Using Discriminant Analysis for Classification of Patient Status after Three Months from Brain Stroke

ID No.3228

(PP 206 - 223)

https://doi.org/10.21271/zjhs.24.2.13

Saman Hussein Mahmood
Department of Statistics & Informatics / College of Administration and Economics/ Salahaddin University-Erbil
saman.mahmood@su.edu.krd

Hoshamin Haji Khidir M.Sc. student
Department of Statistics & Informatics / College of Administration and Economics / Salahaddin University-Erbil
hoshamin.haji@yahoo.com

Received: 02/10/2019
Accepted: 16/01/2020
Published:20/04/2020

Abstract

Brain Stroke is a medical condition in which poor blood flow to the brain results in cell death. It also effects in parts of the brain not functioning properly. This situation is one of the main cause of death and a leading cause of disability in Kurdistan region and in worldwide. The aim of this paper is to predict the membership in two mutually exclusive groups (Group Died and Group Alive) after three months from brain stroke by using discriminant analysis which has used to classify the status of patient. The data of the research has been collected in Rzgari Hospital and Rojhalat Emergency Hospital of Erbil. The questionnaire form consists of 29 variables, one of them dependent and other variables are independent appendix (1). Sample size of the research contains 134 patients which faced brain stroke. In the practical part of the research we have built two models based on variable gender and model of both gender together by using stepwise method of discriminant analysis. After creating the three models we compared the results of the classification. The classification results show that the percentage of patients classified correctly for male and female separately higher than the percentage of patient classified correctly for the model of both genders together. For analysis, (SPSS V25) was used.

Keywords: Brain Stroke, Discriminant Analaysis, Stepwise Method.

1. Introduction

A brain stroke occur when the blood supply to part of your brain is interrupted or reduced, depriving brain tissue of oxygen and nutrients. Within seconds, brain cells begin to die. The good news is that strokes can be cured and prevented, and many fewer people pass away of stroke now than in the past.

In this research discriminant analysis is a technique that is used to analyze the research data because of the dependent variable is categorical and the predictor or the independent variable is interval and nominal in nature. The term categorical variable means that the dependent variable is divided into a number of categories. In this research the dependent variable is status of patients after three months from brain stroke. The categories are died and alive.

Discriminant analysis (DA) was first presented by R.A. Fisher in 1936. (Fisher, 1938). Previous studies showed the using of discriminant functions to classify three types of lesions in three groups: The normal, the benign and the malignant. It was observed that the correctly classified carcinoma is only 41.6% and for normal are 100% and the correctly classified means for the discrimination of fibroadenoma (which is a type of lump that can develop in the breast and typically appears in females. Fibroadenomas are very common, but they are benign, which means that they are noncancerous.) are 86.2 %. (Mokhtari-Dizadji1∗, Vahed, & Gity, 2003).

Furthermore another studies showed the used of discriminant functions to classify two groups. In one of those studies some researchers have used stepwise method of discriminant Analysis for Bronchitis Cancer patients. The results showed 90%, and 98% of dead and survive patients were classified correctly. (Al-mayali , Dakhil, & Al-Thabhawee , 2012)

In this research we have constructed three Discriminant function models based on the variable sex (Gender).

2. Methodology

This section studied some basic concepts of Discriminant analysis; some tests and methods used to analysis discriminant data analysis.

2.1 Discriminant Analysis:

Discriminant analysis (DA), also known as Discriminant function Analysis is a multivariate statistical tool used to classify a set of objects or individuals into one of the predetermined groups. DA is also a tool to solve the problem of separation where a sample of n objects will be separated into predefined groups. We wish to develop a rule to discriminate between the objects belonging to one group and the other. The result of the classification rule then becomes a random variable and some misclassifications may occur. The objective therefore is to arrive at a rule that would minimize the percentage of misclassification.

Linear Discriminant Analysis (LDA) is a statistical procedure proposed by R.A. Fisher (1936) in which multiple linear regression is used to relate the outcome variable (Y) with several explanatory variables, each of which is a possible marker to determine Y. When the outcome is dichotomous (taking only two values) the classification will be binary. This is not the case with the usual multiple linear regression where Y is taken as continuous. The regression model connects Y with the explanatory variables, which can be either continuous or categorical. In LDA the dichotomous variable (Y) are regressed on to the explanatory variables as done in the context of multiple linear regression.

Let **A** and **B** be two populations (groups) from which certain characteristics have been measured. For instance **A** and **B** may represent case and control subjects. Let $\mathbf{X_1, X_2,...,X_p}$ be the p-variables measured from samples of size $\mathbf{n_1}$ and $\mathbf{n_2}$ drawn from the two groups **A** and **B** respectively. All the samples belonging to **A** can be coded as **'1'** and the other coded as **'2'** or **'0'**. Obviously these two groups are mutually distinct because an object cannot belong to both **A** and **B** and there is no third group to mention. (Sarma & Vardhan, 2019).

2.2 The statistical procedure in LDA is based on three methods:

1. Maximum Likelihood Discriminant Rule (ML method)
2. Bayes Discriminant Rule
3. Fisher's Linear Discriminant Function

In our research Fisher's Linear Discriminant Function have used. Because of we wanted to build three models by Fisher's Linear Discriminant Function. We didnot try to compare these three methods.

Let μ_1 and μ_2 be the means vectors of the two multivariate populations with covariance matrices Σ_1 and Σ_2 respectively. LDA is useful only when H_0: $\mu_1 = \mu_2$ is rejected by MANOVA. In other words, we need the mean vectors in the two groups shall differ significantly in the two populations.

2.3 Assumption:

The following assumptions are assumed in the context of LDA (Sarma & Vardhan, 2019)**.**

1. The two groups have the same covariance matrix ($\Sigma_1 = \Sigma_2 = \Sigma$). In the case of unequal covariance matrices, we use another tool called quadratic discriminant analysis or we can use LDA.
2. A normality condition is not required. When this condition is also true then the Fisher's method gives optimal classification.

3. Data should not have outliers. (If there are a few they should be handled suitably).
4. The size of the smaller group (number of records in the group) must be larger than the number of predictor variables to be used in the model.

2.4 There are several purposes for Discriminant analysis:

- To investigate differences among groups.
- To determine the most parsimonious way to distinguish among groups.
- To discard variables which are little related to group distinction
- To classify cases into groups.
- To test theory by observing whether cases are classified as predicted.

2.5 Some basic concepts of Discriminant analysis:

1- Linear discrimination function

This type of discrimination is one of the simplest cases of discrimination and requires the following conditions:

a. It is assumed that the explanatory variables are distributed in a multivariate natural distribution (Afifi , S., & V., 2012).
b. The variances for all the studied totals (Matrices of Contrast and Contrast) are equal, ie, acceptance of the null hypothesis when testing the hypothesis:

$\mathbf{H_0} = \mathbf{\Sigma_1} = \mathbf{\Sigma_2} = \cdots = \mathbf{\Sigma_k}$

$\mathbf{H_1} = \mathbf{\Sigma_1} \neq \mathbf{\Sigma_2} \neq \cdots \neq \mathbf{\Sigma_k}$

whereas:

Σ: Matrix of Contrast and Contrast

k: Number of totals

c. Categorize existing views to totals ..., **n2, n1** accurately.

n1: cases in group A

n2: cases in group B

In this research group A and B in bellow showed:

A: group of alive

B: group of died

2- The discriminant function for two groups

We assume that the two populations to be compared have the same covariance matrix $\mathbf{\Sigma}$ but distinct mean vectors μ_1 and μ_2.

3- Discriminant analysis for several groups

In discriminant analysis for several groups, we are concerned with finding linear combinations of variables that best separate the k groups of multivariate observations. Discriminant analysis for several groups may serve any one of various purposes (RENCHER, 2002):

4- Group centroids

Group centroids are the mean (average) discriminant scores for each group in the dependent variable for each of the discriminant functions(DF). For 2 groups in the dependent variable there is one dicriminant function. The centroids are in a unidimensional space, one center for both group. And also For 3 groups in the dependent variable there are a couple dicriminant functions. Hence, the centroids are in a 2 dimensional space. Via relating the centroids a canonical plot can be created describing a discriminant function space. (Hahs-Vaughn, 2017)

5- Eigenvalue

Eigenvalue, also named the "characteristic roots", is the variance of the linear combination of variables that results from the decomposition of the correlation matrix. Also Eigenvalue is a ratio between the explained and unexplained variation in a model. For a good model the eigenvalue must be more than one. In discriminant analysis there is one eigenvalue for each discriminant function. The bigger the eigenvalue, the stronger is the discriminating

power of the function. In an analysis with three groups, the ratio between two eigenvalues shows the relative discriminating power of the discriminant function over the other. For instance, if the ratio of two eigenvalues is 1.6, the first discriminant function accounts for 60% more of the between-group variance for the three groups in the dependent variable compared to the 2nd discriminant function. Relative percentage of a discriminant function is the function's eigenvalue divided by the addition of all eigenvalues of all discriminant functions in the model. It represents the percentage of discriminating power for the model associated with a given discriminant function. Typically, the relative percentage of the 1st functions will be high. If values for the subsequent functions are small, then a single function is as good as two or more function in the grouping or classification. (Hahs-Vaughn, 2017)

6- Canonical correlation

The canonical correlation is an amount of the association between the groups in the dependent variable and the discriminant function. A high value shows a high level of association between the groups in the dependent variable and the discriminant function and vice-versa. The canonical correlation coefficient squared represents the proportion of variance in the discriminant function that is explained by the dependent variable's groups. It is important to reminder that the first canonical correlation will not always have the largest value. Because of the canonical correlation measures the relationship between the discriminant function and the groups, if there is a weak association between the groups and first discriminant function, the 1st canonical correlation coefficient will be weak, even though the first discriminant function, by default, represents maximum group differences. When the groups of the dependent variable are similar concerning (with respect to) the independent variables, the canonical correlations will be weak. (Hahs-Vaughn, 2017)

7- Wilks's lambda

In discriminant analysis (DA), the Wilk's Lambda which is the 11th letter of the Greek alphabet Λ is the statistic that measures discrimination via multivariate group differences. Lambda values can range between 0 and +1.0. The larger and closer to one the lambda value is, the less separation or discrimination there is between groups. Lambda values of +1.0 are found when group centroids are the same, reflecting absolutely no differences between groups. Smaller values of lambda show that the independent variables are more effective in discriminating cases into the dependent variable's groups. The smaller and closer to zero the lambda value is, the more separation there is between groups relative to within groups. Wilks's lambda can be interpreted as a test of inference using the chi-square distribution. Statistically significant Λ values indicate the discriminant function is statistically significant, and this can be interpreted as evidence of discrimination between groups. When we have only two groups in the dependent variable, only one Λ value will be computed. (Hahs-Vaughn, 2017)

8- Classification matrix

The classification matrix in DA is a cross-tabulation of the distribution of observed group membership to predicted group membership. For a good prediction, the values in the diagonal must be large and the values off the diagonal must be near to 0. (Hahs-Vaughn, 2017)

9- Box's M

Like in other multivariate data analysis, the Box's M tests are the assumption of equality of variance-covariance matrices in the groups. A large Box's M indicated by a small p-value indicates violation of this assumption. However, when the sample size is big, Box's M is usually large. In such situations, the natural logarithm of the variance-covariance

matrices for the groups are compared. In Box's M tests if the samples don't meet the assumption of normality, we shouldn't use this test. (Hahs-Vaughn, 2017)

10- Sample size

As a rule, A guideline for overall sample size suggests that there should be a sample size of at least 20 for every one predictor, with a more liberal ratio of a sample of size 5 for every predictor. (Hahs-Vaughn, 2017)

2.6 Stepwise Selection of Variables

When in the LDA phenomenon research we have a lot of predictors, the stepwise method can be useful by automatically selecting the "best" variables to use in the model. (SPSS, 2015) In this research we used stepwise .The stepwise method for selecting variables in discriminant analysis is rather like doing a stepwise regression and is especially useful in similar circumstances, namely when we have rather a long list of possible classification variables and it is unlikely that all will make a useful contribution to a set of discriminant functions.

2.7 Advantages and Disadvantages of Discriminant Analysis

A lot of the statistics tools have disadvantages and advantages. And discriminant analysis one those tools which have its own advantages and disadvantages. In below have explained. (Wilson)

2.7.1 Advantages

1) Discrimination of different groups.
2) Accuracy of classification of groups can be determined.
3) Helps for categorical regression analysis.
4) Visual graphics makes clear understanding for the two or more categories with computation logics.

2.7.2 Disadvantages

1) Linear discrimination cannot be used when subgroups are stronger.
2) The selection of the predictor variables are not strong until a strong classification exists.
3) It cannot be used when here is insufficient data to define sample means.
4) If the number of observations or cases are less, the discrimination method cannot be used

3. Application

3.1. Data Collection

The data set for this study about brain stroke was collected in **Rzgari Hospital** and **Rojhalat Emergency Hospital** in the capital of Kurdistan region Erbil**.** The data contains **134** patients or cases have been collected during the year 2018; beginning from **1st** January **2018** through **31st** December **2018** all of those cases were interviewed by the researcher Dr. Azad H. khidir in fifth floor of Rzgari Hospital and the Patient reception room of Rojhalat Emergency Hospital **.** Seventy-eight of the cases are male and 56 of the cases are female. Out of those patients 49 died after three months from interview and 85 survived after the same times. The data collected randomly. The questioner of the research consisted of 29 variables. The dependent variable (Parameters) of the study is the status of those patients which faced stoke after three months from brain stroke. Twenty-eight variables (Parameters) of the research are independent.

3.2. Applied Discriminant Analysis (DA) Method:

In the applied of DA we should investigate the data to get the best model. Also in the apply discriminant analysis stepwise method used to get the best models for classification.

3.3. Investigating Multivariate Normality:

The explanatory variables will be tested whether they have a normal multivariate distribution or not through the use of nonparametric test (Kolmogorov-Smirnov) and parametric test under the significance level (1%) or 0.01. This study contain only one scale variable in the independent variables. That is why in this investigation we check for the variable age for both status of the cases after three months. Also in this study we make three models for male, female and both gender. We must test normality separately for age based on gender. The other variables are categorical variable we cannot test multivariate normality because of the normal distribution deals with scale variable. The test results for classical method are summarized in the following table:

Table (3.1): Test of Normality

One-Sample Kolmogorov-Smirnov Test			
Sex			Age for alive
Male	N		54
	Normal Parameters	Mean	57.59
		Std. Deviation	9.482
	Test Statistic		0.123
	Asymp. Sig. (2-tailed)		0.04
Female	N		31
	Normal Parameters	Mean	55.42
		Std. Deviation	13.532
	Test Statistic		0.131
	Asymp. Sig. (2-tailed)		0.189
One-Sample Kolmogorov-Smirnov Test			
Sex			Age for died
Male	N		24
	Normal Parameters	Mean	65.13
		Std. Deviation	12.495
	Test Statistic		0.182
	Asymp. Sig. (2-tailed)		0.039
Female	N		25
	Normal Parameters	Mean	67.52
		Std. Deviation	13.254
	Test Statistic		0.147
	Asymp. Sig. (2-tailed)		0.173
Investigate normality for age of both	One-Sample Kolmogorov-Smirnov Test		
			Age for alive
	N		85

gender	Normal Parameters	Mean	56.8
		Std. Deviation	11.101
	Test Statistic		0.088
	Asymp. Sig. (2-tailed)		0.099
	One-Sample Kolmogorov-Smirnov Test		
			Age for died
	N		49
	Normal Parameters	Mean	66.35
		Std. Deviation	12.81
	Test Statistic		0.102
	Asymp. Sig. (2-tailed)		0.2

Table (3.1) note that the variable age based on male, female and both gender have a normal distribution because p-values for all tests are greater than 0.01.

3.4. Investigating Multivariate Outlier with Mahalanobis:

To detect outlier values in the research Mahalanobis Distance have used based on variable gender for the reason that we desire create three models for classification at the end of practical part of discriminant analysis will show. In this investigation first of all we search for the outlier for male in the result of the test show that there is not any outlier and then we examine for the female and also there is not any outlier, finally we tested for outlier for both gender the result showed that there is not any multivariate outlier in the data.

3.5. Multiple linear regression:

To test if the classification of groups in a variable status depends on at least one of the independent variables, in the multiple linear regression we have to make three multiple linear regression and the make up our mine to this test based on the sex or gender and both together we must test significantly the influence hypothesis, and estimate parameters of multiple linear regression with the some criteria as in the following table:

Table (3.2): Multiple linear regressions based on variable Gender after Stepwise

Gender	No. of Variable after stepwise	Variables in the model of regression after stepwise	Coefficients of Regressions	t	Sig.	Collinearity Statistics	F
						VIF	
Male	16	(Constant)	1.149	3.590	.001		36.943
		Quadriplegia	-.184	-2.473	.016	2.886	Sig.
		Size (mm)	.218	4.978	.000	2.943	0
		facial or perioral numbness	.768	8.433	.000	4.625	Adjusted R Square %
		Intra-Ventricular extension	-.431	-7.573	.000	2.234	87.483
		Lt Side weakness	-.450	-7.936	.000	2.249	Std. Error of the Estimates

		Disturbed consciousness	.271	4.860	.000	1.831	0.161
		Diplopia or visual disturbance	-.303	-3.353	.001	3.551	
		Chief complaint	.014	4.819	.000	1.647	
		Address	-.100	-2.244	.029	1.349	
		duration spent for reaching hospital	-.104	-5.390	.000	1.735	
		Speech difficulty	-.446	-4.516	.000	1.416	
		midline shift	.234	3.454	.001	3.124	
		Seizure	.170	2.469	.017	1.300	
		Rate Blood Suger	.118	2.365	.021	1.716	
Female	7	(Constant)	1.146	4.923	.000		F
		Address	.199	2.982	.005	1.189	
		midline shift	-.620	-9.871	.000	1.055	31.816
		Age	.008	3.308	.002	1.311	Sig.
		Smoking	-.224	-4.927	.000	1.028	0
		Diastolic Blood Pressure	-.167	-5.180	.000	1.233	Adjusted R Square %
		Systolic Blood Pressure	.253	5.648	.000	1.691	80.576
		Brain imaging (CT scan) for detecting bleeding	.007	4.297	.000	1.114	Std. Error of the Estimates 0.221
Male and Femal	5	(Constant)	1.631	5.768	.000		F
		Intra-Ventricular extension	-.224	-3.125	.002	1.479	28.778
		Lt Side weakness	-.276	-4.484	.000	1.087	Sig.
		midline shift	-.305	-4.517	.000	1.290	0
		Age	.008	3.273	.001	1.197	Adjusted R Square %
		Vertigo	.215	2.812	.006	1.015	52.632

							Std. Error of the Estimates
							0.329

Table (3.2) shows that, the F-test is highly significant for male, female and both gender because of (p-values are equal to zero), thus we can assume that the models explain a significant amount of the variance with the Adjusted R Square male = 87.483%, Adjusted R Square female = 80.576% and Adjusted R Square male and female = 52.632%, these mean that the multiple linear regressions explains 87.483% of the variance in the data of male, and the multiple linear regression explains 80.576% of the variance in the data of female and the multiple linear regression explains 52.632% of the variance in the data of male and female. Also, the "Standard Error (sd. Error) of the Estimate" is the standard deviation of the residuals and measures the efficiency of the estimated model and whenever it's small it indicates the efficiency of the estimated model and it is equal to 0.161, 0.221 and 0.329 respectively for male, female and both gender. Furthermore Table (3.4) tells us how to check for multi-collinearity. The Variance Inflation Factor (VIF) less than 10 this means there is no Multi-collinearty problem (Hair, Black, Babin , & Anderson, 2014). VIF of all variables based on the variable gender less than 10 that's mean we don't have multi-collinearity.

3.6. Tests of Equality of Group Means:

Table (3.3): Tests of Equality of Group Means

Variable	Test of Equality of group Means for Male			Test of Equality of Group Means for Female			Test of Equality of Group Means for Male and Female		
	Wilks' Lambda	F	Sig.	Wilks' Lambda	F	Sig.	Wilks' Lambda	F	Sig.
1Age	.886	9.155	.003	.813	11.759	.001	.849	22.026	.000
2Address	.928	5.528	.021	.977	1.209	.277	.993	.882	.349
duration 3spent for reaching hospital	.984	1.133	.291	.996	.189	.666	.998	.228	.634
4Chief complaint	.885	9.253	.003	.928	3.984	.051	.898	14.053	.000
5Headache	.984	1.180	.281	.979	1.099	.299	.998	.218	.641
6Disturbed consciousness	.918	6.300	.014	.976	1.280	.263	.936	8.490	.004
7Seizure	.999	.061	.806	.999	.051	.822	.999	.110	.741
8Rt side weakness	.919	6.266	.015	.977	1.209	.277	.996	.527	.469
9Lt Side weakness	.761	22.282	.000	.842	9.598	.003	.794	32.223	.000
10Monoplegia	.[a]			.975	1.312	.257	.985	1.877	.173
11Quadriplegia	.524	64.375	.000	.956	2.329	.133	.779	35.078	.000
Speech 12difficulty	.977	1.702	.196	.970	1.581	.214	.973	3.419	.067
13Mouth deviation	.988	.861	.357	.984	.822	.369	.992	.992	.321
14Vertigo	.904	7.525	.008	.992	.434	.513	.953	6.119	.015
15Ataxia	.936	4.863	.031	1.000	.001	.976	.980	2.592	.110

16facial or perioral numbness	.848	12.765	.001	.997	.165	.686	.939	8.058	.005
17Diplopia or visual disturbance	.904	7.525	.008	1.000	.001	.976	.963	4.827	.030
18Systolic Blood Pressure	.993	.477	.492	.961	2.062	.157	.997	.368	.545
19Diastolic Blood Pressure	.942	4.365	.040	.948	2.811	.100	.999	.063	.802
20Rate Blood Suger	.964	2.621	.110	.854	8.746	.005	.916	11.366	.001
21Diabetes	.975	1.848	.178	.992	.395	.533	.986	1.743	.189
22Ischaemic heart disease	.989	.804	.373	.936	3.502	.067	.968	4.038	.047
23Smoking	.960	2.962	.090	.943	3.096	.084	.939	8.038	.005
24Brain imaging (CT scan) for detecting bleeding	.995	.376	.541	.917	4.609	.037	.987	1.597	.209
25Size (mm)	.597	47.876	.000	.897	5.825	.019	.757	39.747	.000
26midline shift	.780	20.054	.000	.624	30.708	.000	.726	46.685	.000
27Intra-Ventricular extension	.712	28.675	.000	.703	21.587	.000	.710	50.606	.000
Sex							.977	2.913	.090
a.Cannot be computed because this variable is a constant.									

Table (3.3) shows that the Test of equality of means for male, female and both gender, the p-value (sig.) for Age, Chief Complaint, Lt Side weakness, Quadriplegia, Vertigo, facial or perioral numbness, Diplopia or Visual disturbance, Size (mm), midline shift and Intra-Ventricular extension variables for male are less than 0.01. Thus there are a significant difference for each Age, Chief Complaint, Lt Side weakness, Quadriplegia, Vertigo, facial or perioral numbness, Diplopia or Visual disturbance, Size (mm), midline shift and Intra-Ventricular extension variables for male between died and alive, but there is no a significant difference in those variables of male which there p-value are greater than 0.01 between the died and alive. Also for female and both gender together we have the same analysis as we showed in the Table (3.3).

3.7. Box's Test of Equality of Covariance Matrices:

Table (3.4): Log Determinants and Box's Test of Equality of Covariance Matrices

Status	Log Determinants for Male		Log Determinants for Female		Log Determinants for Male and Female	
	Rank	Log Determinant	Rank	Log Determinant	Rank	Log Determinant
Alive	1	.100	1	.287	1	.146
Died	1	-.312	1	-.578	1	-.352
(identity matrix)	1	0.000	1	0.000	1	0.000

Test Results for Male			Test Results for Female			Test Results for Male and Female		
Box's M for Male		1.140	Box's M for Female		4.353	Box's M for Male and Female		3.292
F	Approx.	1.119	F	Approx.	4.289	F	Approx.	3.262
	df1	1		df1	1		df1	1
	df2	8758.156		df2	7418.308		df2	35606.661
	Sig.	.290		Sig.	.039		Sig.	.071

Table (3.4) Log Determinants and Box's Test of Equality of Covariance Matrices shows the assumption of variance-covariance matrix across groups and Log determinants for male and female and both gender. The results of the test Box's M separately and together for the variable gender show that the p-values are greater than 0.01.Thus, equality of variance-covariance matrix can be assumed. Also the log determinant values are quite close for male and female respectively and both gender together.

3.8. Variables in the analysis by stepwise method

Table (3.5): Variables in the Analysis based on Gender

Variables in the Analysis for Male				
Step	Variable	Tolerance	F to Remove	Wilks' Lambda
14	Quadriplegia	.547	11.078	.119
	Size (mm)	.404	19.653	.134
	facial or perioral numbness	.107	80.379	.239
	Intra- Ventricular extension	.349	53.413	.193
	Lt Side weakness	.262	67.252	.216
	Chief complaint	.469	24.940	.143
	Disturbed consciousness	.412	26.825	.147
	duration spent for reaching hospital	.331	36.681	.164
	Diplopia or visual disturbance	.261	11.118	.119
	Speech difficulty	.543	17.871	.131
	Seizure	.643	11.671	.120
	midline shift	.336	12.547	.122
	Rate Blood Suger	.477	10.938	.119
	Systolic Blood Pressure	.685	5.410	.110
Variables in the Analysis for Female				
Step	Variable	Tolerance	F to Remove	Wilks' Lambda
7	midline shift	.480	97.430	.532
	Age	.755	10.940	.209
	Smoking	.670	24.278	.259
	Diastolic Blood Pressure	.536	26.835	.268
	Systolic Blood Pressure	.360	31.895	.287
	Brain imaging (CT scan) for detecting bleeding	.694	18.462	.237
	Address	.719	8.895	.201

Variables in the Analysis for Male and Female				
Step	Variable	Tolerance	F to Remove	Wilks' Lambda
5	Intra- Ventricular extension	.880	9.765	.492
	Lt Side weakness	.992	20.105	.531
	midline shift	.912	20.404	.532
	Age	.903	10.715	.495
	Vertigo	.970	7.907	.485

Table (3.5) shows Wilk's lambda, F, and significance values contribute information about difference means for each variable based on variable gender. It is immediately clear that the stepwise rule here was to minimize Wilks' Lambda at each step. It also we should to know that tolerance is the proportion of a variable's variance not accounted for by other independent variables in the equations. A variable with very low tolerance contributes little information to a model and can cause estimation problems. F to remove values is useful for describing what takes place if a variable is removed from the current models (given that the other variables stay). F to remove for the entering variable is the same as F to Enter.

3.9. Summary of Canonical Discriminant Functions based on Gender

Table (3.6): Eigenvalues and Wilks' Lambda based on Gender

Function of Male	Eigenvalue of Male	Canonical Correlation of Male		
1	8.976	.949		
Test of Function(s) of Male	Wilks' Lambda of Male	Chi-square of Male	df	Sig.
1	.100	147.211	14	.000
Function of Female	Eigenvalue of Female	Canonical Correlation of Female		
1	4.949	.912		
Test of Function(s) of Female	Wilks' Lambda of Female	Chi-square of Female	df	Sig.
1	.168	84.705	7	.000
Function of Male and Female	Eigenvalue of Male and Female	Canonical Correlation of Male and Female		
1	1.199	.738		
Test of Function(s) of Male and Female	Wilks' Lambda of Male and Female	Chi-square of Male and Female	df	Sig.
1	.455	95.747	5	.000

There are two groups for male, female and both gender. Therefore number of function = 1 for male, Female and both gender. It is important to note that The Eigen value for male is 8.976 (>1). Canonical correlation for male, r_c=.949(>.35). Wilks' Lambda of male = .100, p-value = .000 (<0.01). Thus, the Function 1 of male clarifies the variation well. For female, the Eigen value of female is 4.949 (>1), Canonical correlation for female, r_c=.912. Wilks' Lambda of Female = .168, P-value=.000(<0.01). Thus, the Function 1 female explains the variation well. For male and female the Eigen value of male and female is 1.199 (>1), Canonical correlation for male and female, r_c=.738. Wilks' Lambda of male and Female .455,

P-value=.000(<0.01). Thus, the Function 1 male and female explains the variation well. We also have to know that the Canonical correlation of male, female separately greater than the Canonical correlation for male and female. That's why the Classification Results of them different. The canonical correlation coefficient measures the association between the discriminant score and the set of independent variables. high Canonical correlation means the function that discriminated well.

3.10. The function:

Table (3.7): The Function According to Gender

	Function Coefficient for male	Standardized Function Coefficient for Male
duration spent for reaching hospital	-.879	-1.140
Chief complaint	.106	.844
Disturbed consciousness	2.072	.924
Seizure	1.700	.538
Lt Side weakness	-3.413	-1.508
Quadriplegia	-1.804	-.571
Speech difficulty	-3.044	-.694
facial or perioral numbness	5.895	2.454
Diplopia or visual disturbance	-2.179	-.827
Systolic Blood Pressure	.885	.372
Rate Blood Suger	1.235	.608
Size (mm)	1.438	.835
midline shift	1.741	.767
Intra- Ventricular extension	-2.909	-1.235
Constant	-6.746	
	Function Coeffiont for Female	Standardized Function Coefficent for Female
Age	-.041	-.558
Address	-1.052	-.525
Systolic Blood Pressure	-1.341	-1.177
Diastolic Blood Pressure	.884	.915
Smoking	1.186	.793
Brain imaging (CT scan) for detecting bleeding	-.038	-.710
midline shift	3.279	1.309
Constant	1.522	

	Function Coeffiont for Male and Female	Standardized Function Coefficent for Male and female
Age	-.035	-.408
Lt Side weakness	1.154	.515
Vertigo	-.900	-.342
midline shift	1.274	.541
Intra- Ventricular extension	.936	.396
Constant	-1.179	

Table (3.7) shows the discriminant equations as follow:

1. Discriminant function model For Male

Status of patient after three months for Male = -0.879 duration spent for reaching hospital +0.106 Chief complaint +2.072 Disturbed consciousness +1.700 Seizure -3.413 Lt Side weakness -1.804 Quadriplegia -3.044 Speech difficulty +5.895 facial or perioral numbness -2.179 Diplopia or visual disturbance +.885 Systolic Blood Pressure +1.235 Rate Blood Suger +1.438 Size (mm) +1.741 midline shift -2.909 Intra- Ventricular extension -6.746

2. Discriminant function model For Female

Status of patient after three months form brain stroke for Female = -0.041 Age -1.052 Address -1.341 Systolic Blood Pressure + 0.884 Diastolic Blood Pressure + 4.080 High Blood Pressure + 1.186 Smoking -0.038 Brain imaging (CT scan) for detecting bleeding –3.279 midline shift +1.522

3. Discriminant function model For Male and Female

Status of patient after three months from brain stroke for Male and Female = –0.035 Age +1.154 Lt Side weakness -0.900 Vertigo + 1.274 midline shift + 0.936 Intra- Ventricular extension -1.179

3.11. Centroid Classification Statistics

Table (3.8): Classification Results According to Gender

Classification Function Coefficients for Male			Functions at Group Centroids for Male	
Variable in the model of Male	Status		Status	Function
	Alive	Died		1
duration spent for reaching hospital	-2.436	-8.172	Alive	-1.878
Chief complaint	1.074	1.768	Died	4.649
Disturbed consciousness	14.376	27.900		
Seizure	28.931	40.024		
Lt Side weakness	-10.823	-33.099		
Quadriplegia	-3.274	-15.047		
Speech difficulty	6.906	-12.962		
facial or perioral numbness	48.258	86.732		
Diplopia or visual disturbance	-1.015	-15.237		
Systolic Blood Pressure	35.390	41.167		
Rate Blood Suger	23.557	31.620		
Size (mm)	21.488	30.872		

midline shift	29.812	41.178		
Intra- Ventricular extension	-10.464	-29.449		
(Constant)	-194.372	-248.353		
Classification Function Coefficients for Female			Functions at Group Centroids for Female	
Variable in the model of Female	Status		Status	Function
	Alive	Died		1
Age	0.118	0.300	Alive	1.911
Address	7.724	12.357	Died	-2.492
Systolic Blood Pressure	3.526	9.430		
Diastolic Blood Pressure	1.716	-2.175		
Smoking	3.818	-1.402		
Brain imaging (CT scan) for detecting bleeding	0.069	0.234		
midline shift	7.080	-7.360		
(Constant)	-28.405	-36.654		
Classification Function Coefficients for Male and Female			Functions at Group Centroids for Male and Female	
Variable in the model of Male and Female	Status		Status	Function
	Alive	Died		1
Age	0.561	0.642	Alive	.796
Lt Side weakness	7.613	4.983	Died	-1.483
Vertigo	12.999	15.049		
midline shift	4.589	1.685		
Intra- Ventricular extension	11.313	9.179		
(Constant)	-47.262	-45.982		

Here the Centroids for male, female and both gender in the table (3.8) are the mean discriminant scores for each group (status). Table of Functions at Group Centroids use to establish the cutting point for classifying cases. The centroids for male, female and both gender are calculated based on the three discriminate function models. Classification Function Coefficients of Two sets (one for each dependent group) for male, female and both gender of unstandardized linear discriminant coefficients are calculated, which can be used to classify cases (patients). Classification Function Coefficients is the classical method of classification.

3.12. Classification Results

Table (3.9): Classification Results According to Gender

Classification Results for Male					
Status			Predicted Group Membership		Total
			Alive	Died	
Original	Count	Alive	52	0	52
		Died	1	22	23
	%	Alive	100	0	100
		Died	4.3	95.7	100
98.7% of original grouped cases for male correctly classified.					

Classification Results for Female					
Status			Predicted Group Membership		Total
			Alive	Died	
Original	Count	Alive	31	0	31
		Died	0	25	25
	%	Alive	100	0	100
		Died	0	100	100
100% of original grouped cases for female correctly classified.					
Classification Results for Male and Female					
Status			Predicted Group Membership		Total
			Alive	Died	
Original	Count	Alive	76	7	83
		Died	9	39	48
	%	Alive	91.6	8.4	100
		Died	18.8	81.3	100
87.8% of original grouped cases for male and female correctly classified.					

Table (3.9) illustrated a simple summary of number and percent of patients classified correctly and incorrectly based on variable gender. It is immediately clear that the classification results for male 100.0%, and 95.7% of dead and alive patients (cases) were classified correctly. Only 4.3% of dead and alive patients were misclassified. It is also clear that the classification results for female 100.0%, and 100.0% of dead and alive patients were classified correctly. It means that we do not have misclassified. It must be pointed out the classification results for both genders 91.6%, and 81.3% of dead and alive patients (cases) were classified correctly. Only 18.8% of dead and alive for both genders and 8.4 of dead and alive for both genders patients were misclassified. We should note that the percentages for both genders are less than the percentages of male and female separately. The main aim of doing analysis for male, female and both gender separately were to get the best model for well forecasting. In other words 98.7% of original grouped cases for male, 100% of original grouped cases for female and 87.8% of original grouped cases for both gender classified correctly.

4. Conclusion

Throughout conducting the discriminant analysis and according to the results from the practical part the following decisions have been drawn:

1. The number of Variables in the discriminant function model for male and the discriminant function model for female are more than the number of variables in the discriminant function model for both genders.
2. In this paper we built three models. Because of the correctly classifications for male and female are higher than for both genders.
3. After three months from the follow up the situation of the patient from facing brain stroke we have got that 63.4% of patients survived and 36.6% of the patients died.
4. Classification results based on the variable gender indicate that
   a) 98.7% of original grouped cases for male correctly classified.
   b) 100% of original grouped cases for female correctly classified.

c) 87.8% of original grouped cases for male and female correctly classified.

5. Because of we had lots of independent variables. Stepwise method has been used to create the best models for forecasting the status of patient after three months from stroke.

5. References

1) Afifi , A. A., S., M., & V., A. C. (2012). Practical Multivariate Analysis. 5th ed. Boca Raton: CRC Press.
2) Al-mayali , Y. M., Dakhil, N. K., & Al-Thabhawee , G. D. (2012). Discriminant Analysis of Bronchitis Cancer Data. Kufa for Mathematics and Computer, 97-106.
3) Fisher, R. A. (1938). The Statistical Utilization of Multiple Measurements. Annals of Eugenics, 376-386.
4) Hahs-Vaughn, D. L. (2017). APPLIED MULTIVARIATE STATISTICAL CONCEPTS. New York and London: Routledge.
5) Mokhtari-Dizadji1∗, M., Vahed, M., & Gity, M. (2003). The application of discriminant analysis in differentiation of fibroadenoma and ductal carcinoma of breast tissue using ultrasound velocity measurement. Iran. J. Radiat. Res., 163-169.
6) RENCHER, A. C. (2002). Methods of Multivariate Analysis. Brigham: A JOHN WILEY & SONS, INC.
7) Sarma, K. V., & Vardhan, R. V. (2019). Multivariate Statistics Made Simple. New Yourk: Taylor & Francis Group.
8) SPSS. (2015). Stepwise Discriminant Analysis. Retrieved from IBM: https://www.ibm.com/support/knowledgecenter/en/SSLVMB_23.0.0/spss/tutorials/discrim_telco_stepwise.html
9) Wilson, L. (n.d.). Discriment analysis. Retrieved March 25, 2015, from SlideShare: https://www.slideshare.net/jitenderanduat/livi-ppt
10) Hair, J. F., Black, W. C., Babin , B. J., & Anderson, R. E. (2014). Multivariate Data Analysis. Harlow: England and Associated Companies throughout the world.

Appendix (1)
Questionnaire

1- Status of Patient after three months from brain stroke 1- Alive 2- Died
2- Age (year)
3- Sex (Gender): 1- Male 2- Female
4- Address: 1- Urban 2- Rural
5- Time spent (required) from 1st symptom and receiving treatment (duration spent for reaching hospital): in (minutes)
   1- less than 30 2- 31-60 3- 61-90 4- 91+
6- Chief complaint (……)
7- Headache 1- Yes 2- No
8- Disturbed consciousness 1- Yes 2- No
9- Seizure 1- Yes 2- No
10- Right side weakness 1- Yes 2- No
11- Left side weakness 1- Yes 2- No
12- Monoplegia 1- Yes 2- No
13- Quadriplegia 1- Yes 2- No
14- Speech difficulty 1- Yes 2- No
15- Mouth deviation 1- Yes 2- No
16- Vertigo 1- Yes 2- No
17- Ataxia 1- Yes 2- No
18- facial or perioral numbness 1- Yes 2- No
19- Diplopia or visual disturbance 1-Yes 2- No
20- Systolic Blood Pressure
21- Diastolic Blood Pressure
22- Rate Blood Sugar

23- Diabetes 1- Yes 2- No
24- Ischaemic heart disease 1- Yes 2- No
25- Smoking 1- Non-smoker 2- Current smoker 3- Ex-smoker
26- Brain imaging (CT scan) for detecting bleeding (Site of bleeding) (……)
27- Size (mm) 1- Small 2- Median 3- Large
28- midline shift 1- Yes 2- No
29- Intra- Ventricular extension 1- Yes 2- No

بەکارهێنانی شیکردنەوەی جیاکردنەوە بۆ پۆلێنکردنی ڕەوشی تەندروستی توشبووی جەڵتەی مێشك لەدوای سێ مانگ

سامان حسین محمود
بەشی ئامار و زانیارییەکان/ کۆلێژی بەڕێوەبردن و ئابووری/ زانکۆی سەڵاحەددین- هەولێر
saman.mahmood@su.edu.krd

هۆشەمین حاجی خدر
قوتابی ماستەر بەشی ئامار و زانیارییەکان/ کۆلێژی بەڕێوەبردن و ئابووری/ زانکۆی سەڵاحەددین- هەولێر
hoshamin.haji@yahoo.com

پوختە

جەڵتەی مێشك حاڵەتێکی پزیشکییە کە تووشبوونی مێشکە بە زیانێکی دەماری بە هۆی نەگەیشتنی خوێن بۆ مێشك، جەڵتەی مێشك دەبێتە هۆی مردنی خانەکان. هەروەها جەڵتەی مێشك دەبێتە هۆی ئەوەی مێشك بەشێوەیەکی دروست کار نەکات. جەڵتەی مێشك لە هەرێمی کوردستان و سەرانسەری جیهان یەکێکە لە هۆکارە سەرەکییەکانی مردن و بوون بە کەسێکی خاوەن پێدوایستی تایبەت. ئامانجی ئەم توێژینەوەیە پێشبینیکردنی ئەندامێتی دوو گروپە (گروپی مردن و گروپی ڕزگاربوون لە مردن) لە دوای سێ مانگ لە تووشبوون بە جەڵتەی مێشك بە بەکارهێنانی شیکرنەوەی جیاکردنەوە (Discriminant Analysis) کە بۆ پۆلێنکردنی ڕەوشی تەندروستی نەخۆش بەکارهاتووە. داتای توێژینەوەکە لە نەخۆشخانەی ڕزگاری هەولێر و نەخۆشخانەی فریاکەوتنی ڕۆژهەڵاتی هەولێر کۆکراوەتەوە. فۆرمی ڕاپرسیەکە بیست و نۆ گۆڕاو لە خۆدەگرێت. یەکێك لە گۆڕاوەکەن ناسەربەخۆیە و گۆڕاوەکانی دیکەش سەربەخۆن. پاشکۆ (1) قەبارەی نموونەکەمان 134 کەیسە (134 نەخۆشە) کە تووشی جەڵتەی مێشك بوون. لە بەشی کرداری توێژینەوەکە سێ مۆدێلمان لەسەر بنەمای گۆڕاوی ڕەگەز دروستکردووە بەبەکارهێنانی ڕێگەی هەنگاویی شیکردنەوەی جیاکردنەوە. لەدوای دروستکردنی سێ مۆدێلەکان بەراوردی ئەنجامەکانی پۆلێنکردنەکەمان کردووە. ئەنجامەکانی پۆلێنکردنەکە نیشانیدەدات کە ڕێژەی سەدی بە دروستی پۆلێنکردن بۆ نێر و بۆ مێ بەجیا بەرزترە لە ڕێژەی سەدی پۆلێنکردن بە دروستی بۆهەردوو ڕەگەزەکە بەیەکەوە. بۆ شیکردنەوەکە (SPSS V25) بەکارهاتووە.

کلیلەوشە: جەڵتەی مێشك، شیکردنەوەی جیاکردنەوە، ڕێگای پلە پلە(ستێپوایس)

استخدام التحليل التمييزي لتصنيف حالة المريض بعد ثلاثة أشهر من جلطة الدماغ

سامان حسین محمود
قسم الاحصاء والمعلوماتية/ كلية الادارة والاقتصاد/ جامعة صلاح الدين-اربيل
saman.mahmood@su.edu.krd

هۆشەمین حاجي خدر
طالب ماجستير قسم الاحصاء والمعلوماتية/ كلية الادارة والاقتصاد جامعة صلاح الدين-اربيل
hoshamin.haji@yahoo.com

ملخص

جلطة الدماغ هي حالة طبية يؤدي الى ضعف تدفق الدم إلى المخ وبالتالي تؤدي الى موت الخلية. كما أنه يؤثر في أجزاء من الدماغ لا يعمل بشكل صحيح. هذا الوضع هو أحد الأسباب الرئيسية للوفاة والسبب الرئيسي للإعاقة في إقليم كردستان وفي جميع أنحاء العالم. هدف البحث هو التنبؤ بحالة المريض المصاب بجلطة الدماغ ينتمي الى اي من مجموعتين (مجموعة ماتوا ومجموعة مازالوا على قيد الحياة) بعد ثلاثة أشهر من جلطة دماغية مستخدما التحليل التمييزي لتصنيف حالة المريض. تم جمع بيانات البحث في مستشفى رزكاري ومستشفى روزهلات للطوارئ في أربيل. يتكون نموذج الاستبيان من 29 متغيراً ، أحد المتغيرات هو متغير تابع . عينة البحث يتكون من 134 مريضاً تعرضوا لجلطة الدماغ. الغرض من البحث هو الحصول على أفضل النماذج للتنبؤ بحالة المريض بعد ثلاثة أشهر حيث يموت المريض أو يبقى على الحياة . في جزء العملي من البحث، تم استخدام طريقة تدريجية للتحليل التمييزي لبناء نموذجين على أساس للذكور والإناث ونموذج كلا الجنسين معًا. بعد إنشاء النماذج الثلاثة قارنا نتائج التصنيف. تظهر نتائج التصنيف أن النسبة المئوية للمرضى المصنفة بشكل منفصل للذكور والإناث أعلى من النسبة المئوية للمرضى المصنفة لنموذج كلا الجنسين معًا. وقد تم استخدام برنامج (SPSS,V25) لاغرض التحليل.

الكلمات الدالة: سكتة دماغية ، تحليل تمييزي ، طريقة تدريجية