

The effect of some independent variables on granting the basic application certificate using multivariate statistical analysis

Assist. Lecturer: Aasha A. Ismael
College of Administration and Economics
Tikrit University

Assist. Lecturer: Zainab A. A. Albairmani
Presidency University
University of Babylon

Abstract:

In this research we study of one of the methods of regression using real data for applicants for the proficiency exam (IC₃), where the basic applications exam is most important test for the certificate of computer basics and the internet which is conducted by the university of Tikrit in the center for computer and informatics. Therefore it was necessary to know and study the impact of some independent variable and the extent of their impact on the approved variable, the score of the exam and has been taken five independent variables, this is, college specialization x_1 , the gender of the examiner x_2 , specialization advanced the examiner x_3 , computer examiner experience x_4 , subscriber in turn or completed x_5 . and the degree of the examiner (dependent variable) The regression method and estimation of regression equations were used according to the gradual regression methods and the standard regression method (Multiple Regression) in order to select the best regression model that includes the variables that actually affect the result of the certification of basic applications and by using the average error boxes differential measurement Mean square error (MSE) (Dean et al.1998).

Keywords: Independent variables, granting a certificate, applications using Multivariate
تأثير بعض المتغيرات المستقلة على منح شهادة التطبيقات الأساسية باستخدام
التحليل الإحصائي متعدد المتغيرات

م.م. زينب عبود أحمد البيرماني
رئاسة الجامعة
جامعة بابل

م.م. عائشة عبدالخالق إسماعيل
كلية الإدارة والاقتصاد
جامعة تكريت

Zanab.abboud@uobabylon.edu.iq

Aasha.A.Alkalek@tu.edu.iq

المستخلص:

يتضمن هذا البحث دراسة أحد أساليب التحليل الإحصائي متعدد المتغيرات وهو تحليل الانحدار باستخدام بيانات حقيقية للمتقدمين لامتحان الكفاءة (IC₃) حيث ان امتحان التطبيقات الأساسية هو احد اهم الاختبارات للحصول على شهادة اساسيات الحاسوب والانترنت والذي تجريه جامعة تكريت في مركز الحاسوب والمعلوماتية، لذلك كان من الضروري معرفة ودراسة تأثير بعض المتغيرات المستقلة ومدى تأثيرها في المتغير المعتمد وهو درجة الامتحان، وقد تم اخذ خمسة متغيرات مستقلة وهي تخصص الكلية x_1 ، الجنس x_2 ، التخصص المتقدم اليه الممتحن x_3 ، خبرة الممتحن بالحاسوب x_4 ، مشترك بالدورة ام مستوفي x_5 ، وقد تم استخدام اسلوب الانحدار وتقدير معادلات الانحدار وفق طريقتين هي الانحدار التدريجي وطريقة الانحدار المتعدد وذلك

لاختيار افضل نموذج انحدار يضم المتغيرات التي تؤثر فعليا على نتيجة شهادة التطبيقات الاساسية وباستخدام مقياس المفاضلة متوسط مربعات الخطأ (MSE).
الكلمات المفتاحية: المتغيرات المستقلة، منح الشهادة، التطبيقات التي تستخدم متعدد المتغيرات.

1. Introduction:

The IC3 exam includes three basic tests: Computing Fundamental Computer Principles, Key applications, and dealing with the Internet. This test was introduced to Iraqi universities to grant computer and Internet certification as one of the requirements for applying for postgraduate studies to obtain a higher diploma, master's or ph.d the researchers chose the regression analysis method to study the effect of independent variables on the adopted variable, as it is one of the statistical analyzes adopted in studying the hypothesis test to choose the influencing factors and clarify the relationship between them and the IC3 test result.

Regression analysis gives an easy and understandable way to research the functional relationships between variables and is used to describe and control data as well as estimate parameters and forecast(Bashir, 2002) Through this analysis we can choose the best model that contains variables of statistical significance in explaining the changes that occur to the dependent variable(Faisal et al., 2017)

2. Previous studies:

The statisticians followed the behavior of the various phenomena and were interested in studying the variables that affect those phenomena under study, whether they are economic, social, or scientific using multivariate statistical analysis, because of its great importance in analyzing data and obtaining good results regarding estimating the parameters of the statistical model. These studies include the following:

The research presented by researchers (Farid et al., 2007), where the research included a study of one of the multivariate statistical analysis methods is the discriminatory analysis, which is one of the advanced statistical methods in the description and distribution of families within the socio-economic structure of society and helps in drawing the economic and social development planned , the state aims to find the most appropriate methods in terms of equitable income distribution, tax burden and government subsidies for society families in a more realistic way. Where the goal of their study was to describe and distribute families and determine the factors and variables that affect their characterization within the socio-

economic structure in society, as well as trying to develop methods and techniques of statistical analysis and their application in economic studies. The researchers have demonstrated the efficiency of discrimination functions to characterize and distribute families by estimating error rates to reach a test Significance differences between societies to ensure the significance of classification and distribution of families, the study reached the possibility of using multivariate statistical analysis methods in drawing development plans in addition to arriving at the best methods that help to achieve the fair distribution of wealth in society to those who deserve it more accurately and fairly.

Also (Haitovsky, 1968) studied missing values and used a regression analysis method to estimate regression equations and obtain the best regression model with missing data. The researcher (Bushra .2008), also estimated (the effect of the interaction between independent economic variables in the multiple linear regression model), where I mentioned in the research that the regression model takes into account the effects of the interaction between the independent variables on the dependent variable and this type of interaction often occurs between independent economic variables, as the effect of one of the independent variables in the dependent variable may be dependent on the level of the other independent variable. In her research, she concluded that the interaction takes place without any doubt between the independent variables, and the effect of the reaction was estimated in several ways, where the best estimate was obtained using the orthogonal interactions method.

(Fatin, 2019) The aim of the study is to identify the explained variance ratio at the multiple regression model of different sample sizes. To achieve this, the researcher used the descriptive approach. Random samples selected from the statistical observations generated using different sample sizes which are observations of one dependent variable (Y) and ten independent variables (X1 , X2 , X3 , X4 , X5 , X6 , X7 , X8 , X9 , X10) to study the efficiency of the multiple regression model and changes in the efficiency of this model.

3. The goal:

The researchers believe that there are some variables that affect the degree of testing the basic applications (IC3), which is one of the requirements to introduce students to graduate studies. Regression analysis

will be used to obtain the best model that includes the variables that most affect the dependent variable in two ways. The first is stepwise multiple regression. The second is the standard multiple regression. We can include the research objective through the following points:

- a. Study the variables affecting students 'performance in the proficiency test (IC3).
- b. To highlight the role of statistical methods, which is the method of regression analysis in determining these variables.
- c. Comparing the regression equations of the statistical model based on the available data.

Research assumes:

- a. Null hypothesis: there is no significant effect of the independent variables on the dependent variable basic applications.
- b. Alternative hypothesis: there is a significant effect of the independent variables on the dependent variable.
- c. The ability to know and classify the set of independent variables affecting the award of the basic applications certificate (IC3) the concept of multivariate statistical analysis

4. Multivariate Statistical Analysis.

The concept of multivariate statistical analysis includes a set of specialized statistical methods and methods in data processing and analysis that govern one or more independent variables, as these methods were developed using computer technologies because some of them may need computational capabilities that cannot be performed except by modern computer devices, as it works to describe and analyze phenomena with multiple variables with its different methods. If we have a vocabulary that has the same characteristics and characteristics with different degrees, then this statistical analysis deals with studying the phenomena observations and expressing them through Impact variables influence in the related data. (Fatten, 2019: 324, Al-jaoni et al., 2007: 14).

The importance of this analysis (multivariate) is highlighted in being one of the most important tools of statistics used in the study of economic, social, natural and scientific phenomena, which is difficult to analyze its data through simple statistical methods and methods due to the large number of variables that may affect a particular phenomenon, as well as the different nature of the variables themselves, as well as what Increasing its

importance is the difference in its methods and methods, which have the ability to analyze the direction of most phenomena, which facilitates obtaining reliable results in the interpretation of the behavior of those phenomena under study.(Bashir, 2002: 55)

From the foregoing, we can say that the multivariate statistical analysis, including its statistical methods, has become a proven necessity for researchers, especially in the economic and social fields, with the aim of standing and solving problems of the phenomena under study in scientific research, whether economic or social, and more importantly, is the use of computers as an aid to apply these methods On these phenomena in order to analyze the data and obtain the results that contribute to solving the problems of the phenomena under study.(Bushra, 2008: 66)

5. Regression models:

In general, regression models are divided into two parts:

a. Simple Linear Regression:

If the relationship between the dependent and independent variables is a linear relationship, i.e. it can be expressed in a straight line called the regression line and the correlation between the two variables, either it is direct (i.e. they increase together or decrease together) or it is inverse (i.e. one of the two variables decreases when the other increases) as well as the points are close to the two variables of the slope, the greater the degree of correlation between them (Richard et al.1998:13,Matthias et al.2011:16)

The form is as follows:

$$Y = \beta_0 + \beta_1 x + \varepsilon \dots \dots (1)$$

b. Multiple Regression:

They are models that include more than one independent variable affecting the dependent variable, which cannot be treated with simple regression by the effect of all independent variables together on the dependent variable, which leads us to multiple regression. (Abaci et al., 2015: 384)

The aim of conducting the analysis of variance is to know the factors that affect the dependent variable by studying a set of data that helps in finding the parameter values and then estimating the value of the dependent variable based on the availability of the estimated values of the independent variables. (Al-Hussein, 2015: 55)

This means that any change that occurs in the independent variable leads to a change in the dependent variable, the researchers have stated that

it is necessary to differentiate between two types of relationships that link the dependent variable with independent variables, the first is called the functional relationship and this means that the change of the dependent variable number is dependent only on the variability of the variables while the second type is called the statistical relationship in which the model does not contain all the independent variables for this The reason had to include the regression model on a term called (error limit) or residuals, which measures the part of the change in the dependent variable and which is caused by independent variables not included in the model, the formula for the multiple linear regression model is as follows:

$$y_i = \beta_o + \sum_{j=1}^m \beta_j x_{ij} + \varepsilon_i \dots (2)$$

where:

$i = 1, 2, \dots, n$ (represents number of views)

$j = 1, 2, \dots, m$ (represents the number of variables)

y_i (dependent variable)

x_{ij} (View value i in explanatory variable j)

β_o (It is the slope of the intersection of the regression line with the y -axis, and when the parameter value is zero, it means that it is not present in the model).

β_j (They are the regression parameters, so that a change of one unit in the independent variables is followed by a change in the dependent variable accordingly, and the signs of the parameters may be positive or negative. If they are positive, this indicates the existence of a direct relationship between the independent variable and the dependent variable, and if the sign of the parameter is negative, this indicates the existence of a relationship Inverse between the independent variable and the dependent variable).

ε (Limit error or residuals).

It should be noted that we can use this method if the following conditions are met. (Namiq et al., 2017: 26)

- ❖ The relationship is linear between the independent variable and the dependent variable to be predicted.
- ❖ The dependent variable to be predicted must be related degrees.

- ❖ The independent variables that we use in the analysis can be at the level, and it is possible to enter variables at the nominal level, such as the type variable, which includes two categories (male and female).
- ❖ Researchers believe that the sample size should be large, so that it is more accurate to the results.

Since we will use two methods of analysis, we will look at them briefly.

First: Simultaneous method: Where its name is in the accredited program SPSS (Inter method) in which the researcher determines the set of independent variables that make up the model.

Second: Statistical method: In this method, the variables are entered or removed from the model based on the amount of correlation between them and the adopted variable. The statistical method includes three types of statistical analyzes:

- ❖ Forwards Selection.
- ❖ Backward Selection.
- ❖ Stepwise Selection.

One of the characteristics of this method is the most complex statistical method whereby independent variables are entered sequentially and the variable is evaluated in the regression model. If one of the variables is significant in the model, it is kept within the regression model. Omitted, so this method will sum up the smallest number of variables that go into the regression model and we can judge the suitability of the equation, which will be estimated by the mean square error.

$$MSE = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n - (m + 1)} \text{ -----3}$$

Whereas, the best regression equation is the one that has the smallest (MSE) equation.

6. Data collection:

The researchers collected the research data from the Computer and Informatics Center/Tikrit University, and it included data (65) students who applied to perform the basic applications exam, where the questions included (50 questions) and may include scientific and human colleges, where the research variables consist of six variables, which is the approved variable (degree of success the student) and the independent variables are:

(college specialization, gender, Specialization advanced the examiner, computer examiner experience, Subscriber in turn or completed) Where the variables were coded as in Table 1, as follows:

Table (1): Search variables

The Number	Code	Saturation	The description	Variable
65		Passing score	Dependent variable	Y
	0	Scientific	College specialization	X ₁
	1	Humanitarian		
	0	Male	Gender	X ₂
	1	Female		
	0	M.A.	Specialization for the examiner	X ₃
	1	Ph.D		
	1	College is studying curriculum	Computer examiner experience	X ₄
	2	College is studying curriculum is few		
	3	College specializing in computer		
	0	Completed	Subscriber in turn or completed	X ₅
	1	Mutual		

Regression analysis was applied to the data to know the effect of the independent variables on the dependent variable, which is the degree of success in the IC3 exam. The results were as follows:

Table (2): Descriptive Statistics

	Mean	Std. Deviation	N
Score of the examiner	81.75	6.953	65
College specialization	.62	.490	65
Specialization advanced the examiner	.53	.425	
Computer examiner experience	1.83	.945	65
Subscriber in turn or completed	.63	.486	65

Where Table (2) describes the descriptive statistics of the variables that entered in regression analysis, which is the dependent variable (exam score) and independent variables (college specialization, Specialization advanced the examiner, computer examiner experience, Subscriber in turn or completed) and table data indicates that the average of the exam score was (81.75) with a standard deviation of (6.953). The table also shows the averages for the independent variables and their standard deviations.

Table (3): Correlation

Spearman's rho correlation	Score of the examiner	College specialization	Gender	Specialization advanced to the examiner	Computer examiner experience	Subscriber in turn or completed
Score of the examiner	1.000	-.401**	-.186	.193	.443**	-.401**
Sig. (2-tailed)	.	.001	.138	.124	.000	.001
College specialization	-.401**	1.000	.277*	-.075	-.915**	.839**
Sig. (2-tailed)	.001	.	.026	.555	.000	.000
Gender	-.186	.277*	1.000	-.157	-.308*	.375**
Sig. (2-tailed)	.138	.026	.	.210	.013	.002
Specialization advanced to the examiner	.193	-.075	-.157	1.000	-.033	.058
Sig. (2-tailed)	.124	.555	.210	.	.796	.648
Computer examiner experience	.443**	-.915**	-.308*	-.033	1.000	-.809**
Sig. (2-tailed)	.000	.000	.013	.796	.	.000
Subscriber in turn or completed	-.401**	.839**	.375**	.058	-.809**	1.000
Sig. (2-tailed)	.001	.000	.002	.648	.000	.

Table (3) show a matrix of correlations between the independent variables and the dependent variable and the links between the independent variables together. The results of the first table represent the method of the standard multiple regression, where the results showed that the correlation between the adopted variable (exam score) and the independent variable (computer examiner experience) was significant and statistically significant, where the value of (sig = 0.00) was smaller than (0.05) while other variables such as (gender and specialization advanced to the examiner) were statistically significant where the values of (sig) equals (0.138, 0.124), respectively It is greater than (0.05), as the two variables (college specialization, joint in the mother's course) showed an inverse correlation, but it is also statistically significant.

Table (4): Variables Entered

Model	Variables Entered	Variables Removed
1	Subscriber in turn or completed Specialization advanced to the examiner Gender Computer examiner experience College specialization	

a. Passing score. Dependent Variable:

b. All requested variables entered.

Table (5): Variables Entered/removed

Model	Variables Entered	Variables Removed	Method
1	Computer examiner experience		Stepwise (Criteria: Probability-of-F-to-enter \leq .050, Probability-of-F- to-remove \geq .100).

a. Passing score. Dependent Variable:

Table (4) explains to us the independent variables involved in the regression, and since the table pertains to the standard multiple regression method, all variables, whether significant or not significant, were entered in contrast to Table (5) where the table relates to the gradual multiple regression. All the non-significant variables, as well as the variables that are inversely related to the dependent variable, and the independent variable (computer examiner experience) remains the statistically significant variable.

Table (6): ANOVA^a

Model	Sum of squares	df	Mean square	F	Sig.
1 Regression	759.651	5	151.930	3.840	.004 ^b
Residual	2334.411	59	39.566		
Total	3094.062	64			

a. Dependent Variable: Passing score.

b. Predictors: (Constant), College specialization, Computer examiner experience, gender, Specialization advanced to the examiner, Subscriber in turn or completed.

Table (7): ANOVA^a

Model	Sum of squares	df	Mean square	F	Sig.
1 Regression	620.493	1	620.493	15.803	.000 ^b
Residual	2473.569	63	39.263		
Total	3094.062	64			

a. Dependent Variable: Passing score.

b. Predictors: (Constant), Computer examiner experience.

The most important table in the analysis is the tables of (ANOVA), where it shows the results of the analysis of variance of the regression significance test, and we notice through the two tables that the value of the regression is (759.651, 620.493) respectively with a probabilistic value of (0.004, 0.00) which is less than (0.05) and therefore we reject the assumption Zero and we accept the alternative hypothesis, which is that the slope is not morally equal to zero and therefore there is a relationship between the dependent variable and the independent variables, but we do not know specifically which of the five independent variables is the one who added a fundamental explanation for the variance in the variables with the dependent variable so we go to the table (Coefficients) to clarify that matter.

Table (8): Coefficients

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.	correlations		
	B	Std. error	Beta			Zero-order	partial	part
1 (constant)	74.029	6.872		10.773	.000			
College specialization	3.912	4.662	.276	.839	.405	-.413	.109	.095
Gender	.856	1.667	.061	.514	.610	-.057	.067	.058
Specialization advanced to the examine	3.216	1.971	.196	1.632	.108	.147	.208	.185
Computer examiner experience	3.636	2.222	.494	1.636	.107	.448	.208	.185
Subscriber in <u>turn</u> or completed	-3.843	3.167	-.269	-1.214	.230	-.420	-.156	-.137

Passing score. Dependent Variable:

Table (9): Coefficients

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.	correlations		
	B	Std. error	Beta			Zero-order	partial	part
1 (constant)	75.721	1.705		44.410	.000			
Computer examiner experience	3.295	.829	.448	3.975	.000	.448	.448	.448

Passing score. Dependent Variable:

From the observation of the two tables and by comparison between them in the value (sig) where it was significant for the independent variable (computer examiner experience), the regression equation extracted in the second method, the gradual regression is better than the regression equation extracted by the standard multiple regression, and we can write the two equations in the following way:

First: The regression equation extracted by the standard multiple regression
 $y = 74.029 + 3.912x_1 + 0.856x_2 + 3.216x_3 + 3.636x_4 - 3.843x_5$

Second: The regression equation extracted by the progressive multiple regression

$$y = 75.721 + 3.295x_4$$

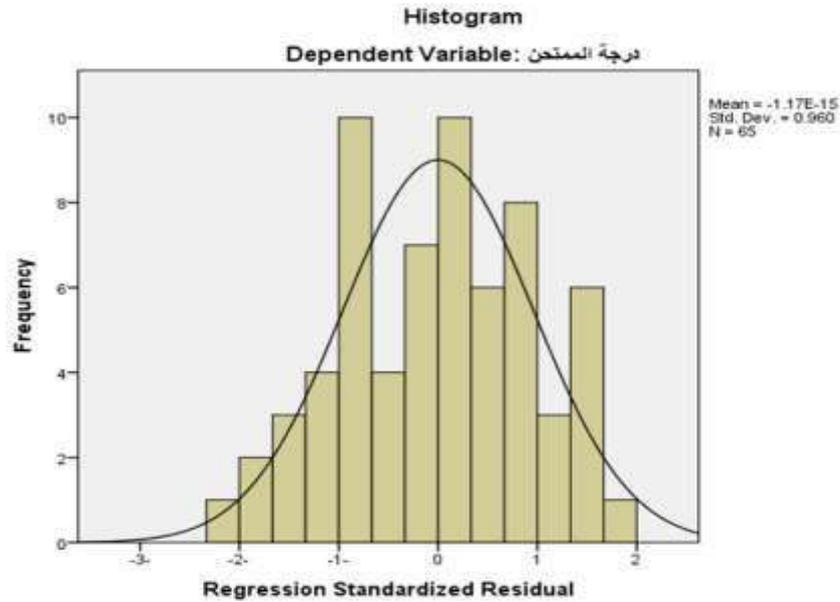


Figure (1): shows that the data follow the normal distribution

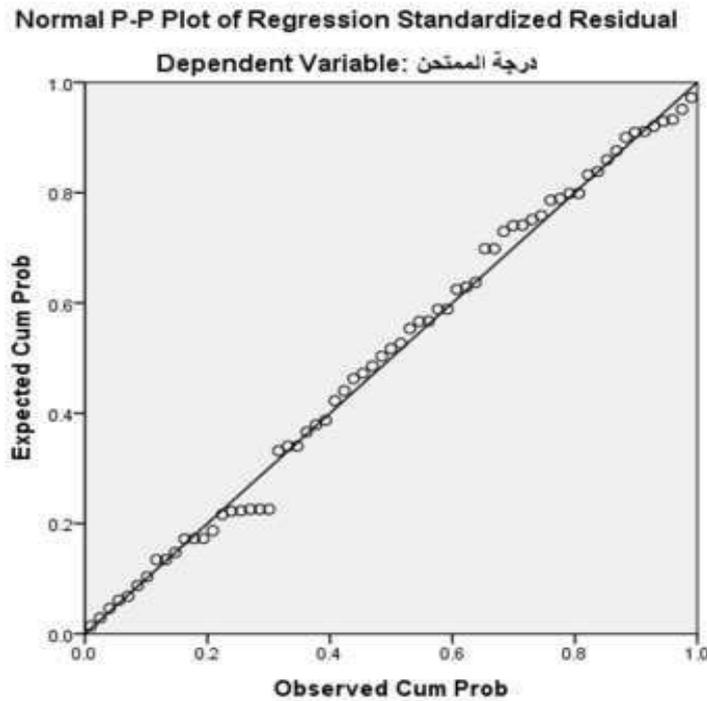


Figure (2): shows that the relationship is linear between the independent and dependent variables

The graphic form tests whether the residues follow the normal distribution or not, and from the figure we find that the points gather around the line and therefore (the residues) are distributed according to the normal distribution.

Table (10): Summary of the results of the two methods for selecting the best model

Var.	MSE	The method used	
X_1, X_2, X_3, X_4, X_5	39.566	Standard Multiple Regression (ENTER)	1
X_4	39.263	Progressive Multiple Regression (Stepwise Regression)	2

From the observation of Table (10) it becomes clear to us that the gradual regression method in finding the regression equation was better than the standard regression method because it entered the variable with a significant statistical significance in the changes that occur for the adopted variable, and from observing the value of (MSE) we find that its value in the regression method Progressive is less than its value in the standard regression method.

7. Conclusions:

The researchers reached, through the results of the statistical analysis, to:

- ❖ The results indicated that there is a statistically significant relationship between the final score of the student who took the Proficiency Exam and his computer experience at a significant level (0.05).
- ❖ The results of the statistical analysis showed that there is no statistically significant relationship between the degree of the examiner and (the college specialization, gender, the student's major, the fourth variable is whether the student is participating in a training course to perform the basic application exam or not) and the results showed that there is no strong moral relationship between them and the result of the student in the exam.

8. Recommendations:

We suggest using multivariate statistical analysis and expanding its field of study due to its efficiency in analyzing and solving the problems of the phenomenon under study. 2-Expanding the study of multivariate statistical analysis methods, including discriminatory analysis.

References:

1. Nasser, Muhammad Bashir, (2002), Multivariate analysis and the investment of the SPSS system in addressing an applied subject in statistical analysis"Master Thesis, Faculty of Administration and Economics-University of Damascus.
2. Namiq, Faisal Najji, Hameed, Majed Rashid, (2017), Using path analysis technology to find out the direct and indirect effects of students failure factors, case study at the Administrative Technical College - Baghdad" Al Kett Magazine for Economic and Administrative Sciences - published by the College of Administration and Economics - Wasit University Issue (26).
3. Al-Jaouni, Farid and Ghanem Adnan, (2007), Multivariate statistical analysis (discriminatory analysis) in the description and distribution of families within the economic structure in society, Damascus University Journal of Economic and Legal Sciences - Volume (23) - No. 2.
4. Abaci , Samet Hasan ,Soner Cankaya, (2015), A comparative study of some Estimation Methods in simple linear Regression Model for Different sample sizes in presence of Outliers" Turkish journal of Agriculture-food science and technology 3(6):380-386
5. Al-Hussein, Yasser Shaaban, (2015), Multivariate statistical analysis and its use to determine the most important factors affecting the effectiveness of banking performance, (field study) -College of Administration and Economics- University of Aleppo.
6. Richard Johnson, Dean and Chernen, (1998), Statistical analysis of multiple variables from an applied point of view, Mars Publishing House, Riyadh, King Saud University.
7. Al-Jubouri, Habib and Ali shalaal, Salah Hamza (2000), A Multivariate Analysis, University of Baghdad Press.
8. Al-Khatib, Bushra Abdul-Bari (2008), The effect of the interaction between independent variables in the multiple linear regression model", MA thesis - College of Administration and Economics-Tikrit University.
9. Templ, Matthias & Kowarik, Alexander and Filzmoser, peter, (2011), Iterative stepwise Regression Imputation Using standard and Robust Method.
10. Assas, Fatten Adnan Mohamed, (2019), Study of variance ratio explained in the progressive multiple regression model in light of different sample sizes" journal of scientific Research in Education, issue 20 (c9), 2019.