Modelling student travel mode choice: a case study in Semarang, Indonesia

B Haryadi¹, A Narendra¹, M Purnomo¹ and A Purnomo²

¹Department of Civil Engineering, Universitas Negeri Semarang, Kampus Sekaran, Semarang, Indonesia
²Department of Architecture, Universitas Negeri Semarang, Kampus Sekaran, Semarang, Indonesia

bambang.haryadi@mail.unnes.ac.id

Abstract. The paper presents an investigation of university students' travel mode choice behaviour for commuting to campus. It uses a dataset collected through a survey among students of four universities in Semarang, i.e., Unnes, UINW, Udinus, and Unissula. Eight hundred and thirty-seven respondents completed the questionnaires. Multinomial logit (MNL) models are used for investigating four transport mode choices, i.e., walk, motorcycle, car, and bus. The independent variables are schools, genders, origin, residential types, vehicle ownership, driving license ownership, monthly allowance, transportation budget, and commuting distance. The model shows that all independent variables, except monthly allowance, are significant predictors. Overall, the model prediction accuracy is about 90 percent. A motorcycle was the foremost prevalent transport for school trips. Empirical models reveal that the mode of choice behaviour of non-native students differs significantly from those that originate in Semarang. The log odds of driving a car, riding a motorcycle, and walking versus riding a bus are higher when students' origin is not from Semarang compared to Semarang.

1. Introduction

The City of Semarang is the fifth biggest city in Indonesia, with over 1.8 million people and 373.78 square kilometres. There are 64 public and private universities in the Semarang metropolis, with 231,636 students [1]. These include large, medium, and small universities and colleges. In term of the student body of these university vary widely. Few of these institutions have more than 30 thousand students, while some others have less than 2000 student population.

University students can make up a significant portion of a region's population, yet university students' travel behaviour is poorly understood or adequately represented [2]. University students has complex and distinctive travel activities. They have freedom in the campus environment, and have autonomy in their decision-making regarding to their activities, with minimal control from university authorities and their parents. Their day-to-day doings of a student are highly influenced by their peers. In general university student are broad-minded and open to new ideas from peers with various backgrounds and various interests [3].

A lot of study has been done on the young generation's travel behaviour, but still little is known about the factors that influence day-to-day trips. University students are at an essential stage in their life-course where they begin to form habitual travel behaviour as young adults [4]. A study in Auckland found that university students' choice of travel mode is different from general workers [5].
A critical reason to study college students’ travel mode choice is that universities represent liveable environments, are friendly to alternative travel modes, have a higher density than other environments, and offer mixed transport modes. Exploration of students' travels behaviour can be instructive and reveal valuable information about associations with the built environment and the extent of differences in travel (e.g., trip generation and mode choices) compared with the general population [2].

The study's objective is to investigate the travel mode behaviour of university students with a case study in the City of Semarang, Indonesia. This study examines various aspects of students' backgrounds and travel characteristics. The results give us a better understanding of mode choice in a campus environment and how university students’ travel modes differ from other social groups.

2. Method
The data was collected during the period of May 2020. Questioners were distributed in four large universities. They were two public universities, i.e., Universitas Negeri Semarang (Unnes) and Universitas Islam Negeri Wali Songo (UINWS), and two private universities, i.e., Universitas Dian Nuswantoro (Udinus) and Universitas Islam Sultan Agung Semarang (Unissula). These academic institutions were chosen as their characteristics differed, and many students were studying in these institutions.

The study expected the respondents to complete a structured survey form. The questionnaire consisted of two parts: socio-demographic and travel characteristics of students. The first part gathered data on the university enrolled, sex, year of student, monthly allowance, transportation expenses, residential types, driving (riding) license, and vehicle possession. Students were also requested to provide their school name, campus location and their usual place of residence, distance from their usual place of residence to campus, transport mode to campus. The second part gathers data on trip behaviour, e.g., the primary mode choice, the trip frequency, and the time they travel to school.

The sampling was conducted in two stages: selection of universities and selection of respondents from the selected universities to take part in the survey. Two public universities (Universitas Negeri Semarang (Unnes) and Universitas Islam Negeri Wali Songo (UINSW)), and two private university (Universitas Islam Sultan Agung (Unissa) and Universitas Dian Nuswantoro (Udinus)) were selected as a sample to take part in the survey. These institutions were chosen to represent the different characteristics in term of size (number of enrolled student), location (Central Business District (CBD) and suburb), status (public and private), and subject matter (Islamic and general).

Regarding the sampling process for the selection of respondents, for a±5% error margin and 95% confidence level, the minimum sample required is 383 for a population of over 10,000 [7]. A purposive sampling technique was used to select students from selected university in order to take part in the survey. The total sample was assigned to be chosen for the interview based on the student population in each university. After errors checking and data cleaning, 837 student respondents remained used in this study.

There are several models available to link a set of transportation mode choices as response variables with a set of variables as regressors [8]. The MNL model is preferred in the present context. The three reasons behind this decision were that the dependent variable has only five distinct and separable choice alternatives, the choice alternatives in question are not nested, and there are no alternative-specific independent variables [9]. The MNL model may be derived within a random utility framework in which the utility $U_{ij}$ derived by the ith individual from the jth choice as follow [10]

$$U_{ij} = \bar{U}_{ij} + e_{ij} = X_{ij}' \beta + e_{ij}$$  \hspace{1cm} (1)

where $\bar{U}_{ij}$ = average utility, $e_{ij}$ = random error term, $X_{ij}$ = the set of independent variables, and $\beta$ = an unknown parameters vector. Concerning the alternatives among the five transportation mode choices, the MNL regression model is thus given by [10]

$$P(Y_i = j) = \frac{e^{X_{ij}' \beta}}{\sum_{k=0}^{4} e^{X_{ik}' \beta}}, j = 0, 1, 2, 3,$$  \hspace{1cm} (2)
where \( P(Y_i = j) \) equals the probability that the student with characteristics \( x_i \) chooses the \( j \)th category of the dependent mode choice variable [11].

3. Data
Table 1 presents the summary statistics of categorical data used for developing the model. Eight hundred thirty-seven students completed the questionnaires. Regarding gender, the sample comprises 58 percent female and 42 percent male. Based on the origin of the students, the "Origin" variable is grouped into two levels, i.e., if the students are Semarang native or not. Most of the students are not originated from Semarang, with the proportion 86.02%. The motorcycle is the transportation mode most preferred by university students, with the proportion being 80.41%. Concerning the residential type, most students (52.45%) live in rented rooms called "kos." It is no surprise that the most students owned motorcycles (78%), and the most license held are riding license Type C (78%).

| Characteristics    | Levels       | N   | %   | Characteristics     | Levels       | N   | %   |
|--------------------|--------------|-----|-----|---------------------|--------------|-----|-----|
| Schools            | Udinus       | 220 | 26.28 | Gender              | Female       | 489 | 58.42 |
|                    | UNIV         | 217 | 25.93 | Male                | 348          | 41.58 |
|                    | Unissula     | 200 | 23.89 | Origin:             | 0 (No)       | 720 | 86.02 |
|                    | Unnes        | 200 | 23.89 | Semarang:           | 1 (Yes)      | 117 | 13.98 |
| Residential types  | Dormitory    | 91  | 10.87 | ComModes            | Bus          | 41  | 4.90 |
|                    | Home         | 307 | 36.68 | Car                 | 42           | 5.02 |
|                    | Rented room  | 439 | 52.45 | Car & MC            | MC           | 673 | 80.41 |
| Veh. Owned         | Bicycle      | 2   | 0.24 | Others              | 4            | 0.48 |
|                    | Car          | 24  | 2.87 | Walk                | 77           | 9.20 |
|                    | Car & MC     | 25  | 2.99 | MC                  |              |     |     |
|                    | MC           | 629 | 75.15 |                      |              |     |     |
|                    | None         | 157 | 18.76 |                      |              |     |     |

Table 2. Summary statistics of numeric variables

| Statistics | Characteristics | Allowance (Rp. 100,000) | Transp. Cost (Rp. 100,000) | Distance (km) |
|------------|-----------------|-------------------------|----------------------------|---------------|
| Min.       |                 | 1.00                    | 0.00                       | 0.15          |
| Median     |                 | 10.00                   | 1.50                       | 2.30          |
| Mean       |                 | 12.46                   | 1.91                       | 7.13          |
| Max.       |                 | 40.00                   | 10.00                      | 300.00        |
| S.D.       |                 | 8.11                    | 1.55                       | 16.19         |

Table 2 summarizes three numeric variables, i.e., student income (monthly allowance), monthly transportation expenses, and commuting distance from residential to campus. The range of students' income is vast, from the minimum of IDR 100 thousand to a maximum of IDR 4 million, with an average of about IDR 1.25 million.

4. Results and discussion
The structure of the MNL model is presented in this section. Analysis was conducted in R [12], and fitting for multinomial logistic regression model was done using mutlinorm, a function in the nnet package [13, p. 204]. The results of the mode choice analysis are shown in Table 3.
Table 3 shows the model estimation results. Variables are selected based on the corresponding parameters' expected signs and statistical significance (95% confidence interval). Some variables in the model are not statistically significant at a 95% confidence interval. Still, they are kept in the model because these variables provide significant insight into the model structures. The dependent variable is transportation modes used for commuting to campus. The "Bus" category does not appear in the model because it is used as the base (reference category). The other categories are compared with the reference category.

Table 3. NML Model estimation results

| Var. names          | Modes       | Estimates | Std. Errors | p-values | Var. names          | Modes       | Estimates | Std. Errors | p-values |
|---------------------|-------------|-----------|-------------|----------|---------------------|-------------|-----------|-------------|----------|
| Intercept           | Car         | 0.768     | 0.835       | 0.358    | Vehicle ownership:  | Car         | 9.937     | 0.550       | 0.000    |
|                     | MC          | 32.500    | 1.044       | 0.000    | Car & MC            | MC          | 9.826     | 0.550       | 0.000    |
|                     | Walk        | 32.327    | 1.531       | 0.000    | Walk                | Walk        | 1.257     | 0.000       | 0.000    |
| Gender: Male        | Car         | -2.030    | 1.218       | 0.096    | Vehicle ownership:  | Car         | -19.115   | 1.163       | 0.000    |
|                     | MC          | -0.819    | 0.623       | 0.188    | Car & MC            | MC          | -12.973   | 0.855       | 0.000    |
|                     | Walk        | -18.416   | 0.000       | 0.000    | Walk                | Walk        | -14.922   | 1.382       | 0.000    |
| Native              | Car         | 0.014     | 0.053       | 0.794    | Vehicle ownership:  | Car         | -20.702   | 1.285       | 0.000    |
|                     | MC          | 0.044     | 0.037       | 0.240    | Car & MC            | MC          | -15.883   | 0.856       | 0.000    |
|                     | Walk        | 0.039     | 0.039       | 0.319    | Walk                | Walk        | -13.659   | 1.399       | 0.000    |
| Monthly allowance   | Car         | 0.918     | 0.249       | 0.000    | License: MC         | Car         | -18.322   | 1.461       | 0.000    |
|                     | MC          | 0.223     | 0.179       | 0.212    | Car & MC            | MC          | -15.540   | 0.853       | 0.000    |
|                     | Walk        | -0.401    | 0.207       | 0.053    | Walk                | Walk        | -16.412   | 1.609       | 0.000    |
| Transport Cost      | Car         | 31.503    | 0.636       | 0.000    | License: None       | Car         | -16.686   | 1.331       | 0.000    |
|                     | MC          | -1.041    | 0.587       | 0.076    | MC                  | MC          | -17.301   | 0.859       | 0.000    |
|                     | Walk        | -1.446    | 0.743       | 0.052    | None                | Walk        | -18.033   | 1.229       | 0.000    |
| Residential type:   | Car         | 32.333    | 0.595       | 0.000    |                    | Car         | 0.036     | 0.380       | 0.925    |
| Home                | MC          | 0.625     | 0.562       | 0.266    |                    | MC          | -0.047    | 0.058       | 0.418    |
|                     | Walk        | 0.998     | 0.582       | 0.086    |                    | Walk        | -0.149    | 0.088       | 0.091    |

There is no systematic relationship between student income (in IDR 100 thousand units) and the choice between bus and other means of transportation. For monthly transportation expenses in IDR 100 thousand units, the positive and significant coefficients for "Car" (p < 0.001) thus show that the more the transportation expenses, the higher probability of "Car" being preferred over bus transportation (reference category). One unit (IDR 100 thousand) increase in transportation expenses is associated with a 0.918 increase in the relative log odds of driving a car compared to riding a bus. The negative coefficients for "Walk" (p = 0.053) show that the more the transportation expenses, the marginally lower probability of walking being preferred over bus transportation (reference category).

Regarding the residential type variable, the reference category is "Dormitory." It does not appear in the model. Other residential type categories, i.e., "Home" and "Rented room," are compared to "Dormitory." The positive and significant coefficients for "Home" compared to "Dormitory," thus show a significantly higher probability of "Car" being preferred over bus transportation (reference category). Also, the negative and significant coefficients for "MC" and "Walk," thus show a significantly lower
probability of motorcycles and walking being preferred over bus transportation. The relative log odds of choosing a car vs. bus will increase by 28.85 if moving from the "Dormitory" residential type to "Home." Residential type is a significant variable, although, for motorcycle choice, the significance is marginal. As for the "Rented Room" category, compared with the reference category (Dormitory), positive coefficients of car, motorcycle, and walking show a higher probability of these modes of transport from dormitory to rented room.

As with the residential type variable, the vehicle ownership variable appears essential in explaining the students' choice between bus transportation and other means of transportation, i.e., car, motorcycle, and walking. The base (reference) category for this variable is the car. The other categories, i.e., Car & MC, MC and None, are compared to the Car. The positive and significant coefficients of the car (or motorcycle, or walk) for the "Car & MC" category show a significantly higher probability of this mode of transport being preferred over bus transportation if the vehicle ownership category moves from "Car" to "Car & MC." Conversely, the negative and significant coefficients for vehicle ownership category "MC" show a significantly lower probability of those modes of transport, i.e., car, motorcycle, and walking being preferred over bus transportation if the vehicle ownership category moves from "Car" to "MC." A similar trend is actual if the vehicle ownership category moves from car to none.

Driving/riding license ownership is a categorical variable with four categories, i.e., "A" (car), "C" (motorcycle), "A and C" (car and motorcycle), and "None" categories. The base category is car driving license ownership. The negative and significant coefficients show a significant decrease in the relative log-odds of choosing a particular mode of transport compared to ride a bus. For example, moving the driving license ownership from car to car and motorcycle is associated with a 15.22 decrease in the relative log-odds of driving a car compared to riding a bus. Regarding the commuting distance variable, there appears to be no systematic relationship between travel distance to campus and the choice between riding a bus and other means of transport, i.e., driving a car, riding a motorcycle, or walking. A positive coefficient for a car shows a higher probability of driving a car being preferred more than taking a bus as the distance increases. A negative coefficient shows a lower probability of riding a motorcycle or walking than riding a bus as the distance increases.

One of the fundamental tasks in analysing data using statistical methods is to understand the model prediction accuracy. It is essential to compare the model prediction and the actual data to see if the match is not there or misclassification. Table 4 presents a misclassification error. The rows represent the prediction of the model, and the columns represent the actual data. The model predicts a total of 21 observations ride buses. Based on the real data, of those observations, in fact only 12 ride buses, five ride motorcycles, and four walk. Based on the actual data, there are a total of 41 observations of riding buses, but the model wrongly predicts 14 of those observations as riding motorcycles and 15 walking. The data in the diagonal represents the correct classification, and off-diagonal numbers are incorrect classification. Based on Table 4, the classification error is about 10 percent.

**Table 4. Prediction accuracies**

| Prediction | Actual observation |
|------------|--------------------|
|            | Bus    | Car   | MC    | Walking |
| Bus        | 12     | 0     | 5     | 4       |
| Car        | 0      | 36    | 5     | 0       |
| MC         | 14     | 6     | 645   | 21      |
| Walking    | 15     | 0     | 17    | 51      |

5. Conclusions
The analysis of this study can be satisfactorily said that the multinomial logit model provides excellent performance in analysing mode choice. It was observed that the multinomial logit model could predict four modes, i.e., bus, car, motorcycle, and walking, with an overall accuracy of 90%. Amongst the
influencing variables, vehicle ownership, driving/riding license, and transport cost are the most significant determinants of modal choice.

References

[1] Anonymous 2020 Jawa Tengah Province in Figures 2020
[2] Khattak A, Wang X, Son S and Agnello P 2011 Transp. Res. Rec. 2255 137–145
[3] Limanond T, Butsingkorn T and Chermkhuenthod C 2011 Transp. Policy 18 163–171
[4] Nash S and Mitra R 2018 J. Transp. Geogr. 76 200–211
[5] Mohammadzadeh M 2020 J. Transp. Geogr. 87 102788
[6] Nguyen-Phuoc D Q, Amoh-Gyimah R, Tran A T P and Phan C T 2018 Travel Behav. Soc. 13 1–10
[7] Rea L M and Parker R A 2014 Designing and Conducting Survey Research 4th Ed (San Francisco: Jossey-Bass)
[8] Ben-Akiva M and Lerman S R 1985 Discrete Choice Analysis: Theory and Application to Travel Demand (Transportation Studies) 1st Ed (London: The MIT Press)
[9] Thrane C 2015 Tour. Manag. Perspect. 15 115–121
[10] Hasnine M S, Lin T Y, Weiss A and Habib K N 2018 J. Transp. Geogr. 70 161–171
[11] Greene W H 2000 Econometric Analysis 4th Ed (New Jersey: Prentice Hall)
[12] R Core Team 2018 R: A language and environment for statistical computing (Vienna: R Foundation for Statistical Computing)
[13] Venables W N and Ripley B D 2002 Modern Applied Statistics with S 4th Ed (New York: Springer)

Acknowledgements

The authors would like to thank Apriliya Wahyuniingsih, Iis Saputra, Nuha Fairuz Arwa, and Riko Hasnianto of Universitas Negeri Semarang (Unnes) for data collection and Unnes, UINW, Udinus, and Unissula students for completing questionnaires.