Transportation mode choice binary logit model: a case study for Johor Bahru city

To cite this article: O. C. Puan *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **527** 012066

IOP Conf. Series: Materials Science and Engineering **527** (2019) 012066 doi:10.1088/1757-899X/527/1/012066

Transportation mode choice binary logit model: a case study for Johor Bahru city

O. C. Puan*[1], Y. A. H. Hassan[1], N. Mashros[1], M. K. Idham[1], N. A. Hassan[1], M. N. M. Warid[1], and M. R. Hainin[1]

[1]School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

*Corresponding author e-mail: othmancp@utm.my

Abstract. The mode choice stage in transportation planning is the analysis process to estimate the number or percentage of trips performed by each mode of transport. In practice, the number of trips is used to estimate the demand for each mode of transport. Such information is important for planning and designing transportation facilities in an urban area. A mathematical model of individuals' mode choice is usually required in such an analysis. This paper discusses the result of a study carried out to evaluate the attributes that influence the mode choice of transport in Johor Bahru city by interviewing the individuals living for more than one year in the study area. The investigation conducted through one of attitudinal survey techniques known as stated preference survey. Stated preference survey conducted on a random sample of 384 respondents. Investigated attributes have been identified in analysis stage by applying binary logistic regression analysis procedure before deriving binary logit model for individuals' mode choice of transport in Johor Bahru city. Based on stated preference survey and logistic regression analysis of data, the model of mode choice of transport in Johor Bahru has been derived and attributes of age, income, vehicle ownership, comfort of car, reliability of bus service, affective motives and instrumental motives were found statistically correlated to the mode choice of transport in Johor Bahru.

1. Introduction

Transportation service considers as one of the important elements that enhance countries' development due to significant role of transportation in many aspects such as social, economic and political. However, transportation activities generate some negative reflections on surrounding environment and communities. Based on many studies conducted by transportation authorities, agencies and individuals it has been proven that most of these reflections such as traffic congestion, air pollution, noise and accidents are caused by more use of private vehicles on roads. Trips conducted by private cars contribute to climate change, dependence on fossil fuels, and traffic congestion. In 2006, roughly 30% of all greenhouse gas emissions in USA came from the transport sector and mainly in form of CO2 emissions from the burning of fossil fuels [1]. Therefore, transportation planners need to control the modal split between available modes of transport and promote particular modes such as public transport, walking and cycling.

Travel forecasting process in transportation planning of urban areas involves several stages, those stages are land use and travel characteristics, trip generation, trip distribution, mode choice and traffic assignment. The mode choice stage in transportation planning is the analysis process used to estimate the number or percentage of trips performs by each mode of transport [2].

The modelling of the mode choice of transport considering the attributes that correlated with the mode choice provides a mathematical reference used to estimate the number of trips conducted by each mode of transport. Desired rates for modal split could be adopted by changing the attributes correlated with the mode choice of transport.

This paper discusses the result of a study carried to develop a binary logit model describing the mode choice of a transportation system for the users in Johor Bahru City, Malaysia. The study was carried out based on the following limitations:

i. Study area only covers Johor Bahru city.
ii. Binary logistic regression model (Logit model) method is used to identify the determinants and derive the model used to estimate the probability of using each mode of transport.
iii. Modes of transportation were categorised into two types, i.e. Private transport mode (Cars) and Public transport mode (Bus).
iv. Stated preference survey method was adopted for data collection exercises.

2. Literature review

Mode choice of transport has been studied in many cities in Malaysia such as Penang, Klang Valley and Johor Bahru in order to identify the attributes influence the mode choice of transport or to model the mode choice of transport. Based on previous studies conducted in this aspect, it found that many attributes are contributing to the individuals' choices of their transport mode. These attributes generally categorized into three groups; Socio-economic and demographic characteristics, characteristics of trip, characteristics of the transport facilities.

2.1. Mode choice in developed countries

A comparative study conducted by Buehler [1] has discussed the important attributes that correlated to the mode choice of transport in Germany and USA which have been categorized into four groups. First group was socio-economic and demographic characteristics which includes attributes such as household income, vehicle ownership, gender and age. Second group was the spatial development patterns which involves population density. The third group was transportation policies which include fuel prices, sales taxes, registration fees, parking cost, tolls, speed limits, integration among public transport services and other modes, restriction on drivers and planning of land uses. Fourth group was culture and attitude which measures by eco-socialization that has been developed for this purpose by Low and Gleeson [3].

In 2006, Shen et al [15] had utilized five attributes to compare between Latent Class Model and Mixed Logit Model for transport Mode Choice based on preference surveys conducted in Japan. These attributes were In-vehicle time, Access time, Frequency, Travel cost, and negative impact on the environment caused by transport modes.

In Italy, 2007 a model has been developed to measure the customer satisfaction of bus service by Eboli and Mazzulla [16] based on several determinants such as bus stop availability, route characteristics, frequency, reliability, bus stop furniture, overcrowding, cleanliness, cost, information promotion safety on board, personal security, personnel, complaints, environmental protection, and bus stop maintenance. Although these attributes had been used to measure customer satisfaction and survey has been conducted for particular category of users (Student) some of these attributes are similar to the determinants had been used for transport mode choice as illustrated above.

2.2. Mode choice of transport in Malaysia

The individual's behaviour on choosing their mode of transport in Malaysia has been studied in some urban areas like Penang, Klang Valley and Johor Bahru in order to identify the determinants that influence traveller's behaviour on choosing their modes of transport or to model the mode choice of transport in particular city.

The determinants that influence the mode choice of transport in Penang have been studied by Chee and Fernandez [4]. Some determinants were found not correlated to the mode choice of travellers such as gender, availability of parking facilities, reliability and personal income. The personal income has been categorized into two groups, i.e. RM3000 or less and above RM3000, however most of respondents

in both groups prefer private vehicles. Although other determinants were found related to mode choice of transport in Penang such as; driving license, comfort of vehicle, overall quality of bus service, regularity of bus service as well as flexibility of reaching multiple destinations.

Study has been conducted by Chiu et al [5] to model the mode choice of transport considering Car, Bus and Rail using the logistic regression analysis. The study area has been comprised the federal territory of Kuala Lumpur, Selangor district of Petaling, Klang, Gombak, and Hulu Langat. Public transportation modes considered in this study were rail transport (LRT, Monorail, ERL and KTM commuter), buses and Taxis. The study categorized the determinants that influence the mode choice of transport in Klang Valley into three categories which are characteristics of travellers which comprised income, vehicle ownership, household structure, age and gender. Characteristics of trip which involve the purpose of trip, time of trip and trip distance. The third category was characteristics of the transport facilities, which included parking cost, car price, fuel price, accessibility distance to rail service, transit time, time spent using the service, accessibility distance to bus service, bus network, buses transit time and travelling time.

Loo et al [6] has studied the mode choice of transport in South East Asia (SEA) by investigating the relationship between transport user's perception and their travel mode of transport in Johor Bahru, Malaysia as a case study. The focus of this study was on psychological determinants that influence the mode choice of transport. These determinants have been categorized into three groups; Instrumental motives, Symbolic motives and Affective motives. Instrument motives are functional qualities on vehicle such as speed, safety and convenience. Symbolic motives mean by car used to express the identity and presenting the social position. Affective motives are emotional desires of individuals on car such as enjoyment of cars driving.

Minhans et al [17] studied the bus service quality in Johor Bahru, 2015 based on several variables under major categories; Reliability, Transit service and facilities, Bus Fare, Bus Characteristics, Conduct, Information, and Suitability.

In 2017, Kamaruddin et al [18] investigated the satisfaction of customer on public transport in Klang Valley considering LRT, Monorail, Bus, and KTMB Komuter. The determinants had been used in this study were Reliability, Safety, Accessibility, Fares, Communication, and Trip experience.

3. Methodology

Using all determinants has been studied before whether in developed countries or in Malaysia the mode choice of transport in Johor Bahru has been modelled through investigating the individuals' behaviour on mode choice of their trips.

3.1. Sampling

The sample size estimation could be done through various methods and references but mainly the estimation depends on confidence level and confidence interval [7].

In this study, the sample size was based on Equation 1 as suggested by Roess et al [8]. Using the standard deviation equal to 5, i.e. based on Chiu et al [5], and tolerance equal to ±0.5, a minimum sample size of 384 was required for the analysis.

$$N \geq (Z^2 . S^2) / \varepsilon^2 \quad (1)$$

Where:

N = Sample size
Z = Confidence level (1.96 for 95%)
S = Standard deviation
ε = Tolerance

3.2. Survey procedure

Stated preference (SP) survey which has been established in 1980 and developed in 1990, nowadays it is one of the most popular survey techniques used in transportation studies [9]. SP is conducted through several techniques such as personal interview, computer or web base and via telephone [10]. Each technique has some benefits as well as disadvantages. For example, using computer base and telephone base interview is cheaper than personal interview and it is expected that a greater number of respondents

could be involved, but the disadvantage is all population may don't have a computer or telephone and they may not respond also. However, the personal interview is more expensive but it is more reliable.

In this research a combination of both techniques has been used, a computer base questionnaire has been developed (online Google forms) then personal interviews have been conducted by using online Google forms and mobile tab which is used by respondents to fill the online questionnaire. The combination of Google forms online questionnaire and personal interview technique has certain benefits such as all population in Johor Bahru have same chance to participate in this survey which is important for random sampling, no need for data entry because once respondent submit the questionnaire the data automatically appear in excel format. Also, people prefer to deal with mobile phones and devices rather than paper and pencil. With this technique it was easy to manage the time because no need to wait for long time till receives the required sample size.

3.3. Analysis procedure

The models used to identify the individual's choices among several alternatives are commonly known as discrete choice models. These models were developed based on utility maximization hypothesis, which assumes an individual select the mode which maximizes his or her utility [11].

There are three discrete choice models' groups; the first group is Logit models which called Logit family models. Logit family models comprised simple binary logit model, nested binary logit model and multinomial logit model. The simple binary model used for two discrete choices only where individuals have two options to choose, while multinomial logit model used for more discrete choices (more than two). The nested binary logit model is basically used for two discrete choices but one or both of two choices comprise subsets of choices.

The utility function is generally represented by a linear model, comprises a group of independent variables which are factors influencing the mode choice and dependent variable which is the utility of the particular mode. Equation 2 demonstrates a utility function of mode (m) for a number of independent variables, based on Ben-Akiva and Bierlaire [11].

$$U_m = C + A_1. X_1 + A_2. X_2 + A_3. X_3 \ldots\ldots + A_i X_i \quad (2)$$

Where:

U_m = Utility function of mode 'm'
C = Constant
A_i = Coefficients (weight of each attribute based on survey's data)
X_i = Independent Variables correlated with the mode choice

The probability of using mode m (P_m) based on binary logit model represents mathematically by Equation 3 while the probability of second mode of transport (P_n) could be obtained from Equation 4.

$$P_m = e^{Um} / (1+ e^{Um}) \quad (3)$$

$$P_n = 1 - P_m \quad (4)$$

3.4. Data coding

Data has been coded for several levels according to Likert scale in order to get a comprehensive evaluation of each attribute. Some variables categorized into two levels such as gender (i.e. Male = 0 & Female = 1) and mode of transport (i.e. Car = 0 & Bus = 1) while other variables for five levels such as parking availability, i.e. very low, low, moderate, high and very high.

Some levels are indicating same concept like 'very low and low' both of them indicate low evaluation. Therefore, for analysis purposes levels with same concept combined into one level. It is because the software deals with all levels as different, which is not reflected the real situation. Table 1 summarises the coding of correlated attributes with mode choice of transport.

Table 1. Coding of correlated attributes.

Attributes	Range & code	Codes to be used in utility function model				
Age	Range	18 – 24	25 – 34	35 - 44	45 - 55	> 55
	Code	1	2	3	4	5
Income	Range	<RM1000	RM1000-1224	RM1225-1743	RM1744-3854	>RM3854
	Code	1	2	3	4	5
Comfort of car	Range	Dissatisfied	Neutral	Satisfied	-	-
	Code	1	2	3		
Reliability	Range	Dissatisfied	Neutral	Satisfied	-	-
	Code	1	2	3		
Affective motives	Range	Disagree	Undecided	Agree	-	-
	Code	1	2	3		
Instrumental motives	Range	Disagree	Undecided	Agree	-	-
	Code	1	2	3		
Vehicle ownership	Range	Own	Don't own	-	-	-
	Code	1	2			
Mode	Range	0	1	-	-	-
	Code	Car	Bus			

4. Results and discussion

In order to identify the attributes that statistically correlated with the mode choice of transport in Johor Bahru, binary logistic regression analysis has been conducted for recoded data which has been obtained through stated preference survey.

4.1. Identification of correlated attributes

According to analysis results in Table 2 the significance of each attribute is estimated. The p-value helps to decide whether there is a statistical correlation between dependent and independent variables or not. The lower p-value indicates the more confident on the existence of correlation between the variables [12, 13]. The origins of p-values come from hypothesis testing in statistics. In hypothesis testing, there are two hypotheses, i.e.:

(a) H_0 (called the null hypothesis): There is no relationship between the variables, and
(b) H_1 (called the alternative hypothesis): There is a relationship between the variables.

If the p-value is less than small threshold (often 0.05 is used), then the null hypothesis (H0) rejected, which means there is a relationship between the variables. The attribute with p-value less than 0.05 is considered as significantly correlated, while greater than 0.05 is considered as insignificantly correlated. Accordingly, seven attributes are significantly correlated with mode choice of transport in Johor Bahru. These attributes are the age, income, vehicle ownership, comfort of car, bus service reliability, affective and instrumental motives.

Table 2. Logistic regression analysis results.

Parameters	Coefficients	p-value
Intercept	-2.829	0.145
Gender	-0.529	0.102
Age	0.367	0.042
Education	-0.186	0.534
Employment	-0.076	0.632
Income	-0.684	0.000
HH Size	0.028	0.877
Vehicle Ownership	2.927	0.000
Purpose of Trip	-0.139	0.232
Travel distance	0.062	0.858
Travel time	-0.007	0.975
Parking cost	0.330	0.174
Parking availability	0.379	0.088
Car Price	0.608	0.062
Fuel price	-0.454	0.290
Toll cost	-0.239	0.537
Comfort of car	0.668	0.031
Number of transfers	-0.203	0.479
Reliability	0.861	0.002
Bus frequency	0.429	0.153
overall quality of bus service	-0.093	0.758
Coverage	0.155	0.588
Affective motives	-0.710	0.002
Instrumental Motives	-0.786	0.004
Symbolic motives	-0.160	0.359

Since the attributes significantly correlated with the mode choice of transport have been identified, all attributes are not correlated have been excluded and analysis has been repeated for those are significantly correlated only for more accuracy and the result of this analysis is presented in Table 3.

Table 3. Repeated Logistic regression analysis results.

Attributes	Coefficients	p-value
Intercept	-2.712	0.019
Age	0.343	0.020
Income	-0.715	0.000
Vehicle ownership	2.757	0.000
Comfort of car	0.619	0.029
Reliability	0.961	0.000
Affective motives	-0.721	0.001
Instrumental motives	-0.808	0.002

4.2. Modeling of mode choice

According to logistic regression analysis and repeated logistic regression analysis results the model of mode choice of transport in Johor Bahru could be presented by following utility function equations, i.e. Equations 5–7.

$$U_{bus} = -2.712 + 0.343(A) - 0.715(INC) + 2.757(VO) + 0.619(CC) + 0.961(RBS) - 0.721(AM) - 0.808(IM) \quad (5)$$

$$P_{bus} = e^{Ubus} / (1 + e^{Ubus}) \quad (6)$$

$$P_{car} = 1 - P_{bus} \quad (7)$$

Where:

U_{bus} = utility function of a bus mode of transport
A = age
INC = income
VO = vehicle ownership
CC = comfort of car
RBS = reliability of bus service
AM = affective motives
IM = instrumental motives
P_{bus} = probability of choosing a bus
P_{car} = probability of choosing a car

4.3. Accuracy of derived model

The accuracy of the derived model is one of the logistic regression analysis outcomes which is showing the number and percentages of correct predictions and false predictions for each mode of transport as well as the overall accuracy as shown in Table 4. Out of 117 choices of bus 62 choices have been predicted correctly while 55 choices were false prediction, therefore the accuracy of predicting bus choice was 53% (62/117). On other hands out of 267 car choices 244 choices were predicted correctly while only 23 choices were false prediction, therefore the accuracy of car choice prediction was 91% (244/267). And overall accuracy of derived model was 80% ((62+244) / 267).

The best indicator for accuracy of logistic regression model is the area under ROC curve (Receiver Operating Characteristic) which is a plot of true rates against false rates. ROC curve is drawn automatically and presented with logistic regression analysis results by using real statistics analysis in Microsoft Excel. If area under ROC curve is between 0.5 and 0.7 the prediction is poor, but if it is between 0.7 and 0.9 the prediction considered as reasonable. The area above 0.9 indicates very good prediction [14]. Since the overall accuracy is 80% which indicates the area under ROC curve then the derived model considers as reasonable.

Table 4. Accuracy of derived model.

	Bus Observed	Car Observed	Overall
Bus-Predicted	62	23	85
Car-Predicted	55	244	299
Total	117	267	384
Accuracy	0.53	0.91	0.80

The following Figure 1 illustrates the ROC curve for the derived model.

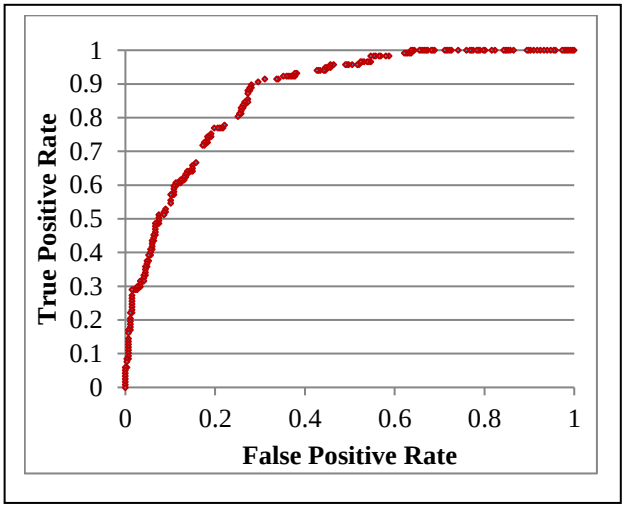

Figure. 1. ROC curve for the derived model.

5. Concluding remarks

The main findings of the study can be summarised as follows:

(i) The mode choice of transport in Johor Bahru city is statistically correlated with age of users, income, vehicle ownership, car-comfortability, reliability of bus service, affective motives and instrumental motives.

(ii) The bus mode choice of transport of travelers in Johor Bahru can be presented in a form of a utility function given by Equation 5 and the probability of bus mode and private car mode are given by Equations 6 and 7, respectively.

(iii) The derived model is considered as reasonable based on overall accuracy of 80% and area under ROC curve.

6. References

[1] Buehler, R. 2011 Determinants of transport mode choice: a comparison of Germany and the USA *Journal of Transport Geography* **19(4)** 644-657

[2] Nicholas J. Garber and Lester A. Hoel 2009 *Traffic and Highway Engineering* Fourth Edition, Toronto, Canada

[3] Low, N and Gleeson, B. 2001 Eco socialisation or counter modernisation, Reviewing the shifting 'storylines' of transportation planning *Int. J. Urban Reg. Res.* **25** 784-803

[4] Chee, W. L., and Fernandez, J. L. 2013 Factors that Influence the Choice of Mode of Transport in Penang: A Preliminary Analysis *Procedia - Social and Behavioral Sciences* **91** 120- 127

[5] Chiu Chuen, O., Karim, M. R., and Yusoff, S. 2014 Mode choice between private and public transport in Klang Valley, Malaysia *The Scientific World Journal* **2014**

[6] Loo, L. Y. L., Corcoran, J., Mateo-Babiano, D., and Zahnow, R. 2015 Transport mode choice in South East Asia: Investigating the relationship between transport users' perception and travel behaviour in Johor Bahru, Malaysia *Journal of Transport Geography* **46** 99-111

[7] Evans, H. K. 2010 *Traffic Engineering Handbook*. Institute of Traffic Engineers, Memphis, Tenn: General Books

[8] Roess, R. P., McShane, W. R., and Prassas, E. S. 1998 *Traffic Engineering* (3rd edition) Upper Saddle River, N.J: Prentice Hall

[9] O'Flaherty, C. (Ed.) (1997). *Transport planning and traffic engineering*. Elsevier.

[10] A. G. Bluman (2011). *Elementary Statistics: A Step by Step Approach*, McGraw-Hill Higher Education, 8th edition.

[11] Ben-Akiva, M. and Bierlaire, Michel 1999 *Discrete Choice Methods and Their Application to Short Term Travel Decisions* The MIT Press, Massachusetts, U.S.A

[12] Morganstein, D. R., R. K. Som 1999 *Practical Sampling Techniques* Technometrics **41(2)** 175-175.

[13] P. S. Levy and S. Lemeshow 2008 *Sampling of Populations: Methods and Applications*, Fourth Edition, Hoboken: Wiley
[14] Swets, J. A. 1986 Indices of discrimination or diagnostic accuracy: Their ROCs and implied models *Psychological Bulletin* **99(1)** 100-117
[15] Shen, J., Sakata, Y., & Hashimoto, Y. 2006 A comparison between latent class model and mixed logit model for transport mode choice: evidences from two datasets of Japan *Discussion Papers in Economics and Business* 06-05
[16] Eboli, L., and Mazzulla, G. 2007 Service quality attributes affecting customer satisfaction for bus transit *Journal of Public Transportation* **10(3)**
[17] Minhans, A., Shahid, S., and Hassan, S. A. 2015 Assessment of bus service-quality using passengers' perceptions *Jurnal Teknologi* **73(4)** 61-67
[18] Kamaruddin, R., Osman, I., and Pei, C. A. C. 2017 Customer expectations and its relationship towards public transport in Klang Valley *Journal of Asian Behavioural Studies* **2(4)** 29-39

Acknowledgments
The authors would like to express deep gratitude and appreciation to Universiti Teknologi Malaysia for providing a research grant under GUP Scheme (Q.J130000.2511.18H29), the opportunity and necessary facilities to support this research.