Indexed by

INVESTIGATING THE DETERMINANTS OF TRAVEL MODE CHOICE ACROSS AGE CLASSES IN LANGSA, INDONESIA UTILIZING LOGIT MODEL

Sofyan M. Saleh
Universitas Syiah Kuala, Department of Civil Engineering, Banda Aceh, Indonesia, 23111

Fadhlullah Apriandy
Universitas Syiah Kuala, Department of Civil Engineering, Banda Aceh, Indonesia, 23111

Sugiarto Sugiarto
Universitas Syiah Kuala, Department of Civil Engineering, Banda Aceh, Indonesia, 23111

Lulusi Lulusi
Universitas Syiah Kuala, Department of Civil Engineering, Banda Aceh, Indonesia, 23111

Alfi Salmannur
Universitas Syiah Kuala, Department of Civil Engineering, Banda Aceh, Indonesia, 23111

Key words: *discrete choice model, determinants, age-classes, trip attributes, socio-demography, accuracy*

doi:10.5937/jaes0-34044

Cite article:

M. Saleh S., Apriandy F., Sugiarto S., Lulusi L., Salmannur A. (2022) INVESTIGATING THE DETERMINANTS OF TRAVEL MODE CHOICE ACROSS AGE CLASSES IN LANGSA, INDONESIA UTILIZING LOGIT MODEL, *Journal of Applied Engineering Science*, 20(2), 511 - 522, DOI:10.5937/ jaes0-34044

Online aceess *of full paper is available at: www.engineeringscience.rs/browse-issues*

Istraživanja i projektovanja za privredu
Journal of Applied Engineering Science

Original Scientific Paper

doi:10.5937/jaes0-34044

Paper number: 20(2022)2, 957, 511-522

INVESTIGATING THE DETERMINANTS OF TRAVEL MODE CHOICE ACROSS AGE CLASSES IN LANGSA, INDONESIA UTILIZING LOGIT MODEL

Sofyan M. Saleh, Fadhlullah Apriandy, Sugiarto Sugiarto, Lulusi Lulusi, Alfi Salmannur*
Universitas Syiah Kuala, Department of Civil Engineering, Banda Aceh, Indonesia, 23111

There are different preferences in the decision-making process of humans due to stochasticity. Therefore, this study was conducted to investigate the preferences in selecting a particular mode of travel. This involved using discrete choice modeling. The predictive performance of the model was also evaluated with the contribution of each variable to the model. This is useful for stakeholders to evaluate which factors have significant contributions enabling them to adjust policy accordingly. This study made use of surveys which incorporate revealed and stated preferences in the City of Langsa, Aceh, Indonesia to produce 13 variables including trip attributes and socio-demographic characteristics. This study employs tree distinguished models based on age classes within the sample: all-data, old-age class, and young-age class. Seven variables namely trip frequency, willingness to travel frequency, level of education, household transport expenditure, number of family members, travel cost , and travel time exhibit significancy in every model albeit with diverse extents. With negative vectors, travel cost appears to have the greatest magnitude of scale parameter among variables in every model. Furthermore, each model managed to predict the outcome of alternative 1 extremely well, scoring nearly a perfect 100% a piece. However, no model yields a good accuracy rate in predicting alternative 2, with all models scoring below 15%. All models exhibit good overall accuracy rates, correctly predicting in at least 7 out of 10 times.

Key words: discrete choice model, determinants, age-classes, trip attributes, socio-demography, accuracy

INTRODUCTION

Significant economic growths have been boosted by the implementation of free movement agreement for people and goods across borders and this is partly aided by low-fare air carrier expansion. This means air transport is mostly preferred due to its minimum low travel time, but seaborne transportation provides an alternative particularly for maritime countries like Indonesia. This is due to its abundance of maritime wealth, more eco-friendliness in term of CO2 emission per cargo ton [1], and the recent reduction in its fare [2]. Therefore, people's willingness to make use of maritime instead of air transport need to be investigated. This is necessary to assist stakeholders in making appropriate investment decision in a sea-borne transportation system due to its economic feasibility. Moreover, it is possible to model the preference of people in selecting a travel mode using the conventional statistical method called discrete choice modeling as well as more advanced approaches such as machine learning. A discrete choice model is usually applied to understand the effects of people's behavior on their ability to decide. It has been implemented in extensive sectors such as transportation, economics, energy consumption, and several others with different goals including preference analysis, market share prediction, etc. [3]. Mode choice modeling has also been used to evaluate policy implications by developing accurate travel demand forecasts [4-7]. Moreover, Random Utility Maximization (RUM) has been widely used in modeling travel choice behavior. It is established on a premise that each person aims to optimize its utility based on a set of attributes after which the most optimized alternative is selected. As the name implies, it is a random process which involves each person having distinct objective functions to optimize certain attributes more than the others. The model has the capacity to reflect an individual's perceived value of these attributes. It is, however, impossible to make a definite and precise prediction of every person's choice but the model can estimate the probability of a person opting for each alternative. This involved the consideration of the choice with the highest value as an individual's preferred choice. In contrast to RUM, Random Regret Minimization (RRM) is driven by aims to elude a case where, on some attribute, some non-chosen alternative outperform the chosen alternative hence its objective is preventing 'regret' [8, 9]. However, empirical findings suggest that there is minimum discrepancy between RRM and RUM in term of performance. RUM aims to investigate the relationship between choice alternatives and their possible determining factors. Previous studies suggested that this relationship tends to be non-linear and one of the most dominantly used techniques to model a non-linearity is the logit model. It is based on the consumer economic theory originally developed by McFadden [10] to explore the probability of a person making a particular choice. This concept is fed on a previously set utility function. It is considered simple but also allows unobserved variables to be represented in the term of stochasticity. According to Aloulou [11], this model performs well in separating determinants and analyzing a person's random behavior towards discovering the tendency in line with motivational and characteristics ground. Moreover, the logit model assumes that each choice is independent and identically

*fadhlullah.apriandy@unsyiah.ac.id

distributed (IID) and there is also the need to satisfy the premise of independence of irrelevant alternatives (IIA) to obtain unbiased predictions and consistent parameter estimations [12]. This model often ignores relationships among dependent variables because they are difficult to be included in the utility function [13]. Most researches in the scope of discrete choice modelling evaluate the model solely on statistical inference on the model and falling short of testing the model. Parady et al. [14] argue that the reliance on statistical goodness-of-fit in discrete choice models is not appropriate. The model should also be validated and has its predictive ability tested. In evaluating the predictive performance of the model, this study made use the k-fold cross-validation. This procedure is aimed to reduce bias and avoid overfitting. K-fold approach divided the data into k-subsets with approximately same size, where k-1 groups of data were used to train the model while the remaining group was fitted into the model to test, make a prediction, and obtain the accuracy score. This process was repeated for k-time as cross-validation in which each subset was treated as the test set exactly once. Therefore, this study aimed to investigate people's preference in selecting their travel mode choice using discrete choice modeling particularly in the context of a medium-sized city of Langsa with a population of 176,811. The effects of every variable on the model were evaluated. Each variable contribution is a useful consideration for stakeholders to evaluate which factors contribute the most hence they can adjust policy accordingly. The models were split based on age allowing for an inference on variables distinguished impacts across age groups. Moreover, the models' predictive capabilities were tested for further evaluation. The remainder of this paper is divided into some sections beginning with survey and data description, methodological approach, parameter estimation, result discussion, and conclusion and future work and the end.

MATERIAL AND METHODS

Material

This study focused on the City of Langsa, Aceh, Indonesia which is located 400 km north-east of provincial capital Banda Aceh and has a population of 176,811 [15] contributing to a 737 people/km2 density. An international passenger sea route connecting Langsa and Penang, Malaysia, as shown in Figure 1, was opened in 2013 but was then terminated after operating for merely 3 months accommodating 35 trips. This study is also part of a project which is originally aimed to evaluate and determine whether the city government should reverse its decision and reoperate the route. This led to the conduct of a survey for almost two weeks in November 2019 using eight surveyors that randomly disseminated questionnaires in-person and assist respondents in filling in the questions attributed to sociodemographic, trip attribute, and travel mode choice. The survey area is illustrated in Figure 2. Moreover, nine different hypothetical scenarios of travel cost and time were designed and included in the paper sheets to determine the preferred mode. The variables used in the logit model were derived from these questions. Only responses returned in complete form were considered for this study. A total of 402 questionnaires were adjudged to be valid for the study and the questions presented are summarized in Table 1. A small number of samples could significantly impair the normality of the dataset [16, 17], however with a sample size of at least 200, the normal distribution of the dataset should be maintained [16]. Using a formula stated in Lwanga and Lameshow [18] to estimate the sufficient number of sample, it is found that with 95% confidence level and 10% sampling error, there should be at least 96 dataset. Hence, the dataset used in this study should represent the population properly thus warrant the validity of this study.

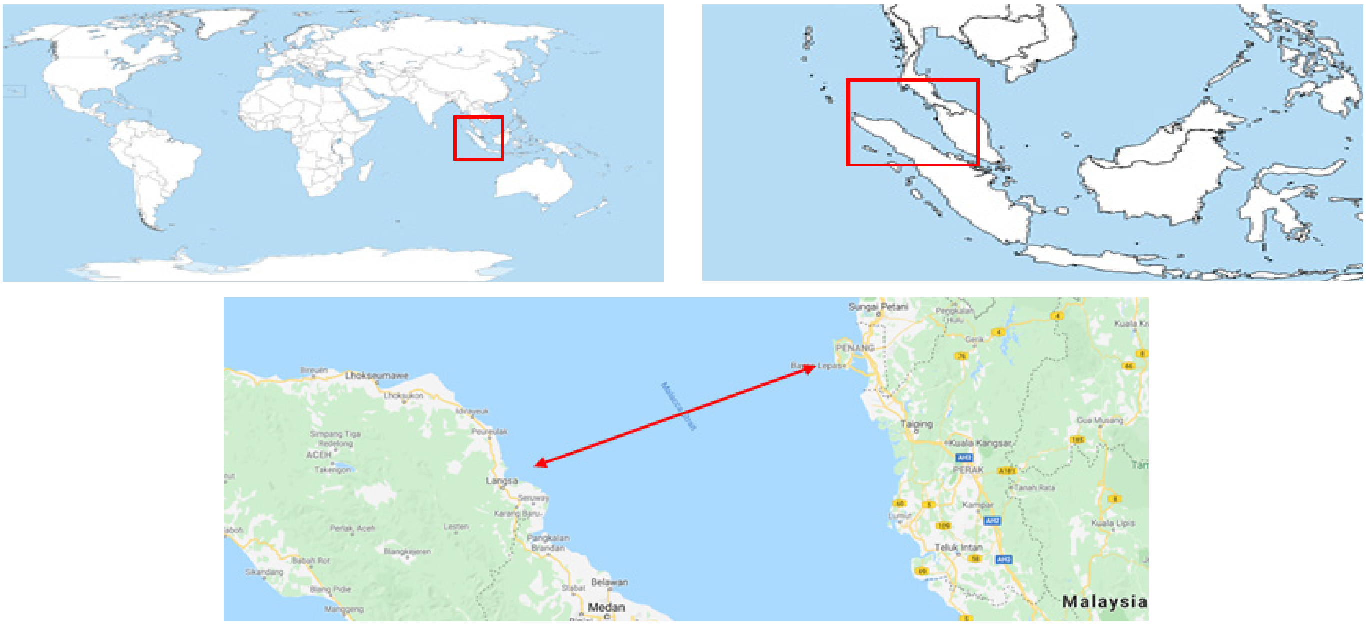

Figure 1: An international passenger sea-route connecting Langsa and Penang

The questionnaire presented three different alternative modes of traveling to the respondents. Alternative 1 refers to ship through Langsa-Penang sea-route while alternatives 2 and 3 correspond to airplane through Sultan Iskandar Muda Airport in Banda Aceh (nearly 450 km northwest of Langsa) and Kuala Namu Airport in Medan (more than 200 km southwest of Langsa) respectively which are reachable through car or bus. The fare for alternative 1 was set at IDR 500,000 while alternative 2 costs IDR 1,110,000 and alternative 3 costs IDR 925,000 in typical days. Moreover, alternative 1 normally takes 6-7 hours to complete, alternative 2 takes 10-12 hours, and alternative 3 takes 8-9 hours depending on the weather condition. There were nine hypothetical scenarios involving combinations of ship travel time and cost which were introduced to respondents and these included unchanged time and 45% cheaper cost, unchanged time and 30% cheaper, unchanged time and 15% cheaper, 10% faster time and 45% cheaper cost, 10% faster time and 30% cheaper cost, 10% faster time and 15% cheaper cost, 20% faster time and 45% cheaper cost, 20% faster time and 30% cheaper cost, and 20% faster time and 15% cheaper cost. The respondents were asked whether they prefer to take the sea route in each of these scenarios. These hypothetical schemes are useful to understand the perception of the respondents about the value of travel time and cost as well as the importance of setting appropriate fare when introducing a travel mode. In total there were ten distinct scenarios each respondent needed to answer which gave an equivalent of 4020 observations. Table 2 shows the descriptive statistics of the responses obtained. Alternative 1 was observed to have been selected in nearly every four out of five observations and on average respondents travel from Langsa to Penang once in every quarter. It reflects the study conducted by Bergantino et al. [19] which indicates accessibility as a primary constituent in airport provision. It confirms that cost and time taken to access affect travelers' preferences.

Figure 2: Area of survey [20]

Table 1: Summary of the questionnaire

Item	Description
Trip	
Mode	Travel mode choice
Frequency	How many times respondent travel per month
Willingness frequency	How many times respondent is willing to travel using alternative two
Individual	
Gender	Gender of respondent
Age	Age of respondent
Education	Education level of respondent
Employment	Employment of respondent
Household	
Income	Household income of respondent per month
Transport expenditure	Transport expenditure of respondent per month
Family size	Number of respondent's family member
Car	Number of cars owned by respondent
Motorbike	Number of motorbikes owned by respondent
Travel time scenarios	unchanged, 10%, or 20% faster
Travel cost scenarios	15%, 30%, or 45% cheaper

Table 2. Descriptive statistics of respondents' responses

Item	Category	Share	Mean	Min	Max	Std. Dev.
Trip						
Mode (total)	Alternative 1	78.73 %	-	-	-	-
	Alternative 2	0.17 %	-	-	-	-
	Alternative 3	21.10 %	-	-	-	-
Existing Frequency	-	-	0.25	0	6	1.06
Willingness frequency	-	-	2.86	0	6	2.48
Individual						
Gender	Male	76.87 %	-	-	-	-
	Female	23.13 %	-	-	-	-
Age	17-19 years	1.00 %	-	-	-	-
	20-29 years	27.61 %	-	-	-	-
	30-39 years	25.62 %	-	-	-	-
	40-49 years	22.39 %	-	-	-	-
	50-59 years	17.66 %	-	-	-	-
	> 60 years	5.72 %	-	-	-	-
Education	Elementary school	8.46 %	-	-	-	-
	Junior high school	7.96 %	-	-	-	-
	Senior high school	62.93 %	-	-	-	-
	Vocational study	4.23 %	-	-	-	-
	Undergraduate	16.17 %	-	-	-	-
	Postgraduate	0.25 %	-	-	-	-
Employment	Civil Servant/ Military officer/ Police officer	2.74 %	-	-	-	-
	General em-ployee	12.19 %	-	-	-	-
	Entrepreneur	50.00 %	-	-	-	-
	Part-time/ retiree	2.98 %	-	-	-	-
	Housewife	8.96 %	-	-	-	-
	Others	23.13 %	-	-	-	-

Item	Category	Share	Mean	Min	Max	Std. Dev.
Household						
Income	IDR 1,000,000 or less	19.15 %	-	-	-	-
	IDR 1,000,000-3,000,000	53.23 %	-	-	-	-
	IDR 3,000,000-5,000,000	23.63 %	-	-	-	-
	IDR 5,000,000-7,000,000	2.49 %	-	-	-	-
	IDR 7,000,000-10,000,000	0.75 %	-	-	-	-
	More than IDR 10,000,000	0.75 %	-	-	-	-
Transport expenditure	IDR 500,000 or less	68.66 %	-	-	-	-
	IDR 500,000-1,000,000	21.14 %	-	-	-	-
	IDR 1,000,000-1,500,000	8.46 %	-	-	-	-
	IDR 1,500,000-3,000,000	0.99 %	-	-	-	-
	IDR 3,000,000-3,500,000	0.50 %	-	-	-	-
	More than IDR 3,500,000	0.25 %	-	-	-	-
Family size	-	-	4.09	1	9	1.64
Car	-	-	0.10	0	4	0.38
Motorbike	-	-	1.45	0	7	0.82

It is important to note that more than three-quarters of the respondents were male and those aged 20-29 years were observed to be the most dominant age. Moreover, more than six-tenths of the respondents were graduates of senior high school while half of the sample set were entrepreneurs. It was also discovered that only 5% of the respondents earned more than IDR 5 million in a month while only one-tenth spent more than IDR 1 million for transportation purposes monthly. The mean of the respondent's family members was approximately 4 while each household averagely had 0.10 cars and 1.45 motorbikes. Meanwhile, the lack of data on alternative 2 led to its combination with alternative 3. There is a need to arrange the data in typical ways to help with model calibration due to its distribution. For example, the income section was sorted into two distinct clusters including the middle-low income (earning less than IDR 3 million) and middle-high income (earning more than IDR 3 million) instead of the previous six classes. Similarly, it also applies to the transport expenditure which was arranged into middle-low expenditure (spending less than IDR 500,000) and middle-high expenditure (spending more than IDR 500,000), employment into entrepreneur and others, age into young (40 years or younger) and old (40 years or older), education level into secondary education and non-secondary education graduates, and mode choice into alternative 1 (ship) and alternative 2 (airplane via airport either in Banda Aceh or Medan). There was small difference between the share of income class with its corresponding transport expenditure class, as households with higher income tend to allocate more resource for transportation expenditure [21, 22]. Table 3 illustrates the data after some adjustments have been made.

Method

Some observed and deterministic factors influence the utility function in the logit model and they relate to the individual's socioeconomic characteristics, the technical aspect of the alternatives, and the environment surrounding the choice [11]. An unobserved factor which reflects the stochasticity of individuals' ultimate choice and likely to be undetected by the model was also added to the model. The logit model assumes each unobserved factor is IID with a distribution of Gumbel type I extreme

Table 3: Data distribution after pre-processing

Item	Category	Share	Mean	Min	Max	Std. Dev.
Trip						
Mode (total)	Alternative 1	78.73 %	-	-	-	-
	Alternative 2	21.27 %	-	-	-	-
Trip frequency (times/month)	-	-	0.25	0	6	1.06
Willingness to travel frequency (times/month)	-	-	2.86	0	6	2.48
Individual						
Gender	Male	76.87 %	-	-	-	-
	Female	23.13 %	-	-	-	-
Age	Young < 40	54.23 %	-	-	-	-
	Old > 40	45.77 %	-	-	-	-
Education	Secondary	91.54 %	-	-	-	-
	Non-second-ary	8.46 %	-	-	-	-
Employment	Entrepreneur	50.00 %	-	-	-	-
	Others	50.00 %	-	-	-	-
Household						
Income	Middle-low	72.38 %	-	-	-	-
	Middle-high	27.62 %	-	-	-	-
Transport expenditure	Middle-low	68.66 %	-	-	-	-
	Middle-high	31.34 %	-	-	-	-
Family size	-	-	4.09	1	9	1.64
Car	-	-	0.10	0	4	0.38
Motorbike	-	-	1.45	0	7	0.82

value and the utility function is illustrated in the following Equation 1.

$$U_i = f(X_i) + \varepsilon_i \quad (1)$$

It can also be written in linear expression as shown by Equation 2

$$U_i = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \cdots + \varepsilon_i \quad (2)$$

Where:

U_i = utility function of each alternative

X_i = factors affecting utility functions which include individual's socioeconomic characteristics, technical aspect of the alternatives, and environment surrounding the choice

α = scale parameters (coefficient of the variables and constant)

ε_i = random unobserved term

$$P_{ij} = \frac{e^{U_{ij}}}{\sum_{j=1}^{J} e^{U_{ij}}} \quad (3)$$

The choice with the highest utility value has the highest probability to be selected by an individual and is considered as the chosen alternative. Data obtained from observations were utilized to calibrate the scale parameters α. Maximum Likelihood Estimator (MLE) was employed in scale parameters calibration process, which was conducted in Python-based module named Biogeme [23]. In calibrating the model, some parameters called the Goodness-of-Fit were set as bars to determine whether the subsequent model fit well enough. These parameters are Pseudo R-squared ($\bar{\rho}^2$) and significant student t-test. $\bar{\rho}^2$ is defined as how far the ratio of the log of convergent model to the log of the initial model deviates from 1. The greater the value of $\bar{\rho}^2$ is, the better the model is. $\bar{\rho}^2$ ranges from 0 to 1, and a 0.2 – 0.4 value of $\bar{\rho}^2$ indicates an excellent model [24]. The scale parameter α was evaluated using a student t-test with a maximum of 10% significance. A high significant error indicates an insignificant parameter, consequently this parameter was then excluded from the model. This assessment was executed repeatedly until all remaining parameters in the model are considered significant. This procedure is illustrated in Figure 3.

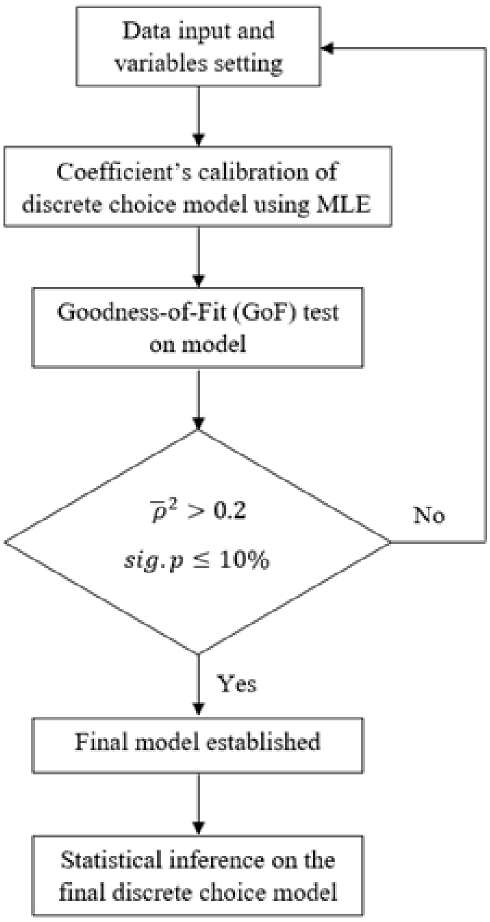

Figure 3: Model calibration flowchart

The factors commonly employed by previous studies in investigating individual travel mode choice include the sociodemographic attributes such as the individual and household characteristics, trip attributes, mode-specific dynamics, built environment, environmental attributes, urban design, health-related characteristics, and meteorological conditions [25-28]. Factors used as the variables in this study are, therefore, presented in Table 4. In specifying the appropriate and good utility functions, trial-and-error was employed until all variables used in the models are significant which is unfortunately often time-consuming and vulnerable to flaws [29].

Table 4: Variables employed

Acronym	Variable
ASC	Alternative Specific Constant
TF	Trip frequency (times/month)
WF	Willingness to travel frequency (times/month)
G	Gender: male = 1, female 0.
A	Young age: 40 years old or younger = 1, otherwise 0
LE	Level of education: secondary or over = 1, otherwise 0
E	Employment: entrepreneur = 1, otherwise 0
I	Household income: at least IDR 3 million = 1, otherwise 0
TE	Household transport expenditure: IDR 500 thousand or more = 1, otherwise 0
FM	Number of family members
C	Number of cars owned
M	Number of motorcycles owned
TC	Travel cost of alternative 1 (million IDR)
TT	Travel time of alternative 1 (hour)

K-fold cross validation was utilized to assess the performance of each model in predicting travel model choice. It allows a minimization of bias and prevention of overfitting. K-fold split the data into k-subsets with approximately same size. The first k-1 classes were used to train the model which ultimately was validated using the remaining class. This would return a prediction which would be compared to the actual outcome to obtain accuracy rate. The value of k used in this research was 10, hence this process was repeated for 10 times as cross-validation. Each subset was treated as the test set in only one corresponding repetition.

RESULTS AND DISCUSSION

The logit model was calibrated using Biogeme developed by Bierlaire [23] on Python. The distribution of mode choice as dependent variable is presented in Figure 4. Of all 4020 observations, 78.73% were in favor

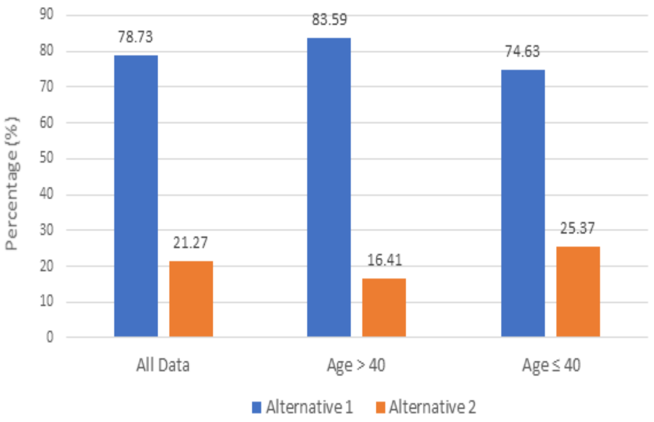

Figure 4: Mode choice distribution

of alternative 1. Should the age classes be taken into account, 83.59% of respondents aged over 40 years preferred alternative 1 than 2, whereas that number decreased to 74.63% among young respondents (40 years old or younger). Table 5 shows the summary of the model alongside the coefficient of each variable on models either while using the entire dataset or when data being classified according to age classes. The value of (ρ)¯^2 are 0.31, 0.45, and 0.24 in all-data model, old-age model, and young model respectively, all of which are deemed as excellent [24]. Of all 14 variables, 13 were found to be significant in the all-data-model, leaving variable "number of cars owned" to be excluded. That figure, though, lessen when the models were built upon age classes: only 9 variables were significant in old-age class and a mere 8 variables in young-age class. 7 variables namely trip frequency, willingness to travel frequency, level of education, household transport expenditure, number of family members, travel cost of alternative 1, and travel time of alternative 1 exhibit significancy in every model albeit with diverse extents. While employment was insignificant in old-age model as was household income in young-age model, the following variables showed importance in exactly one of the models: gender in all-data model; number of cars owned in old-age model; and number of motorcycles owned in all-data model. Travel cost of alternative 1 appears to have the greatest magnitude of scale parameter among variables in every model. Each value is shown to be negative, meaning that it leans toward the opposite alternative. An IDR 1 million increase in alternative 1 travel cost will contribute to an increase of alternative 2 utility function as high as 2.48 in the all-data model, 1.57 in old-age model, and 3.19 in young-age model. More importantly, it is the change of the probability that is of the most interest. A similar travel cost raise result in 92%, 79%, and 96% less likely a person would choose alternative 1 mode in the corresponding models. It indicates a pivotal role a reasonable fare could play in mode choice study, particularly in the context of new travel mode implementation. Considering each variable effect on alternative 1 utility function, level of education provides the biggest positive contribution in all-data model. Should a person has graduated at least a secondary study, the utility is expected to go up by 0.47 which generate a 60% increase in its parallel probability function. Even though it similarly occurs in old-age model with much larger degree (1.00 utility and 172% probability rise), a contrasting trend is observed in young-age model whose negative vector adding up alternative 2 utility by 1.18 and 69% less likely a person would pick alternative 1. A value of 1.23 belonging to trip frequency represents the largest sole positive coefficient on old-age utility model. It accounts for a mammoth 242% soar in alternative 1 probability. However, a reverse tendency is exhibited in all-data and young-age model, both of which contain negative vectors reflecting less likelihood. Furthermore, young-age model sees employment as providing its biggest positive part, with 0.40 coefficient portrays 49% more likely people would prefer alternative 1 if they are entrepreneur. The utility equations generated in this study are written in Equation 4, 5 and 6.

All data model:

$$U = 3.42-0.37TF-0.52WF+0.17G-0.31A+0.47LE + 0.24E+0.22I-1.02TE+ 0.13FM-0.11M - 2.48TC-0.22TT \quad (4)$$

Old age model:

$$U = 1.59+1.23TF-0.51WF+LE+0.59I-1.17TE + 0.52FM-0.44C-1.57TC-0.32TT \quad (5)$$

Young age model:

$$U=5.51-0.44TF-0.41WF-1.18LE+0.40E - 0.89TE- 0.13FM-3.19TC-0.19TT \quad (6)$$

Table 6 describes the variation between the predicted outcome and the actual outcome of travel mode choice across sample size. It can be seen that there are insignificant deviations shown between alternative 1 predicted and actual data, whereas greater discrepancies are displayed in alternative 2 cases. Figure 5 shows the accuracy attain by each model in accordance with the alternatives. Each model managed to predict the outcome of alternative 1 extremely well, scoring nearly a perfect 100% a piece. That is not the case, however, in forecasting alternative 2. No model has shown a good prediction accuracy. The most accurate one, the old-age model, somehow display a mere 13.6% accuracy score, whereas the others score much lower figures (6.8% for all-data and 9.8% for young-age). Limited data on alternative 2 might play a huge part to blame. It might prevent the program to better train and calibrate the models. Old-age model was found to be the best performing model overall, with a rate of 83% indicating a substantial increase from all-data model (78.6%) and young age model (74.2%).

Table 5: Variable coefficient estimation

Name	Variable	All Data			Age > 40			Age ≤ 40		
		Coef.	$e^{coef.}$	Sig.	Coef.	$e^{coef.}$	Sig.	Coef.	$e^{coef.}$	Sig.
ASC	Alternative Specific Constant	3.42	-	0.00	1.59	-	0.05	5.51	-	0.00
α_{TF}	Trip frequency (times/month)	-0.37	0.69	0.00	1.23	3.42	0.01	-0.44	0.64	0.00
α_{WF}	Willingness to travel frequency (times/month)	-0.52	0.59	0.00	-0.51	0.60	0.00	-0.41	0.66	0.00
α_{G}	Gender: male 1, otherwise 0	0.17	1.19	0.08	-	-	-	-	-	-
α_{A}	Age: young 1, otherwise 0	-0.31	0.73	0.00	-	-	-	-	-	-
α_{LE}	Level of education: secondary or over 1, otherwise 0	0.47	1.60	0.00	1.00	2.72	0.00	-1.18	0.31	0.00
α_{E}	Employment: entrepreneur 1, otherwise 0	0.24	1.27	0.00	-	-	-	0.40	1.49	0.00
α_{I}	Household income: middle-high 1, otherwise 0	0.22	1.25	0.03	0.59	1.80	0.00	-	-	-
α_{TE}	Household transport expenditure: middle-high 1, otherwise 0	-1.02	0.36	0.00	-1.17	0.31	0.00	-0.89	0.41	0.00
α_{FM}	Number of family members	0.13	1.14	0.00	0.52	1.68	0.00	-0.13	0.88	0.00
α_{C}	Number of cars owned	-	-	-	-0.44	0.64	0.06	-	-	-
α_{M}	Number of motorcycles owned	-0.11	0.90	0.03	-	-	-	-	-	-
α_{TC}	Travel cost of alternative 1	-2.48	0.08	0.00	-1.57	0.21	0.10	-3.19	0.04	0.00
α_{TT}	Travel time of alternative 1	-0.22	0.80	0.00	-0.32	0.73	0.01	-0.19	0.83	0.06
Statistical summary										
Sample size (N)		4020			1840			2180		
LL (initial)		-2786.45			-1275.39			-1511.06		
LL (β)		-1910.94			-687.30			-1134.38		
$\bar{\rho}^2$		0.31			0.45			0.24		

Table 6: Confusion matrix with k-fold cross-validation

All-data model		Predicted choice		Alternative accuracy	Overall accuracy
		Alternative 1 3901	Alternative 2 119		
Actual choice	Alternative 1 3165	3104	61	98.1%	78.6%
	Alternative 2 855	797	58	6.8%	
Old-age model		Predicted choice		Alternative accuracy	Overall accuracy
		Alternative 1 1750	Alternative 2 90		
Actual choice	Alternative 1 1538	1489	49	96.8%	83.0%
	Alternative 2 302	261	41	13.6%	
Young-age model		Predicted choice		Alternative accuracy	Overall accuracy
		Alternative 1 2062	Alternative 2 118		
Actual choice	Alternative 1 1627	1563	64	96.1%	74.2%
	Alternative 2 553	499	54	9.8%	

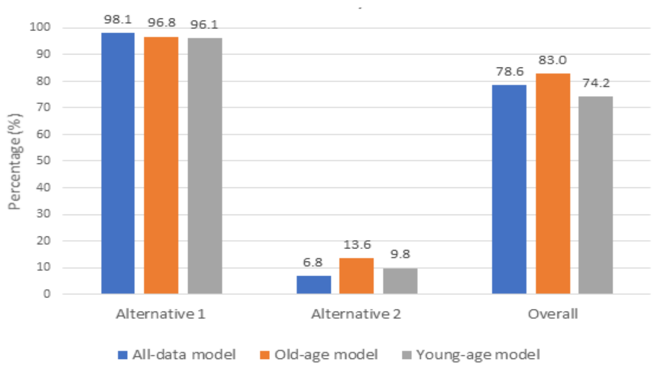

Figure 5: Model accuracy

CONCLUSION AND FUTURE WORKS

People have preferences in making a decision when faced with several distinguished alternatives and these preferences often differ to indicate the existence of stochasticity. The investigation of people's preferences requires methods such as those related to statistics like discrete choice modeling. This study was, therefore, conducted to investigate people's preferences when making choices on the mode of transportation to be used in traveling through the application of discrete choice modeling. The predictive performance of the method was evaluated alongside the contribution of each variable. Each variable contribution is a useful consideration for stakeholders to evaluate which factors contribute the most that could be enhanced to further stimulate travel demand. The surveys including revealed and stated preferences were conducted in the city of Langsa, Aceh, Indonesia, and the results obtained were analyzed. The 13 variables used include trip attributes such as travel cost and time as well as socio-demographic characteristics such as gender, age, household income, household transport expenditure, education level, employment, and others. This study employed tree distinguished model based on age classes within the sample: all-data, old-age class, and young-age class. Each model exhibits an excellent (ρ)̅^2 value. 7 variables namely trip frequency, willingness to travel frequency, level of education, household transport expenditure, number of family members, travel cost of alternative 1, and travel time of alternative 1 exhibits significancy in every model albeit with diverse extents. Travel cost of alternative 1 appears to have the greatest magnitude of scale parameter among variables in every model. A negative vector indicates that travel cost of alternative 1 lean toward the opposite alternative. Level of education, trip frequency, and employment are the biggest sole positive contributors to alternative 1 in all-data, old-age, and young-age models respectively. In analyzing each model's performance metric, a k-fold cross validation was utilized. Insignificant deviations between predicted and actual data are shown in alternative 1, whereas greater discrepancies are displayed in alternative 2 cases. Each model managed to predict the outcome of alternative 1 extremely well, scoring nearly a perfect 100% a piece. However, no model yields a good accuracy rate in predicting alternative 2, with all models scoring below 15%. Limited data on alternative 2 might play a huge part to blame, which might prevent the program to better train and calibrate the models. All models exhibit good overall accuracy rates, correctly predicting in at least 7 out of 10 times. This research analyzes trip and socio-demographic attributes as determinant in trip mode choice. Future works should incorporate other characteristics such as built environment, weather and climate, and health indicator as the variables. It will enable a more comprehensive and extensive understanding of magnitude of impacts from various factors. Furthermore, utilizing other branches of logit models such as nested logit and mixed logit could possibly yield different outcome which will enrich valuable insight into the statistical inferences. Finally, future study should consider focusing the study on different type of area such as big or metropolitan cities that could generate distinguished estimations

ACKNOWLEDGEMENT

The authors would like to send their sincerest gratitude to the Universitas Syiah Kuala for providing financial support for this study under Contract No. 268/UN11/SPK/PNBP/2020 and also appreciate the surveyors for their assistance in the data collection. The authors also claim all remaining oversight in this study.

REFERENCES

1. Grant, D.B., and Elliott, M. (2018). A proposed interdisciplinary framework for the environmental management of water and air-borne emissions in maritime logistics. Ocean & coastal management, vol. 163, 162-172, DOI: 10.1016/j.ocecoaman.2018.06.011.
2. Cordón-Lagares, E., and García-Ordaz, F. (2020). Factors affecting the survival of maritime goods transport firms in Spain. Research in Transportation Business & Management, vol. 37, 100520, DOI: 10.1016/j.rtbm.2020.100520.
3. Bouscasse, H., Joly, I., and Peyhardi, J. (2019). A new family of qualitative choice models: An application of reference models to travel mode choice. Transportation Research Part B: Methodological, vol. 121, 74-91, DOI: 10.1016/j.trb.2018.12.010.
4. Ben-Akiva, M., and Lerman, S.R. (2018). Discrete choice analysis: theory and application to travel demand. MIT Press, Cambridge.
5. de Dios Ortúzar, J., and Willumsen, L.G. (2011). Modelling transport. John wiley & sons, West Sussex.
6. Jara-Díaz, S. (2007). Transport economic theory. Elsevier, Oxford.
7. Schmid, B., Jokubauskaite, S., Aschauer, F., Peer, S., Hössinger, R., Gerike, R., Jara-Diaz, S.R., and Axhausen, K.W. (2019). A pooled RP/SP mode, route and destination choice model to investigate mode and user-type effects in the value of travel time savings. Transportation Research Part A: Policy and Practice, vol. 124, 262-294, DOI: 10.1016/j.tra.2019.03.001.
8. Hess, S., Beck, M.J., and Chorus, C.G. (2014). Contrasts between utility maximisation and regret minimisation in the presence of opt out alternatives. Transportation Research Part A: Policy and Practice, vol. 66, 1-12, DOI: 10.1016/j.tra.2014.04.004.
9. Chorus, C.G. (2010). A New Model of Random Regret Minimization. European Journal of Transport and Infrastructure Research, vol. 10, no. 2, DOI: 10.18757/ejtir.2010.10.2.2881.
10. McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. Frontiers in Econometrics, 105-142.
11. Aloulou, F. (2018). The Application of Discrete Choice Models in Transport. In Statistics - Growing Data Sets and Growing Demand for Statistics.
12. Cheng, L., Chen, X., De Vos, J., Lai, X., and Witlox, F. (2019). Applying a random forest method approach to model travel mode choice behavior. Travel Behaviour and Society, vol. 14, 1-10, DOI: 10.1016/j.tbs.2018.09.002.
13. Lee, D., Derrible, S., and Pereira, F.C. (2018). Comparison of Four Types of Artificial Neural Network and a Multinomial Logit Model for Travel Mode Choice Modeling. Transportation Research Record: Journal of the Transportation Research Board, vol. 2672, no. 49, 101-112, DOI: 10.1177/0361198118796971.
14. Parady, G., Ory, D., and Walker, J. (2021). The overreliance on statistical goodness-of-fit and under-reliance on model validation in discrete choice models: A review of validation practices in the transportation academic literature. Journal of Choice Modelling, vol. 38, DOI: 10.1016/j.jocm.2020.100257.
15. Langsa's Centre for Statistics Board (BPS Langsa). (2020). Langsa Municipality in Figures 2020. BPS Langsa, Langsa.
16. Hair, J.F., Black, W.C., Babin, B.J., and Anderson, R.E. (2019). Multivariate data analysis. Eighth edition. Cengage, Hampshire.
17. Sugiarto, S., Fahlevi, H., Achmad, A., Fajri, L., and Miwa, T. (2021). The relative importance of bus system's Perceived Service Quality (PSQ) attributes among public and private mode users in Indonesia. Journal of Applied Engineering Science, vol. 19, no. 3, 600-609, DOI: 10.5937/jaes0-27957.

18. Lwanga, S.K., Lemeshow, S., and World Health, O. (1991). Sample size determination in health studies : a practical manual / S. K. Lwanga and S. Lemeshow. In. (Geneva, World Health Organization.
19. Bergantino, A.S., Capurso, M., and Hess, S. (2020). Modelling regional accessibility to airports using discrete choice models: An application to a system of regional airports. Transportation Research Part A: Policy and Practice, vol. 132, 855-871, DOI: 10.1016/j.tra.2019.12.012.
20. BAPPENAS. (2021). WebGIS Perencenaan Kementerian PPN, from http://webgis-simrenas.bappenas.go.id, accessed on 2021-11-18.
21. Lulusi, L., Sugiarto, S., Anggraini, R., Apriandy, F., Fadhurrozi, A., Rusdi, M., and Basrin, D. (2021). Travel Cost Budget and Ability of Urban Bus Users to Pay Considering the Income Classes in Indonesia. Transactions on Transport Sciences, vol. 12, no. 1, 19-24, DOI: 10.5507/tots.2021.006.
22. Sugiarto, S., Lulusi, L., Isya, M., Apriandy, F., and Ramadhan, F. (2021). Understanding Household's Travel Costs Budget Frontier in Banda Aceh, Indonesia. Communications - Scientific letters of the University of Zilina, vol. 23, no. 2, A116-A124, DOI: 10.26552/com.C.2021.2.A116-A124.
23. Bierlaire, M. (2020). A short introduction to Pandas-Biogeme. Transport and Mobility Laboratory, ENAC, EPFL, Lausanne.
24. McFadden, D. (1979). Quantitative methods for analysing travel behaviour of individuals: some recent developments. Routledge.
25. Dėdelė, A., Miškinytė, A., Andrušaitytė, S., and Nemaniūtė-Gužienė, J. (2020). Dependence between travel distance, individual socioeconomic and health-related characteristics, and the choice of the travel mode: a cross-sectional study for Kaunas, Lithuania. Journal of Transport Geography, vol. 86, DOI: 10.1016/j.jtrangeo.2020.102762.
26. Cheng, L., Chen, X., Yang, S., Wang, H., and Wu, J. (2016). Modeling Mode Choice of Low-Income Commuters with Sociodemographics, Activity Attributes, and Latent Attitudinal Variables: Case Study in Fushun, China. Transportation Research Record: Journal of the Transportation Research Board, vol. 2581, no. 1, 27-36, DOI: 10.3141/2581-04.
27. Susilo, Y.O., Williams, K., Lindsay, M., and Dair, C. (2012). The influence of individuals' environmental attitudes and urban design features on their travel patterns in sustainable neighborhoods in the UK. Transportation Research Part D: Transport and Environment, vol. 17, no. 3, 190-200, DOI: 10.1016/j.trd.2011.11.007.
28. Böcker, L., Dijst, M., and Prillwitz, J. (2013). Impact of Everyday Weather on Individual Daily Travel Behaviours in Perspective: A Literature Review. Transport Reviews, vol. 33, no. 1, 71-91, DOI: 10.1080/01441647.2012.747114.
29. Ortelli, N., Hillel, T., Pereira, F.C., de Lapparent, M., and Bierlaire, M. (2021). Assisted specification of discrete choice models. Journal of Choice Modelling, vol. 39, DOI: 10.1016/j.jocm.2021.100285.

Paper submitted: 21.09.2021.
Paper accepted: 24.11.2021.