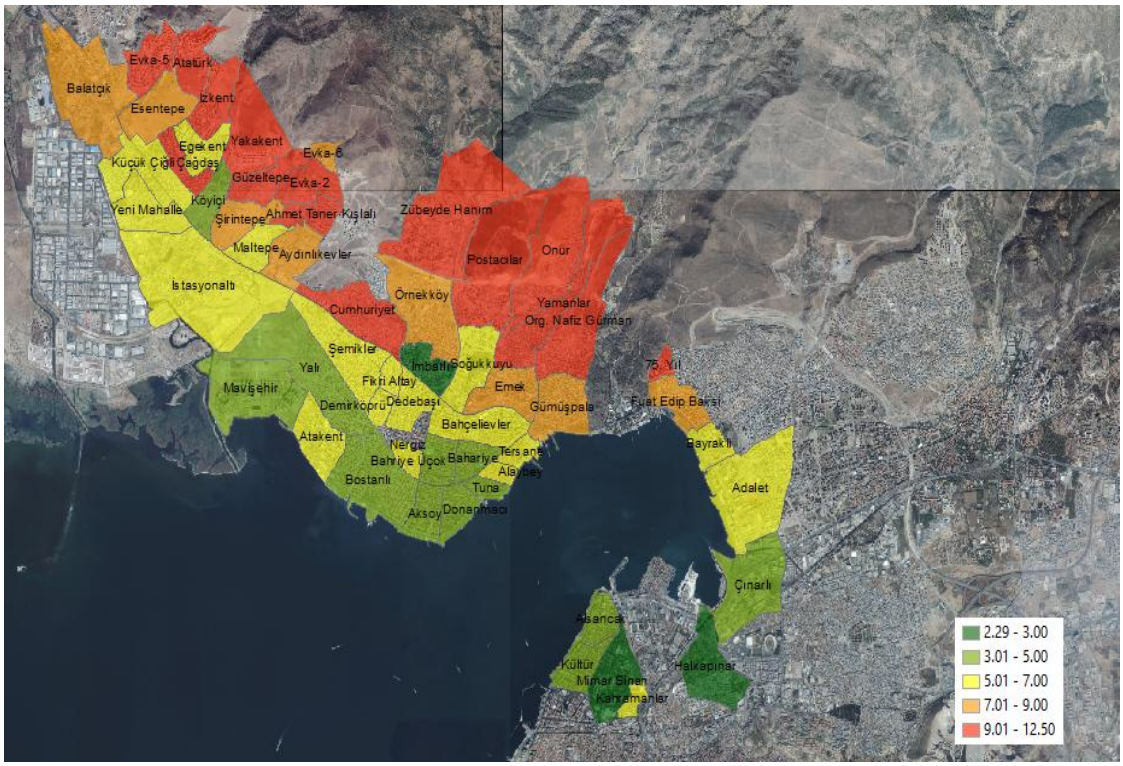


Figure 5.13 The RaWeUScB for all districts in study area.

5.5 Correlation

Correlation is a statistical technique that can show whether and how strongly variables are related.Formally, dependence refers to any situation in which random variables do not satisfy a mathematical condition of probabilistic independence. In loose usage, correlation can refer to any departure of two or more random variables from independence, but technically it refers to any of several more specialized types of relationship between mean values. There are several correlation coefficients, often denoted ρ or r, measuring the degree of correlation. The most common of these is the Pearson correlation coefficient, which is sensitive only to a linear relationship between two variables (which may exist even if one is a nonlinear function of the other).

5.5.1 Pearson's Product-moment Coefficient

The most familiar measure of dependence between two quantities is the Pearson product-moment correlation coefficient, or (Pearson's correlation coefficient), commonly called simply the correlation coefficient. It is obtained by dividing the covariance of the two variables by the product of their standard deviations. Karl Pearson developed the coefficient from a similar but slightly different idea by Francis Galton (Rodgers and Nicewander, 1988).

The population correlation coefficient $\rho_{X,Y}$ between two random variables X and Y with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y is defined as:

$$\rho_{x,y} = corr(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x \sigma_y} \qquad (5.6)$$

Where, *E* is the expected value operator, cov means covariance and, corr a widely used alternative notation for the correlation coefficient.

The Pearson correlation is defined only if both of the standard deviations are finite and nonzero. It is a corollary of the Cauchy–Schwarz inequality that the

correlation cannot exceed 1 in absolute value. The correlation coefficient is symmetric: corr(X,Y) = corr(Y,X).

The Pearson correlation is +1 in the case of a perfect direct (increasing) linear relationship (correlation), −1 in the case of a perfect decreasing (inverse) linear relationship (anticorrelation) (Dowdy and Wearden, 1983), and some value between −1 and 1 in all other cases, indicating the degree of linear dependence between the variables. As it approaches zero there is less of a relationship (closer to uncorrelated). The closer the coefficient is to either −1 or 1, the stronger the correlation between the variables.

If the variables are independent, Pearson's correlation coefficient is 0, but the converse is not true because the correlation coefficient detects only linear dependencies between two variables. For example, suppose the random variable X is symmetrically distributed about zero, and $Y = X^2$. Then Y is completely determined by X, so that X and Y are perfectly dependent, but their correlation is zero; they are uncorrelated. However, in the special case when X and Y are jointly normal, uncorrelatedness is equivalent to independence.

In this study correlation has been used to show the weather and how strongly variables are related. The variables that have been examined in this section are: Reliability 1, Reliability 2, VOT1, VOT2, VOT 3, average income, development level and finally the different smart card statistics. SPSS has been used to calculate the correlation, which gave the following correlations matrix that contains all variables relates to this study. Here in the following Table 5.15, correlation is significant at the 0.05 level (so that the variables having less than 0.05 are significant). Pearson correlation sign (positive or negative) shows if variables are correlated positively or negatively with each other.

Table 5.15 Correlations matrix for the study's main measures and the statistical usage of public transport smart card.

Correlations

		ScNo	ScNoWd	ScNoWe	WdScB	WeScB	RWdScB	RWeScB	RaScNoWd	UScNoWd	RaUScNoWd	UScNoWe	RaUScNoWe	UScNo	RaUScNo	RaWdUScB	RaWeUScB	RaTUScB
VOT1	Pearson Correlation	-.101	-.104	-.095	-.115	-.108	-.277*	-.222	.069	-.076	.447**	-.075	.409**	-.076	.428**	-.172	-.201	-.183
	Sig. (2-tailed)	.453	.442	.483	.395	.423	.037	.097	.608	.576	.000	.579	.002	.576	.001	.202	.133	.173
	Spearman's rho	.114	.125	.112	.124	.105	-.154	-.152	-.021	.122	.126	.114	.167	.122	.136	-.149	-.151	-.154
	Sig. (2-tailed)	.400	.356	.406	.359	.439	.253	.258	.878	.367	.351	.399	.215	.366	.314	.269	.261	.254
VOT2	Pearson Correlation	.354**	.353**	.354**	.323*	.321*	-.155	-.175	-.148	.442**	.206	.414**	.209	.432**	.211	-.179	-.204	-.187
	Sig. (2-tailed)	.007	.007	.007	.014	.015	.250	.192	.273	.001	.125	.001	.119	.001	.115	.183	.127	.163
	Spearman's rho	.297*	.305*	.300*	.307*	.286*	-.100	-.113	-.073	.368**	.151	.317*	.207	.352**	.170	-.162	-.172	-.172
	Sig. (2-tailed)	.025	.021	.023	.020	.031	.458	.404	.588	.005	.261	.016	.122	.007	.205	.228	.202	.201
VOT3	Pearson Correlation	.195	.207	.167	.179	.143	-.255	-.295*	.004	.267*	.250	.217	.293*	.247	.260	-.272*	-.305*	-.280*
	Sig. (2-tailed)	.145	.123	.215	.182	.288	.056	.026	.974	.045	.061	.104	.027	.064	.051	.041	.021	.035
	Spearman's rho	.260	.272*	.268*	.232	.219	-.360**	-.405**	-.041	.328*	.321*	.294*	.376**	.316*	.328*	-.315*	-.390**	-.335*
	Sig. (2-tailed)	.051	.041	.044	.083	.101	.006	.002	.761	.013	.015	.026	.004	.017	.013	.017	.003	.011
Average Income	Pearson Correlation	.336*	.330*	.347**	.271*	.294*	-.484**	-.525**	-.164	.413**	.448**	.411**	.461**	.413**	.455**	-.521**	-.527**	-.521**
	Sig. (2-tailed)	.011	.012	.008	.042	.027	.000	.000	.222	.001	.000	.001	.000	.001	.000	.000	.000	.000
	Spearman's rho	.231	.233	.230	.174	.175	-.512**	-.535**	-.136	.387**	.502**	.318*	.532**	.365**	.514**	-.517**	-.530**	-.520**
	Sig. (2-tailed)	.084	.082	.085	.195	.194	.000	.000	.312	.003	.000	.016	.000	.005	.000	.000	.000	.000
Devlopment Level	Pearson Correlation	-.360**	-.355**	-.367**	-.292*	-.317*	.595**	.605**	.199	-.466**	-.623**	-.449**	-.632**	-.461**	-.628**	.613**	.624**	.619**
	Sig. (2-tailed)	.006	.007	.005	.027	.016	.000	.000	.138	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Spearman's rho	-.256	-.259	-.253	-.211	-.208	.608**	.622**	.152	-.430**	-.623**	-.351**	-.659**	-.402**	-.634**	.632**	.654**	.635**
	Sig. (2-tailed)	.055	.052	.058	.115	.121	.000	.000	.260	.001	.000	.007	.000	.002	.000	.000	.000	.000
Reliability 1	Pearson Correlation	.180	.179	.180	.166	.181	-.268*	-.286*	-.170	.187	.206	.195	.233	.191	.220	-.269*	-.294*	-.278*
	Sig. (2-tailed)	.181	.182	.181	.217	.179	.044	.031	.206	.163	.123	.145	.081	.154	.100	.043	.026	.037
	Spearman's rho	.137	.135	.146	.100	.122	-.362**	-.387**	-.179	.264*	.229	.205	.254	.235	.250	-.262*	-.312*	-.279*
	Sig. (2-tailed)	.311	.318	.280	.461	.365	.006	.003	.183	.047	.086	.125	.057	.078	.060	.049	.018	.036
Reliability 2	Pearson Correlation	.289*	.285*	.295*	.266*	.292*	-.299*	-.312*	-.233	.285*	.219	.294*	.236	.290*	.232	-.296*	-.311*	-.304*
	Sig. (2-tailed)	.029	.032	.026	.046	.028	.024	.018	.081	.032	.101	.026	.077	.029	.082	.026	.018	.021
	Spearman's rho	.175	.172	.189	.129	.158	-.459**	-.471**	-.271*	.328*	.341**	.271*	.342**	.298*	.359**	-.377**	-.393**	-.387**
	Sig. (2-tailed)	.194	.202	.159	.340	.240	.000	.000	.042	.013	.009	.042	.009	.024	.006	.004	.002	.003

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

The most important predictors in this study are the ratios of unique smart card boarding statistics (RaWdUScB, RaWeUScB and RaTUScB) which affect both value of time and reliability. The above correlations matrix shows that VOT 3 is correlated negatively with the ratio of unique smart card boarding for both working days and weekends, which means when the value of time gets bigger or more significant, the ratio of boarding is getting decrease. It also the same for average income, it is correlated negatively with the ratio of unique smart card boarding. In other words, districts that have a high average income mostly the value of time become more significant for them, because the individuals are going to save time by using another modes, such as their private car, taxi specially for the working days. This is logical and reasonable for the value of time case. Whereas, it is illogical for the reliability measures. As it is shown in the correlations matrix, both Reliability 1 and Reliability 2 are correlated negatively with the ratios of unique smart card boarding statistics(RaWdUScB, RaWeUScB and RaTUScB), that mean the boarding ratios will get higher whenever the reliability less.

Linear regression has been done to examine the correlation and the relationship between smart card statistics and both reliability and value of time. Statistics such as working days smart card boarding (WdScB), rate of working days smart card boarding (RWdScB), Ratio of unique smart card number in working days (RaUScNoWd) and ratio of working days unique smart card boarding “RaWdUScB” have been used separately as a dependent variable for the linear regression model, while (VOT 1, VOT 2, VOT 3, Reliability 1, Reliability 2, Nor. Reliability 1 and Nor. Reliability 2) are used as independent variables. Finally, the backward method has been choosing for these models.

5.6 Linear Regression

Linear regression analysis is the most widely used one of all statistical techniques and it is the next step after correlation. Linear regression is used to additive relationships between variables, usually under the assumption of independently and identically normally distributed errors. It is used to predict the value of a variable

based on the value of other variables. The variable that wants to predict is called the dependent variable (or sometimes, the outcome variable). The variables that are used to be predicted the other variable's value is called the independent variable (or sometimes, the predictor variable). Let Y denote the dependent variable, and let X_1, …,X_n denote the independent variables. Then the equation for predicting the value of Y is of the form:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + ...\beta_n X_n (5.7)$$

This formula has the property that the prediction for Y is a straight-line function of each of the X variables, holding the others fixed, and the contributions of different X variables to the predictions are additive. The slopes of their individual straight-line relationships with Y are the constants β_1, β_2, …, βn, the so-called coefficients of the variables. The additional constant α, the so-called intercept, is the prediction that the model would make if all the X's were zero (if that is possible). The coefficients and intercept are estimated by least squares, i.e., setting them equal to the unique values that minimize the sum of squared errors within the sample of data to which the model is fitted.

The main objective of using the linear regression is to predict public transport smart card use by using value of time and reliability measures. Some of illogical relationships direct the study to examine the reverse relation. As it shown in the correlations matrix, both Reliability 1 and Reliability 2 are negatively correlated with the ratios of unique smart card boarding statistics (RaWdUScB, RaWeUScB and RaTUScB) which is illogical and it doesn't make sense. To examine this relationship more clearly linear regression has been used. SPSS Statistics software took a place to determine the variables that are correlating with each other and then Excel solver has been used again to examine these correlated variables. The following tables present the linear regression results of some variables. Linear regression of each variable is going to examine by these values (R square, t-test, F-test and P-value).

Table 5.16 presents the correlation matrix of the study main variables (VOT 1, VOT 2, VOT 3, Average Income, Development Level, Reliability 1 and Reliability 2). It is for preventing the use of correlated variables as predictors in the same regression. As it is shown, both VOT 2 and VOT 3 are correlated with Average Income and Development Level. Because of this relationship one of these four variables are going to be used in the same linear regression model. For instance, if VOT 2 and Average Income are significant at the same model, one of them should be eliminated. Besides, in this study variables that have illogical relationship are going to eliminate. This rule is also valid for both of Reliability 1 and Reliability 2. Only one of these to reliability measures can be used in the same linear regression model.

Table 5.16 Correlations matrix of the study main measures.

Correlations

		VOT 1	VOT2	VOT 3	Average Income	Ddevlopment Level	Reliability 1	Reliability 2
VOT 1	Pearson Correlation	1	.082	-.005	.094	-.246	.021	.018
	Sig. (2-tailed)		.544	.973	.489	.066	.879	.895
	N	57	57	57	57	57	57	57
VOT 2	Pearson Correlation	.082	1	.086	.387**	-.387**	.043	.050
	Sig. (2-tailed)	.544		.525	.003	.003	.751	.709
	N	57	57	57	57	57	57	57
VOT 3	Pearson Correlation	-.005	.086	1	.353**	-.387**	.189	.158
	Sig. (2-tailed)	.973	.525		.007	.003	.159	.240
	N	57	57	57	57	57	57	57
Average Income	Pearson Correlation	.094	.387**	.353**	1	-.909**	.086	.088
	Sig. (2-tailed)	.489	.003	.007		.000	.525	.517
	N	57	57	57	57	57	57	57
Ddevlopment Level	Pearson Correlation	-.246	-.387**	-.387**	-.909**	1	-.143	-.131
	Sig. (2-tailed)	.066	.003	.003	.000		.290	.330
	N	57	57	57	57	57	57	57
Reliability 1	Pearson Correlation	.021	.043	.189	.086	-.143	1	.943**
	Sig. (2-tailed)	.879	.751	.159	.525	.290		.000
	N	57	57	57	57	57	57	57
Reliability 2	Pearson Correlation	.018	.050	.158	.088	-.131	.943**	1
	Sig. (2-tailed)	.895	.709	.240	.517	.330	.000	
	N	57	57	57	57	57	57	57

**. Correlation is significant at the 0.01 level (2-tailed).

5.6.1 WdScB as Dependent Variable

R Square equals 0.62 which is a good fit. 62% of the variation in WdScB is explained by the independent variables VOT 2 and Reliability 2. The closer R squares to 1, the better the regression line. Significance F is 3.5E-12 (less than 0.05) which means that the results are reliable (statistically significant). P-value also should be less than 0.05 and t-test should be more than 2 absolute. According to the above linear regression results, the dependent variable working days smart cards boarding (WdScB) is affected by VOT 2 and Reliability 2. The regression line is:

$$y = \text{WdScB} = 1921245\ \text{VOT 2} + 67635\ \text{Reliability2} \quad (5.8)$$

In other words, for each unit increase in Reliability 2, (WdScB) increases with 67635 units. And for one unit increase in VOT 2, (WdScB) increases with 1921245 units. This model does not include the intercept, because after the first regression the other variables and the intercept had a t-test less than 2 absolute, therefor, these insignificant variables have been eliminated in this final model.

As Table 5.17 shows, Reliability 2 has a logical relationship with the (WdScB). Which means that the boarding in working days are going to increase if the reliability measure satisfies the individuals' needs. While, the value of time which it represented by VOT 2 in this model is correlated illogically with "WdScB". Illogically here means the boarding in working days will increase even if the individuals' value of time is high also. In other words, individuals are not going to take other options (taxi, private car) to go to work or to school, this model shows that they are going to take the public transport even if it crowded or they are going to wait the next service.

Table 5.17 Summary output (Dependent variable is WdScB)

Regression Statistics	
Multiple R	0.78757497
R Square	0.620274333
Adjusted R Square	0.595188412
Standard Error	232169.5003
Observations	57

ANOVA

	df	*F*	*Significance F*
Regression	2	44.92070375	3.24912E-12
Residual	55		
Total	57		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
VOT 2	1921245.119	762859.7448	2.518477521	0.014725093
Reliability 2	67635.37588	12550.71128	5.388967554	1.52605E-06

5.6.2 RaWdUScB as Dependent Variable

As it has been discussed before, the most important smart card statistics are the ratios and the most important of these ratios are the ratios of unique smart cards boarding (RaWdUScB, RaWeUScB and RaTUScB). The linear regression took a place in this study when these three ratios correlate negatively with both Reliability 1 and Reliability 2, while the logical and the reasonable is to correlate positively.

Table 5.18 Summary output (Dependent variable is RaWdUScB)

Regression Statistics	
Multiple R	0.650279429
R Square	0.422863336
Adjusted R Square	0.401487904
Standard Error	5.682307229
Observations	57

ANOVA

	df	*F*	*Significance F*
Regression	2	19.78268026	3.58529E-07
Residual	54		
Total	56		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	10.88619197	3.482766462	3.125731251	0.002853752
Development Level	2.682730076	0.478893508	5.601934529	7.35165E-07
Reliability 2	-1.8992974	0.904483985	-2.09986847	0.040427607

According to the following two tables also, (RaWdUScB, RaWeUScB and RaTUScB) still have negative relationships with Reliability 2, which means that the ratio of boarding will increase even if reliability measures are low. It seems that individuals do not have other options to choose from, so they are taking the public transport in any case.

5.6.3 RaWeUScB as Dependent Variable

According to Table 5.19 reliability is less efficient comparing with development level and intercept. In other words, the use of public transport increases when the provided service is confortable and reliable (in this study reliable means on time). But in this regression, the ratio of unique smart cart boarding increases even if the average income is high at the same time reliability is low. The only logical explanation is that districts that have high average income are the city center or/and the city facilities are located there, for that people from different districts are going there and using the public transport from these districts. For this reason the ratios are increases even they have a low reliability. This means that these smart card statistics are not the real predictor of the individuals of these districts.

Table 5.19 Summary output (Dependent variable is RaWeUScB)

Regression Statistics	
Multiple R	0.665457312
R Square	0.442833434
Adjusted R Square	0.422197635
Standard Error	1.995983676
Observations	57

ANOVA

	df	*F*	*Significance F*
Regression	2	21.45947626	1.38547E-07
Residual	54		
Total	56		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	5.620541356	1.223366623	4.594322952	2.64524E-05
Development Level	0.974046505	0.168217519	5.790398712	3.68927E-07
Reliability 2	-0.723421396	0.317711661	-2.276974641	0.026774192

5.6.4 RaTUScB as Dependent Variable

As it shown from the last three tables development level is correlated positively with these ratios, and that is logical, because in the highest development level districts live poor people and these people do not have other option to choose from, therefor, the boarding ratios are going to increase if development level is increasing also. It has been explained before those districts that have a high value of development level; do not have good services as the districts that have a low value of development level.

Table 5.20 Summary output (Dependent variable is RaTUScB)

Regression Statistics	
Multiple R	0.658663214
R Square	0.433837229
Adjusted R Square	0.412868238
Standard Error	3.994820108
Observations	57

ANOVA

	df	*F*	*Significance F*
Regression	2	20.68946563	2.13509E-07
Residual	54		
Total	56		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	8.821752374	2.448481741	3.602947993	0.00068561
Development Level	1.921213683	0.336675462	5.706426211	5.01882E-07
Reliability 2	-1.394858987	0.635877411	-2.193597324	0.032588955

Again the reliability still has negative relationships with these three ratios. The last thing that should check now is the relationships if the dependent variables are changed. In other words, reliability and other variables have to be checked if they are under the effect of smart card statistics or vice versa. The next tables present the results of linear regression when dependent variables have been changed.

5.6.5 Reliability 1 as Dependent Variable

According to Table 5.21 reliability 1 is not under effect. Because variables insignificant and F test fails, reliability 1 is not affected by public transport smart card use.

Table 5.21 Summary output (Dependent variable is Reliability 1)

Regression Statistics	
Multiple R	0.296108744
R Square	0.087680388
Adjusted R Square	0.053890773
Standard Error	0.799574632
Observations	57

ANOVA

	df	*F*	*Significance F*
Regression	2	2.59489158	0.083940063
Residual	54		
Total	56		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4.48203841	0.892005246	5.02467718	5.87316E-06
RWeScB	-0.565662594	0.320470365	-1.7651011	0.08320115
UScNoWd	1.99589E-06	3.42668E-06	0.58245737	0.56268290

5.6.6 Reliability 2 as Dependent Variable

Here in Table 5.22, reliability 2 is almost correlates with the weekends smart card boarding which it has an insignificant value around 2.9E-12. While, it has a negative relationship with the rate of weekends smart cards boarding (RWeScB). But it has a significant positive coefficient value for the ratio of smart card number of working days (RaScNoWd), these force transit operators to make reliable service. If weekend use increase, reliability decreases. Then weekend use is more effective on reliability.

Table 5.22 Summary output (Dependent variable is Reliability 2)

Regression Statistics	
Multiple R	0.969599174
R Square	0.940122559
Adjusted R Square	0.91938635
Standard Error	0.803433084
Observations	57

ANOVA

	df	*F*	*Significance F*
Regression	3	282.614047	1.421E-32
Residual	54		
Total	57		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
WeScB	2.9102E-06	1.02825E-06	2.83026100	0.006518339
RWeScB	-0.68510199	0.317676363	-2.15660361	0.035507075
RaScNoWd	6.2173729	1.155729837	5.37960755	1.64771E-06

CHAPTER SIX
CONCLUSION

Public transport agencies are increasingly realizing the benefits of collecting and analyzing operational data recorded by public transportation smart card, and using it to increase the quality of their service. This work represents an attempt at quantifying public transport service reliability in Izmir/ Turkey. By using the available data, combined with a small set of observations, some measures of reliability was calculated. Data for public transport bus travel time along each bus line that passes the study area was available on a whole day basis for the study duration (three months, from October 1st to December 31st 2012). Value of time measured as the ratio between the Lagrange multiplier of time restriction μ and the Lagrange multiplier of income restriction λ. While, public transportation usage represented by boarding rates and ratios for both working days and weekends.

Calculations show that people of Izmir prefer to take public transport bus or to wait the next service than use their private vehicles or ride a taxi. The reliability measures in this study do not affect the consuming of public transport use. In other words, the boarding on public transport services increase even if the reliability measures are low and do not satisfy the passengers' needs (travel time). Most of public transport users are students and officers (at least five days a week), these users usually do not have private vehicles and they cannot ride taxies also because it is too much expensive comparing with the public transport cost. As a result, individuals prefer to ride a crowed service or to loss some time than ride a taxi or another option.

The negative relationship between the reliability measures and the rate/ratio of public transport boarding can be also affected by the passengers boarding time during they are getting the public transport service or/and when they are leaving the public transport service. In other words, both of boarding and departure time affect the reliability measures of the public transport service because not only the travel time is going to increase, but also trip time is going to increase. That means the

reliability measures which it is a function of travel time only in this study are going to decrease.

Reliability measures as this research illustrated are high for the city center and commercial zones even if they have high rating of public transport boarding such as Alsancak district. This is because these zones are well served with different kinds of public transport systems, this variation of public transport systems give individuals an opportunity to choose from different public transport modes to serve their needs. On other hand, this relieves the pressure public transport bus service, which reduces the trip time and that definitely increases reliability measures.Also, districts that are located far from city center such as Balatçık district, have high reliability measures because of these districts normally contain the last stations of public transport bus line, which means that the public transport bus departures at planned time. Usually, there are no hindrances at the first stations to delay the bus movement when it is going to the city center. Another explanation for these far district may be; usually the people who living these districts are retired and do not have daily jobs in the city center (point of view). For that, they do not use public transport too much and when they use it they do not care about reliability measures because they enough time to wait.

The last thing that should be clarified in this study is the value of time. As it shown in chapter 5, value of time has insignificant small values, and that is because value of time in this study has been calculated as a function of trip cost which it is normally calculated as a function of individual's wage rate.

REFERENCES

Abkowitz, M., & Tozzi, T. (1987). Research contributions to managing transit service reliability. *Journal Advanced Transportation, 21*, 47-65.

Agard, B., Morency, C., &Trépanier, M. (2006).Mining public transport user behaviour from smart card data.*12th IFAC Symposium on Information Control Problems in Manufacturing – INCOM 2006*, Saint-Etienne, France, May 17-19.

Bagchi, M., & White, P. (2005).The potential of public transport smart card data. *Transport Policy, 12*, 464-474.

Bates, J., Polak, J., Jones, P., & Cook, A. (2001).The valuation of reliability for personal travel.*Transportation Research E 37*, 191–229.

Becker, G. (1965). A theory of the allocation of time.*The Economic Journal, 75*, 493-517.

Beesley, M. (1965). The value of time in travelling: Some new evidence. *Economica, 32*, 174-185.

Ben-Akiva, M., & Lerman, S. R. (1985).*Discrete choice analysis*.Cambridge, MA: The MIT Press.

Biørn, E. (2010). *Four lecture notes, part of syllabus for master course ECON4160: Econometrics* - Modelling and Systems Estimation.

Bowman, L.,& Turnquist, M. (1981). Service frequency, schedule reliability and passenger wait times at transit stops. *Transportation Research A, 15A*, 465-471.

Bruno, A., & Catherine, M. (2007).Mining public transport user behavior from smart card data.*Cirrelt*, 42.

Calfee, J., & Winston, C. (1998). The value of automobile travel time: implications for congestion policy. *Journal of Public Economics*, 83-102.

Camus, R., Longo, G., &Macorini, C. (2005). Estimation of transit reliability level-of-service based on automatic vehicle location data. *Journal of the Transportation Research Board, 1927*(-1), 277-286.

Chapman, R. (1976). Bus reliability – definition and measurement.Research report no.18.*Transport Operations Research Group, University of Newcastle upon Tyne, UK.*

Chen, C., Skabardonis, A., &Varaiya, P. (2003).Travel-time reliability as a measure of service.*Transportation Research Record, 1855*, 74-79.

Christopher, M. (2006). In the terminology of statistics.*Pattern Recognition and Machine Learning*.Springer, 205.

Currie, G., Douglas, N. J., & Kearns, I. (2012).An Assessment of alternative bus reliability indicators.*Australasian Transport Research*, Perth, Australia.

Deakin, E., & Kim, S. (2001). Transportation technologies: implications for planning. *University of California Transportation Center paper, 536*, 27.

DeSerpa, T. (1971).A theory of the economics of time.*The Economic Journal, 81*, 828-846.

Dowdy, S. &Wearden, S. (1983). *Statistics for research*. New York:Wiley.

Evans, A., (1972). On the theory of the valuation and allocation of time.*Scottish Journal of Political Economy, 19* (1), 1-17.

ESHOT General Directorate of İzmir Public Bus System (ESHOT) (2012).Dataset.İzmir.

ESHOT General Directorate of İzmir Public Bus System (ESHOT) (2012).*Bus line number 121 Bostanlı ferry station to Konak*.Retrieved June, 2014, fromhttp://www.eshot.gov.tr/HareketSaatleri.aspx?TIP=TumHatlar&HatNo=5.

Fan, W., & Machemehl, R. (2009). Do transits users just wait for buses or wait with strategies. *Journal of the Transportation Research Board, 2111*(-1), 169-176.

Furth, P., Muller, T. (2006). Service reliability and hidden waiting time - Insights from automatic vehicle location data. *Paper presented at the 85th Annual Meeting of the Transportation Research Board*, Washington, DC

Florida Department of Transportation, (2009).Synthesis of research on value of time and value of reliability.*Center for urban transportation research,* 54, 11-23.

Golshani, F. (1983).System regularity and overtaking rules in bus services.*Journal of the Operational Research Society*, 591-597.

Goverde, R. (1999). *Improving punctuality and transfer reliability by railway timetable optimization*.Doctor of Philosopy, Delft University of Technology.

Gronau, R. (1976). Economic approach to value of time and transportation choice value of travel time. *Transportation Research Record, 587,* 1-5.

Henderson, G., Adkins, H., &Kwong, P. (1991).Subway reliability and the odds of getting there on time.*Transportation Research Record, 1297*, 10–13.

Hensher, A. (1976). Review of studies leading to existing values of travel time, value of travel time. *Transportation Research Record, 587,* 30-41.

Hosmer, W., &Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York:Wiley.

Rodgers, J., &Nicewander, W. (1988). Thirteen ways to look at the correlation coefficient.*The American Statistician, 42*(1),59–66.

Jang, W. (2010). Travel time and transfer analysis using transit smart card data. *Journal of the Transportation Research Board, 2144*(-1), 142-149.

Jara-Díaz, S. (2000).*Allocation and valuation of travel-time savings.*Handbook of Transport Modelling. Oxford, UK: Emerald Group Publishing Limited.

Jara-Díaz, S. (2007).*Allocation and valuation of travel time savings.*Handbook of Transport Modelling 2nd Edition. Oxford: Pergamon.

Jara-Diaz, S. (2003).On the goods-activities technical relations in the time allocation theory.*Transportation 30*, 245-260.

Johnson, M. (1966). Travel time and the price of leisure. *Western Economic Journal, 4*, 135-145.

Kaparias, I., Bell, M., &Belzner, H. (2008). A new measure of travel time reliability for in-vehicle navigation systems.*Journal of Intelligent Transportation Systems, 12*(4), 202-211.

Kittelson, Q. & Associates, H. (2003).A guidebook for developing a transit performance-measurement system.*Transport Research Board*, Washington, D.C.

Lin, J., & Ruan, M. (2009).Probability-based bus headway regularity measure.*Intelligent Transport Systems, 3*(4), 400-408.

Lisco, T. (1967).*The value of commuters' travel time.*A Study in Urban Transportation.University of Chicago.

Liu, R., & Sinha, S. (2007). Modeling urban bus service and passenger reliability. Paper presented at *The Third International Symposium on Transportation Network Reliability*, The Hague, Netherlands.

Lomax, T., Schrank, D., Turner, S., &Margiotta, R. (2003). Selecting travel reliability measures.*Texas Transportation Institute monograph* (May 2003).

Luce, D. (1959). *Individual choice behavior*.New York:Wiley.

Luce, D. & P. Suppes. (1965). *Handbook of Mathematical Psychology*, New York:Wiley,249–410.

Marie, A., & Catherine, M. (2009).Smart card data in public transit planning: A review. *Cirrelt*, 48.

McFadden, D. (1974). *Conditional logit analysis of qualitative choice behavior*, in P.Zarembka, ed., New York:Frontiers in Econometrics, Academic Press, 105–42.

Meyer, M.D. (2002). Measuring system performance: the key to establishing operations as a core agency mission. Paper presented at *TRB Annual Conference*, January, Washington D.C.

Nakanishi, Y. J. (1997). Bus performance indicators: on-time performance and service regularity. *Journal of the Transportation Research Board, 1571*(-1), 1-13.

Oort, N. & Nes, R. (2003). Service regularity analysis for urban transit network design, Paper presented at *TRB Annual Meeting*, January, Washington DC.

Oort, O.,(1969). The evaluation of travelling time.*Journal of Transport Economics and Policy 3* (September), 279-286.

Osuna, E., & Newell, G. (1972). Control strategies for an idealized public transportation system. *Transportation Science, 6*(1), 52.

Polus, A. (1978). Modeling and measurements of bus service reliability. *Transportation Research, 12,* 253-256.

Pullen, W. T. (1993). Definition and measurement of quality of service for local public transport management.*Transport Reviews*, *13*(3), 247-264.

Reichman, S. & Peter R. (1976). Applications of value of travel time to travel Demand estimation. *Transportation Research Record,587,* 6-11.

Reichman, S. (1976).Conceptual problems in evaluation of travel time.*Transportation Research Record*, 24-29.

Shaw, W. (1992).Searching for opportunity cost of an individual's time.*Land Economics, 68*, 107-115.

Small, K. (1982). The scheduling of consumer activities: Work trips.*American Economic Review, 72*, 467-479.

Sterman,B. & Schofer, J. (1976). Factors affecting reliability of urban bus Services. *Transport Engineering Journal*, 147-159.

Strathman J., Kimpel T., &Dueker K. (2001). Evaluation of transit operations: data applications tri-Met's automated bus dispatching system. *Transportation Research Board Annual Meeting,* 2002.

Transport for London (TfL) (2006). *Passenger waittime reliability.*Retrieved June, 20, 2014, fromhttp://www.tfl.gov.uk/buses/about/performance/borough-report.asp.

Thurstone, L. (1927). A law of comparative judgment.*Psychological Review 34*, 273 86.

Train, K. (2009).*Discrete choice methods with simulation* (2nd ed.). New York: Cambridge University Press.

Trépanier, M., Chapleau, R., &Tranchan,t N. (2007). Individual trip destination estimation in transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, Taylor & Francis, 11*(1), 1-15.

Trépanier, M., Morency, C., &Agard, B. (2009). Calculation of transit performance measures using smartcard data. *Journal of Public Transportation, 12*(1), 79-96.

Trépanier, M., Tranchant, N., &Chapleau, R. (2007). Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems, 11*, 1-14.

Trompet, M., Liu, X., & Graham, D. (2011).Development of key performance indicator to compare regularity of service between urban bus operators.*Transportation Research Record: Journal of the Transportation Research Board, 2216*(-1), 33-41.

Turnquist, M., & Bowman, L. (1980). The effects of network structure on reliability of transit service. *Transportation Research Part B: Methodological, 14*(1–2), 79-86.

Türkiye İstatistik Kurumu (TÜİK) (2013).*General information about İzmir city*. Retrieved June, 2014, from http://www.tuik.gov.tr/ilGostergeleri/iller/IZMIR.pdf.

Türkiye İstatistik Kurumu (TÜİK) (2012).*Destricts development level for İzmir city*. Retrieved June, 2014, from http://www.tuik.gov.tr/ilGostergeleri/iller/IZMIR.pdf.

Uniman, D. (2009*). Service reliability measurement framework using smart card data: Application to the London underground*. Master of Science, Massachusetts Institute of Technology.

United States Department of Transportation(USDOT)(2003).*Revised departmental guidance: valuation of travel time in economic analysis*. U.S. Department of Transportation.

Van Lint, J., & Van Zuylen, H. (2005).monitoring and predicting freeway travel time reliability: using width and skew of day-to-day travel time distribution. transportation research record. *Journal of the Transportation Research Board, 1917*(-1), 54-62.

Oort, N., & Nes, R. (2004).Service regularity analysis for urban transit network design. Paper presented at the *Transportation Research Board 83st* Annual Meeting.

Waters, W. (1992).*Value of time savings for the economic evaluation of highway investments in British Columbia*.Reports and Studies, British Columbia Ministry of Transportation.