

# Statistical Analysis of the Fish Death in Babylon Province by using an Interactive Network of Simple and Multiple Linear Regression/Iraq

<sup>1</sup>Asmaa Shaker Ashoor, <sup>2</sup>Ali Abdul Karim Kadim Naji  
University of Babylon  
College of Education for Pure Science  
Babylon/Iraq

**Abstract**—Everyone knows about the importance of fish fortune in Iraq as it is the most prominent vital sectors and it is one of the components of country's nutrition, but it's exposed to a terrible damage as well as the rest of national fortune sources after 2003 which is followed by wars, invasions, the absence of governmental interest, water shortage and unorganized random haunting. The resources of fish fortune in Iraq distributed to the Euphrates and Tigris and their branches, natural lakes, muddy lakes that are founded to fish raising and floated cages on the rivers. Fish fortune has exposed to a calamity in few past days, it is the phenomenon of death of thousands carp and Cyprinus which is considered according to specialists and the organization of international health as an environmental calamity followed by pernicious consequences in a country by which most of its inhabitants work in agriculture, cattle raising, fish and livestock, from what was mentioned, the importance of the present study has been constructed to make a scientific statistical study which is based on the explanation of indicators of increasing and decreasing of kinds and floated fish places according to regions that the case of calamity appeared.

**KeyWords**— *Mathematical Operations, Linear Regression, SPSS, Minitab*

## I. AIM

Construction of statistical research study using systematic scientific methods that explain the phenomenon of dead fish and the knowledge of height and decline indicators in the types and places of dead fish by regions where the case appeared.

## II. METHODOLOGY

Babylon province was chosen as a spatial sample to collect data on fish according to species and weights distributed according to the regions where the phenomenon occurred. The data were collected in the field by the researcher in cooperation with the department of livestock in the Directorate of Agriculture of Babylon.

## III. INTRODUCTION

In this study, the researcher constructs the (simple and multiple) linear regression models for the study and analysis of the data of the dead fish by comparing linear regression models and identifying any of these models with

high morale and explain the phenomenon in a statistical explanation of the two theoretical aspects which use statistical and practical theories using SPSS and Minitab and through a display of the morale basics and efficiency of model assessment.

The determination of a specific form of a relationship between two variables starting from knowing and determining the nature of those variables and the relationship between them such that one of the variables affects the other and causes its occurrence. Hence the possibility of describing the variables in terms of an independent variable, it can be internal and determined from inside.

$$y_i = f(x_i)$$

Then through a statistical theory, the form of the relationship can be determined whether it is linear or not and because two variables depend on another while the other is independent, we can thus call the model a simple linear regression model in the case of the existence of one independent variable, while in case of the existence of more than one independent variable, so this model is known as a multiple linear regression model.

## IV. LINEAR REGRESSION MODELS[1][2]

The simple linear regression model is one of the regression model forms and comes with one independent explanatory variable. It uses many relationships between variables, such as the relationship between the quantities of dead fish in the Babylon province with the number of fish farms or any other variables and the relationship of the variables among them within the model is that one of these variables is dependent on its value from the model, the other is also an independent value determined from outside the model. The simple regression model is described as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \dots (1)$$

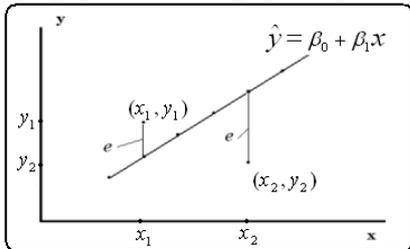
The random variable associated with the  $\varepsilon_i$  The model is the independence of the independent variable  $x_i$  And the distribution of the random variable by natural distribution is given by  $\delta$  and  $\mu$  parameters. Random error represents irregular measurement errors resulting from data behavior and other errors as well as the error of any view independent from other observations.

V. THE METHOD OF ORDINARY SMALL SQUARES (OLS) [3][6]

This method is one of the estimation methods, through which the estimation formulas can be reached for the parameters of the model:  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

Where:  $y_i$  Is the dependent variable  
 $x_i$  is the independent variable

$\beta_0$  And  $\beta_1$  Are the parameters of the model and represent the form of propagation of data graphically as follows:



$\varepsilon_i$  Error in interpreting  $y_i$  Variable from the previous relationship can be written as follows:

$$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i)$$

Through using the simple squares method, it is possible to estimate the parameters of the model which seems to reach the best linear model when the sum of the squares of the deviation between the real and estimated observations is minimized which means the attempt to reduce the following value:

$$SSE = \sum_{i=1}^n e_i^2 \quad \dots\dots (2)$$

VI. ESTIMATING THE PARAMETERS OF THE MODEL [4][8]

In order to estimate the parameters of the model, we need to reduce the sum of the squares of the errors. This is done by using the partial differentiation of the parameters of the linear model and their equivalents by zero and as follows:

$$\frac{\partial f(b_0, b_1)}{\partial b_0} = \sum_{i=1}^n 2(y_i - (b_0 + b_1 x_i))(-1) = 0 \quad \dots\dots\dots (3)$$

$$\frac{\partial f(b_0, b_1)}{\partial b_1} = \sum_{i=1}^n 2(y_i - (b_0 + b_1 x_i))(-x_i) = 0 \quad \dots\dots\dots (4)$$

In mathematical operations, the following natural equations are obtained:

$$nb_0 + \left(\sum_{i=1}^n x_i\right)b_1 = \sum_{i=1}^n y_i$$

$$\left(\sum_{i=1}^n x_i\right)b_0 + \left(\sum_{i=1}^n x_i^2\right)b_1 = \sum_{i=1}^n x_i y_i \quad \dots\dots\dots (5)$$

In order to solve these equations, we obtain the estimated parameters that are the inclination  $b_1$  And the fixed limitation:

$$b_1 = \hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad \dots (6)$$

The fixed limit  $b_0$  Is calculated from its relation to inclination by the use of the following relation:  $b_0 = \hat{\beta}_0 = \frac{\sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i}{n} \quad \dots\dots\dots (7)$

In terms of equations (6) and (7), we obtain the estimated value of the dependent variable:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad \dots\dots (8)$$

4- Efficiency tests for regression model [3] [7]

The judgment is based on the interpretive efficiency of the model through some tests that determine the efficiency and quality of the model or not.

The coefficient dimension of selection  $R^2 = r^2$  which takes the following relationship:

$$"R^2 = r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} " \quad \dots\dots\dots (8)$$

$R^2$  can be calculated in another format as follows:

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \quad \dots\dots\dots (9)$$

$R^2$  Identification coefficient indicates the amount of deviations or changes that occur in  $y_i$  Response variable that is interpreted by the independent variable changes  $x_i$ . The high rate of the coefficient indicates the quality and superiority of the model of the explanatory force of the phenomenon.

The coefficient of limitation is an identified value and belongs to the next scope.

$$0 \leq R^2 \leq 1$$

The closer the value of the coefficient of selection than the one the greater the responsiveness of the model in the statement that the independent variable  $x_i$  Is the one that explains the phenomenon  $y_i$ .

- The null hypothesis ( $H_0$ ) and Alternative hypothesis ( $H_1$ ): These hypotheses can test the submitted hypothesis which is about the parameters of the model  $\beta_0$  And  $\beta_1$ , the null hypothesis ( $H_0$ ) suggests that there is no effect of the independent variable of the model since of the parameters of the model equals zero or equal to another number. The relationship between the variables X and Y is based on the linear model. Here, when this relationship is lacking, the society regression line is a horizontal one, that is:

On the other hand, there is another ( $H_0 : \beta_1 = 0$ ) hypothesis which are called alternative hypothesis: ( $H_1 : \beta_1 \neq 0$ ) The distribution of T statistic is as

$$T_{calc} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}}$$

follows:

Acceptance or rejection of the hypothesis is determined as follows:

If  $\left| \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} \right| \leq t_{n-2, \frac{\alpha}{2}}$   $H_0$  is accepted at a significance level ( $\% \alpha$ )  $0 = \beta_1$ .

If  $\left| \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} \right| \leq t_{n-2, \frac{\alpha}{2}}$   $H_0$  is rejected at a significance level

(%  $\alpha$ ).  $\beta_1$  At a significant level, statistical and not equal to zero with the same test and  $\beta_0$  Parameter. In the F statistic, we can test the overall morbidity and the effect of the independent variable of the simple linear regression model which is described as follows:

$$F = \frac{ESS/1}{RSS/(n-2)} \sim F_{1,n-2} \dots \dots \dots (10)$$

These quantitative methods will be applied in detail in the applied side and reviewed by analyzing the quantitative data obtained by the researcher, through which the quality and strength of the model is determined in the interpretation of the relationship between the variables within the model.

*Applied Side*

The statistical analysis will be carried out through the construction of simple linear and multiple regression models (quadratic and cubic) with a comparison between these models and determine what is better in the interpretation of the phenomenon of fish death I the province of Babylon according to statistical standards. The data of dead fish were collected from the Directorate of Agriculture of Babylon, Department of Animal Resources, Fish section, where the data included the quantities of dead fish and the number of farms where the phenomenon occurred distributed to the agricultural people of the Directorate of Agriculture of Babylon as follows:

Table (1) dead fish in Babylon province

Agricultural branches	Quantities of dead fish	Number of farms
Al – Kifil division	455 Ton	7
Abu Gharaq division	125 Ton	1
Division of the center	138 Ton	5
Hamza al-Garbi	93 Ton	3
Hashemia	56 Ton	8
Al – Shoumaly	21 Ton	1
Al – Taleaa	7 Ton	1
Al- kassim	6.48 Ton	1
Al- Mahaweel	103.2 Ton	5
Al- Exandria	858.3 Ton	22
Al – Musaiab	1737 Ton	16
Al -Sadah	627 Ton	12

VII. LINEAR REGRESSION MODELS OF THE VARIABLE QUANTITIES OF DEAD FISH

The simple linear regression model and the linear regression model of the second and third classes were constructed. The variable used in the quantities of dead fish

in Babylon province was defined by 12 observations representing the agricultural parts in the governorate after their order to match the method of analysis and its relation to the number of farms where the phenomenon occurred as an independent variable Express the growth of the phenomenon under study

*First: the simple linear model*

The model of simple linear regression between the quantity of dead fish in the Babylon Governorate, as a dependent variable  $dth_i$  and the number of farms  $f_i$  as an independent explanatory variable affects the quantity of dead fish. According to the statistical theory, the errors of the model are independent and do not follow any statistical model linking errors. In order to be the basis in the process of estimation and construction of models, scientifically basis, where the model was estimated through the method of squares lower with the extraction of statistical tests of the parameters and model with the efficiency measures of the model and the results are as follows:

*The regression equation is*

$$dth = - 66.1 + 61.22 f$$

S = 323.996 R-sq = 64.08% R-sq(adj) = 60.49%

The variance analysis table for the regression model was as follows:

*Analysis of Variance*

Source	DF	SS	MS	F	P
1	18729551	1872955	17.84	0.002	
Regression					
Error	10	1049737	104974		
Total		11	2922691		

VIII. ANALYSIS OF THE RESULTS

The results indicate that there is a strong correlation between the variable dependent quantities of dead fish in the province of Babylon with the explanatory variable of the number of fish farms with the relationship of adverse effect and this means that the increase in the number of farms leads to the high quantities of dead fish in the province, The explanatory power of the model is medium, with a coefficient of 64.08%. The independent variable has explained the changes in the variant of dead fish in the Babylon province by an approximate 64%. The estimated model as a whole is statistically acceptable, F, which is 17.84 and its probability,  $p = 0.002$ , of the estimated model, which is well below 0.05, which confirms the significance of the simple linear regression model as a whole. This explains the possibility of relying on its results in explaining the relationship between the quantities of dead fish in Babylon province, The parameter of the independent variable is positive and significant and its value 61.22 means that the increase in the number of fish farms by one unit leads to an increase in the amount of dead fish in the province by an amount equal to the estimated model parameter 61.22, and Figure (1) shows the estimated model results.

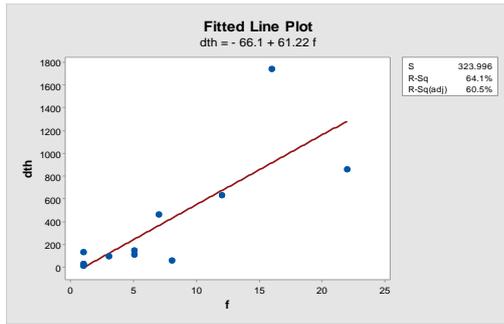


Figure (1) Model of the simple linear trend of the variable quantity of dead fish

### IX. BEHAVIOR OF SIMPLE LINEAR MODEL ERRORS

In order to ensure the efficiency of the quadratic regression model, it is possible to deal with model errors and draw them using a (normal probability plot). We find that the model's errors are very close to the trend line, indicating acceptable quality of the model and lack of extreme values, The beginning of the data or at the end of the data and therefore the quality of the model is relatively acceptable and the distribution of errors through the (Histogram) is not separated from the form of physical distribution Jersi, and the errors of the model against the estimated values are close to zero and around.

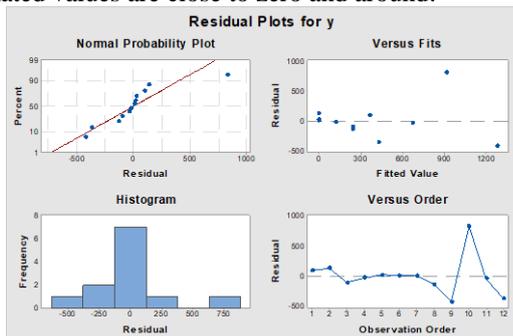


Figure (2) Behavior of the errors of the simple linear model of the variable quantities of dead fish in the province of Babylon

Second: Quadratic Model: The quadratic model was constructed from the second degree to describe the relationship between the mortality of fish in the Babylon governorate as a dependent variable  $dth_i$  and the number of dead fish farms as an explanatory variable  $f_i$ . The results were as follows:

#### Polynomial Regression Analysis

The regression equation is:

$$dth = -132.9 + 87.16 f - 1.250 f^2$$

$S = 336.467$        $R\text{-sq} = 65.14\%$        $R\text{-sq(adjust)} = 57.39\%$

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	1903799	951900	8.41	0.002
Error	9	1018892	113210		
Total	11	2922691			

### X. ANALYSIS OF THE RESULTS

The results indicate that there is a correlation between the dependent variable and the quantities of dead fish in Babylon province with the explanatory variable represented in the number of farms with the relationship of the positive effect with the number of farms. This means that the increase of the number of fish farms by one unit leads to the increase of the dead fish in the governorate about 87.16. The values of the coefficient of determination indicate that the explanatory power of the model is much higher than that of the simple linear model, with a coefficient of 65.14%, indicating that the independent variable number of fish farms has explained the changes in the variable dead fish about 65.14%. The estimated model as a whole is statistically acceptable and according to the value of the test F of 8.41 and its probability,  $p = 0.009$  for the estimated model, which is less than 0.05, which confirms the significance of the linear regression model as a whole in statistical terms. That means it could be depend on the results to explain the dead fish quantity with the fish farms. The other variable, which represents the square number of farms, has an adverse effect on the quantities of fish because it is negative as the parameter is estimated at 1.250, and this effect is due to the decrease in the quantities of dead fish and the quadratic model in terms of Or much better than the simple linear model, but it still does not explain the majority of the changes in variable quantities of dead fish, and Figure 3 shows the results of the estimated model by comparing the estimated with the original data of the time series quantities of dead fish model in the province of Babylon.

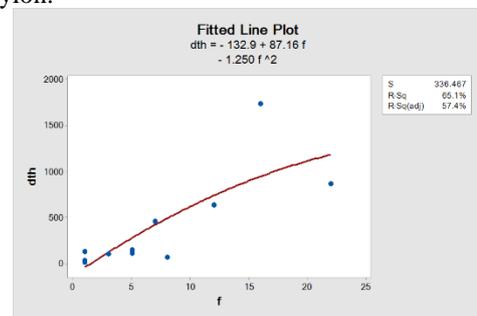


Figure (3) Model of the linear trend of the variable dead fish in the province of Babylon

A sequential analysis of the variance of the quadratic model was carried out according to its linear and quadratic components. The results were as follows

#### Sequential Analysis of Variance

Source	DF	SS	F	p
Linear	1	1872955	17.84	0.002
Quadratic	1	30845	0.27	0.614

We find the mean of the squared form of the estimated model according to the extracted value of 0.002 p

### XI. BEHAVIOR OF SQUARE MODEL ERRORS

In order to ensure the efficiency of the quadratic regression model, it is possible to deal with model errors and draw

them using a normal probability plot. We find that the model's errors are very close to the trend line, indicating acceptable quality of the model and lack of extreme values, The beginning of the data or at the end of the data and therefore the quality of the model is relatively acceptable and the distribution of errors through the (Histogram) is not separated from the form of physical distribution Jersi, and the errors of the model against the estimated values are close to zero and around.

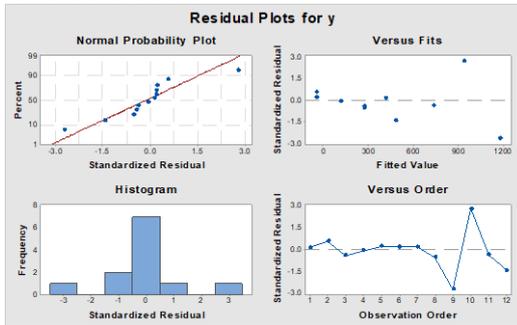


Figure (4) Behavior of the errors of the estimated quadratic model of the variable quantities of dead fish in Babylon province

In this way, it is possible to infer that the model is an acceptable explanation of the quantities of dead fish. The model has been improved to increase its explanatory power and to withdraw the remaining information in the errors. The cuboids model was constructed from the third degree to increase the accuracy of the model.

*Third: Cubic model:* The cuboid model was constructed from the third degree between the quantities of dead fish in the Babylon province as a modified change  $dth_i$  and the preparation of the fish farms as an explanatory variable (independent)  $f_i$ . The aim was to improve the results and reach the preference. The results of the cubism model were as follows :

*Polynomial Regression Analysis*

The regression equation is

$$dth = 219.5 - 171.0 f + 31.71 f^2 - 1.022 f^3$$

$$R\text{-sq}(\text{adj}) = 80.37\% \quad R\text{-sq} = 85.73\% \quad S = 228.355$$

And the table of variance analysis of the regression model

*Analysis of Variance*

Source	DF	SS	MS	F	P
Regression	3	2505522	835174	16.02	0.001
Error	8	417169	52146		
Total	11	2922691			

*Analysis of the results*

The model is generally acceptable from the statistical point of view according to the F test of 16.02 and the probability of 0.001 compared with the quadratic model. It is statistically acceptable and has improved the results more acceptable. The value of the selection factor was 85.73% Analysis, Sequential variation where the results were as follows:

*Sequential Analysis of Variance*

Source	DF	SS	F	P
Linear	1	1872955	17.84	0.002
Quadratic	1	30845	0.27	0.614
Cubic	1	601723	11.54	0.009

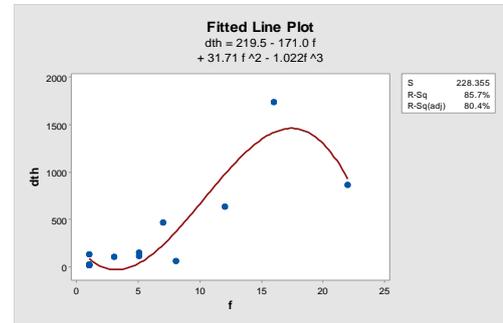


Figure (5) The estimated cubic model of the variable quantities of dead fish in the province of Babylon

**XII. BEHAVIOR OF CUBIC MODEL ERRORS**

In order to ensure the efficiency of the cuboid regression model, it is possible to deal with model errors and draw them through the normal probability plot. We find that the errors of the model (estimated model locks) are approaching the trend line, indicating the high quality of the model compared to its counterparts in the previous models. Either at the beginning or the end of the data. Thus, the quality of the model is relatively high. The distribution of errors through the histogram is not dissimilar to the normal physical distribution. The errors of the model versus the estimated values are close to and around zero.

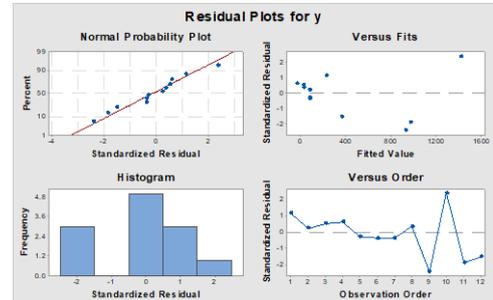


Figure (6): Behavior of errors of the estimated cubic model of the variable quantities of dead fish in Babil Governorate

**XIII. CONCLUSIONS**

The researcher came to a number of conclusions after the theoretical study and then the process, which included the construction of three models of the regression (simple and quadratic and cubes) of the variable quantities of dead fish as a dependent variable (certified) and the number of fish farms as an independent explanatory variable where he came out with the following conclusions:

- 1-The simple linear model was statistically acceptable, with a coefficient of 64.08% and a probability of 0.002, which is less than 0.05 according to test **F**. This confirms the acceptance of the model in terms of statistics. However, the researcher went on to construct the quadratic model to obtain a better model in explaining the phenomenon under study.
- 2- The quadratic model according to the statistical measures was acceptable, but it was not at a higher level

than the previous one. The coefficient of determination of the quadratic model of 65.14% was not much closer to the model of the simple model. Therefore, it did not improve the results so the researcher resorted to constructing a cubistic model to obtain higher results. He was able to compare the three models. The value of the cubism model was 85.73%. According to the F test of 16.02 and the probability of 0.001, which is much lower than 0.05, it was found that the cube model was the best statistical model, although it is often preferred Model of the square and ease meaningful, but it and by the coefficient of determination and the value of one approach, it is a high quality and favorable compared to the rest of the previous models, indicating the gradual decline in the case of mortality.

3-One of the most important conclusions reached by the researcher is that the explanatory variable (independent) is the number of fish farms, which clearly affect the model on the high and low quantities of dead fish as the dependent variable (approved), which were the basis of building three statistical models.

#### XIV. RECOMMENDATIONS

1 - The difficulty of obtaining data was one of the most important obstacles faced by the researcher in the preparation of the study as the lack of disclosure extensively by the official bodies represented by the Directorate of the cultivation of Babylon, despite following the official ways to request data made the researcher has difficulty in providing more research and studies. The statistics concerned with this field.

2. The data were distributed according to the quantities of weights and according to the agricultural divisions of the Directorate and the number of farms within the geographical area, but they are not disaggregated by type and weights, so the recommendation to give the researcher more space to work and provide statistical research on the livestock sector in general and fish on In particular.

3 - The need to direct the agricultural divisions and departments of the Directorate to create a database in accordance with statistical methods of scientific data classification of the livestock sector and away from the informal paper supervision which is not organized well.

4- The need to make recommendations to the institutions of the Supreme State to develop a mechanism of work is working to increase the water releases of rivers Tigris and Euphrates and streams and rivers belonging to them if the reduction of water levels and scarcity and high salinity was a major cause of the phenomenon of death of large quantities of fish, causing loss And damaged an important wealth and is considered one of the most important components of the food basket for the country's population.

5-Work on the development of the agricultural extension program to raise awareness of fish fishes in maintaining the health distance between fish farms and not to increase them at random.

6 - The need to adopt this study and the models that were built in the statistical analysis for the development of health curricula for the breeding model and provide an environment suitable for the return of fish farms to work better.

#### REFERENCES

- [1] Al-Ja'ouni, Faridkhilil, "Multiple Linear Regression Analysis Method in the Study of the Most Important Social, Economic and Demographic Variables Affecting Total Birth Rate" Applied Study of Human Development Report 2006 in 177 Countries, Damascus University, Faculty of Economics, 2006.
- [2] Al-Jiboory, Shalal Habeeb and Abid, " Multiple Regression Analysis", Baghdad: Directorate of Dar A l-Kutub for Printing and Puolishing, 2000.
- [3] Al-Tamimi, Raad Fadhil, Regression and Time Series, Advanced Statistical Methods of Application Using the Minitab system, Al Jazeera, Baghdad, 2013.
- [4] Naji, Ali Abdul Karim Kazem "(Statistical analysis of the determinants of mortality for children under the age of six 2010-2017 in Babylon" (case study)", College of Education Sciences Pure Babylon University Diploma study, 2018.
- [5] Zubi, Mohammed Bilal, Talafah, Abbas "Statistical System SPSS Understanding and Analysis of Statistical Data", Jordan, Dar Wael, 2012.
- [6] Norusis, M., (1986), "UserGide Spss/pc," Chicago (manual), Spss Manual 7.5, "Advanced Statistical", NC, 1997.
- [7] Morrison, D.F., "Multivariate Statistical Method", McGraw Hill, New York, 1976.
- [8] Michael H. Kutner , William Li " Applied Linear Statistical models" McGraw- Hili, 2005.