International Journal on Advanced Science Engineering Information Technology

Vol.10 (2020) No. 5
ISSN: 2088-5334

The Discriminant Analysis in the Evaluation of Cancers Diseases in Iraq

Nabaa Naeem Mahdi[a,1], Haifa Taha Abed[a], Nazik J. Sadik[a]

[a] *College of Management and Economics, University of Mustansiriyah, Iraq*
Email: [1]nabaanaeemmahdi@uomustansiriyah.edu.iq

Abstract— **Cancer diseases are considered one of the most critical problems facing the world's countries, especially the State of Iraq. Many local and international reports indicated that the weapons used in wars and the accompanying nuclear and chemical radiation are among the most prominent reasons for the spread of cancerous diseases in Iraq. This study found that Gender has the highest discriminating power, whereas the Grade variable has the least discriminatory power. Similarly, Behavior has the highest discriminatory power, whereas the Government has the least biased power. It became clear that the third group (those with breast cancer) had the highest probability of the correct classification. The probability of correct classification reached 92%, followed by the second group with brain cancer, where the probability of correct classification was 64%. Finally, the first group with bladder cancer had the lowest probability of correct classification. We conclude that increasing the sample size has a significant impact on the correct classification of observations. The effects of these weapons were tremendously harmful to public health and the environment. Its effect persisted after many years, so three groups of cancer patients (bladder, brain, and breast cancer) were analyzed from 2012 to 2017 using a statistical method to analyze multivariate data. The results showed gender and the nature of the tumor (Behavior) have the highest discriminating power. The results were entirely satisfactory, as the discriminatory predictive capacity obtained a level of success of 72.2%.**

Keywords— **Iraq; discriminant linear function; ROC curve; wilks lambda.**

I. INTRODUCTION

Discriminant Analysis is a multivariate method used to learn pattern recognition and machine learning to find a linear combination of independent quantitative variables. The discriminant analysis distinguishes two or more events by the dependent variable taxonomic, which is a way to classify the observation in two or more groups. It is subject to predict group membership based on a combination of linear quantity where built. This technique of adopting observations with the values of information and with well-known groups and formation model variables allows the prediction of group membership and knowledge independent quantitative variables [1], [2]. The second purpose is to understand the data set closer examination of the model forecasting and the relationship between the membership of the groups and independent variables [3].

The Discriminant Analysis is like the analysis of variance (ANOVA) and regression analysis (RA) that reflect the dependent variable combination and the independent variable. However, the dependent variable is required to be quantified. The study of the discriminatory is the dependent variable taxonomically. This analysis's approach is firmly from the logistic regression and probability regression because they interpret a particular variable classificatory. However, the difference among them is that the logistic regression and regression probability do not assume that the independent variables are distributed normal distribution. In contrast, analysis of the discriminatory assumes that the spread data Distribution naturally and is to impose a basic analysis of the hypotheses discriminatory [4], [5].

Similarly, each of the Principal Component Analysis (PCA) and the factor analysis are all looking for combinations of linear variables that better interpret the data. Discriminant analysis works on the modeling of the difference between the categories of data. In contrast, the factor analysis does not compare the categories where it built combinations of variables based on differences rather than similarities. The second difference is that it does not support the Interdependence technique, distinguishing between independent variables and the dependent variable [6], [7]. Discriminant analysis is applied when independent variables take measurements of continuous quantity, but we use Discriminant Correspondence Analysis when dealing with independent taxonomic [8], [9].

Cancer is considered one of the worlds' biggest causes of mortality, but the chances for its healing appear to strengthen many of its forms [10]. This is because of the approaches and recovery services for early detection. It may be said that cancer is a diagnostic concept that encompasses a wide variety of diseases characterized by the development without censorship of defective cells that are separated that have the potential to infiltrate that kill healthy tissues in the body. The disorder can spread across the body and is often referred to as malignancies, malignant tumors, and cysts. It is named the metastasis syndrome, which is the most severe cause of death from cancer metastases. [12], [13].

Cancer develops due to disruption (change or mutation) in a series of RNA demineralized oxygen (Deoxyribonucleic Acid-DNA). A series of DNA in the human body includes a collection of several commands to identify the body's cells, how to expand and mature and differentiate healthy cells that often appear to make DNA modifications. However, it remains able to correct the bulk of these modifications [14].

The accumulation of these cells generates a cancerous tumor in some types of cancer. Still, not all cancer types produce cancerous tumors (e.g., leukemia Dam- leukemia), a variety of cancer that affects bone marrow blood cells (bone marrow) spleen. [15]. The first is the genetic deviation is only the beginning of the process of the development of the disease; the researchers believe that the development of the disease requires several changes in the cell include

- Entrepreneurial factors are leading to obtaining a genetic change.
- The assistant factor for the growth of cells quickly.
- An encouraging factor that makes the disease more aggressive and helps him to the outbreak.

II. Materials and Method

In this respect, we describe the concept of discriminant analysis, types, and types of discriminant analysis models, and the test of significant choice models.

A. Concept of Analysis Discriminant

In the case of this method, variables in the model are analyzed in an interdependent way. Considering the nesting relationship between these variables, it attempts to form a relationship described by a statistical model, the interaction between different variables highlights its effectiveness. By using the differences between many variables, we can find a group of linear combinations. Variables are called discriminant variables.

The discriminatory analysis model depends on access to discriminant function, which works to maximize the differences between the averages of groups and reduce the similarity of classification errors simultaneously; this is by finding linear combinations of a set of variables. The classification process is the subsequent operation of the process formation discriminant function. They rely on this function in the classification of the new observation and belonging to any group.

The types of discriminatory analysis are as follows:

- Direct discriminatory analysis: all variables are entered once into the analysis.
- Hierarchical discriminatory analysis: In this type, the variables are entered according to a schedule organized by the researcher.
- Stepwise discriminatory analysis: a statistical standard is defined that defines the priority of the variables to be entered into the model.

Discriminatory analysis requires analysis of discriminatory that societies be the subject of study separate and identifiable even that these overlapping communities, among them a certain degree, and that each individual is in every subject community description and selection group of any measures to be independent, in addition to the communities under study data used to vary their midst, and that the analysis is random so that these samples are representative of the communities under study.

Some types of discriminant analysis models are as follows:

- Discriminatory linear analysis in the case of two or more than two groups.
- Analysis of discriminatory non-linear and which is used in the case of the heterogeneity of variance.

B. Linear Discriminant Function-Two Groups

This function is called the Fisher Function and considers nonparametric, where data distribution is not assumed naturally occurring. It can formulate a model function discrimination based on sample indicators that selected the vocabulary developed in two groups. This function can be tested individually and determine its affiliation with any group. If we assume that the area of the sample is K, which will be divided into two parts (Z), which dates to the first group and (K-Z) goes back to the second group while the boundary between the two groups could be due to any combination of these two groups.

To impose our variables (X1, X2…Xp), the general formula special function according to the following formula:

$$Y_{i=} a_1X_{1\,+}\, a_2X_2 \qquad (1)$$

Whereas:

Y_i: Represent Response variable.

X_i: Represent P variables (attributes).

To evaluate the differentiation between the two groups, the means between the two groups need to be derived, and that the process of estimation parameters which makes the function gives a better distinction between the two groups it must be done by making the square differences between the mean of two groups to covariance greatest of any two groups according to the following formula:

$$Q = (Y_2\text{- } Y_2)^2/\ (Y_{ij}\text{-}Y_i)^2 \qquad (2)$$

Where we appreciate the discriminant function parameters by maximizing the ratio Q through partial derivation and equality to zero, we get:

$$a= S^{-1}(X_1\text{- } X_2) \qquad (3)$$

Then, the stage of classification the observation in one of the two groups depending on the middle point of the two groups let (L), which makes the possibility of the wrong classification less than what can be, according to the following formula:

$$L = Y_2\text{- } Y_2 / 2 \qquad (4)$$

And classified seen to the first group if $\hat{Y} > L$ it is classified as a second group if $\hat{Y} < L$ they are classified as seen at random to the first or second group if $\hat{Y} = L$ they were:

$$Y= (X_1- X_2)^{'} S^{-1}X \quad (5)$$

To test the significant characteristic function that can test the hypothesis, which states equal to the averages of the two groups are:

$$D^2= (X_1- X_2)^{'} S^{-1} X_1- X_2 \quad (6)$$

Where D^2 represents (Mahalanobis Distance) to test the significant function discriminatory use F test and degree of freedom according to the following formula:

$$F= (n_1+n_2-p-1)/ (n_1+n_2-2) * T^2 \quad (7)$$

It can also be used scale (Wilks-Criteria) according to the following formula:

$$A= |W| /|T| \quad (8)$$

Where: T is a matrix, variance-covariance total groups, and W variance-covariance matrix within groups. Range A is between the value of (0, 1). If they are close or equal to the one, it indicates that the averages are equal, this indicates that there is no differentiation between the groups. The value is close to zero is indicated by the power of discrimination between the groups. The use of $_X{}^2$ can be more accurate than the scale Wilks, as presented in the following formula:

$$X^2 = NLog(A) \quad (9)$$

With a degree of freedom P(r-1) where P the number of variables and r the number of groups.

To calculate the probability of error classification, this is on two types:

- The probability of error classification P_{12} is the probability of observation classification to the second group and is originally dating back to the first group.
- The probability of error classification P_{21} is the probability of observation classification to the first group and is originally dating back to the second group.

Thus, it will be the probability estimate classification by the following formula:

$$P_{12}= P_{21} – (D/2) \quad (10)$$

Whereas:
D: It represents the root of Mahalanobis Distance.

C. Linear Discriminant Function More than Two Groups

The linear discriminant function can be the case of the two groups to generalization more than two groups, and the greater the number of aggregates compounded the problem of discrimination, to impose our (k) of aggregates (n_i) of observations (i = 1,2, ..., k). Each watch includes (p) of variables (t) represent a matrix variation and variation and common form:

$$T = \sum_{i=1}^{k}\sum_{i=1}^{n}(X_{ij} - \bar{X})(X_{ij} - \bar{X})' \quad (11)$$

We can calculate the variance-covariance matrix of the group (i) according to the formula:

$$W_i = \sum_{i=1}^{m}(X_i - \bar{X})(X_i - \bar{X})' \quad (12)$$

D. Receiver Operating Characteristic Curves (ROC)

It is used in medical tests and in other tests, and the reason for the designation is due to the theory (detection of the Signal); as it was developed in the Second World War, Radar images were analyzed, so the signal detection theory came to measure the ability of radar receiver operators to discover these important differences, and in 1970 the detection theory was Signal is useful for interpreting medical tests. This analysis works through a graph that shows the classification performance at each graphical point. This curve depends on two parameters:

- True Positive Rate is defined as follows:

$$TRP = TP / TP + FN \quad (13)$$

Whereas:
TPR: represent the rate of real positive results.
TP: represent the results of the real positive.
TP + FN: It represents the sum of true positive and false negative results.

- False Positive Rate (FPR) is defined as follows:

$$FRP = FN /(TP+FN) \quad (14)$$

Whereas:
FRP: Represent the rate of false-positive results.
FP: Represents false-positive results.
$FP + TP$: It represents the sum of false results and false negative results.

The goal of this curve is to predict the accuracy of the test when we have classification data. The area can illustrate it under the curve, the more likely the test's positive results are correct. The area under the curve approaches the correct one, meaning that all positive test results are correct. The relationship between Sensitivity and specificity is an inverse relationship. Still, we take Sensitivity on the y-axis, and the specificity is on the x-axis. The specificity is calculated (1-Sensitivity), so the relationship is Positive. The value of the axis is confined between (0-1), so the results are interpreted as follows: if the value is confined between (0.90-1) the result is (excellent), but if it is between (0.80-0.90) it is (good), whereas if the result is between (0.70-0.80) it is (fair) and if it is between (0.60-0.70) the result is (Finally, if it is between (0.50-0.60) then the result is (fail).

III. Results and Discussion

Our study population consists of all people infected with some form of cancer diseases during the period (2012 – 2017) and available at the Iraqi Ministry of Health database. The study included three groups of patients with cancer. The first group included patients with bladder cancer, the size of 932 infected, the second group of patients with brain cancer in 1123 the size of an infected, and the third group included with breast cancer in 1949 infected.

Where represents a variable response variable of illness of cancer has been given the number (0) for a group of people with bladder cancer and the number (1) for a group of people with brain cancer, and the number (3) injury breast cancer concerning variables, have been taken seven variables can which are described as follows:

- Gender ($x_{1)}$: 1 is a symbol for Male and 2 for females.

- Age ($x_{2)}$: rated to 0 for periods less than or equal to 50.1 greater than 50 reconstructions.
- Government ($x_{3)}$: 0 is classified for the central Governorates, 1 is for the northern provinces, and 2 for southern Governorates in Iraq.
- Occupation ($x_{4)}$: 0 is for Housewife, 1 for government employee 1, and 2 for Earner.
- Diagnosis of the disease method (Basis) x_5: 0 is classified for clinical examination and 1 for the death certificate.
- The nature of the tumor (Behavior) x_6: 0 is categorized for Hamid and 1 for malignant
- Degree of disease (Grade) x_7 is classified into first grade (1), second grade (2), third grade (3), and fourth grade (4).

A. Statistical Analysis

After the sample definition, the necessary step to apply in the analysis of the discriminatory is a variable test that feeds the model using the statistical package SPSS 20 [16]. We calculated Kolmogorov-Smirnov to verify the normal distribution of the variables. The test results for the period studied are shown in Table 2, which identifies the variables not distributed normal distribution. The number of people with each group shall be considered a relatively large distribution variables are distributed according to the central limit theorem [17].

$$H_O: \underline{\mu}_o = \underline{\mu}_1 = \underline{\mu}_2$$

TABLE I
KOLMOGOROV-SMIRNOV TEST (P-VALUES)

Variables	Kolmogorov-Smirnov		
	Statistic	df	Sig.
Gender	.422	4004	0.000
Age	.388	4004	0.000
Government	.364	4004	0.000
Occupation	.346	4004	0.000
Basis	.498	4004	0.000
Behavior	.533	4004	0.000
Grade	.217	4004	0.000

TABLE II
THE VIF FOR EACH VARIABLE

Variables	VIF
Gender	2.157
Age	1.043
Government	1.102
Occupation	2.140
Basis	1.097
Behavior	1.013
Grade	1.075

Discriminant Analysis is adversely affected by Multicollinearity, which refers to near multiple linear dependencies (i.e., high correlations) among variables in the data set. Therefore, the Variance Inflation Factor (VIF) was relied on for each variable, so VIF>10, there is an indication of the presence of Multicollinearity between X_j and the rest of the variables. Table 2 explains the result.

From Table 2, the variance inflation factor for all variables is less than 10, and this indicates the absence of the Multicollinearity problem between the variables. Another important test is a significant test difference between the averages of group under the study to test the distinction between group and the formation of discriminant functions acceptable statistically, and we use Box's test results are shown in Table 3, and according to the following hypothesis:

TABLE III
DISCRIMINATORY FUNCTIONS SIGNIFICANCE TEST

Test of Function(s)	**Wilks' Lambda**	**Chi-square**	**df**	**Sig.**
1 through 2	.403	3637.470	14	0.000
2	.840	695.173	6	.000

The results were shown in Table 3, depending on (Wilks) and χ^2, (sig=0.000) there are significant differences between the means of people with bladder cancer and those suffering from brain and breast cancer, which means that discriminatory functions can classify any individual into one of the three groups. As for the homogeneity test matrix variation of the groups studied, it has been used. It is tested Box's M equivalent to testing Bartlett's in multivariate analysis using the following hypothesis:

$$H_O: \textstyle\sum_0 = \sum_1 = \sum_2$$
$$H_1: \textstyle\sum_0 \neq \sum_1 \neq \sum_2$$

Where Σ represents a matrix of common variation and variation of each group, and Table 4 shows the test results of the homogeneity of variance:

TABLE IV
TESTING OF HOMOGENEITY OF VARIATIONS AMONG THE THREE GROUPS

Box's M		1489.043
F	Approx.	70.685
	df1	21
	df2	14442691.377
	Sig.	.000

According to Table 4, the results showed that the value of P-Value <o.o5 This means not accepting the premise of nothingness any lack of homogeneity of variance between groups. However, many researchers discussed the problem of heterogeneity of variance, and among them, Joseph [18] "asymmetric matrix variances negatively affect the classification process. If the sample sizes are small, and the covariance matrices are not equal, then the estimation process's statistical significance is negatively affected. This effect can be minimized by increasing sample size as well as by using group-specific covariance matrices for classification purposes. Still, this approach imposes mutual verification of discriminatory outcomes", Here we have the sample size 4218 and the classification is good as shown below. The equality of variance-covariance matrices of is not necessary when LDA is used [19]. However, the Framingham Heart Study data suggests that the finding might be sensitive to the assumption of normality". Thus, the importance of discriminatory analysis is its ability to separate groups appropriately. Another important test to test Box's M is to test each variable's spirits, any statement the

importance of the variable, and the extent of its impact in building the discriminatory linear function, and the results of this test are shown in Table 5.

TABLE V
TEST THE SIGNIFICANCE OF THE VARIABLES IN THE DISCRIMINANT FUNCTION

Variable	Wilks' Lambda	F	df1	df2	Sig.
Gender	.548	1652.846	2	4001	0.000
Age	.878	278.195	2	4001	.000
Government	.940	126.884	2	4001	.000
Occupation	.590	1389.161	2	4001	0.000
Basis	.991	18.186	2	4001	.000
Behavior	.966	71.012	2	4001	.000
Grade	.999	1.676	2	4001	.187

Through Table 5, it was found that all study variables were important in constructing the discriminatory function, where the value of significance in the sixth column in Table (5) was less than (0.05), while the degree of disease (grade) variable not show any significant effect. To illustrate the percentage of variation for group on the response variable we use canonical correlation box, and the results are shown in Table 6.

TABLE VI
EIGEN VALUES AND CANONICAL CORRELATIONS

Function	Eigenvalue	Variance % of	Cumulative %	Canonical Correlation
1	1.087	85.1	85.1	.722
2	0.19	14.9	100.0	.400

Table 6 shows the eigenvalues and canonical correlation values. The eigenvalue is the index of the overall model fit. The table of eigenvalues gives information about the effectiveness of the discriminant functions. Larger eigenvalues indicate that the discriminant function is more useful in distinguishing between the groups. The canonical correlation is the multiple correlations between the predictors and the discriminant function. This table shows that the values of canonical correlations of functions I and 2 are .722 and .400, respectively. Since the square of the canonical correlation indicates the percentage of variation explained by the model in the grouping variable, hence here the function I indicate 52% (=$.722^2$) and function 2 indicates 16% (=$.400^2$) of the variation in the three different groups is explained by the discriminant model.

B. Linear Discriminant Function Estimation

The classification function Method in a linear function is defined for each group. Classification is performed by calculating a score for each observation on each group's classification function and then assigning the observation to the group with the highest score. It differs from the calculation of the discriminant Z score, which is calculated for each discriminant function as shown in table 7.

TABLE VII
STANDARDIZED CANONICAL DISCRIMINANT FUNCTION COEFFICIENTS

Variables	Function	
	1	2
(Constant)	3.225	1.913
Gender	-1.785	.990
Age	.428	-1.631
Government	.265	-.023
Occupation	.659	.759
Basis	.087	.929
Behavior	-1.178	-3.634
Grade	-.021	.098

The discriminating power of the variables in the model is shown in the above table. The Standardized Canonical Discriminant Function Coefficients can be used to rank the importance of each variable. In other words, the variables having a higher magnitude of the absolute function value is more powerful in discriminating against the three groups. Since for function I, the absolute function value of the Gender is 1.785; hence this variable has the highest discriminating power, whereas the Grade has the least discriminating power as its absolute function value is .021. Similarly, for function 2 the absolute function value of the variable Behavior is 3.634 is hence this variable has the highest discriminating power, whereas the variable Government has the least discriminating power. Its absolute function value is .023. then the discriminant functions are as follows:

$$D_1 = 3.225 - 1.785X_1 + 0.428X_2 + 0.265X_3 + 0.659X_4 + 0.087X_5 - 1.178X_6 - 0.021X_7$$
$$D_2 = 1.913 + 0.990X_1 - 1.631X_2 - 0.023X_3 + 0.759X_4 + 0.929X_5 - 3.634X_6 + 0.098X_7$$

Calculating a score for each observation on each group's classification function and then assigning the observation to the group with the highest score, the result shown in Table 8.

TABLE VIII
FISHER'S LINEAR DISCRIMINANT FUNCTIONS

Variables	LDF		
	Fuction1	Fuction2	Fuction3
Gender	21.669	19.337	23.891
Age	1.632	3.833	2.282
Government	.981	1.163	.592
Occupation	7.872	7.381	6.067
Basis	1.398	.338	.423
Behavior	75.369	79.049	80.554
Grade	2.915	2.787	2.866
(Constant)	-62.633	-63.347	-69.928

C. Classification accuracy

Another important step in this analysis is that the classification of three groups of cancer is a discriminatory function. It is worth noting that the classification process may lead to classification errors, a single classification probability. Table 9 shows the classification results of a specific group that belongs to another group.

TABLE IX
CLASSIFYING OBSERVATIONS BY LINEAR DISCRIMINATION FUNCTIONS

Actual Population	Predicted			Sample Size	Correct Classification Ratio
	0	1	2		
0	467	196	269	932	50.10%
1	191	723	209	1123	64.40%
2	112	51	1882	2045	92%

Table 9 we note that the probability of correct classification of an infected back to the first group reached 50.1% For the second group 64.4% the third group 92 % While the probability of misclassification for the first group 49.9% For the second group 35.6% For the third group of 8%, the probability has reached the overall correct classification of 72.8% and the probability of misclassification group amounted to 27.2%.

Two LDFs were required for the three-group discrimination. Since two discriminant functions were required to discriminate between the three groups, each observation had two discriminant function scores which were used for identification. When the two scores for each observation were plotted on the x and y axes of a rectangular coordinate system, the clear spatial distinction between bladder, brain, and breast became apparent as shown in Figure. 1.

/

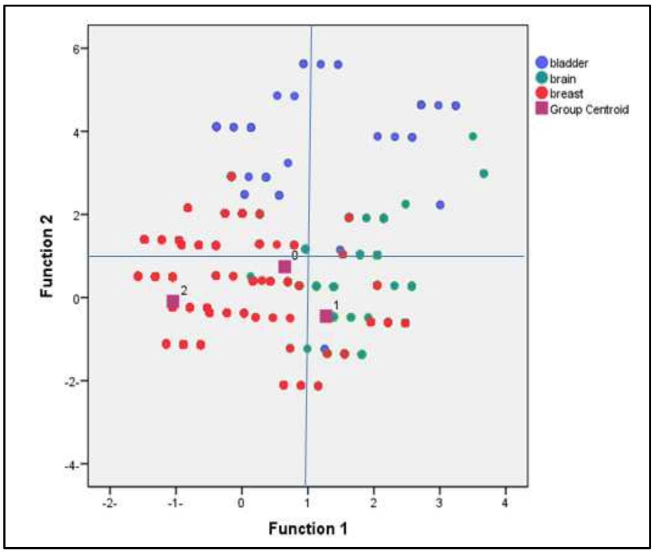

Fig. 1 Function space distribution of LDF score

Roc can draw a curve for an extra measure of support functions of classification, depending on the appropriate cutting point for discriminant score observation can be classified, and Figure 2 shows the ROC curve.

Figure 2, which shows the area under the curve of up to 0.754 which shows the probability of correct classification of the functions of discrimination and is a good value.

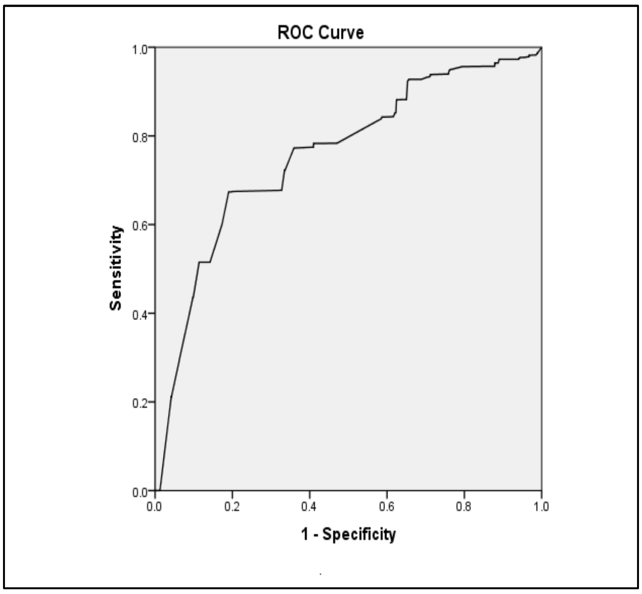

Fig. 2 ROC curve for +Discriminant Function

IV. CONCLUSION

This study proved discriminatory analysis using the method of linear classification function and formula probabilistic its high ability to classify data even if you do not check the assumptions of linear function discriminant. The other finding is that the linear discriminant function shows that all the variables studied have the importance of establishing the discriminant function. Besides the disease degree variable, it has not proved its contribution to the construction of the function.

This study found that Gender has the highest discriminating power, whereas the grade has the least discriminating power. Similarly, the behavior has the highest discriminating power, whereas the Government has the least discriminating power. It became clear to us that the third group (those with breast cancer) had the highest probability of the correct classification, where the probability of correct classification reached 92%, followed by the second group with brain cancer, where the probability of correct classification was 64%, and finally, the first group with bladder cancer had the lowest probability of correct classification, and from it, we conclude that increasing the sample size has a significant impact on the correct classification of observations.

In classifying the data, according to the linear discriminant function formula and probability formula, it gives the smallest data with the wrong classification, in which the probability of correct total classification reaches 72.2% although it is possible Macro error classification 28. Ultimately, this study proved that ROC curve as a measure to measure the accuracy of classification is very necessary to support our findings using linear discrimination function.

REFERENCES

[1] Raykov T, Marcoulides GA. An introduction to applied multivariate analysis. Routledge; 2008 Mar 10.
[2] Rencher AC. Methods of multivariate analysis. John Wiley & Sons; 2003 Apr 14.
[3] Härdle WK, Simar L. Cluster Analysis. InApplied Multivariate Statistical Analysis 2019 (pp. 363-393). Springer, Cham.

[4] Sharaf HK, Ishak MR, Sapuan SM, Yidris N. Conceptual design of the cross-arm for the application in the transmission towers by using TRIZ–morphological chart–ANP methods. Journal of Materials Research and Technology. 2020 Jul 1;9(4):9182-8.

[5] Dubrov A. Applied Multivariate Data Analysis. Statistica, Moscow. 1992.

[6] GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet (London, England). 2016 Oct 8;388(10053):1659.

[7] Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. Global burden of cancers attributable to infections in 2012: a synthetic analysis. The Lancet Global Health. 2016 Sep 1;4(9):e609-16.

[8] Huberty CJ, Olejnik S. Applied MANOVA and discriminant analysis. John Wiley & Sons; 2006 May 12.

[9] McLachlan GJ. Discriminant analysis and statistical pattern recognition. John Wiley & Sons; 2004 Aug 4.

[10] Catoiu I, Tichindelean M. Using discriminant analysis in relationship marketing. Annales Universitatis Apulensis: Series Oeconomica. 2013 Jul 1;15(2):727.

[11] Dinca G, Bociu M. Using discriminant analysis for credit decision. Bulletin of the Transilvania University of Brasov. Economic Sciences. Series V. 2015 Jul 1;8(2):277.

[12] Manly BF, Alberto JA. Multivariate statistical methods: a primer. CRC press; 2016 Nov 3.

[13] Agresti A. Categorical data analysis. John Wiley & Sons; 2003 Mar 31.

[14] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982 Apr;143(1):29-36.

[15] Sharaf HK, Ishak MR, Sapuan SM, Yidris N, Fattahi A. Experimental and numerical investigation of the mechanical behavior of full-scale wooden cross arm in the transmission towers in terms of load-deflection test. Journal of Materials Research and Technology. 2020 Jul 1;9(4):7937-46.

[16] Zwed JG. Rabab Adnan Al-Rubaye Assessment of Auditor's Responsibility on the Strategic Planning and Controls Using SWOT Analysis: An Ethical Approach. Journal of Engineering and Applied Sciences. 2019(14):600-9.

[17] Flayyih HH, Mohammed YN, Talab HR, Radhi NR. Integration of the system of Activity Based Costing and liability accounting. Integration. 2020 May;1(2):1-9.

[18] Sharaf HK, Ishak MR, Sapuan SM, Yidris N. Conceptual design of the cross-arm for the application in the transmission towers by using TRIZ–morphological chart–ANP methods. Journal of Materials Research and Technology. 2020 Jul 1;9(4):9182-8.

[19] Kadhim Sharaf H, Jalil NA, Salman S. A simulation on the effect of ultrasonic vibration on ultrasonic assisted soldering of Cu/SAC305/Cu joint. Journal of Advanced Research in Applied Mechanics. 2017;36(1):1-9.