International Journal of Early Childhood Special Education (INT-JECSE)
DOI: 10.48047/INT-JECSE/V13I2.211217 ISSN: 1308-5581 Vol 13, Issue 02 2021

# Passenger's Satisfaction on different modes of transportation in upcoming airport cities – Multiple Discriminant Analysis

Authors:
[1]**Dr. Yaram Srinivasa Reddy,** Associate Professor & HOD,Scient Institute of Technology
[2]**Dr. Vaddula Venkat Reddy**, Associate Professor,Scient Institute of Technology
[3]Dr. Nalla Ramakrishna, Business Analyst

**Abstract:**
Many variables will influence the satisfaction levels of the customers. In that, some of the variables are correlated and they are giving the same results in the multivariate model, and these will make the model more complex. So, reducing a correlated number of variables into fewer factors is called Dimension Reduction Technique. In general, one of the popular usage dimension reduction techniques is factor analysis, where variables will be transformed into fewer factors using the principal component analysisi.e an unsupervised algorithm and it works on maximizing the data points variation and it will not consider the cases or categories as well some of the data will be lost. In this context, one of the most used supervised classification algorithms transforms multiple variables into a few factors by finding the discriminants and classifying them according to the variation of data set and cases or categories is the discriminant analysis. It works as a classifier and satisfaction prediction model. The main objective of the paper is by using a discriminate analysis to identify the satisfaction levels of passengers on factors to the modes of transportation that they are using and classifying them into the respective group according to the cases as well as building a predictive satisfaction model.

**Keywords:** Satisfaction, factor analysis, discriminate analysis, prediction, transportation,

**Introduction:**
Discriminant Analysis:
In general, the cost of computation, inference, and overfitting is more while performing any statistical model learning with many independent variables or features. To overcome those problems the most popular technique is dimension reduction. Using this technique can reduce the number of variables or features into limiting factors that highly influence the dependent variables by combining highly correlated variables into factors with maximizing the variation of the dataset. Dimension reduction can be possible with unsupervised and supervised learnings to bring the principal components.
There are three types of dimension reduction techniques used for reducing the multiple dimension into a linear dimension and are possible for feature selection and extraction.

- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)
- Generalized Discriminant Analysis (GDA)

The most commonly used technique by the researchers is Principal component analysis. Where it is an unsupervised algorithm for reducing the number of variables into fewer factors as principal components as a cluster and further it will be used for the decision-making model. It works by maximizing the dataset variation but not considering the cases or classes of the dataset as well it will not find the discriminates of the selected factors.
The discriminant analysis will help to overcome the limitation of the principal component analysis. It's a supervised machine learning algorithm useful for the dimension reduction from many features into fewer factors by considering the data set similarities and case separation. This analysis is similar to ANOVA, Logistic regression but in ANOVA dependent variables for continuous, in logistic regression dependent variables are dichromatic and categorical variables for discriminant analysis.
For discriminant analysis, there are some assumptions before implementing models are independent variables that should be normal to each other, the variance of the different groups should be the same for the dependent variable, samples should collect randomly from the respondents, and it shouldn't have an autocorrelation problem.
The calculation of the score of the discriminant of the dataset is based on the combination of predicted values and independent variables.

$$Di = a + b1x1 + b2x2 + \ldots + bnxn$$

- Di is discriminated predicted score
- x is predictor

- b is the discriminant coefficient

**Customer Satisfaction:**

(Deneke Dana 2016)Customer satisfaction is the judgment or opinion of the customer by consuming the product or service as per their requirements and fulfillment. Encounter satisfaction and cumulative satisfaction are the two levels of satisfaction. Cumulative satisfaction is judgmental, it comes from not a single experience with the same experiences with many encounters. Whereas encounter satisfaction is a one-time experience without multiple occurrences.

Customer satisfaction is one of the key metrics for any business unit or organization, it tells about where the business stands today. There are several ways to measure a customer's fulfillment. It can be done with the questionnaire by the researchers as surveys, direct feedback of customers to the business or organization, and the organization itself going for the satisfaction levels of the customers on their product or service usage. Satisfaction levels of the end-user bring new insights that are hidden for the growth-causing factor.

(B. Manikandan 2016)The measurement of satisfaction has several dimensions of the customer or organization. For the passenger, the measurement will be according to transportation and their circumstances. It will be economical, comfort, safety, information, time, frequency of services, cleanliness, availability and there are many factors or dimensions.

Literature review:

1. (Dr. B. Rameshwaran, 2018). The author made a study on the developing discriminations of the behaviors of the consumers of two wheels in India of rural and urban consumers while choosing a brand. The outcome of the study concluded that social factors are the main discriminating factor amongst rural and urban two-wheeler consumers while choosing a brand. The sampling method used here stratified random sampling method i.e 100 sample size and conceptual framework prepared based on the literature reviews, secondary data research, and qualitative data. These 4 factors( Primary reference group, secondary reference group, social factor, and role in the society) and the categorical variable is rural and urban are used in the study. Another objective here is finding the major promotional tools for the customer while choosing a brand using the linear discriminant analysis.
2. (Benjamin. Oghojafor, Godson. Mesike,2012). The authors made a study on customer churn of the telecom customers. The main objective of the paper is to find the discriminants(factors) across the customers who are ready to churn and continue the service and also finding the effect of demographic and socioeconomic factors with the discriminants using discriminate analysis of a random sample of 800 network providers. The identified factors are poor service, high call rate, the medium of advertisement, off-beam advertisement medium, service plans, and service provider.
3. (BD. Verdessia, G. Jaraa, R. Fuentesa, JC. Gonzaleza, F. Espejoa, and AC. de Azevedob.2001). The main objective of the authors is to build a prediction model from the customer satisfaction survey as a decision-making model. The two categorical variables are satisfied and unsatisfied and the used sample size is 350. Data collected on the Likert scale on various factors like physician care, pharmacy, nursing, registry, and appointment conditions. The discriminant analysis provided the effect of independent variables on the dependent variables and was able to create a significant decision model.

**Research Gap:**

Regarding the present study reviewed literature related to customer satisfaction and discriminant analysis. From past studies, it is identified that discriminate analysis is applicable for the classification data as well as reducing the number of variables into factors. Studies related to identifying the passenger's satisfaction major factors and discriminates concerning the modes of transportation to reach the nearest airport city are not found by using the discriminant analysis. By this, it will be a research gap for the study to continue.

**Research Objectives:**

- To study the passenger satisfaction levels of various factors of responded passengers those who travel to the nearest airport city
- To find the discriminates of the dependent variables based on the passenger satisfaction levels, reducing the number of variables into fewer factor and classifying them accordingly
- Building the prediction model for passenger satisfaction

Research design:

The present study used descriptive statistics for understanding the respondents. This was followed by applying the discriminant analysis using the tool SPSS 25.0. Independent variables for the study are identified based on the literature review and prepared the questionnaire accordingly. The research variables or factors and dependent variables as the mode of transportation of respondents. This study followed the search methodology steps for findings. Included steps data collection, processing, analyzing then result, and conclusion.

Steps for performing discriminant analysis:

- Design the Questionnaire, sample size and collect the data
- Data should be according to the assumptions of the regression
- Box's Test of Equality of Covariance
- Computing the discriminant functions
- Cross-validation of classification
- Creating a group membership of independent variables

Population and Sample size:

The population of the study is the passengers from tier 2 & tier 3 cities of South India where air mode of transportation is not available and where the government proposed new airports in these cities.

The 500 random passengers are the respondents for the study as sample size those who travel to airport cities by the different modes of transportations. The cities included study are Shimoga, Hassan, Gulbarga, Kannur, Nellore, Vizag, Kothagudem, Karaikal, Bijapur, and Kurnool.

Variables of the study & Scale for measurement:

For the research objectives considered factors as independent variables are related to:

- Economical: Statements related Economical condition of the service and passengers
- Frequency of services: How frequent services are there
- Time: How much time have to get service and how to much time will it take, etc.,
- Comfort: It's about the luxury and comforts of the service and various things
- Information: Passenger required details during the travel
- Safety. Safety and security during the travel

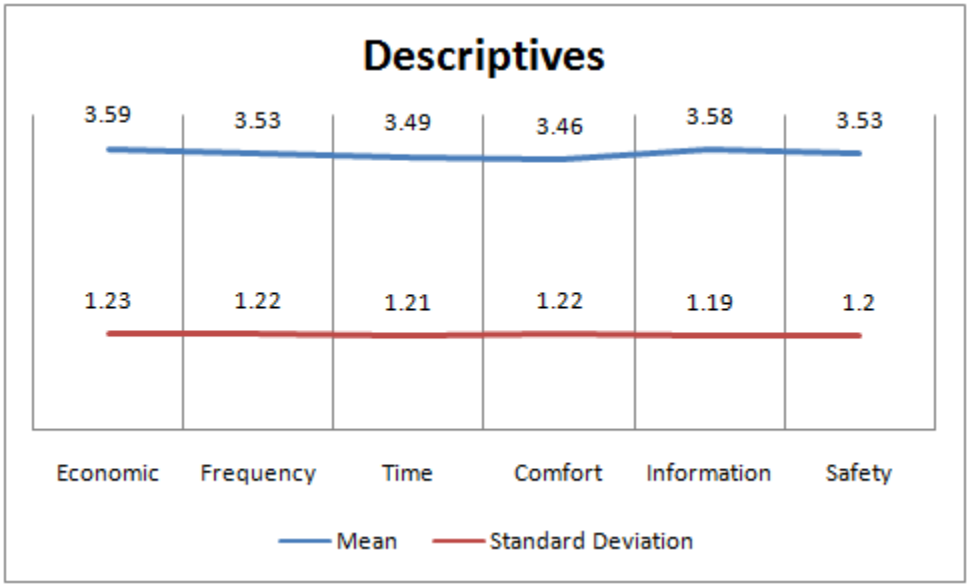
These variables created 18 statements for getting Passenger's satisfaction levels towards the current transportation modes as a dependent variable collected on a 5 points rating scale is 5: Highly Satisfied, 4: satisfied, 3: neutral,2: dissatisfied, and 1: Highly dissatisfied by using a Likert scale.

Data Analysis:

The basic foremost condition that needs to fulfill the requirement of the research is the reliability of the measuring scale used for the collection of complete data. Reliability test used for checking the internal consistency of the developed questionnaire. The questionnaire used for the study having excellent internal consistency i.e 0.97 for 18 items which is higher than the minimum required test value is 0.6.

Descriptive analysis:

The first research objective of the paper is to study the average values on the satisfaction of the factors are Economic, comfort, time, information, safety, and frequency. Each factor has some Likert scale statements and designed with 5 levels of satisfaction score for the 18 statements.




From the above graph, it is understood that the mean satisfaction scores of various factors are more than neutral and less than the satisfaction. The satisfaction level of respondents is not varying much from their satisfaction. The respondents' mean scores are approximately varying 1.2 deviations from the mean. So, the satisfaction levels are not varying from factor to factor.

The mean satisfaction score of the various types of modes are Own Vehicle is 3.7, Private bus is 3.7, Public bus is 3.5, the train is 3.3 and taxi is 3.6. The mean satisfaction values of various types are transportation modes are more than neutral and less than the satisfaction.

To see the discriminates of the dependent variable following analysis carried.

Box's M Test of Equality of Covariance:

Box's M test is used to find the variances among the multiple groups to the dependent variables. For the Box's Test the null hypothesis of equal population covariance matrices.

| Box's M Test Results | | |
|---|---|---|
| Box's M | | 364.02 |
| F | Approx. | 5.907 |
| | df1 | 60 |
| | Sig. | 0 |
| Tests null hypothesis of equal population covariance matrices | | |

From the above test results, it is identified that the p-value is very small i.e., 0 which is less than the 5 % significance level. Hence unable to accept the null hypothesis. The covariance is different for the groups/various types of transportation modes. So, the tastes and preferences are varying from each mode.

**Canonical Discriminant Analysis**

Canonical Discriminant analysis is to differentiate the categorial groups of dependent variables by the lowest dimensional discriminant space. Canonical Discriminant Analysis is the technique for multiple variables.

**EigenValues:**

To find the canonical relationship between the linear combination of variables Eigenvalues are useful. It tells about the discriminating power of functions to separate. The following results show the relationship between the variables and their discriminated power.

| Summary of Canonical Discriminant Functions - EigenValues | | | | |
|---|---|---|---|---|
| Function | Eigenvalue | % of Variance | Cumulative % | Canonical Correlation |
| 1 | 2.104a | 98.9 | 98.9 | 0.823 |
| 2 | .013a | 0.6 | 99.5 | 0.114 |
| 3 | .010a | 0.5 | 100 | 0.099 |
| 4 | .001a | 0 | 100 | 0.031 |
| a. First 4 canonical discriminant functions were used in the analysis. | | | | |

From the above test results, we can observe that function one has the highest discriminate value among the other four functions derived from the independent variables. Function one showing the highest variance i.e., 99% and having a perfect positive correlation between the independent variables.

Wilk's Lambda:

This test shows the significance of the derived discriminant function. As well as the contribution of the variables to discriminant function.

| Wilks' Lambda | | | | |
|---|---|---|---|---|
| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
| 1 through 4 | 0.314 | 571.456 | 20 | 0 |
| 2 through 4 | 0.976 | 11.845 | 12 | 0.46 |
| 3 through 4 | 0.989 | 5.397 | 6 | 0.49 |
| 4 | 0.999 | 0.486 | 2 | 0.78 |

DOI: 10.48047/INT-JECSE/V13I2.211217 ISSN: 1308-5581 Vol 13, Issue 02 2021

From the above test results, it is identified that Wilk's value is near zero which means that the contribution of variables to the discriminant function is more. The p-value for the test is 0.00 which is less than a 5 % level of significance. So, the independent variables are significant to the discriminant function.

Standardized Canonical Coefficients

The derived coefficients of the discriminant function are to know the individual contribution to the equation and their importance.

| Standardized Canonical Discriminant Function Coefficients | | | | |
|---|---|---|---|---|
| **Function** | **1** | **2** | **3** | **4** |
| Comfort | 0.274 | -0.138 | 0.233 | -0.08 |
| Frequency | 0.29 | -0.399 | -0.543 | 0.852 |
| Density | 0.23 | -0.281 | 0.971 | 0.005 |
| Information | 0.321 | -0.22 | -0.642 | -0.94 |
| Maintenance | 0.301 | 1.062 | 0.076 | 0.239 |

From the above table, it is identified that out of 18 variables only five independent variables are useful for discriminating the dependent variable or classification of categories. All the five variables having approximately equal canonical coefficients for function 1 which is significant among the four functions.

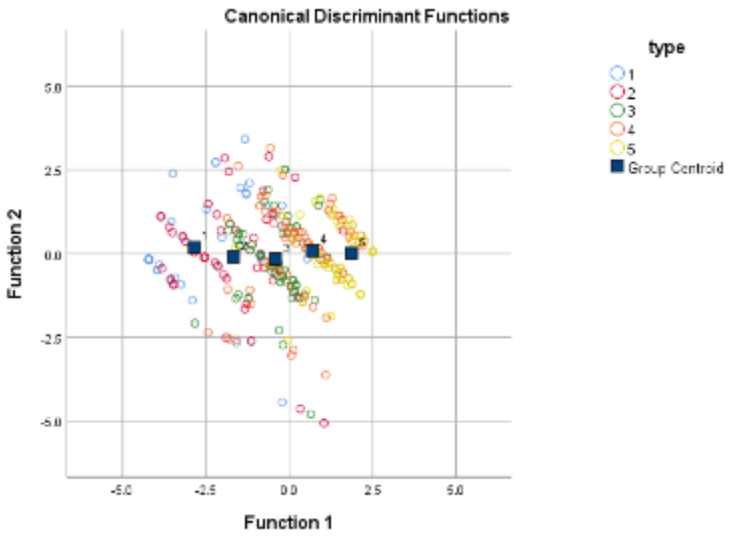
Canonical Structure Matrix

To know the relationship between the discriminant function and the independent variables of the model. Canonical Structure Matrix compares the relationship between the other functions. Here the relationship is about the correlation. The correlation should be more than 0.3 is considerable for the model and functions. Here likewise factor analysis structure matrix coefficients are used for assigning the names for the discriminant functions from the independent variables.

| Structure Matrix | | | | |
|---|---|---|---|---|
| **Function** | **1** | **2** | **3** | **4** |
| Comfort | .713* | -0.127 | 0.167 | 0.06 |
| Frequency | .701* | -0.315 | -0.318 | 0.545 |
| Density | .664* | -0.263 | 0.614 | -0.082 |
| Information | .752* | -0.093 | -0.298 | -0.555 |
| Maintenance | 0.688 | .718* | 0.003 | 0.076 |
| Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions | | | | |
| *. Largest absolute correlation between each variable and any discriminant function | | | | |

As presented in the above table it is observed that out of 18 independent variables only 5 are significant at the 5 % level of significance. Out of 5 variables, 4 are significant to function one and the remaining one significant to the second function. And all five variables are having more than 0.5 correlation coefficients.

Score1: 0.271*Comfort+0.28*Frequency+0.23*Density of travel+0.321*Information+0.31*Maintenance of vehicles




Classification of predictions:

The following classification table helps to understand the comparison between the actual and predicted cases of dependent variables based on the standard coefficients of the derived variables.

From the above table, it is identified that the model can predict 70% of cases correctly when compared to the actual observed cases. Category 1 ( Public Bus) has classification accuracy is about 74%, category 2 (Private Bus) has 50%, category 3(Own Vehicle) has 70%, category 4(Taxi) has 73% and category 5th(Trains) has 73%. Among all, category 2 has less than others.

Conclusion:

The analysis was carried out based on the objectives of the paper. The satisfaction levels of the passengers of the various modes of transportation are between neutral and satisfied. It means there is scope for the development of services. The satisfaction levels on the various factors also vary between neutral & satisfied. So, there is improvement required in all the factors to achieve the satisfaction levels of the passengers. The discriminant analysis can reduce 18 studying variables to 5 variables which are classified under four functions with multiple coefficients and their significance based on the dependent variable's classification. Function one is significant at the 5% level of significance for all the derived variables.

The discriminant analysis reduced the number of variables into fewer based on the variance produced by the respondents on the various statements of multiple factors. With the derived function coefficients from the analysis, developed a prediction equation to classify the passenger category. This means based on the scoring on the statements predicting or classifying the passenger's mode of transportation by reducing the number of study variables. So, the analysis can classify 70% of passenger groups correctly then compared them to the observed groups.

Classification of predictions:

The following classification table helps to understand the comparison between the actual and predicted cases of dependent variables based on the standard coefficients of the derived variables.

| Predicted Group Membership | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Type of Mode | Public Bus | Private Bus | Own Vehicle | Taxi | Trains | Total |
| Original | Count | Public Bus | 39 | 7 | 6 | 1 | 0 | 53 |
| | | Private Bus | 12 | 30 | 15 | 4 | 0 | 61 |
| | | Own Vehicle | 2 | 5 | 81 | 24 | 4 | 116 |
| | | Taxi | 6 | 6 | 16 | 126 | 19 | 173 |
| | | Trains | 2 | 1 | 4 | 19 | 71 | 97 |
| | % | Public Bus | 73.6 | 13.2 | 11.3 | 1.9 | 0 | 100 |
| | | Private Bus | 19.7 | 49.2 | 24.6 | 6.6 | 0 | 100 |
| | | Own Vehicle | 1.7 | 4.3 | 69.8 | 20.7 | 3.4 | 100 |
| | | Taxi | 3.5 | 3.5 | 9.2 | 72.8 | 11 | 100 |
| | | Trains | 2.1 | 1 | 4.1 | 19.6 | 73.2 | 100 |
| a. 69.4% of original grouped cases correctly classified. | | | | | | | | |

From the above table, it is identified that the model can predict 70% of cases correctly when compared to the actual observed cases. Category 1 ( Public Bus) has classification accuracy is about 74%, category 2 (Private Bus) has 50%, category 3(Own Vehicle) has 70%, category 4(Taxi) has 73% and category 5th(Trains) has 73%. Among all, category 2 has less than others.

**Conclusion:**

The analysis was carried out based on the objectives of the paper. The satisfaction levels of the passengers of the various modes of transportation are between neutral and satisfied. It means there is scope for the development of services. The satisfaction levels on the various factors also vary between neutral & satisfied. So, there is improvement required in all the factors to achieve the satisfaction levels of the passengers. The discriminant analysis can reduce 18 studying variables to 5 variables which are classified under four functions with multiple coefficients and their significance based on the dependent variable's classification. Function one is significant at the 5% level of significance for all the derived variables.

The discriminant analysis reduced the number of variables into fewer based on the variance produced by the respondents on the various statements of multiple factors. With the derived function coefficients from the analysis, developed a prediction equation to classify the passenger category. This means based on the scoring on the statements predicting or classifying the passenger's mode of transportation by reducing the number of study variables. So, the analysis can classify 70% of passenger groups correctly then compared them to the observed groups.

**Reference**


- (Dr. B. Rameshwaran, 2018). A Discriminant Analysis of the Social Factors Influencing the Consumer Behavior in the Two-wheeler Market in India- An Empirical Study. Global journals, 18(6), 2249-4588.
- (Benjamin. Oghojafor, Godson. Mesike, 2012).DiscriminantAnalysis of Factors Affecting Telecoms Customer Churn. International Journal of Business Administration.3(2), 1923-4007,10.5430/ijba.v3n2p59.
- (BD. Verdessia, G. Jaraa, R. Fuentesa, JC. Gonzaleza, F. Espejoa, and AC. de Azevedob.2001).The role of discriminant analysis in the refinement of customer satisfaction assessment. Journal of Public Health, 6(34),0033-3506. 10.1590/S0034-89102000000600010
- https://stats.idre.ucla.edu/spss/dae/discriminant-function-analysis/#:~:text=Version%20info%3A%20Code%20for%20this,needed%20to%20describe%20these%20differences
- Manikandan, B. &Dr.Vanniarajan, T. (2016). SERVICE QUALITY IN BUS SERVICES: AN EMPIRICAL STUDY IN TAMILNADU. shanlax journals. 4(2)
- Deneke, Dana. Million. Nane, Mebratu. Belete, Teshome. Ergado&Teferi, Labis. (2016). A SURVEY OF PASSENGERS SATISFACTION ON SERVICE QUALITY OF PUBLIC TRANSPORT SECTOR: THE CASE OF SNNPR, ETHIOPIA. International Journal of Multidisciplinary Academic Research (IJMDAR). 4(5). 2309-3218