Discriminant Analysis as an Aid to Human Resource Selection and Human Resource Turnover Minimization Decisions

Nasir Uddin[1], Md. Siraz Meah[2] & Raihan Hossain[3]

[1] Department of Business Administration, Yokohama National University, Japan

[2] Faculty of Business Administration, BGC Trust University, Bangladesh

[3] Credit Approver, SME, Standard Chartered Bank, Bangladesh

Correspondence: Nasir Uddin, Department of Business Administration, Yokohama National University, Tokiwadai 79-4, Hodogaya-Ku, Yokohama 240-8501, Japan. Tel: 81-45-339-3659. E-mail: uddin-nasir-xc@ynu.ac.jp

Abstract

A discriminant analysis is conducted in order to estimate a discriminant function to determine the expected status of the faculty post candidates in a private university in Bangladesh. The explanatory variables are age of the candidate, salary offered for the post, whether the candidate has foreign degree-dummy variable and result in masters' examination of the candidate in Bangladesh. Statistically significant differences are observed in the group means of the variables of the two groups: not stayed faculty & stayed faculty. The log determinants are found approximately equal in size for the groups while the Box's M value shows that the assumption of the equal co-variances is violated. However, the uni-variate normality tests are conducted and found the variables follow approximately normal distribution. Consequently, we proceeded to estimate the discriminant function. The estimated function is significant at 1 per cent level of significance and can explain 50 per cent of the variations in the group memberships. The structure matrix shows that the variables: result (0.526), f-degree (-0.489) and salary (0.408) are very important and the age (0.127) is the least important determinants of the expected status of the faculties. Finally, the prediction matrix of the holdout sample shows that 83 per cent of the cases are classified correctly.

Keywords: human resources selection, turnover, two-group discriminant analysis

1. Introduction

Like developed countries, in Bangladesh, most of the organizations have introduced a human resource management department in order to efficient management of human resources and to increase the productivity of the organization in recent years. The suitable human resources selection is one of the main functions of the human resources management department. At present, in the most of the organization, human resources are selected in a traditional way-based on a written test and/or an oral interview. But this selection process is creating a large human resources turnover ratio in the organizations. In fact, a discussion with the head of the human resource department of the organization studied in this study reveals that the human resource turnover ratio is about 30-40 per cent per year for the organization and the post studied in this research. As a result, this high human resources turnover ratio is a big concern to the organization as it creates substantially higher human resources management cost.

High human resources turnover ratio generates substantial high cost to the concerned organization. As for example, a dairy firm owner asserts, "Every time a milker leaves, I loss about a cow" (Billikope 2003). Furthermore, if a salesperson leaves the organization, the lost sale with the unmanned area is ranging from $50, 000-$75,000 (Futrell and Parasuraman, 1984). They also reported that some organizations face 50 percent turnover in their sales force recruited in the less than last two years. In addition to financial loss; there may be an overall reduction in organizational welfare, leakage of confidential information to competitors and others. Hence taking initiative for reducing human resource turnover ratio by any organization is justified for economic and non-economic reasons. Like many other developing countries, use of discriminant analysis in human resources selection/faculty selection is not found in Bangladesh.

In this study, an attempt is made to model human resources selection process in order to minimize human the resources turnover ratio and to decrease the human resources management costs. An old technique called linear discriminant analysis (originally developed by Fisher, 1936) is used in this study in order to achieve the objective. The discriminant analysis is look like regression analysis in terms of dependent variable, number of independent and nature of independent variables. For instance, in both of the methods, independent variables can be metric or non-metric. In addition, one dependent variable and single/multiple independent variables can be used in both methods. Furthermore, discriminant analysis uses OLS in order to estimate parameters α and β_i by minimizing the within group sum of squares. But discriminant analysis differs significantly from regression analysis in terms of nature of dependent variables. In linear discriminate analysis, dependent variable is a categorical variable/non-metric whereas in regression analysis; the dependent variable is a metric variable. However, binary logistic regression analysis is similar to two-group discriminant analysis.

The linear discriminant function takes form of a linear combination of coefficients of variables and their respective variables in the study as equation 1. The variable coefficients are estimated such that the function maximizes the distance between the two centroids. That happens when a ratio (λ)-between group sum of squares to within group sum of squares is maximized. For any other combination, the function will not be optimal because the validity of the model will not be justified. The coefficients are unstandardized coefficients in the equation 1. However, the standardized coefficients also can be estimated. But in a standardized function, there is no constant term as the mean of a standardized variable is Zero. The larger the coefficient the better the independent variable in discriminating between the groups. A good independent variable should have large weight.

$$Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + - - + \beta_k X_k = \alpha + \sum_{i=1}^{k} \beta_i X_i \qquad (1)$$

where, Z = discriminant score, α = a constant term, β_i = the discriminant coefficient or weight of the variable, X_i = predictor or independent variable, i = number of predictor variables; $i = 1,2,3,....k$.

By substituting the unstandardized values of a new case in the estimated unstandardized coefficients equation and by substituting the standardized values of a new case in the estimated standardized coefficients equation, the expected position of the case is determined. In this study, the main focus is on unstandardized discriminant function. The number of discriminant functions (NDF) those could be estimated from a discriminant analysis is less than or equal to the minimum number of number of categories (G) in the dependent variable minus one or the number of predictor variables (P):

$$NDF \leq Min\{G - 1, P\} \qquad (2)$$

So, in a two-group discriminant analysis, only one function is estimated. As stated earlier, the between group sum of squares (SSB) to within group sum of squares (SSW) which is maximized at the time of the estimation of the discriminant function is as follows:

$$Max\ \lambda = \frac{\sum_{g=1}^{G} N_g[\bar{y}_g - \bar{y}]^2}{\sum_{g=1}^{G}\sum_{p=1}^{N_g}[y_{pg} - \bar{y}_g]^2} = \frac{SSB}{SSW} \qquad (3)$$

Where G = Number of groups, g = Group g, $g = 1, 2, ..., G$, N_g = Number of firms in group g, y_{pg} = Firm p in group g, $p = 1.......Ng$, $\bar{y}_g$ = Group mean/centroid, $\bar{y}$ = Over all sample mean/grant mean.

Figure- I shows the pictorial presentation of the data collected on two variables: X1 and X2 for a two groups: G1 and G2 discriminant analysis. If there are two variables, then the data can be presented in the graph like figure-1 and could have a look at the two clusters and their characteristics profiles. Upper part of figure-1 shows that some cases of G1 group are also in G2 group. The cases which are actually member in G1 but included in G2 also are misclassified. Similarly, some cases of the G2 are also in the G1 group. The cases those are actually members of G2 but included in G1 also are misclassified by the figure 1. In other words, 13 cases are common in the two groups in figure-1. Symbolically, n (G1∩G2) = 13. Actually, one case should be in one group as group member because group membership is mutually exclusive and collective exhaustive but the figure showed in both which is misclassified presentation by the upper part of the figure.

Like upper part of the figure, separating the two groups graphically is not possible when the data are collected on more than two variables as we have two axes to present two variables in the graph. But the problem is easily solved by the discriminant analysis like the lower part of the figure 1. For the purpose, discriminant analysis generates Z-scores: negative or positive value for the cases and a new axis-Z is created when data is collected on more than two variables. In Figure 1, Z scores are computed for the two groups: G1 and G2. The Z values of the G1 group is expected to be negative as centroid for this group is negative and denoted by a new set G1$^{/}$ But some G1$^{/}$ cases have positive value which means that the discriminant function misclassified those cases.

Similarly, the Z values for G2 group is expected to be positive as the centroid for this group is positive and denoted by G2$^/$. But some G2$^/$ cases have negative value which means that those cases are misclassified by the discriminant analysis. The miss-classified cases are represented by the shaded area. The smaller the shaded area, the greater accuracy of the discriminant function is ensured. (Uddin, 2013; Malhotra & Das, 2011). The average of the Z values of the all groups is denoted by G0.
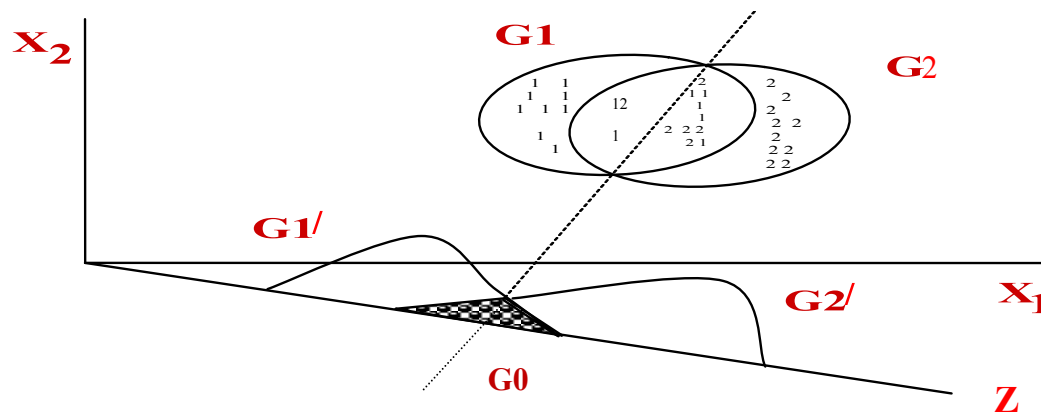

Figure 1. Discriminant analysis

The broad objective of this study is to estimate a two-group discriminant function in order to identify efficient human resources-faculty who will stay in the University for the long run. In consistent with the broad objective, the specific objectives are as follows: (i) to prepare the characteristics profiles of the groups in the dependent variable, (ii) to check whether significant difference exists between the group means of the variables, (iii) to estimate a discriminant function to estimate the expected status of the candidate for the faculty post in the university, (iv) to point out which variables contributing most in determining the group membership. (v) to check the validity/acceptability of the estimated function, (vi) to recommend policy implications to the policy maker.

The rest of the study is organized as follows. Section 2 is about the literature review of the study. Section 3 describe the methodology of the study. Section four deals with the conducting discriminant analysis and section 5 is about conclusion of the study. The study report ends with a references list.

2. Literature Review & Variables Selection for the Study

Literature in discriminant analysis for decision making is very large. For instance, discriminant analysis is widely used for insolvency prediction, and sales force turnover management. However, any research paper presenting discriminant function for faculty selection for a university is not found by us. As a result, it is not possible to review any paper in this research field. But, in order to design this research, several papers in other fields are studied. The papers in other field helped the author substantially to design and conduct this research. Specifically, the several papers are examined to notice the kind of variables used in their study.

Lucas et.al (1987) argued that by reducing human resource turnover ratio, the costs of human resource selection, training & development, and maintenance can be reduced substantially. To achieve the purpose, they conducted a discriminant analysis with a view to estimating a discriminant function to determine the expected status of the sales force of a selling organization. To estimate the function, they have used three employee characteristics variables and four attitude variables. Three employee characteristic variables are age, education, tenure and four attitude variables are intrinsic job satisfaction (IJS-satisfaction with work), extrinsic job satisfaction (EJS-pay, fringe benefits, job security etc), supervisory consideration (SC-employee employer relationship), and task-specific-self-esteem (TSSE-self perceived quality and quantity of performance). They have used three techniques to analyze their data: multiple regression analysis, MANOVA, and discriminant analysis. The discriminant analysis part of their study shows a more than 70 per cent hit ratio that ensures the validity of using discriminant analysis in the field (Boyd, Westfall & Stasch 2005).

Welker (1974) proposed to use discriminant function for a CPA firm to select candidate from a large number of applicants who might stay in the firm for the long-run. He proposed to use the general information variables: age, asset balance, debt balance; performance information measures: college point average, class ranking in the

college of business, number of hours of accounting class in a business college, number of hours of quantitative class in a business college, achievement test scores (professional courses), average expected hour to stay away from home per week, expected hours of overtime per week, and size of the firm (no. of employees or clients); psychological test variables: I.Q test results, dominance-submission rank, conformity-nonconformity rank, passivity–tension rank, extroversion-introversion rank; and interviewer's subjective rank of the applicant variables (0-10): appearance, communication, and overall in the discriminant analysis function for the large CPA firm. Jensen & Bailey (1975) supported Welker's (1974) proposal for employee selection but suggested to use alternative method if the assumptions are not satisfied by the data.

The above two study shows that wide range of variables can be included in human resource turnover study by using discriminant analysis. The variables like age, income, salary are easily available than variables like psychological variable. But the final variable selection could be done based on the significance of contribution of the variables in the discriminant function or discriminant score. In addition to the papers of Lucas et.al (1987) and Welker (1974), the research papers helped in design current research are: Awh and Waters (1974), Capon (1982), Credit Card Redlining (1979), Wekesa, Samuel and Peter (2012), Dinh and Kleimeier (2007), Grablowsky (1975), Hand and Henley (1997), Lusk (1972), Miodonski (2004), Moncrief, Hoverstad & Lucas (1989), Aktan (2011), Walker (1998), Chancharat (2008), Fern (1989), Davis *et. al* (1992) and Wiginton (1980).

In this research, only four variables are included in the discriminant analysis because of lack of data. The variables are age of the applicant (age), salary per month for the post (salary), whether the applicant has foreign degree-if has foreign degree then1, otherwise 0 (F-Degree), and result of masters of the candidate in Bangladesh in CGPA (result). The discriminant analysis shows that the variables could predict the group memberships by more than 75 per cent correctly.

3. Methodology

3.1 Data

The study is mainly based on primary data. It is collected from job application forms of the faculty by filling up a predetermined questionnaire. Secondary data is collected from books, journals, magazine, websites and SPSS manual of George & Mallery (2006). The collected data is divided into two samples: analysis sample and holdout sample. There is no specific rule about the proportion of each sample in total sample. The division may 50-50, 60-40, or 75-25. In this study, the collected primary data is divided into two samples as (1) analysis sample consists of 70 per cent and (2) holdout/split/validation sample consists of 30 per cent and each sample contains equal proportion of the groups: not stayed (1), and stayed (2) as proportionately stratified random sampling rule.

Analysis sample consists of 15 not stayed (1) faculty & 15 stayed (2) faculty and the holdout sample consists of 6 not stayed (1) faculty and 6 stayed (2) faculty. The analysis sample is used to estimate the discriminant function and the holdout sample is used to check the accuracy of the prediction power of the model. If possible, it is better to collect data for a larger number of cases then the total data set should be divided into two groups: analysis sample and holdout sample equally. The analysis sample should be used to estimate the discriminant function and the holdout sample should be used to forecast the accuracy of the estimated model. Then the role of the data set should be reversed. In other words, the holdout sample should be used to estimate the model and the analysis sample should be used to check the accuracy of the estimated model. This process is well known as cross-validation approach.

3.2 Description of the Variables

The variables used in this study are divided into two types: dependent variable and independent variables. The only dependent variable is the status of the faculty which is a categorical variable. The employee who has left the university is denoted by not stayed (1) and who stayed in the university is denoted by stayed (2). The independent variables are, result in GPA in masters examination of the candidate for the faculty post in Bangladesh (result), whether the faculty has foreign degree (dummy variable). Based on the historical data collected from the resumes, if the faculty has a foreign degree, that is denoted by 1 and if the faculty does not have any foreign degree that is denoted by 0 (f-degree), salary is the per month financial benefit of the faculty for the faculty post. Finally, age means the age of the faculty at the time of application.

3.3 Data Analysis Technique and Software Used

In order to answer the research questions of the study, direct method of discriminant analysis is used as an analysis technique in this study. According to the direct method in discriminant analysis, all of the independent variables are included in the model simultaneously. The direct method discriminant analysis is used based on

previous researches or by other sources, if the researcher knows that discriminant analysis should be based on all of the independent variables. Besides, Graham (2001) argued not to use step-wise discriminant analysis for several limitations. According to the stepwise method discriminant analysis, the independent variables are included in the model according to the discriminating power of the variables.

3.4 Assumptions and Guidelines

3.4.1 Equal Variance-Covariance Matrix

The main assumption to conduct the discriminant analysis is that the groups have equal variance-covariance matrices although their means are substantially different. This assumption is tested by using a transformed value of Box's M, which compares the equality of log determinants of the categories in the dependent variable, called F ratio. Theoretically, this F is equivalent to the F of ANOVA analysis which is a ratio of between group variability to within group variability. In the test, the null hypothesis (H_0) is the variance-covariance matrices of the groups are the same in the population. If the p-value (Sig.) of the test is less than 0.05, the null hypothesis is rejected at 5 per cent level of significance. As a result, the assumption of equal variance-covariance matrices is violated. This problem can be overcomed as follows. (i) The violation is not a problem if the violation is because of skewness but not by outliers (Tabachnick and Fidell, 1996). (ii) If the sample size is large than this violation cannot be a big problem and the validity of estimating the discriminant function can be checked by hit ratio of the holdout sample.

3.4.2 Multivariate Normality

An important assumption of estimating discriminant analysis is that all of the groups in the dependent variable are selected randomly from a multivariate normal population. To test this assumption, a variable by variable normality test of the variables can be conducted using histogram with normal curve, if the variables follow normal distribution then proceeding to estimate the function is justified on the multivariate normality ground.

3.4.3 No Multicollinearity

There should not be multicollinearity in the independent variables. The correlation matrix can be used to check the multicollinearity of the variables. In addition, the relationship of one independent variable with the rest of the independent variables could be checked by running regression of one independent variable on the rest of the independent variables and checking the R^2. The multicollinearity problem is also can be solved by using stepwise discriminant analysis. If multicollinearity is found in the data, should be corrected.

3.4.4 Linearity

Unlike regression analysis, in discriminant analysis, the dependent variable is non-metric. Consequently, there is no linear relationship between dependent (non-metric) and independent (metric) variables. The linear relationship is required between the independent variables. To test this assumption, we will check the degree of relationship of one independent variable with another independent variable and we will check the degree of relationship of one independent variable with the rest of the independent variables. If one variable is consistently found non-linear with the other variables than only the variable should be attended.

3.4.5 Outliers

Discriminant analysis is very outlier sensitive. Hence outlier should be identified and excluded from the analysis. According to the *SPSS Base 10.0 Applications Guide*, page 259, "cases with large values of Mahalanobis distance from their group mean can be identified as outliers." In addition to Mahalanobis distances, the Box plots and histogram can be used to identify the outliers. The outliers should be excluded from the analysis.

3.4.6 Sample Size

Some researchers argue to have 20 sample sizes per predictor variable. But many times, that could not be achieved because of lack of data. However, Burns and Burns (2008) argued that the sample size in the smallest group of the dependent variable should be at least 5 times higher than the number of the predictor variables. Like regression analysis, discriminant analysis should be estimated based on a data set as large sample size as possible. Small sample size may produce wrong discriminant function.

3.4.7 Other Points to Remember

The observations must be random sample. The categories in the dependent variable should be well defined, categories should be well defined before data collection and all group cases must be mutually exclusive and collectively exhaustive. A contentious variable or a variable on which data is collected on a scale should not be divided into group just to use discriminant analysis.

4. Conducting the Discriminant Analysis

4.1 Group Means

Group means and standard deviations for each variable for not stayed faculty (1) and stayed (2) faculty are calculated in table 1. Group mean provides an idea about whether the means of the variables differ between the groups. In addition, group means and group standard deviations can be used as characteristics profile for the two groups. The table 1 shows that the group means are different for the variables: age, salary, f-degree and result. In the section 4.2, the statistical significance of the mean differences is tested.

Table 1. Group statistics

Status	Not Stayed (1)		Stayed (2)		Total	
Variables	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Age	29.67	3.11	31.07	7.16	30.37	5.47
Salary	28,067	4,399	34,733	10,687	31,400	8,716
F-Degree	0.47	0.52	0.07	0.26	0.27	0.45
Result	3.3	0.32	3.59	0.21	3.45	0.3

4.2 Tests of Equality of Group Means

The Wilk's lambda and the F ratio are used to test the equality of means of the groups for the same variable. The Wilk's lambda for the each predictor is equal to the ratio of the within group sum of squares to the total sum of squares. It is estimated from one way analysis of variance by considering status variable as independent variable and the predictor variable as dependent variable. The Wilk's lambda is also known as U statistic. The range of Wilk's lambda value is 0 to 1. If a variable's wilk's is less 0.95, it is revealed that the group means are significantly different. The larger the value, the smaller significance and the smaller the value, the larger significance is ensured. The Wilk's lambda and the transformation of its value to F is done as under and presented in the table 2.

$$Wilk's\ Lambda\ (\Lambda) = \frac{|SSW|}{|SSB+SSW|} \tag{4}$$

Transformed value of Λ to an equivalent F is calculated as in MANOVA:

$$F_{approx}(df_1, df_2) = \left(\frac{1-y}{y}\right)\left(\frac{df_2}{df_1}\right) \tag{5}$$

For n is equal in all groups in the dependent variable:

$y = \Lambda^{\frac{1}{s}}$, P is the number of independent variables.

$$s = \sqrt{\frac{p^2(df_{effect})^2 - 4}{p^2 + (df_{effect})^2 - 5}} \tag{6}$$

df_{effect} = number of groups minus one: $(k-1)$.

$df_1 = p(df_{effect})$

df_{error} = number of groups times $(n-1)$: $k(n-1)$

$$df_2 = s\left[(df_{error}) - \frac{p-df_{effect}+1}{2}\right] - \left[\frac{p(df_{effect})-2}{2}\right] \tag{7}$$

For unequal n in groups in dependent vraiable, $df_{error} = (N - k)$ = Total sample size − no. of categories.

Table 2. Tests of equality of group means

Variables	Wilks' Lambda	F	Sig.
Age	0.983	0.483	0.493
Salary	0.849	4.991	0.034
F-Degree	0.795	7.2	0.012
Result	0.771	8.309	0.007

Table-2 shows that all of the means of the variables have very significant differences between the groups. The lowest Wilk's lambda presents highest importance in the discriminant function. Hence, the most important variable in discriminant function is result and the lowest important variable is the age of the faculty. The result is exactly supported by the p-value of the F-test.

4.3 Estimate the Discriminant Function Coefficients

Test of Equality of Covariance Matrices by Using Box's M: In order to estimate a valid discriminant function, an important assumption is that the variance-covariance matrices of the groups should be the same. The variance-covariance matrices are presented in table 3. The pooled within group is computed by taking the average of the variance-covariance matrices of the groups. An overview of the variance-covariance matrices shows that the variance-variance matrices are substantially different. Now, we will test statistically in the table-4 whether the variance-covariance matrices are same or different.

Table 3. Variance-covariance matrices[a]

Status	Variables	Age	Salary	FDegree	Result
Not Stayed (1)	Age	9.67	4452	0.45	-0.26
	Salary	4452.38	19352381	-533.33	-545.81
	FDegree	0.45	-533	0.27	0.07
	Result	-0.26	-546	0.07	0.1
Stayed (2)	Age	51.21	68162	-0.08	0.39
	Salary	68161.9	114209524	376.19	739.43
	FDegree	-0.08	376	0.07	0.01
	Result	0.39	739	0.01	0.04
Pooled Within-Groups Matrices	Age	30.44	36307.143	0.19	0.06
	Salary	36307.14	66780952.38	-78.57	96.81
	FDegree	0.19	-78.571	0.17	0.04
	Result	0.06	96.81	0.04	0.07

Note: a. the covariance matrix has 28 degrees of freedom.

In table 4, rank 4 means the size of the covariance matrix, the 4 means that this is a (4x4) matrix, and the 4 is the number of independent variables in the discriminant function.The covariance metrics are same if the log determinants of not stayed group covariance matrix and the log determinant of stayed group covariance matrix are the same. Whether the log determinants are the same is tested by the Box's M. Box's M for this study is 30.37. The transformed F. ratio is computed in order to test the equality of the two covariance matrices. The F ratio is similar to ANOVA which is the ratio of between group variability to within group variability. In the test, the null hypothesis (H_0) is that the variance-covariance matrices of the groups are the same in the population. A P-value (sig) of 0.00 means that the null hypothesis is rejected; consequently, the assumption is violated. However, a value less than 0.05 do not automatically cancel the prospect of the estimation of discriminant function. Although the assumption is violated, most of the time, the discriminant function can be found valid at the time of validity check. This is surprisingly true for many cases. This happens if the violation of the

assumption is because of skewness not for outliers. In case of a large sample size, the violation is not a problem in the model's high accurate forecasting rate. However, since the significance ratio (p-value) is very low it is justified to check the uni-variate normality of the variables. A uni-variate normality test of the variables show that some variables do not follow normal distribution. As our sample is very small, we do not drop any cases from the analysis for which the variables become not-normal. If we can drop the cases for which the variables do not follow normal distribution, we can have a better estimated discriminant function with higher hit ratio. This point should keep in mind in real life discriminant analysis with utmost importance.

Table 4. Test of equality of covariance matrices by using box's M

Status	Rank	Log Determinant	Box's M	Approx. F	df1	df2	Sig.
Not Stayed (1)	4	14.686	30.37				
Stayed (2)	4	14.752		2.56	10	3748	0
Pooled within-groups	4	15.804					

Note: Tests null hypothesis of equal population covariance matrices.

4.4 Determine the Significance of the Discriminant Function

Function-1 in the table 5 means that one function is estimated for a two-group discriminant analysis. The eigen value means a ratio of between group sum of squares to within group sum of squares as equation 3. The larger value means the better estimation of the discriminant function. The minimum acceptable value of eigen value is 1.00. The higher the better. Another way to test the significance of the Eigen value is to check the canonical correlation of the function. The canonical correlation(r) for the estimated function is 0.72. Consequently, the coefficient of determination (=SSB/SST) of the function is $(0.72^2) = 0.50$ which means that 50% of variation in the group memberships in the dependent variable is explained by the estimated function. Furthermore, Wilk's lambda (=SSW/SST) is also used to check the significance of the estimated function. The Wilk's lambda for the estimated function is 0.482.The smaller the Wilk's lambda, the more significance of the estimated function is assumed. The transformed chi-square $((\chi^2 = - [(n - 1) - 0.5 (m + P + 1)] \ln \Lambda$, df= (k - 1), m = number of discriminant function extracted, P = number of predictor variables) is 18.96 with 4 degrees of freedom and level of significance (Sig.) is 0.001. In the chi-square test, the null hypothesis (H_0) is the centroids of the categories and the grand centroid all are equal. So at 1% level of significance, the null hypothesis is rejected. Thus estimating the discriminant function and interpreting the results are statistically significant.

Table 5. Determine the significance of the discriminant function

Function	Eigen value	% of Variance	Cumulative %	Canonical r	Test of Function(s)	Wilks' λ	χ	df	Sig.
1	1.073[a]	100	100	0.72	1	0.482	18.96	4	0.001

Note: a First 1 canonical discriminant function is used in the analysis.

4.5 Interpreting the Results

4.5.1 Structure Matrix

The coefficients of structure matrix (table 6) are known as factor loadings or canonical loadings. A factor loading represents the correlation between the variable and the estimated discriminant function. By squaring a factor loading, we can determine the variation the variable can explain in the dependent variable. The larger the coefficient, the more important in determining the group membership. Thus the absolute size of coefficient represents relative importance of the variable. The factor loadings are calculated by using equation-8 and equation-9 and presented in the table-6. The importance of a variable is calculated first to determine the relative importance of a variable in the discriminant function by using equation-8. The importance of a variable is equal to the weight or coefficient of the variable times the mean difference value of the two groups of the variable.

Symbolically:

$$I_i = W_i(\bar{X}_{i1} - \bar{X}_{i2}) \tag{8}$$

where, I_i = importance of the variable i, W_i = coefficient of the variable i, $\overline{X}_{i1}$ = average value of variable i for group 1, $\overline{X}_{i2}$ = average value of variable i for group 2, I = number of predictors: *1, 2, 3,---k*. A variable's relative importance is equal to the variable's importance divided by the sum of all variables importance as equation-9. Symbolically,

$$R_i = \frac{W_i(\overline{X}_{i1} - \overline{X}_{i2})}{\sum W_i(\overline{X}_{i1} - \overline{X}_{i2})} \qquad (9)$$

Where, R_i = relative importance of the variable i.

Alternatively, the structure matrix can be calculated as equation-10:

$$L = R_W D \qquad (10)$$

where, L is the loadings matrix, R_w is the within groups correlation matrix, and D is the vector of standardized discriminant function coefficients.

The structure matrix, table 6, is arranged by showing the variables as an order of highest important in determining group memberships to lowest important in determine group memberships. In our study, result is the most important variable followed by foreign degree and salary. The lowest important variable is the age of the applicant. By squaring the coefficient of a variable, the variation in the dependent variable by a variable can be explained can be determined. For instance, result can explain 28 (= 0.526^2) per cent variation in the group memberships. In addition to structural matrix, standardized coefficients also are used to check the relative importance of the variables. Table-6 shows the relative importance of the variables in the study.

Table 6. Structure matrix

Variables	Function 1
Result	0.526
F-Degree	-0.489
Salary	0.408
Age	0.127

4.5.2 The Function

The main objective of this study is to estimate a discriminant function to estimate the expected position of the faculty post candidate in a private university in Bangladesh. Table-7 shows the coefficients of the variables in the unstandardized discriminant function. The coefficients are the multiplier of the variables when they are in original measurement units. By using the variables and the coefficients, the required discriminant equation 1 is in the form equation-11. The equation is often known as discriminator.

Z = -10.001354 - 0.068052 Age + 0.0000805 Salary -1.815383 F-Degree + 2.909966 Result   (11)

Table 7. Canonical discriminant function coefficients (unstandardized coefficients)

Variables	Function1
Age	-0.0680519
Salary	0.0000805
F-Degree	-1.815383
Result	2.9099664
(Constant)	-10.001354

By using the discriminator, the HR manager can estimate the expected position of an applicant applied for the faculty post. For the purpose, from the application document, the values of variables will have to be substitute in the equation-11. If the Z score for a candidate is negative then his expected position is not stayed (1) as the centroid for not stayed group is negative and the higher the distance between Z and 0, the higher possibility of

leaving the university. If the Z score is positive then the expected position of the applicant is stayed (2) as the centroid for stayed group is positive and the larger the distance between Z and 0, the larger possibility that the applicant for the post will stay in the University for the long run. Thus management can select persons whose expected status is stayed for the long run in the University and can minimize the faculty turnover ratio.

4.5.3 Group Centroids

The centroid is the mean value of the discriminant scores for a particular group. There are as many centroids as there are groups, as there is one for each group. In our two group discriminant analysis, we have two centroids. The discriminant scores for the analysis sample (Zs) are presented in the last column of the table-9. Symbolically, they are calculated as equation 12:

$$\bar{y}_g = \frac{1}{N_g} \sum_{p=1}^{N_g} y_{pg} \qquad (12)$$

The centroid for the not stayed group (1) is -1.001 and the centroid for the stayed (2) group is 1.001. In absolute term, the larger the Z value the better estimate is estimated. These centroids are used to determine the expected status of the new applicants for the faculty post. If a new applicant comes, the values of the variables from his application form or resume will be substituted in the estimated function and then the Z value for the applicant will be computed. Then, the Z score will be compared with the centroids for a decision. If the Z score for the new applicant is negative then the expected position is not stayed (1) as the centroid is negative for the not stayed (1) group and if the Z score for the new applicant is positive then the expected position is stayed (2) as the centroid is positive for the stayed (2) group. The cutting point for the decision making is zero as the average of the centroids of the groups are equal in absolute term. This is the case always if the sample sizes of the groups in dependent variable are equal and the analysis is a two-group discriminant analysis. If the sample sizes of the groups in the dependent variable are not equal, then the cutting point will be the weighted average of the centroids. The groups' centroids are presented in table 8.

Table 8. Functions at group centroids

Employee Type	Function1
Not Stayed (1)	-1.001
Stayed (2)	1.001

Note: Unstandardized canonical discriminant functions evaluated at group means.

4.5.4 Casewise Statistics

Table 9 presents an important summary of part of the analysis. In the table, the case number means the sample serial number in the study. Actual group means the cases on which actual data is collected and predicted group means the predicted group membership of the cases in the actual group. The double asterisk (**) in the predicted group means the miss-prediction by the estimated discriminant function. For instance, the case number 6 in the original analysis sample was actually belongs to group 1 but the estimated model forecasted the case number 6 as a member of the group 2. The probability of the highest group is the probability of being in the first predicted group. The group in the second highest group is the second highest possibility of being in a predicted group. Since our analysis is a two-group analysis, this group membership is the just alternative of the predicted group membership. For example, the model predicted the actual group case 1 as group 1 case , so the second highest group here will be definitely group 2 for the case one as our analysis is a two-group analysis. The second highest group probability is the possibility of being in the second highest group. The last column of the table presents Z score for the analysis sample cases. The average of the Z scores of the two groups: 1 and 2 are called centroid for group 1 and centroid for group 2 respectively. The lower part of the table represents the above statistics except Z for the cross-validated sample. In cross-validated sample, statistics for each case is estimated based on the function derived from all of the cases other than that case.

In addition, Mahalanobis (1893–1972) distance (Mahalanobis, 1936) is a distance measure used to measure the distances between a case and the centroid. A new case will have the measures by which the case is compared with the centroids and is included in the group for which the distance is smallest. The distance calculation method takes into account- the variances in each direction are different and co-variance between the variables at the time of distance calculation. Mahalanobis distance is calculated as equation 13:

$$d_j(x) = \left(x - \bar{x}_j\right)^T S_{pl}^{-1}\left(x - \bar{x}_j\right) \tag{13}$$

where, S_{pl} = pooled estimate of the covariance matrix.

A case x is classified by using one Mahalanobis distance for each centroid to one of the G groups. A case will be included in a group for which the distance is minimum. Mahalanobis distance is measured in standard deviation (SD) unit. If the distance is more than 1.96 between a case and a centroid, then the probability is less than 5 per cent that the case will be in the group of the centroid.

Table 9. Casewise statistic

	Case Number	Actual Group	Highest Group					Second Highest Group			Discriminant Scores
			Predicted Group	P(D>d | G=g) p	df	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid	Group	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid	Function 1
Original	1	1	1	.778	1	.809	.079	2	.191	2.960	-.720
	2	1	1	.778	1	.808	.080	2	.192	2.956	-.718
	3	1	1	.436	1	.609	.607	2	.391	1.495	-.222
	4	1	1	.570	1	.704	.323	2	.296	2.055	-.433
	5	1	1	.554	1	.694	.351	2	.306	1.987	-.409
	6	1	2**	.351	1	.534	.870	1	.466	1.142	.068
	7	1	1	.471	1	.637	.520	2	.363	1.641	-.280
	8	1	1	.017	1	.999	5.741	2	.001	19.341	-3.397
	9	1	1	.317	1	.500	1.002	2	.500	1.002	.000
	10	1	1	.059	1	.997	3.555	2	.003	15.112	-2.886
	11	1	1	.212	1	.989	1.557	2	.011	10.559	-2.249
	12	1	1	.436	1	.610	.606	2	.390	1.497	-.223
	13	1	1	.459	1	.627	.549	2	.373	1.590	-.260
	14	1	1	.870	1	.911	.027	2	.089	4.687	-1.164
	15	1	1	.262	1	.986	1.256	2	.014	9.750	-2.122
	16	2	2	.562	1	.699	.336	1	.301	2.023	.421
	17	2	2	.668	1	.946	.184	1	.054	5.911	1.430
	18	2	2	.254	1	.986	1.303	1	.014	9.881	2.142
	19	2	2	.933	1	.862	.007	1	.138	3.676	.916
	20	2	2	.083	1	.996	3.011	1	.004	13.965	2.736
	21	2	1**	.374	1	.556	.791	2	.444	1.237	-.111
	22	2	2	.468	1	.634	.528	1	.366	1.627	.275
	23	2	2	.494	1	.653	.469	1	.347	1.735	.316
	24	2	2	.448	1	.971	.577	1	.029	7.624	1.760
	25	2	2	.254	1	.986	1.302	1	.014	9.877	2.142
	26	2	2	.758	1	.800	.095	1	.200	2.871	.693
	27	2	2	.364	1	.547	.823	1	.453	1.198	.094
	28	2	2	.918	1	.858	.011	1	.142	3.607	.898
	29	2	2	.401	1	.580	.706	1	.420	1.350	.161
	30	2	2	.890	1	.907	.019	1	.093	4.581	1.139
Cross-validated [a]	1	1	1	.264	4	.723	5.236	2	.277	7.157	
	2	1	1	.350	4	.736	4.439	2	.264	6.487	
	3	1	2**	.135	4	.606	7.008	1	.394	7.870	
	4	1	1	.255	4	.579	5.330	2	.421	5.966	
	5	1	1	.821	4	.659	1.529	2	.341	2.843	
	6	1	2**	.753	4	.601	1.904	1	.399	2.720	
	7	1	2**	.143	4	.559	6.868	1	.441	7.343	
	8	1	1	.000	4	1.000	26.557	2	.000	46.369	
	9	1	2**	.694	4	.571	2.228	1	.429	2.801	
	10	1	1	.163	4	.998	6.524	2	.002	19.113	
	11	1	1	.197	4	.990	6.034	2	.010	15.220	
	12	1	1	.776	4	.567	1.783	2	.433	2.323	
	13	1	1	.770	4	.584	1.812	2	.416	2.490	
	14	1	1	.664	4	.893	2.392	2	.107	6.635	
	15	1	1	.580	4	.986	2.869	2	.014	11.323	
	16	2	2	.744	4	.654	1.956	1	.346	3.232	
	17	2	2	.766	4	.938	1.834	1	.062	7.254	
	18	2	2	.613	4	.986	2.681	1	.014	11.226	
	19	2	2	.975	4	.850	.481	1	.150	3.952	
	20	2	2	.000	4	1.000	162.575	1	.000	192.776	
	21	2	1**	.701	4	.640	2.190	2	.360	3.344	
	22	2	1**	.017	4	.730	12.025	2	.270	14.014	
	23	2	2	.848	4	.621	1.379	1	.379	2.369	
	24	2	2	.635	4	.968	2.556	1	.032	9.374	
	25	2	2	.076	4	.987	8.475	1	.013	17.191	
	26	2	2	.973	4	.786	.504	1	.214	3.110	
	27	2	1**	.696	4	.510	2.215	2	.490	2.296	
	28	2	2	.846	4	.837	1.390	1	.163	4.661	
	29	2	2	.792	4	.540	1.693	1	.460	2.017	
	30	2	2	.897	4	.895	1.084	1	.105	5.366	

For the original data, squared Mahalanobis distance is based on canonical functions.
For the cross-validated data, squared Mahalanobis distance is based on observations.
**. Misclassified case
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.

4.5.5 Histogram and Box's plots of Z Values of Status 1 (Not stayed) & Status 2 (Stayed)

Figure 2 shows the pictorial presentation of the Z scores of the cases of the analysis sample presented in the last column of the table 9. The left hand side histogram presents the Z−scores of the not stayed (1) group. Table 9 and the histogram shows that the minimum Z is 0.068, maximum is -3.397, average is -1.001 and the standard deviation 1.115. Here, it is notable that the expected sign of Z score of a case from not stayed (1) is negative as

the centroid for not stayed (1) is negative. The larger value of Z with negative the better. One case generated positive value- indicates the miss-classification of the case by the estimated model. So, 14 cases out of 15 of not stayed (1) group is correctly classified by the estimated model.

The right hand side histogram presents the Z scores of the stayed (2) group reported in the last column of table 9. The table and the histogram show that the minimum Z score of the cases in stayed (2) group is -0.111, the maximum Z score is 2.736, average of the cases of stayed (2) is 1.001 and the standard deviation is 0.87. The larger value of the Z the better estimate is assumed as the centroid for the stayed (2) group is positive. One case of this group generated negative value, reports misclassification. Hence, 14 cases out of 15 of stayed (2) group are classified correctly. In total, 28 cases of 30 cases are classified correctly. Thus the accuracy rate or hit ratio of the estimated model for the analysis sample is 93 per cent.
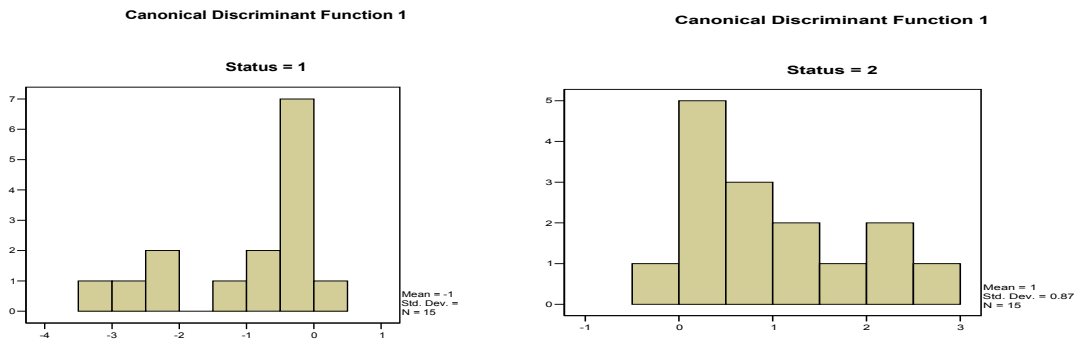
Figure 2. Histogram of Z values of status 1 (not stayed) & status 2 (stayed)

In addition to the above histogram, the Z scores of the cases of the analysis sample can be presented in the Box plots as figure 3. The Box plots are used to check the distribution of the Z scores and to find out the outliers if exist any. The Box plots show that the Z scores of the not stayed (1) group does not follow normal distribution and the Z scores of the stayed (2) group follows approximately normal distribution. In spite of this violation of rule, the discriminant function is found significant by significantly higher hit ratios. The hit ratio can be higher if we could exclude outliers and the cases for which the variables become not-normal. We could not do that because of limitation of very small size.
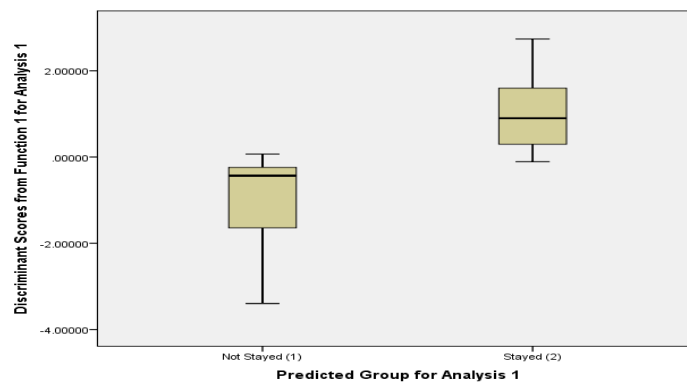
Figure 3. Box plots illustrating the distribution of discriminant scores for the two groups

4.6 Assessing the Validity of the Model

4.6.1 Classification Matrix of the Analysis Sample

In order to check the validity of the model and to know the accurate forecasting power of the estimated model, at this stage, will look at the classification/prediction/confusion matrix. The matrix is constructed based on the prediction of the analysis sample by the estimated model and presented by the table 10. The primal diagonal of the matrix presents the accuracy rate oh the model and the off diagonal of the matrix presents the misclassification rate of the estimated model. The first element of the primal diagonal presents the rate-a case is actually in group 1 and the estimated model forecasted as group 1 divided by the number of the cases in group 1 and the rate is 93 percent. The second element of the primal diagonal presents the rate-a case is actually in group 2 and the estimated model forecasted the case as group 2 divided by the number of cases in group 2 and the rate is 93 per cent. The total of the primal diagonal divided by the total number of the cases in the analysis sample is equal to correct prediction rate which is also known as hit ratio. In addition, the off-diagonal presents the misclassification rate of the estimated model. The first element of the off-diagonal matrix presents the rate- a case actually is in group 1 but the model forecasted as group 2 divided by number of cases in group 1 and the rate is 7 per cent. The second element of the off diagonal element present the rate- a case actually is in group 2 but the model forecasted as group 1 divided by the number of cases in group 2 and the rate is 7 per cent. In aggregate, the accuracy rate of the model is 93 per cent.

The classification matrix of the original sample may be biased because the model is estimated by including the case for which expected status is estimated. So, cross-validation classification matrix is prepared and it is better to compare the accuracy rate of the cross-validation to the standard accuracy rate to check the validity of the model. In the cross-validation analysis, the case for which expected status is estimated is not included in the analysis sample to estimate the discriminant function. Thus, the process is continued as many times as many cases in the analysis sample. To the end of the analysis, the cross-validation matrix is constructed. The aggregate accuracy rate of the model is 77 per cent which is quiet high compared to the minimum standard rate 70 per cent.

Table 10. Classification results (b, c)

			Predicted Group Membership		Total
		Status	Not Stayed (1)	Stayed (2)	
Original	Count	Not Stayed (1)	14	1	15
		Stayed (2)	1	14	15
	%	Not Stayed (1)	93	7	100
		Stayed (2)	7	93	100
Cross-validated(a)	Count	Not Stayed (1)	11	4	15
		Stayed (2)	3	12	15
	%	Not Stayed (1)	73	27	100
		Stayed (2)	20	80	100

Note: 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. 93% of original grouped cases correctly classified. c. 77% of cross-validated grouped cases correctly classified.

4.6.2 (a) Classification Matrix of the Holdout Sample

The holdout sample is used to check the validity of the sample furthermore. By substituting the values of the cases of the holdout sample in the estimated model, the Z scores are calculated for the cases. Based on the Z scores and the centroids, the expected status of the cases are estimated and table-11 is constructed. The table shows that 83 per cent of the cases are correctly classified. In compare to the standard, this hit ratio is very high. Hence, estimating & using the discriminant function in human resource selection is justified.

Table 11. Classification results-holdout sample

		Status	Predicted Group Membership		Total
			Not Stayed (1)	Stayed (2)	
Holdout Sample	Count	Not Stayed (1)	5	1	6
		Stayed (2)	1	5	6
	%	Not Stayed (1)	83	17	100
		Stayed (2)	17	83	100

Note: a. 83.0% of holdout cases correctly classified.

4.6.2 (b) Casewise Statistics of the Holdout Sample

By substituting the values of the cases of the holdout sample in the estimated function, the casewise Z values are computed as shown in the table 12 for the cases. In the table, the double asterisk (**) means misclassified cases by the estimated discriminant model. It is notable that 5 cases out of 6 cases of group 1 are correctly classified and 5 cases out of 6 cases of group 2 are correctly classified. Thus, the estimates in the table 12 shows that 83 per cent of cases are correctly classified.

Table 12. Casewise statistics-holdout sample

SL No.	Status	Z Value	Predicted Status
1	Not Stayed (1)	-1.12192	Not Stayed (1)
2	Not Stayed (1)	0.105333	Stayed (2)**
3	Not Stayed (1)	-1.23349	Not Stayed (1)
4	Not Stayed (1)	-1.50551	Not Stayed (1)
5	Not Stayed (1)	-2.22399	Not Stayed (1)
6	Not Stayed (1)	-0.58345	Not Stayed (1)
7	Stayed (2)	0.77339	Stayed (2)
8	Stayed (2)	0.163024	Stayed (2)
9	Stayed (2)	1.498397	Stayed (2)
10	Stayed (2)	0.153629	Stayed (2)
11	Stayed (2)	-0.74409	Not Stayed (1)**
12	Stayed (2)	3.484578	Stayed (2)

Note: ** misclassified case.

The accuracy rate of the estimated function must be compared with the standard accuracy rate set by the statisticians and the scholars in the field to justify the logic of estimating and using the discriminant function (Uddin 2013). Furthermore, some researchers argued that the accuracy rate of the estimated function should be compared with the probability of selecting right employee, if the employee is selected randomly from the analysis sample. If the groups are equal in size of the employee than the probability of selecting expected employee is 1/number of groups in the analysis. In our study, number of groups is 2 and the number of not stayed and stayed faculty in each group is equal. In details, in the analysis sample 15 are not stayed faculty & 15 are stayed faculty and in the hold out sample 6 are not stayed faculty and 6 are stayed faculty. So if randomly selected, the possibility of selecting an expected employee from the analysis sample is 0.5. Joseph, William, Barry, & Ralph, (2010) and Glen (2001) argued that accuracy rate of 0.25 higher than the random chance is justified to estimate a discriminant function. In addition, Boyd, Westfall & Stasch (2005) argued that more than 70 per cent accuracy is acceptable to estimate and to use discriminant function in decision making. In this case, the minimum accuracy rate is 77 per cent, which ensures the validity of the model.

4.6.3 τ-Test to Compare Model's Classification Rate and Random Classification Rate

A test statistic, tau (τ), can be computed to check the acceptability of the overall classification of the model. The statistic generates a number which can be explained as 'fewer errors compare to the random classification'. The

value of τ is large enough for the original analysis sample (= 86.67%), the cross-validated sample (= 53.33%) and the holdout sample (= 66.67%), those values prove the substantial improvement in right faculty selection by using the estimated function. Assume: n_c = total number correctly classified, p_g = prior probability of membership for group g, N_g = number of cases for group g, N = total sample. Now, tau (τ) is defined as equation-14.

$$\tau = \frac{n_c - \sum_{g=1}^{G} p_g N_g}{N - \sum_{g=1}^{G} p_g N_g} \tag{14}$$

4.6.4 T-Test for Equal Size Groups in Dependent Variable

If both of the groups are equal in size, a t test can be conducted to test whether the hit ratio is higher than the chance as used by Altman (1968) like equation (15). where, the null hypothesis (H_0) is the models hit ratio is not higher than the chance ratio, P is equal to the proportion of the correctly predicted cases and the degree of freedom (df) is total sample size (N) minus 2: N-2. The t values for original analysis sample (t = 4.71), cross-validated (t = 2.96), and holdout sample (t = 2.29) are statistically significant. So, discriminant analysis produced significant higher accuracy in forecasting expected status of the faculty post candidate.

$$t = \frac{P - 0.5}{\sqrt{\frac{0.5(1-0.5)}{N}}} \approx t(N-2) \tag{15}$$

4.6.5 Press's Q for Uneven Sample in Groups

Press's Q (Press & Wilson, 1978) is also used to compare the accuracy rate in the classification by the estimated function with the random chance rate. In order to estimation and use the discriminant function, Press's Q should be extremely high compared to Chi Squared value at 1 degree of freedom which is 6.63. However, decision maker should be very careful when using Press's Q ratio for decision making if the sample size is small. The decision making will be misleading, if the sample size is small. The Press's Q ratio is not large enough (5.33) compared with table value for the holdout sample of this study although the hit ratio (83%) is significant. So, the estimation of the discriminant analysis and the use of the function in faculty selection are not justified using Press's Q ratio. But estimation and using discriminant function is justified using hit ratio and chance criterion. This problem will not exist if large sample is used as the holdout sample and the analysis sample in the study. Assume, N = sample Size, n = number of observations classified correctly, K= number of groups in the dependent variable. Hence, Press's Q is defined as equation-16.

$$Press's \ Q = \frac{[N - (nK)]^2}{N(K-1)} \approx \mathcal{X}^2(1) \tag{16}$$

4.6.6 Histogram of Discriminant Scores

Another available tool in hand to check goodness of fit of the model is histogram of the discriminant scores of the analysis sample and holdout sample. Figure-2 shows the histogram of the discriminant scores of the analysis sample. The overlap area is very small hence estimating and using discriminant analysis are justified. By including more variables, and larger sample size,a better function can be estimated and used in decision making.

4.6.7 Casewise Plots of the Predictors

The casewise histograms of the predictors can be drawn to check the normality of the variable. The results show that all of the variables are approximately normally distributed. In addition, descriptive statistics- mean, median, mode, skewness and kurtosis is used to check the uni-variate normality of the variables. For a normally distributed variable, mean = median = mode and skewness = kurtosis = 0. If the variables are found to follow uni-variate normal, the estimation is worthwhile. In this case, some variables are found not to follow uni-variate normal distribution. We did not skip the cases liable for this problem because of limited sample size problem.

4.6.8 Eigen Value, Canonical Correlation, and Wilk's Lambda

Eigen value is an important criterion to predict validity of the model. Theoretically, eigen value is a ratio of between group sum of squares to within group sum of squares. The eigen value 0 means that the discriminant function has no discriminatory power. The higher eigen value means the higher accurate forecasting power of the discriminant function. The minimum acceptable Eigen value is 1.00. In this research the Eigen value is 1.073, the canonical correlation-r is 0.72 and the Wilk's lambda is 0.482. All of the statistics ensure reasonable acceptance of the estimated function. To recapitulate, by increasing sample size and more variables, HR manager can increase the validity of the function.

5. Conclusion

This study estimates a two group discriminant function in order to determine efficient faculty who will stay in the organization for the long run and consequently to reduce human resource management costs: recruitments, training & development, and managing a human resource costs etc. The function is significant at a one per cent level of significance and the function can explain 50 per cent of the variations in the dependent variable. A better function can be estimated by using more variables and larger sample size in the analysis sample. For example, the variables- like job average (average job duration in the last there jobs), gender, race, and time variable included in the study of Walker (1998) and socio economic standing variable suggested by Reiss (1961) can be included in the study. For the purpose, head of the human research department can take responsibility of the functions. Thus, productivity of the human research management could be increased.

Acknowledgement

References

Aktan, S. (2011). *Early Warning System for Bankruptcy: Bankruptcy Prediction.* PhD Thesis, Faculty of Economics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany. Retrieved from http://digbib.ubka.uni-karlsruhe.de/volltexte/documents/2001113

Altman, E. I. (1968, September). Financial Ratios, Discriminant Analysis and Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 589–609. http://dx.doi.org/10.2307/2978933

Awh, R. Y., & Waters, D. (1974). A Discriminant Analysis of Economic, Demographic and Attitudinal Characteristics of Bank Charge-Card Holders: A Case Study. *The Journal of Finance*, 29(3), 973–980. http://dx.doi.org/10.2307/2978604

Billikopf, G. E. (2003). *Labor Management in Agriculture: Cutting Personal Productivity* (2nd ed.). Retrieved from http://nature.berkeley.edu/ucce50/ag-labor/7labor/16.pdf

Boyd, H. W. Jr., Westfall, R., & Stasch, S. F. (2005). *Marketing Research: Test and Cases* (7th ed.). Richard D. Irwin, Inc. Homewood, Illinois.

Burns, R. B., & Burns, R. A. (2008). Business Research Methods and Statistics Using SPSS. Sage Publications. Retrieved from http://www.uk.sagepub.com/burns/default.htm

Capon, N. (1982, Spring). Credit Scoring Systems: A Critical Analysis. *Journal of Marketing*, 46, 82–91. http://dx.doi.org/10.2307/3203343

Chancharat, N. (2008). *An Empirical Analysis of Financial Distressed Australian Companies: The Application of Survival Analysis.* Ph.D Thesis, School of Accounting and Finance, University of Wollongon. Retrieved from http://ro.uow.edu.au/cgi/theses/401

Credit Card Redlining. (1979). Hearings Before the Subcommittee on Consumer Affairs of the Committee on Banking, Housing and Urban Affairs, United States Senates, 96th Congress, Fist Session, on S15, June 4 & 5, 1979, Washington DC, U.S. Government Printing Office. pp. 183–184.

Davis, R. H., Edelman, D. B., & Gammerman, A. J. (1992). Machine-Learning Algorithms for Credit Applications. *IMA Journal of Management Mathematics*, 4, 43–51. http://dx.doi.org/10.1093/imaman/4.1.43

Dinh, T. H. T., & Kleimeier, S. (2007). A Credit Scoring Model for Vietnam's Retail Banking Market. *International Review of Financial Analysis*, 16(5), 571–495. http://dx.doi.org/10.1016/j.irfa.2007.06.001

Fern, E. F., Avila, R. A., & Grewal, D. G. (1989). Sales-force Turnover: Those Who Left and Those Who Stayed. *Industrial Marketing Management,* 1–9.

Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics, 7*, 179–188. http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x

Futrell, C. M., & Parasuraman, A. (1984, Fall). The Relationship of Satisfaction and Performance to Sales Force Turnover. *Journal of Marketing, 48*, 33–40.

George, D., & Mallery, P. (2006). *SPSS for Windows Step by Step: A Simple Guide and Reference, 13.0 Update* (6th ed. pp. 278–292). Pearson Education.

Glen, J. J. (2001). Classification Accuracy in Discriminant Analysis: A Mixed Integer Programming Approach.

The Journal of Operational Research Society, *52*(3), 328. http://dx.doi.org/10.1057/palgrave.jors.2601085

Grablowsky, J. B. (1975). A Behavioral Risk in Consumer Credit. *The Journal of Finance*, *30*(3), 915–916. http://dx.doi.org/10.2307/2326880

Graham, J. M. (2001, January). *Problems with Step-wise Procedures in Discriminant Analysis.* Paper Presented at the Annual meeting of the Southwest Educational Research Association, New Orleans, LA. Retrieved from http://myweb.facstaff.wwu.edu/~graham7/articles/SERAStepwiseDA.pdf

Hand, D. J., & Henley, W. E. (1997). Statistical Classification Methods in Consumer Credit Scoring: A Review. *The Open University*, MK 7 6AA, UK. http://dx.doi.org/10.1111/j.1467-985X.1997.00078.x

Jensen, D. L., & Bailey, A. D. Jr. (1975, July). Discriminant Analysis as an Aid to Employee Selection: A Comment. *The Accounting Review, 50*(3), 588–592.

Joseph, F., Hair, Jr., William, C. B., Barry, J. B., & Ralph, E. A. (2010). *Multivariate Data Analysis with Readings* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Lucas, G. H., Parasuraman, A., Davis, R. A., & Enis, B. M. (1987, July). An Empirical Study of Sales force Turnover. *Journal of Marketing*, *51*, 34–59.

Lusk, E. J. (1972, July). Discriminant Analysis as Applied to the Resource Allocation Decision. *The Accounting Review*, 567–575.

Mahalanobis, P. C. (1936, January 4). On the Generalized Distance in Statistics. *Proceedings of the National Institute of Sciences of India, 2*(1), 49–55.

Malhotra, N. K., & Dash, S. (2011). *Marketing Research: An Applied Orientation* (6th ed., pp. 552–582). Prentice Hall.

Miodonski, B. (2004, October). Retaining Good Employees Starts at the Top. *Contractor, 51*(10), 7–8.

Moncrief, W. C. III., Hoverstad, R., & Lucas, G. H. (1989, Summer). Survival Analysis: A New Approach to Analyzing Sales Force Retention. *The Journal of Personal Selling and Sales Management*, *9*(2), 19–30.

Reiss, A. J. Jr. (1961). Socio-economic Index for Occupations in the Detailed Classification. *Occupations and Social Status,* 114–138. New York: The Free Press.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using Multivariate Statistics.* New York: Harper Collins College Publishers.

Press, S. J., & Wilson, S. (1978). Choosing Between Logistic Regression and Discriminant Analysis. *Journal of the American Statistical Association, 73*, 699–705.

Uddin, N. (2013). Consumer Credit Customers' Financial Distress Prediction by Using Two-Group Discriminant Analysis: A Case Study. *International Journal of Economics and Finance, 5*(6), 55–66. http://dx.doi.org/10.5539/ijef.v5n6p55

Walker, K. D. (1998, August). *A Survival Analysis Approach to Employee Turnover: Its Application and Advantages.* Ph. D Thesis. Graduate Faculty of Texas Tech University. Retrieved from http://repositories.tdl.org/ttu-ir/bitstream/handle/2346/19664/31295013268429.pdf?sequence=1

Wekesa, O. A., Samuel, M., & Peter, M. (2012). Modelling Credit Risk for Personal Loans Using Product-Limit Estimator. *International Journal of Financial Research*, *3*(1), 1–22. http://dx.doi.org/10.5430/ijfr.v3n1p22

Welker, R. B. (1974, July). Discriminant Analysis as an Aid to Employee Selection. *The Accounting Review,* 514–523.

Wiginton, J. C. (1980). A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior. *Journal of Financial and Quantitative Analysis*, *15*(3), 757–770. http://dx.doi.org/10.2307/2330408