المجلة الإلكترونية الشاملة متعددة التخصصات
العدد الثامن والعشرون شهر(9) 2020

# Application of Discriminant Analysis Techniques to Study the Absenteeism of Work phenomenon for Blue Nile Cigarette Company Workers

**Majda Mohamed**[1], ***Abdelaziz G. M. Musa***[2] ***Fath ELrhman Elsmih***[3] ***Salem alzahrani*** [4]

1-Assistant Professor at Alneelain University - Faculty of mathematical sciences and statistics
E-mail: majdamm2010@gmail.com

2- Assistant Professor at Alneelain University - Faculty of mathematical sciences and statistics
E-mail: azizgibreel@yahoo.com
3- Associate Professor at Al-Baha University, Faculty of Science & Arts baljurshi, dep of Math. KSA
E-mail: fathyelrhman@gmail.com
4- Assistant Professor at Al-Baha University, Faculty of Science & Arts, dep of Math, Al-Mandag, KSA
E-mail: Salem.bb@hotmail.com

## Abstract:

This paper aims at using linear discriminant analysis to study the relationship between the absenteeism of work as dependent variable and type, age, social situation and transportation status as independent variables. Discriminant analysis technique was applied to primary data collected through a questionnaire with related questions to absenteeism of work phenomena distributed to population consist of 69 workers in Blue Nile Cigarette Company. The study presented a discrimination function used for discrimination of any new worker to one of the two discriminations "absent" or "not absent" as well as constructed visual basic program to help in classification of any new cases using specific discrimination function.

## Keywords:
Discriminant Analysis, Blue Nile Cigarette Company, Absenteeism of Work Phenomenon.

# 1- Introduction:

Discriminant analysis is a multivariate statistical technique used to analyze the data. The objective of discriminant analysis is to develop discriminant functions that are nothing but the linear combination of independent variables that will discriminate between the categories of the dependent variable in a perfect manner. It enables us to examine whether significant differences exist among the groups, in terms of the predictor variables. It also evaluates the accuracy of the classification. Discriminant analysis assumes that the data (for the variables) represent a sample from a multivariate normal distribution, the variance/covariance matrices of variables are homogeneous across groups, Correlations between means and variances and the variables that are used to discriminate between groups are not completely redundant.

# 2- The problem and objective:

An increasing number of workers become a great challenge in Blue Nile Cigarette Company; therefore, studying the absenteeism of work phenomenon in companies becomes an important issue in these companies. The importance of this study is to determine the factors that affect absenteeism of work phenomenon in Blue Nile Cigarette Company that helps to tackle such a problem. This study aims to investigate whether there is statistically significance between the variables. Also, to construct a new discriminant function that helps in the prediction of a new cases status concerning the absenteeism of work phenomenon. Moreover, to find a logistic regression function that link between the absenteeism of work and its different related factors by finding the probability of absenteeism for a new worker.

# 3- The hypothesis:

This study hypothesizes that; there is a statistically significant relationship between absenteeism of work and worker type, age, social status and transportation status in Blue Nile Cigarette Company. Also the discrimination function helps in the prediction and classification of the absenteeism of work according different related factors.

# 4- Data:

Data used in the analysis of this paper is primary data collected through a questionnaire distributed to Blue Nile Cigarette Company workers contains the related questions to absenteeism of work phenomena. The data collected through a census method from population of size 69 workers.

# Linear Discriminant Function:

Linear discriminant analysis is a method for identifying the classification of individuals based on a series of explanatory variables, for instance, suppose the researcher wanted to know how height and weight contribute to the classification of males and females Linear discriminant analysis does this by producing a series of $k-1$ discriminants where $k$ is the number of groups, linear discriminant analysis sometimes called MANOVA The number of linear discriminant functions is equal to the number of levels minus 1, $(k-1)$.

Sometimes the researcher wants to classify a unit for one of for instance $k$ populations with the vector of observations$\underline{x} = [x_1, x_2, \dots, x_p]$, in this case there are $k$ populations and the aim is to construct a formula or rule helps in classifying the new cases with a minimum loss of wrong classification. The following assumptions are required:

**i-** Probability density functions $p_1(\underline{x}), p_2(\underline{x}), \dots, p_k(\underline{x})$ where $\underline{x}$ is a vector of observations and $p_i(\underline{x})$ is the probability density function of population $i$ of observation $\underline{x}$.

ii- The probabilities of drawing $1,2, \dots, k$ (the relative frequencies in units of the study) $\pi_1, \pi_2, \dots, \pi_k$.

iii- Loss functions $r_{ij}$ which give the loss of classification of unit belong to population $j$ instead of the population$i$.

The discriminate score for any person has a observed $\underline{x}$ unit belong to the population $i$ computes from the following expression:

$$S_i = -[\pi_1 p_1(\underline{x}) r_{1i} + \pi_2 p_2(\underline{x}) r_{2i} + \cdots + \pi_k p_k(\underline{x}) r_{ki}] \quad (1)$$

Where: $i = 1,2, \dots, k$,

$\pi_1 p_1(\underline{x}) r_{1i}$the product of loss classification of a unit belong to the population $i$ by the probability of a unit belong to the population $i$ and it has an observation $\underline{x}$ . it is possible to proof that if a unit could be classified to population has a maximum discriminate score, then the expected loss classification can be minimize. In most cases the researchers un known the wrong loss classification, in such case a criteria of minimizing frequency (probability) of minimizing wrong classification can be used. The optimum criterion in such case is to classify the unit which has an observation $\underline{x}$ to the population which has a most probability that the observation is unit in that population, which means that the most classification is the classifications to the population $i$ *is expressed as follows:*

$$S_i = \pi_i p_i(\underline{x}) \quad (2)$$

If $\underline{x}$ in the population $i$ distributed as a multivariate normal distribution with the

probability density function:

$$p_i(\underline{x}) = \frac{1}{((2\pi)^{-p/2}\ |\Sigma_i|^{-1})}\ exp(-\frac{1}{2}\left(\underline{x} - \underline{\mu_i}\right)'\Sigma_i\left(\underline{x} - \underline{\mu_i}\right) \quad (3)$$

Where: $i = 1,2, \dots, k$

Then logarithm $\pi_i p_i(\underline{x})$ after dropping $(2\pi)^{-p/2}$ term and substituting with respect to $p_i(\underline{x})$ in multivariate normal distribution, then the discriminate score of the population $i$ becomes:

$$S_i = -\frac{1}{2}\log_e|\Sigma_i| - \frac{1}{2}\left(\underline{x} - \underline{\mu_i}\right)'\Sigma_i^{-1}\left(\underline{x} - \underline{\mu_i}\right) + \log_e \pi_i \tag{4}$$

Where: $\underline{\mu_i}$ and $\Sigma_i$ are mean vector and covariance matrix of $\underline{x}$ population. A unit can be classifies to the population which has a maximum discriminate score (Max $S_i$ ). The above equation is a quadratics discriminate score. If all populations has an equal covariance matrix such that $\Sigma_i = \sum$ for all $i$ , then the value of:

$\underline{x}' \sum^{-1} \underline{x} - log|\sum| - \frac{1}{2}$ can be the same for each $S_i$ and therefore it can be drops to gives:

$$S_i = (\underline{\mu_i} \Sigma^{-1})\ \underline{x} - \frac{1}{2}\underline{\mu_i'} \Sigma^{-1} \underline{\mu_i} + \log \pi_i \tag{5}$$

which is linear discriminate score. If $\underline{\mu_i}$ and $\sum$ are unknown the estimates $\underline{\bar{x}}$ and $\hat{\sum}$ are used instead. In case of two populations there is one comparison, and the decision can be taken by computing the difference such as:

$$S_1 - S_2 = L(\underline{x}) - C \tag{6}$$

where

$$L(\underline{x}) = \underline{\mu_1'} - \sum^{-1} \underline{x}\, \underline{\mu_2'} \tag{7}$$

$$C = \frac{1}{2}\left(\underline{\mu_1'} \sum^{-1} \underline{\mu_1'} - \underline{\mu_2'} \sum^{-1} \underline{\mu_2'}\right) \log \pi_2 - \log \pi_1 \tag{8}$$

The decision is to:

Classify to population 1 if $L = (\underline{x}) \geq C$, otherwise classify to population 2 if $L = (\underline{x}) \leq C$.

## 5- Results and Discussion:

This section provides empirical analysis results of applying discriminant analysis function to study the relationship between the absenteeism of work as dependent variable and type, age, social situation and transportation status as independent variables in Blue Nile Cigarette Company. The study presents the application of linear discriminant function to the two groups of data representing non absent and absent workers in Blue Nile Cigarette**.** Table (1) bellow reports the standardized canonical discriminant function coefficients.

Table (1) Discriminant Function Coefficients

**Standardized Canonical Discriminant Function Coefficients**

| | Function |
|---|---|
| | 1 |
| Type | 0.941 |
| Social Status | 0.505 |
| Transportation Status | -0.106 |
| 1st Age Group (Less than 25) | 0.058 |
| Second Age Group (25-35) | 0.272 |
| Third Age Group (35-50) | -0.119 |

Table (1) above reports the linear discriminant function coefficient, the estimated discriminant function model is written as:

$$S_i = 0.941x_{1i} + 0.505x_{2i} - 0.106x_{3i} + 0.058x_{41i} + 0.272x_{42i} - 0.119x_{43i} \qquad (9)$$

Where: the counter I indicate to one of the two groups "0"and"1". Table (15) bellow show functions at group centroids which means the average discriminant function in the discriminant analysis of the two groups "0"and" . It explores that these two groups lie differently from each other.

Table (2) Functions at Group Centroids

**Functions at Group Centroids**

| Absence | Function |
|---|---|
| | 1 |
| Not Absent | -0.121 |
| Absent | 0.243 |

From the above table the average is: $\frac{(0.243-0.121)}{2} = 0.061$ which means that if the researcher wants to know new worker in terms of absence or not absence, the researcher must determine new worker type, age, social status and transportation status and substituting these values in the estimated discriminant function model ($S_i$ $equation$). If the substituting result of $S_i$ is less than the average (0.061) the result indicates that this worker discriminate as non absent worker and if the substituting result of $S_i$ is greater than the average (0.061) the result indicates that this worker discriminate as an absent worker. Table (3) bellow shows the equality of two matrix variation of population test which is assumption used in computing discriminant function.

Table (3) Results of equal variation of population Test the variations of the two matrix groups

| | | |
|---|---|---|
| Box's M | | 23.713 |
| F | Approx. | 1.000 |
| | df1 | 21 |
| | df2 | 7453.501 |
| | Sig. | .459 |

According to Box-M test the sig value is 0.459 which is greater than 0.05 significant level; these results conclude that there is no statistically significant test among the variation of the two matrix groups. Table (4) indicates Classification Results report the classification result of discriminant function quality.

Table (4) Classification Results

**Classification Results (a)**

| | | Absence | Predicted Group Membership | | Total |
|---|---|---|---|---|---|
| | | | Not Absent | Absent | Not Absent |
| Original | Count | Not Absent | 45 | 1 | 46 |
| | | Absent | 21 | 2 | 23 |
| | % | Not Absent | 97.8 | 2.2 | 100.0 |
| | | Absent | 91.3 | 8.7 | 100.0 |

The classification test result gave a great idea about forecasting accuracy of discrimination function. The result shown that 97%of the first group (not absent) were discriminated correctly however, 2.2% were wrongly discriminated.

A visual basic program also was designed to help the Blue Nile Cigarette Company to enter a new worker data in order to predict his future status relates to absent or not absent. Suitable tools and the required commands to construct a program for computing the logistic function has been done. Figure (1) bellow shows the discriminant analysis visual basic program constructed to classify and predict the new cases status relates to absent or not absent in Blue Nile Cigarette Company.

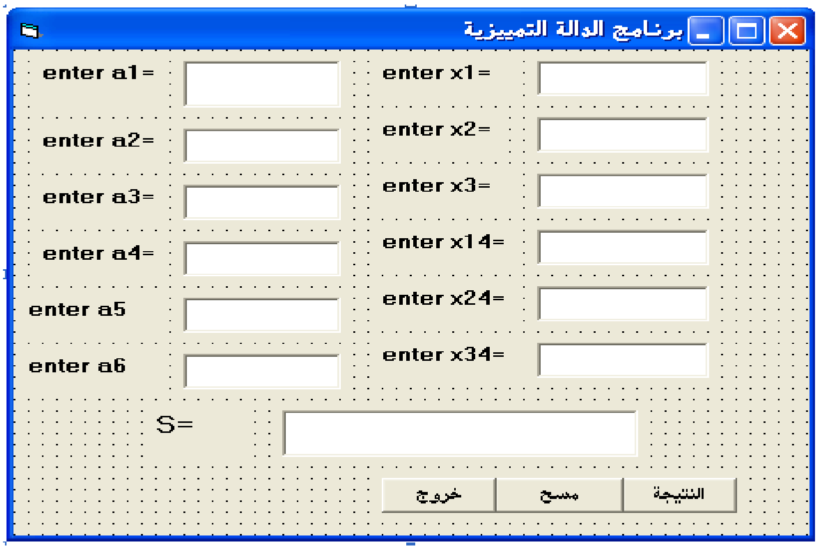


Figure (1) The visual basic logistic regression program

In the visual basic discriminant analysis program, the following values $a_1 = 0.941$ , $a_2 = 0.505$ , $a_3 = -0.106$ , $a_4 = 0.058$ , $a_5 = 0.727$ *and* $a_6 = -0.0119$ , $x_1 = 1, x_2 = 1, x_3 = 1, x_{41} = 1, x_{42} = 1, x_{43}$ where $a_1$ , $a_2$ , $a_3$ , $a_4$ , $a_5$ *and* $a_6$ represents the coefficient values after the analysis of discriminant function and $x_1 = type\ (female), x_2 = single\ merital\ status, x_3 = dificult\ transportation\ status, x_{41} = age\ class\ one\ less\ than\ 25\ years, x_{42} = ahe\ class\ two\ (25-35),$ $x_{43}\ \ age\ class\ tree\ (35-50)$.


Figure (2) The visual basic discriminant analysis program

The above values are entered in all corresponding text book, the final result was $S = 1.551$ which means that the worker absent because the $S$ value is greater than the average discriminant value (0.061) which appear in table (2) above.

## 6- Conclusions and future limitations:

This study investigates the absenteeism of work phenomenon in Blue Nile Cigarette Company col.td. A questionnaire distributed to 69 workers in Blue Nile Cigarette Company col.td consist of questions related to the absenteeism phenomena. The empirical analysis results indicate that there is no statistically significant relationship between the absenteeism of work and type, age, social situation and transportation status in the company. The paper also presented a discrimination function used for discrimination of new worker to one of the two discriminations "absent" or "not absent" as well as program helps in application discrimination function using the study variables. The study also recommends the following suggestions:

a- Taken into account the advantages of discriminant analysis in modeling dependent variables, it's highly interpretation powerful as well as its simple requirements because of ignoring some of general regression model assumptions.

b- Expanding the uses discriminant analysis in economic and social studies and not using it only in medical and education sciences.

.

## declaration of competing interest

The authors declare that they have no known competing financial Interests or personal relationships that could have appeared to influence the work reported in this paper

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References:

1- George, H. (1984). Introduction to multivariate analysis. *Thousand Oaks, CA: Sage Publications.*

2- Klecka, R. (1980). Discriminant Analysis. Quantitative Applications in the Social Sciences Series, No. 19. *Thousand Oaks, CA: Sage Publications*.

3- Lachenbruch, P. (1975). Discriminant Analysis. NY: Hafner. For detailed notes on computations.

4- Morrison, F. (1967). Multivariate Statistical Methods. *McGraw-Hill. New York*.

5- Overall, E. and Klett, C. (1972). Applied Multivariate Analysis. McGraw-Hill: New York.

6- Joseph O. Onu and Onyedikachi O. John. An Application of Discriminant Analysis to the Classification of Students on the Basis of their Academic Performances. International Journal of Applied Science and Mathematics Volume 3, Issue 5, ISSN (Online): 2394-2894
7- Pradeep Singh and Sandeep Kumar. Application of Discriminant Analysis in Examining Influence of Television Advertisement on Women's Buying Behavior. International Journal of Research in Economics and Social Sciences (IJRESS). Vol. 8 Issue 2, February- 2018
8- Carlo Ricciardi etc. Linear discriminant analysis and principal component analysis to predict coronary artery disease. Health Informatics Journal 2020, Vol. 26(3) 2181–2192
9- Stephen J. etc .The determinants of absenteeism: evidence from Australian blue–collar employees .Journal The International Journal of Human Resource Management Volume 6, 1995 - Issue 4.
10- Hakimzadi Wagan. Etc. Corporate Bankruptcy Prediction In Pakistan By Employing Multiple Discriminant Analysis Technique. Developing Country Studies. Vol.6, No.2, 2016
11- Wurim Ben Pam. (2013) Discriminant Analysis and the Prediction of Corporate Bankruptcy in the Banking Sector of Nigeria, International Journal of Finance and Accounting, Vol. 2 No.6, pp. 319-325.
12- Sengur A. An expert system based on linear discriminant analysis and adaptive neuro-fuzzy inference system to diagnosis heart valve diseases. Expert Syst Appl 2008; 35(1–2): 214–222
13- Press, J. and Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association.73: 699-705.
14- Baudat, G. and Anouar, F. (2000) Generalized Discriminant Analysisusing a Kernel Approach. Neural Computation, 12: 2385–2404.
15- Nicholas Pindar Dibal, Christopher Akas Abraham. On the Application of Linear Discriminant Function to Evaluate Data on Diabetic Patients at the University of Port Harcourt Teaching Hospital, Rivers, Nigeria. American Journal of Theoretical and Applied Statistics. 2020; 9(3): 53-56
16- Abdullah, N. A. H., Halim, A., Ahmad, H., and Rus, R. Predicting corporate failure of Malaysia's listed companies: Comparing multiple discriminant analysis, logistic regression and the hazard Model. International Research Journal of Finance and Economics, 15, 201-21
17- Alayande, S. A. and Bashi, K. A. (2015). An Overview and Application of Discriminant Analysis in Data Analysis. IOSR Journal of Mathematics (IOSR-JM). 11 (1), 12-15. DOI: 10.9790/5728-11151215.
18- Erimafa, J. T. Application of Discriminant Analysis to Predict the Class of Degree for Graduating Students in a University System. International journal of physical science, 4 (1), 16 – 21.

19- Hafez, E. I., Abdel-Fatah, E. M., Abdel-Nabi, S. M. and Zeidan, A. S. A. Discriminant Analysis in View of Statistical and Operations Research Techniques. Journal of Multidisciplinary Engineering Science and Technology (JMEST). 2 (11), 3039-3047
20- Mbanasor J. A. and Nto, P. O. O. (2008). Discriminant Analysis of Livestock Farmers' Credit Worthiness Potentials under Rural Banking Scheme in Abia State, Nigeria. Nigerian Agricultural Journal, 39 (1), 1-7.
21- Micheal, A. B. (2014). Application of Discrimination andClassification on Diabetes Mellitus Data. International Journal of Applied Science and Technology, 4 (6). 292-298