

PAPER • OPEN ACCESS

Using Logistic Regression Model to Study the Most Important Factors Which Affects Diabetes for The Elderly in The City of Hilla / 2019

To cite this article: Zainab Abood Ahmed Al_Bairmani and Aasha Abdulkhaleq Ismael 2021 *J. Phys.: Conf. Ser.* **1818** 012016

View the [article online](#) for updates and enhancements.

You may also like

- [Automatic sleep stages classification using respiratory, heart rate and movement signals](#)
Maksym Gaiduk, Thomas Penzel, Juan Antonio Ortega et al.
- [Bearing fault diagnosis based on spectrum images of vibration signals](#)
Wei Li, Mingquan Qiu, Zhencai Zhu et al.
- [Factors associated with emerging multimodal transportation behavior in the San Francisco Bay Area](#)
Emily McAuliffe Wells, Mitchell Small, C Anna Spurlock et al.



The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Presenting more than 2,400
technical abstracts in 50 symposia



**ECS Plenary Lecture
featuring
M. Stanley Whittingham,**
Binghamton University
Nobel Laureate –
2019 Nobel Prize in Chemistry



Register now!



Using Logistic Regression Model to Study the Most Important Factors Which Affects Diabetes for The Elderly in The City of Hilla / 2019

¹Zainab Abood Ahmed AL_Bairmani, ²Aasha Abdulkhaleq Ismael

¹University of Babylon, Dep. Quality Assurance & Academic Performance

²Tikrit University - College of Administration & Economics

Email: aasha.a.alkalek@tu.edu.iq, zanab.abboud@uobabylon.edu.iq

Abstract. The aim of this paper is to study the most important factors affecting diabetes using the logistic regression method and to conduct all tests for this method (Hosmer and Lemeshow test, Omnibus tests of model coefficients, ...etc.). The randomized sample included (150) people among the elderly in Al-Hilla city, the research included focusing on (14) independent variables and most of these variables were found to have significance, effect and contribution to the logistic regression - binary response (not sick(0), sick(1)) model are (4) variables (cigarette smoking, exercise, vitamin (D), blood pressure), Which affects diabetes, and the rest of the variables have no significance or effect. The classification of observations using logistic regression-binary response model was accurate, as the overall correct classification rate was (92.7%) while the overall wrong classification rate was (7.3%).

1. Introduction

Logistic regression is defined as a type of regression used to predict of descriptive variables or order variables based on a set of mixed independent variables as if some of them are continuous variables), the other section being discrete variables (descriptive, order)[1].

It is usual in human, social and economic studies for the dependent variable to be discrete, so that it takes a dichotomous value or more. Which is somewhat restricted by the requirement that the dependent variable is quantitatively continuous instead of being descriptive separately[2]. (Iea, 1997) also states that the logistic analysis technique should be used in these cases and that although there are many statistical methods that have been developed to analyze descriptive data, such as discriminant function analysis, the regression logistic regression has several advantages that make it suitable for use in such situations [3].

The reason for using logistic regression instead of linear regression is due to two reasons:

The first reason: The dependent variable is, in linear regression, a continuous variable, while the logistic regression is a discrete variable (descriptive, order).



The second reason: When using linear Regression, there are a number of assumptions that must be taken into account (normality, homoscedasticity, ... etc). When these assumptions are violated, the best solution is to use logistic regression.

2. Previous studies

In recent years, applications of some modern statistical methods have increased in the analysis of classified data, particularly in the areas of medical, social, agricultural and other research. Analysis logistic regression is one of those statistical methods used to describe the relationship between two or more variables[9].

In the year 2015, the researcher (Ayed, Yasser Abdullah) examined the factors affecting polygamy in the Palestinian territories by comparing the models of neurological networks and logistic regression. The results were that the model neural networks outperformed the logistical regression model but with a small difference in the accuracy of the classification, and there were (4) variables common to the two methods, and they were of high relative importance in building the two models[1].

In 2016, the researcher (Suleiman, Ali Absher et al.) compared the logistic regression- binary response model and the discriminant function model to study the factors influencing the adequacy of family income and to find that the two models are significant, but the logistic regression outperformed the discriminative function in terms of classification accuracy, as the total percentage of correct prediction using the estimated logistic model was (65.3%), while the overall percentage of predicting the use of the discriminative function reached (57.1%) [4].

In 2018, the researcher (Saleh, Meitham Abdul Wahab) took the logistic regression-binary response method to represent the estimated function and the estimation of the parameters in a method (M.L.E) and the results indicated the most important factors affecting the lives of premature infants mortality (dep. variable). The logistic model was a good model for explaining the relationship between premature infant mortality and the variables that explain the behavior of this variable. All criteria gave good indications of the quality of the logistic model [5].

3. Research goal

The research aims to find logistic regression- binary response model and to identify the most important factors affecting model construction.

logistic regression method

The use of the logistic regression analysis has become more important day after day because it is concerned with the analysis of data with a binary response that is usually a binary response variable, in which the success variable takes a value (1), and a failure state takes a value (0). The logistic model is used to describe the relationship between the response variable (y), and one explanatory variable (x) or multiple explanatory variables(x1,x2,...,xn) and that relationship is expressed in the following form:

$$p(x) = \frac{1}{1 + e^{-\alpha - \beta x}} \quad (1)$$

This formula is known as the logistic response function, so that [9]:

β , α : Two model parameters to be evaluated, whereas $\beta > 0$ They are not restricted.

$p(x)$: The probability of response and be specific between(1,0).

(x): The explanatory variable where $(-\infty < x < \infty)$.

Type of Logistic Regression

There are several types of logistic regression :

Binary Logistic Regression: This type when the dependent variable takes only two values (0,1).

Multinomial Logistic Regression: This type is the subject of this research, where the dependent variable is more than two values (0,1,2) for example, or more than that.

Ordinal Logistic Regression: As for this type, the dependent variable is order, not descriptive and it can be binary or multiple.

Logistic Regression Model (Binary Response)

The model logistic regression model builds on a basic imposition: The variable (y) response variable that we study is a binary variable that follows Bernoulli distribution, and takes a value (1) with a probability of (p) and a value (0) with the probability (1-p=q), as we know that the linear regression whose explanatory variables and the response variable take continuous values, the model linking the variables is as follows:

$$y = B_0 + B_1x \quad (2)$$

y: It represents a continuous variable assuming average values (y) at a certain value of the variable (x): E(y) The model can be written as follows [6]:

$$E(y|x) = \beta_0 + \beta_1x \quad (3)$$

It is known in regression that the right-hand side of this model takes values from $(-\infty)$ to (∞) , but when we have two variables, one of which is binary (y), the simple linear regression model is not appropriate because:

$$E(y/x) = P_r(y = 1) = p \quad (4)$$

The right-hand end value is thus confined to the two digits (1,0), and the model is not applicable from the regression point of view, One way to solve this problem is to enter a suitable mathematical trans on the (y), and it is known that $(0 \leq p \leq 1)$ and the ratio $(p/(1-p))$ is a restricted positive amount between $(0, \infty)$, i.e. $(0 \leq p/(1-p) \leq \infty)$, by taking the natural logarithm of the basis (e) of transformation $(P/(1-P))$, the limits of its value become confined to $(-\infty \leq \log_e (P/(1-P)) \leq \infty)$ and accordingly, the regression model in the case of the independent variable one can be written as follows:

$$\log_e \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1x \quad (5)$$

If we have more than one explanatory variable, then the multiple logistic regression model becomes as follows:

$$\log_e \left(\frac{P}{1-P} \right) = \beta_0 + \sum_{j=1}^J \beta_j x_{ij} \quad (6)$$

Since:

$$i = 1, 2, \dots, n, j = 1, 2, \dots, J$$

The response probability formula for logistic regression - binary response model can be obtained by raising both sides of equation (6) to base (e), and we get:

$$\frac{P}{1-P} = e^{\beta_0 + \beta_1x_i} = \frac{1}{\frac{1}{P} - 1} = e^{\beta_0 + \beta_1x_i} \quad (7)$$

By using algebraic methods, the response probability formula for the logistic regression - binary response model is written as follows:

$$P = \frac{1}{1 + (e^{\beta_0 + \beta_1x_i})^{-1}} \quad (8)$$

Thus the probability of the response variable (y) taking the value (1) is

$$P(y = 1 | x) = \frac{1}{1 + (e^{\beta_0 + \beta_1x_i})^{-1}} \quad (9)$$

The probability that the response variable (y) takes the value (0) is

$$P(y = 0 | x) = \frac{1}{1 + e^{\beta_0 + \beta_1x_i}} \quad (10)$$

Note that

$$P(y = 1|x) + P(y = 0|x) = 1$$

Logistic regression model evaluation

The first method: Examining the statistical significance of each independent variable separately:

Wald Statistics is the measure of statistical significance testing for each of the logistics regression coefficients it is the measure for testing the statistical significance of each of the logistical regression coefficients, where the hypotheses for this test are:

$$H_0: b = 0$$

$$H_1: b \neq 0$$

The null hypothesis that the logistic regression factor associated with the independent variable equal zero, the following formula is calculated and coded by (w^2):

$$w^2 = \left(\frac{\hat{b}}{S.E_b} \right)^2 \quad (11)$$

whereas:

\hat{b} : Represents the value of the logistic regression coefficient of the independent variable (x).

$S.E_b$: The value of the standard error represents the logistic regression coefficient of the independent variable (x) [7].

(w^2): Tracing chi-square distribution (χ^2), as for if we want an account (w): So the distribution is a normal distribution (z) the formula becomes:

$$w = \left(\frac{\hat{b}}{S.E_b} \right)$$

- If the value of (w) is statistically significant, This means rejecting the null hypothesis which states that the logistic regression coefficient is not equal to zero (That is, the independent variable x has an effect on predicting the value of the dependent variable).

- If the value of (w) is not statistically significant, This means that the independent x can be deleted from the model because it has no statistical significance. [3]

The second method: Check the fit of the logistic regression model as a whole:

1- Maximum Likelihood Ratio Test

To determine the significance of parameters from their insignificance in logistic regression is through the maximum likelihood ratio, And it is based on finding that sample results can occur when estimating model parameters, let us assume that a good model is the one that has a large maximum likelihood ratio value of the observation results that we have n observations for each variable, (if the model is ideal, the value is = 1 and the value (-2lnl=0)), let's say we have (n) observation per variable, so the deviation (D) for a model defined is as follows[7]:

$$D = -2 \log \frac{\text{Likelihood of the fitting model}}{\text{Likelihood of the saturated model}}$$

Hypotheses are tested:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_K = 0$$

$$H_1: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_K \neq 0$$

Where the comparison is between the value of (D) for the model that includes all variables, with the value of (D) for the model including the constant only:

$$G = D(\text{model without variable}) - D(\text{model with variable})$$

D-value compares with the tabular critical value with P degrees of freedom (p: number of variables) if the value observation is > critical value, the null hypothesis is rejected, i.e. variables are significant[8].

2- Hosmer-Lemeshow test

The Hosmer-Lemeshow test is used for the purpose of evaluating the fit of the estimated model, and there are two strategies for this:

First strategy: grouping sample cases based on percentages of expected possibilities.

Second strategy: grouping sample cases based on fixed values of expected probabilities.

The first strategy is preferred to the second strategy when the expected probabilities are small,

The statistic, symbolized by (H), is calculated according to the following formula:

$$H = \sum_{k=1}^g \frac{(O_k - \hat{n}_k \hat{p}_k)}{\hat{n}_k \hat{p}_k (1 - \hat{p}_k)} \quad (12)$$

whereas:

\hat{n}_k : Represents the total number of cases in group k.

O_k : The number of responses represents $Y=1$, where $(O_k = \sum_{i=1}^{\hat{n}_k} Y)$

\hat{p}_k : Represents the mean probability available for group k, where $(\hat{p}_k = \sum_{i=1}^{\hat{n}_k} \frac{Y}{\hat{n}_k})$.

If the (calculated value of $H >$ the tabular value Chi-square) with a degree of freedom (g-2) at a level of significance (0.05), This means that the model is identical to the observation data [2].

3- $(R^2_{cox-snell}), (R^2_{Nagelkerke})$ tests

For the purpose of testing the explanatory strength of the logistic regression model, and this is done with the following two scales:

The first scale/ statistics $(R^2_{cox-snell})$: This statistic is a scale in the square of the geometric mean per observation and takes the formula:

$$R^2_{cox-snell} = 1 - \left(\frac{L_0}{L_m}\right)^{2/N} \quad (13)$$

whereas:

L_0 : The weight function represents only the model containing the constant term.

L_m : The weighting function of the model containing all explanatory variables is represented.

N: Represents the total number of observations.

• The second scale/ statistics $(R^2_{Nagelkerke})$: This statistic is the measure of improvement in the square of the geometric mean per observation, and takes the formula:

$$R^2_{Nagelkerke} = \frac{1}{\text{Maximum possible } R^2_{cox-snell}} \quad (14)$$

4. The Practical Work

Diabetes is the accumulation of glucose in the blood, which leads to high blood glucose levels that cause health problems associated with diabetes. The reason for diabetes is either that pancreas cannot produce insulin or insulin that it produces is not enough, and cannot function properly and without insulin doing its job.

According to The World Health Organization, diabetes will be the leading killer disease, ranking seventh in death by 2030. Moreover, one of the most expensive diseases is that the low-income sugar patient spends about 25% of his family's income on the care of sugar, and the cost of global sugar treatment is estimated at 215-375 billion\$.

The World Health Organization says that the most important factors are (exercise, healthy food, blood pressure control...etc) and These factors have been taken into account in this research to determine their impact, And 150 random samples were selected from the elderly (56 and over) in Hilla, the logistic regression- binary response model, is chosen[sick (1)/ not sick (0)] and the most important factors (14 independent variables) affecting diabetes are chosen.

Table 1. for the response variable (dependent variable).

Repetitive Values for Response Variable Classifications			
Original value	Internal value	Frequency	Percentage
Not sick	0	59	39.3%
Sick	1	91	60.7%
Total		150	100%

Note from Table No. (1) that the dependent variable is a binary response variable (Sick:1) and (Not sick:0), Where the highest percentage of Frequency was for people with diabetes

(60.7%) This means that the comparison will be according to the patients suffering from the disease.

As for Table No. (2), it shows the classification of the independent variables into categories, and their frequencies and percentage were calculated:

Table 2. The classification of the independent variables.

Descriptive Statistics for Independent Variables			
Explanatory Variables		Frequency	Precentag%
x_1 : Gender	(1) Male	60	40.0
	(2) Female	90	60.0
x_2 : Living	(1) City	96	64.0
	(2) Village	54	36.0
x_3 : Age	(1) 65-69	84	56.0
	(2) 70-74	42	28.0
	(3) 80-	24	16.0
x_4 : Social status	(1) Unmarried	3	2.0
	(2) Married	82	54.7
	(3) Widower/Separated	65	43.3
x_5 : Monthly income	(1) Enough	98	65.3
	(2) Somewhat enough	40	26.7
	(3) Enough Not	12	8.0
x_6 : Educational level	(1) Elementary and lower	82	54.7
	(2) Medium and junior high	43	28.7
	(3) Institute/University and above	25	16.7
x_7 : Cigarette smoking	(1) Yes	97	64.7
	(2) No	53	35.3
x_8 : Having another chronic disease	(1) Yes	120	80.0
	(2) No	30	20.0
x_9 : The patient feels pain in the body	(1) Yes	103	68.7
	(2) No	47	31.3
x_{10} : Exercise	(1) Yes	88	58.7
	(2) No	62	41.3
x_{11} : Take corticosteroids	(1) Yes	51	34.0
	(2) No	99	66.0
x_{12} : Vitamin (D)	(1) Natural	41	27.3
	(2) Little	45	30.0
	(3) A little bit	64	42.7
x_{13} : Body mass	(1) Natural	22	14.7
	(2) Ghee	50	33.3
	(3) Very obese	78	52.0
x_{14} : Blood pressure	(1) Natural	70	46.7
	(2) High	51	34.0
	(3) Very high	29	19.3

As for Table No. (3) below, it shows the estimation of the parameters of the logistic regression model according to the Maximum Likelihood method for all the significant parameters of the model (\hat{b}) excluding other non-significant factors, The factors were significant (Sig.<0.05) as well as the value of the constant and standard error for each parameter, the (Wald) statistic for each of the features, and the number of degrees of freedom.

Table 3. The estimation of the parameters of the logistic regression model according to the Maximum Likelihood method

Variables in the Equation						
Variables	B	S.E.	Wald	Df	Sig.	Exp(B)
Cigarette smoking : $x_{7.1}$	4.362	1.338	10.627	1	.001	78.375
Exercise : $x_{10.1}$	-4.474	1.379	10.530	1	.001	.011
Vitamin (D) : $x_{12.1}$	-4.514	1.595	8.012	1	.005	.011
Blood pressure : $x_{14.1}$	-5.416	2.109	6.596	1	.010	.004
Constant	-1.386	3.487	.158	1	.000	.250

Where it was found that the factors included in the model are only four variables, namely (cigarette smoking, exercise, vitamin D, blood pressure) Thus, the binary logistic regression model can be formulated with the following equation:

$$\hat{y} = \log_e \left(\frac{P}{1-P} \right) = -1.386 + 4.362 x_{7.1} - 4.474 x_{10.1} - 4.514 x_{12.1} - 5.416 x_{14.1}$$

Where we notice that the smoking factor (x_7) was ranked first in terms of relative importance in affecting diabetes, Since the regression coefficient for this variable is (4.362), and this is interpreted by changing the value of (x) with an increase of one unit from a non-smoker to a smoker, increasing the probability of developing diabetes by (4.362) in the logarithm of preference for the dependent variable, Then the exercise factor (x_{10}) came second, followed by the vitamin D factor (x_{12}), and finally, the blood pressure factor (x_{14}).

The (Wald) statistic for chi-square distribution (x_2) with a degree of freedom (1) has a formula for smoking factor (x_7):

$$w^2 = \left(\frac{\hat{b}}{S.E_b} \right)^2 = \left(\frac{4.362}{1.338} \right)^2 = 10.627$$

As for Exp(B), it represents the odds of success for the explanatory variables (factors) in relation to the response variable, and its value indicates the amount of change that the explanatory variable gets, When (Exp (B) <1) leads to a decrease, and in the case (Exp (B) >1) it leads to an increase, but in the case of its value close to (1), this indicates that any change in the explanatory variable by one unit will not affect on the value of the response variable, For example, the odds of success for the smoking variable (x_7) would be in the formula:

$$\text{Exp}(4.362) = 78.375$$

Thus, the odds of success for the smoking factor is greater than one, and therefore any change in the smoking factor by one unit will increase the incidence of diabetes.

For the purpose of classifying any of the vocabulary of the research sample according to the response variable that takes one of the two classifications (0, 1) the results of classifying observations according to the probabilities of responding to the logistic regression model (sick, not sick) and as shown in Table No.(4):

Table 4. Classification by Response Probabilities.

Classification by Response Probabilities			
Observed	Predicted		
	Not sick(0)	Sick(1)	Percentage Correct
Not sick (0)	53	6	89.8
Sick (1)	5	86	94.5
Overall Percentage	38.7	61.3	92.7

Table No. (4) shows that the probability of correct classification for not sick people (0) is 89.8% and the probability of classification error is 10.2%. The probability of correct classification for sick persons (1) is 94.5% and the probability of classification error is 5.5%. As for the overall correct classification rate: 92.7%, the overall correct classification error: 7.3%.

Table 5. Omnibus Tests of Model Coefficients

Omnibus Tests of Model Coefficients			
	Chi-square	Df	Sig.
Model	147.478	21	.000

Table No. (5) shows the test of the significance of the relationship between the response variable and the explanatory variables depending on the statistical analysis of the final value of chi-square which is 147.478 in the model with a degree of freedom of 21 and a significant level ($p\text{-value} = 0.00 < 0.05$) This means rejecting the null hypothesis, meaning that there are no significant differences between the model in terms of the constant term only, without the explanatory variables The model is in terms of the explanatory variables and thus the model is statistically significant and matches the data well (that the explanatory variables included in the model have an impact, significance, and contribution to the classification).

Table 6. Hosmer and Lemeshow Test.

Hosmer and Lemeshow Test			
Step	Chi-square	Df	Sig.
1	.350	8	1.000

Table No. (6) illustrates the Hosmer-Lemeshow test, which is one of the important tests for measuring the quality of fit of the logistic regression model It depends on how close the observed probabilities are to the expected probabilities, The value for chi-square (χ^2) was .35 with a degree of freedom 8 and a significant level ($p\text{-value} = 1.00 > 0.05$) It is the significant fitting quality of the model (That is, the model fits well with the data).

Table 7. Contingency Table for Hosmer and Lemeshow Test.

Contingency Table for Hosmer and Lemeshow Test						
		Not sick: 0		1: Sick		Total
		Observed	Expected	Observed	Expected	
Step 1	1	15	14.999	0	.001	15
	2	15	14.960	0	.040	15
	3	14	13.871	1	1.129	15
	4	9	9.413	6	5.587	15
	5	5	4.572	10	10.428	15
	6	1	1.001	14	13.999	15
	7	0	.166	15	14.834	15
	8	0	.016	15	14.984	15
	9	0	.003	15	14.997	15
	10	0	.000	15	15.000	15

Table No. (7) shows the validation of the model that was reconciled to the data, where the data were divided into ten groups, and these groups are defined in ascending order to estimate the risk. The first group corresponds to the less serious vocabulary group (other than the diabetic) and the second group corresponds to the more serious vocabulary group (diabetic).

Table 8. Model Summary.

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	53.587	.626	.848

Table No. (8) shows the statistic value of each: $(R^2_{Nagelkerke}) = .848$ and $(R^2_{cox-snell}) = .626$.

5. Conclusions

1-The research variables that have significance, influence and contribution to the binary logistic regression model are four variables (cigarette smoking, exercise, vitamin D, blood pressure), and the rest of the variables have neither significance nor effect.

2- The classification of observations using the binary logistic regression model was accurate, as the overall correct classification percentage was 92.7% while the overall wrong classification rate was 7.3%.

6. Recommendations

1- Expand by utilizing the logistic regression method in the medical fields.

2- If the response variable is two-response, and the independent variables are mixed (quantitative, qualitative) then it is necessary to use the logistic regression method.

3- Finally, given the fact that diabetes is one of the most expensive diseases for the affected person, his family and the state, we recommend that the results of this research be approved in the endocrinology and diabetes centers.

7. References

- [1]. Abed, Yasser Abdullah (2015), "*A Comparison of Statistical Methods to Study the Factors Affecting Polygamy in the Palestinian Territories*", Master Thesis, Faculty of Economics and Administrative Sciences, Al-Azhar University/ Gaza.
- [2]. Abbas, Ali (2012), "*Using the Logistic Regression Model in Predicting Functions with Qualitative Dependent Economic Variables*", Kirkuk University Journal of Administrative and Economic Sciences, Volume 2, Issue 2, Iraq.
- [3]. Babtain, Adel (1990), "*Logistic Regression and How to Use It in Building Prediction Models for Data with Two-Value Dependent Variables*", Unpublished PhD thesis specializing in Statistics and Research, Umm Al-Qura University, College of Education, Department of Psychology, Saudi Arabia.
- [4]. Suleiman, Ali Absher and others (2016), "*A Comparison Between the Binary Logistic Model and the Discriminatory Function in Classification*", El-Neelain University, College of Economic and Social Studies, Journal of Economic Sciences, Vol. 17 (2) 2016, Sudan.
- [5]. Najm al-Din, Adnan Karim and Salih, Maytham Abdul-Wahab (2018), "*Using Logistic Regression to Study the Causes of Mortality in Premature Babies in Babil Governorate*", Karbala University Scientific Journal, Volume 16, Issue 2, Iraq.
- [6]. Ghanem, Adnan and Farid (2011), "*The Use of the Logistic Regression - Binary Response Technique in the Study of the Most Important Economic and Social Determinants of Household Income Adequacy*", Journal of Economic and Legal Sciences, Damascus University, Volume 27, Issue 1, Syria.
- [7]. Hasina, Jawani (2018), "*A Comparative Study between Logistic Regression and Discriminant Analysis of Predicting Failure of Institutions*", Master Thesis, College of Science and Commerce, University of Elaraby Ben Mahidi /Oum El Bouaghi, Algeria.
- [8]. Al-Habil, Abdullah and Majed (2014), "*A Comparative Study Between Linear Discrimination Analysis and Multiple Logistic Regression*", Journal of Human Sciences Research, An-Najah University, Volume 28, Issue 6.
- [9]. jbara, Azhar kadhim, (2014), "*Analysis of Multi- Response Data to Diagnosis Eyes Diseases Using Discriminant Function and the Logistic Regression*", Master Thesis, College of Administration and Economics AL- Mustansiriah University, Iraq.