ISSN 2303-4521
Periodicals of Engineering and Natural Sciences *Methodology/Method*
Vol. 9, No. 4, October 2021, pp.481-490

Constructing a model to determine the most important factors affecting diabetes disease

Auday Taha Raheem[1], Husam Abdulrazzak Rasheed [2], Ghiath Hameed Majeed [3]
[1,2] College of Management and Economic, Mustansiriyah University
[3] College of Basic Education, Mustansiriyah University

ABSTRACT

Diabetes occurs as a result of the inability of the pancreatic gland to produce an adequate amount of insulin, or as a result of the body's inability to use the insulin it produces as required. For the purpose of presenting an explanatory study on blood sugar disease and determining the most important factors that affect the incidence of this disease, a mathematical model was built, which is a two-response logistic regression model when the correlation function is of type (logit, Normal, or Gompartz). Then, they were compared to determine the best model using Aic standard and R^2. It has been shown that the two-response logistic regression model with a correlation function (logit) is the best because it has the lowest Aic coefficient and the highest R^2. Likewise, it was found that the effect of the age variable x1 is more significant than the effect of the gender variable x2 on the incidence of diabetes, and that the risk is high between the response variable blood sugar (y) and the age variable (x1). While the risk is low between the response variable (y) and the gender variable (x2).

Keywords: Binary logistic regression, Link function, Maximum likelihood function, Odd ratio.

Corresponding Author:

Auday Taha Raheem
College of Management and Economic
Mustansiriyah University
E-mail: uday_adm@uomustansiriyah.edu.iq

1. Introduction

Diabetes is a chronic disease that occurs as a result of the inability of the pancreatic gland to produce enough insulin or because of the body's inability to use the insulin it produces as required. There are many causes of this disease that vary according to the types of this disease, such as (environmental factors, family history, genetic factors, geographical factors, hunted organisms, and others). In order to know whether the gender factor and age factor affect the blood sugar level, a suitable mathematical model must be built to represent this phenomenon, then predict the possibilities of developing blood sugar, as well as determine the factors that affect the incidence of this disease, including age and gender. The model used for this is the two-response logistic regression model. Whereas, the response variable (dependent variable y) that represents blood glucose level is dual response. The dual response logistic regression model is characterized by its ability to analyze disaggregated data, especially in the medical and social fields in which the response variable (y) is a binary response or a nominal variable or a response variable.
Researcher Saleh [21] in (2014) used a logistic regression model to study the survival time of patients with leukemia. It found that the model that contains the treatment variable and the anemia variable is the most appropriate model for the data. In (2015), the Sarai researcher [12] used a logistic regression model to estimate the continuation of marital life in the city of Kut, and it was concluded that the factors (the age difference between the spouses, the number of children, sugar) have a significant impact on the divorce between the spouses as well as the same Impact on the continuation of marital life. In the same year (2015), the Zarkani researcher [15] applied a multiple logistic regression model to anemia data for the purpose of

identifying the most important factors affecting anemia. As for the year (2016), the researcher Al-Hussainawi [16] built a logistic regression model for deformed births in Iraq, and the researcher found that there are several factors that affect determining the type of deformity (gender, degree of kinship, father's profession, type of birth). In (2017) the Gezzy researcher [5] applied a dual-response logistic regression model to acute and chronic anemia disease. In (2018), the researcher Al-Zaidi [16] built a dual-response logistic regression model, and the model was applied to patients with high and low blood pressure.

2. Aim of the Research

The research aims to build a dual-response logistic regression model to study the effect of age and gender variables on the incidence of blood sugar disease.

3. Binary logistic regression model

It is one of the nonlinear regression models in which the non-linear relationship is between the response variable (y) and the independent explanatory variables (x1, x2,…, xn). The basic premise of the binary logistic model is that the (dependent) response variable (y) binary response takes one of the two values (0,1). That is, upon success, he takes (1) with a probability (p), and upon failure he takes (0) with a (1-p) probability. Therefore, the variable yi distributes the Bernoulli distribution y ~ Ber (pi). To describe the relationship between the response variable y and the influencing variables x1, x2, ..., xn, we will use the binary logistic regression model with the following formula [1] [4]:

$$P(x) = \frac{1}{1+e^{-\beta_o-\beta_1 x_i}} \qquad ...(1)$$

Where β_o, β_1 represent the two parameters of the model, while P(x) represents the probability of response, and *xi* represents the influencing independent variables where (i = 1, 2,…, n) and that $(-\infty < x_i < \infty)$.

The formula in equation (1) defines the logistic response function and has the advantage that P(x) is finite between (0,1) and that the two parameters (β_0,β_1) are not restricted, and that the probability of responding to the logistic regression model takes the value (1) according to the following formula [9]:

$$P(y=1|x) = \frac{1}{1+e^{\beta_o+\sum_{j=1}^{k}\beta_j x_{ij}}} \qquad ...(2)$$

The probability of responding to the model when it takes the value (0) is:

$$P(y=0|x) = \frac{1}{1+e^{\beta_o+\sum_{j=1}^{k}\beta_j x_{ij}}} \qquad ...(3)$$

The two-response logistic regression model can be expressed as (5):

$$y_i = P_i + \varepsilon_i \quad , \ i=1,2,...,n \quad ...(4)$$

Where: Pi represents the logistic regression function or the logistic response function, and ε_i represents the random error. Thus, the logistic regression function is written in the following form [21]:

$$P_i = \frac{e^{\beta_o+\sum_{j=1}^{k}\beta_j x_{ij}}}{1+e^{\beta_o+\sum_{j=1}^{k}\beta_j x_{ij}}} \qquad ...(5)$$

And the estimated logistic regression function is as follows:

$$\hat{P}_i = \frac{e^{\hat{\beta}_o + \sum_{j=1}^{k} \hat{\beta}_j x_{ij}}}{1 + e^{\hat{\beta}_o + \sum_{j=1}^{k} \hat{\beta}_j x_{ij}}} \qquad ...(6)$$

Where $\beta_o, \beta_1, ..., \beta_k$, (j = 1, 2, ..., k) represent the unknown parameters in the model. And that equation (5) shows that the shape of the relationship between the influencing variables (x_{ij}) and the response probability (P_i) is nonlinear [9] [21].

4. Link function

Response with a variable A binary logistic regression model is described for the response of g(p), which is binary (0,1), (yes, no) or (success, failure) and others. And the equation of the model will be as follows [13]:

$$g(p) = \beta_o - x'\beta \qquad ...(7)$$

Where (p) represents the probability of success (response probability), and g(p) represents the link function which is related to the response probability, as for x, it represents the explanatory variables that affect the response associated with the feature vector of the model β [15].

It is possible to describe several link functions, such as the (logit) function, which represents the inverse of the general logistic cumulative distribution function. As well as the link function (probit), this is the inverse of the cumulative distribution function of the normal distribution. The (gompit) function is the inverse of the cumulative distribution function of the Cumbers distribution. This can be explained in the following table [16] [4]:

Table 1. some link functions for logistic regression

Name of the link function	Link function g(p)	Distribution
logit	$g(p) = \log_e\left(\frac{p}{1-p}\right)$	Logistic
normit	$g(p) = \Phi^{-1}(Z)$	Normal
gomit	$g(p) = \log_e(\log_e(1-p))$	Gompertz

5. Maximum likelihood function

The method of the Maximum Likelihood Function is used to estimate the parameters of a two-response logistic regression model. This method finds values ($\hat{\beta}$) they are estimates of vector (β). So if we have r of explanatory variables $x_1, x_2, ..., x_r$, since these variables are distributed in a binomial distribution with the two parameters (ni , pi) and that (yi) represents the sum of the success cases in each attempt from (ni), and there are (k) explanatory variables affecting each group of sums, so the probability density function for (y_i) can be represented as follows [3] [6]:

$$P_i(X_i = x_i) = C_{x_i}^{n_i}\, p_i^{x_i}\, (1 - p_i)^{n_i - x_i} \qquad ...(8)$$

Where xi = 0,1,2, ..., n_i, i = 1,2, ..., r, and E (y_i) = $n_i.p_i$, V (y_i) = $n_i.p_i$ ($1-p_i$), , as well as the ($p_i = \frac{x_i}{n_i}$) and ($q_i = 1 - p_i = \frac{n_i - x_i}{n_i}$). and the logarithm of the maximum likelihood function for the joint distribution of the data (y_i) can be represented as the following formula:

$$\ln(L(p)) = \sum_{i=1}^{r}\left[\ln C_{x_i}^{n_i} + \ln\left(\frac{p_i}{1-p_i}\right) + n_i \ln(1-p_i)\right] \quad \ldots(9)$$

After performing some mathematical operations, we will have:

$$p_i = \frac{\exp(X_L'\beta)}{1+\exp(X_L'\beta)} \quad \ldots(10)$$

$$1 - p_i = \frac{1}{1+\exp(X_L'\beta)} \quad \ldots(11)$$

Substituting formulas (10) and (11) into equation (9), we get:

$$L(P) = \sum_{i=1}^{r}\left[\ln C_{y_i}^{n_i} + y_i\, x_i'\,\beta + n_i\left(\frac{1}{1+\exp(x_i'\,\beta)}\right)\right] \quad \ldots(12)$$

and by using the Newton-Raphson method to find estimates of the greatest possibility, we will get [12]:

$$\hat{\beta}_{S+1} = \hat{\beta}_S + (x'v\,x)^{-1}x'\,y \quad \ldots(13)$$

where ($\hat{\beta}_{S\ +1}$) is the column vector for the values of loop (S+1) of the order [(K+1)*1]
$\hat{\beta}_S$ the column vector for the values of loop (S) of the order [(K+1)*1],
X represents the matrix of influential variables of the order [(r*(k+1)],
and V represents the matrix of diagonal variances of the order (r*r).

6. Odd ratio

It is a positive value used to describe the relationship between the response variable y and the explanatory variables (X's) of the model. When the odds ratio is equal to one, this indicates that there is no relationship between the response variable and the explanatory variables of the model. If the value of the odds ratio is greater than one, the success rate is high when compared to the reference level of the factor, or it is at a high level for the continuous explanatory variable. But when the odds ratio is less than one, the success rate is low when compared to the reference level of the factor or it is at a high level for the continuous explanatory variable [8] [13].

7. Collecting the data

The study sample included (160) people who were randomly selected from Al-Binook Private Hospital in Al-Rusafa District - Baghdad for the year (2019). Information was collected from this sample, which included those with and without diabetes, which represents the dependent variable (y). When the blood sugar level for one person is less than (126) mg/dL, this means that the person is healthy and the blood sugar level is within its normal levels, especially if it is less than (100) mg/dL. But if the ratio is confined between (125-100) mg/dL, this means that there is a high possibility of the person contracting the disease, but he is not infected yet. Therefore, we will symbolize the blood sugar level that is less than (126) mg/dL with the number (0). But when the person's blood sugar level is greater or equal to (126) mg/dL, this means that this person is sick and has diabetes, and we will symbolize this case with the number (1). As for the explanatory variables, x_1 represents age and x_2 represents gender, where we will symbolize the male with the number (1) and the female with the symbol (2). The data has been tabulated as in Table 2.

Table 2. The data on blood sugar levels, broken down by age and gender

Gender	Age	blood sugar level	Gender	Age	blood sugar level	Gender	Age	blood sugar level
female	65	262	male	33	500	male	80	400
male	79	456	male	53	324	male	45	234
male	56	345	female	43	495	male	33	276
female	66	162	female	45	234	male	32	110
male	56	400	male	32	532	male	36	119
male	43	347	male	21	120	male	78	320
male	54	114	female	27	123	female	56	346
female	23	109	female	56	118	female	65	432
male	20	90	male	54	200	male	43	120
male	27	134	female	76	287	male	47	346
male	19	98	female	35	367			
female	29	100	male	29	322			
male	57	132	male	57	126			
male	48	208	male	63	187			
female	24	143	female	24	90			
female	45	117	male	19	98			
female	42	97	male	45	180			
male	46	98	male	44	178			
male	30	100	male	24	200			
male	32	126	male	68	409			
male	20	143	male	25	432			
male	28	207	male	52	188			
female	24	280	male	76	120			
male	40	234	female	56	114			
male	41	256	male	68	187			
female	45	435	male	47	287			
female	65	345	male	89	489			
male	40	128	female	76	434			
male	30	135	female	45	300			
female	39	176	female	33	243			

8. Data analysis

Using the SPSS 17 program on the data of Table (1), we have shown that the response to the dependent variable (y) is binary (0,1), when the value (0) means that the person does not have diabetes and the number of people without this disease has reached (55). And when you take the value (1), it means that the person has the disease and the number of people with diabetes has reached (105) as in Table 3.

Table 3. The number of people with and without diabetes for the dependent variable (y)

Case	No. of Person	Category	Variable
Uninfected	55	0	blood sugar level
Infected	105	1	(y)
	160		Sum

9. Building models

To estimate the effect of age x_1 and gender x_2 on the binary response variable (y), which represents blood sugar. If the blood sugar level is greater than or equal to (126) mg/dL, the person is considered infected and symbolized by (1). But if the blood sugar level is less than (126) mg/dL, the person is considered uninfected and we symbolize him as (0). And to estimate this effect, we will use the Binary Logistic Regression Model when the link functions are (logit), (Compartz), and (Normal). And by using the program (Minitab 17), we get the table of analysis of variance (ANOVA) shown as below.

Table 4. Shows ANOVA table for the parameters of the binary logistic regression model when the link function is (Gomit, Normit, logit)

Link fun.	source	d.f.	adj. dev.	adj. mean	chi-square	P-value
logit	Regression	2	32.115	16.058	32.73	0.000
	Old	1	23.218	23.218	23.22	0.000
	Gender	1	7.544	7.544	7.54	0.000
	Error	157	173.802	1.107		
	Total	159	205.917			
Normit	Regression	2	31.732	15.866	31.73	0.000
	Old	1	22.835	22.835	22.84	0.000
	Gender	1	7.171	7.171	7.17	0.000
	Error	157	174.185	1.109		
	Total	159	205.917			
Gomit	Regression	2	30.357	15.178	30.36	0.000
	Old	1	21.46	21.46	21.46	0.000
	Gender	1	6.08	6.08	6.08	0.014
	Error	157	175.561	1.118		
	Total	159	205.917			

Table 4 shows the results of ANOVA for the parameters of the model and for each of the three link functions (Gomit, Normit, logit). Where the P-value and the Chi-square value for both independent variables, x_1 (age) and x_2 (gender) shown above. We found that the P-value for both independent variables is less than (0.05) and for the three link functions. This indicates that the estimated parameters of the model do not agree with the null hypothesis $H_o: \beta_1 = \beta_2 = 0$, thus, the alternative hypothesis is accepted $H_1: \beta_1 and\ \beta_2 \neq 0$. That means the multiple parameters are significant, and there is a significant effect of each of age variable (x_1) and gender variable (x_2) on the dependent variable (y) on the blood sugar level and for all the used link functions (Gomit, Normit, logit). The following table shows the estimated equations for the possibility of developing high blood sugar for the impact of the variables of age (x_1), gender (x_2), and for each distribution.

Table 5. Shows the estimated equations according to each link function

Link fun.	Gender	Regression equation	P(1)
logit	1	$Y' = -1.131 + 0.05408$ (old)	$= \frac{\exp(Y')}{1+\exp(Y')}$
	2	$Y' = -2.135 + 0.05408$ (old)	
Normit	1	$Y' = -0.6599 + 0.03165$ (old)	$\phi(Y')$
	2	$Y' = -1.244 + 0.03165$ (old)	

Link fun.	Gender	Regression equation	P(1)
Gomit	1	$Y' = -0.986 + 0.02972$ (old)	$1 - \exp(-\exp(Y'))$
	2	$Y' = -1.524 + 0.02972$ (old)	

Pearson test, Deviance test, and Homser-Lemeshow test were used to test the goodness of fit for the model at each link function, and Table 6 represents the results.

Table 6. The tests of goodness of fit (Person, Deviance, and Hosmer-Lemeshow) for each link function of the model

Test		Link fun.		
		logit	Normit	Gomit
Pearson	d.f.	157	157	157
	Chi-square	165.7	167.74	174.08
	P-value	0.302	0.264	0.167
Deviance	d.f.	157	157	157
	Chi-square	173.8	174.19	175.56
	P-value	0.17	0.165	0.148
Hosmer-Lemeshow	d.f.	8	8	8
	Chi-square	7.72	8	9.4
	P-value	0.461	0.433	0.31

From Table 6, we find that these tests confirmed to us the extent of the model's compatibility with the data, as the p-value was higher than the specific level of significance which is (0.05) for each of the Pearson test, Deviance test and Homser-lemeshow test, this confirms that the model is efficient for the data under study and for all three link functions (Gompit, Normit, logit).

10. Comparison of models

To compare among the models, we find the value of the coefficient of determination (R^2), the adjusted coefficient of determination adj. (R^2), and Aic coefficient for the model and according to each of the three link functions Gomit, Normit, logit, and this was clarified in the following Table (7)

Table 7. The values of R^2, adj. R^2, and Aic for the model and according to the link functions (Gomit, Normit, logit)

Link fn.	R^2	Adj. R^2	Aic
Logit	15.60%	14.62%	179.80
Normit	15.41%	14.44%	180.19
Gomit	14.74%	13.77%	181.56

- Coefficient of determination R^2

Table 7 shows the results of the coefficient of determination R^2, and its ratio represents the proportion of the independent variables x_1 and x_2 that are affected by the model, where the highest percentage is in a logistic model when the link function is of the type (logit) and the value of its coefficient of determination has reached (R^2 = 15.6%), then came a logistic model when the link function is of the type (Normit) and its value was (R^2 = 15.41%), then the logistic model came in the last rank when the link function is of the type (Gomit) and its value was (R^2 = 14.74%).

- Adjusted Coefficient of Determination adj. R^2

As for the adjusted coefficient of determination adj. R^2 which represents the explanatory value of the effect of the variables x_1 (age) and x_2 (gender) on this model. We note that the highest value of the adjusted coefficient of determination adj R^2 is when the link function of the model is of the type (logit) and its value is (adj. R^2 = 14.62%), and this value represents the percentage explained by the independent variables: age (x_1)

and gender (x_2) that affects the model. Then the ratio (14.44%) when we use the link function of type (Normit) for the model, and in the last rank (13.77%) when using the link function of the type (Gomit) for the model.

- Akike information criteria (Aic Coefficient)

As for the Aic coefficient, it is a measure that represents the efficiency of the model, as the lower the value of this coefficient the model be the best. As we note from Table (7) that the lowest value of the Aic coefficient appears when using (logit) as the link function of the model and its value is (Aic = 179.8), then the value (Aic = 180.19) when using the link function of type (Normit), finally the value (Aic = 181.56) when using the link function of type (Gomit).

Now after calculating these three criteria, R^2, adj. R^2, and Aic, we can determine the best model, which is the model in which the link function used is of the type (logit). It has the advantage that it has the highest coefficient of determination R^2, the highest adjusted coefficient of determination adj R^2, and the lowest Aic coefficient compared to the rest of the link functions, and this indicates that it is an efficient model.

- Odd Ratio

When calculating the odd ratio for the variable age (x_1), we find that its value (1.0556) and this value is greater than (1), This means that the risk is high. That is, the degree of agreement is high between the response variable (y), which represents blood sugar, and the age variable (x_1).

As for the odd ratio for the gender variable (x_2), the value of the odd coefficient was (0.3664) which is less than (1), this indicates that the risk is low. That is, the degree of agreement is less between the response variable (y), which represents blood sugar, and the gender variable (x_2).

11. Conclusions

1. The best estimated model for diabetes is the logistic regression model when its correlation function is of type (logit) because it had the lowest value of the Aic coefficient and the largest value of the coefficient of determination.
2. The compatibility was high between the response variable (y), which represents blood sugar, and the age variable (x_1), because the odd rate for the age variable was (1.0556), which is a value greater than (1), and this means that the risk is high.
3. We note that the compatibility is low between the response variable (y), which represents blood sugar, and the gender variable (x_2), because the odd rate for the gender variable has reached (0.3664) and this value is less than (1), and this means that the risk is low.
4. Through the results of the analysis of variance (ANOVA) for the model variables, we found that the variables of age (x_1) and gender (x_2) have a statistically significant effect, as the P-value of both variables was less than (0.05). Also, the chi-square test indicates that the estimated parameters of the model were significant. And this means that there is a significant effect of both variables, age (x_1) and gender (x_2), on the dependent variable (y) diabetic disease.
5. The effect of the variable age (x_1) was more significant than the effect of the gender variable (x_2).

12. Recommendations

1. Adopting the model that has been reached in order to develop scientific plans that reduce the incidence of diabetes in the blood.
2. Study other variables and analyze them to find out the extent of their impact on diabetes disease.
3. Using statistical and mathematical methods to follow up and study this disease in order to develop health aspect.
4. Interest in recording various health data and facilitating its access to those concerned with health affairs and researchers.

References

[1] A.Y. Abdulridah, “Using The Logistic Regression For Study of The Phenomenon of Unemployment Among Young People In Baghdad Province”. A research of Higher Diploma in applied statistics, College of Administration and Economics, University of Baghdad, 2015.

[2] A.A. Adetumji, “Application of Ordinal Logistic Regression in the Study of Students Performance”. *Mathematical Theory and Modeling*, vol. 3, no. 11, 2013.

[3] S. M. Ahmed, “Comparison of the Ordinal, Linear Regression Models with the Logistic Regression Model with an application”. M.Sc. thesis in Statistics, College of Administration and Economics, University of Baghdad, 2019.

[4] A. J. Al-Obaidy, “Logistic regression in the study of the adequacy of family income in Diwaniya”. *Al Kut Journal of Economics Administrative Sciences*, Scientific Conference of the College of Administration and Economics, Wassit University, 2012.

[5] M. S. M.Al-Ezzy, “Comparison of the Classification Process with the Linear Characteristic Function Method and the Logistic Regression in the Presence of the Multicollinearity with an application”. M.Sc. thesis in Statistics, College of Administration and Economics, University of Baghdad,2017.

[6] A. K. B. Al-Bajlan, “Using (cox) and (logistic) Models in Survival Analysis with Application”. M.Sc. thesis in Statistics, College of Administration and Economics, University of Baghdad, 2007.

[7] H. K. M. Al-Bahadly, “Study of Child Mortality in Maysan Governorate”. A research of Higher Diploma in applied statistics, College of Administration and Economics, University of Baghdad, 2018.

[8] H. I. S. Al-Baiaty, “Path Analysis in a Logistic Regression Model with Application”. Sc. thesis in Statistics, College of Administration and Economics, University of Mustansiriyah, 2005.

[9] Z. J. K. Al-Hasnawi, “Structure logistic regression model of anomalies birth in Iraq except Kurdistan region, for 2015”. A research of Higher Diploma in applied statistics, College of Administration and Economics, University of Baghdad, 2017.

[10] H. F. T. Al-Nuaimi, “Estimation of the Logistic Regression model and the Severity function of Cox Regression models - Case study”. A research of Higher Diploma in applied statistics, College of Administration and Economics, University of Mustansiriyah, 2018.

[11] K. M. Al-Rawi, “Introduction to Regression Analysis”. University of Mosul, 1987.

[12] A. A. K. Al-Sarai, “An Applied Study to Estimate the Logistic Regression Model on the Continuation of Marital Life in Kut City”. A research of Higher Diploma in applied statistics, College of Administration and Economics, University of Baghdad, 2015.

[13] R. F. Al-Temeemy, “Regression and Time Series- Advanced Applied Statistical Methods using the Minitab System”. Baghdad, 2014.

[14] A. Al-Wardi, “Statistical Forecasting Methods”. Publications of the Ministry of Higher Education and Scientific Research, Baghdad, Iraq, 1998.

[15] A. H. S. Al-Zarkani, “Estimate the Parameters of the Multiple Logistic Regression Model in the Case of a Separation and Multicollinearity Problem with Application in the Medical Field”. M.Sc. thesis in Statistics, College of Administration and Economics, University of Baghdad, Iraq, 2015.

[16] M. A. Kh. Al-Zaidi, “Modeling the Effect of Treatment and Age for Blood Pressure Patients (Comparative Study)”. A research of Higher Diploma in applied statistics, College of Administration and Economics, University of Mustansiriyah, 2018.

[17] A.S. Fullerton, and J. Xu, “The Proportional Odds with Partial Proportionality Constraints Models for Ordinal Response Variables”. Social Science Research, no. 41, 2012.

[18] A. K. Jbara, “Analysis of Multi - response Data to Diagnosis Eyes Diseases using Discriminant Function and the Logistic Regression (a comparative study)”. M.Sc. thesis in Statistics, College of Administration and Economics, University of Mustansiriyah, 2014.

[19] M. M. Kadhim, “A Comparative Study of Some Nonlinear Model Estimation Methods with Application”. M.Sc. thesis in Statistics, College of Administration and Economics, University of Mustansiriyah, 2017.

[20] T. H. Muhammed, “Maximizing the Efficiency of the Analysis for Logistic Curve by Using Power Transformation”. M.Sc. thesis in Statistics, College of Administration and Economics, University of Mustansiriyah. 2006.

[21] A. H. Salih, “Logistic Regression Analysis to Study Survival Time of Leukemia Patients”. *Journal of the College of Administration and Economics*, University of Mustansiriyah, vol. 3, Issue 9,2014.