# The Relationship of Passenger Characteristics to Electrical Train Modes Selection in Jakarta

Adhi Muhtadi, Indrasurya B. Mochtar , and Hera Widyastuti

***Abstract*— Electric Railway (KRL) is the backbone of public transportation in Jakarta. But KRL share capital is still 4.1%. Therefore this paper seeks to reveal what characteristics play a role for travellers to use KRL regularly or temporarily. The study was conducted in October 2017 at several Jakarta KRL stations. The sampling technique uses convenience and accidental sampling which chooses 66 respondents of KRL users. The analysis technique used Binary Logistic Regression. The results of this research show that the characteristics of age, education, and income / month influence the choice of travellers to use KRL regularly and temporarily. The resulting model is able to explain the possibility of someone using KRL regularly or temporary as much as 54.8% according Value Nagelkerke R Square. To increase the percentage need to add a number of other free variables in the future.**

***Index Terms*— passenger characteristics, KRL, binary logistic regression, Jakarta**

## I. Introduction

THE history of KRL Jakarta is actually very long which started in 1925. But in 1960 - 1972 KRL operation was stopped because it was considered as the cause of traffic jam in Jakarta. Effective starting in 1972 KRL Jakarta operated again and in the year 2008 formed PT. KAI Commuter Jabodetabek (KCJ) to handle Jabodetabek area operations.

KRL expected to be the backbone of public transportation in Jakarta. The growth of KRL passengers each year can reach 13.8% [1]. But modal share from KRL only 4.1% from 18.7 million movement / day in Jakarta. The modal share obtained not optimal, therefore it is very necessary research on the characteristics of KRL passengers to know what characteristics affect someone so to use KRL regularly or temporarily. In the end of paper can be known what characteristics should be improved so that private vehicle users switch to KRL.

Many of the papers reviewed discussed the importance of a range of socio-demographic variables that can influence the choice of travel mode and the frequency of trips. The most important variables are the following: gender, age, income, employment status, retirement status, educational status, ethnicity, English or non-English speaking background, car, ownership including number of cars in household, household composition (includes presence or absence of children, age of children, students, number of adults in employment)[2].

There is an interesting study in 2009 with the title of possible transition of private transport mode to public transport mode for Depok - Jakarta travel. The study involved some characteristics such as average / monthly income, occupation, education, travel intent, age, and consistency of private vehicle use [3]. The results showed that the willingness of private transportation users to switch and using public transport is too big. The transition from private car to bus by 75%, private car to KRL by 63%, motorcycle to bus by 80%, and motorcycle to KRL for 72%. However, until 2018 that number is inaccurate because only a few switch to KRL. This is considering only 4.1% of the 18.7 million movements that use KRL every day.

In other publications it is stated that KRL is chosen by travelers because it is cheap, fast and safe. Variables suspected to be linked include age, recent education, intent and purpose using KRL, old as KRL users, and reason for using KRL [4]. Although KRL is considered a cheap, fast and secure mode of transportation, but modal share still not optimal.

Other findings from the mass transit system in Bangkok and Manila reveal that those who live and have destinations in transit-based mass transit areas are not users of regular public transport. This means that the transit mass transit system only gets a limited number of passengers. The results also show that the distance of access and availability of cars has a very important influence on trends in the use of mass transit on a regular basis. For further mileage, improving the feeder bus service is still indispensable [5].

Therefore, this paper attempts to model the independent variables associated with the selection of KRL modes. The variables studied were age, sex, education, occupation, income, number of motorcycle ownership, number of car ownership, ownership of driving license, distance to station, distance with KRL, end station distance to destination location, number of luggage, and regularly or temporary KRL user status. The results of this study will examine the independent variables that are closely related to the regular and temporary selection of KRL modes. Another result is a binary logistic regression model capable of predicting someone will use regular or temporary KRL modes.

## II. Research Methodology

This research will determine the problem, purpose of research, collecting primary and secondary data. The primary data collected from the questionnaires distribute and has been

Adhi Muhtadi is with the Department of Civil Engineering and Environment, Sepuluh Nopember Institute of Technology, Surabaya, East Java, Indonesia (e-mail: adhimuhtadi1974@gmail.com).

Indrasurya B. Mochtar is with the Department of Civil Engineering and Environment, Sepuluh Nopember Institute of Technology, Surabaya, East Java, Indonesia .

Hera Widyastuti is with the Department of Civil Engineering and Environment, Sepuluh Nopember Institute of Technology, Surabaya, East Java, Indonesia (e-mail: hera.widyastuti@yahoo.co.uk).

Copyright © 2019, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

answer by KRL user respondents. The sample selection using convenience sampling and accidental sampling method. The number of respondents who returned answers as many as 66 respondents. Location of questionnaires distributed in several KRL stations in Jakarta, such Jakarta Kota Station, Manggarai Station, Tangerang Station, Duri Station, Kampung Bandan Station, Depok Station, Bogor Station and Bekasi Station. Data collection was done in October 2017.

The questionnaires consist of age, sex, education, occupation, income, number of motorcycle ownership, number of car ownership, ownership of driving license, proximity to station, distance with KRL, end station distance to destination, number of goods brought and user status Regular or temporary KRL. After the respondents' answers are collected, then the data is coded into the SPSS software and process through binary logistic regression. The results of this study will examine what variables are closely related to the regular and temporary selection of KRL modes. Another result is a binary logistic regression model capable of predicting someone will use regular or temporary KRL modes. The discussion is trying to compare with some preliminary studies that have been done specifically about urban trains.

## III. Mathematical Background of The Analysis

Respondents' answers will be used to make a probability model of someone who will use KRL regularly or temporarily. The modelling will be done by binary logistic regression analysis technique. Binary logistic regression analysis is used to explain the relationship between response variables in the form of dichotomic / binary data with independent variables in the form of interval and or categorical data [6].

Binary logistic regression is actually the same as multiple regression analysis, only the dependent variable is dummy variable (0 and 1). For example, the effect of some characteristics of KRL users against regular or temporary selection of KRL modes. Then the dependent variable is 0 if selecting KRL temporarily and 1 if choose KRL regularly. Binary logistic regression does not require normality assumption, although outliers data screening can still be performed.

Assumptions in binary logistic regression:

- Not assuming linear relationship between dependent and independent variables
- The dependent variable must be dichotomous (2 variables)
- Independent variables should not have the same diversity between groups of variables
- Categories in independent variables must be separate from each other or are exclusive
- Samples required in relatively large quantities, minimum required up to 50

sample data for a predictor variable (free).

Unlike ordinary linear regression, binary logistic regression does not assume the relationship between independent and linear dependent variables. Binary logistic regression is a non-linear regression where the specified model will follow the curve pattern as shown below.

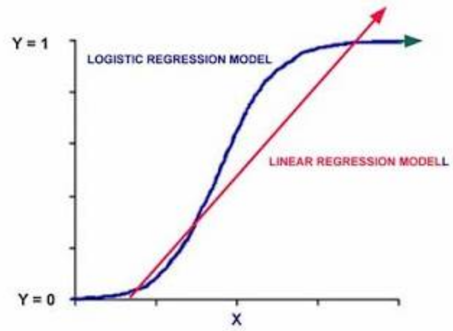


**Fig. 1.** Logistic Curve Regression Model and Linear Regression Model

The models used in binary logistic regression are:

$$Log\ (P/1-p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (1)$$

Where p is the probability that Y = 1, and $X_1$, $X_2$, $X_k$ are independent variables, and β is the regression coefficient.

The steps in using binary logistic regression analysis are:

1. Model Significance Test

To know the effect of free variable to independent variable together in model, can use likelihood ratio test. The hypothesis is as follows:

$H_o$: $\beta_1 = \beta_2 = .... = B_p = 0$ (no effect of free variable simultaneously to dependent variable)

$H_1$: there is at least one $\beta_j \neq 0$ (there is an effect of at least one independent variable against dependent variable)

For j = 1,2, ..., p

Test statistics used are:

$$G^2 = -\ 2ln + L_0/L_p \quad (2)$$

Note:

$L_o$ = Maximum likelihood of the reduction model

$L_p$ = Maximum likelihood of the full model or with all independent variables.

This $G^2$ statistic follows the chi-square distribution with degrees of p-free so the hypothesis is rejected if p-value <α, which means the independent variable X together affects the dependent variable Y.

2. Partial Test and Model Formation

In general, the purpose of statistical analysis is to find a suitable model and strong linkage between the model and the existing data. Testing the significance of parameters (coefficient β) partially can be done through wald test with the hypothesis as follows:

$H_o$: $\beta_j = 0$ (the independent variable to j has no significant effect on dependent variable

$H_1$: $\beta_j \neq 0$ (independent variable to j has significant effect on dependent variable)

For j = 1,2, ...., p

With the following test statistics:

Hypothesis will be rejected if p-value <α which means free variable Xj partially affect the dependent variable Y.

3. Odds Ratio

Odds ratio is a measure of the risk or tendency to experience a successful event between one category with another category, defined as the ratio of odds to xj = 1 to xj = 0. This Odds ratio states the risk or trend of the observational influence with xj = 1 is the number of times if compared with the observation with $x_j = 0$. For independent variables of continuous scale then the interpretation of the coefficient βj in the logistic regression model is that each increase of unit c on the independent variable will cause the risk of occurrence Y = 1, is exp (c.$\beta_j$) times greater. Odds ratio is denoted by θ, defined as the ratio of two values of odds $x_j = 1$ and $x_j = 0$, thus:

$$\theta = [\pi(1)/[1-\pi(1)]] / [\pi(0)/[1-\pi(0)]] \quad (3)$$

Binary logistic regression also generates odds ratios with respect to the value of each predictor. The odds of an event are defined as the probability of an arising outcome divided by the probability of an event not occurring. In general, the odds ratios ratio is a set of opportunities shared by other opportunities. The chance ratio for the predictor is defined as the relative amount by which the yield probability increases (the probability ratio> 1) or decreases (the probability ratio <1) when the predictor variable value increases by 1 unit.

Binary logistic regression will form a predictor / response variable (log (p / (1-p)) which is a linear combination of independent variables The value of this predictor variable is then transformed into probability by logit function Logistic regression also produces odds ratios is related to the value of each predictor The odds of an event are interpreted as the probability of an arising outcome divided by the probability of an event not occurring. Generally, the odds ratios ratio is a set of opportunities shared by the other opportunity The opportunity ratio for the predictor is defined as the relative amount where the yield probability increases (the ratio of chances> 1) or decreases (the probability ratio <1) when the value of the predictor variable increases by 1 unit.

## IV. RESULTS AND DISCUSSION

The results showed that the age of respondents between the ages of 17 years to 67 years. Gender of male respondents as many as 48 people and 18 women. The last education level of the respondent is equal to elementary school (1 person), junior high (3 people), high school (19 people) and bachelor (43 people). Employment of respondents as employees (59 people), entrepreneurs (4 persons) and students (3 persons). Unmarried respondents were 21 people and married as many as 45 people. Number of respondent families 1 (3 people), 2 (13 people), 3 (24 people) and 4 or more (26 people). Respondents who do not have a motorcycle as many as 8 people, have 1 unit of motorcycle (33 people), have 2 motorcycles (22 people), and who has 3 or more motorcycles (3 people). Respondents who did not own a car were 34, had 1 car (29 people), had 2 cars (1 person) and had 3 cars or more (2 persons). Respondents who do not have a driving license as many as 20 people, who have a license to drive as many as 46 people. Luggage carried by the respondent as much as 1 bag (59 people), carry 2 bags (6 people) and carry 3 bags (1 person). Respondents who chose KRL vehicles temporarily as many as 39 people and who chose KRL regularly as many as 27 people.

After knowing the various characteristics of KRL user respondents, then made a binary logistic regression model. From the available variables, it turns out that only 3 have relationship with regular or temporary mode selection of KRL ie age, education, and income / month. So the binary logistic regression model obtained is as follows:

TABLE I
BINARY LOGISTIC REGRESSION MODEL

| | Variable | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1 | Age | 0.205 | 0.060 | 11.830 | 1 | 0.001 | 1.228 |
| | Education | 1.877 | 0.715 | 6.884 | 1 | 0.009 | 6.532 |
| | Income/month | -0.515 | 0.175 | 8.637 | 1 | 0.003 | 0.598 |
| | Constant | -11.254 | 3.799 | 8.774 | 1 | 0.003 | 0.000 |

The initial model of logistic regression equation as follows:

$$Log\ (P/1\text{-}p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \quad (3)$$

or

$$P = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)} / 1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)} \quad (4)$$

While the binary logistic regression model obtained as follows:

$$Log\ (P/1\text{-}p) = -11.254 + 0.205.X_1 + 1.877.X_2 - 0.515.X_3$$

Note: $X_1$ = age; $X_2$ = education ; $X_3$ = Income/month

To simplify the interpretation of the results of logistic regression equation above, then the equation has been obtained will be entered into the data that have been obtained from the respondents.

TABLE II
EXAMPLE CALCULATION OF ODDS RATIO

| Age | Education | Income/month | KRL Regular /Temporary | Odds Ratio |
|---|---|---|---|---|
| 31 | 4 | 1 | 2 | 0.895 |
| 45 | 2 | 1 | 1 | 0.775 |
| 24 | 3 | 1 | 1 | 0.223 |
| 23 | 4 | 1 | 1 | 0.614 |
| 33 | 4 | 1 | 2 | 0.928 |

In Table 1 there is a classification of educational data (1) for respondents with primary or equivalent education, (2) for respondents with junior high school or equivalent, (3) high school education or equivalent, and (4) undergraduate education. While income / month is classified as (1) below Rp. 5 million, (2) Rp. 5.000.001 - Rp. 10 million, (3) Rp. 10,000.001 - Rp. 15 million, and (4) more than Rp. 15 millions. For the classification of passenger KRL passengers are given the code (0) and regular KRL passengers are given the code (1).

In Table 2 it can be seen that there is a respondent with age 31 years, undergraduate education, income / month below Rp. 5.000.000, and use KRL on a regular basis. Then get the value of odds ratio of 0.895 or in other words 89.5% of respondents will use KRL on a regular basis. The resulting model is able to predict the accuracy of the respondent's answers up to 66.5% of the initial respondent's answer.

After obtaining binary logistic regression model, the model is tested through model significance test, partial test and model formation and odds ratio test. Model significance test can be done from Hosmer and Lemeshow Test. The following test results Hosmer and Lemeshow Test obtained from the data processing respondents.. Tabel 1 informs education backgrounds are (1) graduation from elementary school, (2) graduation from junior high school, (3) graduation from senior high school (4) graduation from bachelor. Incomes per month are classified: (1) below Rp. 5 millions, (2) Rp. 5.000.001 – Rp. 10 millions, (3) Rp. 10.000.001 – Rp. 15 millions, and (4) > Rp. 15 millions. For passenger groups, KRL temporary is code (0) and KRL regular passgenrs is code (1).

TABLE III
MODEL SIGNIFICANCE TEST

| Step | Chi-Square | df | Sig. |
|---|---|---|---|
| 1 | 6.532 | 7 | 0.479 |

Hosmer and Lemeshow test (Table 3) is to see whether the empirical data match or not with the model or in other words expected no difference between the empirical data with the model. The model will be declared feasible if significance above 0.05 or -2 Log Likelihood under Chi Square Table. It appears in Table 3 that the value of Hosmer and Lemeshow Test is 6.532 with significance of 0.479> 0.05. Means the model is fit and the model is declared feasible and may be interpreted. When tested through -2 log likelihood, then the value as follows.

TABLE IV
MODEL SUMMARY

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 6.532 | 7 | 0.479 |

-2 Log-Likelihood value is 53.604 which will be compared with Chi Square value at 0.05 significance level with df of n-1 with n is the number of samples, mean 66 - 1 = 65. The value of chi-square table with df = 65 is equal to 84.8065. Since the -2 log-likelihood value (53.604) is still below the value of chi-square table, the model is declared eligible to use.

Partial tests are also performed and the results are listed in Table 1. It is shown in table 1 that the significance of age (0.001), education (0.009), income / month (0.003) and constant (0.003) overall below 0.05, expressed fit / eligible to form a binary logistic regression model as described.

In Table 4 also obtained value Nagelkerke R Square of 0.548. It can be concluded that the free variable that is age, education and income / month explains the possibility of someone using KRL regularly equal to 54.8%.

## V. CONCLUSION

The results showed that the independent variables that affect the selection of modes of KRL regularly or temporarily are age ($X_1$), education ($X_2$) and income / month ($X_3$). The $\beta$ values for each variable are 0.205, 1.877 and -0.515. The constant value is -11.254. The binary logistic regression model that was produced was able to predict the respondent's answer with 66.5% accuracy. Hosmer and Lemeshow test results, -2 log likelihood, and partial test show that the resulting model is feasible to generalize. The value of Nagelkerke R Square generated is 0.548. It can be concluded that the independent variables (age, education and income / month) explain the possibility of someone using KRL regularly at 54.8%.

So for future research, it is necessary to add some independent variables to be able to predict the probability of someone using KRL regularly better. Added independent variables can be derived from the quality of services such as the number of carriages, KRL headlines, KRL feeders, KRL speeds, integration with TransJakarta, pedestrian integration, access for bicycle users and KRL fares [7], travel time and KRL tariff [8] security, comfort, order and hygiene [9], availability of accurate information in real time, service coverage and handling of professional complaints [10], availability of security officers within carriages [11], service delivery and credibility [12] KRL, departure and arrival schedules [13], passenger buildup or queue length and carriage capacity during peak hours [14], and the latter is integration with on-line public transport such as online motorcycles or taxi [15].

## REFERENCES

[1] Nazwirman, Hulmansyah, "Karakteristik Penumpang Pengguna KRL Commuter Line Jabodetabek," *Journal of Economics and Business Aseanomics* (JEBA), 2 No. 1, 2017.
[2] C. Curtis, T. Perkins, "Travel Behaviour: A review of Recent Literature, Impacts of Transit Led Development In a New Rail Corridor," Working Paper No. 3, Department of Urban and Regional Planning, Curtin University, 2006.
[3] R. Ferdiansyah, "Kemungkinan Peralihan Penggunaan Moda Angkutan Pribadi Ke Moda Angkutan Umum Perjalanan Depok-Jakarta," Jurnal Perencanaan Wilayah dan Kota, 20, 183 – 198, 2009.
[4] Nazwirman, Hulmansyah, *Karakteristik Penumpang Pengguna KRL Commuter Line Jabodetabek*, *Journal of Economics and Business Aseanomics* (JEBA), 2 No. 1, 2017.
[5] S. S. Wibowo, S. Chalermpong, "Characteristics of Mode Choice within Mass Transit Catchments Area*," Journal of the Eastern Asia Society for Transportation Studies*, 8, 2010.
[6] H. Huang, *Analisis Regresi Logistik Biner*, http://www.globalstatistik.com, diakses pada 18 Juni 2018, 2017.
[7] A. Muhtadi, I.B. Mochtar, H. Widyastuti, "Best Practice BRT for Increase TransJakarta Modal Share," *Proceeding Regional Conference in Civil Engineering ICCER,* 619 – 627, 2017.
[8] A. Muhtadi, H. Widyastuti, "Prediksi Waktu Tempuh dan Tarif yang Diinginkan Pengguna Sepeda Motor Untuk Menggunakan Trem Surabaya Pada Koridor Terminal Joyoboyo – Jl Rajawali Surabaya," *Prosidng Seminar Nasional X Teknik Sipil ITS Surabaya*, 229 – 237, 2014.
[9] A. Prasetyo, H.A. Basri, "Analisis Tingkat Kepuasan Penumpang Terhadap Pelayanan Jasa Kereta Rel Listrik (KRL) Bogor – Jakarta," *Jurnal Ekonomi*, 5, 119 – 137, 2009.
[10] A. D. Saputra, "*Analysis Of Train Passenger Responses on Provided Service Case study: PT. Kereta Api Indonesia and Statens Järnvägar (SJ) AB, Sweden*, Service Science Program, Karlstad University, 2010.
[11] I. W. Maesarini, R.R. Fauzi, "Analisis Tingkat Kepuasan Pelanggan Krl Sistem Commuter Line (Studi Kasus Pada PT Kereta Api Commuter Jabodetabek)," Program Studi Administrasi Publik STIAMI Jakarta, 2012.
[12] F. Zulkifli, Syahputra, "Factor Analysis of Service Quality in KRL Commuter Line Jabodetabek for Women Railway Carriages (Study at PT KAI Commuter Jabodetabek Bogor-Jakarta Kota Route)," *e-Proceeding of Management*, 3, 2007 – 2019, 2016.

[13] D. Saidah, "Service Quality of Commuter Line," *Jurnal Manajemen Transportasi and Logistik*, 4, 51 – 58, 2017.
[14] M. Lusiani, S. Chandra, "Optimization of Total Arrival on Commuter Line Train to Overcome Passenger's Accumulation Bekasi - Jakarta Kota Using Simulation Promodel," *Journal of Industrial Engineering and Management Systems*, 11, 32-38, 2018.
[15] A F Saffan, M. Rizki, "Exploring the Role of Online 'Ojek' In Public Transport Trips: Case of Jakarta Metropolitan Area Rail Users," *The 4th PlanoCosmo International Conference, IOP Conf. Series: Earth and Environmental Science 158*, 2018.