DOI: http://dx.doi.org/10.21123/bsj.2019.16.4(Suppl.).1080

Modeling Human Capital Impact on the Development of the Iraqi Oil Industry

Bushra A. Ali [1] *Hazim M. Gorgees* [2] *Raghad I. Kathum* [3]

Received 2/9/2018, Accepted 5/5/2019, Published 18/12/2019

Abstract:

Iraq has the second largest proven oil reserves in the world. According to oil experts, it is expected that the Iraq's reserves to rise to 200+ billion barrels of high-grade crude.

Oil is a strategic commodity for producing and exporting countries in general, and Iraq in particular, as demonstrated by the international experience that oil is an important means to achieve economic growth, an important tool in the overall economic, social and political development. It is also an important source of hard currency for any national economy and a means to connect the local economy and the global economy. In this paper we focus our attention on selecting the best regression model that explain the effect of human capital represented by different levels of human cadres on the development of the Iraqi oil industry by employing the data of a time series consisting of 21years period. Two popular classical procedures were applied to select such model, specifically, all possible regression procedure and the stepwise regression procedures. The statistical program SPSS was employed to perform the required calculations.

Key words: All possible regression, Coefficient of determination, Mallow's C_P, Ordinary least squares, Stepwise regression procedure.

Introduction:

Iraq shall be deemed to be one of the countries that have a promising future in the industry as it is taken in the modern technological methods of production. And use the financial returns in the development of other economic sectors, and diversify sources of income to avoid economic crises, whether in the fall of world oil prices and other crises facing the world economy. Economical scientists agree that many specific factors play an important role and seem to be the basis of the development of any economic sector such as transcription modern technology, the applied strategies in investment and saving as well as the realization of high levels from accumulate material capital. However, other factors may have a significant impact on the regression of different economic sectors. In this paper, we focus our attention on human capital which is represented by different schooling levels of the employees in the specific establishment we believe that it has a considerable effect on the development of any economic sectors. The Iraqi oil industry is considered.

[1] Ministry of Education, Baghdad, Iraq.
[2] University of Baghdad, College of Education for Pure Science, Ibn Al_Haitham - Baghdad, Iraq
[3] Al-Yarmouk University College, Baghdad, Iraq.
*Corresponding author: hazim5656@yahoo.com

It is important of the overall Iraqi economy we attempt to establish the influence of different schooling grades on the rate of annual production of the pure oil by constructing and analyzing a linear regression model to determine the functional relationship among different variables under consideration. However, in many applications of regression analysis, the set of variables to be included in the regression equation is not predetermined and selected. These variables are actually the first step of the analysis. Different methods for selecting the best regression equation are presented in literature (1). Two traditional methods namely, all possible regression equations and stepwise regression method are used.

General Linear Regression Model:,

Let us begin with the general linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ...\beta_p x_{ip} + \varepsilon_i$$

$$= \sum_{j=0}^{p} \beta_j x_{ij} + \varepsilon_i \qquad ,i = 1,2.....n \qquad ....(1)$$

Using the matrix representation, the model can be rewritten as

$y = X\beta + \varepsilon$ …….(2)

Where y is an n×1 vector of response variable, X is an n× (p+1) matrix of explanatory variables, β is a (p+1) ×1 vector of the unknown parameters and ε is an n×1vector of random errors.

We estimate the vector of parameters β by using the popular least squares method, which minimizes the error sum of squares ε'ε. The estimators is given as (2)

$$\hat{\beta} = (X'X)^{-1} X'Y \quad .... \quad (3)$$

The estimator $\hat{\beta}_j$, j=0, 1,…..p is an unbiased estimator of β_j and has variance $\sigma^2 c_{jj}$ where C_{jj} is the j^{th} diagonal element of the matrix $(x'x)^{-1}$ and σ^2 is the unknown population variance .The least square estimators have the smallest variance among all unbiased estimators that are linear in the observations .Therefore, the least square estimators are the best linear unbiased estimators which is denoted as BLUE(3).

Furthermore, the estimator $\hat{\beta}$ is normally distributed with mean β and variance-covariance matrix $\sigma^2(x`x)^{-1}$. The properties of least square estimators and the statistical analysis are based on many assumptions that have to be satisfied, These assumptions may be summarized as follows: (4)

1-The linearity Assumption: This assumptions refers to the linear relationship between the response variable y and the explanatory variables $x_1, x_2 \ldots\ldots\ldots x_p$ in the regression parameters ,that is

$$y = \beta_0 + \beta_1 x_1 + ...\beta_p x_p + \varepsilon$$

This implies that the i^{th} observation can be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + ...\beta_p x_{ip}, \quad i=1,2\ldots n$$

when the linearity assumption dose not satisfied, specific transformation of the data may sometimes lead to linearity.

2- The normality Assumption: This assumption means that the error terms have normal distribution .Suitable graphs of the residuals can be examined to assess the validity of normality assumption.

3-The homogeneity assumption: According to this assumption, the error terms $\varepsilon_1 \ldots\ldots \varepsilon_n$ have the same but unknown variance σ^2 it is also known by constant variance assumption or the homoscedasticity assumption

4- The independent errors assumption: it means that the errors $\varepsilon_1, \varepsilon_2, \ldots\ldots \varepsilon_n$ are independent of each other, when this assumption does not satisfied ,we face the autocorrelation problem.

5-The values $x_{1j}, x_{2j}, \ldots\ldots\ldots, x_{nj}$, j=1,2….p are assumed to be fixed and measured without error because the error in measurement will affect ,the multiple correlation coefficient ,the residual variance and the individual estimates of the regression parameters.

6-The explanatory variables $x_1, x_2 \ldots. x_p$ are assumed to be linearly independent each other .This assumption guarantees the existence of the least squares solution, when this assumption does not hold the problem is referred to as the multicollinearity problem. The problem can be thought of as a situation where two or more explanatory variables in the data set move together. As a consequence it is impossible to decide which of the explanatory variables is producing the observed change in the response variable.

Selecting the best regression equation

Assuming that the full regression model with q explanatory variables $x_1, x_2 \ldots\ldots x_q$ is given as

$$y = \beta_0 + \beta_1 x_{i1} + ...\beta_q x_{iq} + \varepsilon_i \quad \ldots.(4)$$

when q is large, we might delete some variables and construct a model with a subset of variables $x_1, x_2 \ldots\ x_p$ instead of dealing with the full model . Thus, the subset model will be

$$y_i = \beta_0 + \beta_1 x_{i1} + ...\beta_p x_{ip} + \varepsilon_i \quad .... \quad (5)$$

The issue is assessing the effects of including unsuitable variables in an equation when they should be properly left out and the effect of deleting suitable variables when they must be included.

Criteria for Evaluating Models

A criterion is needed in order to judge the sufficiency of various fitted models. Several criteria have been suggested in the literature (5). We describe the two that we believe the most important.

1-Residual mean square (RMS) with a p term model includes a constant term and (p -1) variables, the RMS denoted by S_P^2 is defined as follows

$$S_p^2 = RMS_P = \frac{SSE_P}{n-p} \quad .... \quad (6)$$

Where SSE_p is the residual sum of squares for a p term model. The one with the smallest RMS is usually preferred, particularly, if the purpose is forecasting .The criterion RMS_p is related to the square of multiple correlation coefficient R^2_P as follows (6)

$$R_p^2 = 1-(n-p)\ \frac{RMS_P}{SST} \quad ....(7)$$

Where $SST = \sum (y_i - \bar{y})^2$ is the total sum of squares .The model with greatest R_p^2 is preferable

2- Mallow's C_P

Let us assume that MSE($\hat{y}_i$) to be the mean squared error of the i th predicated value from a p term model and σ^2 to be the variance of the random errors , then the standardized total mean squared error of predication for the observed data is measured by [6]

$$J_p = \frac{1}{\sigma^2} \sum MSE(\hat{y}_i) \quad .... \quad (8)$$

The MSE ($\hat{y}_i$) consists of two components, namely, the variance of prediction appearing from estimation and a bias component arising due to deletion of variables.

In order to estimate J_p , Mallows (1973) employed the statistic

$$C_P = \frac{SSE_P}{S^2} - (n - 2p) \quad .... \quad (9)$$

Where S^2 is an estimate of σ^2 obtained from the linear full model of q variables.

When there is no bias in the fitted model which contains p terms, it can be shown that the expected value of C_P is equal to p .Accordingly, the deviation of C_P from p can used as a measure of bias. Subsets of variables which implies values of C_P that are close to p are preferable subsets.

Variable selection procedures

In this section we deal only with the use of particular statistical procedures for selecting the best regression equation for a particular response y in terms of the q explanatory variables $x_1, x_2 \ldots x_q$. For this situation ,we have to deal with two opposed criteria represented by including as many x's as possible in order to determine reliable fitted values which make the equation useful for predictive purpose. On the other hand we prefer the equation to involve as few x's as possible in order to reduce the costs associated with obtaining information on a large number of x's. The issue is how to compromise between these two extremes .In fact this is called selecting the best regression equation. Two such procedures are reviewed in this article, especially, fitting all possible regressions and the stepwise regression procedure.

Fitting All Possible Regressions (4,7)

Unless the fast computers have become available in general, this procedure cannot applied in practice. First of all, it requires fitting each possible regression equation .With q explanatory variables, the total number of fitted equations is z^q since each variable x_i ,i= 1,2…..q is either be or not be in the equation (two possibilities).The collection of fitted equations includes an equation that contains all explanatory variables and another that contains no variables .The latter is obtained from fitting the model y = β_0 +ε which yield $\hat{y}_i = \bar{y}$. The benefit of this procedure is providing the maximum amount of information available about the kind of relationship between y and the set of explanatory variables .Each regression equation is assessed on the basis of some criterion which are actually related to one another, such as the value of R^2 achieved by using least square fit, the value of RMS and the Mallow's c_p. The procedure is represented by dividing the whole set of explanatory variables into different subsets, the first subset contains only the constant term β_0, the second subset include equations with only one explanatory variable , the third subset include equations with two explanatory variables and so on until the last subset which contain the equation with all explanatory variables. From each subset we pick out an equation with the largest value of R^2, the smallest value of RMS and look for a regression with low c_p value about equal to p .This smaller subset of equations is then carefully analyzed by examining residuals for outliers, autocorrelation or the need for transformations before deciding for the final model.

Stepwise Regression Procedure (8)

According to this procedure, the variables enter the regression equation in turn on the basis of their importance assessed by the value of the partial F (or the value of partial correlation coefficient). First of all, we have to fit a simple regression equation on each explanatory variable, the value of partial F is determined from ANOVA table .The explanatory variable with largest value of partial F is the first variable enters the regression equation .The next step is to obtain a regression equation on each of the remaining explanatory variables with the existence of the variable entered at the first step. The second entering variable is that who satisfy the largest value of partial F calculated from ANOVA table which have to be significant according to specified level of significance .The process is repeated until the best regression equation is achieved. We have to notice that at each step, significance of each explanatory variable already entered must be examined with the presence of explanatory variables recently entered the equation .Accordingly, the decision of inclusion and deletion of variables is made.

Practical study

In this section, the variable selection procedures stated above should be used to analyze the data concerned with the Iraqi oil industry sector. The data were collected from Ministry of oil Training and Development directorate .The best regression equation is to be determined in an attempt to understand the relationship between the rate of annual production of the pure oil as a dependent variable and different schooling grades represented by x_1 stands for PHD holders, x_2 represents MSC, x_3 for Advanced Diploma, x_4 for Bachelor degree, x_5 for Diploma, x_6 for vocational and secondary academic and x_7 for intermediate stage students . In order to apply the all possible regression fitting procedure , the whole set of explanatory variables was divided into eight subsets the statistical program SPSS was employed to fit all of 2^7=128 regression equations .The ordinary least squares method was used to estimate the parameters of the regression equations. Each regression equation is assessed according to the three criteria mentioned earlier, namely, the value of R^2 achieved by ordinary least squares fit ,the value of residual mean squares (RMS) and the value of Mallow's c_p statistic. The results are presented in Table 1.

Table 1.Variables subsets and the values of R^2, S^2 and C_p

Set	Variables	R^2	S^2	C_P
A	No variable	0	482441.4	78.3707
	X1	.615	202275.056	18.46982
	X2	**.697**	**153725.666**	**10.475006**
	X3	.304	353321.548	48.74507
	X4	.581	212752.308	21.79263
B	X5	.544	231446.046	25.37690
	X6	.541	233120.782	25.6980
	X7	.631	187423.432	16.93614
	X_1X_2	.702	159717.840	12.01215
	X_1X_3	.623	202275.056	19.74251
	X_1X_4	.661	181759.406	16.015918
	X_1X_5	.658	183445.132	16.322123
	X_1X_6	.640	192972.648	18.05296
	X_1X_7	.669	177333.626	15.21199
	X_2X_3	.791	111895.937	3.32547
	X_2X_4	.747	135392.591	7.593555
C	X_2X_5	.745	136425.976	7.781276
	$\mathbf{X_2X_6}$	**.807**	**103203.242**	**1.746482**
	X_2X_7	.703	158995.569	11.88095
	X_3X_4	.689	166897.264	`13.31627
	X_3X_5	.685	169077.659	13.712330
	X_3X_6	.653	185901.914	16.768387
	X_3X_7	.694	164082.648	12.805007
	X_4X_5	.596	216709.723	22.364516
	X_4X_6	.626	200474.405	19.415428
	X_4X_7	.632	197502.634	18.8756160
	X_5X_6	.545	243803.539	27.2860107
	X_5X_7	.654	185734.743	16.7380227
	X_6X_7	.661	181972.690	16.054659
	X_1,X_2,X_3	.830	96544.622	1.5627002
	X_1,X_2,X_4	.761	135377.998	8.224749
D	X_1,X_2,X_5	.768	131741.721	7.6009255
	X_1,X_2,X_6	.828	97885.700	1.7927653
	X_1,X_2,X_7	.707	166516.506	13.5667144
	X_1,X_3,X_4	.708	165493.320	13.391174
	X_1,X_3,X_5	.719	159242.671	12.318844
	X_1,X_3,X_6	.668	188384.212	2.1545483
	X_1,X_3,X_7	.701	169434.871	14.067371
	X_1,X_4,X_5	.661	192333.699	17.995775
	X_1,X_4,X_6	.746	144322.467	9.7592164
	X_1,X_4,X_7	.669	187720.347	17.204514
	X_1,X_5,X_6	.735	150237.120	10.773902
	X_1,X_5,X_7	.672	186384.182	16.975105
	X_1,X_6,X_7	.690	176001.875	15.193972
	X_2,X_3,X_4	.793	117697.297	5.1915389
	X_2,X_3,X_5	.792	118097.696	5.2602315
	X_2,X_3,X_6	.813	106082.211	3.1989195
	X_2,X_3,X_7	.791	118443.076	5.3194784
	X_2,X_4,X_5	.750	142003.756	9.3614313
	X_2,X_4,X_6	.833	94518.127	1.2150398
	X_2,X_4,X_7	.747	143315.863	9.5865314
	$\mathbf{X_2,X_5,X_6}$	**.838**	**91714.133**	**0.7339995**
	X_2,X_5,X_7	.750	141647.724	9.3003507
	X_2,X_6,X_7	.820	102067.023	2.5100865
	X_3,X_4,X_5	.697	172223.940	14.545848
D	X_3,X_4,X_6	.691	175357.170	15.083370
	X_3,X_4,X_7	.717	160675.612	12.56467
	X_3,X_5,X_6	.686	178492.801	15.621305
	X_3,X_5,X_7	.703	168326.779	13.877269
	X_3,X_6,X_7	.695	173136.237	14.702356
	X_4,X_5,X_6	.627	211970.681	21.364597
	X_4,X_5,X_7	.687	177723.968	15.489400
	X_4,X_6,X_7	.759	137005.922	8.5040305
	X_5,X_6,X_7	.661	192611.640	18.043457
	X_1,X_2,X_3,X_4	.850	90632.244	1.6337872
	X_1,X_2,X_3,X_5	.837	98584.493	2.9177852
	X_1,X_2,X_3,X_6	.855	87676.070	1.1564722
	X_1,X_2,X_3,X_7	.830	102521.008	3.5533858
	X_1,X_2,X_5,X_6	.839	97014.736	2.6643215
	X_1,X_2,X_5,X_7	.804	118014.043	6.0549478
	X_1,X_2,X_4,X_5	.774	136222.836	8.9949939
	X_1,X_2,X_4,X_6	.838	97543.728	2.7497427
E	X_1,X_2,X_4,X_7	.762	143414.987	10.156268
	X_1,X_2,X_6,X_7	.847	92026.217	1.8588580
	X_1,X_3,X_4,X_5	.720	168697.122	14.238405
	X_1,X_3,X_4,X_6	.746	153335.348	11.758045
	X_1,X_3,X_4,X_7	.721	168165.040	14.152497
	X_1,X_3,X_5,X_6	.752	149602.086	11.155252
	X_1,X_3,X_5,X_7	.720	169031.463	14.292384
	X_1,X_3,X_6,X_7	.701	180024.492	16.0673618
	X_1,X_4,X_5,X_6	.795	123517.619	6.94357158
	X_1,X_4,X_5,X_7	.687	188547.714	17.4435444
	X_1,X_4,X_6,X_7	.798	121958.630	6.69185019
	X_1,X_5,X_6,X_7	.735	159544.534	12.7606022
	X_2,X_3,X_4,X_5	.800	120383.502	6.4375252
	X_2,X_3,X_4,X_6	.835	99448.883	3.05735511
	X_2,X_3,X_4,X_7	.793	125043.238	7.18990412
	X_2,X_3,X_5,X_6	.853	88415.203	1.27581384
	X_2,X_3,X_5,X_7	.793	124693.865	7.13349286
	X_2,X_3,X_6,X_7	.822	107534.176	4.3628319
	X_2,X_4,X_5,X_6	.849	91049.539	1.70116899
E	X_2,X_4,X_5,X_7	.752	149458.122	11.132012
	$\mathbf{X_2,X_4,X_6,X_7}$	**.857**	**86503.046**	**0.9670757**
	X_2,X_5,X_6,X_7	.841	95858.022	2.4775554
	X_3,X_4,X_5,X_6	.702	179756.509	16.0240896
	X_3,X_4,X_5,X_7	.718	170249.577	14.4890675
	X_3,X_4,X_6,X_7	.763	143152.103	10.1138240
	X_3,X_5,X_6,X_7	.704	178361.724	15.7988886
	X_4,X_5,X_6,X_7	.772	137332.952	9.17423844
	X_1,X_2,X_3,X_4,X_5	.851	96137.457	3.5525115
	X_1,X_2,X_3,X_4,X_6	.855	93520.462	3.15637131
	X_1,X_2,X_3,X_4,X_7	.851	95534.981	3.4613150
	X_1,X_2,X_3,X_5,X_6	.862	88895.521	2.4562863
	X_1,X_2,X_3,X_5,X_7	.845	99894.718	4.1212562
	$\mathbf{X_1,X_2,X_3,X_6,X_7}$	**.864**	**87482.896**	**2.2424482**
	X_1,X_2,X_4,X_5,X_6	.849	97014.379	3.6852546
F	X_1,X_2,X_4,X_5,X_7	.804	125881.344	8.0548973
	X_1,X_2,X_4,X_6,X_7	.863	87927.881	2.3098087
	X_1,X_2,X_5,X_6,X_7	.848	98069.648	3.8449923
	X_1,X_3,X_4,X_5,X_6	.795	131742.757	8.9421486
	X_1,X_3,X_4,X_5,X_7	.722	178952.651	16.088406
	X_1,X_3,X_4,X_6,X_7	.806	125041.183	7.9277253
	X_1,X_3,X_5,X_6,X_7	.753	159157.118	13.091919
	X_1,X_4,X_5,X_6,X_7	.804	125873.548	8.0537166
	X_2,X_3,X_4,X_5,X_6	.858	91653.516	2.8737700
	X_2,X_3,X_4,X_5,X_7	.810	122433.780	7.5330377
	X_2,X_3,X_4,X_6,X_7	.857	91894.147	2.9101901
	X_2,X_3,X_5,X_6,X_7	.853	94262.681	3.2687195
F	X_2,X_4,X_5,X_6,X_7	.860	90227.830	2.6579540
	X_3,X_4,X_5,X_6,X_7	.773	145774.297	11.066128

	X_1,X_2,X_3,X_4,X_5,X_6	.862	95205.434	4.450665
	X_1,X_2,X_3,X_4,X_5,X_7	.852	102269.056	5.4486200
G	**X_1,X_2,X_3,X_4,X_6,X_7**	**.866**	**92041.066**	**4.0036036**
	X_1,X_2,X_3,X_5,X_6,X_7	.865	93339.197	4.1870058
	X_1,X_2,X_4,X_5,X_6,X_7	.863	94096.148	4.2939450
	X_1,X_3,X_4,X_5,X_6,X_7	.808	132528.758	9.7237259
	X_2,X_3,X_4,X_5,X_6,X_7	.860	96282.340	4.6028143
H	**$X_1,X_2,X_3,X_4,X_5,X_6,X_7$**	**.866**	**99093.685**	**6.0000009**

The next step is represented by determining the equation that yield the maximum value of R^2 and minimum of RMS and C_p from each set of equations, as shown in Table 2.

Table 2.The selected equation from each subset

Subset	Variables	R^2	S^2	C_P
B	X2	.697	153725.666	10.475006
C	X_2X_6	.807	103203.242	1.746482
D	X_2,X_5,X_6	.838	91714.133	0.7339995
E	X_2,X_4,X_6,X_7	.857	86503.046	0.9670757
F	X_1,X_2,X_3,X_6,X_7	.864	87482.896	2.2424482
G	X_1,X_2,X_3,X_4,X_6,X_7	.866	92041.066	4.0036036
H	$X_1,X_2,X_3,X_4,X_5,X_6,X_7$	.866	99093.685	6.0000009

It is obvious that when the comparison is made between the values of $R^2$for the selected equations from each set we prefer choosing the selected equation from set (c) with explanatory variables(x_2,x_6) since these variables can explain 80.7% of the variation in response variable y. We believe that adding any other remaining variables will not satisfy a significant increment to the value of R^2 and then its inclusion in the equation is not necessary. Moreover, the difference between the RMS for this equation and the RMS for the equation selected from the sets which include extra variables seem to be small .Also this equation has the smallest value of c_p. All these reasons lead us to consider this equation as the best regression equation.

The stepwise regression procedure was also applied as another approach in an attempt to obtain the best regression equations .The first step is represented by evaluating the value of F statistic for regression equation of y on each explanatory variable x_i, i=1,2…7 separately .It was found that the largest value of F statistic is (43.767)obtained from ANOVA table of the regression equation of y in (x_2) (Table 3) which is significant at 5% level of sig .and 1df , therefore (x_2) is the first variable enter the equation .

Table 3. ANOVA table for the regression equations of y on X_2

S.O.V	d.f	SS	MS	F
$R(X_2)$	1	6728039	6728039.491	43.767
$Error(X_2)$	19	2920787	153725.666	
Total	20	9648827		

The next step is evaluating the value of partial F statistic from the ANOVA table when y is regressed on each other of the remaining explanatory variables with existence of (x_2) , the largest value was found to be (24.918) obtained from the ANOVA table of the regression equation of y on (x_6) with existence of (x_2) (Table 4) which is larger than the tabulated F(0.05,1,18) which is equal to (4.41),hence,(x_6)is the second entering variable.

Table 4. ANOVA table for the regression equations of y on X_6 with existence of X_2

S.O.V	d.f	SS	MS	F
$R(X_2X_6)$	2	7791168	3895584.390	
$R(X_6)$	1	5219532		
$R(X_2/X_6)$	1	2571636	2571636	24.91817
$Error(X_2X_6)$	18	1857658	103203.242	
Total	20	9648827		

Now ,before nominating any other explanatory variable to enter the equation, we have to check whether the variable already entered is still significant with the existence of (x_6) , this can be accomplished by evaluating the value of the partial F statistic for the regression equation of y on (x_6) with existence of (x_2) (Table 5).

Table 5. ANOVA table for the regression equations of y on X_2 with existence of X_6

S.O.V	d.f	SS	MS	F
$R(X_2X_6)$	2	7791168	3895584.390	
$R(X_2)$	1	6728039		
$R(X_6/X_2)$	1	1063129	1063129	10.30131
$Error(X_2X_6)$	18	1857658	103203.242	
Total	20	9648827		

It was found to be (10.30131) which is larger than the value of F (0.05, 1, 18), hence, the variable (x_2) will stay in the equation. In order to nominate the third variable to enter the equation ,we have to find the value of partial F statistic from regressing y on each of the remaining explanatory variables (x_1,x_3,x_4,x_5,x_7) with the existence of (x_2) ,(x_6) the results are presented in Tables 6,7,8,9,10.

Table 6. ANOVA table for the regression equations of y on X_1 with existence on X_2 X_6

S.O.V	d.f	SS	MS	F
$R(X_1,X_2,X_6)$	3	7984770	2661590.081	
$R(X_2\ X_6)$	2	7791168		
$R(X_1/\ X_2\ X_6)$	1	193602	193602	1.97783
$Error(X_1,X_2,X_6)$	17	664056	97885.700	
Total	20	9648827		

Table 7. ANOVA table for the regression equations of y on X_3 with existence on X_2 X_6

S.O.V	d.f	SS	MS	F
$R(X_2,X_3,X_6)$	3	7845429	2615143.187	
$R(X_2\ X_6)$	2	7791168		
$R(X_3/\ X_2\ X_6)$	1	54261	54261	0.51149
$Error(X_2,X_3,X_6)$	17	1803397	106082.211	
Total	20	9648827		

Table 8. ANOVA table for the regression equations of y on X_4 with existence on X_2 X_6

S.O.V	d.f	SS	MS	F
$R(X_2,X_4,X_6)$	3	8042018		
$R(X_2\ X_6)$	2	7791168		
$R(X_4/\ X_2\ X_6)$	1	250850	250850	2.6539
$Error(X_2,X_4,X_6)$	17	1606808	94518.127	
Total	20	9648827		

Table 9. ANOVA table for the regression equations of y on X_5 with existence on X_2 X_6

S.O.V	d.f	SS	MS	F
$R(X_2,X_5,X_6)$	3	8089686		
$R(X_2\ X_6)$	2	7791168		
$R(X_5/\ X_2\ X_6)$	1	298518	298518	3.25487
$Error(X_2,X_5,X_6)$	17	1559140	91714.133	
Total	20	9648827		

Table 10. ANOVA table for the regression equations of y on X_7 with existence on X_2 X_6

S.O.V	d.f	SS	MS	F
$R(X_2,X_6,X_7)$	3	7913687		
$R(X_2\ X_6)$	2	7791168		
$R(X_7/\ X_2\ X_6)$	1	122519	122519	1.20037
$Error(X_2,X_6,X_7)$	17	1735139	102067.023	
Total	20	9648827		

We found that all of the remaining variables are not significant at 5% level of significant which means that these variables can be ignored.

Discussion and Conclusions:

From our practical study, a rational conclusions have been achieved which implies that the two explanatory variables x_2,x_6 are more important than other variables since it is most capable to describe the variations in the amount of production (response variable) than the other explanatory variables as it is shown in Table 2. This leads the decision makers to assign a great attention to these categories of human capital. From a statistical point of view, two different approaches were used in an attempt to select the best regression equations in particular, fitting all possible regression approach and stepwise regression approach which comprise between the forward selection procedure and backward elimination procedure. The latter approach is more flexible and more convenient in practical study. However, many other procedures may be used to select the best regression equation such as the ridge regression procedure and the principal components procedure which are more efficient with the collinear data.

References:

1. Wang K, Chen Z. Stepwise Regression and All Possible Subsets Regression in Education. EIJEAS.2016; (l 2), No 60-81 .
2. Ucenna O E , Addeyeye A C, Abiobaragha.S.R .Multiple Regression of Students performance Using forward Selection Procedure ,Backward Elimination and Stepwise procedure. IJIRD. October 2016;(5) .
3. Daoud J. Multicollinearity and Regression Analysis .International coference on mathematical application in engineering .2017 (949).
4. Chaterjee S, Hadi A S. Regression Analysis by Exampke. Fourth Edition, John Wily and Sons, Inc. 2006.
5. Hewig N E. Model Selection and Diagnostics. university of Minnesota Jan-2017 available online
6. Paul A M , Perfoemance of several variable-selection methods applied to real ecological data. Ecology. 2009;(12) 1061-1068
7. Shao J. An Asymptotic Theory for linear Model Selection. Statistic. 1997; 221-264.
8. Speed T P, Yu B. Model selection and Prediction Normal Regression. Ann. Inst. Statist .Math 1993 ;.(45), No 35-54.

نمذجة تاثير رأس المال البشري على تطور صناعة النفط العراقية

بشرى عبد لرسول علي [1] **حازم منصور كوركيس[2]** **رغد ابراهيم كاظم [3]**

[1] وزارة التربية، بغداد، العراق.
[2] جامعة بغداد، كلية التربية للعلوم الصرفة، ابن الهيثم ، بغداد، العراق.
[3] كلية اليرموك الجامعة، بغداد، العراق.

الخلاصة:

يمتلك العراق ثاني أكبر احتياطي نفطي مثبت في العالم. ووفقًا لخبراء النفط ، من المتوقع أن يرتفع احتياطي العراق إلى أكثر من 200 مليار برميل من النفط الخام عالي الجودة.

يعد النفط سلعة استراتجية للدول المنتجة والمصدرة بشكل عام وللعراق بشكل خاص اذ برهنت التجارب الدولية بان النفط وسيلة مهمة لتحقيق النمو الاقتصادي واداة مهمة في التنمية الاقتصادية والاجتماعية والسياسية الشاملة كما انه مصدرا مهما للعملات الصعبة لاي اقتصاد وطني ووسيلة لربط الاقتصاد المحلي بالاقتصاد العالمي .

في هذا البحث سنركز اهتمامنا على اختيار أفضل نموذج انحدار من خلال توظيف قياس أثر راس المال البشري ممثلا بمستويات مختلفة من الكوادر البشرية على انتاجية العمل في تطوير صناعة النفط العراقية من خلال توظيف بيانات سلسلة زمنية مؤلفة من 21 سنة .
سنطبق طريقتين كلاسكيتين شائعتيين لاختيار النموذج خصوصا أسلوب كل الانحدارات الممكنة وكذلك أسلوب الانحدار المتدرج وقد استخدم البرنامج الاحصائي SPSS لاجراء الحسابات المطلوبة .

الكلمات المفتاحية : كل الانحدارات الممكنة، معامل التحديد، $Mallow's\ C_p$ ، طريقة المربعات الصغرى، أسلوب الانحدار المتدرج.