A Study on Prediction of Output in Oilfield Using Multiple Linear Regression

by

Izni binti Mustafar

Dissertation submitted in partial fulfilment of

the requirements for the

Bachelor of Engineering (Hons)

(Electrical and Electronic Engineering)

MAY 2011

Universiti Teknologi PETRONAS

Bandar Seri Iskandar

31750 Tronoh

Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

A Study on Prediction of Output in Oilfield Using Multiple Linear Regression

by

Izni binti Mustafar

A project dissertation submitted to the
Electrical and Electronic Engineering Programme
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
BACHELOR OF ENGINEERING (Hons)
(ELECTRICAL AND ELECTRONIC ENGINEERING ENGINEERING)

Approved by,

(AP DR. RADZUAN RAZALI)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
November 2010

ABSTRACT

This report basically discusses about method that has been done to validate this study, which is A Study on Prediction of Output Oilfield Using Multiple Linear Regression. The objective of the project is to come out with a solution on how to predict the output of oilfield using multiple linear regressions. The basic method is to establish the linear relationship between oil output in oilfield and the influencing factors. An oilfield is an area with reserves of recoverable petroleum, especially one with several oil-producing wells [1]. The challenge in this project is to find the variable for the output of oilfield because here are numerous factors affecting output in an oilfield. The relationship between field output and one of the affecting factors is unscientific and are not precise, which it is needed to come out with a simple and more accurate. After a several studies have been made, there are some factors affecting the output that were identified as variables for the output in an oilfield. There are 8 parameters that have been identified to predict the oilfield of output. Then the author constructs full calculation in order to find the data and value for regression coefficient, β to predict the output oilfield. The author also used the ordinary least squares method to minimize the sum of variables by eliminating the least important variables. Last but not least, the author uses different set of data with the same parameters to do some comparison and verified the validity of this method.

ACKNOWLEDGEMENT

First and foremost, I would like to express my utmost gratitude to my Final Year Project (FYP) supervisor, Associate Professor Dr. Radzuan bin Razali for helping and assisting me throughout these two semesters by giving opinions, suggestions and advices as well as his continuous encouragement to me. I do believe without his supervising, this study would not have been successful. He was greatly inspired me to work in this project, and as much as I appreciate things, his determination in motivating me contributed tremendously in this project.

Besides, my coursemates have been a very good company throughout my journey doing this project. They have been helping me a lot especially the Petroleum Students for helpin to gain a better understanding about the Oilfield background

Other than that, I would love to thank to all parties who contributed directly or indirectly for making this project a success.. Last but not least, I would like to thank to the course coordinators for giving me such opportunity to explore creativity and innovativeness through this course in UTP.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Background Of Study

The production of oil is very significance as a world energy source. Every year, the increasing of oil production has been by far as the major contribution to the growth in energy production. The oil production is generally from an oilfield. An oilfield is an area of sedimentary rocks under the ground or called as crude oil. Oil is created in a source rock along with water and gas.

The oilfields typically extend over a large area, possibly several hundred kilometres across. Therefore, full exploitation entails multiple wells scattered across the area. In addition, there may be exploratory wells probing the edges, pipelines to transport the oil elsewhere and support facilities.

The term oilfield is also used as shorthand to refer to the entire petroleum industry. However, it is more accurate to divide the oil industry into three sectors which are upstream, midstream and downstream. Upstream is a crude production from wells and separation of water from oil meanwhile midstream is a pipeline and tanker transport of crude and downstream is a refining and marketing of refined products[2].

For a major reason, it is crucial to predict the oilfield output for oil production. Thus, studies have been making to predict the output using multiple linear regression method.

1.2 Problem Statement

In a business world, the production planning management is extremely important. It purposes is to efficiently organize the use of resources and maximize efficiency in workplace. Therefore, it is crucial to predict the oilfield output accurately in the oil production process to easily plan the oilfield production planning management system.

At present, the main methods of oilfield output prediction include injection production relation model, production decline model and logistic model. However, the input variables are fixed in the above methods and significant factors which influence the oilfield system dynamic is not considered [3]. Thus the multiple linear regression method is used to predict the output of oilfield because this method is more simple and accurate.

1.3 Objective

The main objectives of this research are:

1. To analyse important parameters or the influencing factors in oil production
2. To predict the output of oilfield production using Multiple Linear Regression method.
3. To use the Multiple Linear Regression method using MATLAB programme.
4. To identify the most significant influencing factors in oil product.
5. To validate the Multiple Linear Regression forecasting model.

1.4 Scope Of Study

The scope of study will evolve around the programming on MATLAB to calculate the output of oilfield using multiple linear regression method. Learning on the method is also needed as the calculation process is required to be implemented on the coding. The author also utilised the Microsoft Office Excel to construct a data and at once to construct the Multiple Linear Regression equation.

The author also conducted a study on topic cover from Oil Field, Well Production, Reservoir and Oil Recovery to have a better understanding on their behaviour. It is crucial in determining the important factors of influencing oilfield output which have been identified as the variables to predict the output of oilfield.

Overall, the project scope has been divided into two stages whereby the first stage is the study of the theories behind the oilfield and well production as well as the method used in estimating the output of oilfield. Meanwhile, the second stage is to simulate the calculation of Multiple Linear Regression to predict the output of oilfield using MATLAB. The simulation used to calculate the output of oilfield that will be going used in the downstream process.

CHAPTER 2
LITERATURE REVIEW

2.1 Oilfield

An oilfield is an area with reserves of recoverable petroleum, especially one with several oil-producing wells [1]. It is full of oil or gas or both. Mostly oilfield has sizes ranging from 10-20 km wide. An oilfield needs a special combination of geological features. Oil is created in a source rock, along with water and gas. Over millions of years, the oil and gas float upwards above the water along a migration path. Oil and gas often rises all the way to the surface of the earth. In other cases, it collects in a reservoir which is a rock that has spaces where oil can collect [3].

Some oil is in structural traps for example where rocks have been folded into a dome shape, in which oil will collect at the top of the dome. Other structural traps are sealed by faults, where tilted blocks of rock have slid up or down and become sealed by layers of clay. Some oil is in stratigraphical traps, where areas of sand have been laid down within areas of clay for example in ancient river channels [3].

2.2 Factors Affecting Oilfield Production

In oil production, there are two major factors affecting oilfield production which are geological factors and human factors. Therefore, these two factors are being considered to predict the output of oilfield. Considering the geological factors, the oil wells are the utmost important element in predicting oilfield's output directly determines the yield of oilfield [4].

Next, the water content of oil also is considered as major factor that affect the oilfield production. These due to some of oil well in our country are non self spraying. Thus, the respective oil wells need steam or injecting water to drive oil. It also can be used to increase pressure and thereby it will stimulate production of oilfield. The available oil reserves are also a factor because underground reserves of oil are basically unchanged [4].

The basic method is to establish the linear relationship between oil output and the influencing factors such as moisture content. Then the linear system is established according to the experience. To predict future output of an oilfield, the influencing factors combined with actual production are selected and analysed deeply.

2.3 Water Injection Process

Water injection is one in an oil recovery process where water is injected back into the reservoir. Mainly the water is injected into the reservoir is to increase pressure, thus it will exhilarate the production. The purpose of this process is to support pressure of the reservoir and to displace oil from reservoir and push it towards a well [5].

2.4 Oil Well Drilling Process

Oil well is the main factors affecting the output of oilfield because it determines the yield of oilfield. An oil well is a well that supplies either naturally or by means of a pumping system. A hole drilled or dug in the earth surface that is designed to find and acquire petroleum oil hydrocarbons [6]. Usually some natural gas is produced along with the oil. The well is created by drilling a hole into earth with an oil rig turning as drill bit.

After the hole is drilled, a metal pipe called 'casing' is cemented into the hole. In order to get access to the hydrocarbon producing interval, the 'casing' and cement are either perforated or additional section of earth is drilled below the 'casing'. In most cases several casings are set in the well, starting with large shallow 'casing' and the deeper 'casings' are set in smaller holes drilled through the upper 'casings' [6].

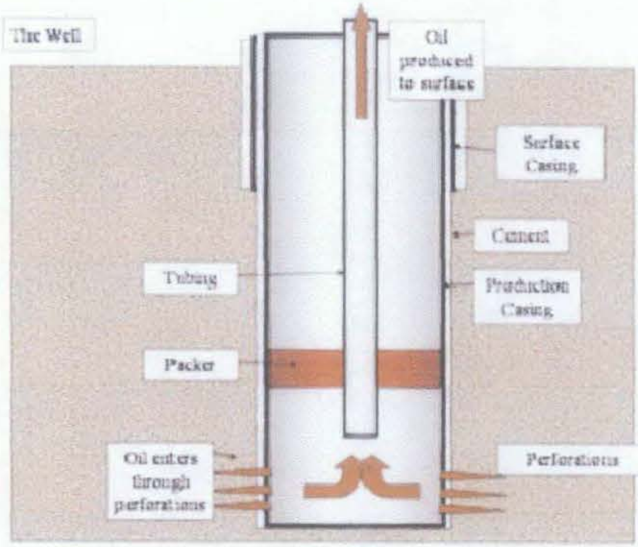


***Figure 1** : Drilling Process*

2.5 Enhanced Oil Recovery

Enhanced Oil Recovery (EOR) is a method to increase the amount of crude oil that can be extracted from an oilfield. This can be related as a factors affecting oilfield output because it determine the value of new oil every year in oil well. There are three stages of oil field development which are primary, secondary and tertiary. In primary recovery, oil is forced out by pressure generated from gas present in the oil meanwhile in secondary the reservoir is subjected to water flooding or gas injection to maintain a pressure that continues to move oil to the surface [7].

While tertiary recovery also known as EOR, introduces fluids that reduce viscosity and improve flow. These fluids could consist of gases that are miscible with oil, steam, air or oxygen, polymer solutions, gels, surfactant-polymer formulations, alkaline-surfactant-polymer formulations, or microorganism formulations [7].

The primary recovery typically provides access to only a small fraction of a reservoir's total oil capacity. Secondary techniques can increase productivity to a third or more. Tertiary recovery enables producers to extract up to over half of a reservoir's original oil content, depending on the reservoir and the EOR process applied [7].

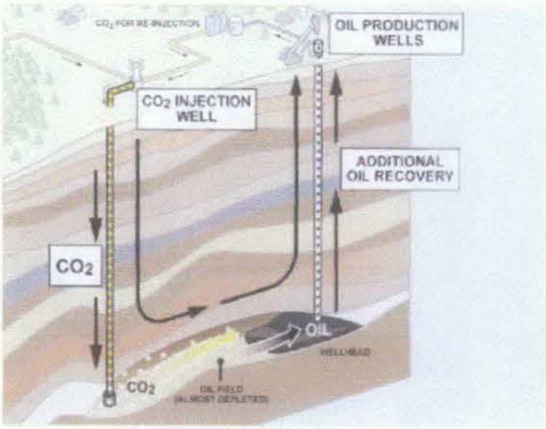


Figure 2 *: Oil Recovery*

2.6 Model Used To Predict the Output of Oilfield

The oilfield development of predicting the output of an oilfield is the basis of the optimal decision making of oilfield manager [8]. By far, there are many methods to predict the output of oilfield such as Multiple Linear Regression, Artificial Neural Network, Grey Prediction method, and Logistic Curve Method which have different applicable environments and limits [8].

At present time, there are several major models are being used to predict the oilfield output such as logistic model [9], production decline model [10] and logistic model [11]. But the problem is, there are several input variables in the above models and significant factors influencing dynamic system is not considered. Thus, the prediction result was affected and is not accurate. Meanwhile, the Multiple Linear Regression model is more simple and accurate.

In the process of predicting the oilfield output using Multiple Linear Regression model, several model factors related to oilfield output are often identified as the model variables.

By using this model, the Multiple Linear Regression equation is constructed. Therefore, the most significant factors that influence the oilfield output are determined by using the Multiple Linear Regression model. The model is applied to the actual production and the satisfying predictions are obtained.

Some model factors related to oilfield output are often taken as the model variables in the process of prediction oilfield output with the multiple linear regression method. The key factors which influence oilfield output is determined by comprehensive multiple regression analysis and then the multiple linear regressions are built. The model is applied to actual production and the satisfying predictions are obtained [3].

2.7 Multiple Linear Regression

2.7.1 Linear Model

Multiple linear regressions are one of the most widely used of all statistical methods. Multiple regression analysis is also highly useful in experimental situation where the experimenter can control the predictor variables. A single predictor variable in the model would have provided an inadequate description since a number of key variables affect the response variable in important and distinctive ways [13].

It attempts to model the relationship between two or more variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable x is associated with a value of the dependent variable y. The population regression line for p explanatory variables $x_1, x_2, \ldots, x_p$ is defined to be :

$$\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

This line describes how the mean response μ_y changes with the explanatory variables. The observed values for y vary about their means μ_y and are assumed to have the same standard deviation σ. The fitted values $b_0, \ldots, b_p$ estimate the parameters $\beta_0, \beta_1, \ldots, \beta_p$ of the population regression line [13].

β_0 is the mean of y when all x's are 0. Meanwhile, β_j is the change in the mean of Y associated with a unit increase in x_j, holding the values of all the other x's fixed. Coefficient estimated via least squares.

2.7.2 Confidence and Prediction Intervals

Variance of mean response at x_o :

$Var\,(\hat{y}_0) = \; Var\, x'_0\hat{\beta} = \sigma^2\, x'_0\,(X'X)^{-1}\, x_0 = \sigma^2 v_0$ [14]

Variance of new observation at x_0 , $y_0 = \hat{y}_0 + \varepsilon_0$ [14];

$Var\,(\,\hat{y}_0 + \varepsilon_0\,) \; = \; Var\,(\hat{y}_o) + \, Var\,(\varepsilon_0) = \sigma^2\, x'_0\,(X'X)^{-1}\, x_0 + \sigma^2 \; =$
$\sigma^2\,(x'_0\,(X'X)^{-1}\, x_0 + \;1) = \; \sigma^2\,(v_0 + \;1)$

An estimate of σ^2 is s^2 = MSE = $\frac{y'(1-H)y}{(n-k-1)}$

The $(1-a)$ Confidence Interval on Mean Response at x_0 is defined as below [14]:

$$\hat{y}_0 \pm cd$$

Where;

$c = \; t_{n-(k+1),a/2}$ and $d = \sqrt[s]{v_0}$

Meanwhile, the $(1-a)$ Confidence Interval on New Observation at x_0 is defined as below [14] :

$$\hat{y}_0 \pm cd$$

Where;

$c = \; t_{n-(k+1),a/2}$ and $d = \sqrt[s]{v_0 + \;1}$

2.7.3 Sum of Squares

Sum of squares is a concept that permeates much of inferential statistics and descriptive statistics. More properly, it is the sum of squared deviations. Mathematically it is an unscaled , or unadjusted measure of dispersion. When scaled for number of degrees of freedom, it estimates the variance, or spread of the observations about their mean value [14].

Based on sample $i = 1, 2, \ldots, n$ containing n observations;

Sum of Squares Total (SST): $\sum_{i=1}^{n}(y_i - \bar{y})^2$

Sum of Squares for Error (SSE): $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2$

Sum of Squares for Regression (SSR): $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$

SSR = SST – SSE

2.7.4 Overall Significance Test

To see if there is any linear relationship we test [14]:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$
$$H_1 : \beta_j \neq 0 \text{ for some } j$$

Compute the equation as below:

SSE = $\sum(y_i - \hat{y}_i)^2$

SST = $\sum(y_i - \hat{y}_i)^2$

SSR = SST – SSE

The F statistic is :

$$\frac{\frac{SSR}{k}}{\frac{SSE}{(n-k-1)}} = \frac{MSR}{MSE}$$

With F based on k and $(n - k - 1)$ degrees of freedom.

Reject H_0 when F exceeds$F_{k,n-k-1(a)}$.

2.7.5 ANOVA Table

The ANOVA Table for regression:

Source	SS (Sum of Squares, the numerator of the variance)	DF (the denominator)	MS (Mean Square, the variance)	F
Regression	$SSR = \sum_{i=1}^{n}((\hat{\beta}_0 + \hat{\beta}_1 x_i) - \bar{y})^2$	$k-1$	$MSR = \frac{SSR}{k-1}$	$F = \frac{MSR}{MSE}$
Error	$SSE = \sum_{i=1}^{n}(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$	$n-k$	$MSE = \frac{SSE}{n-k}$	
Total	$TTS = \sum_{i=1}^{n}(y_i - \bar{y})^2$	$n-1$		

***Table 1**: ANOVA Table*

2.8 Matlab

MATLAB is a program for computation and visualization. It is widely used and is available on all kinds of computers, ranging from personal computers to supercomputers. MATLAB is controlled by commands and it is programmable. There are hundreds of predefined commands and functions and these functions can be further enlarged by user-defined functions [15].

Furthermore, MATLAB has a really powerful command. For instance, it can solve linear systems with one single command, and perform a lot of advanced matrix manipulations. It also has powerful tools for graphics in two and three dimensions. The MATLAB can be used together with other programs. The graphic capabilities of MATLAB can, for instance, be used to visualize computations performed in a FORTRAN program [15].

MATLAB also can be used to calculate the curve fitting and interpolation. It provides interpolation functions for both two and three dimensions. MATLAB can return the values of a set of points to its intermediate points by interpolating the data. This can be done in a different ways. Thus, MATLAB is the suitable software to be used in this project.

CHAPTER 3
METHODOLOGY

3.1 Project Work Flow

The project activities flow is shown in Figure 3.

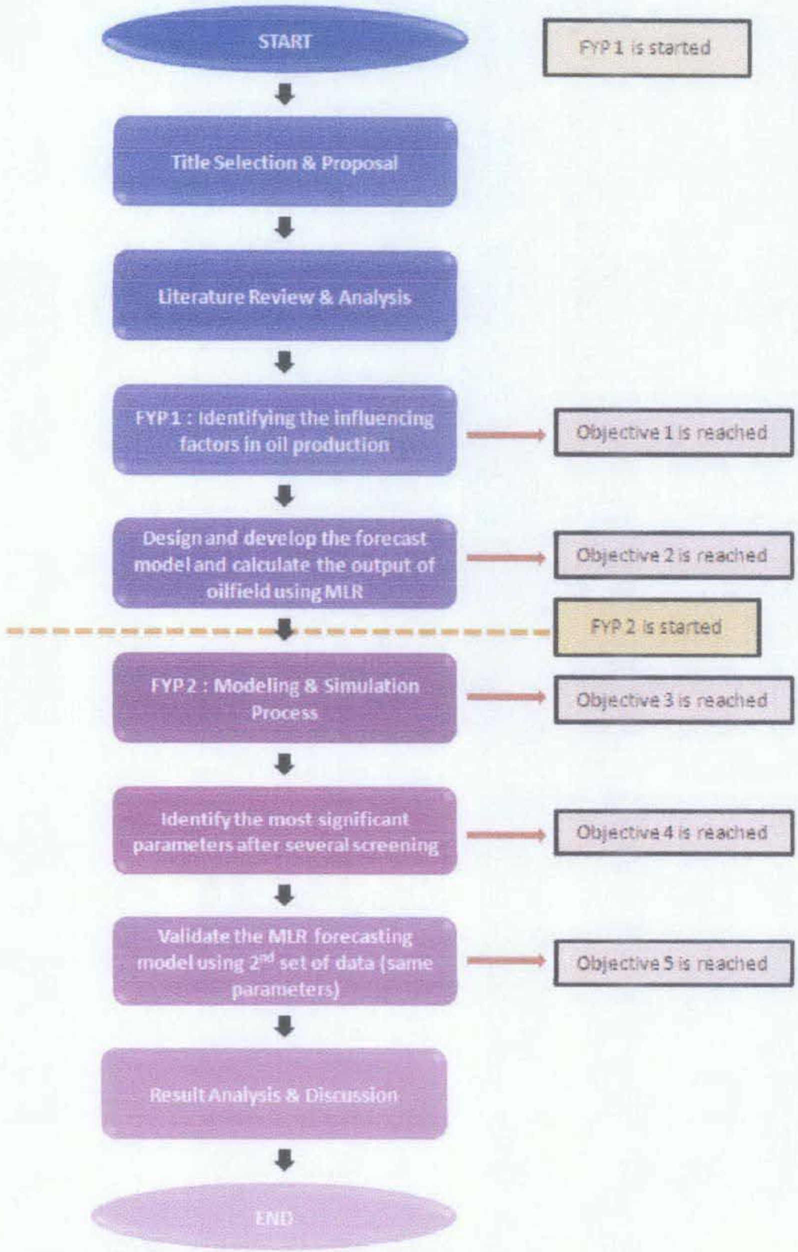


***Figure 3:** Project Activities Flow Chart*

3.2 Research Methodology

In order to achieve the aim of the project, some research has been done on several resources from books, technical papers and internet. For the first step, the gathering information needs to be done on the Oilfield, Well Production, Reservoir Behaviour and Multiple Linear Regression method. After all the studies have been done and the parameters have been identified, the author started constructs a Multiple Linear Regression calculation using Microsoft Excel and obtain the oilfield output.

The next stage is the simulation stage whereby the calculation will be simulated in order to make it easier to achieve the oilfield output. During this stage, knowledge of MATLAB software is a requirement. Apart from that, the author identifies the most significant parameters after several screening and validates this model using 2^{nd} set of data with the most significant parameters only.

3.3 Tools Required

For the accomplishment of the project, there are needs for a certain software application especially for Modelling and Simulation process for our design. For this project, the author use Microsoft Excel to construct the calculation and do modelling and simulation using MATLAB software.

3.4 Determining Factors Affecting Oilfield Production

In oil production, there are two major factors affecting oilfield production which are geological factors and human factors. Therefore, these two factors are being considered to predict the output of oilfield. Considering the geological factors, the oil wells are the utmost important element in predicting oilfield's output directly determines the yield of oilfield [3].

Next, the water content of oil also is considered as major factor that affect the oilfield production. These due to some of oil well in our country are non self spraying. Thus, the respective oil wells need steam or injecting water to drive oil. It also can be used to increase pressure and thereby it will stimulate production of oilfield. The available oil reserve is also a factor because an underground reserve of oil is basically unchanged [3].

The basic method is to establish the linear relationship between oil output and the influencing factors such as moisture content. Then the linear system is established according to the experience. To predict future output of an oilfield, the influencing factors combined with actual production are selected and analysed deeply. Eight factors are selected as follows [3]:

1. The total numbers of wells
2. The start up number of wells
3. The number of new adding wells
4. The injected water volume last year
5. The oil moisture content of previous year
6. The oil production rate of previous year
7. The recovery percent of previous year
8. The oil output of previous year

3.5 Construct Calculation Using Multiple Linear Regression Model

In this project, the basic method is to establish the linear relationship between oil output and the influencing factors such as moisture content. To predict future output of an oilfield, the influencing factors combined with actual production are selected and analysed. Eight factors affecting oilfield outputs are:

Parameters	X variables
Total numbers of wells	x_1
Start up well number	x_2
Number of new adding wells	x_3

The latest injected water volume	x_4
Moisture content of previous year	x_5
Oil production rate of previous year	x_6
Recovery percent of previous year	x_7
Oil output of previous year	x_8

***Table 2** : Factors affecting oilfield outputs*

The multiple linear regression formula can be expressed as follows:

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_3 x_3$$

y = annual oilfield output

The matrix form is written by $y = x\beta$

β is the unknown parameter of the multiple linear regression equation that is the regression parameter. Based on the 8 identified variables, the multiple linear regression can be expressed as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

$$+ \beta_7 x_7 + \beta_8 x_8$$

3.6 Applying Calculation Using Excel

To do regression in Microsoft Excel, the author used data analysis functions which include multiple regressions.

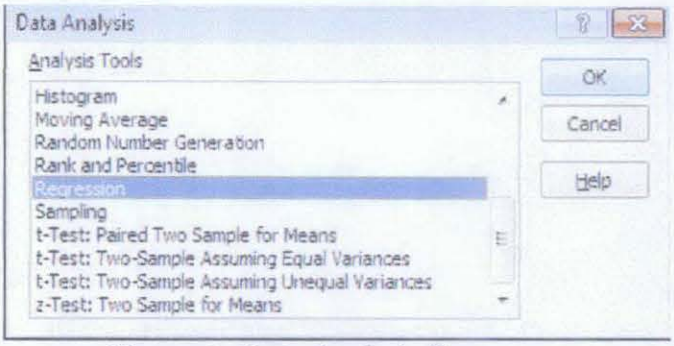


***Figure 4**: Data Analysis Box*

By select Regression, the following dialog appears:

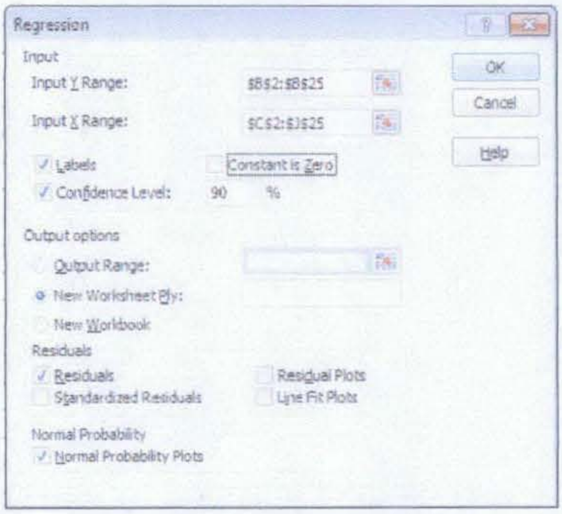


***Figure 5** : Regression Application Table*

In here, the all data will be place in respective column to be analyzed.

3.7 Simulation Using Matlab

MATLAB imports date fields from Excel files in the format in which they were stored in the Excel file. If stored in string or date format, `xlsread` returns the date as a string. If stored in a numeric format, `xlsread` returns a numeric date[15].

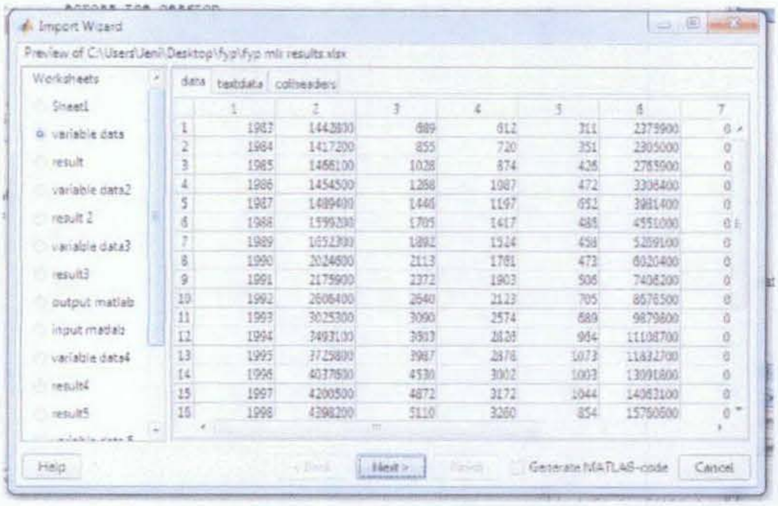


***Figure 6** : Import Wizard in Matlab*

To do regression in Matlab, the author used regstats functions which include multiple regressions. Regstats (y,X,*model*) performs a multi-linear regression of the responses in y on the predictors in X. X is an *n*-by-*p* matrix of *p* predictors at each of *n* observations. y is an *n*-by-1 vector of observed responses[16]. By default, regstats automatically adds a first column of 1s to the design matrix to compute the *F* statistic and its *p*-value so a constant term should not be included explicitly as for regress. For example[16]:

```
X2 = moore(:,1:5);
stats = regstats(y,X2);
```

These functions will create structure stats containing regression statistics. To identify the computed statistics, call 'regstats' without an output argument. An optional input argument allows you to specify which statistics. For example [16]:

```
regstats(y,X2)
```

When the 'regstats' function was called, it will open the following interface.

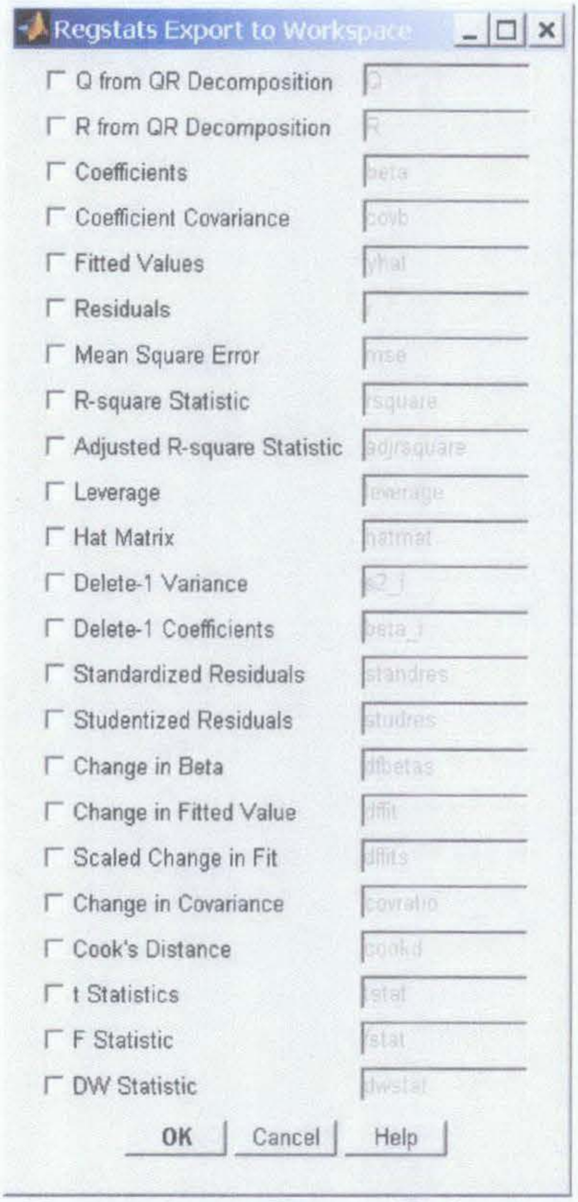


***Figure** 7 : Regstat Export to Workspace*

3.7.1 Tabulating Diagnostic Statistic

The 'regstats' function computes statistics are typically used in regression diagnostics. Statistics can be formatted into standard tabular displays in a variety of ways. For example, the 'tstat' field of the stats output structure of 'regstats' is itself a structure containing statistics related to the estimated coefficients of the regression. Dataset arrays provide a natural tabular format for the information [16]:

```
t = stats.tstat;
CoeffTable = dataset({t.beta,'Coef'},{t.se,'StdErr'}, ...
                     {t.t,'tStat'},{t.pval,'pVal'})

CoeffTable =

    Coef              StdErr            tStat        pVal
     2.0197e+006      1.3887e+006        1.4544        0.16645
          177.71           135.64        1.3102        0.20984
          218.25           85.414        2.5552       0.021969
           193.7           216.54       0.89454        0.38516
        0.076829         0.041965        1.8308       0.087072
    -5.4502e+006      2.1771e+006       -2.5035       0.024339
    -9.8346e+007       5.786e+007       -1.6997        0.10982
     2.7193e+007      6.7288e+006        4.0412      0.0010661
        0.025911          0.20441       0.12676        0.90081
```

The MATLAB 'fprintf' command gives you control over tabular formatting. For example, the 'fstat' field of 'thestats' output structure of 'regstats' is a structure with statistics related to the analysis of variance ANOVA of the regression. The following commands produce a standard regression ANOVA table [16]:

```
f = stats.fstat;

fprintf('\n')
fprintf('Regression ANOVA');
fprintf('\n\n')

fprintf('%6s','Source');
fprintf('%10s','df','SS','MS','F','P');
fprintf('\n')

fprintf('%6s','Regr');
fprintf('%10.4f',f.dfr,f.ssr,f.ssr/f.dfr,f.f,f.pval);
fprintf('\n')

fprintf('%6s','Resid');
fprintf('%10.4f',f.dfe,f.sse,f.sse/f.dfe);
fprintf('\n')

fprintf('%6s','Total');
fprintf('%10.4f',f.dfe+f.dfr,f.sse+f.ssr);
fprintf('\n')
```

After tabulating the data the result looks like this [16]:

```
Source          df          SS          MS          F          P

  Regr    8.00001451424645618 53.06001814280807023 1.6330
1456.6428    0.0000

 Resid   15.0000186828316480.326012455221098.6884

 Total   23.000014532929287 8333.3700
```

CHAPTER 4
RESULT AND DISCUSSION

The author obtains a list of data parameters from China's Oilfield. Please refer to ***Appendix B***. Based from data in ***Appendix B***, the calculation was constructed using Microsoft Excel and Matlab based on Multiple Linear Regression(MLR) model where :

$x1$ = The total numbers of wells

$x2$ = The start up number of wells

$x3$ = The number of new adding wells

$x4$ = The injected water volume last year

$x5$ = The oil moisture content of previous year

$x6$ = The oil production rate of previous year

$x7$ = The recovery percent of previous year

$x8$ = The oil output of previous year

y = The oil output

From basic MLR equation $y = x\beta$, the basic form MLR can be expressed as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

$$+ \beta_7 x_7 + \beta_8 x_8$$

Linear regression method was used to calculate the regression coefficients with 8 independent variables. The regression coefficients from β_0 to β_8 are respectively given as follows:

$\beta_0 = 2019687.48$

$\beta_1 = 177.71$

$$\beta_2 = 218.255$$

$$\beta_3 = 193.70$$

$$\beta_4 = 0.077$$

$$\beta_5 = -5450242.21$$

$$\beta_6 = -98346111.91$$

$$\beta_7 = 27192743.26$$

$$\beta_8 = 0.026$$

The mathematical regression model is obtained as:

$$y = 2019687.48 + 177.71x\_1 + 218.255x\_2 + 193.70x\_3 +$$

$$0.077x\_4 - 5450242.21x\_5 - 98346111.91x\_6 +$$

$$27192743.26x\_7 + 0.026x\_8$$

Based on ***Appendix D***, the less significant variables are rejected one by one based on P-value. The significant indicator $a = 0.1$ is considered as the screening index. When P – value > 0.1, the item is the less significant items and should be removed. Otherwise, the result is opposite. For instance, the P – value for $x8$ which is oil output last years is 0.9008 in the first round screening and P – value > 0.1, therefore, this item should be rejected. After 5 rounds screening, the variables rejected are as follows:

$x1$ = The total numbers of wells
$x3$ = The number of new adding wells
$x6$ = The oil production rate of previous year
$x8$ = The oil output of previous year

From the calculation, the P – value of all variables left satisfy the significance requirements of $a = 0.1$. After the screening of P – value, the four most important factors which affect the oilfield output are determined. They are (in most significant order) :

$x2$ = The start up number of wells

$x7$ = The recovery percent of previous year

$x4$ = The injected water volume last year

$x5$ = The oil moisture content of previous year

Therefore, the new mathematical model is obtained with the screening of P – value. The model may written as

$$y = -259910 + 352.2079x_2 + 0.123018x_4 -$$

$$3660564x_5 + 27702445x_7$$

After the result obtain, the author compared two kinds of model (four parameters model and eight parameters) that were used to predict oilfield production from 2000 to 2006.

Year	Actual Output	Four Parameters Output	Error (%)	Eight Parameters Output	Error (%)
2000	4712500	4755212	0.413101	4819842	1.191005
2001	5205000	5197864	0.069019	5180056	0.276769
2002	6115500	6155542	0.387279	6169154	0.595312
2003	7158700	7146255	0.120367	7195551	0.408875
2004	8109500	8030987	0.759365	7966909	1.582103
2005	9051000	8856198	1.884093	8956653	1.046812
2006	9623000	9822647	1.93096	9752501	1.436868
Average Total			5.564183		6.537743

***Table 3**: The Comparison Of Prediction Results Of Two Models*

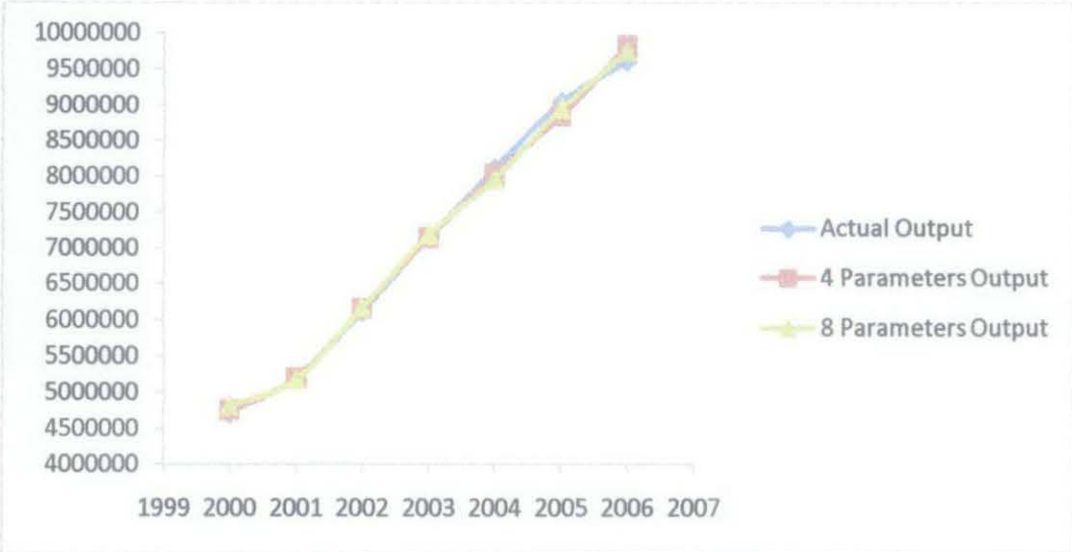


Figure 8: *The Relationship Between Actual Output And 8&4 Parameters Output*

Thus, to verify this method the author obtain new list data parameters obtain from another China's oilfield (Please refer to ***Appendix** C*) but by only using 4 parameters that have the most significant value in calculation that were made using the first data which are:

$x1$ = The start up number of wells

$x2$ = The recovery percent of previous year

$x3$ = The injected water volume last year

$x4$ = The oil moisture content of previous year

After the screening of P – value in the new set of data, all the P- value still satisfy the significance requirement $a = 0.1$ which is P – value > 0.1. (Please refer to ***Appendix** T*)

Thus, from the result obtain, the author calculate the percentage error from the latest model. We can see that from Table 4 that the total percentage error is less than 4.57%. This validate that the MLR method can be use to forecast oilfield data.

Year	Actual Output	Latest Model	Error (%)
2000	4712500	4660515.787	0.408306
2001	5205000	5108470.348	0.06392
2002	6115500	6067627.998	0.4519
2003	7158700	7058753.792	0.100139
2004	8109500	7968840.953	0.531737
2005	9051000	8815560.913	1.536501
2006	9623000	9672381.567	1.482886
Average Total			4.57539

***Table 4**: The error calculated from the latest model*

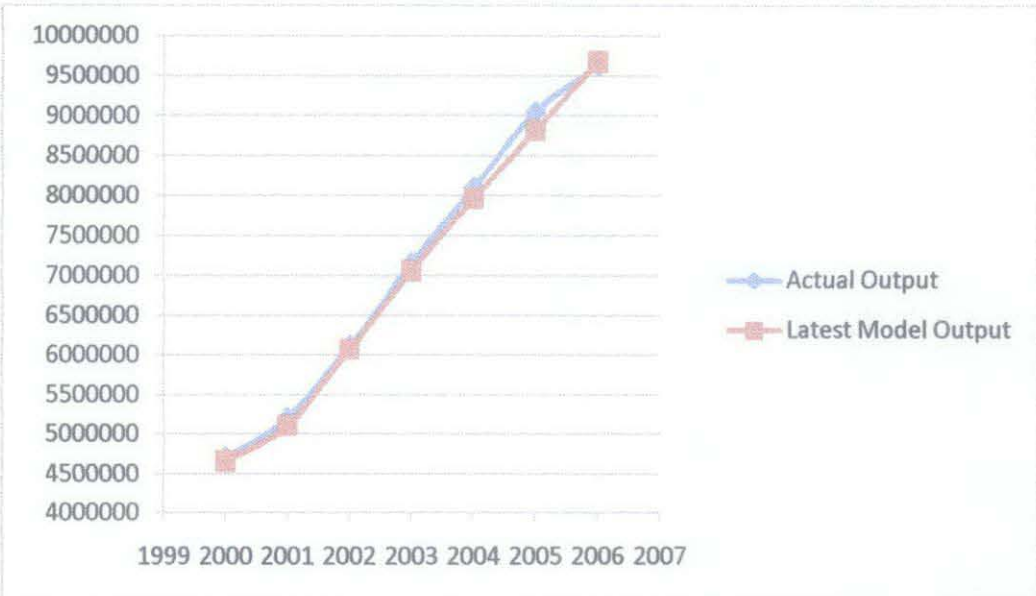


***Figure 9**: The Relationship Between Actual Output And Latest Model Output*

CHAPTER 5
CONCLUSION AND RECOMMENDATION

As for the conclusion, the variables that affecting the performance of oilfield's output has been identified and the full calculation were already constructed in order to find the value for regression coefficient, β and to predict the output of oilfield. By implementing this method, output of oilfield can also obtained by using MATLAB simulation.

Since there are too many variables that affecting the performance of oilfield's output, the author used the ordinary least squares method to minimize the sum of variables by eliminating the least important variables.

The author also uses different set of data with the most significant parameters that have been identified in first calculation using first set of data to do some comparison and verified the validity of this method. From the result and discussion it shown that the percentage error of predicted y value from the actual output is only 4.57%. This validate that this method can be implement to forecast the oilfield output.

REFERENCES

[1] Princeton University. (2003). *The Definition Of an Oilfield.* Retrieved September 19, 2010, from http://www.thefreedictionary.com/oilfield

[2] Daniel Yergin. 1999. *"The Prize: The Epic Quest for Oil, Money, and Power"*,United Sates, Simon Schuster

[3] Ling Guo and Xianghui Dei , 2009, "Application of Improved Multiple Linear Regression Method in Oilfield Output Forecasting," *International Conference on Information Management, Innovation Management and Industrial Engineering.*

[4] Larry W. Lake. 2007. "Petroleum Engineering Handbook Volume VII",United States, Society of Petroleum Engineers.

[5] Wikipedia. (2008). *Water Injection (Oil Production).* Retrieved September 19, 2010, from http://en.wikipedia.org/wiki/Water_injection_(oil_production)

[6] Hybrid Energy Holdings. (2003). *Oil & Gas – Drilling & Production.* Retrieved September 19, 2010, from http://www.hybridenergyholdings.com

[7] Dyne , 2007, "Enhanced Oil Recovery," *International Technology Centers.*

[8] Changjun Zhu and Xiujuan Zhao, 2009, "Application of Artificial Neural Network in the Prediction of Output in Oilfield," *International Joint Conference on Artificial Intelligence*

[9] Chen Yuanqian, Hu Jianguo and Zhang Dongjie. 1999. *Derivation of Logistic Model and Its Self-Regression Method*, China, Citic Press.

[10] Wang Junkui. 1993. *A Theoretical Discussion of Production Decline Curve of Oil and Gas Reservoirs*, China, Petroleum Exploration and Development.

[11] Wang Tao, Chen Xiang Guang and Li Yu Feng. 2006. *Optimization of Multivariate Model in Oilfield Output Prediction*, New York, Computer Simulation.

[12] Yale University. (1997). *Multiple Linear Regression*. Retrieved September 19, 2010, from http://www.stat.yale.edu/courses/1997-98/101/linmult.htm

[13] Massachusetts Institutes of Technology. (2006). 'Multiple Linear Regression' Retrieved September 19, 2010, from http://ocw.mit.edu/courses/sloan-school-of-management/15-075-applied-statistics-spring-2003/lecture-notes/lec14and15_chap11.pdf.

[14] Kutner, Natchsheim and Neter. 2004. *Applied Linear Regresion Model*,(4), New York, Mc Graw Hill

[15] Part Enander and Sjoberg, 1999, *The Matlab Handbook*, (1), London, Addision Wesley Longman.

[16] The Mathworks, Inc. (1984-2011). 'Linear Regression Models'. Retrieved September 19, 2010, fromhttp://www.mathworks.com/help/toolbox/stats/bq_676m-2.html#bq_676m-4

APPENDICES

Appendix A – Suggested Milestone for the Final Year Project

No.	Detail/ Week	1	2	3	4	5	6	7		8	9	10	11	12	13	14
1.	Selection of Project Topic															
2.	Determining Parameters															
3.	Submission of Preliminary Report															
4.	Seminar 1 (optional)															
5.	Construct draft for calculation															
6.	Submission of Progress Report															
7.	Seminar 2															
8.	Construct and analyzed calculation and data															
9.	Submission of Interim Report Final Draft															
10.	Oral Presentation															

Gantt Chart FYP 1

No.	Detail/ Week	1	2	3	4	5	6	7		8	9	10	11	12	13	14
1.	Modelling & Simulation Process															
2.	Identifying The Most Significant Parameter															
3.	Submission of Progress Report															
4.	Validate the MLR Forecasting															
5.	Result & Analysis															
6.	Poster Submission & Presentation															
7.	Submission of Draft Report															
8.	Finalized Final Report															
9.	Submission of Final Report															
10.	Oral Presentation															

Gantt Chart FYP2

Appendix B - Parameters Data from China Oilfield

year	y	x1	x2	x3	x4	x5	x6	x7	x8
1983	1442800	689	612	311	2375900	41.80%	1.45%	9.07%	1421900
1984	1417200	855	720	351	2305000	42.33%	1.53%	9.54%	1442800
1985	1466100	1028	874	426	2765900	42.93%	1.60%	9.49%	1417200
1986	1454500	1268	1087	472	3306400	46.21%	1.55%	10.25%	1466100
1987	1489400	1446	1197	652	3981400	45.80%	1.49%	9.35%	1454500
1988	1559200	1705	1417	486	4551000	47.80%	1.43%	9.08%	1489400
1989	1652300	1892	1524	458	5269100	49.30%	1.31%	9.31%	1559200
1990	2024600	2113	1761	473	6020400	52.15%	1.37%	10.13%	1652300
1991	2175900	2372	1903	506	7406200	55.46%	1.26%	10.88%	2024600
1992	2606400	2640	2123	705	8676500	59.83%	1.18%	11.54%	2175900
1993	3025300	3090	2574	689	9879800	60.87%	1.11%	12.07%	2606400
1994	3493100	3603	2826	964	11108700	63.39%	1.11%	12.96%	3025300
1995	3725800	3987	2878	1073	11832700	63.12%	1.20%	13.57%	3493100
1996	4037600	4530	3002	1003	13091800	64.79%	1.20%	14.76%	3725800
1997	4200500	4872	3172	1044	14063100	67.45%	1.07%	14.59%	4037600
1998	4398200	5110	3260	854	15760600	68.89%	1.01%	14.88%	4200500
1999	4649700	5400	3375	686	16760300	70.12%	0.95%	15.40%	4398200
2000	4712500	5524	3497	758	16519000	71.88%	0.88%	15.82%	4649700
2001	5205000	5653	3704	891	18083400	71.88%	0.91%	16.46%	4712500
2002	6115500	6958	5523	1043	19267300	72.95%	0.83%	17.22%	5205000
2003	7158700	8680	7805	1181	19580500	72.83%	0.83%	17.74%	6115500
2004	8109500	9864	8263	1319	25365000	72.28%	0.89%	17.71%	7158700
2005	9051000	11805	9522	1946	30032000	72.01%	0.84%	16.98%	8109500
2006	9623000	12314	11092	2347	32987000	72.31%	0.85%	17.20%	9051000

Appendix C - Parameters Data from China Oilfield 2

year	y_{new}	$x2_{new}$	$x4_{new}$	$x5_{new}$	$x7_{new}$
1983	1352300	407	1564500	40.96%	8.92%
1984	1326700	515	1493600	41.49%	9.39%
1985	1375600	669	1954500	42.09%	9.34%
1986	1364000	882	2495000	45.37%	10.10%
1987	1398900	992	3170000	44.96%	9.20%
1988	1468700	1212	3739600	46.96%	8.93%
1989	1561800	1319	4457700	48.46%	9.16%
1990	1934100	1556	5209000	51.31%	9.98%
1991	2085400	1698	6594800	54.62%	10.73%
1992	2515900	1918	7865100	58.99%	11.39%
1993	2934800	2369	9068400	60.03%	11.92%
1994	3402600	2621	10297300	62.55%	12.81%
1995	3635300	2673	11021300	62.28%	13.42%
1996	3947100	2797	12280400	63.95%	14.61%
1997	4110000	2967	13251700	66.61%	14.44%
1998	4307700	3055	14949200	68.05%	14.73%
1999	4559200	3170	15948900	69.28%	15.25%
2000	4622000	3292	15707600	71.04%	15.67%
2001	5114500	3499	17272000	71.04%	16.31%
2002	6025000	5318	18455900	72.11%	17.07%
2003	7068200	7600	18769100	71.99%	17.59%
2004	8019000	8058	24553600	71.44%	17.56%
2005	8960500	9317	29220600	71.17%	16.83%
2006	9532500	10887	32175600	71.47%	16.61%

SUMMARY OUTPUT FOR 1ST SCREENING TEST

Regression Statistics	
Multiple R	0.999357017447175
R Square	0.998714448320913
Adjusted R Square	0.998028820758733
Standard Error	111602.961872382
Observations	24

***Appendix D** : Regression Statistic for Step 1*

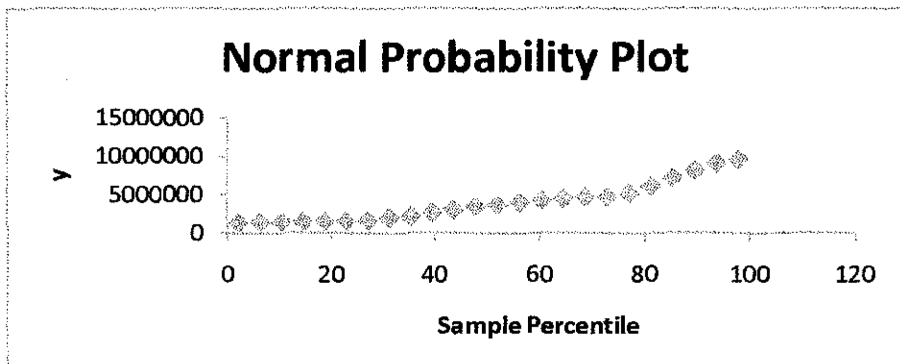
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	145142464561853	18142808070231.6	1456.64279473467	2.93003926241061E-20
Residual	15	186828316480.325	12455221098.6883		
Total	23	145329292878333			

***Appendix E** : ANOVA Table for Step 1*

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2019687.48476573	1388711.40697671	1.45436083740586	0.166450856551788
x1	177.710516292587	135.637554023751	1.31018667780952	0.209842209016502
x2	218.252106546175	85.4137604296871	2.55523355309759	0.0219687940525236
x3	193.700000254855	216.536807092068	0.894536143097815	0.385164157635301
x4	0.0768290297978363	0.0419653259492141	1.83077405119678	0.0870717795566058
x5	-5450242.21088406	2177080.6436923	-2.50346362991889	0.024338816101881
x6	-98346111.9121125	57860214.5782375	-1.69971910109546	0.109820060815556
x7	27192743.2554067	6728812.78207853	4.04123938889067	0.00106609849437169
x8	0.0259111034198541	0.204406871010171	0.126762389599735	0.900812086617006

***Appendix F** : Coefficients Table for Step 1*

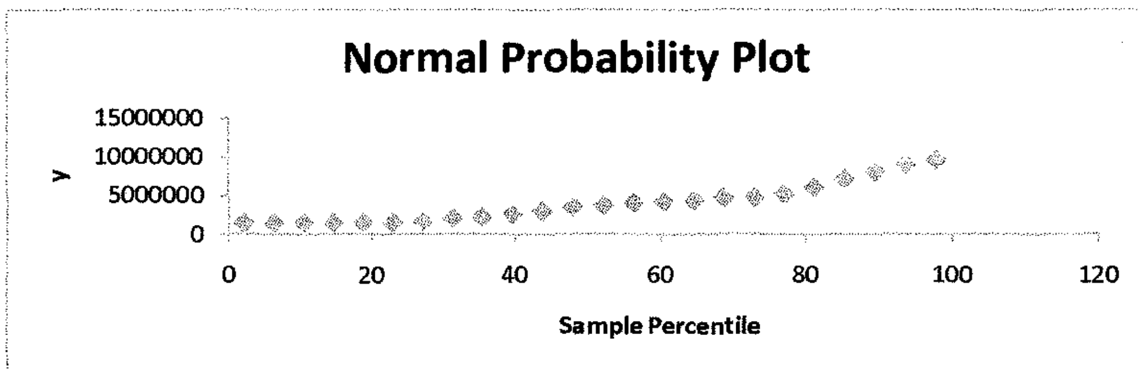


***Appendix G** : Normal Probability Plot for Step*

SUMMARY OUTPUT FOR 2ND SCREENING TEST

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2088494.8445488	1238307.13449178	1.68657256861074	0.111082462232387
x1	183.825923459356	122.80841445987	1.49685120736922	0.153897760431364
x2	217.083428276937	82.2623658960372	2.63891544951778	0.017868507309661
x3	204.281960208501	193.557788124047	1.05540553128031	0.306916687166185
x4	0.0803453600803463	0.0305061796912042	2.63374047139413	0.0180579158664623
x5	-5644643.92525134	1496976.75126212	-3.7706957843482	0.00167305561156279
x6	-100458953.735882	53676808.7504075	-1.8715522788064	0.0796674747464904
x7	27876515.1767855	3896836.65784194	7.15362680667849	2.29208861843739E-06

***Appendix H** : Coefficients Table for Step 2*

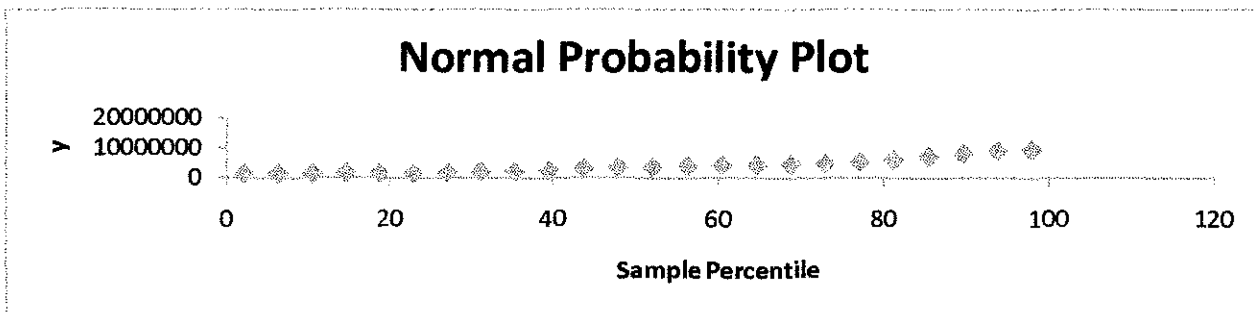


***Appendix I** : Normal Probability Plot for Step 2*

SUMMARY OUTPUT FOR 3RD SCREENING TEST

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1381124.54020234	1044725.66402452	1.32199733170327	0.203682564452449
x1	157.132290384466	120.577579737046	1.30316341335709	0.209894000078776
x2	253.739059657766	74.8203400588456	3.39131123245635	0.00347246043603412
x4	0.0949045600615735	0.0272999534297249	3.47636344164012	0.00288830551467339
x5	-4858026.68595958	1302583.82692887	-3.72953094113984	0.00166741130842961
x6	-64950637.1059472	41964068.422787	-1.54776787730806	0.140091888227209
x7	26397651.1421381	3648351.29229577	7.23550147100199	1.39323602214027E-06

***Appendix J** : Coefficients Table for Step 3*

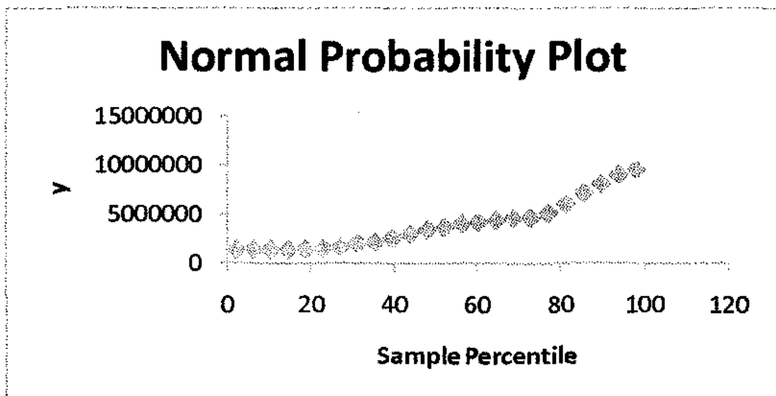


***Appendix K**: Normal Probability Plot for Step 3*

SUMMARY OUTPUT FOR 4TH SCREENING TEST

	Coefficients	Standard Error	t Stat	P-value
Intercept	1238195	1058912	1.169309	0.257531
x2	339.4357	36.37376	9.331884	2.56E-08
x4	0.124103	0.015896	7.8071	3.46E-07
x5	-5084611	1315727	-3.86449	0.001136
x6	-6.2E+07	42711987	-1.45387	0.163198
x7	28559343	3311861	8.623351	8.28E-08

***Appendix L** : Coefficients Table for Step 4*



***Appendix M** : Normal Probability Plot for Step 4*

SUMMARY OUTPUT FOR 5TH SCREENING TEST

Regression Statistics	
Multiple R	0.999154
R Square	0.998308
Adjusted R Square	0.997952
Standard Error	113756.8
Observations	24

***Appendix N** : Regression Statistic for Step 5*

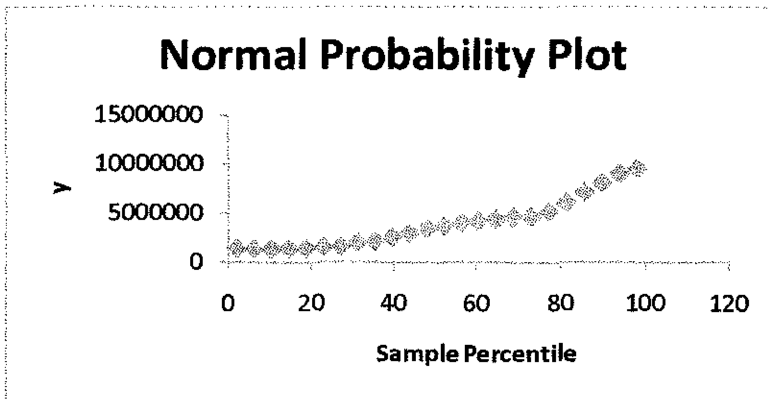
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	1.45E+14	3.63E+13	2802.873	4.9E-26
Residual	19	2.46E+11	1.29E+10		
Total	23	1.45E+14			

***Appendix O**: ANOVA Table for Step 5*

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-259910	250998.7	-1.0355	0.313434
x2	352.2079	36.31672	9.698228	8.6E-09
x4	0.123018	0.016337	7.529832	4.07E-07
x5	-3660564	903866.8	-4.04989	0.000684
x7	27702445	3353151	8.261616	1.04E-07

Appendix P *: Coefficients Table for Step 5*



Appendix Q *: Normal Probability Plot for Step 5*

SUMMARY OUTPUT FOR THE LATEST DATA

Regression Statistics	
Multiple R	0.999295627
R Square	0.998591749
Adjusted R Square	0.998295275
Standard Error	103786.237
Observations	24

***Appendix R**: Regression Statistic for latest model*

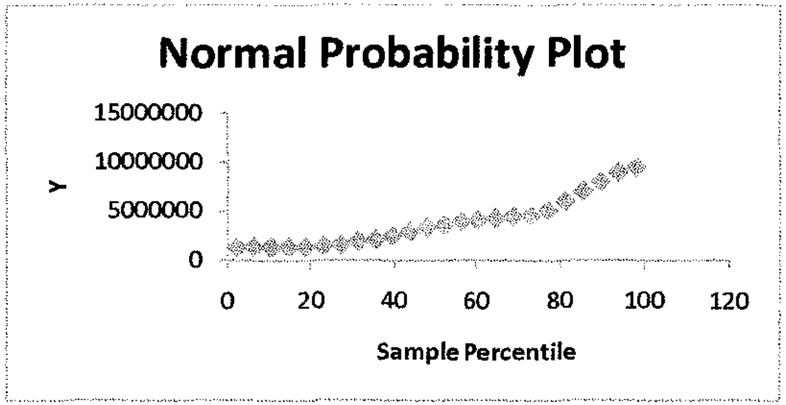
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	1.45125E+14	3.62812E+13	3368.228999	8.57226E-27
Residual	19	2.0466E+11	10771582994		
Total	23	1.45329E+14			

***Appendix S**: ANOVA Table for latest model*

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-79077.80482	230847.479	-0.342554336	0.735694265
X1	352.1687732	33.09573993	10.64090949	1.91727E-09
X2	0.127223503	0.014897999	8.539637213	6.272E-08
X3	-3840142.614	828426.6044	-4.635465102	0.000180394
X4	27504246.4	2968894.512	9.264137303	1.77654E-08

***Appendix T**: Coefficients Table for latest model*



***Appendix U** : Normal Probability Plot for latest model*

Appendix V - Matlab Coding for 8 parameters

```
>> A
=[1442800,689,612,311,2375900,0.418000000000000,0.0145000000000000,0.0907000000000000,142
1900;1417200,855,720,351,2305000,0.423300000000000,0.0153000000000000,0.0954000000000000,
1442800;1466100,1028,874,426,2765900,0.429300000000000,0.0160000000000000,0.0949000000000
000,1417200;1454500,1268,1087,472,3306400,0.462100000000000,0.0155000000000000,0.10250000
0000000,1466100;1489400,1446,1197,652,3981400,0.458000000000000,0.0149000000000000,0.0935
000000000000,1454500;1559200,1705,1417,486,4551000,0.478000000000000,0.0143000000000000,0
.0908000000000000,1489400;1652300,1892,1524,458,5269100,0.493000000000000,0.0131000000000
000,0.0931000000000000,1559200;2024600,2113,1761,473,6020400,0.521500000000000,0.01370000
00000000,0.101300000000000,1652300;2175900,2372,1903,506,7406200,0.554600000000000,0.0126
000000000000,0.108800000000000,2024600;2606400,2640,2123,705,8676500,0.598300000000000,0.
0118000000000000,0.115400000000000,2175900;3025300,3090,2574,689,9879800,0.60870000000000
0,0.0111000000000000,0.120700000000000,2606400;3493100,3603,2826,964,11108700,0.633900000
000000,0.0111000000000000,0.129600000000000,3025300;3725800,3987,2878,1073,11832700,0.631
200000000000,0.0120000000000000,0.135700000000000,3493100;4037600,4530,3002,1003,13091800
,0.647900000000000,0.0120000000000000,0.147600000000000,3725800;4200500,4872,3172,1044,14
063100,0.674500000000000,0.0107000000000000,0.145900000000000,4037600;4398200,5110,3260,8
54,15760600,0.688900000000000,0.0101000000000000,0.148800000000000,4200500;4649700,5400,3
375,686,16760300,0.701200000000000,0.00950000000000000,0.154000000000000,4398200;4712500,
5524,3497,758,16519000,0.718800000000000,0.00880000000000000,0.158200000000000,4649700;52
05000,5653,3704,891,18083400,0.718800000000000,0.00910000000000000,0.164600000000000,4712
500;6115500,6958,5523,1043,19267300,0.729500000000000,0.00830000000000000,0.172200000000
00,5205000;7158700,8680,7805,1181,19580500,0.728300000000000,0.00830000000000000,0.177400
000000000,6115500;8109500,9864,8263,1319,25365000,0.722800000000000,0.00890000000000000,0
.177100000000000,7158700;9051000,11805,9522,1946,30032000,0.720100000000000,0.00840000000
000000,0.169800000000000,8109500;9623000,12314,11092,2347,32987000,0.723100000000000,0.00
850000000000000,0.172000000000000,9051000;]
```

```
>> x1 = A(:,2:9);

>> y = A(:,1);

>> stats = regstats(y,x1);

>> t = stats.tstat;

>> CoeffTable = dataset({t.beta,'Coef'},{t.se,'StdErr'}, ...
                        {t.t,'tStat'},{t.pval,'pVal'})
```

CoeffTable =

Coef	StdErr	tStat	pVal
2.0197e+006	1.3887e+006	1.4544	0.16645
177.71	135.64	1.3102	0.20984
218.25	85.414	2.5552	0.021969
193.7	216.54	0.89454	0.38516
0.076829	0.041965	1.8308	0.087072
-5.4502e+006	2.1771e+006	-2.5035	0.024339
-9.8346e+007	5.786e+007	-1.6997	0.10982
2.7193e+007	6.7288e+006	4.0412	0.0010661
0.025911	0.20441	0.12676	0.90081

```
>> f = stats.fstat;

fprintf('\n')
fprintf('Regression ANOVA');
fprintf('\n\n')

fprintf('%6s','Source');
fprintf('%10s','df','SS','MS','F','P');
fprintf('\n')

fprintf('%6s','Regr');
fprintf('%10.4f',f.dfr,f.ssr,f.ssr/f.dfr,f.f,f.pval);
fprintf('\n')

fprintf('%6s','Resid');
fprintf('%10.4f',f.dfe,f.sse,f.sse/f.dfe);
fprintf('\n')

fprintf('%6s','Total');
fprintf('%10.4f',f.dfe+f.dfr,f.sse+f.ssr);
fprintf('\n')
```

```
Regression ANOVA


Source       df        SS        MS         F         P
  Regr    8.0000145142464561853.060018142808070231.6330 1456.6428    0.0000
 Resid   15.0000186828316480.326012455221098.6884
 Total   23.0000145329292878333.3700
```

Appendix W - Matlab Coding for 7 parameters

```
>> A
=[1442800,689,612,311,2375900,0.418000000000000,0.0145000000000000,0.0907000000000000;141
7200,855,720,351,2305000,0.423300000000000,0.0153000000000000,0.0954000000000000;1466100,
1028,874,426,2765900,0.429300000000000,0.0160000000000000,0.0949000000000000;1454500,1268
,1087,472,3306400,0.462100000000000,0.0155000000000000,0.102500000000000;1489400,1446,119
7,652,3981400,0.458000000000000,0.0149000000000000,0.0935000000000000;1559200,1705,1417,4
86,4551000,0.478000000000000,0.0143000000000000,0.0908000000000000;1652300,1892,1524,458,
5269100,0.493000000000000,0.0131000000000000,0.0931000000000000;2024600,2113,1761,473,602
0400,0.521500000000000,0.0137000000000000,0.101300000000000;2175900,2372,1903,506,7406200
,0.554600000000000,0.0126000000000000,0.108800000000000;2606400,2640,2123,705,8676500,0.5
98300000000000,0.0118000000000000,0.115400000000000;3025300,3090,2574,689,9879800,0.60870
0000000000,0.0111000000000000,0.120700000000000;3493100,3603,2826,964,11108700,0.63390000
0000000,0.0111000000000000,0.129600000000000;3725800,3987,2878,1073,11832700,0.6312000000
00000,0.0120000000000000,0.135700000000000;4037600,4530,3002,1003,13091800,0.647900000000
000,0.0120000000000000,0.147600000000000;4200500,4872,3172,1044,14063100,0.67450000000000
0,0.0107000000000000,0.145900000000000;4398200,5110,3260,854,15760600,0.688900000000000,0
.0101000000000000,0.148800000000000;4649700,5400,3375,686,16760300,0.701200000000000,0.00
950000000000000,0.154000000000000;4712500,5524,3497,758,16519000,0.718800000000000,0.0088
0000000000000,0.158200000000000;5205000,5653,3704,891,18083400,0.718800000000000,0.009100
00000000000,0.164600000000000;6115500,6958,5523,1043,19267300,0.729500000000000,0.0083000
0000000000,0.172200000000000;7158700,8680,7805,1181,19580500,0.728300000000000,0.00830000
000000000,0.177400000000000;8109500,9864,8263,1319,25365000,0.722800000000000,0.008900000
00000000,0.177100000000000;9051000,11805,9522,1946,30032000,0.720100000000000,0.008400000
00000000,0.169800000000000;9623000,12314,11092,2347,32987000,0.723100000000000,0.00850000
000000000,0.172000000000000;]
```

```
>> x1 = A(:,2:8);

>> y= A(:,1);

>> stats = regstats(y,x1);

>> t = stats.tstat;

>> CoeffTable = dataset({t.beta,'Coef'},{t.se,'StdErr'}, ...
                   {t.t,'tStat'},{t.pval,'pVal'})
```

```
CoeffTable =


    Coef              StdErr          tStat       pVal

      2.0885e+006     1.2383e+006     1.6866        0.11108

           183.83          122.81     1.4969         0.1539

           217.08          82.262     2.6389       0.017869

           204.28          193.56     1.0554        0.30692

         0.080345        0.030506     2.6337       0.018058

     -5.6446e+006      1.497e+006    -3.7707      0.0016731

     -1.0046e+008     5.3677e+007    -1.8716       0.079667

      2.7877e+007     3.8968e+006     7.1536    2.2921e-006
```

```
>> f = stats.fstat;

fprintf('\n')
fprintf('Regression ANOVA');
fprintf('\n\n')

fprintf('%6s','Source');
fprintf('%10s','df','SS','MS','F','P');
fprintf('\n')

fprintf('%6s','Regr');
fprintf('%10.4f',f.dfr,f.ssr,f.ssr/f.dfr,f.f,f.pval);
fprintf('\n')

fprintf('%6s','Resid');
fprintf('%10.4f',f.dfe,f.sse,f.sse/f.dfe);
fprintf('\n')

fprintf('%6s','Total');
fprintf('%10.4f',f.dfe+f.dfr,f.sse+f.ssr);
fprintf('\n')
```

```
Regression ANOVA


Source         df          SS          MS           F           P
  Regr    7.0000145142264422599.2500207346092O3228.4650 1773.8143      0.0000
 Resid   16.0000187028455734.154911689278483.3847
 Total   23.0000145329292878333.4100
```

Appendix X - Matlab Coding for 6 parameters

```
>> A
=[1442800,689,612,2375900,0.418000000000000,0.014500000000000,0.090700000000000;1417200
,855,720,2305000,0.423300000000000,0.015300000000000,0.095400000000000;1466100,1028,874
,2765900,0.429300000000000,0.016000000000000,0.094900000000000;1454500,1268,1087,330640
0,0.462100000000000,0.015500000000000,0.102500000000000;1489400,1446,1197,3981400,0.4580
00000000000,0.014900000000000,0.093500000000000;1559200,1705,1417,4551000,0.47800000000
0000,0.014300000000000,0.090800000000000;1652300,1892,1524,5269100,0.493000000000000,0.
013100000000000,0.093100000000000;2024600,2113,1761,6020400,0.521500000000000,0.0137000
00000000,0.101300000000000;2175900,2372,1903,7406200,0.554600000000000,0.01260000000000
0,0.108800000000000;2606400,2640,2123,8676500,0.598300000000000,0.011800000000000,0.1154
00000000000;3025300,3090,2574,9879800,0.608700000000000,0.011100000000000,0.120700000000
000;3493100,3603,2826,11108700,0.633900000000000,0.011100000000000,0.129600000000000;372
5800,3987,2878,11832700,0.631200000000000,0.012000000000000,0.135700000000000;4037600,45
30,3002,13091800,0.647900000000000,0.012000000000000,0.147600000000000;4200500,4872,3172
,14063100,0.674500000000000,0.010700000000000,0.145900000000000;4398200,5110,3260,157606
00,0.688900000000000,0.010100000000000,0.148800000000000;4649700,5400,3375,16760300,0.70
1200000000000,0.009500000000000,0.154000000000000;4712500,5524,3497,16519000,0.71880000
0000000,0.008800000000000,0.158200000000000;5205000,5653,3704,18083400,0.71880000000000
0,0.009100000000000,0.164600000000000;6115500,6958,5523,19267300,0.729500000000000,0.00
8300000000000,0.172200000000000;7158700,8680,7805,19580500,0.728300000000000,0.00830000
0000000,0.177400000000000;8109500,9864,8263,25365000,0.722800000000000,0.008900000000
000,0.177100000000000;9051000,11805,9522,30032000,0.720100000000000,0.008400000000000,0
.169800000000000;9623000,12314,11092,32987000,0.723100000000000,0.008500000000000,0.172
000000000000;]
```

```
>> x1 = A(:,2:7);

>> y = A(:,1);

>> stats = regstats(y,x1);

>> t = stats.tstat;

>> CoeffTable = dataset({t.beta,'Coef'},{t.se,'StdErr'}, ...
                        {t.t,'tStat'},{t.pval,'pVal'})
```

```
CoeffTable =


    Coef              StdErr           tStat        pVal

      1.3811e+006     1.0447e+006       1.322          0.20368

           157.13          120.58      1.3032          0.20989

           253.74           74.82      3.3913        0.0034725

         0.094905          0.0273      3.4764        0.0028883

      -4.858e+006     1.3026e+006     -3.7295        0.0016674

    -6.4951e+007     4.1964e+007     -1.5478          0.14009

      2.6398e+007     3.6484e+006      7.2355      1.3932e-006
```

```
>> f = stats.fstat;

fprintf('\n')
fprintf('Regression ANOVA');
fprintf('\n\n')

fprintf('%6s','Source');
fprintf('%10s','df','SS','MS','F','P');
fprintf('\n')

fprintf('%6s','Regr');
fprintf('%10.4f',f.dfr,f.ssr,f.ssr/f.dfr,f.f,f.pval);
fprintf('\n')

fprintf('%6s','Resid');
fprintf('%10.4f',f.dfe,f.sse,f.sse/f.dfe);
fprintf('\n')

fprintf('%6s','Total');
fprintf('%10.4f',f.dfe+f.dfr,f.sse+f.ssr);
fprintf('\n')
```

```
Regression ANOVA


Source        df          SS          MS          F          P

  Regr    6.000014512924395931 6.250024188207326552.7070 2055.4949     0.0000

 Resid   17.00002000489190 17.117311767583471.5951

 Total   23.000014532929287 8333.3700
```

Appendix Y - Matlab Coding for 5 parameters

```
>> A =
[1442800,612,2375900,0.418000000000000,0.0145000000000000,0.0907000000000000;1417200,720,
2305000,0.423300000000000,0.0153000000000000,0.0954000000000000;1466100,874,2765900,0.429
300000000000,0.0160000000000000,0.0949000000000000;1454500,1087,3306400,0.462100000000000
,0.0155000000000000,0.102500000000000;1489400,1197,3981400,0.458000000000000,0.0149000000
000000,0.0935000000000000;1559200,1417,4551000,0.478000000000000,0.0143000000000000,0.090
8000000000000;1652300,1524,5269100,0.493000000000000,0.0131000000000000,0.093100000000000
0;2024600,1761,6020400,0.521500000000000,0.0137000000000000,0.101300000000000;2175900,190
3,7406200,0.554600000000000,0.0126000000000000,0.108800000000000;2606400,2123,8676500,0.5
98300000000000,0.0118000000000000,0.115400000000000;3025300,2574,9879800,0.608700000000000
0,0.0111000000000000,0.120700000000000;3493100,2826,11108700,0.633900000000000,0.01110000
00000000,0.129600000000000;3725800,2878,11832700,0.631200000000000,0.0120000000000000,0.1
35700000000000;4037600,3002,13091800,0.647900000000000,0.0120000000000000,0.147600000000000
00;4200500,3172,14063100,0.674500000000000,0.0107000000000000,0.145900000000000;4398200,3
260,15760600,0.688900000000000,0.0101000000000000,0.148800000000000;4649700,3375,16760300
,0.701200000000000,0.0095000000000000,0.154000000000000;4712500,3497,16519000,0.71880000
0000000,0.0088000000000000,0.158200000000000;5205000,3704,18083400,0.718800000000000,0.0
091000000000000,0.164600000000000;6115500,5523,19267300,0.729500000000000,0.008300000000
00000,0.172200000000000;7158700,7805,19580500,0.728300000000000,0.0083000000000000,0.177
400000000000;8109500,8263,25365000,0.722800000000000,0.0089000000000000,0.177100000000000
0;9051000,9522,30032000,0.720100000000000,0.0084000000000000,0.169800000000000;9623000,1
1092,32987000,0.723100000000000,0.0085000000000000,0.172000000000000;]
```

```
>> x1 = A(:,2:6);

>> y = A(:,1);

>> stats = regstats(y,x1);

>> t = stats.tstat;

>> CoeffTable = dataset({t.beta,'Coef'},{t.se,'StdErr'}, ...

                        {t.t,'tStat'},{t.pval,'pVal'})
```

```
CoeffTable =


    Coef              StdErr         tStat       pVal

     1.2382e+006      1.0589e+006     1.1693         0.25753

          339.44           36.374     9.3319     2.5604e-008

          0.1241         0.015896     7.8071     3.4633e-007

    -5.0846e+006      1.3157e+006    -3.8645       0.0011358

    -6.2098e+007      4.2712e+007    -1.4539          0.1632

     2.8559e+007      3.3119e+006     8.6234     8.2838e-008
```

```
>> f = stats.fstat;

fprintf('\n')
fprintf('Regression ANOVA');
fprintf('\n\n')

fprintf('%6s','Source');
fprintf('%10s','df','SS','MS','F','P');
fprintf('\n')

fprintf('%6s','Regr');
fprintf('%10.4f',f.dfr,f.ssr,f.ssr/f.dfr,f.f,f.pval);
fprintf('\n')

fprintf('%6s','Resid');
fprintf('%10.4f',f.dfe,f.sse,f.sse/f.dfe);
fprintf('\n')

fprintf('%6s','Total');
fprintf('%10.4f',f.dfe+f.dfr,f.sse+f.ssr);
fprintf('\n')
```

```
Regression ANOVA


Source        df          SS          MS            F           P

  Regr     5.00001451092598388588.940029021851967717.7890 2374.1586     0.0000

 Resid    18.0000220033039744.398212224057763.5777

 Total    23.0000145329292878333.3400
```

Appendix Z - Matlab Coding for 4 parameters

```
>> A =
[1442800,612,2375900,0.418000000000000,0.0907000000000000;1417200,720,2305000,0.423300000
000000,0.0954000000000000;1466100,874,2765900,0.429300000000000,0.0949000000000000;145450
0,1087,3306400,0.462100000000000,0.102500000000000;1489400,1197,3981400,0.458000000000000
,0.0935000000000000;1559200,1417,4551000,0.478000000000000,0.0908000000000000;1652300,152
4,5269100,0.493000000000000,0.0931000000000000;2024600,1761,6020400,0.521500000000000,0.1
01300000000000;2175900,1903,7406200,0.554600000000000,0.108800000000000;2606400,2123,8676
500,0.598300000000000,0.115400000000000;3025300,2574,9879800,0.608700000000000,0.12070000
0000000;3493100,2826,11108700,0.633900000000000,0.129600000000000;3725800,2878,11832700,0
.631200000000000,0.135700000000000;4037600,3002,13091800,0.647900000000000,0.147600000000
000;4200500,3172,14063100,0.674500000000000,0.145900000000000;4398200,3260,15760600,0.688
900000000000,0.148800000000000;4649700,3375,16760300,0.701200000000000,0.154000000000000;
4712500,3497,16519000,0.718800000000000,0.158200000000000;5205000,3704,18083400,0.7188000
00000000,0.164600000000000;6115500,5523,19267300,0.729500000000000,0.172200000000000;7158
700,7805,19580500,0.728300000000000,0.177400000000000;8109500,8263,25365000,0.72280000000
0000,0.177100000000000;9051000,9522,30032000,0.720100000000000,0.169800000000000;9623000,
11092,32987000,0.723100000000000,0.172000000000000;]
```

```
>> x1 = A(:,2:5);

>> y = A(:,1);

>> stats = regstats(y,x1);

>> t = stats.tstat;

>> CoeffTable = dataset({t.beta,'Coef'},{t.se,'StdErr'}, ...
                    {t.t,'tStat'},{t.pval,'pVal'})
```

```
CoeffTable =


    Coef             StdErr        tStat       pVal

    -2.5991e+005     2.51e+005     -1.0355       0.31343
          352.21        36.317      9.6982     8.602e-009
         0.12302      0.016337      7.5298    4.0715e-007
    -3.6606e+006    9.0387e+005    -4.0499    0.00068367
     2.7702e+007    3.3532e+006     8.2616    1.0358e-007
```

```
>> f = stats.fstat;

fprintf('\n')
fprintf('Regression ANOVA');
fprintf('\n\n')

fprintf('%6s','Source');
fprintf('%10s','df','SS','MS','F','P');
fprintf('\n')

fprintf('%6s','Regr');
fprintf('%10.4f',f.dfr,f.ssr,f.ssr/f.dfr,f.f,f.pval);
fprintf('\n')

fprintf('%6s','Resid');
fprintf('%10.4f',f.dfe,f.sse,f.sse/f.dfe);
fprintf('\n')

fprintf('%6s','Total');
fprintf('%10.4f',f.dfe+f.dfr,f.sse+f.ssr);
fprintf('\n')
```

```
Regression ANOVA


Source        df          SS          MS           F           P

  Regr     4.0000145083421483031.160036270855370757.7890 2802.8728      0.0000

 Resid    19.0000245871395302.173412940599752.7460

 Total    23.0000145329292878333.3400
```

Appendix AA - Matlab Coding for Latest Model

```
A =
[1352300,407,1564500,0.409600000000000,0.0892000000000000;1326700,515,1493600,0.414900000
000000,0.0939000000000000;1375600,669,1954500,0.420900000000000,0.0934000000000000;136400
0,882,2495000,0.453700000000000,0.101000000000000;1398900,992,3170000,0.449600000000000,0
.0920000000000000;1468700,1212,3739600,0.469600000000000,0.0893000000000000;1561800,1319,
4457700,0.484600000000000,0.0916000000000000;1934100,1556,5209000,0.513100000000000,0.099
8000000000000;2085400,1698,6594800,0.546200000000000,0.107300000000000;2515900,1918,78651
00,0.589900000000000,0.113900000000000;2934800,2369,9068400,0.600300000000000,0.119200000
000000;3402600,2621,10297300,0.625500000000000,0.128100000000000;3635300,2673,11021300,0.
622800000000000,0.134200000000000;3947100,2797,12280400,0.639500000000000,0.146100000000
00;4110000,2967,13251700,0.666100000000000,0.144400000000000;4307700,3055,14949200,0.6805
00000000000,0.147300000000000;4559200,3170,15948900,0.692800000000000,0.152500000000000;4
622000,3292,15707600,0.710400000000000,0.156700000000000;5114500,3499,17272000,0.71040000
0000000,0.163100000000000;6025000,5318,18455900,0.721100000000000,0.170700000000000;70682
00,7600,18769100,0.719900000000000,0.175900000000000;8019000,8058,24553600,0.714400000000
000,0.175600000000000;8960500,9317,29220600,0.711700000000000,0.168300000000000;9532500,1
0887,32175600,0.714700000000000,0.166100000000000;]
```

```
>> x1 = A(:,2:5);

>> y = A(:,1);

>> stats = regstats(y,x1);

>> t = stats.tstat;

>> CoeffTable = dataset({t.beta,'Coef'},{t.se,'StdErr'}, ...
                    {t.t,'tStat'},{t.pval,'pVal'})
```

```
CoeffTable =


    Coef               StdErr          tStat         pVal

          -79078       2.3085e+005     -0.34255          0.73569

          352.17            33.096       10.641      1.9173e-009

         0.12722          0.014898       8.5396       6.272e-008

    -3.8401e+006       8.2843e+005      -4.6355       0.00018039

     2.7504e+007       2.9689e+006       9.2641      1.7765e-008
```

```
>> f = stats.fstat;

fprintf('\n')
fprintf('Regression ANOVA');
fprintf('\n\n')

fprintf('%6s','Source');
fprintf('%10s','df','SS','MS','F','P');
fprintf('\n')

fprintf('%6s','Regr');
fprintf('%10.4f',f.dfr,f.ssr,f.ssr/f.dfr,f.f,f.pval);
fprintf('\n')

fprintf('%6s','Resid');
fprintf('%10.4f',f.dfe,f.sse,f.sse/f.dfe);
fprintf('\n')

fprintf('%6s','Total');
fprintf('%10.4f',f.dfe+f.dfr,f.sse+f.ssr);
fprintf('\n')
```

```
Regression ANOVA


Source        df         SS         MS          F          P
  Regr    4.00001451246328014449.1200362811582003 62.2810 3368.2290     0.0000
 Resid   19.0000204660076884.201810771582993.9054
 Total   23.00001453292928783 33.3100
```