IMM

Journal of New Frontiers in Mathematics and Statistics

ISSN 2719- 2415 (Online)

Volume 1 (1) 2020, 16 - 25

Home page: www.imathm.edu.lk/publications

Identifying Factors Associated with Price Categories of Motorcycles in Sri Lanka using Discriminant Analysis

P.A.L. Chanika[1]*, A.M.C.H. Attanayake [1], M.D.N. Gunaratne [2]

Department of Statistics & Computer Science, University of Kelaniya, Sri Lanka[1]
Trade & Productivity Commission Secretariat, Ministry of Industrial Export and Investment Promotion, Sri Lanka[2]

laknichanika1213@gmail.com*

ABSTRACT

The motorcycle is one of the most popular modes of transportation in Asia because it is a fast, economic and convenient mode of transportation. According to the motor vehicle registration statistics of the department of motor traffic, motorcycle is the best-selling vehicle in Sri Lanka. Pricing of the motorcycles is one of the major concerns for motorcycle byers and importers. The purposes of this study are to identify the factors which are associated with the prices of motorcycles and to construct a model to determine the price of a motorcycle based on significant factors. Specifications and features of selected motorcycles were collected from secondary data sources. 'Motorcycle Category', 'Fuel Efficiency', 'Engine Power', 'Brake System', 'Torque', 'Length' were the significant factors which had an association between the price categories of motorcycles among 15 properties of motorcycles considered under study. The canonical discriminant analysis was used to develop the model and to predict the price categories of motorcycles. The significant categorical variables were used as the dummy variables. Torque was omitted from the model because of the presence of multicollinearity. The two significant discriminant functions were classified the data into four categories of price. The 75% responds were correctly classified into four price categories. Motorcycle category, Fuel Efficiency, Engine Power, Brake System and Length were found as significantly associated factors to the prices of motorcycles. Decision makers in the field can make good use of those factors in developing their pricing strategies in the motorcycle industry.

Keywords: Canonical Discriminant Analysis, Associated Factors, Pricing of Motorcycles

1. INTRODUCTION

1.1 Background of the Study

One of the most common modes of transportation in world is the motorcycle, also called a motorbike. Further, the motorcycle uses for the racing, goods transport and military activities. In addition, most people use the motorcycle for the day to day transportation. Sri Lanka is the one of countries in the Asian region, where a lot of people use motorcycles for their day to day transport. Most people do not like to travel by public transport because of many reasons such as public transport takes more time, it is crowded, hard to go the terminal, and it rarely services in remote areas etc. Especially most of people often face the dense traffic and bad roads then the resulting impact can be minimized by using motorcycles. Motorcycle gives mobility, overall economy in transportation, productivity and reducing time losses.

According to statistics of Department of Motor Traffic, Sri Lanka motorcycle population of the country has increased to 4.3 million by 2018 and, 338 433 number of motorcycles were registered in 2018. There are different motorcycle brands in Sri Lankan motorcycle market. Currently Honda, Yamaha, Bajaj, Hero Honda, KTM, Suzuki, and Demak are the highest demanded motorcycle brands in Sri Lanka. The manufacturing country of Honda, Yamaha, Bajaj, Hero Honda, KTM, Suzuki and Yamaha motorcycles is India and Demak is a Malaysian motorcycle brand.

Prices of motorcycles may vary from one country to another. This is because of local taxes and due to transport and handling charges etc. In Sri Lanka motorcycle prices may vary according to the production country, exchange rate, government import taxes, dealers profit margin and also, extra accessories, etc. Specially motorcycles prices may vary according to the accessories of motorcycle namely motorcycle category (standard, commuter deluxe, sports etc.), appearance of motorcycle, brake type, presence of alloy wheels, start type (electric/ kick start), meter panel type (digital/ analog), engine power, torque, frame, suspension, light system, fuel efficiency, brand, resale value, and fuel tank capacity, etc.

1.2 Problem Statement

According to Motor vehicle registration statistics of the department of motor traffic, motorcycles is the best-selling vehicle category in Sri Lanka. There are many players in the Sri Lankan Motorcycle industry and Pricing strategy is a very important feature in the marketing mix. Pricing is a very complex decision and statistical techniques can be applied for this decision-making process. This is very important especially for in the process of new model launch. Therefore, the predicting appropriate price range of motorcycles as per product attributes with statistical techniques would be more reliable and technical approach instead of informal customary applications.

1.3 Objectives

The main objectives of this study are,

- To identify the statistically significant factors associated for pricing of Motorcycles.

ISSN 2719-2415 (Online)

- To develop a methodology to predict the price category of a newly introduced Motorcycle under existing scenario.

1.4 Significance of the Study

Motorcycle prices is very important for both importers and buyers. This research finds out factors associated with the pricing in different motorcycle models and develop a model will be able to determine the price range of the motorcycle accordingly. Determination of relevant price range would be very useful for preparation of a solid technical basis for pricing strategy, initial discussions with principals and developing marketing plans.

2. LITERATURE REVIEW

The literature review is based on the Automobile Pricing and application of the Discriminant Analysis.Iseri & Karlik, (2009) have examined the automobile pricing using artificial neural network. The main objective of this study is to find an automobile pricing model using artificial neural networks (ANN). The data set consisted of 159 samples each of 15 input values and one output value. Overall, 60% of this data was used for training, 20% was used for cross-validation, and 20% was used for analysis. The data are selected randomly from the data set used for each class. The input entities are wheelbase, length, width, height, average loss payment per insured vehicle year, total weight, Engine size, stroke, diameter of a cylinder in a piston engine, compression ratio, horse power, city mpg, highway mpg, peak rpm. The price of automobile is the target entity. Mean absolute error of outputs is 8%.

Peerun, et al., (2015) have predicted the the price of second-hand cars using artifical neural networks.They have used the 200 cars as the sample of their study.According to this study price prediction of second-hand cars mainly depend on the factors namely manufactering year,make,model,mileage,horsepower and country of origin and etc.They have analysed the data using four different machine learning technique with six explanotory variables namely manufactering year,make,engine capacity,paint type,transmission type and mileage. For all the four methods, the mean residual value was significantly lower.

Venkatasubbu & Ganesh, (2019) have conducted the study on used cars price prediction using supervised learning techniques. They have proposed three models using lasso regression, multiple regression and regression tree techniques. The data was gathered from the 2005 Central Edition of the Kelly Blue Book and 804 records of cars were used for the analysis. Their main objective was to use these three techniques to predict the price of used car. Eleven explanatory such as mileage, make, model, type of car model, cylinder, body type, fuel capacity, number of doors, cruise, sound, and leather interior were applied as explanatory variables. Considering mean error of models, mean error of regression tree model is greater than the mean error of multiple regression and lasso regression models.

ISSN 2719-2415 (Online)

Furthermore, Noor & Jan, (2017) proposed the multiple linear regression technique for vehicle price prediction sysytem. This study presents a methodology where price is a dependent variable, and the independent variables such as vehicle's model, make, city, version,color, mileage, alloy rims and power steering. Resrachers have applied the variable selection method on the all input variables and select the most significant explanotory variables namely model year,model and engine type.The proposed multiple linear regression model accuracy is 98%.

Yamada, et al., (2001) have analyzed the road condition towards understanding of vehicle driving environment by using discriminant analysis. They focused on features related to water and snow on the road, and we extracted these features by image analysis for discriminating the road condition. They used a multivariate algorithm to differentiate between five types of road conditions: "Dry," "Wet," "Slushy," "Icy" and "Snowy," based on these characteristics as well as temperature derived from the road images. The researchers performed field tests to verify this algorithm's accuracy and obtained an overall favorable discrimination accuracy 92.3%.

Pardo, et al., (2006) have examined the Classification of adolescent psychotic disorders using linear discriminant analysis. They were used linear discriminant analysis to classify 28 adolescents in to three categories. The categories are healthy, schizophrenia, biopar and the examiners were used the 12 predictor variables for the study. According to the conclusion of the study Linear Discriminant Analysis (LDA) with 12 explanatory variables could provide a correct classification and this robust classification relies upon both neuropsychological and brain structural information. The results of this study were presented that despite overlapping clinical symptoms, schizophrenia and bipolar disorder can be differentiated early in the course of disease. This study has been presented the application of linear discriminant analysis by using the dependent variable which consists 3 categories.

This study is mainly focused on identifying the factors which affect to price of motorcycles and find a model to predict the prices of selected motorcycles. The artificial neural network model has used in price determination of automobiles. Many researchers widely used the technique of Discriminant Analysis for numerical predictor variables. In this study categorical variables are used as the predictor variables of the Discriminant Analysis.

3. METHODOLOGY

Mainly, secondary data were used for the study. Specifications and features of selected Main motorcycle models which accounted around 85% in terms of volumes of motorcycles in Sri Lanka. Attributes of motorcycle models were obtained through the internet promotional materials (leaflets) of different players in the automotive industry. In addition to that some qualitative information obtained from selected subject expertise of the automotive industry. All the attributes were chosen under the supervision of selected subject expertise of Sri Lankan automotive industry. Motorcycle Category, Length, Width, Height, Brake System, Presence of Alloy-wheel, Start, Engine Power, Torque,

ISSN 2719-2415 (Online)

Frame, Fuel Efficiency, Brand, Country of Origin, Resale Value, Fuel tank Capacity attributes were used as independent variables and the dependent variable consists of 4 groups (LKR 200001-300000 (Category 1), LKR 300001-400000 (Category 2), LKR 400001-500000 (Category 3) and Above LKR 500000 (Category 4)). The independent variables consisted of numeric and categorical variables. The dependent variable, motorcycle price was a categorical variable.

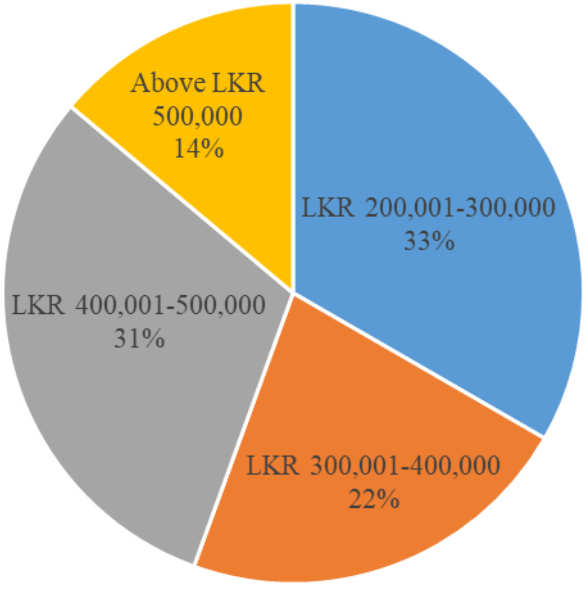
The association between the price category and each explanatory variable should be examined and significant variables should be identified using statistical techniques. The Chi square test was used to assess whether there is a significant relationship between two nominal (categorical) variables. (Agresti, 2007) Upon violation of the rule of chi square, the Fisher's exact test was used to verify the relationship between these two variables.

The canonical discriminant analysis was used to develop the model and to predict the price categories of motorcycles (Alayande & Adekunle, 2015). The significant categorical variables were used as the dummy variables. Variance Inflation Factor (VIF) was used to check the multicollinearity of the model.

4. RESULTS

The results of the analysis are given in following tables and figures. Microsoft Excel (version 2016) and IBM SPSS Statistical analytical software (version 20) were used to analyze and present the results.

Figure 1. Proportions of price categories of motorcycles in the sample



According to the Figure 1, sample consists 4 price categories. This figure indicates that the motorcycles which are priced LKR 200,001-300,000 have the highest proportion (33.33%) of sample. The minimum proportion of motorcycles (13.89%) are belongs to

ISSN 2719-2415 (Online)

the Above LKR 500,000 of the study.

Fisher's Exact test and Kruskal Wallis test were used to assess the association between the dependent variable (price category) and each explanatory variable. Due to the results of the tests there were six significant variables in this study namely motorcycle category, fuel efficiency, engine power, brake system, torque and length.

All the six number of significant variables were used to construct a discriminant analysis model. The Canonical Linear Discriminant method was used to analyze the variables. The categorical predictor variables namely motorcycle category and brake system used as the dummy variables. Variance Inflation Factor (VIF) was used to test the multicollinearity of the explanatory variables. Considering multicollinearity torque was omitted from the model (VIF>10). Motorcycle category, fuel efficiency, engine power, brake system and length variables were used to construct the model.

Table 1. Tests of Equality of group means for the model

	Wilks' Lambda	F	df1	df2	Sig.
Length	0.692	4.740	3	32	0.008
Fuel Efficiency	0.219	38.028	3	32	0.000
Engine Power	0.164	54.387	3	32	0.000
MotorcycleCategory_CommuterStandard	0.545	8.889	3	32	0.000
MotorcycleCategory_CommuterDeluxe	0.720	4.139	3	32	0.014
MotorcycleCategory_SportClassic	0.728	3.977	3	32	0.016
BrakeSystem_Dualdrum	0.126	74.074	3	32	0.000
BrakeSystem_Frontdisc	0.530	9.468	3	32	0.000
BrakeSystem_Dualdisc	0.526	9.606	3	32	0.000

Table 1 illustrates that the test of equality of the means for each independent variable. The all variable's significance value of is less than 0.05 implies that the null hypothesis of the variable is not significant can be rejected for all predictor variables in the model, according to the Wilks' Lambda, the engine power and fuel efficiency variable is better at discriminating between groups, followed by length, motorcycle category and brake system.

Table 2. Wilk's Lambda for the model

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 3	0.011	130.339	24	0.000
2 through 3	0.195	47.409	14	0.000
3	0.722	9.465	6	0.149

ISSN 2719-2415 (Online)

Table 2 indicates that the Wilks' Lambda ratio for each function. Three pairs of canonical variates, the number of the pairs needed to describe the relationship between groups of variables was decided based on the results of above table.

According to results in above table it was concluded that first and second pairs of canonical covariates are significant (P value<0.05) and the third pair was not significant. Thus, it can be concluded that of the three canonical dimensions, only the first two was statistically significant between sets of groups.

Table 3. Canonical Discriminant Function Coefficients of the model

	Function		
	1	2	3
Length	-0.002	-0.005	-0.018
Fuel Efficiency	0.010	-0.016	0.061
Engine Power	-0.117	0.545	0.303
MotorcycleCategory_CommuterStandard	3.930	1.798	-0.410
MotorcycleCategory_CommuterDeluxe	4.093	1.441	0.110
MotorcycleCategory_SportClassic	0.256	-0.926	-0.326
BrakeSystem_Dualdrum	3.832	2.237	0.016
BrakeSystem_Frontdisc	0.378	-0.061	2.412
(Constant)	2.816	5.196	29.306

According to the Table 3 the predictive equations was constructed using the canonical discriminant function coefficients. The predictive equations are,

$$\begin{aligned}\text{Score } 1 = {} & 2.816 + 3.930\,(\textit{Motorcycle Category} - \textit{CommuterStandard}) \\ & + 4.093\,(\textit{Motorcycle Category} - \textit{CommuterDeluxe}) \\ & + 0.256\,(\textit{Motorcycle Category} - \textit{SportsClassic}) - 0.002\,(\textit{Length}) \\ & - 0.117\,(\textit{Engine Power}) + 3.832\,(\textit{Brake System} - \textit{Dualdrum}) \\ & + 0.378\,(\textit{Brake System} - \textit{Frontdisc}) + 0.010\,(\textit{Fuel Efficiency})\end{aligned}$$

$$\begin{aligned}\text{Score } 2 = {} & 5.196 + 1.798\,(\textit{Motorcycle Category} - \textit{CommuterStandard}) \\ & + 1.441\,(\textit{Motorcycle Category} - \textit{CommuterDeluxe}) \\ & - 0.926\,(\textit{Motorcycle Category} - \textit{SportsClassic}) - 0.005\,(\textit{Length}) \\ & + 0.545\,(\textit{Engine Power}) + 2.237\,(\textit{Brake System} - \textit{Dualdrum}) \\ & - 0.061\,(\textit{Brake System} - \textit{Frontdisc}) - 0.016\,(\textit{Fuel Efficiency})\end{aligned}$$

$$\begin{aligned}\text{Score } 3 = {} & 29.306 - 0.410\,(\textit{Motorcycle Category} - \textit{CommuterStandard}) \\ & + 0.110\,(\textit{Motorcycle Category} - \textit{CommuterDeluxe}) \\ & - 0.326\,(\textit{Motorcycle Category} - \textit{SportsClassic}) - 0.018\,(\textit{Length}) \\ & + 0.303\,(\textit{Engine Power}) + 0.016\,(\textit{Brake System} - \textit{Dualdrum}) \\ & + 2.412\,(\textit{Brake System} - \textit{Frontdisc}) + 0.061\,(\textit{Fuel Efficiency})\end{aligned}$$

ISSN 2719-2415 (Online)

Table 4. Functions at group Centroids of the model

Price Category of Motorcycle	Function		
	1	2	3
1	5.687	0.350	-0.060
2	-1.513	-1.845	0.813
3	-2.707	-0.475	-0.694
4	3.592	3.321	0.498

The Table 4 shows that the mean discriminant scores for each of the Price categories for each of the discriminant functions. According to the function 1, the mean scores for each price category are 5.687, -1.513, -2.707 and -3.592. According to the function 2, the mean scores for each price category are 0.350, -1.845, -0.475 and 3.321. According to the function 3, the mean scores for each price category are -0.060,0.813, -0.694 and 0.498.

Figure 2. Graphical illustration of classification results

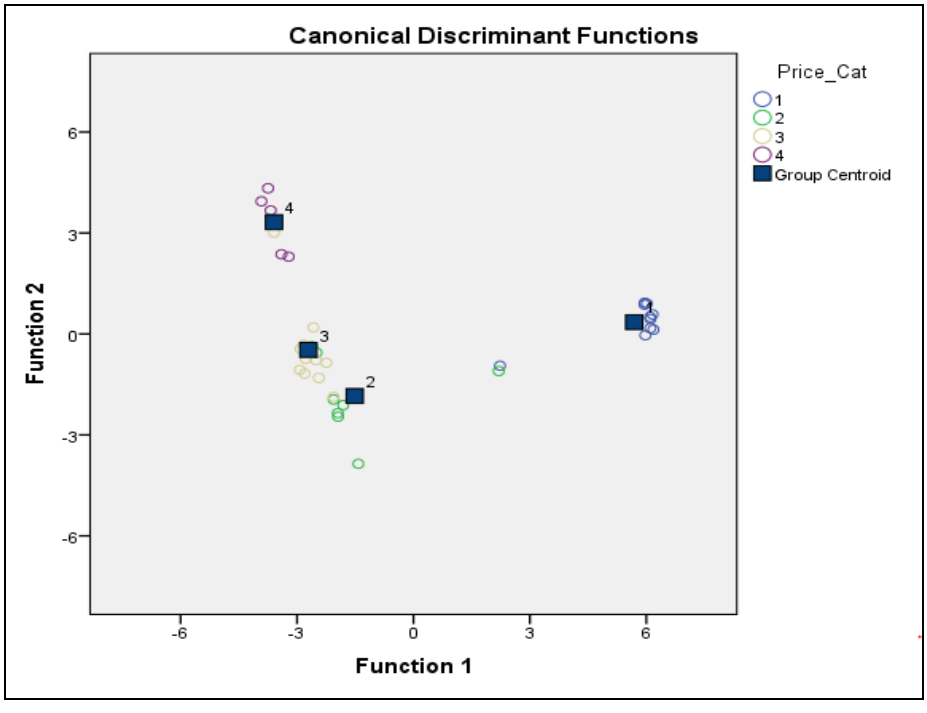


Figure 2 shows the spread of the four price categories between canonical function 2 and canonical function 1.

ISSN 2719-2415 (Online)

Table 5. Classification results for the model

	Price Category	Predicted Group Membership				Total
		1	2	3	4	
Original	1	10	1	0	0	11
	2	0	6	2	0	8
	3	0	2	9	1	12
	4	0	0	0	5	5
Cross Validated	1	10	1	0	0	11
	2	1	5	2	0	8
	3	0	4	7	1	12
	4	0	0	0	5	5

The classification results (Table 5) reveal that 83.3% of motorcycles were classified correctly into four price categories. Price category 4 was classified with better accuracy (100%) than price category 1 (90.9%), price category 3 (75%) and price category 2 (75%). By using the cross validation, the 75% responds were correctly classified into four price categories and price category 4 was classified with better accuracy (100%) than price category 1 (90.9%), Price Category 2 (62.5%) and Price Category 3 (58.3%).

5. CONCLUSION & RECOMMENDATION

This study based on the attributes belongs to the accessories of motorcycle to evaluate the price of motorcycles. Attributes were Motorcycle Category, Length, Width, Height, Brake System, Presence of Alloy-wheel, Start type Engine Power, Torque, Frame, Fuel Efficiency, Brand, Country of Origin, Resale Value, Fuel tank Capacity. These attributes were selected by considering the ideas of subject expertise in automotive industry in Sri Lanka. The Motorcycle category, Fuel Efficiency, Engine Power, Brake System, Torque and Length provided a significant association between the price categories. The Discriminant Analysis was used to analyze whether there are substantial differences between groups in terms of predictor variables. Most often Discriminant Analysis deals with numerical predictor variables and categorical dependent variable, but it is also possible to use categorical variables as predictor variables. Initially, all the six number of significant variables were used to construct a Discriminant Analysis model. The Canonical Linear Discriminant method was used to analyze the variables. The categorical predictor variables namely Motorcycle Category and Brake System used as the dummy variables. There is high correlation between predictor variables makes the error of classification. The extreme multicollinearity is shown by the high Variance Inflation Factor (VIF>10). By removing the Torque variable which presents the multicollinearity, the remaining five predictor variables were used to develop the Discriminant Analysis model and selected as the best model by using Wilk's Lambda and Overall classification accuracy. The final model contains Engine Power, Fuel Efficiency, Motorcycle Category, Brake System and Length. The two significant Discriminant Functions were classified the

ISSN 2719-2415 (Online)

data into four categories of Price. The purposes of the study are identifying the factors which is associated with the price of motorcycles and constructing the pricing model for the motorcycles. According to the Discriminant Analysis, Motorcycle category, Fuel Efficiency, Engine Power, Brake System and Length were significantly affected to the motorcycle prices.

REFERENCES

Agresti, A. (2007). *An Introduction to Categorical Data Analysis.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Alayande, S., & Adekunle, B. (2015). An Overview and Application of Discriminant Analysis in Data. *IOSR Journal of Mathematics*, 12-15.

Iseri, A., & Karlik, B. (2009). An artificial neural networks approach on automobile pricing. *Expert Systems with Application*, 2155-2160.

Noor, K., & Jan, S. (2017). Vehicle Price Prediction System using Machine Learning Techniques. *International Journal of Computer Applications*, 27-31.

Pardo, P., Georgopoulos, A., Kenny, J., Stuve, T., Findling, R., & Schulz, S. (2006). Classification of Adolescent Phychotic Using Linear Discriminant Analysis. *Science Direct*, 297-306.

Peerun, S., Chummun, N., & Pudaruth, S. (2015). Predicting the Price of Second-hand Cars using Artificial Neural Networks. *Second International Conference on Data Mining, Internet Computing, and Big Data*, (pp. 17-21). Reduit, Mauritius.

Venkatasubbu, P., & Ganesh, M. (2019). Used Cars Price Prediction using Supervised. *International Journal of Engineering and Advanced Technology (IJEAT)*, 216-223.

Yamada, M., Ueda, K., Horiba, I., & Sugie, N. (2001). Discriminantion of road condition toward understanding of vehicle driving environments. *IEEE Transactions on Intelligent Transportation Systems*, 26-31.

ISSN 2719-2415 (Online)