And., Intl.,J., Appl Sci., Eng., Tech, Vol 06, No 03, November 2026
e-ISSN: 2797-0442

Probability Modelling of Motorcycle and Private Car Mode Choice in Sa'dan District, North Toraja

Azril Tangke Sombolinggi*, Rilva Toding Bua'
Department of Civil Engineering, Faculty of Engineering, Universitas Kristen Indonesia Toraja, Indonesia
*Corresponding author, e-mail: azrilsombolinggi@ukitoraja.ac.id

Abstract— This study aimed to develop a binary logistic regression model to examine transportation mode choice between motorcycles and private cars among respondents in Sa'dan District, North Toraja Regency, Indonesia. A quantitative approach was employed using 162 valid observations collected through a household survey. Five explanatory variables were examined: vehicle ownership, household size, number of working household members, number of school-attending household members, and household income. Motorcycles accounted for 70.4% of the observed mode choices, while private cars accounted for 29.6%. Household income was the only statistically significant predictor of private-car choice ($B = 2.303$, $p < 0.001$, $Exp(B) = 10.003$). The overall model was statistically significant ($\chi^2 = 51.014$, $p < 0.001$), with a Nagelkerke R^2 of 0.384 and a non-significant Hosmer–Lemeshow test ($p = 0.639$). The model achieved an apparent classification accuracy of 78.4% and an area under the receiver operating characteristic curve (AUC) of 0.830. The mean predicted probability of private-car choice was 29.63%. These findings indicate that household income was positively associated with private-car choice in the study sample. The study demonstrates the application of binary logistic regression to examine transportation mode choice in a semi-rural area, while highlighting the need for further validation using independent data.

Keywords: transportation mode choice, binary logistic regression, and household characteristics.

This article *is licensed under the CC–BY-SA license.*

1. Introduction

Transportation mode choice is an important component of travel behavior because it reflects how individuals and households select transportation alternatives to satisfy their daily mobility needs. The choice of transportation mode is influenced by the availability of transportation alternatives, household characteristics, socioeconomic conditions, vehicle ownership, and the spatial context in which travel takes place [1]. At the household level, travel decisions are not isolated individual decisions because household composition, residential characteristics, and access to vehicles can influence the allocation of travel and transportation resources among household members [2]. Socioeconomic characteristics also play an important role in shaping travel behavior, particularly in rural and semi-rural areas where transportation alternatives and accessibility may differ substantially from those in urban areas [3].

Motorcycles and private cars are particularly relevant modes of transportation in regions where private vehicle ownership provides greater flexibility for daily activities. Vehicle ownership determines the availability of private transportation alternatives, while household size and the composition of household members may influence the frequency, purpose, and spatial distribution of household trips. The presence of working and school-attending household members can generate different travel requirements and consequently influence household transportation decisions [2]. Previous research has also demonstrated that household characteristics and regional transportation conditions are associated with travel generation in semi-rural areas of North Toraja, indicating the importance of considering household-level characteristics when examining transportation behavior in the region [4].

Received: September 24, 2026, Revised: October 13, 2026, Accepted: November 9, 2026
https://doi.org/10.25077/aijaset.v6i3.375

Transportation mode choice can be investigated using discrete choice models that represent the selection of one transportation alternative from a set of available alternatives. Discrete choice modelling provides a theoretical and analytical framework for relating transportation decisions to characteristics of travelers, households, trips, and transportation alternatives [5]. Logistic regression is particularly suitable when the choice consists of two alternatives because the model estimates the probability of selecting one alternative relative to another while allowing the effects of explanatory variables to be statistically evaluated [6]. The application of statistical and discrete choice approaches has been widely used to examine travel mode choice and to identify the socioeconomic, individual, trip-related, and spatial characteristics associated with transportation decisions [7], [8].

The literature on travel mode choice has developed through several analytical perspectives. Some studies have emphasized household and socioeconomic characteristics as important determinants of travel behavior, while others have incorporated individual, trip, neighborhood, and spatial characteristics into mode choice models [2], [8]. Other studies have compared discrete choice models with machine learning approaches to evaluate their ability to predict travel mode selection, demonstrating the continuing relevance of probabilistic and predictive modelling approaches in transportation research [7]. Research on intercity travel has also examined the effects of data characteristics, travel-chain attributes, and socioeconomic factors on transportation mode choice, indicating that mode selection can be influenced by multiple dimensions of travel behavior [9]. These studies demonstrate that transportation mode choice is multidimensional and that the appropriate explanatory variables depend on the characteristics of the study area and the transportation alternatives being examined.

Within the Indonesian context, recent research has applied logistic modelling to investigate transportation mode preferences in Tana Toraja. Such research demonstrates the applicability of probabilistic modelling to local transportation decisions and provides an important regional reference for understanding mode preference in the Toraja area [10]. However, the transportation context of North Toraja differs from that of more concentrated urban areas because mobility takes place across a combination of urban, rural, and semi-rural environments. Differences in settlement characteristics, household structure, accessibility, and private vehicle availability may consequently produce different patterns of transportation choice. Understanding these differences is important for developing transportation models that reflect the characteristics of local communities rather than relying exclusively on models developed in metropolitan or highly urbanized settings.

Despite the growing body of research on travel behavior and transportation mode choice, a specific research gap remains in modelling the choice between motorcycles and private cars based on household characteristics in semi-rural North Toraja. Previous studies have examined household travel behavior, socioeconomic determinants, intercity mode choice, and transportation preferences in various contexts, but their findings cannot necessarily be transferred directly to Sa'dan because transportation behavior is shaped by the characteristics of the local population, transportation system, and spatial environment [3], [8], [9]. In addition, existing research in the Toraja region has considered transportation mode preferences, but a focused probability model that examines the household characteristics associated with the choice between motorcycle and private car in Sa'dan District remains limited [10]. This gap indicates the need for an empirical model that specifically represents the binary mode choice structure and household characteristics of a semi-rural community in North Toraja.

The novelty of this study lies in developing a binary probability model to examine the choice between motorcycles and private cars based on household characteristics in Sa'dan District, North Toraja. Rather than examining transportation preference only through descriptive comparisons, this study estimates the probability of selecting a private car relative to a motorcycle using binary logistic regression. The model incorporates five household characteristics—vehicle ownership, household size, number of working household members, number of school-attending household members, and household income—as explanatory variables. This formulation provides a more specific representation of household mode choice in a semi-rural North Toraja context and enables the contribution of each household characteristic to be evaluated statistically [6].

A further contribution of the study is the use of predicted probability as the principal behavioural output of the model. In addition to identifying statistically significant explanatory variables, the model provides an estimated probability of choosing a private car for each respondent based on the reported household characteristics. This probability-based output allows the model to move beyond identifying associations toward quantifying the likelihood of observing a particular mode choice under different household characteristics. Such an approach is consistent with the probabilistic nature of discrete choice modelling, in which observed characteristics are used to estimate the probability of selecting available transportation alternatives [5]. The model therefore contributes empirical evidence that can be used to understand differences in private vehicle mode choice within a semi-rural community.

The study also contributes to the growing body of transportation research in the Toraja region by providing evidence at a more localized spatial scale. While broader regional statistics can describe the growth of private vehicles and demographic characteristics, the present study connects these regional conditions with respondent-level mode-choice observations and household characteristics from Sa'dan District. This combination provides a context-specific understanding of how household characteristics are associated with the choice between motorcycles and private cars. The findings can therefore complement previous research on household travel behavior and transportation in North Toraja by focusing specifically on mode choice probability rather than trip generation alone [4].

Accordingly, the objective of this study is to develop and evaluate a binary logistic regression model for estimating the probability of motorcycle and private car choice among households in Sa'dan District, North Toraja. Specifically, the study examines the relationship between vehicle ownership, household size, working household members, school-attending household members, and household income and the probability of choosing a private car over a motorcycle. The study further evaluates the statistical significance, model fit, classification performance, and discriminatory ability of the resulting probability model. The results are expected to provide empirical evidence on transportation mode choice based on household characteristics in a semi-rural area and contribute to the application of probability-based transportation modelling in the North Toraja context.

2. Method

This study employed a quantitative research design to examine the association between household characteristics and transportation mode choice in Sa'dan District, North Toraja Regency, South Sulawesi, Indonesia. The study focused on two private motorized transportation alternatives: motorcycles and private cars. A quantitative approach was adopted to examine the factors associated with transportation mode choice and estimate the probability of choosing a private car rather than a motorcycle. Binary logistic regression was selected because the dependent variable consisted of two mutually exclusive categories, allowing the probability of choosing one alternative to be estimated from a set of explanatory variables [6].

The study was conducted in Sa'dan District, North Toraja Regency, South Sulawesi, Indonesia, a semi-rural setting selected to examine transportation behavior in relation to household characteristics, vehicle ownership, and socioeconomic conditions. Primary data were collected through a questionnaire survey from April to May 2026. One questionnaire was collected from each household, with one respondent providing information about the household's characteristics. The unit of analysis was the individual respondent, and each observation represented the transportation mode most frequently used by that respondent, classified as either a motorcycle or a private car. A total of 162 valid observations were included in the final analysis, with no missing observations for the variables used in the logistic regression model. The questionnaire collected information on household characteristics, vehicle ownership, household income, and trip characteristics, including travel distance and travel time [11]. Although travel distance and travel time were collected, they were excluded from the final model because the study specifically focused on the association between household characteristics and transportation mode choice. Consequently, the model does not account for potentially relevant trip-level influences on mode choice.

Participation in the survey was voluntary, and respondents agreed to participate before completing the questionnaire. Household income and other respondent information were treated confidentially and used solely for research purposes. The findings are presented in aggregate form to protect respondents' privacy.

The dependent variable was transportation mode choice. In the original questionnaire, the transportation mode variable was coded as 1 = motorcycle and 2 = private car. For the binary logistic regression analysis, the original coding was transformed into a binary coding scheme in which motorcycle was coded as 0 and private car as 1. This transformation was performed for statistical modelling so that the estimated probability represents the probability of choosing a private car relative to a motorcycle. Five household characteristics were used as explanatory variables: vehicle ownership (X1), household size (X2), working household members (X3), school-attending household members (X4), and household income (X5). These variables represent household transportation resources, household composition, household activity characteristics, and socioeconomic conditions that may be associated with travel behavior and transportation mode choice [2], [3]. The operational definitions of the variables are presented in Table 1.

Table 1. Variables and Operational Definitions

Variable	Symbol	Variable Type	Operational Definition
Transportation mode choice	Y	Dependent, binary	Most frequently used transportation mode
Vehicle ownership	X1	Independent	Number of vehicles owned by the household
Household size	X2	Independent	Total number of household members
Working household members	X3	Independent	Number of household members who are working
School-attending household members	X4	Independent	Number of household members attending school
Household income	X5	Independent, ordinal	Monthly household income category

Vehicle ownership, household size, working household members, and school-attending household members were measured as actual numerical counts based on the conditions reported by each household. These variables were not assigned predetermined maximum values because they represent the actual number of vehicles or household members. Household income was treated as an ordinal variable because it was recorded in four ordered income categories. The selection of household characteristics as explanatory variables is consistent with previous research showing that household composition, socioeconomic conditions, and transportation resources can influence travel behavior and transportation decisions [3].

Descriptive statistics were used to summarize the characteristics of the respondents and the distribution of transportation mode choices. Binary logistic regression was subsequently estimated using the Enter method, in which all five explanatory variables were entered into the model simultaneously. The dependent variable was coded as 0 for motorcycle choice and 1 for private-car choice. The regression coefficients were estimated using maximum likelihood estimation. The resulting coefficients and odds ratios were used to examine the associations between the explanatory variables and the odds of choosing a private car rather than a motorcycle.

Binary logistic regression was applied to estimate the probability of choosing a private car relative to a motorcycle. Because private car was coded as 1 in the regression model, the dependent variable represents the probability of the event of choosing a private car. The logistic regression model was formulated as:

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

where P(Car) represents the probability of choosing a private car, β0 represents the constant, β1–β5 represent the regression coefficients, and X1–X5 represent vehicle ownership, household size, working household members, school-attending household members, and household income, respectively. The predicted probability of choosing a private car was calculated using:

$$P(\text{Car}) = \frac{1}{1+\exp[-(\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+\beta_4 X_4+\beta_5 X_5)]} \quad (2)$$

The probability of choosing a motorcycle was obtained as the complement of the predicted probability of choosing a private car. The regression coefficients were estimated using maximum likelihood estimation. A positive coefficient indicates an increase in the log-odds of choosing a private car as the corresponding predictor increases, whereas a negative coefficient indicates a decrease in the log-odds. The odds ratio, expressed as Exp(B), was used to interpret the change in the odds of choosing a private car associated with a one-unit increase in a numerical predictor or a one-category increase in household income, while the other predictors were held constant [6], [12], [13].

The performance of the logistic regression model was evaluated using several complementary measures. The Omnibus Test of Model Coefficients was used to determine whether the model containing the five explanatory variables significantly improved the prediction compared with the intercept-only model. Cox and Snell R^2 and Nagelkerke R^2 were used to describe the explanatory performance of the model, while the Hosmer–Lemeshow goodness-of-fit test was used to evaluate the agreement between observed and predicted outcomes [6], [12]. Classification performance was assessed by comparing the observed transportation mode with the predicted mode, and the overall proportion of correctly classified observations was used to describe classification accuracy. The discriminatory ability of the model was further evaluated using the Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC), which indicate the ability of the predicted probabilities to distinguish between private-car and motorcycle choices across different classification thresholds [14], [15], [16].

The predicted probability generated by the logistic regression model was used as the principal probabilistic output of the study. For each observation, the model generated an estimated probability of choosing a private car, while the corresponding probability of choosing a motorcycle was obtained from the complement of that probability. The predicted probabilities were used to describe differences in mode-choice likelihood among households with different characteristics. This probability-based approach is consistent with the discrete choice modelling framework, in which observed characteristics are used to estimate the likelihood of selecting a particular transportation alternative [5].

3. Result and Discussion

3.1 Characteristics of Respondents and Mode Choice

A total of 162 valid observations were analyzed in this study. Regarding vehicle ownership, households owning two vehicles represented the largest proportion (38.3%), followed by three vehicles (31.5%), one vehicle (15.4%), four vehicles (11.7%), and five vehicles (3.1%). Household size varied from one to eleven members, with households consisting of four members representing the largest proportion (31.5%), followed by five members (26.5%) and three members (24.7%). In terms of working household members, most households had one working member (66.7%), followed by two working members (28.4%), while households with three and seven working members accounted for 4.3% and 0.6%, respectively.

For school-attending household members, households with one school-attending member constituted the largest proportion (48.8%), followed by two members (38.3%), three members (8.6%), four members (3.1%), five members (0.6%), and six members (0.6%). Household income was concentrated in the Rp 1,500,000–Rp 3,000,000 category (53.7%), followed by Rp 3,000,000–Rp 5,000,000 (38.3%), above Rp 5,000,000 (5.6%), and below Rp 1,500,000 (2.5%). Regarding the most frequently used transportation mode, motorcycles were the dominant mode, with 114 respondents

(70.4%), while 48 respondents (29.6%) reported private cars as their most frequently used mode. Figure 1 illustrates the distribution of household characteristics and the most frequently used transportation modes among the 162 respondents.

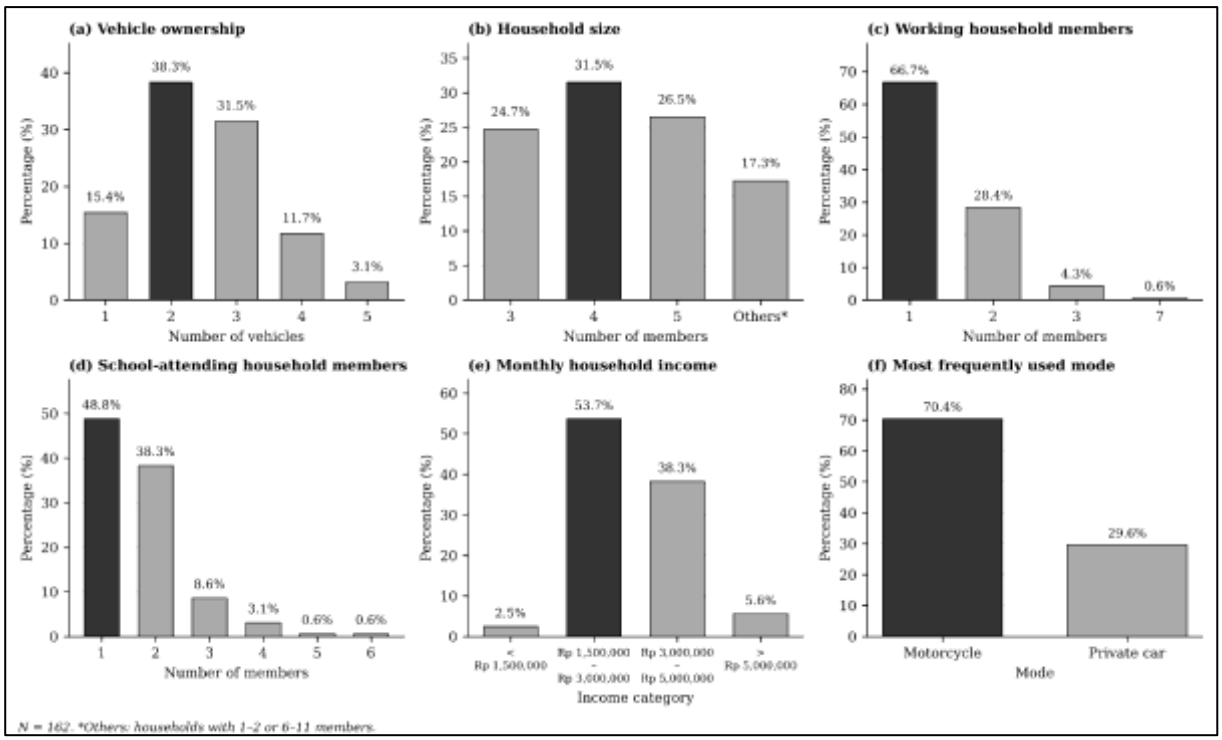


Figure 1. Household Characteristics and Mode Choice (N = 162)

3.2 Binary Logistic Regression Results

A binary logistic regression analysis was conducted to examine the relationship between household characteristics and transportation mode choice. In the original dataset, transportation mode was coded as 1 = motorcycle and 2 = private car. For the logistic regression analysis, SPSS internally recoded the dependent variable as 0 = motorcycle and 1 = private car. A total of 162 observations were included in the analysis, with no missing cases. The predictors entered into the model were vehicle ownership (X1), household size (X2), working household members (X3), school-attending household members (X4), and household income (X5).

Table 2 presents the binary logistic regression results, including the estimated coefficients, standard errors, Wald statistics, significance levels, and odds ratios for the five explanatory variables.

Table 2. Binary Logistic Regression Results

Predictor	B	S.E.	Wald	df	Sig.	Exp(B)
Vehicle ownership (X1)	0.265	0.261	1.030	1.000	0.310	1.304
Household size (X2)	-0.036	0.262	0.019	1.000	0.890	0.964
Working household members (X3)	-0.465	0.367	1.607	1.000	0.205	0.628
School-attending household members (X4)	-0.255	0.386	0.438	1.000	0.508	0.775
Household income (X5)	2.303	0.445	26.838	1.000	<0.001	10.003
Constant	-6.309	1.193	27.949	1.000	<0.001	0.002

Note: Original mode coding: 1 = motorcycle and 2 = private car. For logistic regression, SPSS internally recoded the dependent variable as 0 = motorcycle and 1 = private car. B = regression coefficient; S.E. = standard error; Exp(B) = odds ratio.

The estimated logistic regression model is expressed as:

$$\ln\left(\frac{P(\text{Car})}{1-P(\text{Car})}\right) = -6.309 + 0.265X_1 - 0.036X_2 - 0.465X_3 - 0.255X_4 + 2.303X_5$$

The results show that household income (X5) was the only statistically significant predictor of transportation mode choice (B = 2.303, Wald = 26.838, $p < 0.001$). The odds ratio of 10.003 indicates that a one-category increase in the household income variable was associated with approximately 10 times higher odds of choosing a private car rather than a motorcycle, holding the other predictors constant.

Vehicle ownership (X1) was not statistically significant ($p = 0.310$), nor were household size (X2) ($p = 0.890$), working household members (X3) ($p = 0.205$), and school-attending household members (X4) ($p = 0.508$). Thus, among the five household characteristics included in the model, only household income showed a statistically significant association with transportation mode choice.

Household income was the only significant predictor of transportation mode choice (B = 2.303, $p < 0.001$; Exp(B) = 10.003). Higher income may increase a household's financial capacity to purchase and maintain a private car, making it a more feasible alternative to a motorcycle. This finding is consistent with Aderibigbe and Gumbo [3], who identified household-head income as a significant predictor of trip frequency in rural Nigeria. However, Srinivasan and Ferreira [2] emphasized that travel behaviour is also influenced by residential location and the spatial context of household activities. Therefore, income appears important in Sa'dan, although its effect should be interpreted as an association rather than a causal relationship.

3.3 Model Fit and Classification Performance

The overall significance of the binary logistic regression model was evaluated using the Omnibus Test of Model Coefficients. The model produced a chi-square value of 51.014 with 5 degrees of freedom and a significance level of $p < 0.001$. This result indicates that the model containing vehicle ownership, household size, working household members, school-attending household members, and household income significantly improved the prediction of transportation mode choice compared with the intercept-only model.

The model summary showed a −2 Log Likelihood value of 145.879, with a Cox and Snell R^2 of 0.270 and a Nagelkerke R^2 of 0.384. These values indicate the explanatory performance of the logistic regression model in representing the observed variation in transportation mode choice.

The Hosmer–Lemeshow test produced a chi-square value of 6.074 with 8 degrees of freedom and a significance value of 0.639. Since the significance value was greater than 0.05, the test did not indicate a statistically significant lack of fit between the observed outcomes and the outcomes estimated by the model.

The classification results showed an overall apparent classification accuracy of 78.4% on the full dataset. The model correctly classified 95 of 114 motorcycle observations (83.3%) and 32 of 48 private-car observations (66.7%). These results indicate that the model classified motorcycle choice more accurately than private-car choice on the dataset used for model estimation.

Table 3 presents the classification accuracy obtained for each fold in the five-fold cross-validation procedure.

Table 3. Five-Fold Cross-Validation Classification Accuracy

Fold	Test samples	Correct predictions	Accuracy (%)
1	33	24	72.7
2	33	26	78.8
3	32	26	81.3
4	32	21	65.6
5	32	23	71.9
Total	162	120	74.1

The five-fold cross-validation results yielded an overall classification accuracy of 74.1%, with 120 of 162 observations correctly classified across the five test folds. This accuracy was lower than the apparent classification accuracy of 78.4% obtained from the full dataset. The difference suggests that classification performance may be lower when evaluated on observations excluded from model fitting. This result should be interpreted as an internal validation estimate only if predictions for each fold were generated using a model trained exclusively on the other four folds.

The discriminatory performance of the fitted model was further evaluated using the Receiver Operating Characteristic (ROC) curve. The analysis yielded an Area Under the Curve (AUC) of 0.830, indicating good discrimination between motorcycle and private-car choices on the data used for model estimation. This AUC should be interpreted as apparent discrimination unless it was calculated from independent test data or out-of-fold predictions. The ROC curve is presented in Figure 2.

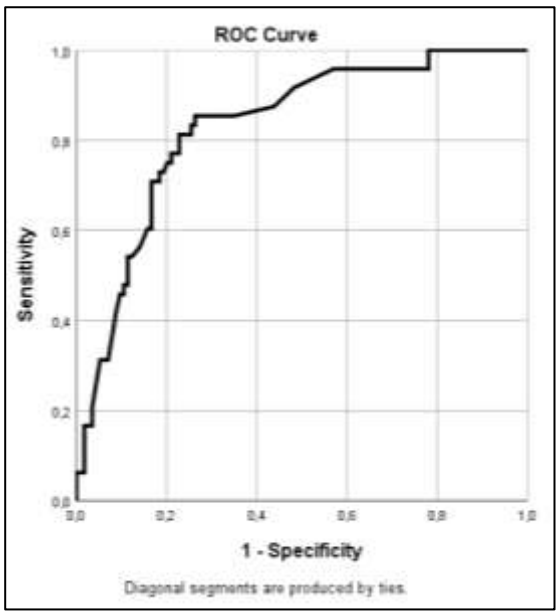


Figure 2. ROC Curve for Mode Choice

The ROC curve illustrates the trade-off between sensitivity and specificity across different classification thresholds. Sensitivity measures the proportion of actual private-car users correctly classified by the model, whereas specificity measures the proportion of actual motorcycle users correctly classified. Both measures depend on the classification threshold and the model's ability to distinguish between the two modes. Increasing sensitivity may reduce specificity, and vice versa. Therefore, the threshold should be selected according to the study objectives and the consequences of misclassification.

3.4 Predicted Probability of Mode Choice

The descriptive statistics of the predicted probability of private-car choice are presented in Table 4.

Table 4. Predicted Probability of Private-Car Choice

Statistic	Predicted Probability
N	162
Minimum	0.00281
Maximum	0.93522
Mean	0.29630
Std. Deviation	0.25121

The binary logistic regression model generated predicted probabilities of private-car choice for all 162 observations. The predicted probability ranged from 0.00281 to 0.93522, with a mean of 0.29630 and a standard deviation of 0.25121. This indicates that the average predicted probability of choosing a private car was approximately 29.63% across the observed households.

The relatively wide range of predicted probabilities indicates variation in the estimated propensity for private-car choice among households. A higher predicted probability represents a greater estimated likelihood of private-car choice, whereas a lower probability represents a greater estimated likelihood of motorcycle choice.

3.5 Discussion, Interpretation, and Limitations

Household income was the only statistically significant predictor of transportation mode choice in Sa'dan District ($B = 2.303$, $p < 0.001$, $Exp(B) = 10.003$). A one-category increase in household income was associated with approximately ten times higher odds of choosing a private car rather than a motorcycle, holding the other variables constant. This result suggests that economic capacity may be an important consideration in household transportation decisions. However, this association should not be interpreted as evidence of causation.

The relationship between income and private-car choice may reflect differences in households' ability to purchase, maintain, and operate private cars. In areas where motorcycles are widely used, private cars may represent an alternative that requires greater financial resources. Previous research has shown that socioeconomic characteristics are associated with travel behavior, although their influence may vary according to local transportation conditions and household circumstances [2], [3].

Vehicle ownership, household size, the number of working household members, and the number of school-attending household members were not statistically significant predictors in the present model. These findings do not necessarily mean that the variables are irrelevant to transportation decisions. Their effects may depend on other factors, including trip purpose, travel distance, travel time, vehicle availability, and the availability of public transportation. Household-level travel behavior is influenced by relationships between household characteristics and residential or transportation conditions, which may not be fully represented by the five predictors included in this study [2].

The findings should also be considered in relation to previous transportation studies. Research on socioeconomic characteristics and travel behavior has demonstrated that mode choice is influenced by the context in which travel takes place, meaning that results from other regions cannot automatically be generalized to Sa'dan District [3]. Research on household travel behavior also emphasizes the relationship between household circumstances and transportation decisions [2]. In the Toraja region, previous studies have examined transportation behavior and mode preferences, providing a regional basis for further investigation of household-level transportation choices [4], [10]. The present study

extends this line of research by estimating the probability of choosing a private car rather than a motorcycle among households in Sa'dan District.

Within the survey sample, motorcycles accounted for 70.4% of observed mode choices, while private cars accounted for 29.6%. This distribution indicates that motorcycles were the dominant transportation mode among the surveyed households. The predominance of motorcycles provides an important context for interpreting the regression results, particularly the association between household income and private-car choice. However, the observed distribution describes the study sample and should not automatically be generalized to all households in North Toraja.

The logistic regression model demonstrated statistically significant overall performance (χ^2 = 51.014, $p < 0.001$), with a Nagelkerke R^2 of 0.384. The Hosmer–Lemeshow test was not statistically significant ($p = 0.639$), indicating no evidence of poor fit based on this test. The model correctly classified 78.4% of observations, while the ROC analysis produced an AUC of 0.830. These results indicate that the model can distinguish between motorcycle and private-car choices reasonably well within the observed sample. Nevertheless, classification accuracy and AUC calculated from the same data used to estimate the model may provide an optimistic assessment of performance on new observations. Internal validation is therefore important when evaluating the model's generalizability [6], [14].

The mean predicted probability of private-car choice was 29.63%, which was close to the observed proportion of 29.6%. Although this similarity is encouraging, it does not by itself establish that the model is well calibrated across individual households or probability levels. Further assessment using calibration measures and validated predictions would provide a more complete evaluation of predictive performance [6].

Several limitations should be acknowledged. First, the analysis was based on 162 valid observations from one district, so the findings may not represent all households in North Toraja or other regions. Second, although trip distance and travel time were collected, they were not included in the final model. Their exclusion means that the analysis primarily represents household characteristics rather than the full range of factors affecting transportation mode choice. Third, the model considered only motorcycles and private cars and did not explicitly model public transportation or other available alternatives. Future research should include a broader range of household and trip-level variables, consider additional transportation alternatives, and validate the model using independent data or appropriate internal validation procedures.

Overall, the findings show that household income was the only statistically significant predictor among the five variables examined in Sa'dan District. Binary logistic regression provides a probability-based approach for examining motorcycle and private-car choice, but the model should be interpreted within the limits of the sample, selected variables, and validation procedures. Further research is needed to assess its predictive performance and applicability in other transportation contexts.

4. Conclusion

This study developed a binary logistic regression model to examine transportation mode choice between motorcycles and private cars in Sa'dan District, North Toraja, using 162 valid observations. Motorcycles accounted for 70.4% of the observed choices, while private cars accounted for 29.6%. Among the five household characteristics examined, household income was the only statistically significant predictor of private-car choice ($B = 2.303$, $p < 0.001$, $Exp(B) = 10.003$). A one-category increase in household income was associated with approximately ten times higher odds of choosing a private car rather than a motorcycle, holding the other predictors constant. This interpretation assumes a linear relationship between the ordered income categories and the log-odds of private-car choice.

The overall model was statistically significant ($\chi^2 = 51.014$, $p < 0.001$), with a Nagelkerke R^2 of 0.384 and a non-significant Hosmer–Lemeshow test ($p = 0.639$). The apparent classification accuracy was 78.4%, while the reported five-fold cross-validation accuracy was 74.1%; the ROC analysis yielded an AUC of 0.830. These findings should be interpreted in light of the sample size, the focus on a single district, and the exclusion of trip-level variables such as travel distance and travel time. The results indicate associations rather than causal relationships. Future research should include additional household and trip characteristics, examine the treatment of income categories, and evaluate predictive performance using appropriate validation procedures.

Acknowledgment

The authors would like to thank the Government of Sa'dan District, North Toraja Regency, and all respondents who participated in the household survey and provided the information required for this research. The authors also acknowledge Universitas Kristen Indonesia Toraja for its institutional affiliation.

References

[1] R. T. Bua' and A. T. Sombolinggi, *Permodelan Transportasi untuk Ibu Kota Kabupaten*. Makassar: Arsy Media, 2026.

[2] S. Srinivasan and J. Ferreira, 'Travel behavior at the household level: understanding linkages with residential choice', *Transp. Res. Part D Transp. Environ.*, vol. 7, no. 3, pp. 225–242, 2002, doi: 10.1016/S1361-9209(01)00021-9.

[3] O.-O. Aderibigbe and T. Gumbo, 'Influence of Socio-economic Attributes on Travel Behaviour in the Rural Areas of Nigeria: Towards a Sustainable Rural Planning and Development', *Urban, Plan. Transp. Res.*, vol. 10, no. 1, pp. 181–199, Dec. 2022, doi: 10.1080/21650020.2022.2072946.

[4] Y. Pakiding, A. T. Sombolinggi, R. T. Bua', and A. T. Arrang, 'The Role Of Household Characteristics And Regional Transportation In Trip Generation In Semi-Rural Areas Of North Toraja', *J. Penelit. Sekol. Tinggi Transp. Darat*, vol. 17, no. 1, pp. 100–112, Jul. 2026, doi: 10.55511/jpsttd.v16i2.794.

[5] J. de D. Ortúzar and L. G. Willumsen, *Modelling Transport*, 4th ed. Chichester: John Wiley & Sons, 2011. doi: 10.1002/9781119993308.

[6] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed. Hoboken, New Jersey: John Wiley & Sons, 2013.

[7] A. F. Mohd., S. F. Mohd., M. P. P. Abdul, R. A. Mohd., Z. Aizzat, and N. F. Ab., 'Travel Mode Choice Modeling: Predictive Efficacy between Machine Learning Models and Discrete Choice Model', Dec. 20, 2021, *Bentham Science*. doi: 10.2174/1874447802115010241.

[8] J. Jeong, J. Lee, and T. T. Gim, 'Travel mode choice as a representation of travel utility: A multilevel approach reflecting the hierarchical structure of trip, individual, and neighborhood characteristics', *Pap. Reg. Sci.*, vol. 101, no. 3, pp. 745–766, 2022, doi: https://doi.org/10.1111/pirs.12665.

[9] Q. Deng and I. Henderson, 'Travel Mode Choice for Domestic Intercity Travel: A Case Study in Suzhou, China', *ASEAN J. Hosp. Tour.*, vol. 20, no. 2 SE-, pp. 1–26, Aug. 2022, doi: 10.5614/ajht.2022.20.2.01.

[10] A. T. Sombolinggi and R. T. Bua', 'Pemodelan Preferensi Pemilihan Moda Transportasi Menggunakan Regresi Logistik Nominal di Makale, Kabupaten Tana Toraja', *J. Komposit J. Ilmu-ilmu Tek. Sipil*, vol. 10, no. 2 SE-Articles, pp. 297–308, Aug. 2026, doi: 10.32832/komposit.v10i2.23044.

[11] S. B. Green, 'How Many Subjects Does It Take To Do A Regression Analysis', *Multivariate Behav. Res.*, vol. 26, no. 3, pp. 499–510, 1991, doi: 10.1207/s15327906mbr2603_7.

[12] A. T. Sombolinggi, R. T. Bua', A. T. Arrang, and Y. Pakiding, 'Integrated Traffic Safety Assessment Using Accident Equivalent Numbers And Road User Perceptions In Tana Toraja', *J. Penelit. Sekol. Tinggi Transp. Darat*, vol. 17, no. 1, pp. 10–28, Jun. 2026, doi:

10.55511/jpsttd.v17i1.784.
[13] R. D. Riley *et al.*, 'Calculating the sample size required for developing a clinical prediction model', *BMJ*, vol. 368, 2020, doi: 10.1136/bmj.m441.
[14] T. Fawcett, 'An introduction to ROC analysis', *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006, doi: 10.1016/j.patrec.2005.10.010.
[15] F. S. Nahm, 'Receiver operating characteristic curve: overview and practical use for clinicians', *Korean J Anesth.*, vol. 75, no. 1, pp. 25–36, Feb. 2022, doi: 10.4097/kja.21209.
[16] D. Dey *et al.*, 'The proper application of logistic regression model in complex survey data: a systematic review', *BMC Med. Res. Methodol.*, vol. 25, no. 1, p. 15, 2025, doi: 10.1186/s12874-024-02454-5.