**Journal of ISOSS**
**2017 Vol. 3(2), 243-252**

# CLASSIFYING HEALTHCARE CONSUMERS' DECISION MAKING ABOUT HEALTHCARE PROVIDERS: DISCRIMINANT ANALYSIS APPROACH

**Zaheer Abbas** and **Safia Khalid**
Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: zaheerabbas@uog.edu.pk

## ABSTRACT

The purpose of this study is to identify factors that affect people's choice about selection of healthcare provider and healthcare services. In health care market it is more demanding to understand how individual makes their decisions. Moreover, selection of healthcare services is also very complex. For this study, primary data was collected from 350 respondents through two-stage stratified random sampling. For identification of important factors, we used discriminant analysis to identify factors that best distinguish respondent's decision while choosing health care services. The result of this study reveals that the Word-of-Mouth and Personal Information Source are best discriminating factors that affect individual's decision making. Discriminant model also predict that ninety-six percent (96 %) of estimated grouped cases are correctly classified.

## KEY WORDS

Healthcare provider, Health care services, Discriminant Analysis, Reliability Analysis, Decision Making.

## INTRODUCTION

Healthcare marketing is comparatively more challenging now a days. To understand how medicine consumers' behavior affects marketing, it is very important to identify such factors that have direct effect on their behavior. Marketer's goal is to understand the customer behavior because they seek which products and services they buy or not.

Generally consumer behavior defined as when, where and how people use and dispose of products. Study of consumers facilitates organizations to advance their marketing policy. Islam and Farooqi (2013) referred that companies continually struggle to observe that which products and services customers want to buy and how we can modify their demands. However it's necessary for organizations to understand the decision making of consumers to improve their business. Today everyone have more than one option, reflects the difficulty in consumer decision making. Hence marketers try to examine the factors that influence peoples to make their purchase decision. In any market consumer behavior is very tricky to understand but in health care industry it is more complex due to many reasons. As it's consist series of visit and generally people are also more sensitive about their health. So mostly their decision has been affect by personal and social factors. When the product was more complex and private or we have no

 

capability to evaluate that in such circumstances we prefer to get suggestion from family, friends or some skilled persons.

Friends and family discussion are more powerful for example their referred doctor has great effect on my selection of doctor decision. Different types of electronic medium have influencing power in our decision making process. As we are more conscious about health so we often influenced by opinion leaders. Opinion leaders are leading the opinion of others and direct other decision. All these social factors (Social Structures, Communication Medium, and Opinion Leader) have great influencing power on our decision. Now we discuss some personal factors (Personal Information Source, Perceived Risk, and Word-of-Mouth).

Person's individual effort to gather information for the selection of doctor for example searching information from newspapers, bulletin boards is referred as personal information source. Perceived risk considered consumer level of uncertainty during buying a product or negative consequences that can be take place regarding a services or product. There are different types of perceived risk such as financial, functional, social, psychological, time and security risk. We cannot take risk about our health so we collect information from different resource. People twist to others for suggestion and advice before purchasing a product is not new.

Word-of-mouth is a powerful source of information. Mair, et al. (2013) suggested that traditional word-of-mouth is characterized by a high level of credibility and influence when it comes to purchase decisions. Companies often expend millions of rupees on marketing or advertising but mostly word-of-mouth from trustworthy sources has great influence on purchase decision (Sernovitz, 2012). Across the world word-of-mouth considered more influencing power in making a decision. Word-of-mouth from reliable sources such as family, peers and opinion leaders are more powerful.

The purpose of the study is to identify the factors that discriminant the individual who make their decision or not based on the considered factors while selecting a doctor or health care services using stepwise discriminant analysis. Results also improve our perspective whether considered personal and social factors can classify decision making of people while selecting health care services.

## OBJECTIVE

- To classify consumer decision in medical market on the bases of considered factors.
- To identify the factor(s) that discriminated between individuals decision making by discriminating function.

## MATERIAL AND METHODS

In this study cross sectional study design was used. Population includes all students of Social Sciences, CS&IT and Management Sciences, Faculty and Administrative staff of University of Gujrat in Hafiz Hayat Campus. The students, faculty and staff of university of Gujarat are came from all over the Pakistan, especially from Gujrat, Gujranwala,

Sialkot and Mandi Bahauddin districts. A sample of 400 selected from 3269 population size in our selected campus using (Yamane, 1967) formula with 0.05 margin of error.

We used Two-Stage Stratified random sampling design for sample selection. We have used equal allocation at first stage of stratified random sampling and proportional allocation method at second stage. From total 400 respondents we have selected 200 faculty, administrative staff and 200 students. 25% non-response received from selected sample. For analysis purposes we used sample of 350 respondents. For data collection, a well-structured questionnaire was used as research instrument. There are total thirty eight items were asked at five-point Likert-Scale. A stepwise discriminant analysis are used to analyzing data.

## RESULTS AND DISCUSSION

The value of Cronbach's alpha based on standardized item is 0.854 in Table 1 which indicates that the data is reliable for further statistical analysis.

Table 2 shows the average rank of the respondents on the social structure is 3.8, it means on the average respondents are agree on the above factor. 4.2% respondent strongly disagree, 8.3% respondents disagree, 13.6% neutral, 53.9% agree and 20.0% strongly agree about the social structure factor**.** 9.1% respondent strongly disagree, 23.5% respondents disagree, 22.4% neutral, 33.9% agree and 11.2% strongly agree about communication medium factor**.**

Discriminant analysis used to model the value of a dependent categorical variable based on its relationship to one or more predictors.

$$d_{ik} = b_{0k} + b_{1k}x_{ik} + ... + b_{pk}x_{ip}$$

where

$d_{ik} =$ is the value of the k$^{th}$ discriminant function for the i$^{th}$ case, $p =$ is the number of predictors $b_{jk} =$ is the value of the j$^{th}$ coefficient of the k$^{th}$ function, $x_{ij} =$ is the value of the i$^{th}$ case of the j$^{th}$ predictor). Two group discriminant analyses used for prediction of Decision Making of people for the selection of doctor. Decision making is two-group categorical variable (No, Yes). Social Structure, Opinion Leaders, Communication Medium, Word-of-Mouth, Personal Information Source, Perceived Riskareused as independent variable in the model.

Table 3 shows the group means for each of the independent variables, based on 350 observations. In profiling the two groups, we first identify the two variables with the largest difference in the group means (Word-of-Mouth, Personal Information Source). Table 1 also shows the Wilk's lambda, univariate ANOVA and p-values used to assess the significance between the means of the independent variables for the two groups. All the independent variable has significant p-values, Wilks's lambda and univariate F values represent the separate or univariate effects of each variable, not considering multicollinearity among the independent variables. These tests indicate that the two independent variables (Word-of-Mouth, Personal Information Source) the only two variables with significant univariate differences between the groups. Although greater

statistical significance corresponds to higher overall discrimination (i.e., the most significant variables have the lowest Wilk's lambda values), it does not always correspond to the greatest discrimination between all the groups. All of these measures combine to help identify the sets of variables that form the discriminant function. The data in the Table 3 show that the first variable to enter in the model is Word-of-Mouth because it meets the criteria for statistically significant differences across the group and has the smallest Wilk's lambda value.

In Table 3 from the review of group differences, we saw that Word-of-Mouth had the largest significance difference between groups. Thus in Table 4 word-of-mouth is entered as the first variable in the stepwise procedure. Only one variable enters in the discriminant model at this time, so the significance levels and measures of group differences match those of the univariate tests. After Word-of-Mouth enters the model, the remaining variables are evaluated on the basis of their incremental discriminating ability (group mean differences after the variance associated with Word-of-Mouth is removed). Again, variable with significance level greater than .05 are eliminated from consideration for entry at the next step.

Examining the univariate difference shown in Table 1 identifies Personal Information Source as the variable with the second most significant differences. Yet the stepwise process does not use these univariate results when the discriminant function has one or more variables in the discriminant functions. It calculates the $D^2$ values and statistical significance test of group differences after the effect of the variable(s) in the model is removed (in this case only Word-of-Mouth is in the model). As shown in the last portion of Table 4 Personal Information Source remain the next best candidates to enter the model because it has the highest Mahalanobis $D^2$ (6.225) and the largest $F$ to enter value.

In step 2 (see Table 5) Personal Information Source enter the model as expected. The overall model is significant $(F = 252.801)$ and improves in the discrimination between groups as evidence by the decrease in the Wilks' lambda from 0.438 to 0.407. Moreover the discriminating power of both variables included at this point is also statistically significant ( $F$ values of 409.039 for Word-of-Mouth and 26.260 for Personal Information Source). With Word-of-Mouth and Personal Information Source all statistically significant, the procedure moves to identifying any remaining candidates for inclusion.

As seen in the last portion of Table 5 none of the remaining four independent variables meet the entry criterion. Thus the estimation process stops with two variable (Word-of-Mouth and Personal Information source) constituting discriminant function. Table 6 provides the overall stepwise discriminant analysis results after all the significant variables are included in the estimation of discriminant function. This summary table describes two variables (Word-of-Mouth and Personal Information Source) that were significant discriminators based on their Wilk's Lambda and minimum Mahalanobis $D^2$ values. A number of different results are shown addressing both overall model fit and the impact of specific variables. Canonical Discriminant Functions reported the multivariate measures of overall model fit. Discriminant function is highly significant (.000) and displays a canonical correlation of .770. We interpret this correlation by squaring it.

$(0.770)^2$ = .5929). Thus, almost 60 percent of the variance in the dependent variable (Decision Making) can be accounted for by this model, which include two independent variables.

Discriminant weights are available in unstandardized and standardized forms. The unstandardized weights (plus the constant) are used to calculate the discriminant score, but can be affected by the scale of the independent variable. Thus, the standardized weights more truly reflect the impact of each variable on the discriminant function. Standardized weights show the relative importance of each variable in discriminant function. Table 6 provides the standardized weights (coefficients) for the two variables included in the discriminant function. The discriminant loadings are reported under the headings "Structure matrix" and are ordered from highest to lowest by the size of loading. Discriminant loadings, in contrast to discriminant weights, are less affected by multicollinearity. Discriminant loadings are calculated for all variables, ever for variables not included in the discriminant function. In this study the strongest effect in the discriminant function based on the loading value, is Word-of-Mouth.

The classification function coefficients, also known as Fisher's linear discriminant function, are used in the classification. Group centroids represent the mean of the individual discriminant function score for each group. Group centroids provide a summary measure of the relative position of each group on the discriminant function. Table 6 reveals that the group centroids for the No (group 1) is -1.575, where the group centroids for the yes (group 2) is .920. To compute the overall mean, multiply the number in each group by its centroids and add the result (e.g., 129 × -1.575 + 221 × .871 = -10.684). The signs reflect the relative mean profile of the two groups. The positive signs are associated with variables that have higher score for group 2. The negative weights and loadings are for those variables with the opposite pattern (i.e. higher values in group 1). Thus, the sign indicate the pattern between groups.

Our next step is to assess the classification accuracy with the overall model statistically significant and explaining 60 percent of the variation between groups (See in proceeding discussions and Table 4).

Cutting Score can be calculated as

$$Z_{cs} = \frac{N_A Z_B + N_B Z_A}{N_A + N_B} = \frac{(129 \times 0.920) + (221 \times -1.575)}{129 + 221} = -.6554$$

By substitution of the appropriate values in the formula we can obtain the critical cutting score (assuming equal prior probabilities and equal cost of misclassification) of $Z_{cs} = -.6554$. The procedure for classifying Decision Making of respondents with the optimal cutting score is as follows:

- Classify a respondent as being in group 1 (No category) if its discriminant score is less than $-.6554$.
- Classify a respondent as being in group 2 (Yes category) if its discriminant score is greater than $-.6554$.

Classification matrices for the observation in analysis were calculated and the results are shown in Table 7. The classification table shows the practical results of using the model. Cells on the diagonal of the cross-classification are correct classifications. Cells off the diagonal of the cross classification are incorrect classifications. 125 of the 129 cases who not make decision are classified correctly. 214 of the 221 cases who make their decision are correctly classified. 96 % of estimated grouped cases are correctly classified. Ninety six percent (96%) of cross-validated grouped cases are correctly classified on the other hand only four percent (4%) estimated and cross-validated cases are incorrect classified by the model.

## CONCLUSION

In this study have classified healthcare consumers' decision making based on considered factors by using Discriminant Analysis. Study findings reveal that Word-of-Mouth (Oral Communication) and Personal Information Source are most significant factors that affect people decision making. These two factors provide more powerful distinction between the two groups. The models have 96 percent correct classification power it means that this model can be used for predictions.

## APPENDIX

**Table 1**
**Test of Reliability**

| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of items |
|---|---|---|
| **0.850** | 0.854 | 38 |

**Table 2**
**Overall Percentages, Mean and Standard Deviation of all Ordinal Scale Variables**

| Factor / Variables | SD% | DA% | N% | A% | SA% | Mean | S.D |
|---|---|---|---|---|---|---|---|
| Social Structure | 4.2 | 8.3 | 13.6 | 53.9 | 20.0 | 3.8 | 1.0 |
| Communication Medium | 9.1 | 23.5 | 22.4 | 33.9 | 11.2 | 3.1 | 1.2 |
| Opinion Leaders | 3.4 | 10.6 | 19.6 | 45.6 | 20.8 | 3.7 | 1.0 |
| Personal Information Source | 9.6 | 27.5 | 20.7 | 31.7 | 10.4 | 3.1 | 1.2 |
| Perceived Risk | 8.6 | 17.1 | 16.7 | 37.5 | 20.1 | 3.4 | 1.2 |
| Word-of-Mouth (Oral Communications) | 4.3 | 14.5 | 21.8 | 46.9 | 12.5 | 3.5 | 1.0 |
| Decision Making | 4.6 | 17.3 | 22.0 | 44.0 | 12.1 | 3.4 | 1.1 |

**Table 3**
**Group Descriptive Statistics and Tests of Equality in the Two-Group Discriminant Analysis**

| Independent Variables | Dependent Variable Group Means: | | Test of Equality of Group Means*: | | |
|---|---|---|---|---|---|
| | Group 1: No ( *n* =129) | Group 2: Yes ( *n* = 221) | Wilks' Lambda | F Value | p-value |
| Social Structure | 10.7287 | 11.6606 | .950 | 18.413 | .000 |
| Communication Medium | 17.2791 | 19.8054 | .933 | 24.829 | .000 |
| Opinion Leader | 13.9225 | 15.2986 | .921 | 29.931 | .000 |
| Personal Information Source | 19.3256 | 22.6244 | .887 | 44.448 | .000 |
| Perceived Risk | 19.6667 | 21.1538 | .957 | 15.715 | .000 |
| Word-of-Mouth | 17.0698 | 23.1719 | .438 | 446.903 | .000 |

**Table 4**
**Results from Step 1 of Stepwise Two-Group Discriminant Analysis**

**Overall Model Fit**

| | Value | F value | Degree of Freedom | p-value |
|---|---|---|---|---|
| Wilks' Lambda | 0.438 | 446.903 | 1,348 | 0.000 |

**Variable Entered/Removed at Step 1**

| Variable Entered | Minimum $D^2$ | F | |
|---|---|---|---|
| | | Value | p-value |
| Word-of-Mouth | 5.487 | 446.903 | 2.163E-64 |

Note: At each Step, the variable that maximizes the Mahalanobis distance between the two closet groups is entered.

**Variable in the Analysis After Step 1**

| Variable | Tolerance | F to Remove |
|---|---|---|
| Word-of-Mouth | 1.000 | 446.903 |

**Variables Not in the Analysis After Step 1**

| Variable | Tolerance | Minimum Tolerance | F to Enter | Minimum $D^2$ | Between Groups |
|---|---|---|---|---|---|
| Social Structure | .939 | .939 | 0.407 | 5.498 | no and yes |
| Communication Medium | .973 | .973 | 1.000 | 5.515 | no and yes |
| Opinion Leader | .911 | .911 | 0.321 | 5.496 | no and yes |
| Personal Information Source | .997 | .997 | 26.260 | 6.225 | no and yes |
| Perceived Risk | .969 | .969 | 0.021 | 5.487 | no and yes |

**Table 5**
**Results from Step 2 of Stepwise Two Group Discriminant Analysis**

| **Overall Model Fit** | | | | |
|---|---|---|---|---|
| | **Value** | **F value** | **Degree of Freedom** | **p-value** |
| Wilks' Lambda | .407 | 252.801 | 2,347 | 0.000 |

| **Variable Entered/Removed at Step 2** | | | |
|---|---|---|---|
| **Variable Entered** | **Minimum $D^2$** | **F** | |
| | | **Value** | **p-value** |
| Personal Information Source | 6.225 | 252.801 | 1.831E-68 |

Note: At each Step, the variable that maximizes the Mahalanobis distance between the two closet groups is entered.

**Variable in the Analysis After Step 2**

| **Variable** | **Tolerance** | **F to Remove** | **$D^2$** | **Between Groups** |
|---|---|---|---|---|
| Word-of-Mouth | .997 | 409.039 | 0.546 | no and yes |
| Personal information Source | .997 | 26.260 | 5.487 | no and yes |

**Variables Not in the Analysis After Step 2**

| **Variable** | **Tolerance** | **Minimum Tolerance** | **F to Enter** | **Minimum $D^2$** | **Between Groups** |
|---|---|---|---|---|---|
| Social Structure | .937 | .935 | .741 | 6.248 | no and yes |
| Communication Medium | .756 | .756 | 2.406 | 6.298 | no and yes |
| Opinion Leader | .890 | .890 | 1.715 | 6.277 | no and yes |
| Perceived Risk | .875 | .875 | 2.149 | 6.290 | no and yes |

**Table 6**
**Summary Statistics for Two Group Discriminant Analysis**

| **Overall Model Fit: Canonical Discriminant Function** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Function** | **Eigen Value** | **Percent of variance** | | **Canonical Correlation** | **Wilks' Lambda** | **Chi-Square** | **df** | **p-value** |
| | | **Function %** | **Cumulative %** | | | | | |
| 1 | 1.457 | 100 | 100 | .770 | .407 | 311.942 | 2 | 0.00 |

| **Discriminant Function and Classification Function Coefficients** | | | | |
|---|---|---|---|---|
| | **Discriminant Function** | | **Classification Function** | |
| | | | **Group 1:** | **Group 2:** |
| Independent Variables | Unstandardized | Standardized | No | Yes |
| Word-of-Mouth | .367 | .956 | 2.607 | 3.523 |
| Personal Information Source | .077 | .345 | 1.047 | 1.239 |
| Constant | -9.335 | | -33.055 | -55.526 |

| **Structure Matrix**[a] | |
|---|---|
| Independent Variables | Function 1 |
| Word-of-Mouth | .939 |
| Opinion Leader | .329 |
| Communication Medium | .315 |
| Personal Information Source | .296 |
| Perceived Risk | .272 |
| Social Structure | .249 |

*This variable not used in the analysis

| **Group Means (Centroids) of Discriminant Function** | |
|---|---|
| Decision Making | Function 1 |
| No | -1.575 |
| Yes | .920 |

[a] Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function

**Table 7**
**Classification Results for Two-Group Discriminant Analysis**

| Classification Results [b,c] | | | | |
|---|---|---|---|---|
| **Sample** | **Actual Group** | **Predicted Group Membership** | | **Total** |
| | | **No** | **Yes** | |
| Estimation Sample | No | 125<br>96.9% | 4<br>3.1% | 129 |
| | Yes | 7<br>3.2% | 214<br>96.8% | 221 |
| Cross-Validated[a] | No | 125<br>96.9% | 4<br>3.1% | 129 |
| | Yes | 7<br>3.2% | 214<br>96.8% | 221 |

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.
b. 96.9% of original grouped cases correctly classified.
c. 96.9% of cross-validated grouped cases correctly classified.