Journal of Engineering and Applied Sciences 13 (Special Issue 13): 10602-10608, 2018
ISSN: 1816-949X


# Evaluation of the Birth Rate of Children in Babil Governorate for the Period 2010-2017 by using Network Security through Linear Regression Models/Iraq

Asmaa Shaker Ashoor and Ali Abdul Karim Kazem Naji
College of Education for Pure Science, University of Babylon, Hillah, Iraq

**Abstract:** The present study is about the statistical determinants for children's births in Babil Governorate by a process which contains several steps that based on data basics for a certain period of time (2010-2018) and they have been collected as follows: problem determination data collection data tab and data analysis. This is done by applying statistical tools and (SPSS). The steps, statistical study and birth's analysis determinants are scientifically correlated to techniques and methods used by scientists and researchers that they preceded usin these fields. There is no doubt that we need reliable data to conduct the research as well as the period of time to study the behavior of births in this governorate in terms of height, decline and knowledge of the reasons behind it.

**Key words:** Linear regression, statistical, evaluate, sequential analysis, estimated model, SPSS

## INTRODUCTION

Children are the real wealth of any society because they are the pillar of societies and they are the leaders of the future. Also, the increase in the birth rate and the lack of it is one of the most important in elements of tripartite population change (fertility, mortality, migration). Fertility is one of the most important elements of natural population growth. Therefore, we need to develop extensive studies on the nature of the increase in birth rates or decrease and try to explain the difference between birth rates through a scientific and systematic statistical process.

Disruption of health and preventive education programs on mothers and non-follow-up, including the examination of parents before and during pregnancy and the lack of primary health care programs supposed to be provided by the government and the spread of epidemics, diseases, low standard of living for the family and other reasons and factors that lead to a large disparity in birth rates in the governorate of Babil. In order to provide important information for this study, data on births were collected from the Statistics section of the Health Department of Babil as well as the Statistics section of maternity hospitals and public hospitals in the governorate. Through this information, the researcher looks for the factors that affect on the increase and decrease of births. The simple and multiple regression equations were used in two dimensions. The data collected from the governorate of Babylon and they have been studied. This research includes two aspects. The first is the theoretical aspect of the presentation of statistical models, namely, simple and multiple linear regression models and the focus on the mechanism of constructing these models (quadratic and cubic) with the means of comparison between them and the best of these models. The second is the practical side of constructing statistical models and total database of births for the Babylon Governorate using statistical analysis application (Mahfouz, 2008; Al-Baldawi, 2009) (Version 17 SPSS, Minitab Version 24).

### Literature review

**Linear regression model assumptions (Khlil, 2006; Al-Jubouri and Abd, 2000):** Simple linear regression is the simplest type of regression models used to describe many relationships between variables, for example, the relationship of Babil Governorate births with time or another variable. The model represents a relationship between two variables, one is a dependent variable and its value is determined by the inner inside of the model and the other a variable is independently which determines its value outside of the form. The form is submitted to a set of basic assumptions about the random error and is based on the normal distribution. It is as follows: the random variable is associated with $\varepsilon_i$ a model which is independent of the independent variable $x_i$, it means that the common variation between the independent variable and the error is zero:

$$\mathrm{Cov}(X_i, \varepsilon_i) = 0, \forall_i = 1, ..., n$$

**Corresponding Author:** Asmaa Shaker Ashoor, College of Education for Pure Science, University of Babylon, Hillah, Iraq

Random error variable is $\varepsilon_i$ normally distributed, it represents measurement errors, irregular errors resulting from data behavior and other errors. The mean of the random variable $\varepsilon_i$ is assumed to be zero:

$$E(\varepsilon_i) = 0, \forall_i = 1, ..., n$$

The random error of any view is independent of other views. The random variable $\varepsilon_i$ has a constant variation $\sigma^2$

$$Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2, \forall_i = 1, ..., n$$

## MATERIALS AND METHODS

For the purpose of birth's data collecting, the researcher was provided with a statement for the purpose of facilitating his mission from the Deanship of the Faculty of Education for Pure Sciences at University of Babylon entitled to the Department of Babylon Health and also to the hospitals of the governorate. Data relating to the number of births were collected according to the sample of the time under study and by 8 annual observations. The linear regression model (simple, quadratic, cubic) was used to evaluate birth rates in the governorate. Data were analyzed using the statistical analysis program (SPSS Version 24).

**Method of Ordinary Lower Squares (OLS) (Fadel, 2013; Timm, 1975):** One method of estimation is the Ordinary Lower Squares (OLS) method and the maximum possibility method which aims to obtain estimation formulas for the parameters of the model.The statistical relationship of the variables $X_i$ and $Y_i$ can be described as follows Eq. 1:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, 2, 3, ..., n \tag{1}$$

Where:

$x_i$ = The dependent variable

$y_i$ = The independent variable

$\beta_0$ and $\beta_1$ are two parameter model and graphically represented the followings: $\beta_0$ represents the fraction by the y-axis which reflects the value of the dependent variable in case the value of the independent variable is missing that is the case, x = 0 (Fig. 1).

As for $\beta_1$ is Straight line inclination ($\beta_0+\beta_1 x$). The amount of change is reflected if it changes in one unit and the change may be increased if the inclination signal is positive or downward if the inclination is negative. As for $\varepsilon_i$ represents the error in interpretation of $Y_i$ and it can be written out of the relationship in Eq. 2:

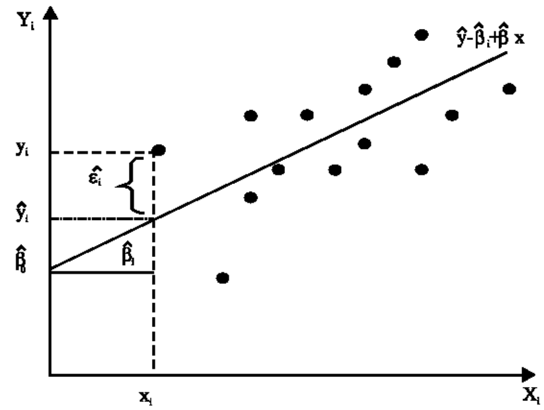

Fig. 1: $\beta_0$ and $\beta_1$ are two parameter model

$$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i) \tag{2}$$

The process of estimating the parameters of the model using the method of squares is smaller as this method tries to find the best linear model by reducing the total squares of deviation (between the actual views and estimated) (Eq. 3):

$$s = \sum_{i=1}^{n} e_i^2 \tag{3}$$

Where:

$$\hat{\varepsilon} = Y_i - \hat{Y}_i$$

This can be written mathematically by:

$$Min \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \underset{\hat{\beta}_0,\hat{\beta}_1}{Min} \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2 \tag{4}$$

Regression coefficients ($\beta_1$, $\beta_0$) can be estimated in the model using the lower squares method and this estimate is the sum of the squares of random errors (Eq. 5):

$$\sum e^2 = \sum \left(y - (\beta_0 + \beta_1 X)\right)^2 \tag{5}$$

In order to estimate the parameters of the model, the sum of the error boxes is minimized by the partial derivation of the parameters of the linear model and their equivalents by zero as follows (Eq. 6):

$$\begin{aligned} \frac{\partial f(b_0, b_1)}{\partial b_0} &= \sum_{i=1}^{n} 2(y_i - (b_0 + b_1 x_i))(-1) = 0 \\ \frac{\partial f(b_0, b_1)}{\partial b_1} &= \sum_{i=1}^{n} 2(y_i - (b_0 + b_1 x_i))(-x_i) = 0 \end{aligned} \tag{6}$$

The application of some calculations for the purpose of getting natural (Eq. 7):

$$nb_0 + \left(\sum x_i\right) b_1 = \sum y_i$$
$$\left(\sum x_i\right) b_0 + \left(\sum x_i^2\right) b_1 = \sum x_i y_i \tag{7}$$

Linear equations with unknown parameters $b_0$ and $b_1$ and provide us with at least two different values $x_i$'s. The capabilities of the lower squares represent the only solution to the system of linear equations. In solving these equations, we obtain the estimated parameters of the slope and the next fixed limit (Eq. 8):

$$b_1 = \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s_{xy}}{s_{xx}} \tag{8}$$

As for:

$$s_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}, \quad s_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} \tag{9}$$

Thus, the tendency is equal to Eq. 10:

$$b_1 = \hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \tag{10}$$

The fixed limit is calculated from its relation to the tendency in relation (Eq. 11):

$$b_0 = \frac{\sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i}{n}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{11}$$

Because of the natural distribution hypothesis, $\hat{\beta}_0$ and $\hat{\beta}_1$ are also consider the greatest potential. As for $\bar{x}$ is the mean of the values of the independent variable $\bar{y}$, x is the mean of the dependent variable values y, the mean of the dependent variable values is $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, this estimate is called "regression equation estimation y on x.

**Some regression model tests (Morrison, 1976; Muhammad, 1988):** The explanatory power of the model is tested through the coefficient of the determination coefficient $R^2 = r^2$ according to the following relation:

$$R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \tag{12}$$

$R^2$ is one of the most important factors that measure the correlation between two variables and the existence of such a relationship implies that one of these variables depends on the change on the other variable. The coefficient of determination is a definite value and belongs to the following field: $0 \le R^2 \le 1$.

The coefficient of value determination ranges between 0 and 1 as the fundamental difference between the coefficient of determination and correlation coefficient lies in causality where correlation measures the relationship between two variables, regardless of the role played by each variable. The coefficient of determination also measures the correlation but takes into account causation as the variable $X_i$ is the one explaining the phenomenon $Y_i$. The overall significance test of the simple linear regression model or the effect of the independent variable $X_i(H_0: \beta_1 = 0)$. It can be shown through the table of analysis of variance according to the following statistical formula:

$$F = \frac{ESS/1}{RSS/(n-2)} \sim F_{1,n-2} \tag{13}$$

As for ESS is the Sum of the Error Squares of Analysis of the Variance Table) (ANOVA) RSS represents the sum of squares interpreted through the linear regression model. The test can be described by its relation to the limiting factor according to the following relationship:

$$F = \frac{R^2/1}{\left(1-R^2\right)/(n-2)} = \frac{R^2}{\left(1-R^2\right)}.(n-2) \sim F_{1,n-2} \tag{14}$$

It is through the relationship with the distribution F, all of the parameters of the model are tested if the model is equal to zero at the same time against the hypothesis of the significance of the tendency. The calculated value is compared with the tabular value of a statistic F with two degrees of freedom 1 (in the numerator) and n-2 (in the denumerator). These quantitative methods must be addressed in detail and through the analysis of health data in Babil Governorate. It can be an independent variable $x_i$ which expresses time (months, years, ...,) then, the model is called the linear trend model.

## RESULTS AND DISCUSSION

**Practical part:** Statistical analysis will be done by building a simple linear and multiple regression models to form the cube and square with the comparison between models and draw the best models according to statistical measurements where data have been collected from the

Table 1: The number of births in the Governorate of Babylon

| No. of births | Years |
|---|---|
| 60517 | 2010 |
| 61741 | 2011 |
| 61887 | 2012 |
| 62532 | 2013 |
| 66544 | 2014 |
| 62119 | 2015 |
| 57774 | 2016 |
| 53449 | 2017 |

Table 2: Analysis of variance

| Source | df | SS | MS | f-values | p-values |
|---|---|---|---|---|---|
| Regression | 1 | 24842545 | 24842545 | 1.91 | 0.217 |
| Error | 6 | 78228271 | 13038045 | | |
| Total | 7 | 103070816 | | | |

Ministry of Health/Department of Health of Babylon and after tab data by the researcher was the data in a Table 1.

**Linear regression models for total number of births:** Simple linear regression models were built and the linear regression model of second and third class where the variable description adopted by the total number of births in the Governorate of Babylon for the years (2010-2017) with 7 annual views compiled from the Ministry of Health Data rearranged to suit analysis method of the gradient as shown in Table 1. Its relation to time as independent variable which represents direction and growth of births in the governorate during the study period.

The simple linear regression model (general trend) was constructed between the total number of births in Babil Governorate as a dependent variable, bar, and time y, as an independent explanatory variable affects the number of births on the hypothesis and according to the statistical theory that the errors of the model are independent and donot follow any statistical model linking the errors with each other and the distribution of errors follows the natural distribution to be the basis in the process of estimation and construction of models of scientific basis where the model was estimated through the method of lower squares with the extraction of statistical tests of the parameters and model with the efficiency measures of the model and the results were as follows: the regression equation is:

$$\text{bar} = 64281 - 769.1\ y$$

$$S = 3610.82,\ R^2 = 24.1\%,\ R^2(\text{adj}) = 11.5\%$$

Table 2 of variance analysis of the regression model. The results indicate that there are very weak correlation between the dependent variable and the number of births in Babil Governorate with the time-variant variable and the effect of the reverse effect.

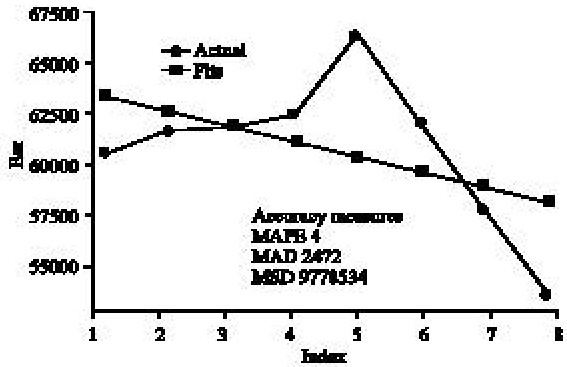


Fig. 2: The general trend model of the variable number of births (Linear Trend Model bar = 64281-769xy

This means that the increase in time leads to a decrease in the number of births in the governorate. The 24.2%, indicating that the independent time variable has explained the changes in the birth rate variable in Babil Governorate by an approximate rate of 24%. The estimated model as a whole is statistically unacceptable and the F test value of 1.91 and its probability is $p = 0.217$ for the estimated model which is much $\geq 0.05$ and confirms the insignificance of the simple linear regression model as a whole. This explains the fact that it is not possible to rely on its results ininterpreting the relationship between the number of births in Babil Governorate from its relationship with time. In addition, the parameter of the independent variable (time) isnegative, insignificant and is worth 769.1. It means that the increase in time by one unit (1 year) leads to a decrease in the number of births in the governorate which is equal to the estimated model parameter of 769.1. Figure 2 shows the results of the estimated model that the deviation of the estimated model can be observed from the description of the real data. So, it is considered a weak model.

The efficiency measures of the model were calculated as the mean absolute percentage of errors and the mean absolute deviations, the mean error squares. They are as follows:

| Accuracy Measures | |
|---|---|
| MAPE | 4 |
| MAD | 2472 |
| MSD | 9778534 |

**Square model:** Square model was built from a second class to describe the relationship between the approved variable to the number of birthsin the Governorate of Babylon, the independent variable and time form results were as follows:

Table 3: Analysis of variance of regression model

| Source | df | SS | MS | F-values | p-values |
|---|---|---|---|---|---|
| Regression | 2 | 84510490 | 42255245 | 11.38 | 0.014 |
| Error | 5 | 18560326 | 3712065 | | |
| Total | 7 | 1030708 | | | |

**Polynomial regression analysis: bar versus y:** The regression equation is:

$$bar = 55342 + 4595y - 596.0y^2$$
$$S = 1926.67 R^2 = 82.0\% R^2(adj) = 74.8\%$$

Table 3 of variance analysis of the regression model. The results indicate the existence of a correlation between the dependent variable (number of births in the Governorate of Babil with explanatory variable, direct impact relationship and time. That means the increased time in one time unit takes the highest number of births in the governorate by the estimated parameter 4595. The values of the coefficient of determination indicate that the explanatory power of the model has increased significantly compared to the simple linear model with a coefficient of 82.0% indicating that the independent time variable has interpreted the changes in the birth count by 82% and the estimated model as a whole is statistically acceptable. According to the F-value of 11.38 and its probability, p = 0.014 for the estimated model which is<0.05 which confirmsthe significance of the linear regression model as a whole in terms of statistics. This explains the possibility of relying on the results in the interpretation of the number of births in time while the other variable which represents the time box, its effect reverse on the number of births because the negative indication reached the estimated parameter 596. This effect is intended to decrease the number of births and that the quadratic model in general is much better than the linear model but still does not explain the majority of changes in the number of births. Figure 3 shows the estimated model results by comparing the estimated model with the original data for the time series of numbers of births in the Governorate of Babil. Efficiency measures were as follows:

| Accuracy Measures | |
|---|---|
| MAPE | 2 |
| MAD | 1242 |
| MSD | 2320041 |

It is smaller than its counterparts in the simple linear model. A sequential analysis of the variance of the quadratic model was carried out according to its linear and quadratic components and the results were as follows Table 4.

Table 4: Linear and quadratic components

| Source | df | SS | F-values | p-values |
|---|---|---|---|---|
| Linear | 1 | 24842545 | 01.91 | 0.217 |
| Quadratic | 1 | 59667944 | 16.07 | 0.010 |

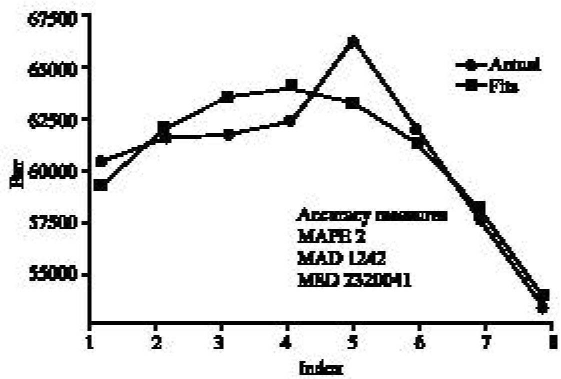


Fig. 3: The general estimated trend of the number of births (Quadratic trend Mdel bar = 55342+4595xy-596xy$^2$

The significance of the linear composition of the model was not significant by which p-values showed that their values were >0.05 with a significant difference to the quadratic composition. To ensure the efficiency of the model has been dealing with errors of the model in several formatsthrough a normal probability plot, it is foundthat the form errors (estimated model protector) be close to much by the direction line indicating the quality of form and the absence of extreme values either at the beginning or end of data and hence, the estimated model quality high. Hence, the quality of the estimated model is as high as the distribution of errors through (Histogram). It is not dissociated from the normal form of physical distribution and model errors versus estimated values are close to and around zero (Mahfouz, 2008; Al-Baldawi, 2009) (Fig. 4).

In addition, the model has been well explained and improved to raise the explanatory power and to withdraw the remaining errors in the information. A third class cubic model has been constructed to increase the accuracy of the model. The cube model (Third Class): Cube model is constructed from third class to describe the relationship between the approved variable to the number of births in the governorate of Babil and the independent variable and time.The main goal is to improve the model and achievethe optimization. It is as follows: the regression equation is:

$$bar = 60755 - 1037y + 880.2y^2 - 109.3y^3$$
$$S = 1692.49 R^2 = 88.9\% R^2(adj) = 80.5\%$$

Table 5 of variance analysis of the regression model. The model constitutes a generally accepted statistical

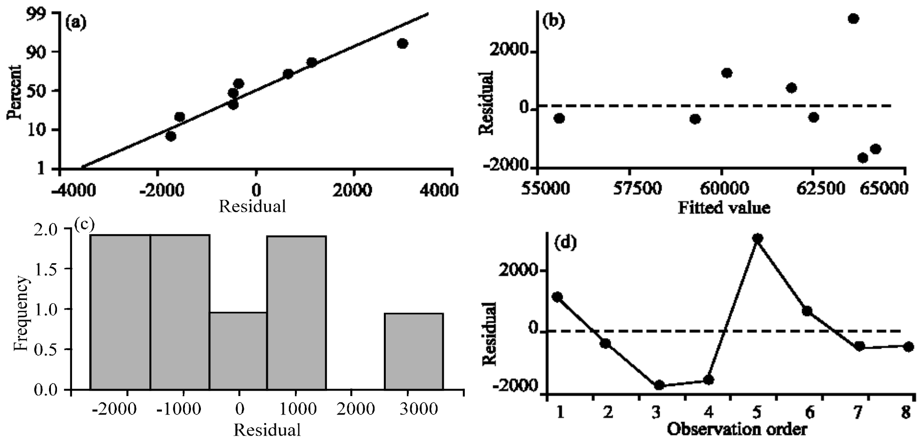


Fig. 4: Behavior of errors of the estimated quadratic model of the variable number of births: a) Normal Probability plot; b) Versus fits; c) Histogram and d) Versus order

Table 5: Analysis of variance according to optimization

| Source | df | SS | MS | F-values | p-values |
|---|---|---|---|---|---|
| Regression | 3 | 91612674 | 30537558 | 10.66 | 0.022 |
| Error | 4 | 11458142 | 2864536 | | |
| Total | 7 | 103070816 | | | |

Table 6: Sequential analysis of variance

| Source | df | SS | F-values | p-values |
|---|---|---|---|---|
| Linear | 1 | 24842545 | 01.91 | 0.217 |
| Quadratic | 1 | 59667944 | 16.07 | 0.010 |
| Cubic | 1 | 7102184 | 02.48 | 0.190 |

and according to F test. It was 10.66 with a probability of 0.022. In comparison with the quadratic model, the estimated model improved the results to the approximation of the values of the coefficient of selection from the (one) at 88.9%. For the sake of the model test, the results were as follows Table 5.

**Sequential analysis of variance:** The results indicate the significance of the quadratic composition and the insignificance of the linear and cubism according to F test and the probability value associated with it. The data for the series of birth were drawn with the regression line of the cuboid model and the results were as follows (Table 6 and Fig. 5).

After the theoretical and applied study on the data of the number of births that included the construction of multiple models, the researcher reached the following conclusions: the estimated model of births was quadratic and cubic that these models converge in terms of interpretation of births with selection coefficient values that vary between 82 and 88%, since, births were rising until 2014 and began a rapid decline by the year 2015. The quadratic model is preferred for not being complicated statistical model, the decline reflects the difficulties of economic life and reluctance of marriage and reproduction that affects the births.

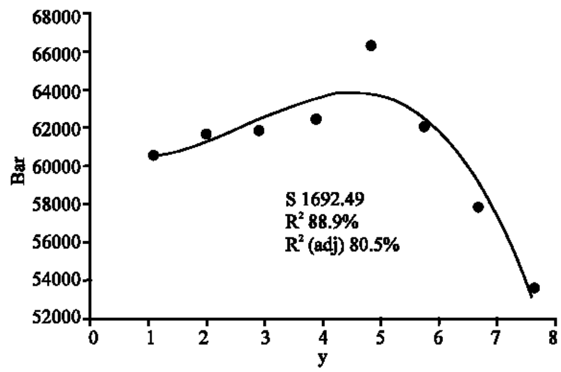


Fig. 5: The estimated general cubism trend model of the variable number of births

## CONCLUSION

The estimated model was the quadratic and cubic model that these two models converged in terms of the interpretation of births with the values of the coefficient of selection and ranged from 28-88% birth to the rise until 2014 and the rapid decline began in 2015. It is preferable to use the quadratic model because of the complexity of its statistical form and the decline reflects the difficulties of economic life and the reluctance to marriage and reproduction which affects births.

## REFERENCES

Al-Baldawi, A.A., 2009. Methods of Statistics for Economic Sciences and Business Administration with the use of SPSS. Dar Wael for Publishing, Amman, Jordan.

Al-Jubouri, S.H. and S.H. Abd, 2000. Multivariate Analysis. Directorate of Dar Al Kutub for Printing and Publishing, Baghdad, Iraq.

Fadel, T.R., 2013. Regression and time series, advanced statistical methods by using Minitab system. Al Jazeera Media Company, Baghdad, Iraq..

Khlil, J.F., 2006. Multiple linear regression analysis method in studying the most important socio-economic and demographic variables affecting total birth rate. Master Thesis, Department of Statistics, University of Damascus, yDamascus, Syria.

Mahfouz, J., 2008. The Book of Advanced Statistical Analysis using SPSS. Dar Wael for Publishing, Oman, Middle East.

Morisson, D.F., 1976. Multivariate Statistical Methods. McGraw-Hill Book Co. Inc., New Yark.

Muhammad, D.M., 1988. Book of Linear Regression Analysis. Dar Osama for Publishing, Jordan, Amman..

Timm, N.H., 1975. Multivariate Analysis with Applications in Education and Psychology. 1st Edn., Brooks, Cole, Montery, California, ISBN-13: 9780818500961.