ENERGY MANAGEMENT IN THE JORDANIAN CEMENT INDUSTRY

Thesis Submitted for the degree of

Doctor of Philosophy

Department of Management and Organisation

University of Stirling

By

Hatem Al- Halawani Al - Tamimi

March 2000

VOLUME ONE

Part Two

Contents:

Chapter Seven

Statistical and Economical

Modelling and Analysis

Contents

Chapter 7

Statistical and Economic Modelling and Analysis

7.1 Introduction

In chapter six, preliminary detailed empirical analysis of the main factors affecting energy consumption was carried out. These factors include availability, production rate, average number and duration of stoppages and other factors affecting the quality of cement produced.

In essence, this preliminary empirical analysis indicates, in a clear manner, that the activation of energy control factors within an EMS is not only possible but also beneficial. However, a detailed statistical and economical analysis to define the relationship between energy consumption and the control factors is needed and will be covered in this chapter.

For this purpose several statistical models will be developed to represent the relationships between energy consumption on one hand, and management functions and practices as given by the variables on the other hand. Detailed statistical analysis of the factors affecting the consumption of energy in the cement industry in order to develop the necessary mathematical functions relating energy consumption to the selected significant factors, will be established, which will help to demonstrate, mathematically, the significance of energy management in controlling the energy consumption and cost. The statistical model will use historical data for certain selected independent factors and the desired dependent factors; namely electricity and fuel consumption. The outcome of this statistical analysis will be a model, which may be used, to a certain degree of confidence, to describe the relationship between each of the two dependent variables (electricity and fuel consumption) and the selected independent variables.

Such derivation of relationships between energy consumption (as a dependent variable) and management functions (as independent variables) would allow managers to plan and control these functions in order to optimise management variables and consequently the consumption of electrical or thermal energy.

The statistical models will be, in turn, employed as a basis for developing economic models and translating the impact of sound management practices and tools into quantifiable economic variables. Managers can immediately use such variables for developing "best practices" policies and decisions. Economic variables such as the net monetary savings and the rate of return on investment can be used for this purpose. Transforming the statistical model into a practical economic model is preferable to relate the saving of energy consumption as a function of the cost of improving the independent variables affecting energy consumption. The objective of building the economic model is to verify whether there will be financial gains from improving the control variables as predicted by the statistical model or not. Furthermore, the size of such gains can be assessed using the economic model.

However, before delving into statistical treatments, it is quite useful at this stage of development to present the basic aspects of economic modelling that will illustrate how key factors (inputs and outputs) interact to produce the net effect on the operational efficiency of the production facility. The following section addresses the economics of production and the causes of inefficiency in the cement manufacturing.

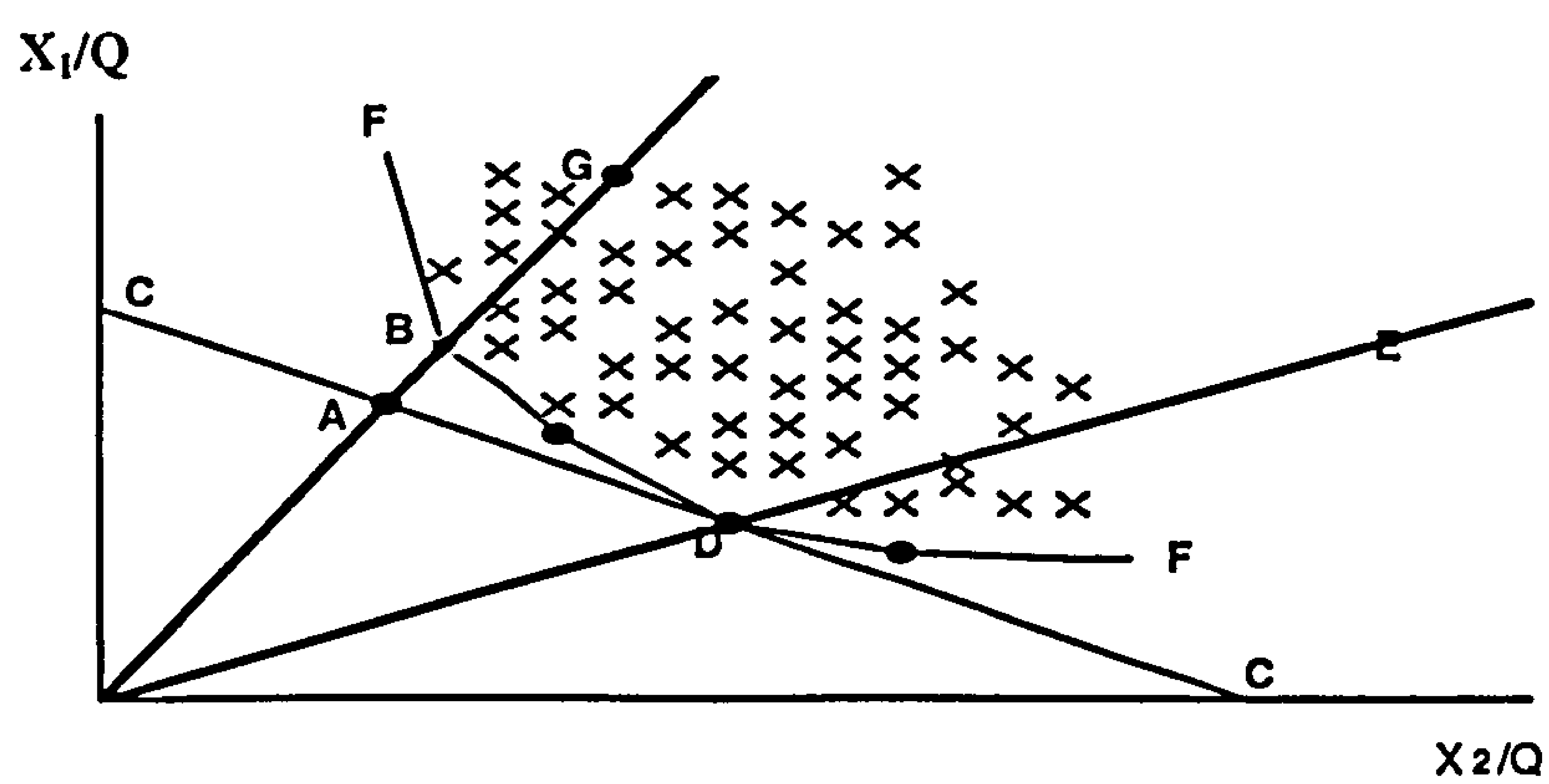
7.1.1 Economics of Production

As emphasised in the previous chapters, the cement industry is highly energy-intensive. The cost of energy represents more than 35% of the total cost and more than 75% of the variable cost in this industry. Thus, efficiency in using energy inputs (electricity and fuel) would result in significant savings in the production cost of cement and would increase its competitiveness in local and regional markets. It is important, therefore, to identify the magnitude and sources in efficiency and waste in energy consumption so that we can recommend appropriate policy measure that would promote efficient use of energy in the industry.

We shall use a *frontier production function* model that was originally proposed by Farrell (1957), and is often used to measure technical efficiency in manufacturing plants. The concept of the "frontier" or the "best practice" production function represents a set of observations from which the maximum output is achievable from a given stock of measured inputs. The amount by which measured output is less than potential output is usually defined as technical inefficiency. In actual application to the cement industry of this concept we can consider that running the production line with partial load i.e., reduced production rate and continuous stoppages of production lines and lower availability can be considered major factors leading to technical inefficiency.

Consider a production function in the form

$$q = f(x_1, x_2, \ldots, x_j)$$



This function summarises the various technical possibilities of converting inputs $x_1, x_2, \ldots, x_j$ into outputs. This represents all technically efficient combinations of inputs used to produce a given level of output. A technically inefficient firm may use the same amount of inputs to produce less output, so observations of inputs and outputs may be described by the function.

$$q \leq f(x_1, x_2, \ldots, x_j)$$

However, only input combinations on the frontier represent efficient production function.

This can be illustrated in the above figure. Consider an industry using two inputs: X_1 and X_2. The two axes measure per unit factor inputs, X_1/Q and X_2/Q, respectively. Each point on the graph represents an observation of various input combinations used to produce one unit of output. Given all production points, we can construct a production frontier, often called frontier unit isoquant, which is represented by the curve FF. Clearly; this curve consists of all minimum input combinations, and thus represents the most efficient points.

This model does not make any assumptions about the degree of returns to scale. Instead, it allows for these returns to be tested based on efficiency measures at various scales. In fact, one of the major purposes of using this model is to test whether economies of scale are achieved in the cement industry in Jordan or not. Furthermore, the model does not identify the sources of inefficiency in the industry, although it acknowledges the existence of various types of inefficiency. Indeed, this model has been widely used by researchers as a theoretical framework to explain different types of inefficiency, including technical and X-inefficiencies. The sources of these types has ranged from exogenous random shocks such as machine breakdown, disruption of supplies, and workers' strikes, to internal factors such as pursuing goals other than profit maximisation, lack of competitive pressure, over use of inputs, low production rate, low availability etc. These sources of inefficiency differ from one industry to another and from one firm to another.

7.1.2 Causes of inefficiency in the cement manufacturing

There are various causes of inefficiency in the cement manufacturing processes; many of the technical inefficiency in the operations of the cement production lines resulted from the emergency stoppages of the machines in the plants. The number and duration of these stoppages have negative impact on the consumption of both types of energy: electricity and fuel and on the average production rate. These factors and others are considered in this research, and a set of variables is going to be selected for statistical analysis. The potential impact of these factors will be assessed and, accordingly, the main hypotheses of the thesis will be formulated. The following sections present the main statistical tool used in this research, that is, linear regression.

7.2 Regression Analysis

Preface

This chapter provides statistical analysis of the available data. It consists of four main parts. **The first part** deals with the theoretical statistical background needed in the analysis. This part considers the following points.

1) A justification of the use of regression analysis is based on the close relationship between the purposes of regression analysis and the main objectives of our research problem.
2) The main steps of model construction have been stated.
3) A review of ordinary least square linear regression is given together with its assumptions and the methods used in checking the adequacy of the model. We explained the meaning of each assumption, the problems that may arise if an assumption fails to hold, and some possible remedy solution of that failure. Moreover, we discussed methods of testing the obtained model as a whole and testing each term separately.

4) Alternative regression methods have been discussed including ridge regression, which is useful if multicollinearity is a problem, robust regression that is useful in case there is a problem in the normality assumption and/or there are outliers in the data.

5) Some nonlinear regression models have been explored. These models are multiplicative models, polynomial models with and without interaction, and logarithmic models.

The second part deals with the available data. This part considers the following points.

1) The dependent and independent variables have been defined. Some physical considerations for their selection have been considered. Limitations of the available data are stated together with the researcher expectations about the behavioural effects of the independent variables on the dependent variables.

2) Exploration of the available data has been considered. This exploration depends on scatter plots of dependent variables against independent variables in order to explore possible linear relationships and to detect possible outliers. Correlations between dependent and independent variables are obtained in order to explore possible linear relationships. Moreover, correlations between independent variables are obtained to explore possible collinearity problems. Finally the normality of the dependent variables has been explored using histograms, normal probability plots and three normality tests. Some transformed data has also been explored.

The third part deals with the problem of fitting regression models for electricity and fuel consumptions of each kiln based on the available independent variables. This part went as follows.

1) For kiln 1, detailed work is considered with discussion of all obtained results. For other kilns, only summary of the results have been considered. The detailed computer output is reported in the appendices.

2) The analysis for any kiln includes the following

   a) Data exploration including screening of scatter plots, transformed data, normality of data, correlation coefficients, and possible outliers.

   b) Selection problem of significant independent variables is based on all possible regression models and their Cp-value, stepwise regression procedures, and t-test of the coefficients of the fitted models.

   c) Exploration of residuals includes histograms, normal probability plots, plot of residuals against predicted values and against each of the independent variables together with skewness, kurtosis and omnibus normality tests.

   d) Randomness of errors is tested through serial correlations and Durbin-Watson test.

   e) Multicollinearity is tested through R-squared vs. other X's, variance inflation factor, eigenvalues and condition numbers.

3) Ordinary least square multiple linear regression, ridge regression, robust regression, and nonlinear regression models have been fitted.

4) The multiple linear regression and the Andrew's robust regression models were run twice for each kiln. First with all available independent variables included in the model. The obtained model is called the full model. In the second run we have included only the significant independent variables. The obtained model with only these significant independent variables is called the final model.

The fourth part concentrates on summarizing the obtained results, comparing the obtained models, stating the main conclusions. It also compares the results with those obtained from the analysis of HoldarBank data.

7.2.1 Introduction

According to Netter and Wasserman (1974) the regression analysis is a statistical tool used by researchers when investigating relationships of a behavioural and economic nature. In other words it can be used to examine data and draw conclusions about the functional relationships that exist between or among dependent and independent variables, whereby such relationships are expressed in a form of mathematical functions that demonstrate how the variables are interrelated (Ostle and Mensing, 1979)

In this section we justify the use of regression by comparing the general purposes of regression analysis with the specific objectives of our research problem. Since these two issues are strongly related we will apply regression analysis to our available sample size data. A parametric regression model will also help us to control the factors that affect energy consumption, and to do cost analysis and economic modelling analysis.

Some of the purposes of this research are .

1. To select the factors which influence the levels of electric power consumption and fuel consumption in the active cement kilns in Fuhais and Rashadiya cement plants - Jordan Cement Factories Company (JCF). Based on the previous analysis and the factories practical experience that was discussed previously in chapter 6, the researcher has in mind the following factors that may influence energy consumption. These factors are average number of stoppages (AvNO), average duration of stoppages (AvHOURS), production rate (PRORATE), availability (AVL), Alumina Ratio (Aratio), Silica Ratio (Sratio), and Lime Saturation Factor

(LimeSF). These will be called the independent variables. On the other hand electricity and fuel consumption will be called the dependent variables. The inclusion of these variables is based on the presence of logical and/or physical relationships, or may be suspected from experience or literature and preliminary data.

2. To analyse the independent variables and discover their relative importance in affecting the dependent variables.
3. To discover the type of influence of each of the independent variables on the dependent variables, e.g. does the increase in an independent variable produce an increase or a decrease in the dependent variable? For example the researcher expects that both electricity and fuel consumption increase with the increase of average number of stoppages, average duration of stoppages, Silica Ratio, and Lime Saturation Factor, but they decrease with the increase of production rate, availability, and Alumina Ratio. The reasons for these expectations will be discussed in Section 7.3.2.
4. To find a functional relationship which may be used to explain the variability in the dependent variables by the independent ones.
5. To build a procedural method, which may help in planning the expected budget needed for the expenses of the dependent variables, based on the levels of the independent variables. That is to find a regression model that expresses each of electricity and fuel consumption as a function of the influential independent variables that may affect the cost of the energy and the cement produced.

Montgomery (1982) and Glantz and Slinker (1990) outline the following purposes for running a regression analysis.

1. **Description:** The analyst is seeking to find an equation that describes or summarizes the relationships in a set of data.
2. **Coefficient Estimation:** The analyst may have a theoretical relationship in mind, and the regression analysis will confirm this theory. Most likely, there is specific interest in the magnitudes and signs of the coefficients.
3. **Prediction:** The prime concern here is to predict some dependent variable. Prediction may be very crucial in planning, monitoring, or evaluating some process or system.
4. **Variable Selection or Screening:** In this case, a search is conducted for those independent variables that explain a significant amount of the variation in the dependent variable. In most applications, this is not a one-time process but a continual model-building process.
5. **Partial Influence:** In laboratory experiments, one can generally control all the variables. One changes one variable, measures another, and then analyses the data with one of the standard statistical tests. But in some kinds of experiments, and many observational studies, one needs to analyse the interaction of several variables. In some situations, the goal may really be to examine several variables at once in order to find out which X variable has the largest influence on Y. In other situations, one really only cares about one of the independent variables, but the analysis needs to adjust for differences in other variables. For example, one might ask: Does Y vary with a specified X variable, after correcting for other X variables?

Comparing the general purposes of regression analysis with the specific objectives of our research problem, it can be observed that regression is practically useful in our problem. Statistical regression models are therefore used to analyse the available data. Once these models are developed, whether linear or non-linear, simple or multivariate, they may be used to predict the value of the dependent variables based on the values of the independent variables. Several types of regression procedures such as multiple linear and non-linear regression, stepwise regression, ridge regression, and robust regression analysis will be used as statistical tools to determine the presence and significance of potential relationships among the variables of the studied data.

Since the treatment depends on regression analysis, the rest of this section provides a short review of regression analysis and its terminology.

7.2.2 General Methodology for Model Construction

The general methodology of the process of constructing a model essentially requires the following steps (Draper and Smith (1981, pp412-422)):

1. State the objectives, hypotheses, and limitations of the study. Based on these select the type of the model and the type of the analysis that is most suitable for them.
2. State and define the dependent variable(s).
3. Start with a list of possible independent variables that may affect the dependent variable. Accurately define all variables included in the study and state the units of measurement of these variables.
4. Design an experiment or use historical records to collect and prepare data to be used to demonstrate the validity of the assumptions of the presence of a relationship between dependent and independent variables.
5. Explore the obtained data to check its validity, i.e. does the obtained data satisfy the assumptions of the model used? This requires the specification of the model and its assumptions.
6. Use the data to estimate the parameters of the assumed model.
7. Use statistical procedures to check the adequacy of the estimated model. Based on this, the analyst can determine whether additional variables should be

considered, some of the variables under consideration should be dropped, or some transformations of the data are needed. Sometimes we may need to fit other models or use other procedures. The process of the fine-tuning model lasts until a final valid and adequate model is generated.

7.2.3 Multiple Regression

In multiple linear regression, a dependent variable (Y) is related to a set of independent variables using the following linear model $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon$

Where Y is a linear function of k independent variables $X_1, ..., X_k$, and ε is an error term. The construction of a multiple regression model essentially requires the estimation of the parameters $\alpha, \beta_1, \beta_2, ..., and \quad \beta_k$ to get the estimated regression equation $\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + ... + \hat{\beta}_k X_k$ and test for the significance of the associated independent variables. The $\hat{\beta}_j$ is called net regression coefficient. It measures the effect of X_j on Y by netting out (controlling for) all other independent variables. This means that $\hat{\beta}_j$ measures the rate of change of Y with respect to X_j in the absence of the effects of all other variables. However, if the independent variables themselves are related, then $\hat{\beta}_j$ actually cannot completely net out the effect of other variables.

7.2.3.1 Assumptions of Multiple Regression

The following assumptions must be considered when using multiple regression analysis (Johnson and Wichern 1988, p. 274).

1. Linearity

Multiple regression models the linear relationship between Y and the X's.

2. Normality of errors

Multiple regression assumes that the error values from the prediction of the model are random and follow a Gaussian distribution with mean zero, i.e. the random errors ε are independent and normally distributed with zero means.

3. Homoscedasticity of error

Multiple regression assumes that scatter of data from the predictions of the model has the same standard deviation for all values of the independent variables. That is the variance of the ε is constant for all values of the X's. The assumption that the standard deviation is the same everywhere is termed homoscedasticity.

4. Independence

The data points are assumed to be uncorrelated with one another.

5. Non-collinearity

One of the main problems in multiple regression is multicollinearity (Willan and Watts 1978). Collinearity, or multicollinearity, is the existence of near-linear relationships among the set of independent variables. The presence of multicollinearity causes all kinds of problems with regression analysis. For example, in multicollinearity, the standard errors of the estimated coefficients in the model tend to have large values (Press 1972, p. 272), which implies that t-tests of the estimated regression coefficients will lead to non-significant value. The standard error of the estimated coefficients increases as the correlation among the explanatory variable increases, giving less precise estimates of the true coefficients in the model. So, one should be careful for the multicollinearity. We will discuss three procedures for detecting multicollinearity in section 7.2.3.2.

6. <u>No Outliers</u>

Outliers are extreme cases on one variable, or a combination of variables, which have a strong influence on the calculation of statistics. Anscombe (1973) gave four sets of data with the same simple regression line and very high R-squared value, even though some of the data sets do not exhibit a linear relationship due to outliers or other reasons.

For multiple regression, it is assumed that the data is free of outliers. Sometimes the data may have outliers that may cause nonconstant variance of error terms, nonnormality of error terms, or other problems with the regression model. In such a case other regression procedures such as robust regression should be used (Rousseeuw and Leroy (1987)).

7. <u>Experimental Errors</u>

The linear regression model assumes that all the X values are exactly correct, and that experimental error or variability only affects the Y values. This is rarely the case, but it is sufficient to assume that any imprecision in measuring X is very small compared to the variability in Y.

7.2.3.2 Model Adequacy

Once the regression output is displayed, one will be tempted to go directly to the probability of the F-test from the regression analysis of variance table to see if that F-value is statistically significant at the required level of significance. One should check for linearity, normality, constant variance, independence, outliers, multicollinearity, and predictability, Belsley, Kuh, and Welsch (1980), i.e. the assumptions are valid, otherwise the analysis is invalid.

1. <u>Check for Linearity</u>

Scatter plots of dependent variable versus each of the independent variables are the most easily procedure to demonstrate linearity. Nonlinear patterns can show up in residual plots. Transformations of the variables may sometimes be used to obtain linearity.

If linearity does not exist, it may be possible to transform either the dependent or independent variable or both to get linearity. It is also possible to use a nonlinear model.

2. Check for Normality

The normality of the residuals should be visually evaluated by graphical displays, such as box plots, normal probability plots, and density plots.

If all of the residuals fall within the confidence bands for the normal probability plot, the normality assumption is likely to be met. One or two residuals outside the confidence bands may be an indicator of outliers, but not nonnormality. The more formal option is to use normality tests based on skewness, kurtosis, and omnibus test based on both skewness and kurtosis, (D'Agostino (1990)).

If the residuals are not normally distributed, a suitable transformation may convert it to normality. Box and Cox (1964) dealt with the power transformation to normality, i.e. if the distribution of a random variable X is not normally distributed then for some suitable power c of X will be approximately normally distributed. Note that c is a real number. In most applications it was found that c may take the values 2, 3, −1, ½, 1/3, or ¼ (see e.g. Bhattacharyya and Johnson (1977, pp. 223-226). Moreover, the same reference recommends using the logarithmic transformation. These transformations are useful if the histogram of the data is not symmetric or if it has some problems at either tail of the distribution.

One of the problems that lead to nonnormality is the existence of a relationship between the mean and the variance of the random variable X. This problem arises if the constant variance assumption is violated. This problem can be resolved by using what is called a variance stabilizing transformation that converts the variance of X to be a constant, which is not related to the mean of X. For more details see e.g. (Mukhopadhyay (1996), pp 686-694).

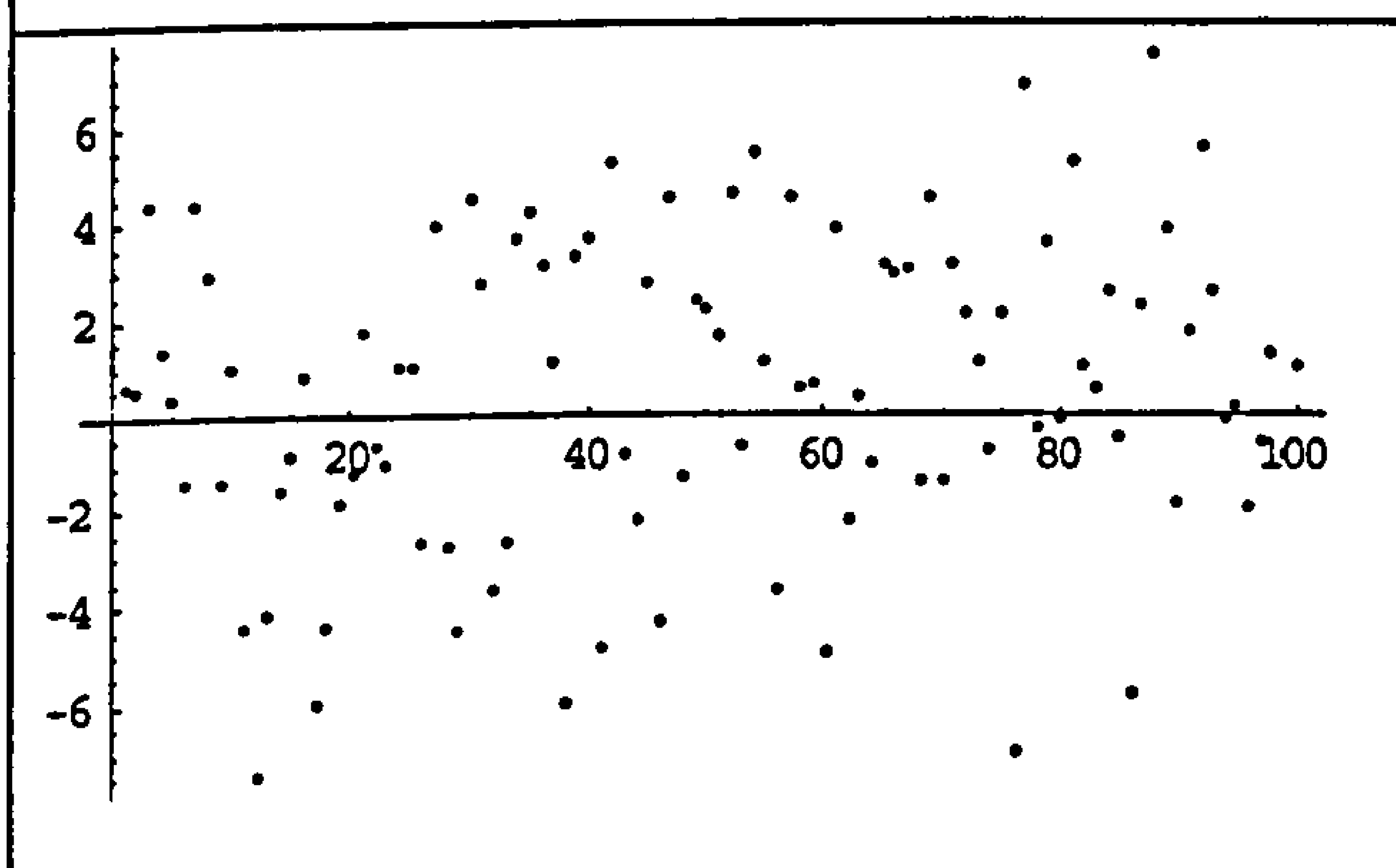
In some cases, if normality does not exist and it is due to existence of outliers, one may remove outliers if their number is small and there is a good reason to believe that they are real outliers, i.e. different in practical, real-life, terms.

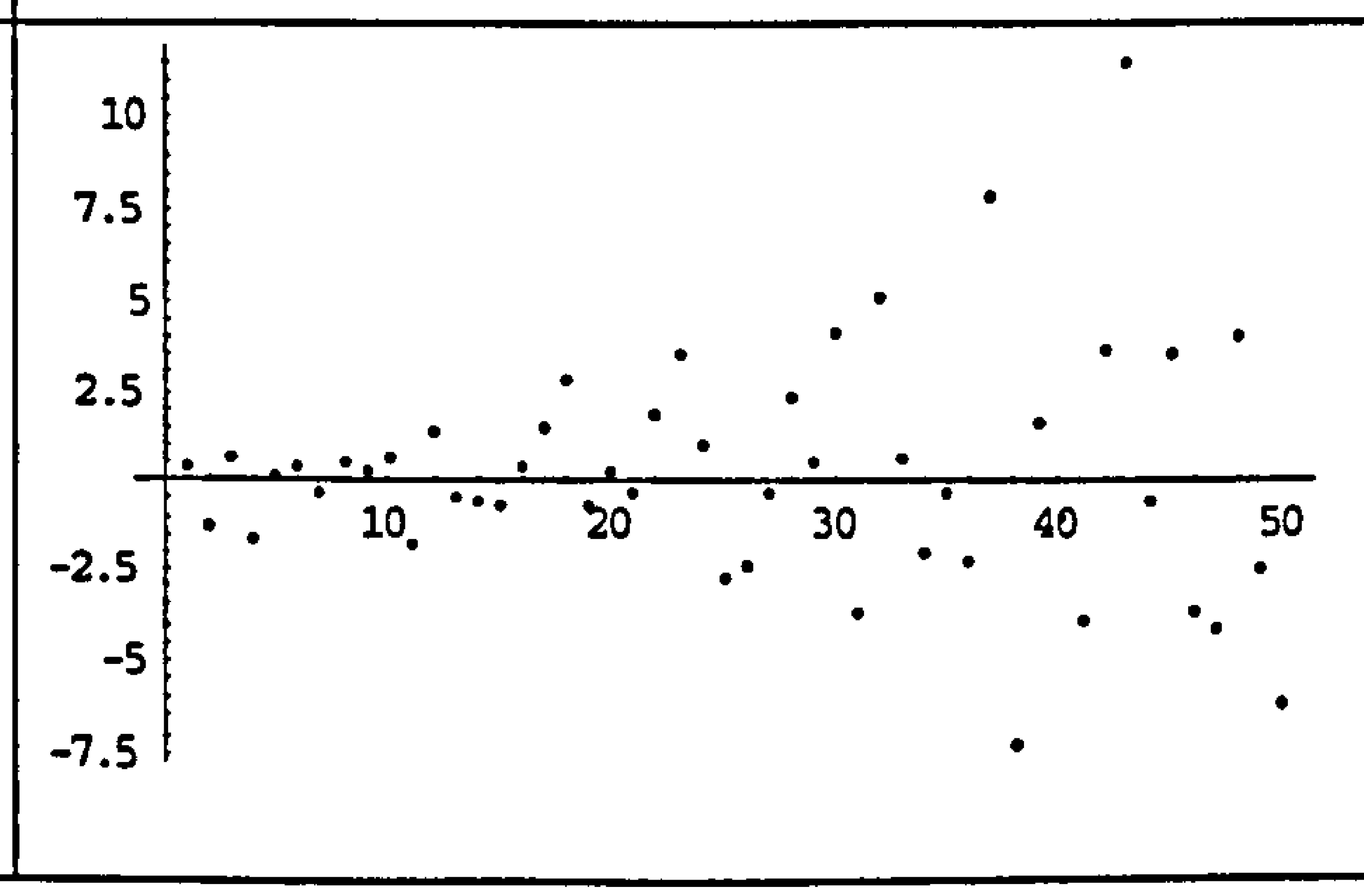
3. Check for Nonconstant Variance (Hetroscedasticity)

Residual plots of errors versus the predicted values or versus each of the independent variables can detect nonconstant variance (hetroscedasticity).

If these residual plots show a rectangular shape, we can assume constant variance. The assumption is violated if the scatter plot spread increases or decreases, as one of the X variables gets larger. The following two plots illustrate these remarks.

Constant Variance	Non Constant Variance





If nonconstant variance does exist, one may use a transform of the dependent variable or use weighted regression.

4. Check for Independence or Serial Correlation

The independence assumption can be violated in two ways: model misspecification or time-sequenced data.

Model misspecification. If an important independent variable is omitted or if an incorrect functional form is used, the residuals may not be independent. The solution to this dilemma is to find the proper functional form or to include significantly influential independent variables.

A. **Time-sequenced data.** Whenever regression analysis is performed on data taken over time, the residuals are often correlated. This correlation among residuals is called serial correlation or autocorrelation. Serial correlation patterns could be identified informally, with the residual plots versus time. A better analytical way would be to compute the serial or autocorrelation coefficient for different time lags and to compare it to a critical value.

If none of the serial correlations are greater than the critical value independence may be assumed.

If independence does not exist, one may use a first difference model. More complicated choices require time series models. Moreover in such a case time should be included as a factor.

5. Check for Outliers

Outliers can completely distort both descriptive statistics and statistical inference. Outliers may contaminate measures of skewness and kurtosis as well as confidence limits and statistical tests. Outliers form a real problem in multiple regression. So, there is a need to detect and handle them. This problem is dealt with by Hawkins et al. (1984) and Ferguson (1961). If the outliers are one-time occurrences (real outliers), they may be removed, but if they represent a certain segment of the population, one should decide between biasing his results (by removing them) or leaving them in and invalidating the normality assumption. In such a case, one may examine changes in the model with and without outliers.

The visualization of univariate outliers can be done in three ways: with the stem-and-leaf plot, with the box plot, and with the normal probability plot. In each of these informal methods, the outlier is far located from the rest of the data.

The box plot shows three main features about a variable: its centre, its spread, and its outliers. A box plot is made up of a box (a rectangle) with various lines and points added to it. The top and bottom of the box are the 25th and 75th percentiles. The length of the box is thus the interquartile range (IQR). That is, the box represents the middle 50% of the data. Values that are under three IQRs from the 25th or 75th percentiles are called mild outliers. Those outside three IQRs are called severe outliers.

Outliers are values that do not follow the pattern of body of the data. They show up as extreme points at either end of a probability plot.

The normal probability plot evaluates the potentiality of an outlier assuming the data are normally distributed. If the variable is not normally distributed, these plots may indicate many outliers. So, one must be careful about checking what distributional assumptions are behind the outliers he may be looking for.

The NCSS2000 Package tests each observation to determine if it is an outlier. The program uses a t-based test statistic, which is given, in terms of the distance of each observation from the sample mean. In regression analysis, this package provides diagnostics section to pin point outliers based on the Cook's D statistic. It also provides a program to implement robust regression procedure, which decides on how to downweight the existing outliers.

6. Check for Multicollinearity

The following procedures may be used to detect collinearity.

A. Pairwise scatter plots of independent variables and the correlation matrix of independent variables may indicate collinearity. Since multicollinearity means the existence of almost perfect linear relationship between independent variables, one should look for near-perfect relationships. Strong pairwise correlation may give some insight as to the variables causing the collinearity. According to Younger (1979) the correlation coefficient with value greater than 0.75 is considered as a strong correlation. This value of the correlation is justified by the discussion given in 7.2.3.3. So, if the absolute value of the correlation coefficient between two independent variables is greater than 0.75 then the two variables are strongly inter-related and should not be used together in the same regression equation.

B. Collinearity may be checked by answering the following question. How well each independent (X) variable can be predicted from the other X variables (ignoring the Y variable)? There are three measures to express the answer to this question, namely

i. R^2 with other X variables, which is the fraction of all variance in one, X variable that can be predicted from the other X variables.

ii. Variance Inflation Factor (VIF) which is defined as $VIF=1/(1-R^2)$.

iii. Tolerance, which is the fraction of the total variance in one, X variable that is not predicted by the other X variables. It is defined as $1 - R^2$

It is clear that the three terms are functionally related and they measure the degree of Multicollinearity. So, one may depend on any of them to check collinearity.

If the X variables contain no redundant information, one expects VIF to equal one. If the X variables are collinear (contain redundant information), then VIF will be greater than one. The large variance inflation factors (VIF's) flag collinear variable.

If any variable has a variance inflation factor greater than 10, collinearity could be a problem (see e.g. the help of the NCSS2000 Package).

C. Eigenvalues of the correlation matrix of the independent variables are useful in detecting collinearity (Younger 1972, p. 340). An eigenvalue of zero or close to zero indicates that an exact linear dependence exists i.e. multicollinearity is a severe problem.

No multicollinearity if eigenvalues are approximately the same size and are not close to zero. If some are much larger than others this is indicative of the related variables being exists together in the sense that they are highly correlated and hence multicollinearity is a problem.

Instead of looking at the numerical size of the eigenvalue, one may use the condition numbers. Condition number of an independent variable is the ratio of the largest eigenvalue to the corresponding eigenvalue of that variable, so it is a sort of ratio of variances. Large condition numbers indicate collinearity. Condition numbers greater than 1000 indicate severe collinearity. Condition numbers between 100 and 1000 imply moderate to strong collinearity (see e.g. the help of the NCSS2000 Package).

If multicollinearity does exist in the model, it could be due to an outlier or due to strong interdependencies between independent variables. In some cases, removing one or more variables from the model will reduce multicollinearity to an acceptable level. In other cases, one may be able to reduce multicollinearity by collecting data over a wider range of experimental conditions.

7. Check for Predictability

The estimated regression model may be used to predict the value of the dependent variable for some specified values of the independent variables. To have reasonable predicted value with some accuracy, the estimated model should have a reasonable ability to do this job. This is what is called the predictability of the model. This is measured by what is called the Press R-squared. The value of this R-squared should be large to have reasonable predictability.

Suppose there are k parameters to be estimated in the regression model based on n observations. For each i=1,2,...,n,

A. Delete the ith observation on the dependent variable and independent variables.

B. Fit regression model to the remaining n-1 data points. (This is jackknife method).

C. Use the fitted model to predict the value of the omitted y, and obtain the predicted error $y_i - \hat{y}_i$

D. Repeat the above steps for i=1,...,n and calculate the predicted sum of squares (PRESS), i.e. the sum of squares of the predicted errors (Draper and Smith (1981, p. 342) and Younger (1979, p 483).

This PRESS value can be used to compute an R^2 -like statistic, called Press R-Squared ,which reflects the prediction ability of the model. This is a good way to validate the prediction of a regression model without selecting another sample or splitting the data provided that the sample is a representative one.

7.2.3.3 Model Testing

The first question one faces after fitting a regression model is that " Is the fitted model acceptable?" If the model is not acceptable one should search for extra independent variables to be included in the model or/and he should search for another form of the model. So, before testing each independent variable separately, one must test for the significance of the regression model as a whole (Younger 1972, p. 340). i.e., it is required to test the hypothesis:

$H_0: \beta_i = 0$ for all i=1,.., k vs.

H_1: some $\beta_i \neq 0$ (at least one coefficient is not zero).

A one sided upper tail F-test through the ANOVA table can be used to test the above hypothesis. The F-statistic has (k, n- (k+1)) degrees of freedom under the null hypothesis. One compares the calculated F value to the (1 - α) th percentile of the F-distribution with the above stated degrees of freedom. This percentile is denoted by F (k, n- (k+1), α) where α is the level of significance and it is called the critical value. One rejects H_0 if the calculated F value is greater than this critical value. The overall p-value corresponding to this calculated F might be used to obtain same result by comparing it to α. If the p-value is less than α, the null hypothesis is rejected; otherwise it cannot be rejected. If one can reject H_0, he is at least $100(1-\alpha)\%$ sure that at least one of the *k* independent variables is significant, and therefore can proceed to look for it. Moreover, the model might have some use. If H_0 cannot be rejected, then none of the independent variables is strong enough to be concerned with.

It is interesting to note that the level of significance α may be usually taken as 0.01, 0.05, or 0.1. However, we will assume that $\alpha = 0.05$ since this is the mostly used value of α. If one assumes that $\alpha = 0.05$, then a p-value that is less than 0.05 indicates that the model is statistically significant and it may be of some use. If a p-value is greater than 0.05, one concludes that the model is not statistically significant and it has no use. However , the same procedure applies for any value of α since the statistical package provides the p-value in all procedures.

The coefficient of determination R^2 represents the proportion of the variability in the dependent variable that can be explained by its relationship to the independent variables. The larger the value of R^2 the better the model is provided that the underlying assumptions are satisfied. The question, which arises here, is that: How large must R^2 be? The answer to this question is given by Younger (1979, PP. 236-244) for the case of simple linear regression, which can be extended to the multiple regression in the same way. It is known that the absolute value of the correlation coefficient ranges between zero and one. She suggested partitioning this range into four subintervals with equal lengths. These intervals are (0, 0.25), (0.25, 0.50), (0.50, 0.75), and (0.75,1.00). Then she said that if the absolute value of the correlation coefficient belongs to the first interval then it is moderately weak, if it belongs to the second interval then it is moderate, if it belongs to the third interval then it is moderately strong, and finally if it belongs to the fourth interval then it is strong. Based on this type of classification, one may suggest that:

a) If SQRT (R^2) belongs the interval (0, ¼) then R^2 is moderately weak, i.e. if $0 < R^2 < 1/16$, i.e. $0 < R^2 < 0.063$

b) If SQRT (R^2) belongs to the interval (¼, ½) then R^2 is moderate, i.e. if $1/16 < R^2 < 1/4$, i.e. $0.063 < R^2 < 0.25$

c) If SQRT (R^2) belongs to the interval (½, ¾) then R^2 is moderately strong, i.e. if $1/4 < R^2 < 9/16$, i.e. $0.25 < R^2 < 0.567$

d) If SQRT (R^2) belongs to the interval (¾, 1) then R^2 is strong, i.e. if $9/16 < R^2 < 1$, i.e. $0.567 < R^2 < 1$

7.2.3.4 Which variable(s) make a significant contribution?

Assuming that the underlying model assumptions are satisfied, if the overall p-value corresponding to the F-test is high e.g. greater than $\alpha = 0.05$, one can conclude that the multiple regression model is not significant and concludes that the model does not fit the data. In this case, there is not much point in looking at the results for individual variables. If the overall p-value is low e.g. less than $\alpha = 0.05$, one probably will next want to find out which independent variables in the model are significant and which are not significant.

To check which independent variables are significant, it is required to test the set of k hypotheses:

$H_{01} : \beta_1 = 0, \quad vs. \quad H_{11} : \beta_1 \neq 0,$

$H_{02} : \beta_2 = 0, \quad vs. \quad H_{12} : \beta_2 \neq 0,$

....

....

....

and

$H_{0k} : \beta_k = 0, \quad vs. \quad H_{1k} : \beta_k \neq 0,$

A two-sided t-test for each of these individual hypotheses can be used.

For each independent variable in the model, one should report a p-value that answers this question. After accounting for all the other independent variables, does adding this variable to the model significantly improve the ability of the model to account for the data? If the p-value is small, the variable contributes in a statistically significant manner. If the p-value is large, then the contribution of the variable is no greater than one would expect to see by chance alone.

If one assumes the standard threshold (alpha) value of 0.05, then a p-value that is less than 0.05 indicates that the variable made a statistically significant contribution to the fit. If a p-value is greater than 0.05, one concludes that the influence of that variable is not statistically significant and it should be omitted from the model.

7.2.4 All Possible Regressions

There are several criteria for model selection, but we give here the most popular three criteria to assess each of the fitted equations (Draper and Smith (1981, p. 296). These are:

a) The value of R-square. The larger the value of R-square the better the model is.

b) The residual mean square. The smaller the residual mean square, the better the model is.

c) The Cp statistic. The smaller the value of Cp the better the model is.

All Possible Regressions procedure applies these three criteria and fits every possible regression equation. Based on the obtained values of the above statistics one can select the best model, in a statistical sense, among the fitted ones.

The Cp- criterion developed by Mallows (1964, 1966, 1973) measures the total squared error (TSE) of a regression model with p parameters. The TSE consists of a bias component and a random error component. The bias component represents the difference in the predicted Y values obtained from the fitted regression model and the "true" regression model. The random error component represents the variability around the fitted regression line.

Because Cp is a measure of total squared error, (Berenson et. all, 1983. P. 371), one should attempt to find the set of independent variables that leads to the smallest Cp-value while at the same time minimizing the bias component (Daniel and Wood (1980)).

The Cp statistic is defined as Cp=(SSE(p)/MSE(T))-(n-2p), where p=number of parameters included in a particular model, SSE(p) = error sum of squares for a regression model with p parameters, T=total number of parameters to be considered for inclusion in the regression model, MSE(T)=mean square error (variance) of a regression model containing all T parameters, and n=sample size.

Cp can be expressed in terms of R2(p), where R2(p) is R^2 for the model with p parameters. Moreover, the Akaike information criterion (AIC) is the likelihood version of Cp. It can be shown that AIC=(Cp+n) SSE. The model with the smallest AIC is the best among the given models (Sakamoto 1986, p. 172).

7.2.5 Stepwise Regression

The stepwise regression procedure in NCSS2000 Package applies each of the following four searching methods. These methods among other methods were treated by Bendel and Afifi (1977), Cochran (1938), Goldberger (1961), Hocking (1972).

1- Forward, or Step-Up, Selection

This method is often used to provide an initial screening of the candidate variables to obtain the best independent variables and then apply the all-possible regression algorithm to the variables in this subset.

The forward selection begins with no candidate variables in the model. It selects the variable that has the highest R-Squared. At each step, it adds the candidate variable that increases R-

Squared the most. It stops adding variables when none of the remaining variables are significant. Note that once a variable enters the model, it cannot be deleted.

2- Backward, or Step-Down, Selection

This method is less popular because it begins with a model in which all candidate variables have been included. However, because it works its way down instead of up, one is always retaining a large value of R-Squared.

The backward selection model starts with all candidate variables in the model. At each step, the variable that is the least significant is removed. This process continues until no nonsignificant variables remain.

3- Stepwise Selection

Stepwise regression is a combination of the forward and backward selection techniques. Stepwise regression is a modification of the forward selection so that after each step in which a variable was added, all candidate variables in the model are checked to see if their significance has been reduced below the specified tolerance level. If a nonsignificant variable is found, it is removed from the model.

Stepwise regression requires two significance levels: one for adding variables and one for removing variables. The cutoff probability for adding variables should be less than the cutoff probability for removing variables so that the procedure does not get into an infinite loop.

4- Min MSE

This procedure is similar to the Stepwise Selection search procedure. However, instead of using probabilities to add and remove, one should specify a minimum change in the root mean square error. At each step, the variable whose status change (in or out of the model) will decrease the mean square error the most is selected and its status is reversed. If it is currently in the model, it is removed. If it is not in the model, it is added. This process continues until no variable can be found that will cause a change larger than the specified minimum change amount.

7.2.6 Robust Regression

Regular multiple regression is the best approach when all of its assumptions are valid. When some of these assumptions are invalid; least squares regression can perform poorly. Thorough residual analysis can point to these assumption breakdowns and can in many circumstances allow us to correct these problems.

Robust regression provides an alternative to least squares regression that works with less restrictive assumptions (Draper and Smith (1981, p. 342), Atkinson and Riani (2000)). Robust regression is a better approach than the ordinary least square approach if the normality assumption breaks down and/or the data has some outliers. Outliers violate the assumption of normally distributed residuals in least squares regression. They tend to pull the least squares fit too much in their direction by receiving much more "weight" than they deserve (Abraham and Box (1978)). The assumptions of multiple regression apply for robust regression, but the normality of residuals and no outliers assumptions are relaxed since this is what the method was designed for. Because of the problems due to the existence of outliers, these outliers may be difficult to identify. When only one or two independent variables are used, these outlying points may be visually detected in various scatter plots. However, the complexity added by additional independent variables might hide the outliers from view in these scatter plots. Robust regression down weights the influence of outliers, i.e. robust regression applies a procedure to test for outliers and it gives a weight equals one for non-outliers and smaller weights for the outliers. These weights differ with the severity of the outlier. The weight may be zero (i.e. the outlier will be omitted if is too severe). This is done only if there are one or two outliers. This makes their residuals larger and easier to spot. Robust regression techniques use iterative procedures that seek to identify these outliers and minimize their impact on the coefficient estimates. The amount of weighting assigned to each observation in robust regression is controlled by a special curve called an influence function (Rousseeuw and Leory (1987), and O'Leary (1990)).

Outliers not only influence the estimation of the regression coefficients, they can also have an even larger effect on standard errors, t-tests, F-tests, R^2, and other regression statistics. Ordinary least squares analysis does not perform well when outliers occur. It is not resistant to changes in one or two observations. A robust estimate is one that is resistant to even drastic changes in one or two observations.

Several families of robust estimators have been developed in the literature. The robust methods fall into the family of M-estimators. This type of maximum-likelihood estimator minimizes the sum of a function of the residuals. M-estimation is usually approximated through the use of iteratively reweighed least squares. This method proceeds as follows (Hamilton (1991) and Montgomery and Peck (1992)

1. A set of starting values for the regression coefficients is obtained. The starting values are the ordinary least squares (OLS) coefficient estimates. Thus, this step amounts to estimating the regression coefficients by standard multiple regression.
2. Using the regression coefficients from step 1, a set of residuals is calculated-one for each observation. From these residuals, a set of weights is developed. These weights range from zero to one. Observations with large residuals receive small weights. The relationship between the residuals and the weights is specified through the influence function. The NCSS200 package uses three influence functions, namely, Andrew's sine, Tukey's biweight, and least absolute deviation functions.
3. Using the weights calculated in step 2, weighted least squares method is used to estimate a new set of regression coefficients.

4. Using the regression coefficients from step 3, calculate a new set of residuals. Go to step 3.

5. Continue iterating steps 3 and 4 until there is little or no change from one iteration to the next. Because of the masking nature of outliers, it is a good idea to run through at least five iterations to allow the outliers to be found.

7.2.7 Ridge Regression and Principal Components Regression

Ridge Regression procedure, first suggested by Hoerl (1962). This procedure is intended to overcome "ill-conditioned" situations in which the estimated parameters may be unstable, i.e. the estimates may have the wrong sign or may be much larger than the practical considerations would deem reasonable (Mullet (1976) and Draper and Smith (1981, p. 313)).

Ridge Regression is a technique for analysing multiple regression data that suffers from Multicollinearity. When multicollinearity occurs, least squares estimates are unbiased, but their variances are large so they may be far from the true value. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors. It is hoped that the net effect will be to give more reliable estimates (Hoerl and Kennard (1970)). Another biased regression technique is principal components regression (Hawkins (1973)). Ridge regression is the more popular of the two methods.

7.2.8 Nonlinear Regression

In regression analysis, one searches for a mathematical model to fit the available data. From a mathematical point of view, there are different methods that allow us to approximate a given function that satisfies some conditions with a suitable function. Some of these methods are the Taylor series expansion of a function. The Taylor series method searches for a polynomial of some degree that best fits the data. The simplest approximation (a first approximation) is a linear relationship, which leads to linear regression. If the linear model does not provide an acceptable approximation, one moves to quadratic model. If the quadratic model does not provide an acceptable approximation, one moves to cubic model, and so on till one gets an acceptable model. This procedure may lead to a nonlinear regression model.

In some cases, one may use other types of (transcendental) functions such as exponential, logarithmic, or trigonometric functions. The pairwise scatter plots of the dependent variable against the independent variables may help in specifying the type of needed function(s).

The regression model may also be classified as additive or multiplicative (nonadditive) model. The multiplicative model can be transformed to an additive model by taking the logarithm of both sides. In other words, the multiplicative model is an additive model in the logarithms of the independent variables rather than in the variables themselves. That is to say, one may use a multiplicative model if scatter plots of the logarithm of the dependent variable is linearly related to the logarithm of each of the independent variables.

Regression models may or may not contain interaction terms. A model is said to have interaction terms if it involves products of two or more independent variables. For example $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$ is a regression model with interaction term $X_1 X_2$.

Multiple linear regression deals with models where the parameters are linear. Nonlinear regression models are those that are not linear in the independent or/and dependent variables, (Drapert and Smith (1981,pp 458-517)). Polynomial regression with or without interaction terms is a sort of nonlinear regression. Moreover, multiplicative models are another sort of nonlinear regression models.

7.3 Model Construction

7.3.1 Variable Selection

7.3.1.1 Introduction

Preliminary detailed empirical analysis of the main factors affecting energy consumption was carried out in chapter six. One purpose of this research is to analyse the factors affecting the levels of electric power consumption and fuel consumption in the active cement kilns in Fuhais and Rashadiya cement plants - Jordan Cement Factories Company. The major focus is on empirical investigation of the relationships between power consumption and fuel consumption as dependent variables, on one hand, and a set of other predictor variables, on the other hand. The latter variables include average number of stoppages, average duration of stoppages, production rate, availability, Alumina Ratio (Aratio), Silica Ratio (Sratio), and Lime Saturation Factor (Lime SF) as independent variables.

Proper statistical tools will be used to examine data and draw conclusions about the functional relationships that exist between the selected dependent and independent variables, whereby such relationships are expressed in a form of mathematical functions that demonstrate how the variables are interrelated.

Statistical regression models will be used to construct empirical relationships among variables that exhibit a form of interaction or relation. Once these models are developed, whether linear or non-linear, simple or multivariate, they are used to predict the value of one variable based on the values of other variables. Predictor variables are called control variables, causes, or independent, while predicted variables usually represent the effects of the causes, and called the dependent variable.

7.3.1.2 Variable Selection Background

The selection of the variables is based on logical and or physical relationship, and from practical experience and from observing and studying the preliminary data. The nature of the cause and effect relation between the dependent variables (electrical and fuel consumption) and the independent variables can be summarized as follows:

When the production process is analyzed and the energy consumption centers are determined, one can predict the factors affecting energy consumption. The burning process in the kiln was analyzed for thermal and electrical consumption. The kiln uses the fuel to produce the needed heat to complete the chemical reaction between the raw materials to produce new material called "clinker". The theoretical thermal heat needed to complete the reaction to be approximately 400 kcal/kg equivalent to 42 kg fuel /tonne clinker while in actual term the consumption is around 80 – 85 kg fuel/tonne cincker. One can study the production line and make heat balance to estimate and locate the heat losses. From this analysis we can see that the thermal consumption is higher than the expected one due to different reasons, among the most important of it are the stoppages (number and duration), production rate, availability and quality control factors. When the kiln stops, the production stops simultaneously leaving all the heat energy inside the kiln to be lost without production i.e.: leading to losses in heat energy inside the kiln. Also when the kiln is restarted, there is a lot amount of fuel is consumed in order to heat up the kiln and reaches the optimal temperature suitable for the chemical reaction. In such a case there is energy consumed (electricity and fuel) with zero production because the raw materials can't be fed to the kiln before it is heated up. As a result of stoppages we have unstable operational condition where we lose production and the average production rate will be decreased, joined by the increase of the energy consumption and drop in the availability of the kiln.

Analyzing the electrical consumption for equipment, one can calculate the needed electrical energy to operate this equipment. Electricity is needed to rotate the kiln, rotate the mills (electrical motors) or operate the auxiliary parts in the production line: pumps, conveyors.... etc.

From the theoretical background it is known that the starting up of any electrical machine will consume a higher electrical energy than the normal operating mode. This indicates that when number of stoppages increases, starts up of machine increase with higher electricity consumption. And when duration of stoppages increase there is a need to heat up the kiln for a long period of time without production, which means consuming more energy (electricity and fuel) at zero loads without production. Also when the kiln or mill stops for a short time, some auxiliary equipment continues operating and consuming electrical energy with zero production. In general, in practical terms increasing number of stoppages, duration of stoppages will decrease the production rate and decrease the availability and increase the energy consumption.

The experience shows that the ball mills are constant power machines, i.e.: the energy consumption is constant whatever the production amount. Increasing the production rate will increase the amount produced while the electrical consumption is fixed which decreases the electricity consumed per tonne of product. Another example of the theoretical justification for the selection of the control variables is the operation of the auxiliary equipment (pumps fans, convey or belts) while the main equipment (kiln, mill) is stopped. The engineers find it is useful to stop some of the auxiliary equipment, as there is no need for them while the main equipment is stopped. If it is difficult to stop auxiliary equipment they can be operated at lowest speed or capacity level with the minimum electrical energy consumption level.

From practical plant operational experience we found that when the stoppages are programmed the production unit energy consumption is less than during emergency stoppages, because emergency stoppages are caused by problems that may increase energy consumption. For example, in case of a blockage, or a failure in a bearing or a gear there will be high current, which results in the consumption of more electricity before the emergency stops, occurs. Also in the case of leakage of gases in the system, there will be an increase in the thermal energy consumption due to the heat losses accompanied the hot gas leakage. The previous issues lead us to select availability as an independent variable, which covers the important effects of emergency stoppages on the energy consumption.

Also the preliminary data sustain and give information on the selection of the control variables. Data was collected and analyzed to choose the best control variables that affect the energy consumption. Not only the selected variables were chosen in the preliminary stage, there were some other variables analyzed and the result of this preliminary study was the selection of these seven variables.

7.3.1.3 Classification of Stoppages

From the above stated background of the variable selection and as it is revealed in the analysis of factors affecting energy consumption in chapter six, it is clear that the stoppages: number, duration and nature of stoppages affect almost all the variables related to the energy consumption, so it is quite appropriate to discuss and analyze the types and the causes of stoppages in JCF.

Stoppages of the production line in JCF essentially include two types of stoppages: breakdown stoppages and programmed stoppages. The raw mill, the kiln and the cement mill can be considered as separate production units (Although sometimes the raw mill and the kiln are considered as one production line.)

A) <u>Breakdowns Stoppages:</u>

These stoppages are unexpected stoppages since they occur because of a failure in the production line. These failures indicate that the production process is not functioning in an appropriate manner. At some time the problem reaches a point where it becomes noticeable (abnormal sound, abnormal operational parameters readings/alarm signalling in the control room), which forces the operator to stop the failed equipment, or the automatic control will initiate an automatic stopping order for this equipment. If the equipment subject to emergency breakdown is essential to the production line the automatic control system will shutdown the whole production line automatically (most of the essential equipment in the production line are interconnected with each others through automatic control in accordance to a programmed operational sequence).

The types of breakdowns in JCF are classified into three types as follows:

i) Mechanical Stoppages

These stoppages are due to mechanical failures (Failures in the mechanical parts of the equipment). Examples of these failures include: failure of a bearing, crack in a mechanical part and a drift in a belt conveyor...etc. These failures are fixed by the Mechanical department.

ii) Electrical Stoppages

These stoppages are due to electrical machines failures (Failures in the electrical parts of the equipment) or electrical and electronic control failures. Examples of these failures include: failure of an electrical motor, failure of an electronic card and failure of a logic circuit...etc. These failures are fixed by the Electrical department.

iii) Production Stoppages

These stoppages are due to production failures (Failures in the production process). Examples of these failures include: hot spots on the kiln body, fallen of thermal bricks inside the kiln, high erosion rate in iron balls inside the mill, material blockages because of wet material...etc. These failures are fixed by the Production department.

All of these stoppages are caused mainly by the failure of the machines and/or electrical and electronic control systems and /or failure in the production process due to inability to supply the proper quality of materials to meet the production requirements. These stoppages will cause unsteady operational conditions and will reduce the production rate of the production line and will reduce the availability, and all of these factors together will cause an increase in the energy consumption (thermal and electrical) as discussed in chapter six.

The main reasons behind the above-mentioned classification of the breakdown stoppages are:

i) To draw the attention of the plant management to the type of breakdowns and its effects on energy consumption and cost.

ii) To define the responsibilities of the concerned departments of these stoppages (i.e. Electrical; Mechanical; Production department). So as to rise their awareness about this very important matter, and that they will be accounted for any stoppages caused by their departments without any justified reasons.

iii) To improve the performance of the maintenance and production departments through the demonstration and clarification of the effects of these stoppages on the energy consumption; loss of production opportunities; and production cost.

iv) To help the decision makers for planning the future upgrading projects based on the continuous analysis of the stoppages and their causes, which may reveal the necessity for technical modification or upgrading projects to rectify any existing continuous technical problem, which may be discovered by the repetition of the stoppages.

B) Programmed Stoppages

Programmed stoppages are decided by the JCF production line staff in cooperation with the planning and maintenance departments. Although the production line is working there would be some reasons for stopping it. The reasons for programmed stoppages in kilns are different from that of the mills.

i) **Kilns programmed stoppages:**

Kilns are working 365 days/year and 24 hours/day continuously all over the year and stop only in the case of emergency breakdown or programmed stoppages (usually there is a period of 30 to 60 days per year of complete shutdown for the purpose of the major annual shutdown, this period is depending on the age and the type of the kiln concerned). The first reason, which is the breakdowns, has been discussed previously, and the second reason is programmed stoppages, which are divided into three types as follows:

a - Stoppages of kilns because of planned maintenance

The Planning department prepares an annual maintenance plan for all the production lines. The objective of this plan is to perform preventive maintenance in a suitable time (usually depends on the age of the thermal bricks inside the kiln). Sometimes the production line is stopped because of a break down, but the maintenance personnel implement a short program to perform some programmed jobs since the kiln is in shutdown condition; this maintenance is called (Opportunity maintenance). The additional stoppage time is added to the planned stoppages time duration. The benefit from this approach is to reduce the stoppage time for the programmed maintenance by eliminating the additional kiln cooling and firing times from the total stoppage time.

b - Stoppages of the kilns, which are caused, by the stoppages of mills

The kiln is stopped when the raw mill is stopped for a long time since the raw mill is the only source of raw materials, which is usually stored in a feed tank to feed the kiln. This is a rare stoppage due to the fact that the mill doesn't stop usually for a long time (the kilns and mills are stopped together in the planned maintenance program). There is a safety margin supply of raw material stored in the feed tank, which usually cover short period of stoppages of raw mill by continuing feeding the kiln during that period.

c - Stoppages of kilns because of low sales

When there are low sales and the output of the kilns is large. The top management may decide to stop one of the kilns or more. This is also a rare type of stoppages and if it happened, usually took months (not hours or days).

ii) Mills programmed stoppages

Raw mills work 24 hour/day as they feed the kiln, but as we mentioned above they can be stopped for some time without forcing the kiln to be stopped, since there is a large tank to feed the kiln and another raw mill, which could help in feeding the kiln.

Cement mills are usually stopped during the day and they work only at nights to take the advantage of the low tariff of electricity at night. Sometimes the mills work during the day when there is a high sales demand.

We have previously explained why the stoppages hours (mainly planned stoppages) of the mills is greater than the kilns and shows that it is not critical in producing the needed amount of cement because of the designed spare capacities and the designed operational procedures. The types of programmed stoppages of the mills are:

a -Stoppages of mills because of planned maintenance

The planning department prepares an annual plan for the production lines. Usually raw mills are planned to be stopped with the kilns, while cement mills have totally separated planned dates.

Planned maintenance for mills is less critical for the continuity of production than the kilns because of the designed spare capacities and the designed operational procedures (especially cement mills), since there is an opportunity to stop these mills to perform preventive maintenance without stopping the production.

b- Stoppages of mills because of no need for production

When the raw material tank or the cement silos are full, the mills output must be stopped, since there is no place to store it. This type of stoppages usually happens in cement mills when there is low sale, and happens in raw mills when the kiln is stopped for a certain period of time, which is enough to fill the kiln, feed tank.

c - Stoppages of cement mills during the day

This type of stoppages is for cement mills only (which is usually designed with over capacity to enable it to grind the 24 hours kiln production in 8 or 16 hours period of time) in order to reduce the electricity cost by utilizing the low electrical tariff in the off peak period during the night and day time. But sometimes when there is a high sale and enough stored quantity of clinker, cement mills works 24 hours/day to produce the needed amount of cement.

d- Stoppages of raw mills because of kiln stoppages

This is a special type of the above mentioned stoppages type (no need for production). It occurs when the kiln is stopped for a long time (days) especially in the case of planned maintenance, in which the raw mills are planned to be stopped since there is no place to store their output.

7.3.2 Definitions of Selected Variables

In order to build a regression model, one would need to define the selected relevant variables of the problem, define and prepare the historical data sets, generate and test several forms of models, and finally interpret the obtained results. From all the above analysis the following variables were selected:

<u>Dependent variables:</u>

1- Electricity consumption (EL)

2- Fuel consumption (FUEL)

<u>Independent variables:</u>

1- Averages duration of stoppage (AvHOURS)

2- Availability (AVL)

3- Average number of stoppage (AvNO)

4- Production rate (PRORATE)

5- Alumina Ratio (Aratio)

6- Silica Ratio (Sratio)

7- Lime Saturation Factor (Lime SF).

The specific definitions and units of measurements of the selected variables are given next.

A) The dependent variables are:

(1) Electricity consumption (EL):

This represents the first dependent variable, which is the average monthly electric energy consumed per one tonne of product measured in kilowatt-hour per tonne of product (KWH/tonne). Its formula is:

$$EL = \frac{\text{total electricity consumed during the month}}{\text{total amount of product during the same month}} \quad \text{(kwh/tonne)}$$

The electrical energy is used usually to crush or grind the raw materials and the clinker produced, and also to rotate the kiln and to run all the auxiliary systems.

The electrical energy is consumed in the kilns and the mills and auxiliary equipment, while fuel is consumed only in kilns. This variable represents consumption rate for each equipment (raw mill, kiln and cement mill), the overall consumption for one tonne of cement is the sum of the consumption rate for each equipment in the production line.

(2) Fuel consumption (FUEL)

This represents the second dependent variable, which is the average monthly thermal energy consumed per one tonne of product produced (kg fuel/tonne of product). Fuel consumption exists only in the kilns, where fuel is burned to generate thermal energy, which in turn burns the raw material, and the outcome of this burning and transformation process produces clinker. The method to calculate this variable is to divide the total fuel amount consumed during the month by the total amount of product (clinker) produced during the same month.

$$FUEL = \frac{\text{total fuel amount consumed during the month}}{\text{total amount of clinker produced during the same month}} \quad \text{(kg fuel/ tonne product)}$$

B) The independent variables are

(1) Average duration of stoppages (AvHOURS)

This represents the total stoppages hours for the equipment or the production unit during the month including all types of stoppages (down time + programmed stoppages) divided by the number of days in the month, the unit is (hour/day).

$$\text{Average duration of stoppages} = \frac{\text{duration of total stoppages per month}}{\text{number of days in the month.}} \text{ (hour/day)}$$

We explained previously in this section in details all types of stoppages and their causes and effects on the energy consumption. When we have shutdown condition in the kiln we need to restart and heat it up again. So we consumed additional energy without production, which lead to increase the average energy, consumed per tonne of product. Same consequences applied into raw and cement mill and the auxiliary equipment where in case of stoppages we need to restart again the equipment applied where very high starting current is needed and electrical energy consumed before reaching the operational steady state to start the production.

The two types of stoppages were combined together because those stoppages (what ever the type of them) are affecting the energy consumption as explained before, and mainly to avoid any statistical complication which may arise from an expected functional relation between breakdown stoppages (if it is used alone as an independent variable) and availability. The effect of the breakdown stoppages will be reflected through the analysis of the other independent variable "availability" which we are going to use in our analysis.

The method that was used to calculate the average duration of stoppage was as follows: -

- Determining the working hours during the month.
- Determining the total hours in the month depending on the number of

days in the months multiplying by 24 hr/day

- Total stoppage hours during the month is the difference between them
- Total stoppage hours are divided by the number of days in the month

$$\text{AvHOURS} = \frac{\text{total hours in the month} - \text{working hours during the month}}{\text{number of days in the month}} \text{ (hr/day)}$$

(2) Availability (AVL)

Availability is a function of both reliability and maintainability and answers the question, (is a system available to perform its intended function when it is needed?). Availability measures are time related and some of the time elements are (1) storage free or off time (2) operating time (3) stand by time and (4) downtime consisting of corrective and preventive maintenance as well as logistics and administrative delay time. (Kailash C. Kapuvand Leonard R. Lamberson, 1986).

According to Mickel P. Groover (1987), the term availability is sometimes used as a measure of reliability for equipment. It is especially germane for automated production equipment and it is defined using two other reliability terms, the mean time between failures (MTBF) and the mean time to repaire(MTTR). The mean time between failures(MTBF) indicates the average length of time between breakdowns of the piece of equipments. The MTTR indicates the average time required to service the equipment and place it back into operation when a break down is occure.

$$\text{Availability} = \frac{MTBF - MTTR}{MTBF}$$

The availability is an indicator of stable plant performance; the availability of production equipment is essential for the following reasons:

- To achieve high level of production, quality and productivity.
- To maintain uniform quality of finished product.
- To maintain production cost at lowest level.

Good availability implies the following:

- continuous control of material quality
- production equipment suitable for products
- high quality equipment operation and good maintenance practice

- competent personnel

In cement industry, because all equipment and production units are running in sequence, it is very important to achieve good availability in order to get high level of production at the lowest cost especially the energy cost.

The adopted definition of Availability is that it represents the percentage of the total available monthly hours in which the production line is producing or available for production and can be operated once there is a need for it, over the total hours in the month. The only reason for the production line that prevents it from not being available is the breakdowns stoppages. Sometimes the equipment is available but there are some external causes preventing the usage of the production line. These include: material shortage, programmed preventive maintenance, absent workers, low sales, strikes...etc. Although the equipment is stopped due to these reasons, the production line is available and could be operated once these external reasons are removed. This definition facilitate to JCF top management to control and supervise the performance of the factories by measuring their performance based on the availability parameter by excluding the influence of any external factors from affecting the value of this parameter

$$\text{Availability (JCF)} = \frac{\text{total hours in the month} - \text{breakdown time hours}}{\text{total hours in the month}}$$

$$= \frac{\text{Total available monthly hours}}{\text{Total hours in the month}}$$

HolderBank as an international cement producer, use the term " availability index "as a comparison factor between its factories to measure its performance.

$$\text{Availability index (holderbank)} = \frac{\text{Operating time}}{\text{Total available time}}$$

But availability index gives correction for low sales and strikes ...etc for the same reason used by JCF to isolate any external causes which may affect the availability index parameter, thus the operating time becomes closer to the definition of the available time and this makes the availability index closer to the availability in JCF definition.

(3) Average number of stoppages (AvNO)

This represents the total number of stoppages of the production line per a month divided by the number of days in the month. This number of stoppages includes all types of stoppages (break down: mechanical, electrical, production and programmed maintenance).

$$\text{Average number of stoppages} = \frac{\text{Number of stoppages during the month}}{\text{Number of days in the month}}$$

(number of stoppages/day)

We mentioned previously in details the types of stoppages and their implications mainly during stoppages. The kilns consume energy without production and if the number of stoppages increases we need to restart up again the production unit and consume energy in each start up condition (electrical and fuel) up till kilns reach a steady state operational condition where kilns can start produce again. All these stoppages circumstances will result in increasing the average energy (fuel and electricity) consumption per unit of product.

(4) Production rate (PRORATE)

This represent the average monthly amount of product produced per hour, its unit is tonne/hour. Production rate is measured every hour and it is changing in accordance with the operational conditions. The method to calculate the average production rate for the whole month is by dividing the total production amount during the month by the total operating hours in the month.

$$\text{production rate (PRORATE)} = \frac{\text{Total production amount per month}}{\text{Total operating hours during the month}} \quad \text{(tonne/hour)}$$

Usually the kiln should produce the basic design capacity in case of optimum operational condition. But in case of unsteady operational conditions or stoppages, the production rate will decrease causing the energy consumption to increase per unit of product as we found out when we discussed the factors affecting the energy consumption in chapter six.

(5) Alumina ratio (Aratio)

Alumina ratio is defined as the ratio of Alumina to iron oxide. Although iron oxide has a great effect in creating the liquid phase which helps in forming mineral clinker, especially when free silica is found in raw meal, experiments of the production and quality control departments have proven that an alumina ratio of (1.38) gives better burning conditions inside kilns, due to the creation of a greater liquid phase with the least possible temperature. Alumina ratio is in the range of 1.5 to 2.5.

(6) Silica ratio (Sratio)

Silica ratio is the ratio of $SiO2$ to iron oxide plus alumina oxide. Controlling the silica ratio to keep it within the required range (2.1 – 2.5) guarantees that the kiln will not be fed with a hard burning raw meal, which increases thermal energy consumption, producing uncontrolled operation, and a product, which contains free lime.

A higher silica ratio impairs the burnability of clinker by reducing liquid phase content and tendency toward formation of coating in the kiln. An increased silica ratio also causes a slow setting and hardening of the cement while a decreased ratio increases the content of liquid phase thus improving the burnability of the clinker.

(7) Lime Saturation Factor (Lime SF)

Lime saturation factor is defined as the ratio of the effective lime content to the maximum possible lime content in the clinker. In cement manufacturing calcium carbonate is the chemical, which requires the highest amount of heat to be produced, as in the following equation:

$$CaCO_3 = CaO + CO_2 - 698 \text{ Kcal / kg.}$$

As calcium carbonate forms (70-75%) of raw meals for cement manufacturing, working with a calciner which has a relatively low saturation factor leads to a decrease in fuel quantity needed for carbonate, and a drop in combinability temperature.

Lime SF ranges, normally between 90 and 98. The higher the Lime SF the more heat is needed for clinker burning because it becomes hard to burn and forms dusty clinker containing free lime. Therefore, it is important to have an optimum value for this factor according to specifications and operating conditions.

7.3.3 Data Sets

Data are taken for kilns at the Fuhais Cement Manufacturing Plant, located 15 kilometers west of Amman and kilns in Rashadiya Cement Plant, located 200 kilometers south of Amman. Historical data on dependent and independent variables were collected from various relevant records of both plants, with the help of engineer Jalal Sader, and it covers an investigation period of four years (1990-1993). Data related to the study variables are included in Appendix (01), including description and summarized specifications of the kilns, which are subject to our analysis.

Multiple linear regression analysis was used as a statistical tool to determine the presence and significance of potential relationships among the variables of the studied data.

It should be emphasized that this research is the first attempt for using statistical analysis for energy management purposes at JCF. Also exhaustive effort to try to trace any previous similar analysis in the cement or other heavy industry was done without any success. The extent of this search of any other experience related to our research was reported previously in chapter three.

7.4 Limitations and Assumptions

7.4.1 Limitations of this Work

This work has the following limitations.

i) The data that is used in this study is reported monthly and we found that some of the data is rounded by the clerks who reported them.

ii) The data has some outliers, their number is very limited and we managed to handle them using exploration techniques that identify the outliers. Moreover, robust regression analysis is used to fit regression models to the data since this is the suitable procedure the existence of the outliers in the data.

iii) The maximum number of observation used in this analysis is 46, which seems to be not a big number but acceptable and expected to give a reasonable regression.

iv) This work may be considered as a good starting point for future research in this direction. There is also a need to collect information on more independent variables to expand the explanation power of the statistical model.

7.4.2 Expectations of the Researcher

In chapter six, preliminary detailed empirical analysis of the main factors affecting energy consumption was carried out. These factors include availability, production rate, average number and duration of stoppages and other factors affecting the quality of produced cement.

In essence, this preliminary empirical analysis indicates, in a clear manner, that the activation of energy control factors within an EMS is not only possible but also beneficial. However, a detailed statistical and economical analysis to define the relationship between energy consumption (dependent variable) and the control factors (independent variables) is needed and will be covered and the researcher expect to achieve the following:

A) Several statistical models will be developed to represent the relationships between energy consumption on one hand, and management functions and practices as given by the variables on the other hand.

B) Detailed statistical analysis of the factors affecting the consumption of energy in the cement industry in order to develop the necessary mathematical functions relating energy consumption to the selected significant factors, will be established, which will help to demonstrate, mathematically, the significance of energy management in controlling the energy consumption and cost.

C) The statistical model will use historical data for the above selected independent factors and the dependent factors; namely electricity and fuel consumption. The outcome of this statistical analysis will be a model, which may be used, to a certain degree of confidence, to describe the relationship between each of the two dependent variables (electricity and fuel consumption) and the selected independent variables. Such derivation of relationships between energy consumption (as a dependent variable) and management functions (as independent variables) would allow managers to plan, forecast and control these functions in order to optimise management variables and consequently the consumption of electrical or thermal energy can be predicted.

D) The statistical models will be, in turn, employed as a basis for developing economic models and translating the impact of sound management practices and tools into quantifiable economic variables. The objective of building the economic model is to verify whether there will be financial gains from improving the control variables as predicted by the statistical model or not. Furthermore, the size of such gains can be assessed using the economic model.

E) Based on the empirical analysis in chapter six and on practical experience in the factories, the researcher expects that

i) The consumptions of each of electricity and fuel affected by the selected independent variables AvNO, AvHOURS, PRORATE, AVL, Aratio, Sratio, and LimeSF, possibly with different degrees.

ii) Each of the independent variables AvNO, AvHOURS, Sratio, and LimeSF has an increasing relationship with each of the dependent variables (EL and FUEL).

iii) Each of the independent variables PRORATE, AVL and Aratio has a decreasing relationship with each of the dependent variables.

So the main objective of this research is to study the validity of these expectations.

7.5 Statistical Analysis

This section reports a statistical analysis of the data under consideration. The complete computer output is given in the Appendices 03-21.

7.5.1 General Procedure

To analyze the available data, we have run the following procedures from the NCSS2000 Package for each kiln. A summary of the terminology and procedures in this package are given in Appendix (02).

1. Scatter Plots Procedure that is used to obtain scatter plots of each of the dependent variables against each of the independent variables. These plots are used to explore the possible existence of linear relationship between each independent variable and each of the dependent variables. They also may indicate the possible outliers in the data. Moreover, they may suggest some transformations of the data to achieve linearity.
2. Correlation Matrix Procedure, which is used to obtain the correlation coefficients between the dependent variables and the independent variables. These correlation coefficients may pin the independent variables that should be removed from the regression model because they have non-significant correlations with the dependent variable. The signs of the significantly correlation coefficients may indicate the type of effect of the independent variable on the dependent variable Y.

   This procedure is also used to find the pairwise correlations between the dependent variables for each kiln. Very high correlations between these variables may indicate the existence of multicollinearity problem.
3. A data Screening Procedure is used to detect the existence of possible outliers for each variable for each kiln. Moreover, it is used to explore normality of the dependent variables through tests based on skewness and kurtosis. It also plots the histograms of

the dependent variables.

4. Multiple Regression Procedure is used to fit a full model of the dependent variable as a function of all the available independent variables. This procedure provides the estimated coefficients of the model with t-test for the significance of each coefficient together with standard deviations and confidence intervals of these coefficients. It also provides diagnostics of the fitted model together with checking all the underlying assumptions of the linear multiple regression model. These include normality tests of residuals, testing serial-autocorrelations, Durbin-Watson test, and testing multicollinearity using variance inflation and condition numbers.

Moreover, multiple regression procedure provides residual analysis through graphical plots, which may be used to check the homogeneity, and zero mean of residuals. These plots may also be used to spot possible outliers. For the basic scatter plots of residuals versus either the predicted values of Y or the independent variables, Hoaglin (1983) explains that there are several patterns to look for. These patterns include, point cloud, wedge, bowtie, sloping band, and curved band.

A point cloud, basically in the shape of a rectangle or a horizontal band, would indicate no relationship between the residuals and the variable plotted against them. This is the preferred condition.

An increasing or decreasing wedge would be evidence that there is increasing or decreasing (nonconstant) variation. A transformation of Y may correct the problem, or weighted least squares may be needed.

Bowtie is similar to the wedge above in that the residual plot shows a decreasing wedge in one direction while simultaneously having an increasing wedge in the other direction. A transformation of Y may correct the problem, or weighted least squares may be needed.

Sloping Band is a kind of residual plot that suggests using a linear version of the independent variable to the model

Curved Band is kind of residual plot which may be indicative of a nonlinear relationship between Y and the independent variables that was not accounted for. The solution might be to use a transformation on Y to create a linear relationship with the X's. Another possibility might be to add quadratic or cubic terms of a particular independent variable.

It is worth noticing that the NCSS2000 Package reports a PRESS R-squared value only for the ordinary multiple regression model.

5. All Possible Regression Procedure, which searches for the best model of size 1,2,..., k, where k is the number of independent variables included in the model. According to NCSS2000 package, this procedure depends on R-Squared, Root MSE, and Cp as selection criteria. The Cp criterion selects the model with the smallest value of Cp, while the other two criteria select the model with the largest value of R-Square or/and Root MSE.

   When the selected variables are fitted for each size 1,2,...,k, one can observe if the estimated coefficients from a model with smaller size to a model of higher size are comparable or not, If they are comparable, then one may conclude that the multiple linear regression may be suitable for the available data.

   It should be noted that the relative importance of the independent variables may be obtained from this procedure. However, this relative importance may not hold for other regression procedures.

6. Stepwise Regression Procedure, which searches for the best model. According to the NCSS2000 package, this procedure applies four searching procedures, namely, Forward, Backward, Stepwise, and MSE procedures.

7. Robust Regression Procedure that is useful in case that normality assumption is not satisfied or/and in case of the existence of outliers in the data. This procedure re-weights the data. According to the NCSS2000 package, there are three methods (influence functions) to run the robust regression analysis, namely, Least Absolute Deviation with robust truncation factor 1.0, Tukey's biweight with robust truncation factor 6.0, and Andrew's sine with robust truncation factor 2.1. (See the help of the NCSS Package). It is interesting to note that these procedures may lead to different results because they apply different trancation factors that affect the weights given to outliers in the data. We will use the procedure that provides the largest value of R2. For the sake of comparison, we will apply the robust regression even if normality is satisfied and/or there are no outliers in the data. If the results of the multiple regression and the robust regression are close this will confirm that there are no outliers and/or normality assumption is satisfied.
8. Ridge Regression Procedure, which is useful in case that there is a doubt of multicollinearity. If there are cases were multicollinearity may not be a problem we will not apply ridge regression. However, for the sake of comparison we will apply the ridge regression model for Kiln1 regardless of multicollinearity being a problem or not.
9. Some nonlinear regression models will be produced and compared with the other linear models. These will include multiplicative models, quadratic, polynomial with interaction terms, and logarithmic models.
10. It may happen that the full multiple, ridge, or robust regression models contain independent variables whose estimated coefficients are not significantly different from zeros. Therefore, there is a need to remove such variables from the model and run the procedure again. The obtained model after this procedure will be called the

final model. So, the Regression analysis is applied at two stages. Stage one is a screening procedure to select the significant variables through the t-test of the coefficients of the full model, i.e. the model that contains all available independent variables. Moreover, the all-possible regression procedure and the stepwise regression procedure help in this screening process. The second stage is used to obtain the final model as a function of only the significant independent variables. It should be remarked here that different selction procedures may lead to different sets of selected independent variable. In such a case the selected variables based on the t-values will be used if there is no major problem in the underlying assumptions. However, if there is a major problem we will also use the independent variables that are selected according to the Cp criterion.

It is interesting to note that the robust regression procedure depends heavily on the obtained set of residuals of the fitted model. Therefore, in the case of using a final model after some nonsignificant variables have been omitted, it is quite possible to have results of the final model that are noncomparable wth the results of the full robust model.

11. Finally according to the R-squared value and the validity of the assumptions on the model and the procedure used to obtain that model, we will recommend a model for each case. Moreover, the sensibility of the estimated coefficients will be another factor in this recommendation.
12. Finally the recommended model will be stated.

A detailed analysis will be given for EL and FUEL in Kiln1, and a summary of the analysis of other kilns and mills will be given in the rest of this Chapter. This is because the same analysis is performed for all kilns and mills and since the results seems to be similar. The full computer output for all kilns, mills and data from the reputable manufacturer will be given in Appendices 03-21 at the end of the thesis. These appendices are:

1. Appendix 01: Data Used
2. Appendix 02: NCSS2000 Statistical Package
3. Appendix 03: Complete Computer Output of Kiln 2
4. Appendix 04: Screening of data of other Kilns
5. Appendix 05: All Possible Regression of Kilns
6. Appendix 06: Stepwise Regression of Kilns
7. Appendix 07: Multiple Regression of Kilns
8. Appendix 08: Robust Regression of Kilns
9. Appendix 09: Screening of Transformed Data of Kilns
10. Appendix 10: Multiple Regression of Transformed Data of Kilns
11. Appendix 11: Multiplicative Models of Kilns
12. Appendix 12: Polynomial Regression with Interaction Terms for Kilns
13. Appendix 13: Quadratic and Linear Models with Four Variables
14. Appendix 14: Screening of Mills Data
15. Appendix 15: All Possible Regression of Mills
16. Appendix 16: Stepwise Regression of Mills
17. Appendix 17: Multiple Regression of Mills
18. Appendix 18: Robust Regression of Mills
19. Appendix 19: Analysis of Holder Bank Data
20. Appendix 20: Analysis of Nonlinear Regression
21. Appendix 21: Effect of Rounding Data on the Regression Models

7.5.2 Exploration of the Data

We are planning to fit linear multiple regression models to EL (and FUEL) in terms of the independent variables under consideration. In this section we will explore some of the properties of the available data (for each kiln), which are related to the assumptions of such models. So, we start by producing scatter plots of EL (and FUEL) against each of the independent variables for each kiln. These plots will help in exploring the linearity assumption. Secondly, we calculate the pairwise correlation coefficients between EL (and FUEL) and each of the independent variables. These coefficients provide a more quantitative method to judge the linearity assumption. To have an idea about the possibility of the existence or non-existence of multicollinearity problem, we calculate the correlation coefficients between the independent variables. If there are very high (strong) correlations between some independent variables , then multicollinearity could be a problem. Finally, to have an idea about the distribution of the dependent variables (EL and FUEL) we have run normality tests and plotted both the histogram and the normal probability plot of each of EL and FUEL for each kiln. Other tests of the underlying assumptions will be considered when we fit the regression models because those tests depend on the obtained residuals.

7.5.2.1 Scatter Plots of Dependent variables vs. Independent Variables

Tables (7.1.a), (7.1.b), (7.2.a) and **(7.2.b)** represents the scatter plots of the dependent variables against the independent variables for all kilns (kiln 4,5 and 6 in Fuhais Plant and Kiln 1 and 2 in Rashadiya Plant). It is seems reasonable from the scatter plots presented in these tables that the data has some outliers. For example, in the case of EL of kiln6 there is an obvious outlier. Moreover, regardless of the outliers, there are cases where one may claim that there is a linear relationship between EL (or FUEL) and some of the dependent variables. For example, in case of kiln4 there is a linear relationship between EL and each of AvHOURS, and PRORATE. But such a relationship is not clear with other independent variables. Moreover, there are cases where it seems that there is no relationship between EL (or FUEL) and some of the independent variables. For example in case of EL of kiln5, it seams there is no relationship between EL and Sratio. These descriptions depend on personal judgments. So, there is a need for some quantitative methods to describe the possible types of relationships. This will be done latter.

Table (7.1.a): Scatter plots of EL vs each of the independent variables/Fuhais

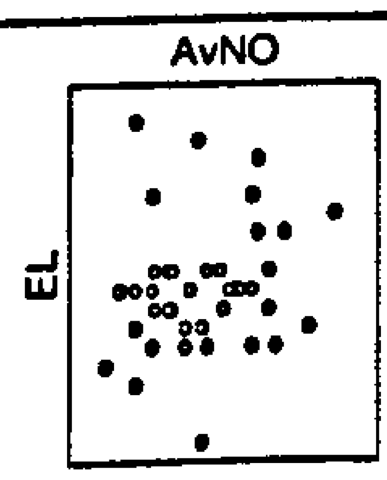
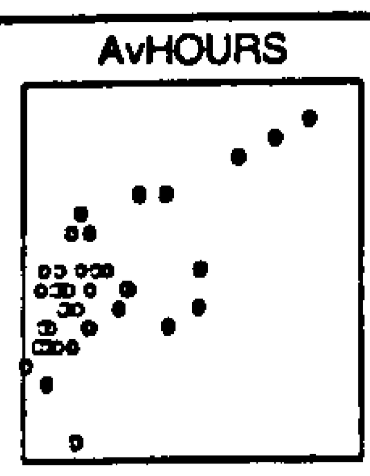
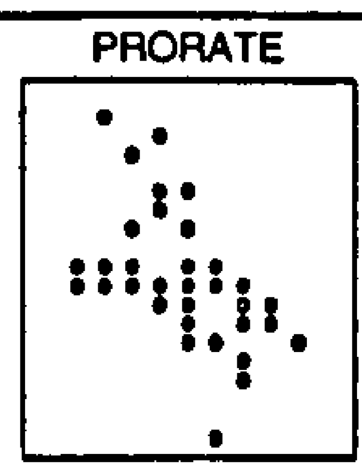
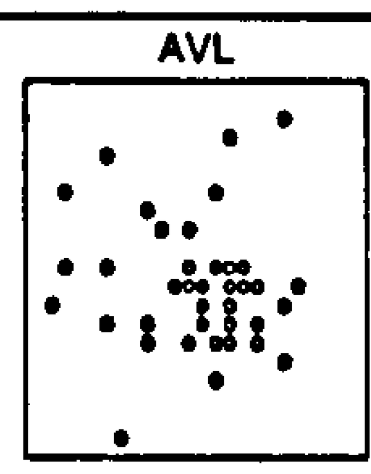
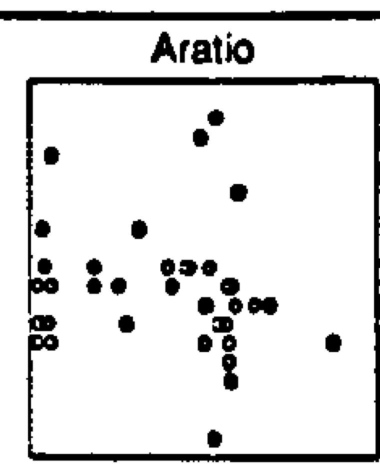
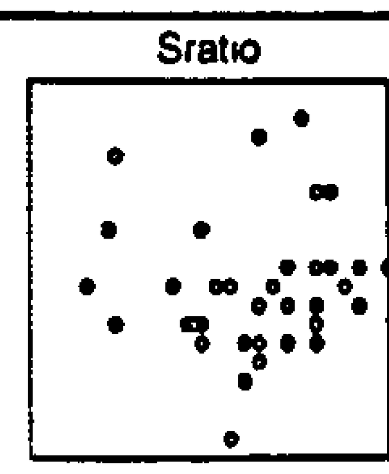
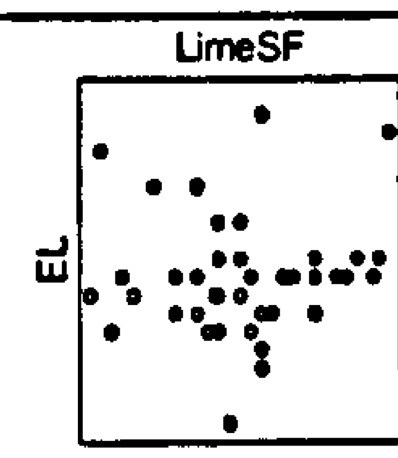
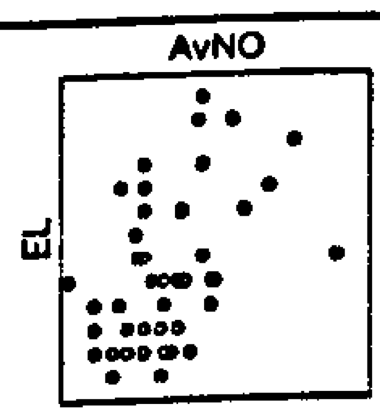
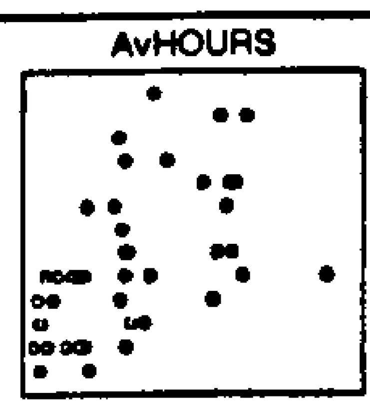
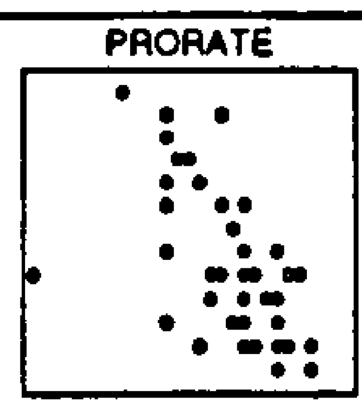
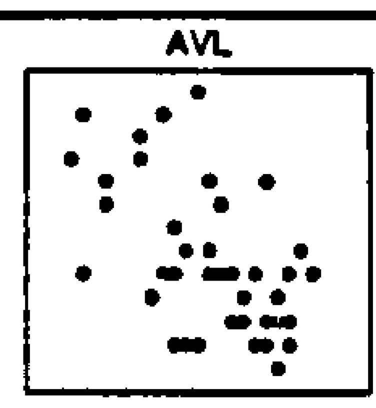
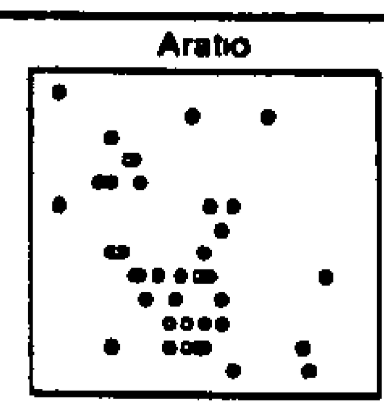
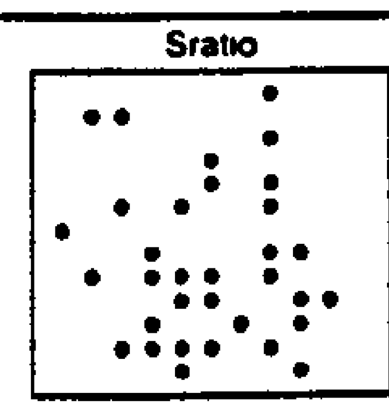
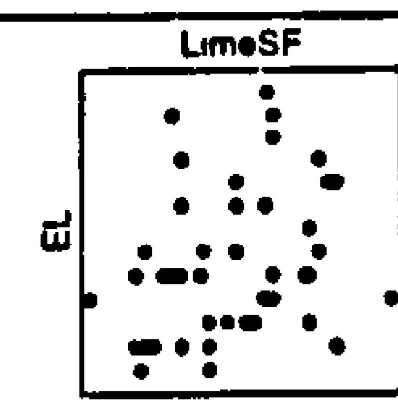
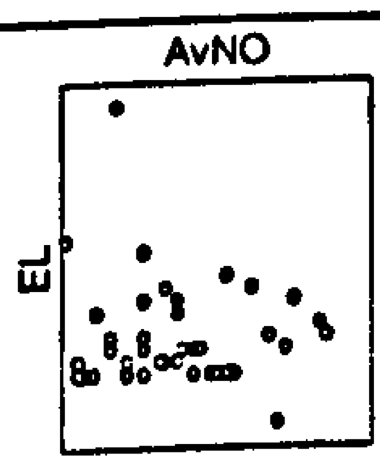
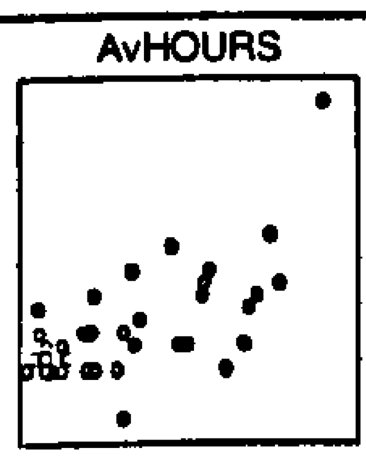
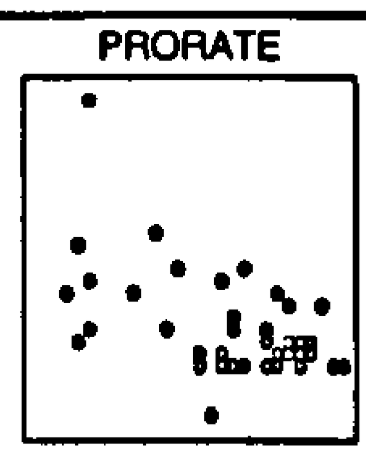
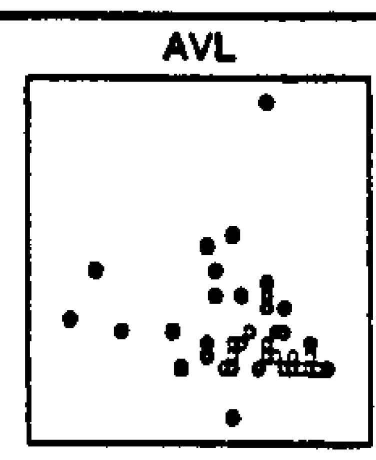
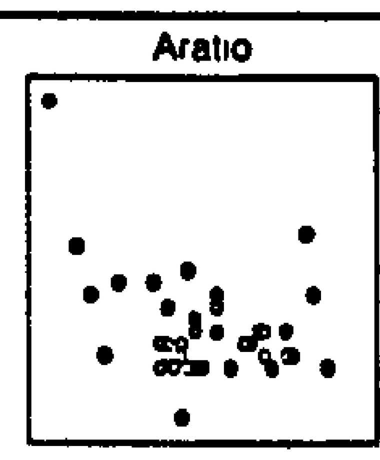
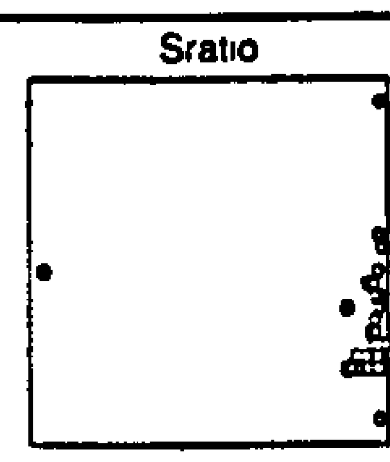
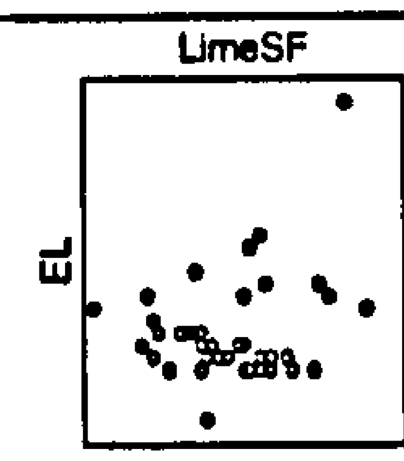
Kiln	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF
K4							
K5							
K6							

Table (7.1.b) Scatter plots of EL vs. each of the independent variables/Rashadiya

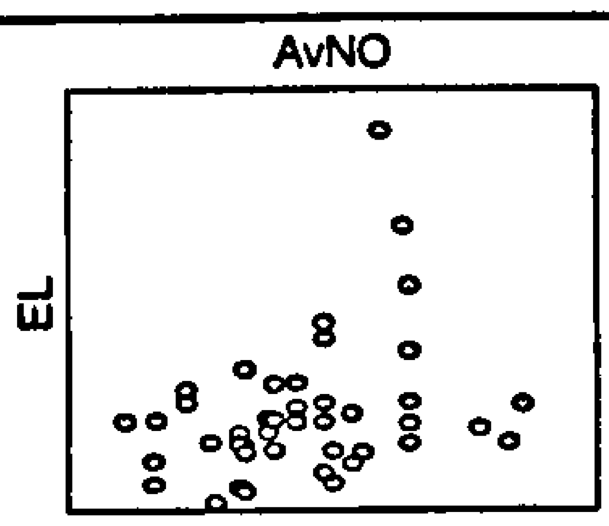
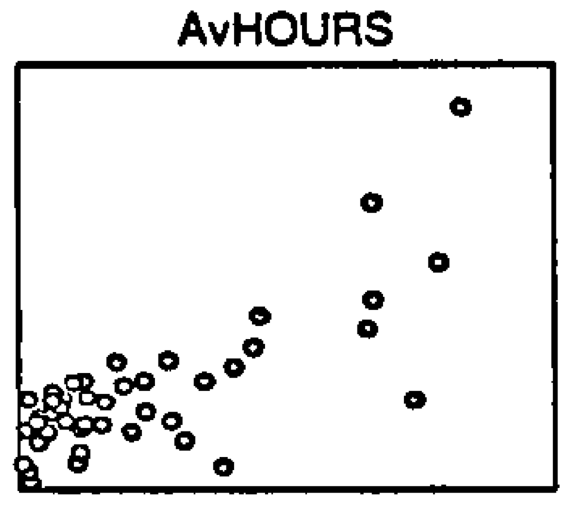
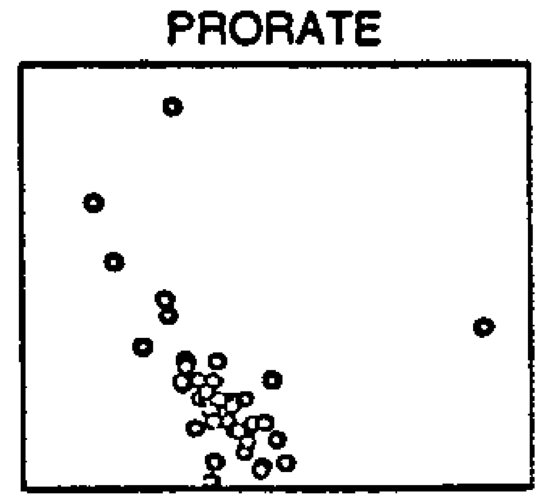
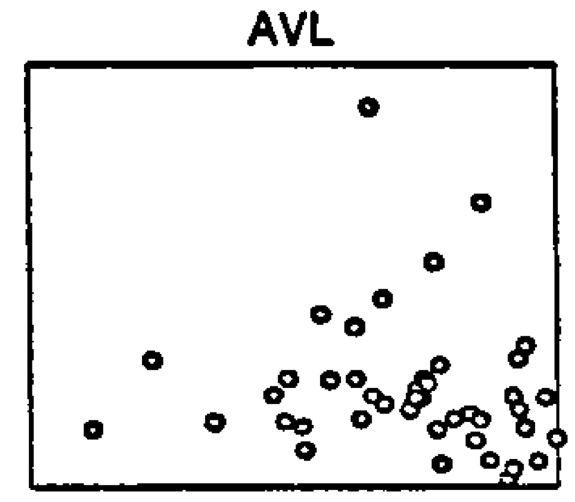
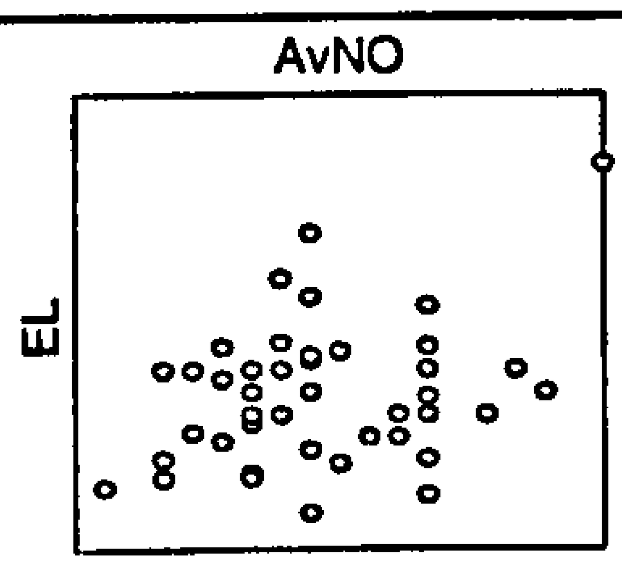
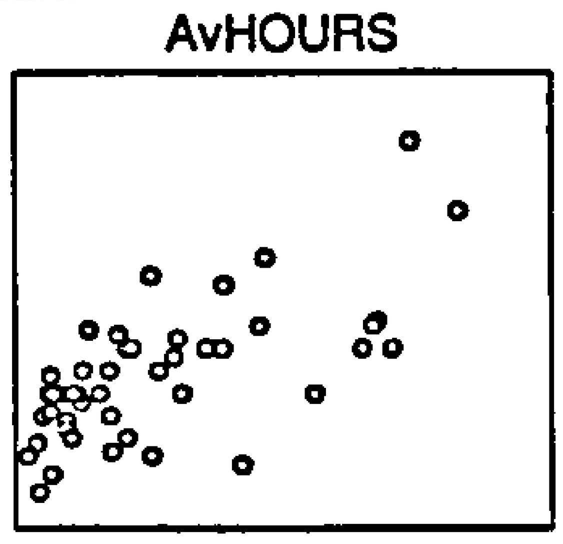
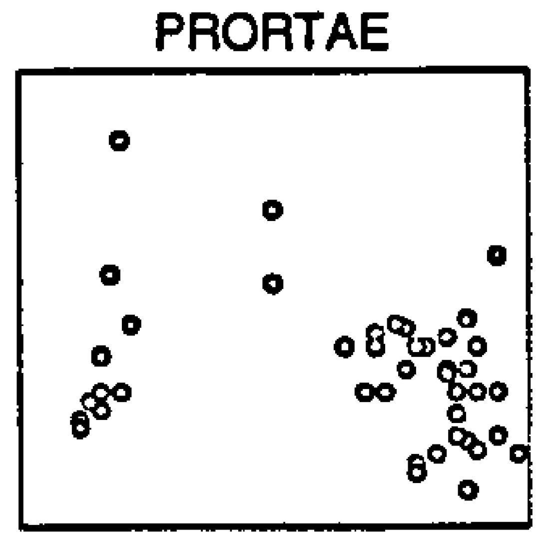
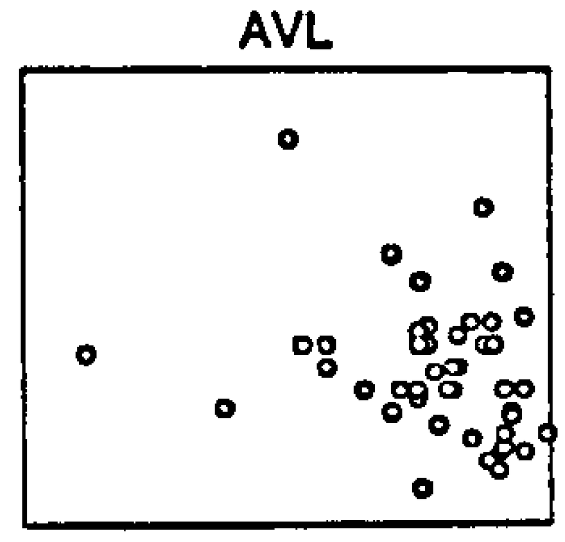
	AvNO	AvHOURS	PRORATE	AVL
K1				
K2				

Table (7.2.a): Scatter plots of FUEL vs. each of the independent variables/Fuhais

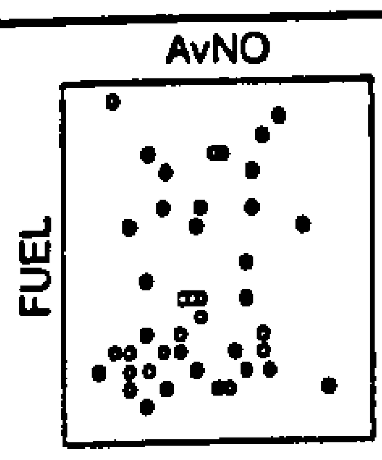
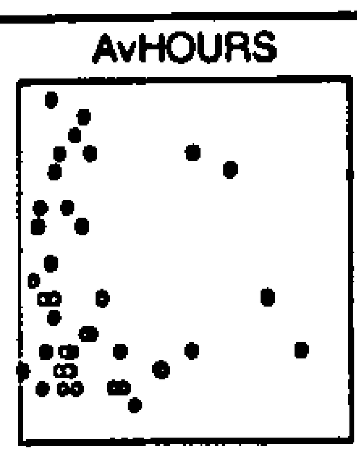
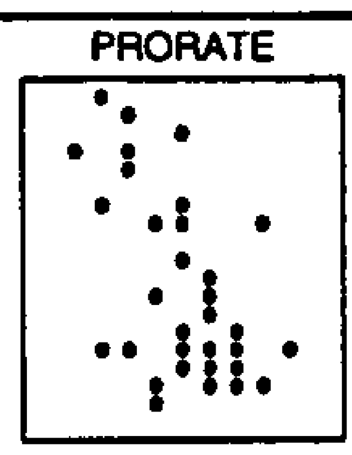
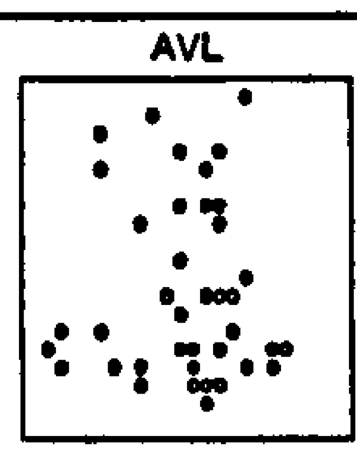
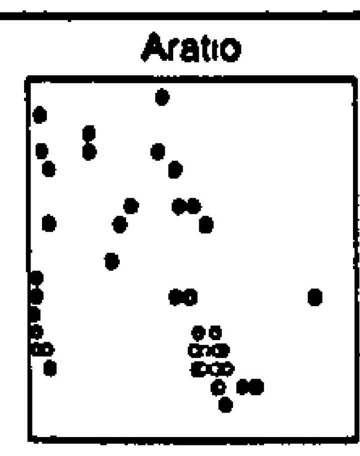
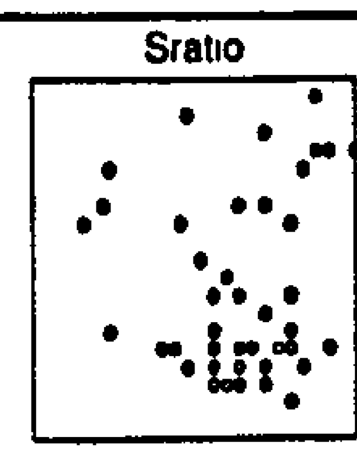
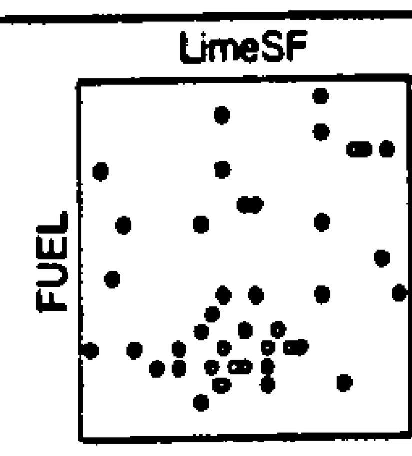
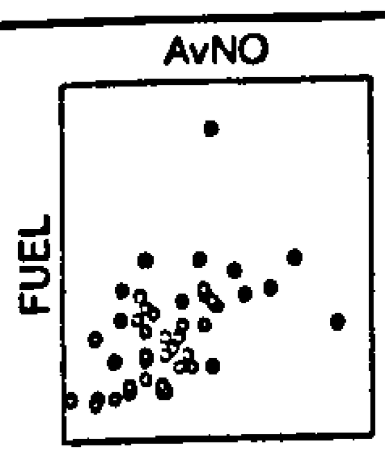
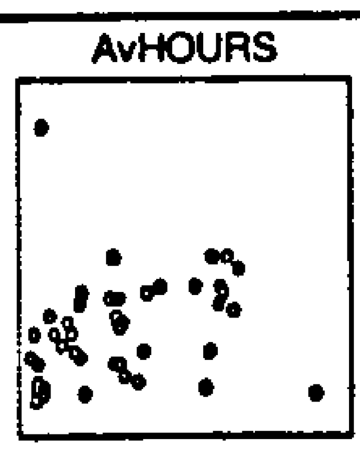
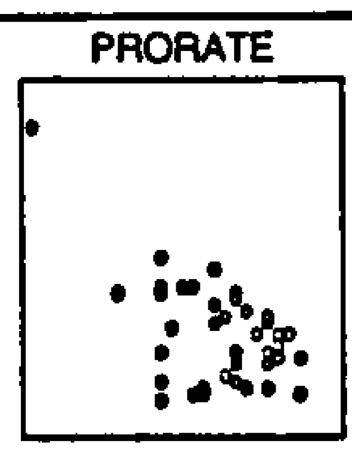
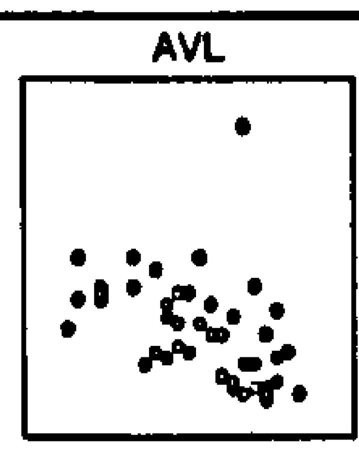
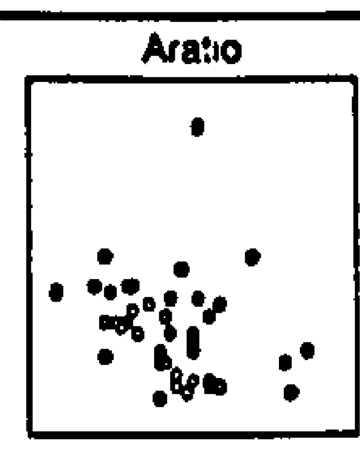
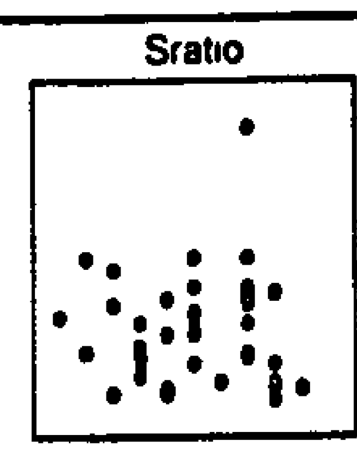
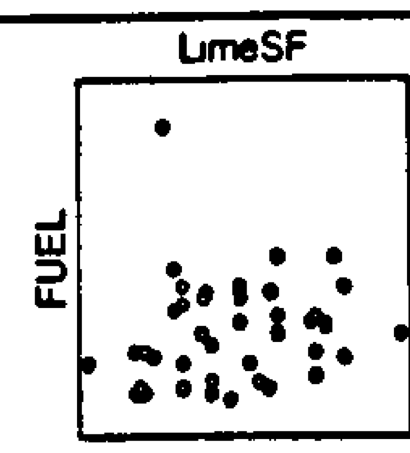
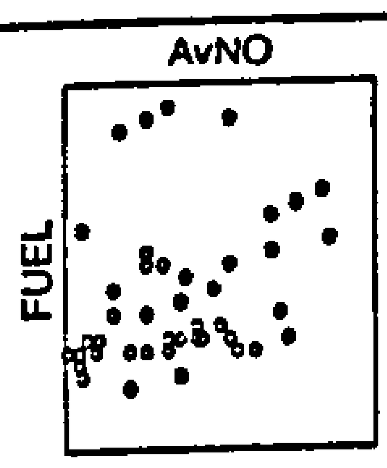
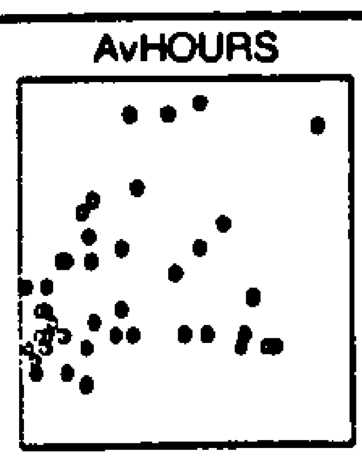
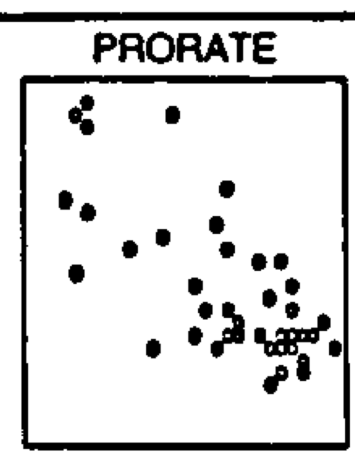
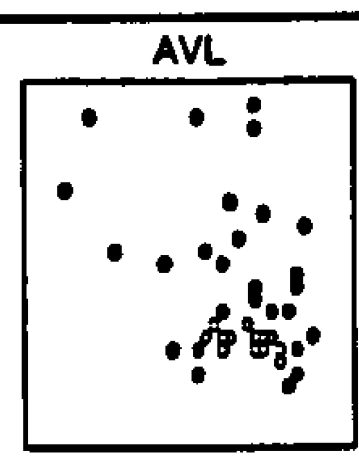
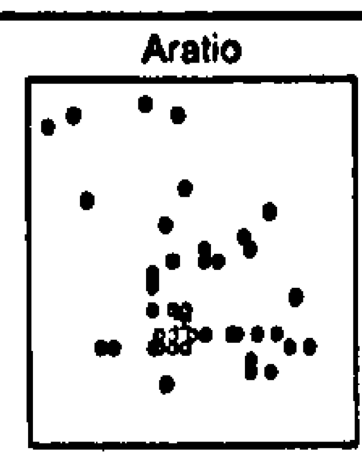
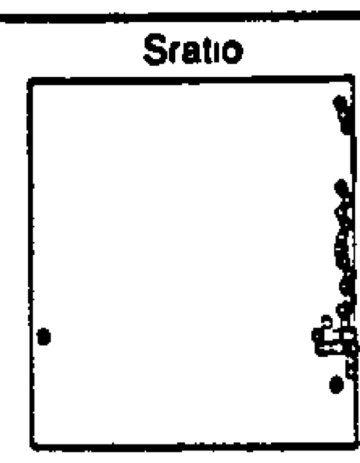
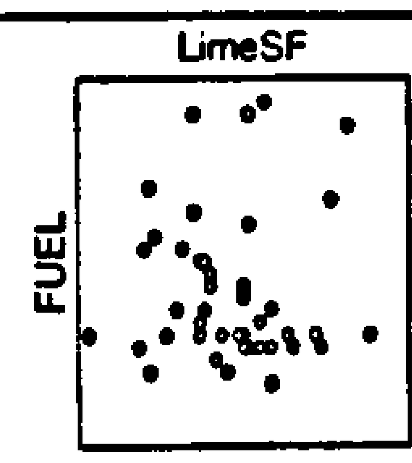
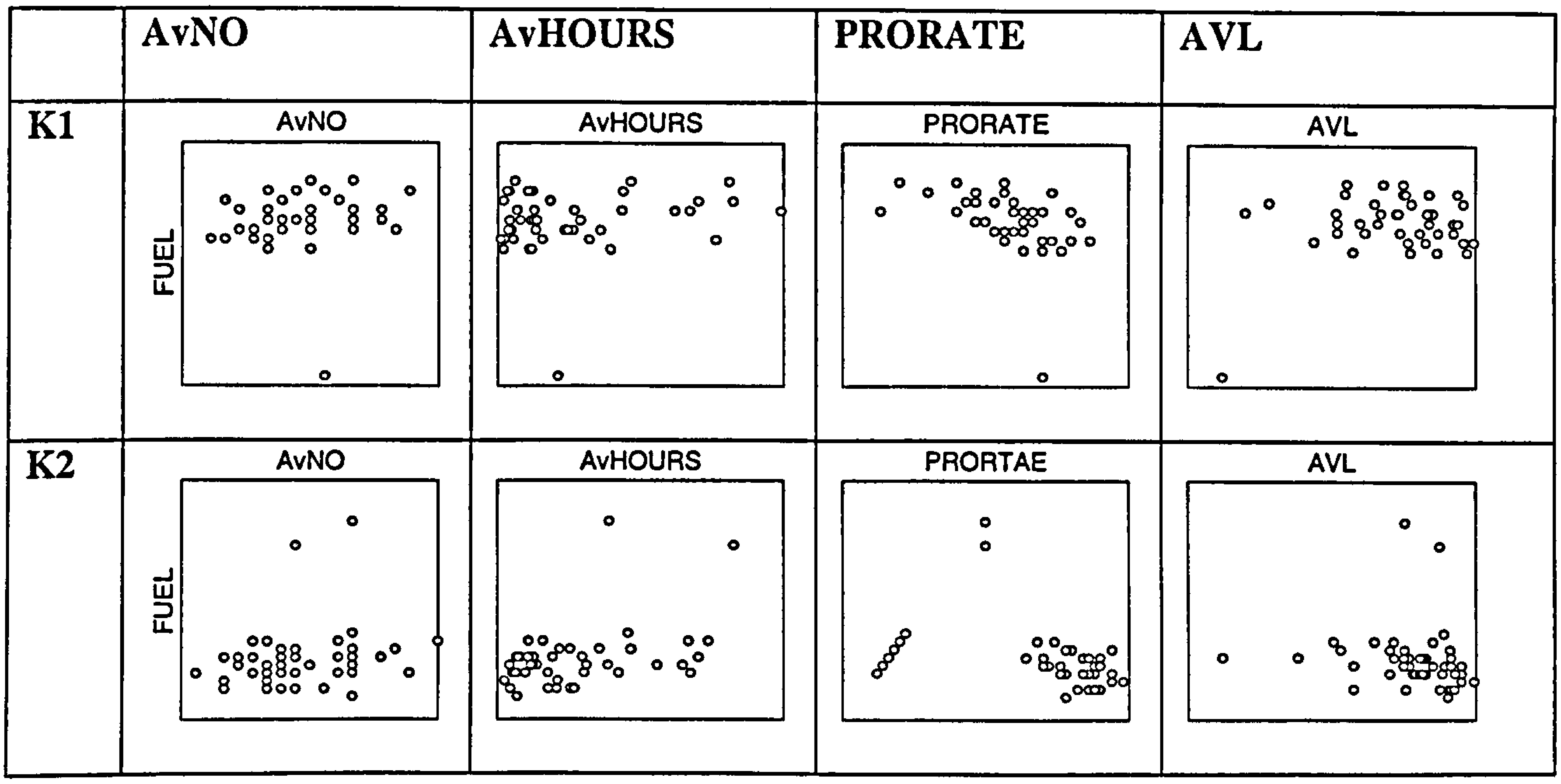
Kiln	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF
K4							
K5							
K6							

Table (7.2.b): Scatter plots of FUEL vs. each of the independent variables/Rashadiya



7.5.2.2 Outliers in the data

To search for outliers, we have used the Data Screening Program in NCSS2000 which locates outliers based on a T-test type. The results are given in the Appendices for each kiln. Table (7.3) reports the numbers of the outlying observations. For example, in kiln4, the AvNO has only one outlying observation. It is observation number 5. A blank cell indicates that there are no outliers. It is seems reasonable from Table (7.3) that the data has some outliers. The existance of these outliers suggests using regression methods that are not affected by outliers and/or minimize there effects.

Table (7.3): Numbers of observations, which are outliers

Kiln	EL	FUEL	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF
4	1, 34, 35, 38	21	5	1,34,38	9,22,23	3,16,17	28	31,39	8
5	6,14,23	35	2,13	29	35	33	36,41,42	22	12,30
6	24	46	34,39	24	23,25,28	37,39,40	24	6	12,44
1	23,24	34	31,33	12,17,24	11,23	15,38			
2	15	16,24	25,42	3,16,25	16	28			

7.5.2.3 Correlation Between Dependent Variables and Independent Variables

Assuming normality, the Pearson's correlation coefficient is a quantitative measure of linear relationship. However, it is affected by the existance of outliers. So, one should take its values with some caution.

To have a more clear idea, there is a need to calculate the correlation matrix between EL (or FUEL) and the independent variables for each kiln.

Tables (7.4) and (7.5) provide the pairwise correlation coefficients between the dependent variable EL (or FUEL) and each of the independent variables, together with the corresponding p-value, which is used to test the significance of correlation. If the p-value is less than say 0.05, then at the 0.05 level of significance the null hypothesis that says, "that the correlation coefficient is zero" is rejected, otherwise we cannot reject that hypothesis.

Table (7.4): Correlation coefficients (and their p-values) between EL and independent variables

Kiln	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF
4	0.113564, 0.468395	0.711479, 0.000000	-0.514869, 0.000412	-0.04275 0.785471	-0.04722 0.76361	0.048582, 0.757042	0.091321, 0.560294
5	0.439431, 0.002841	0.546581, 0.000124	-0.534700, 0.000185	-0.57695 0.00004	-0.38526 0.009809	-0.112646, 0.466603	0.268562, 0.077952
6	-0.082809, 0.593067	0.654625, 0.000001	-0.554805, 0.000093	-0.20699 0.17759	-0.31889 0.03487	0.153171, 0.320882	0.250252, 0.101342
1	0.317961, 0.035437	0.763674, 0.000000	-0.400497, 0.007062	-0.09699 0.53109			
2	0.179687, 0.232121	0.364040, 0.012880	-0.339608, 0.020941	-0.06676 0.659315			

Table (7.5): Correlation coefficients (and their p-values) between FUEL and independent variables

Kiln	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF
4	0.237053, 0.125876	-0.087562, 0.576613	-0.575602, 0.000054	-0.068766, 0.661272	-0.420117, 0.005035	0.084271, 0.591075	0.266252, 0.084385
5	0.483033, 0.000897	0.215094, 0.160864	-0.558598, 0.000081	-0.431385, 0.003460	-0.299988, 0.047880	0.018793, 0.903625	0.139483, 0.366518
6	0.265155, 0.081948	0.340764, 0.023604	-0.713741, 0.000000	-0.394591, 0.008035	-0.377072, 0.011632	0.178126, 0.247347	0.004251, 0.978152
1	0.256898, 0.092299	0.254850, 0.095016	-0.393835, 0.008168	-0.119446, 0.439950			
2	0.176976, 0.239359	0.462303, 0.001218	-0.551312, 0.000072	-0.179437, 0.232783			

It is seems reasonable from this table that a p-value that is less than or equals to 0.05 indicates that the correlation is not significantly different from zero at 0.05 level of significance. This means that some of the independent variables may be not significant if they are included in the regression model for some Kilns. However, for the sake of comparison, all the stated independent variables will be included in the first stage of applying multiple regression models for all kilns, and the t-test will be used to exclude those, which are not influential in predicting the dependent variable, from the final model.

Moreover, the signs of the correlation coefficients agree with the expectations of the researcher as has been stated in Section, 7.4.2.

7.5.2.4 Correlation Between Independent Variables

To have an idea about the possible existence of multicollinearity, Tables (7.6.a)-(7.6.e) report the correlation coefficients between all pairs of the independent variables. Multicollinearity is a problem if the correlation is very high, i.e. if its value is greater than 0.75

Table (7.6.a): Correlations between independent variables for kiln 4

Kiln 4	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF
AvNO	0.017915	0.017096	-0.674955	-0.039477	0.050851	0.016702
AvHOURS		-0.176623	0.005222	0.121681	0.131835	0.018611
PRORATE			-0.007470	0.259653	-0.190055	-0.264637
AVL				0.256171	0.092373	0.088349
Aratio					0.444158	-0.042203
Sratio						0.241475

Absolute max correlation is 0.674955

Table (7.6.b): Correlations between independent variables for kiln 5

Kiln 5	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF
AvNO	0.058096	-0.158303	-0.585609	-0.235468	0.202800	0.111812
AvHOURS		-0.237467	-0.150328	-0.018158	-0.114003	-0.091950
PRORATE			0.116095	0.211031	-0.185676	-0.030708
AVL				0.167042	0.018137	-0.185101
Aratio					-0.358697	-0.253471
Sratio						-0.225996

Absolute max correlation is 0.585609

Table (7.6.c): Correlations between independent variables for kiln 6

Kiln 6	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF
AvNO	0.060785	-0.211198	-0.510319	0.081642	0.096136	-0.161600
AvHOURS		-0.395775	-0.192361	-0.120933	0.122261	0.210710
PRORATE			0.056486	0.393853	-0.211708	-0.075056
AVL				-0.155872	-0.230385	0.371611
Aratio					0.141898	-0.360947
Sratio						-0.434453

Absolute max correlation is 0.510319

Table (7.6.d): Correlations between independent variables for kiln 1

Kiln 1	AvHOURS	PRORATE	AVL
AvNO	0.057085	-0.064054	-0.325681
AvHOURS		-0.171846	0.033761
PRORATE			-0.113734

Absolute max correlation is 0.325681

Table (7.6.e): Correlations between independent variables for kiln 2

Kiln 2	AvHOURS	PRORATE	AVL
AvNO	-0.106441	-0.160228	-0.364133
AvHOURS		-0.617968	-0.017128
PRORATE			0.246447

Absolute max correlation is 0.617968

It is seems reasonable from these tables that, for each kiln, the maximum absolute value of these Pairwise correlation coefficients is less than 0.75 which, indicates that multicollinearity may not be a problem.

7.5.2.5 Normality of Dependent Variables

The multiple regression model $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$ assumes that ε has a normal distribution with zero mean and constant variance. Hence if the X's are assumed deterministic variables then the dependent variable Y should follow a normal distribution. So, it seems reasonable to test Y for normality. Moreover, after fitting the regression model, one will obtain the values of the residuals that can then be tested for normality.

Tables (7.7) and (7.8) provide the computer output of the normal probability plots for both dependent variables (EL and FUEL) for each of the six kilns.

The normality plots, which are given in these tables, indicate that the data has some outliers and the data of EL and FUEL are not normal in several cases. Even in the cases where normality seems to be resonably accepted should be take with caution, since the existance of outliers may lead to false conclusions. Normality tests will be reported when we deal with transformations of dependent variables in the nonlinear regression section.

Table (7.7) Normal Probability Plots of EL

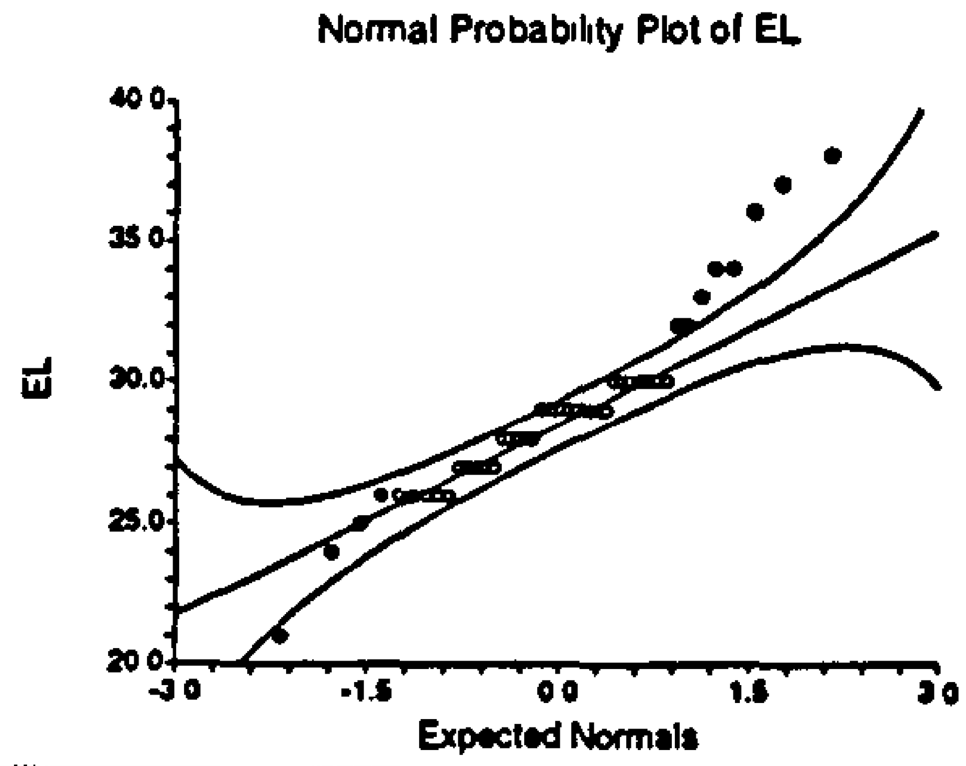
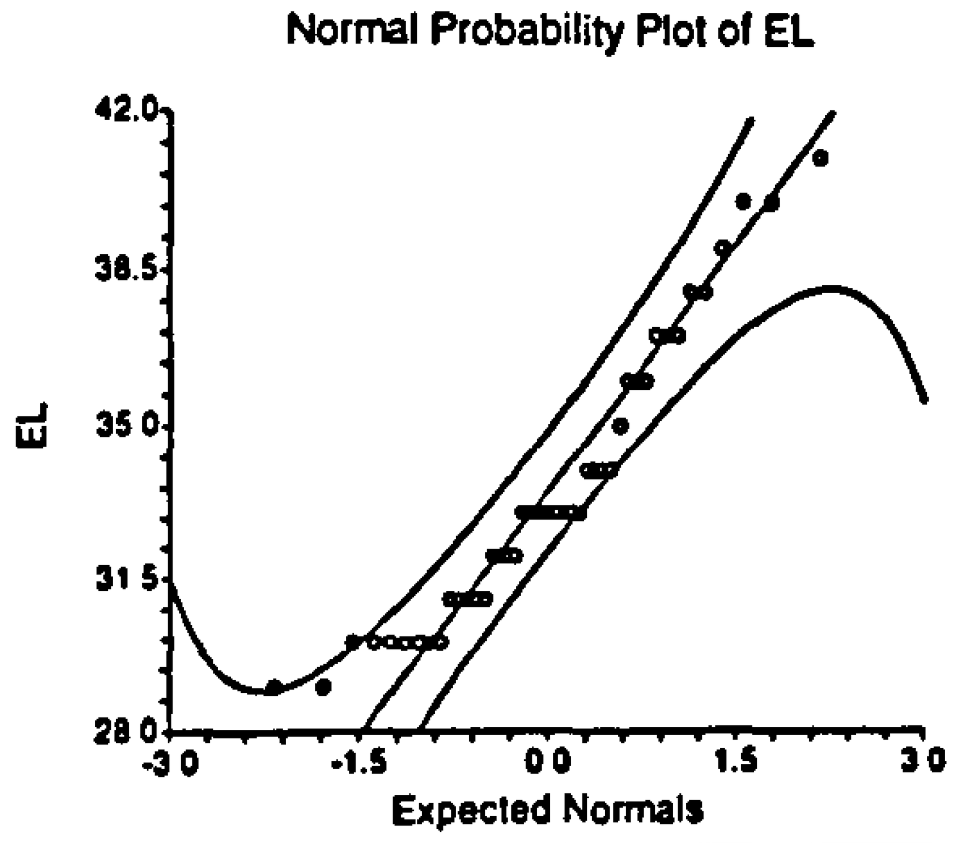
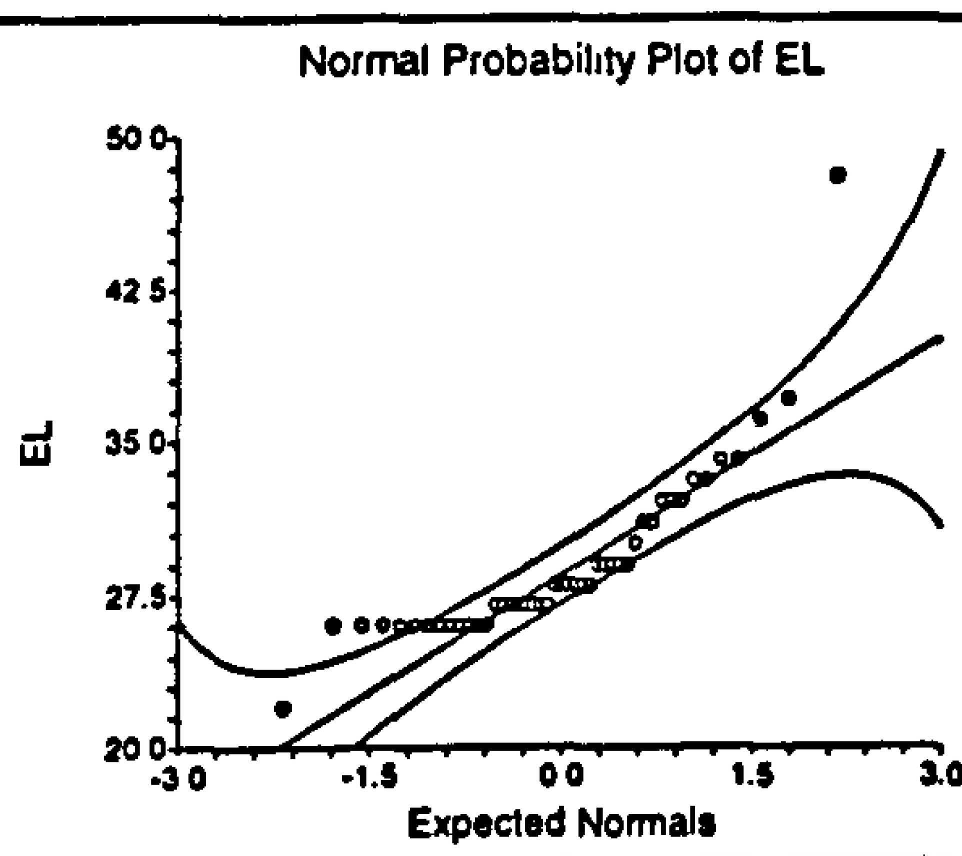
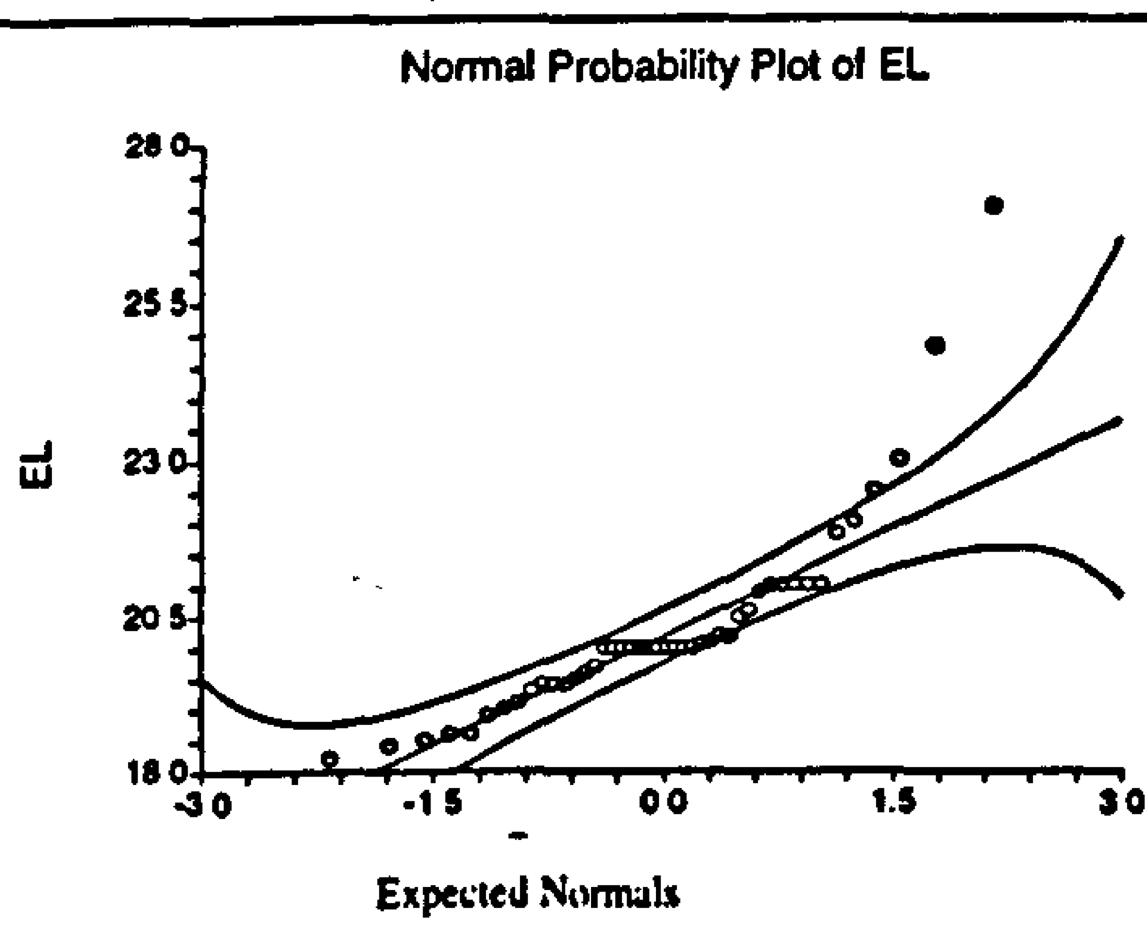
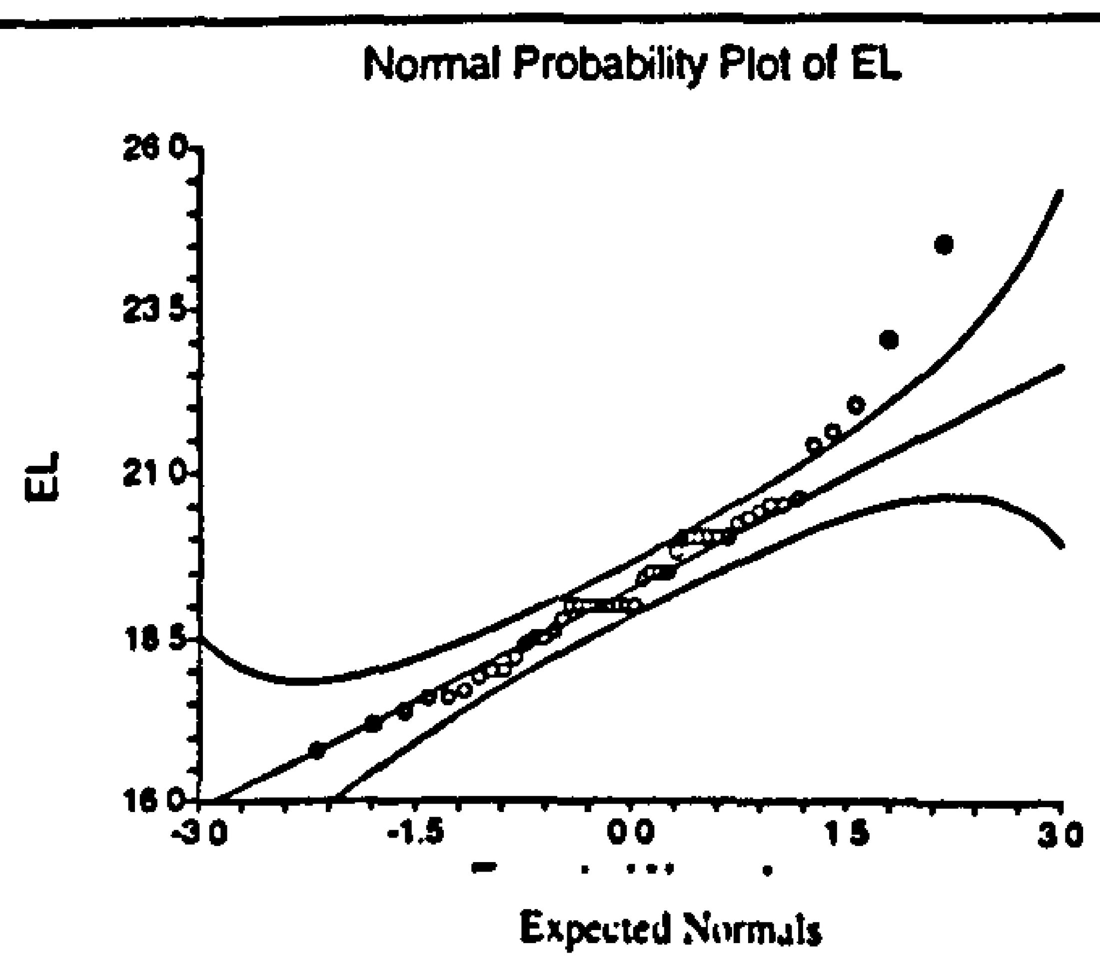
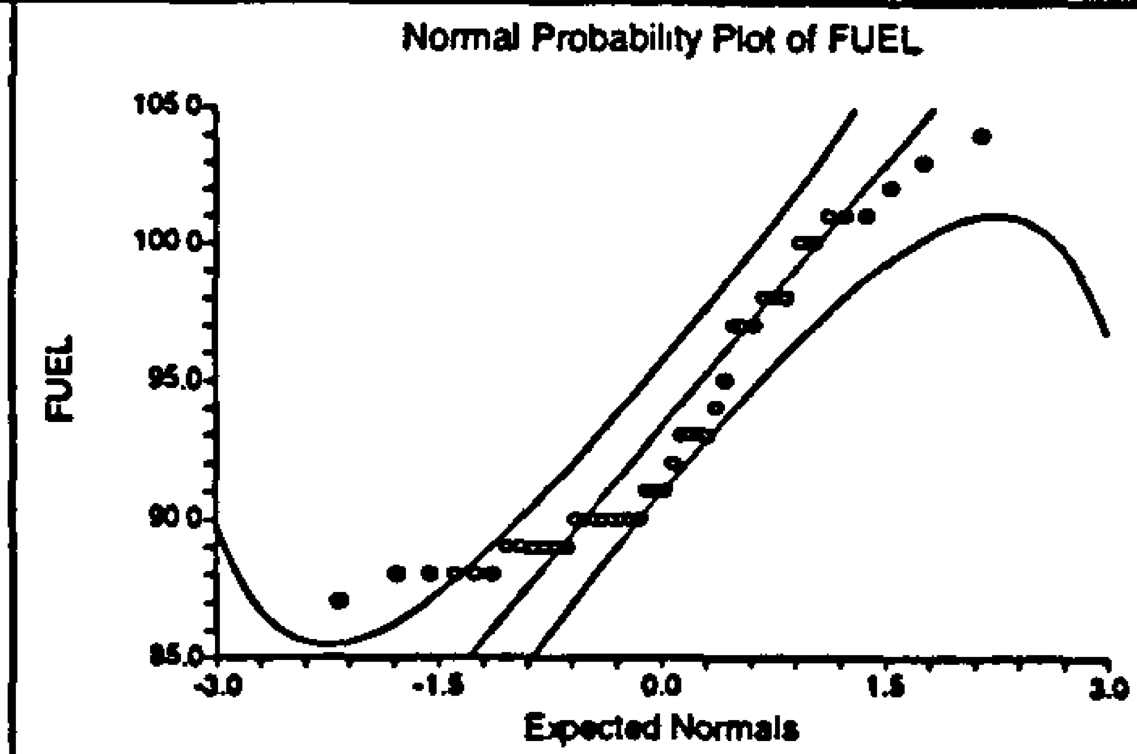
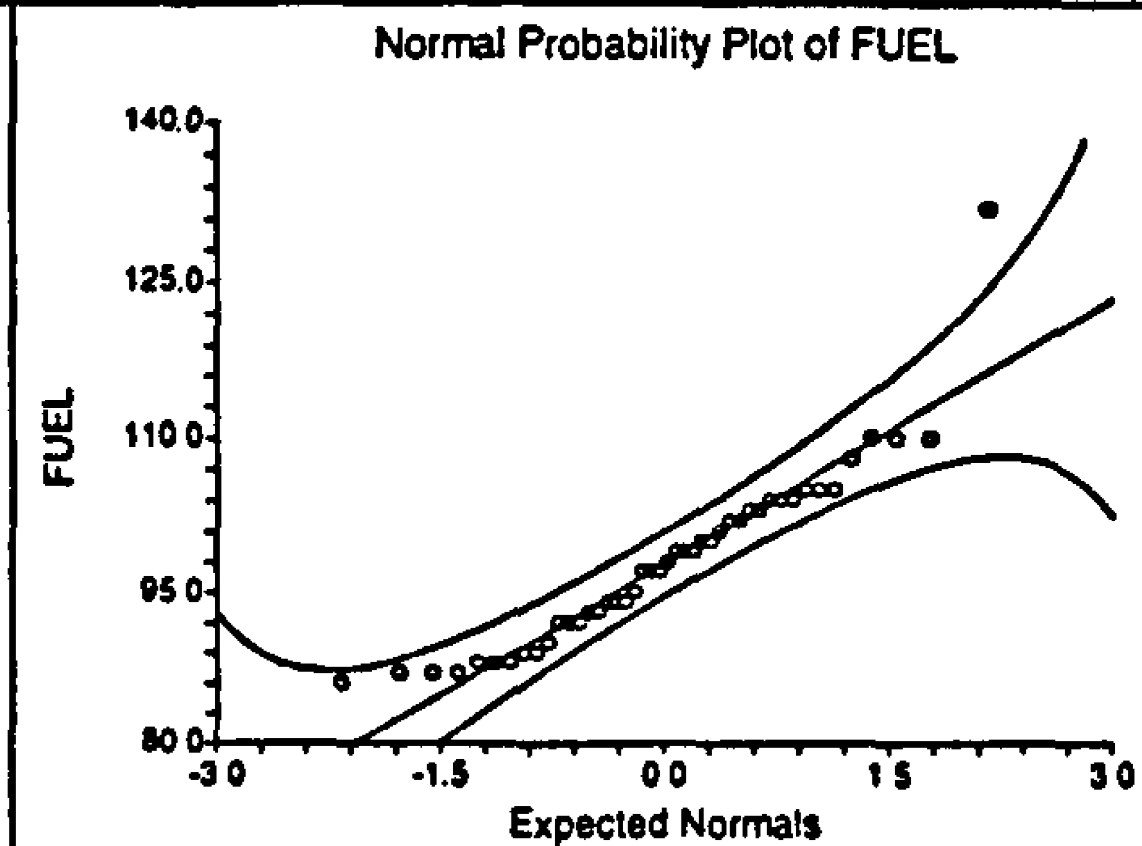
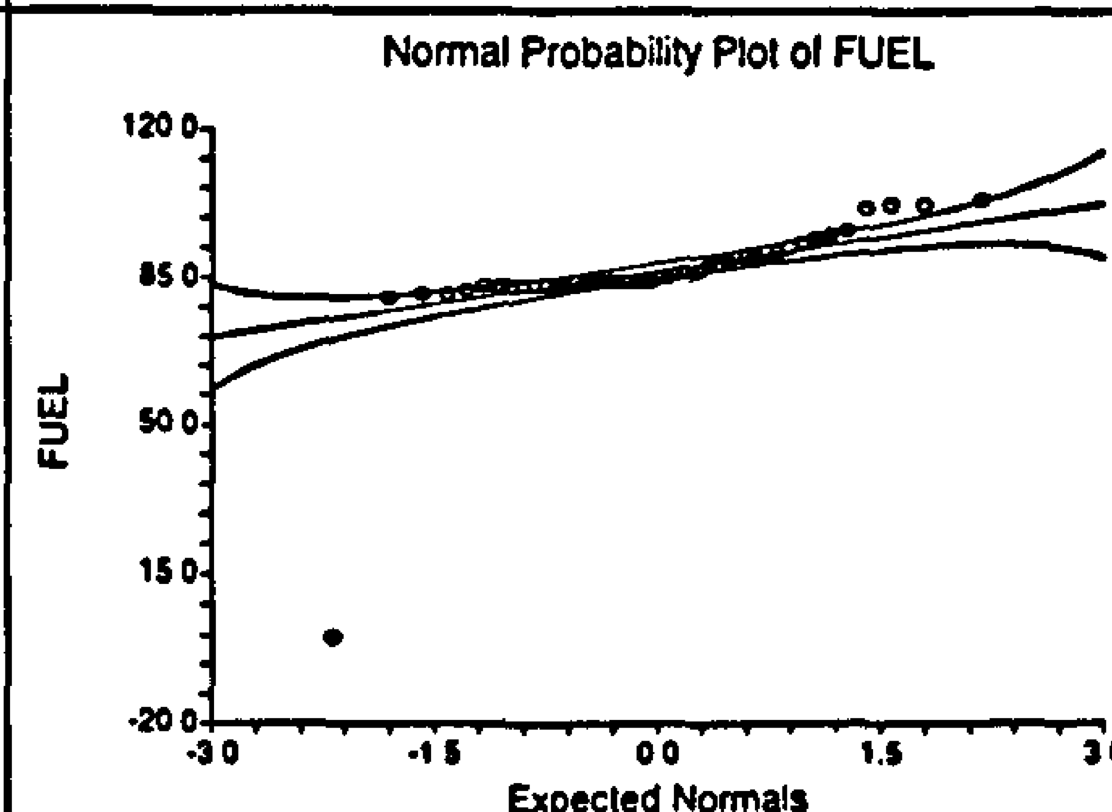
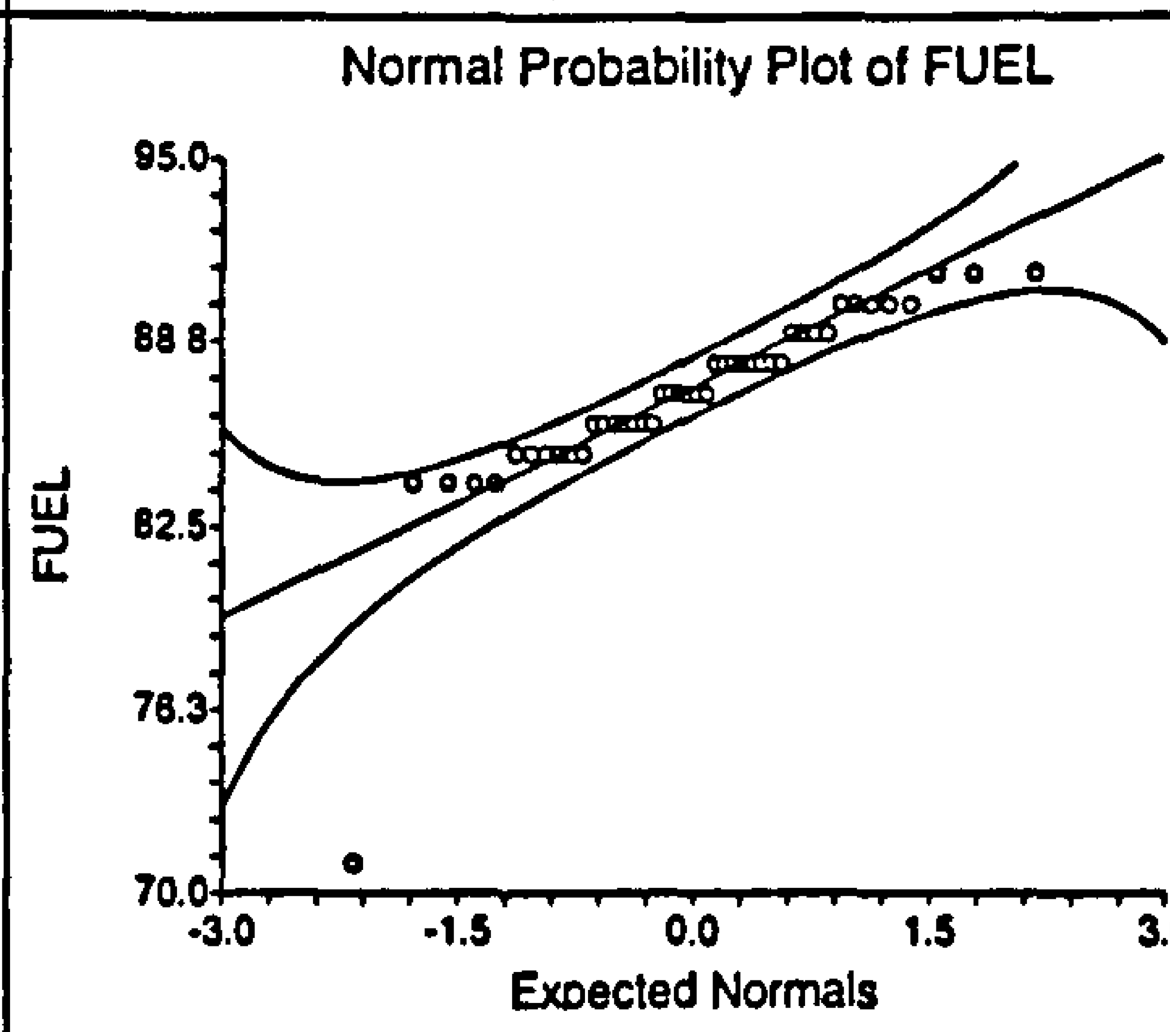
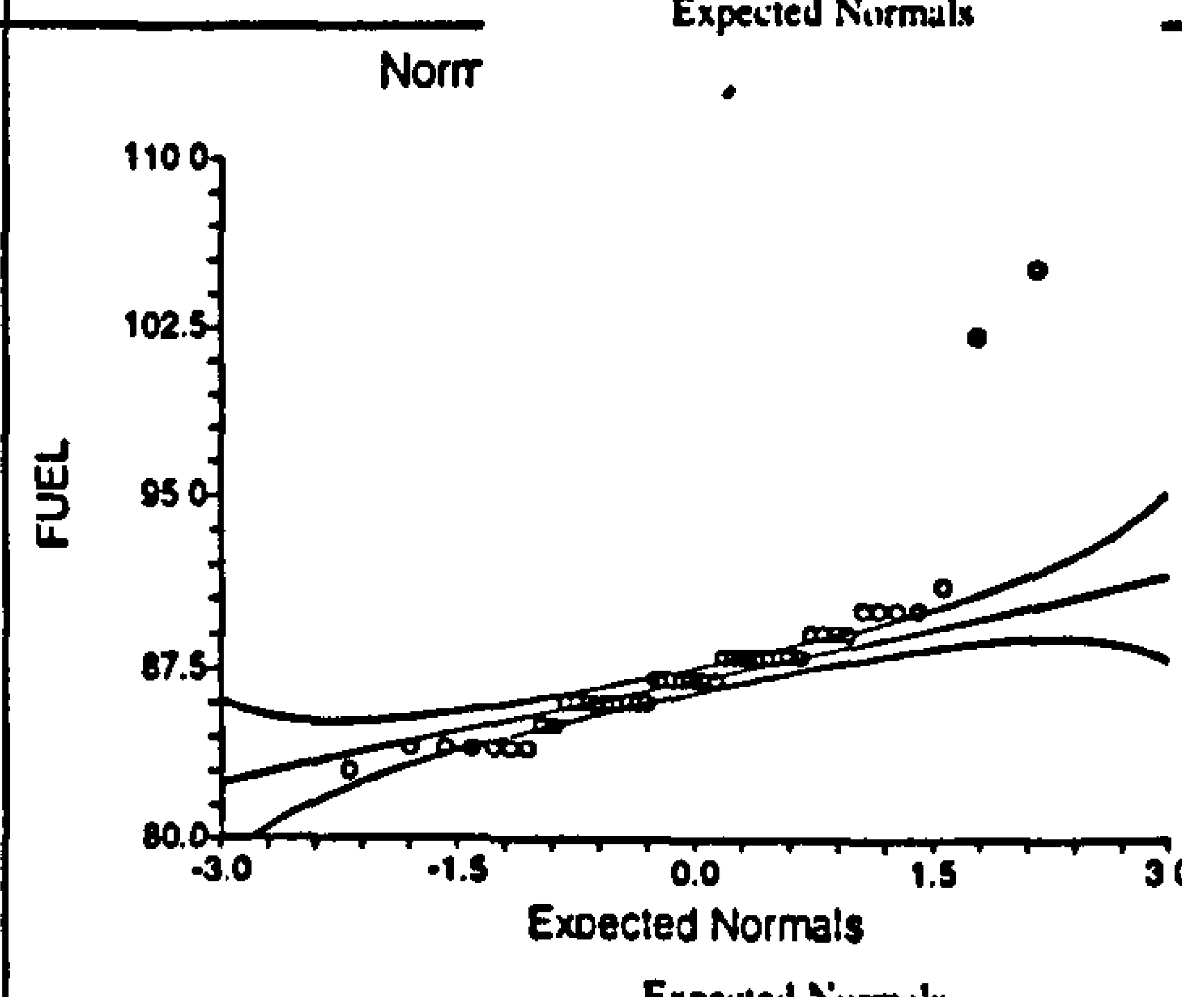
EL	Normal Probability Plot
K4	
K5	
K6	
K1	
K2	

Table (7.8) Normal Probability Plots of FUEL

FUEL	Normal Probability Plot
K4	
K5	
K6	
K1	
K2	

7.5.3. Presentation of Regression Analysis of Kiln 1

7.5.3.1 EL for Kiln 1

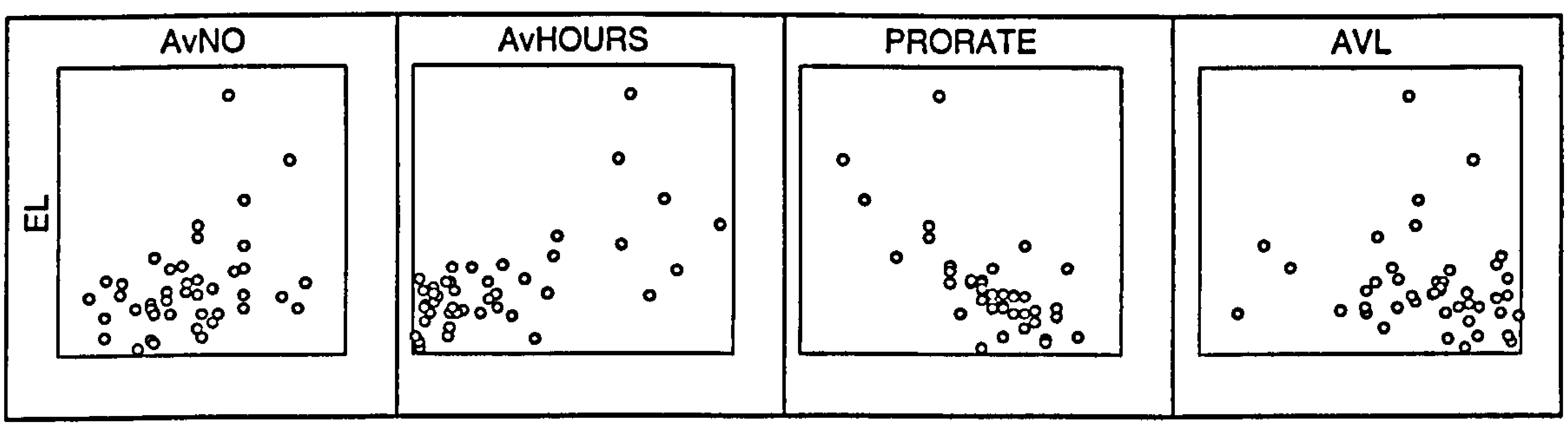
7.5.3.1.1 Data Exploring Process

(I) Data Screening

As a first step in analyzing the EL data of Kiln1 we start exploring the available data. This exploration includes:

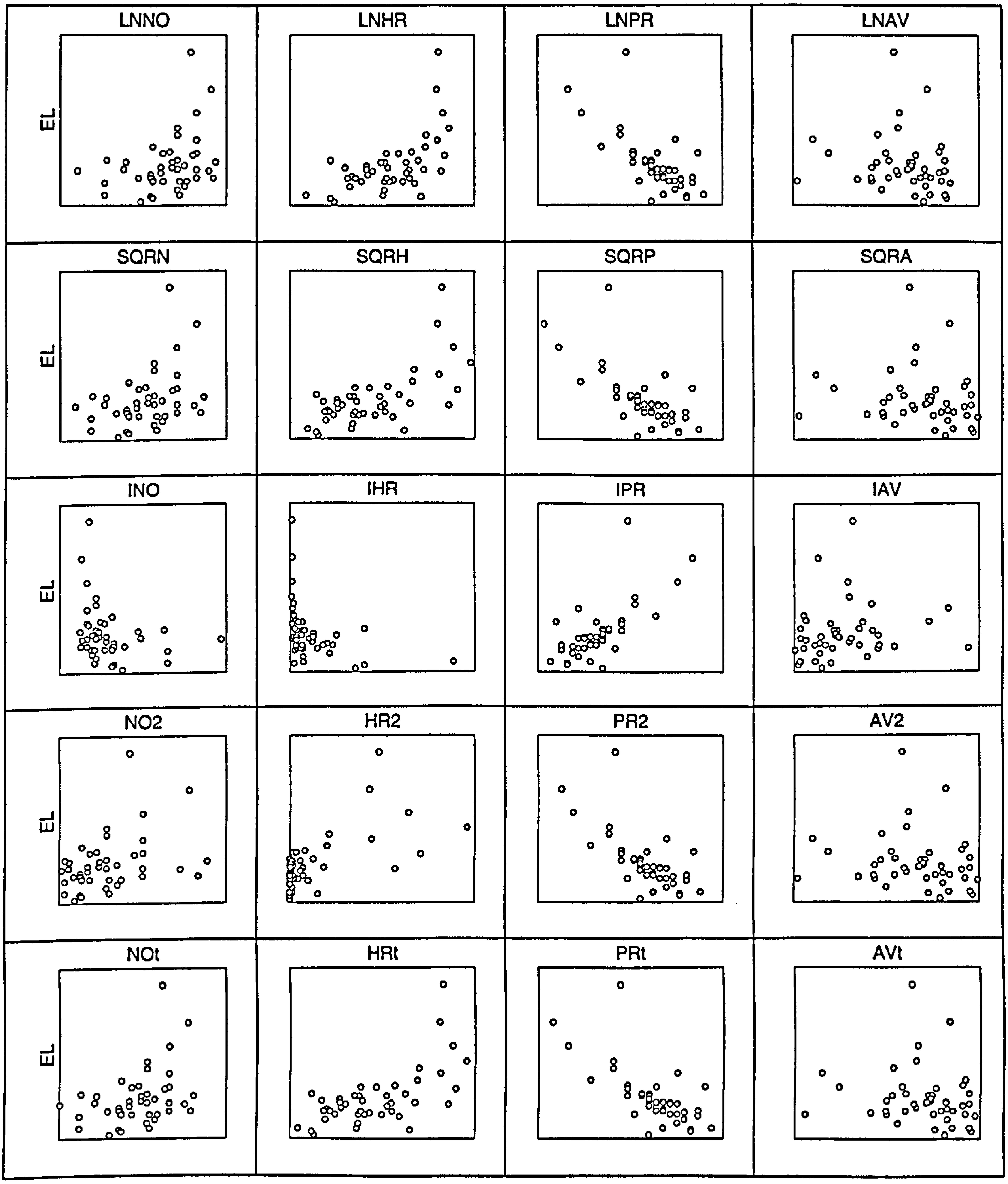
1. The *scatter plots* of the dependent variable EL against each of the independent variables (AvNO, AvHOURS, PRORATE, and AVL) are shown in table (7.9a). It seems reasonable from these plots that there may be a linear relationship between EL and each of AvNO, AvHOURS and PRORATE. Moreover, there is an indication of possible outliers.

Table (7.9a) Scatter Plots of EL vs. Independent Variables



One may try some ***transformation*** of the independent and/or the dependent variables hoping to get more clear cases of linear relationships. We have considered five transformations, namely
logarithm, square, square root, cubic root, and inverse of each of the independent variables. If X denotes the name of a variable, then the names of these transformations are LNX=Ln(X), X2=the square of X, SQX=Square root of X, Xt=the cubic root of X, and IX=1/X. The following is a continuation of Table (7.9a) that provides the scatter plots of the transformed data.

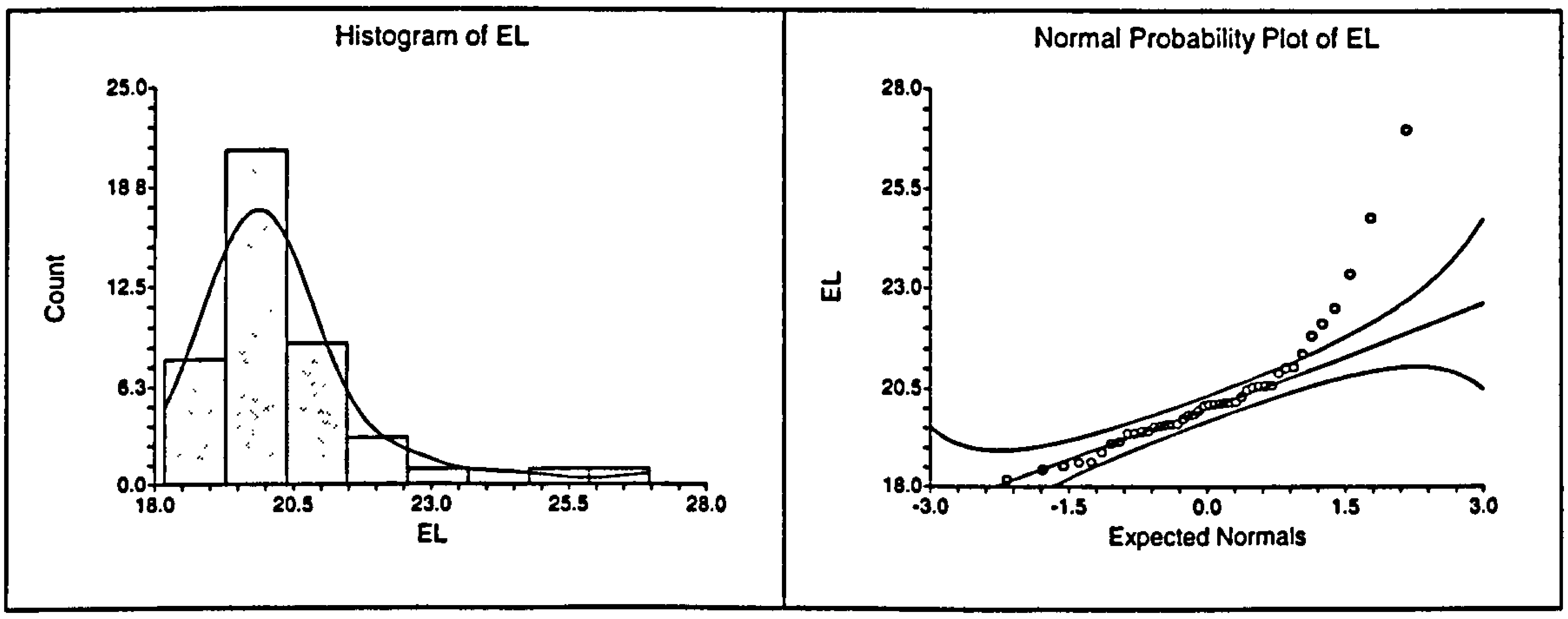
<u>**Continuation of Table (7.9a) Scatter Plots of EL vs. Transformed Independent Variables**</u>



It seems reasonable that these transformations do not improve the situation since they have not pointed out any further transformed independent variables that have linear relationship with the dependent variables. Moreover these plots still indicates some outliers in the data.

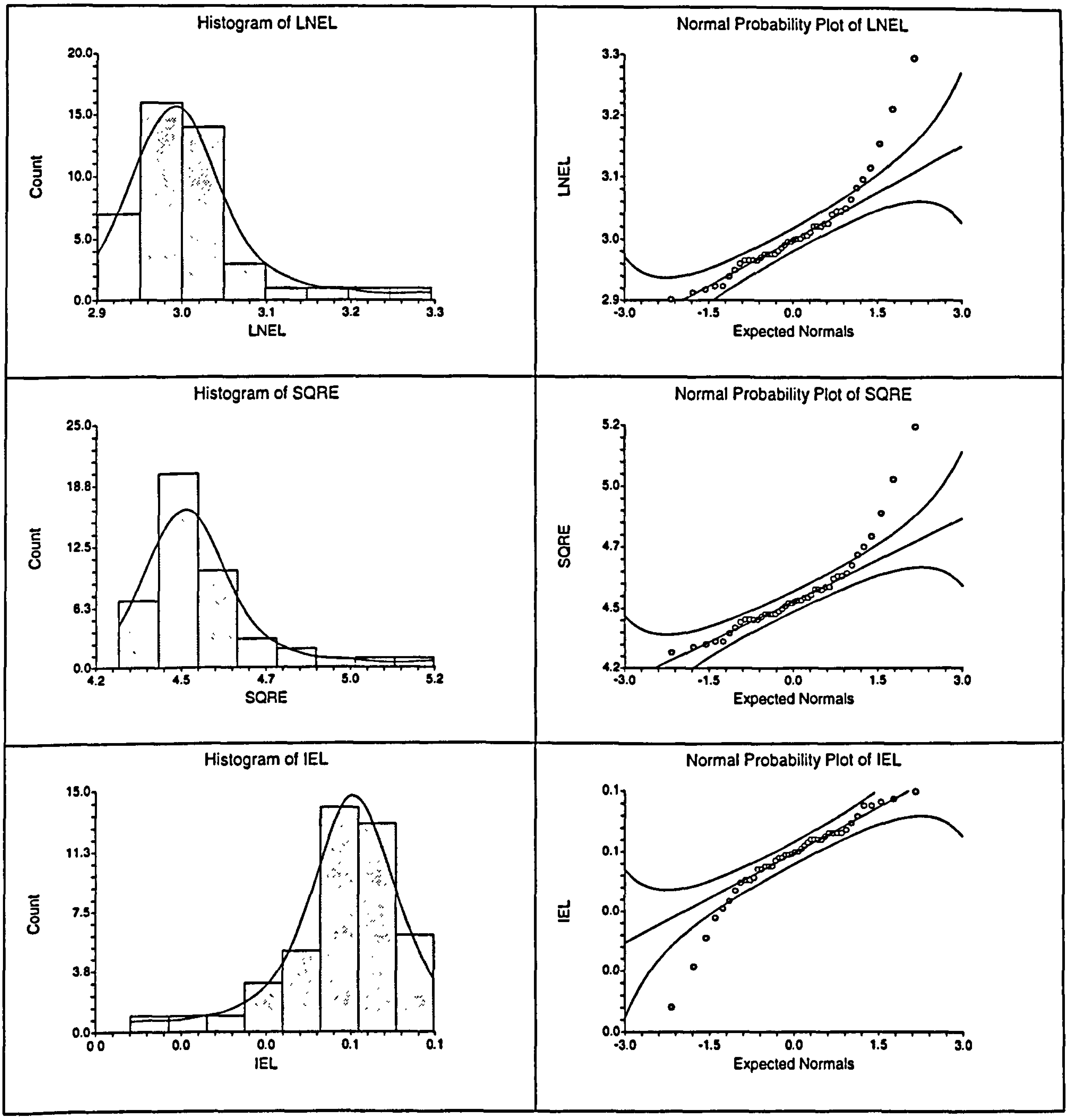
2. The ***histogram and the normal probability plot*** of EL are given in table (7.9b). It seems reasonable from these plots that there are some outliers in the data. These outliers may disturb the normality of the distribution of EL because the distribution has a long tail to the right.

Table (7.9b) Histogram and Normal Probability Plot of EL



3. ***Normality tests*** rejected the normality of EL. This is clear from the table (7.9c). Therefore, we have tried the transformed data. A continuation of Table (7.9b) is given bellow, which provides histograms, and normal probability plots of the transformed data. The situation is still the same, i.e. normality is not acceptable.

Continuation of Table (7.9b) Histogram and Normal Probability Plot of EL and its Transformations



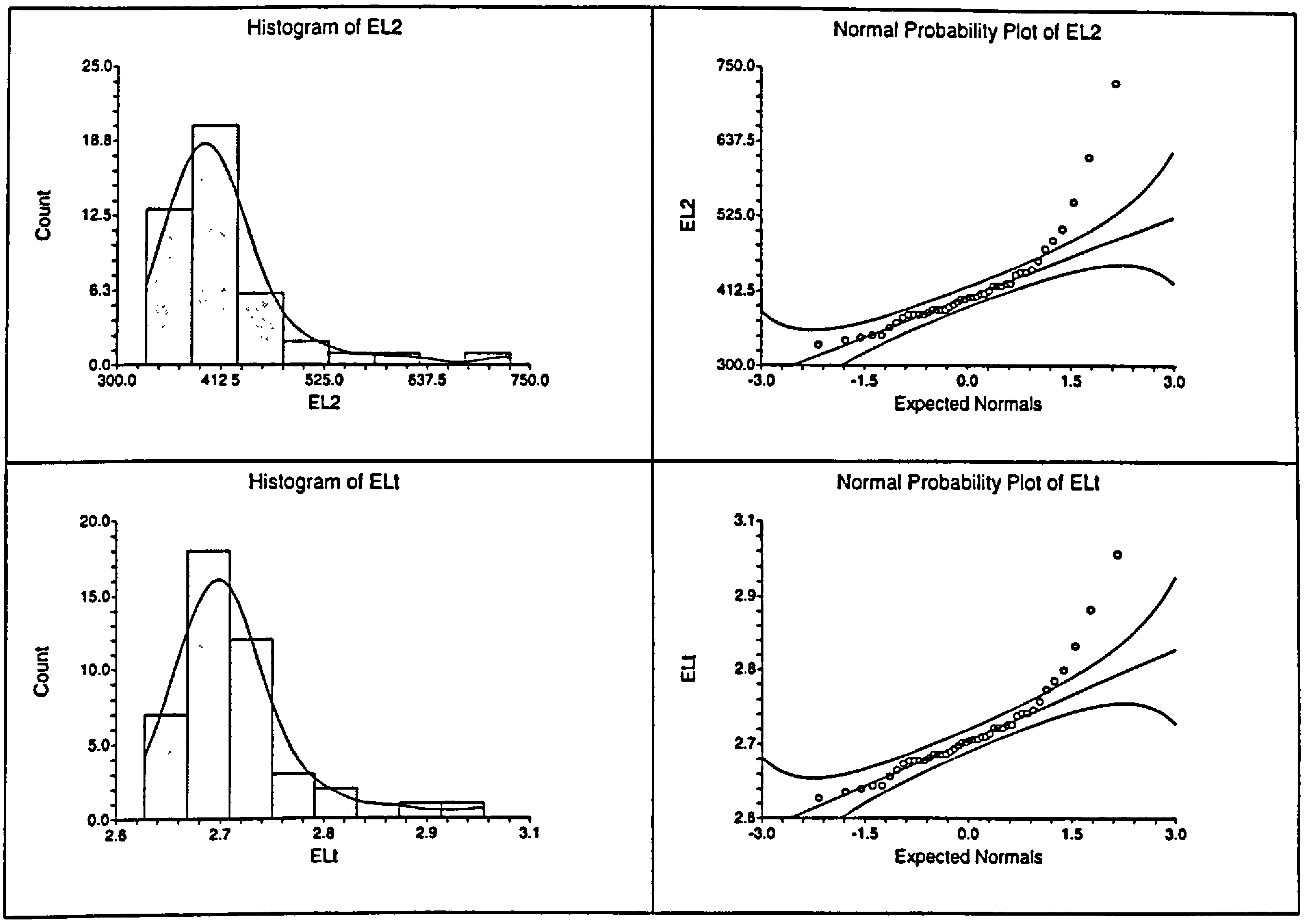


Table (7.9 c) represents the results of three normality tests (Skewness test, Kurtosis test and Omnibus test) for each of the five suggested transformation of the dependent variables (Logarithm, Square root, Inverse, Square and cube of Electricity). The p-values of each of these three tests indicate that the transformed variables are not normally distributed.

The reason for that may be the existence of the outliers at the tail of the distribution of the transformed variables as it can be seen from table (7.9b). So, one should use a regression procedure that treats the problem of the existence of outliers and the non-normality assumption.

Table (7.9c) Normality Tests of Transformations of EL

Test	Prob	10% Critical	5%	Critical	Decision	Normality Tests	Section		
		--- Skewness Test ----		Kurtosis Test - --			Omnibus Test -		Variable
Variable	Value	Z	Prob	Value	Z	Prob	K2	Prob	Normal?
EL	2.03	4.48	0.0000	8.25	3.63	0.0003	33.21	0.0000	No
LNEL	1.69	3.98	0.0001	6.80	3.18	0.0015	25.93	0.0000	No
SQRE	1.86	4.24	0.0000	7.49	3.41	0.0007	29.55	0.0000	No
IEL	-1.37	-3.43	0.0006	5.63	2.69	0.0072	18.99	0.0001	No
EL2	2.38	4.93	0.0000	9.97	4.02	0.0001	40.51	0.0000	No
ELt	1.80	4.15	0.0000	7.25	3.33	0.0009	28.33	0.0000	No

4. A search for the *outliers* is performed using an exploration procedure from NCSS2000 statistical package. Table (7.9d) reports in the first column the position numbers of the observations in the data set, the second and third columns show the test statistics and their probabilities and the fourth column indicates the positions of the outlying observations of the data. This table shows that there are only two outliers in the data, namely 23rd and 24th observations. Comparing this result with the normal probability plot and the histogram one concludes that the observations outside the bands in that plot indicated something else other than the outliers. They indicate that the distribution of EL has a heavy tail to the left, and a long tail to the right.

Table (7.9d) Outliers in EL

	T2	T2	
Row	Value	Prob	Outlier?
1	0.21	0.6492	
2	0.01	0.9254	
3	0.40	0.5296	
4	1.69	0.2002	
5	0.02	0.8777	

6	1.25	0.2693	
7	1.12	0.2958	
8	0.77	0.3857	
9	1.39	0.2446	
10	1.12	0.2958	
11	0.78	0.3823	
12	3.43	0.0707	
13	0.33	0.5696	
14	0.16	0.6930	
15	0.33	0.5696	
16	0.57	0.4543	
17	0.04	0.8353	
18	0.26	0.6110	
19	0.15	0.6976	
20	0.02	0.8777	
21	0.01	0.9304	
22	1.71	0.1980	
23	7.30	0.0098	Yes
24	16.30	0.0002	Yes
25	0.11	0.7379	
26	0.02	0.8827	
27	0.01	0.9304	
28	0.03	0.8684	
29	0.33	0.5696	
30	0.20	0.6537	

31	0.20	0.6537
32	0.01	0.9254
33	0.01	0.9254
34	0.00	0.9785
35	1.13	0.2930
36	0.41	0.5255
37	0.11	0.7427
38	0.16	0.6930
39	0.20	0.6537
40	0.07	0.7886
41	0.33	0.5696
42	0.20	0.6537
43	0.04	0.8353
44	0.02	0.8827

5. To check the validity of our observation from the scatter plot, we calculate the ***correlation coefficient*** between EL and each of the independent variables. Table (7.9e) provides these correlations together with their p-values. It seems reasonable from these results that, at the 0.05 level of significance, there is a linear relationship between EL and each of AvNO, AvHOURS, and PRORATE since their p-values are less than 0.05. This result confirms the observation from the scatter plots.

Table (7.9e): Correlations and p-values Between EL and Independent Variables

	AvNO	AvHOURS	PRORATE	AVL
EL	0.317961	0.763674	-0.400497	-0.096994
	0.035437	0.000000	0.007062	0.531091

6. To explore the possibility of ***multicollinearity***, we reported the correlations between the independent variables, which are given in table (7.9f). As it was pointed out in the previous section, mulicollinearity may occur if there is a very high correlation between some of the independent variables. The threshold, which is recommended here, is 0.75. It seems reasonable that all the obtained correlations are less than 0.75. So, multicollinearity may not be a problem.

Table (7.9f): Correlations and p-values Between Independent Variables

	AvNO	AvHOURS	PRORATE	AVL
AvNO	1.000000	0.057085	-0.064054	-0.325681
	0.000000	0.712828	0.679549	0.030980
AvHOURS	0.057085	1.000000	-0.171846	0.033761
	0.712828	0.000000	0.264678	0.827771
PRORATE	-0.064054	-0.171846	1.000000	-0.113734
	0.679549	0.264678	0.000000	0.462279
AVL	-0.325681	0.033761	-0.113734	1.000000
	0.030980	0.827771	0.462279	0.000000

(II) Variable Selection Process

7. Based on the above exploration we raise the following question: *Which variables should be included in the regression models to be fitted to the given data?* To answer this question, the following steps were implemented. It should be mentioned here that we are not looking for the fitted model at this stage of the analysis because this will be done latter, but we are searching for the best independent variables to be included in the final regression model.

As a first selection procedure, we applied the ***All-Possible Regression Procedure***. The results of this procedure are given in table (7.9g). Based on the Cp criterion, it seems reasonable that the best model is that with the independent variables AvNO, AvHOURS, PRORATE. This result confirms the result obtained from the correlation matrix.

Table (7.9g): All Possible Results Section

Model Size	R-Squared	Root MSE	Cp	Model
1	0.583198	1.070914	20.317491	B (AvHOURS)
1	0.160397	1.51994	81.503081	C (PRORATE)
1	0.101099	1.572699	90.084390	A (AvNO)
1	0.009408	1.650962	103.353549	D (AVL)
Variables in Best Model AvHOURS				
2	0.658721	0.9807921	11.388138	AB
2	0.657906	0.9819623	11.506063	BC
2	0.598289	1.064092	20.133561	BD
2	0.246193	1.457647	71.087110	AC
2	0.180983	1.519389	80.524114	CD
2	0.101148	1.59172	92.077423	AD
Variables in Best Model AvNO, AvHOURS				
3	0.725577	0.8904189	3.713087	ABC
3	0.681541	0.9592037	10.085734	BCD
3	0.659914	0.9912402	13.215586	ABD

3	0.248487	1.473509	72.755238	ACD

Variables in Best Model
AvNO, AvHOURS, PRORATE

4	0.730505	0.8936297	5.000000	ABCD

Variables in Best Model
AvNO, AvHOURS, PRORATE, AVL

8. As another selection procedure we have applied the four criteria in the ***Stepwise Regression***, and obtained the results in table (7.9h). It seems reasonable from all these results that the selected variables to be included in the model are AvNO, AvHOURS, and PRORATE. They are the same as those obtained in the above discussions.

Table (7.9 h): Stepwise Regression Report

a) Forward

Iter. No.	Max R-Squared Action	Variable	R-Squared	Sqrt(MSE)	Other X's
0	Unchanged		0.000000	1.639382	0.000000
1	Added	AvHOURS	0.583198	1.070914	0.000000
2	Added	AvNO	0.658721	0.9807921	0.003259
3	Added	PRORATE	0.725577	0.8904189	0.032483
4	Unchanged		0.725577	0.8904189	0.032483

List of Variables Selected
AvNO, AvHOURS, PRORATE

b) Backward

Iter. No.	Max R-Squared Action	Variable	R-Squared	Sqrt(MSE)	OtherX's
0	Unchanged		0.730505	0.8936297	0.125165
1	Removed	AVL	0.725577	0.8904189	0.032483
2	Unchanged		0.725577	0.8904189	0.032483

List of Variables Selected
AvNO, AvHOURS, PRORATE

c) Stepwise

Iter. No.	Max R-Squared Action	Variable	R-Squared	Sqrt(MSE)	Other X's
0	Unchanged		0.000000	1.639382	0.000000
1	Added	AvHOURS	0.583198	1.070914	0.000000
2	Unchanged		0.583198	1.070914	0.000000
3	Added	AvNO	0.658721	0.9807921	0.003259
4	Unchanged		0.658721	0.9807921	0.003259
5	Added	PRORATE	0.725577	0.8904189	0.032483
6	Unchanged		0.725577	0.8904189	0.032483

List of Variables Selected
AvNO, AvHOURS, PRORATE

d) Min MSE

Iter. No.	Max R-Squared Action	Variable	R-Squared	Sqrt(MSE)	Other X's
0	Unchanged		0.000000	1.639382	0.000000
1	Added	AvHOURS	0.583198	1.070914	0.000000
2	Added	AvNO	0.658721	0.9807921	0.003259
3	Added	PRORATE	0.725577	0.8904189	0.032483
4	Unchanged		0.725577	0.8904189	0.032483

List of Variables Selected
AvNO, AvHOURS, PRORATE

9. A third selection procedure depends on the *t-test* of the coefficients of the full multiple linear regression model. Table (7.9i) provides the computer output of the Multiple Regression procedure. It seems reasonable from this output that the estimated coefficients of the same three variables (AvNO, AvHOURS, and PRORATE) are significant and hence they should be included in the model but AVL should be removed from the final model. This is clear from the p-values corresponding to the tests of the coefficients. This result confirms the previous discussion.

Table (7.9 i): t-tests for Coefficients of Multiple Regression Model

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	28.78952	4.143018	6.9489	0.000000	Reject Ho	0.999999
AvNO	3.605562	1.354505	2.6619	0.011227	Reject Ho	0.737584
AvHOURS	0.2005477	2.401205E-02	8.3520	0.000000	Reject Ho	1.000000
PRORATE	-6.698804E-02	2.095876E-02	-3.1962	0.002760	Reject Ho	0.876280
AVL	-2.482429E-02	2.939716E-02	-0.8444	0.403571	Accept Ho	0.130629
R-Squared	0.730505					

Model
EL= 28.78952+ 3.605562*AvNO+ .2005477*AvHOURS-6.698804E-02*PRORATE-2.482429E-02*AVL

(III) Exploring the Residuals

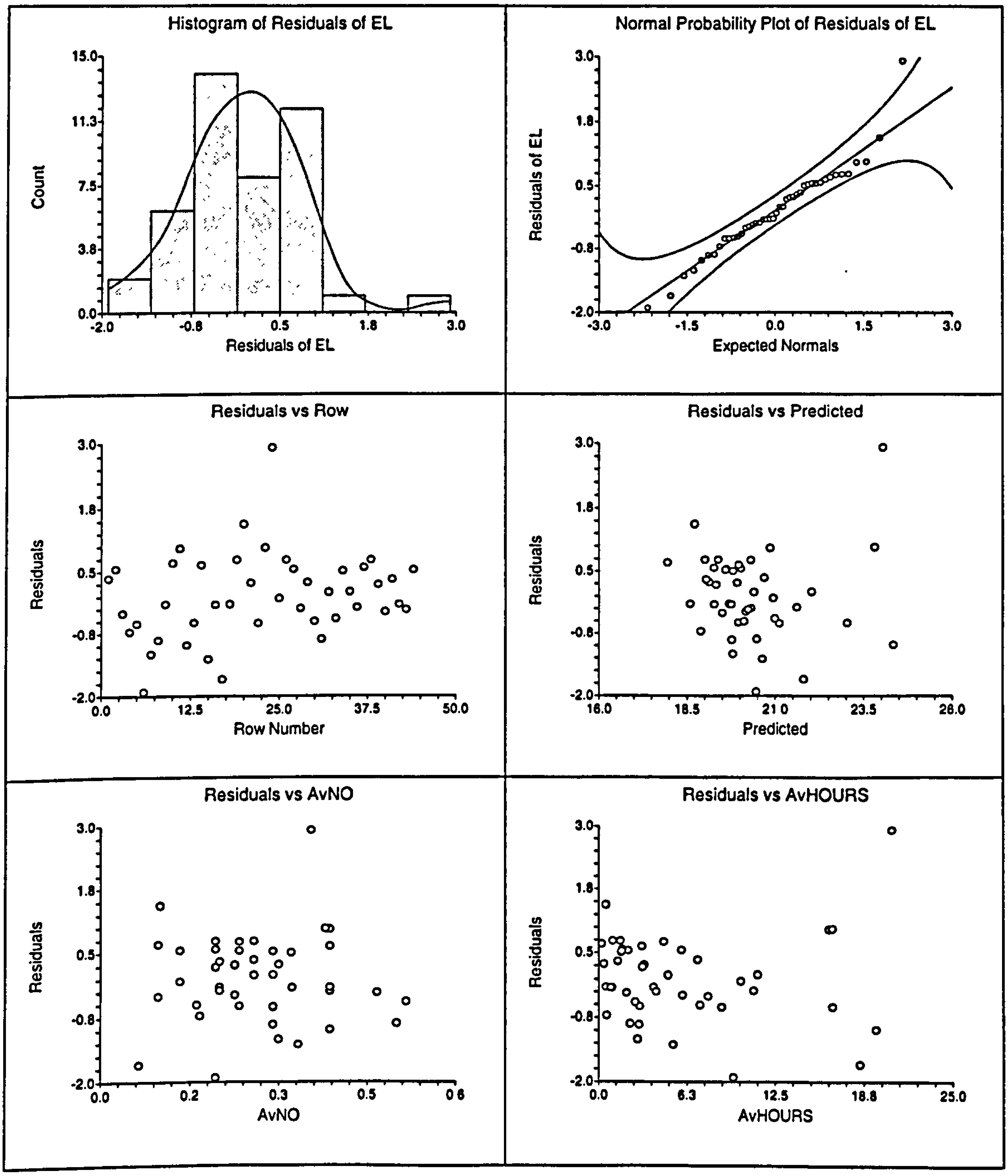
10) The next step is to do an *error analysis of the residuals* of the above full multiple regression model. This analysis includes a plot section, which reports the histogram, and the normal probability plot of the residuals. It also gives the scatter plots of residuals against each of row number, predicted values, and each of the independent variables. These plots may be used to explore the normality, the constant variance, the zero means, the outliers, and the randomness assumptions of the errors. It seems reasonable from the plots given in table (7.9j) that

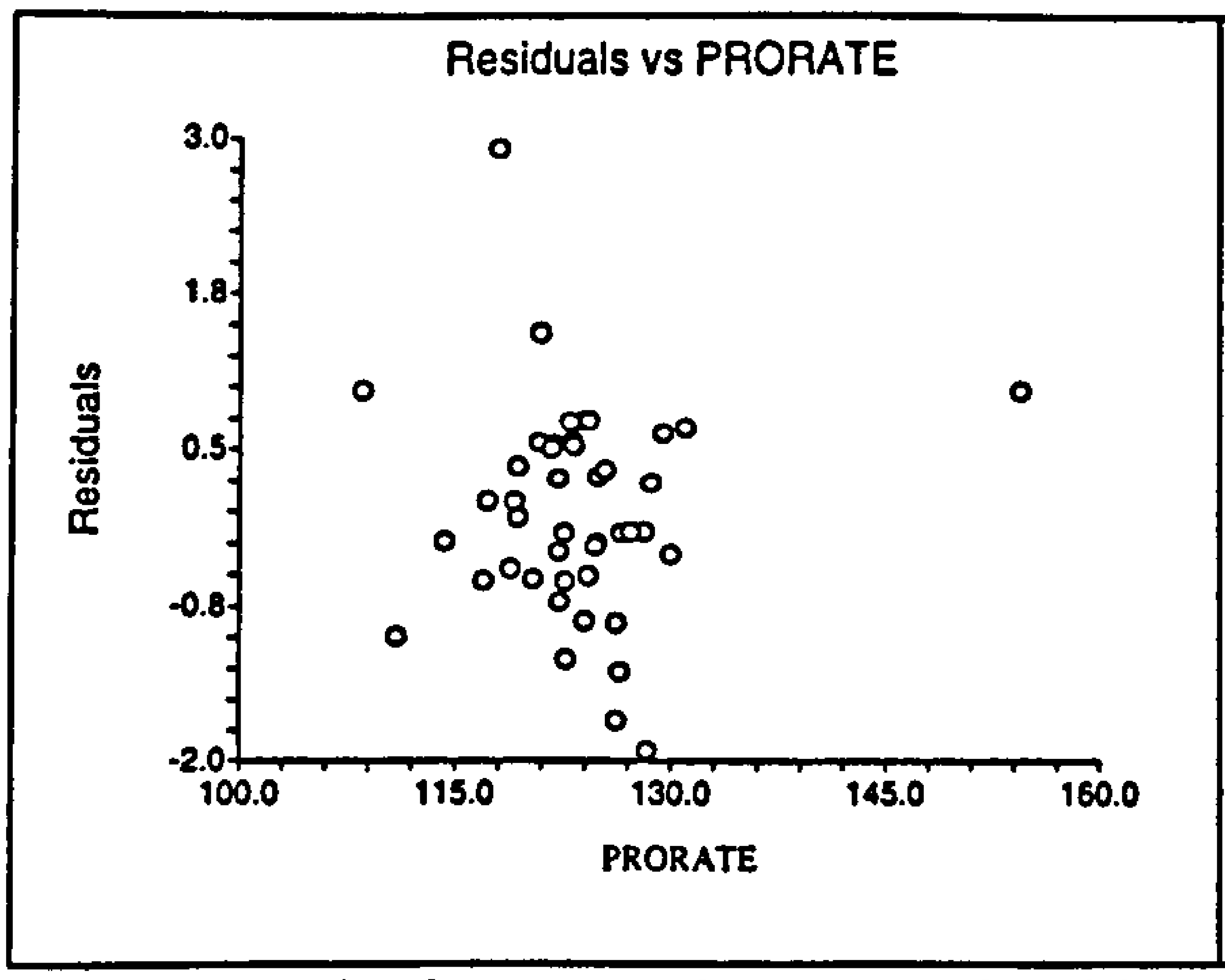
a) Since all points except one fall inside the confidence band of the normal probability plot we conclude that the errors are almost normally distributed. The histogram also shows the existence of some outliers and that the distribution of the errors is symmetric but it does not look like a bell shaped.

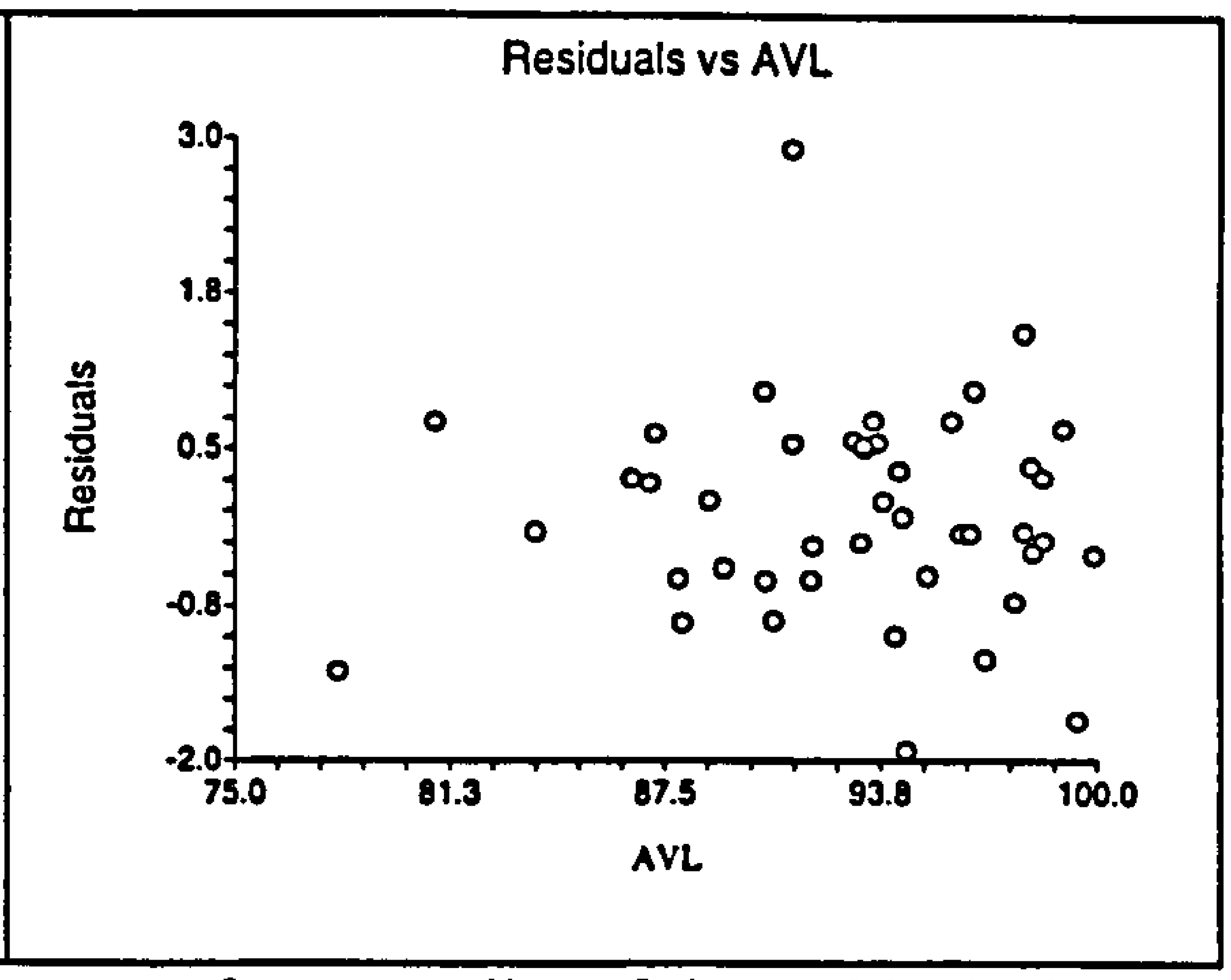
b) The scatter plot of the residuals against their row numbers does not show any pattern, which means that the residuals are not time dependent, i.e. they are random errors and they are independent. Moreover, this plot shows that the residuals are evenly distributed around zero. It also spots an outlier.

c) The scatter plots of the residuals against the predicted values and the independent variables show the same results that the residuals have zero mean and constant variance since they are distributed in a rectangle with no pattern. They also spot an outlier.

Table (7.9 j): Plot of Residuals







As a quantitative procedure, one should run a test for normality of the residuals. Table (7.9k) is the output of the normality section. It seems reasonable from this output that the skewness test accepts normality, which agrees partially with the above observations from the normal probability plot and the histogram of the residuals. However, the kurtosis test rejects normality, this is due to a problem in the shape of the histogram of the data that is the middle part of the histogram is lower than the adjacent parts. Moreover the omnibus test rejects normality since it is a function of the skewness and kurtosis. So, to be on the safe side we say that we are not sure that the residuals are almost normally distributed.

<u>**Table (7.9 k): Normality Tests of Residuals**</u>

Assumption	Value	Probability	Decision(5%)
Skewness	1.5051	0.132311	Accepted
Kurtosis	2.3099	0.020896	Rejected
Omnibus	7.6006	0.022364	Rejected

11) To have a quantitative test of the *randomness (independence) of the residuals*, we run the serial correlation (autocorrelation) procedure, which reports the serial correlations between the residuals together with the Durbin-Watson test. Table

(7.9l) provides the computer output of this section. It seems reasonable from this output that all the absolute values of the serial correlations are less than the critical value 0.301511, which means that they are independent. The same result is obtained based on the Durbin-Watson Value, which is 1.4401. this result agrees with our observation for the scatter plots of the residual.

Table (7.9 l): Serial-Correlation of Residuals

Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.273231	9	0.072060	17	-0.167306
2	0.073929	10	0.166710	18	-0.064823
3	0.101218	11	-0.035819	19	-0.101486
4	0.000211	12	-0.111026	20	-0.163901
5	0.012446	13	0.089410	21	-0.077368
6	-0.114798	14	0.055811	22	-0.022804
7	-0.275749	15	-0.012384	23	0.035184
8	-0.061765	16	-0.123682	24	0.089482

Above serial correlations significant if their absolute values are greater than 0.301511

Durbin-Watson Value 1.4401

(IV) Checking Multicollinearity

12) The final screening process is to explore the ***multicollinearity*** in a quantitative way. It was observed on a previous paragraph that multicollinearity might not be a problem based on correlations between independent variables. Table (7.9m) reports the computer out put of the multicollinearity section of the fitted model. It seems reasonable from this output that since the R-squared vs. other X's which is the R-

squared of regressing each independent variable on the remaining other independent variables and the maximum R-squared value is 0.13 which is too low, multicollinearity should not be a problem.

a) The maximum value of the variance inflation factor is 1.14, which is much less than 10, and then multicollinearity could not be a problem.

b) The same conclusion is obtained from eigenvalues and condition numbers.

Table (7.9 m): Multicollinearity Problem

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance	Diagonal of X'X Inverse
AvNO	1.135189	0.119090	0.880910	2.29745
AvHOURS	1.033768	0.032665	0.967335	7.220103E-04
PRORATE	1.052887	0.050230	0.949770	5.500677E-04
AVL	1.143073	0.125165	0.874835	1.08217E-03

Eigenvalues of Centered Correlations

No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.329820	33.25	33.25	1.00
2	1.204565	30.11	63.36	1.10
3	0.834443	20.86	84.22	1.59
4	0.631171	15.78	100.00	2.11

All Condition Numbers less than 100. Multicollinearity is NOT a problem.

7.5.3.1.2 Regression Analysis

According to the above screening work, it seems reasonable to fit a regression model to the available data. So, the question is which is the most suitable procedure to fit that model. In this section we will use multiple regression, ridge regression, robust regression. Moreover, we will fit some non-linear regression models. Then we select the model with highest R-squared value. It is interesting to note that we are still screening the data to reach the best regression procedure, which is not affected, be the previously discovered problems in the data and to reach a final model which includes only the significant independent variables. It is true that we have observed that multicollinearity may not be a problem and so there is no need for ridge regression. Moreover, some independent variables do not have significant contributions to the regression model, and we are not sure about the normality assumption. But, for the sake of comparison we will apply both ridge regression and multiple regression based on all available independent variables then compare their results. If they are almost the same then this will confirm our previous observation about multicollinearity. At a latter stage we will consider only the significant variable to obtain the final model.

<u>1) Multiple Regression</u>

In the above screening process we reported a part of the output of the multiple regression procedure. *Here we concentrate on both the full model and the model including only the three influential independent variables that were selected in the screening process.* Table (7.10a) and (7.10b) report the computer output of this procedure. It seems reasonable from this output that

a) The R-squared value for the full model is 0.730505 while that of the model containing the three selected independent variables is 0.725577. These two values are almost the same and they indicate that the model explains about 73% of the variability in EL using the three independent variables AvNO, AvHOURS, and PRORATE. This is a reasonable proportion. Moreover, AVL has no influence on EL.

b) The signs of the estimated coefficients agree with the expectations of the researcher, i.e. both AvNO, and AvHOURS have an increasing relationship with EL but AVL has a decreasing one.

c) The F-test in the ANOVA section indicates that the fitted model is significant.

d) The PRESS R-squared value is 0.5421, which means that this model can predict about 54% of the variability in EL. This percentage seems to be low for good prediction but the usual R-squared value for the model is 0.730505. The difference between the Press R-squared and the multiple R-squared may be due to outliers.

e) All the assumptions on the model are met except possibly the existence of outliers and a problem in normality of the errors.

Table (7.10 a): Regression Equation Section of *Full Model*

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	28.78952	4.143018	6.9489	0.000000	Reject Ho	0.999999
AvNO	3.605562	1.354505	2.6619	0.011227	Reject Ho	0.737584
AvHOURS	0.2005477	0.02401205	8.3520	0.000000	Reject Ho	1.000000
PRORATE	-0.06698804	0.02095876	-3.1962	0.002760	Reject Ho	0.876280
AVL	-0.02482429	0.02939716	-0.8444	0.403571	Accept Ho	0.130629
R-Squared	0.730505					

Model

EL= 28.78952+ 3.605562*AvNO+ .2005477*AvHOURS-6.698804E-02*PRORATE-2.482429E-02*AVL

Table (7.10 b): Regression Equation Section of the *Three Selected Variables*

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	26.08888	2.624284	9.9413	0.000000	Reject Ho	1.000000
AvNO	3.99093	1.270731	3.1407	0.003166	Reject Ho	0.865303
AvHOURS	0.1998955	0.0239134	8.3591	0.000000	Reject Ho	1.000000
PRORATE	-0.0064591	-0.2206910	-3.1217	0.003334	Reject Ho	0.861252
R-Squared	0.725577					

Model

EL= 26.08888+ 3.99093*AvNO+ .1998955*AvHOURS-6.459102E-02*PRORATE

Regression Coefficient Section

Independent Variable	**Regression Coefficient**	**Standard Error**	**Lower 95% C.L.**	**Upper 95% C.L.**	**Standardized Coefficient**
Intercept	26.08888	2.624284	20.78501	31.39276	0.0000
AvNO	3.99093	1.270731	1.422687	6.559173	0.2610
AvHOURS	0.1998955	0.0239134	0.1515647	0.2482262	0.7036
PRORATE	-6.459102E-02	2.069104E-02	-0.1064092	-2.277287E-02	-0.2629
T-Critical	2.021075				

Analysis of Variance Section

Source	**DF**	**Sum of Squares**	**Mean Square**	**F-Ratio**	**Prob Level**	**Power (5%)**
Intercept	1	18178.63	18178.63			
Model	3	83.85179	27.9506	35.2535	0.000000	1.000000
Error	40	31.71384	0.7928459			
Total (Adjusted)	43	115.5656	2.687573			

Root Mean Square Error	**0.8904189**	**R-Squared**	**0.7256**
Mean of Dependent	**20.32611**	**Adj R-Squared**	**0.7050**
Coefficient of Variation	**4.380666E-02**	**Press Value**	**52.91578**
Sum |Press Residuals|	**34.66043**	**Press R-Squared**	**0.5421**

Normality Tests Section

Assumption	**Value**	**Probability**	**Decision(5%)**
Skewness	1.5083	0.131482	Accepted
Kurtosis	2.3156	0.020579	Rejected
Omnibus	7.6370	0.021961	Rejected

Serial-Correlation Section

Lag	**Correlation**	**Lag**	**Correlation**	**Lag**	**Correlation**
1	0.283537	9	0.056036	17	-0.181133
2	0.072522	10	0.142279	18	-0.091180
3	0.126111	11	0.008171	19	-0.083109
4	-0.002391	12	-0.101986	20	-0.170742
5	0.027351	13	0.089098	21	-0.097844
6	-0.083602	14	0.089763	22	-0.035138
7	-0.304161	15	-0.026073	23	0.031559
8	-0.080390	16	-0.127008	24	0.109477

Above serial correlations significant if their absolute values are greater than 0.301511

Durbin-Watson Value 1.4227

Multicollinearity Section

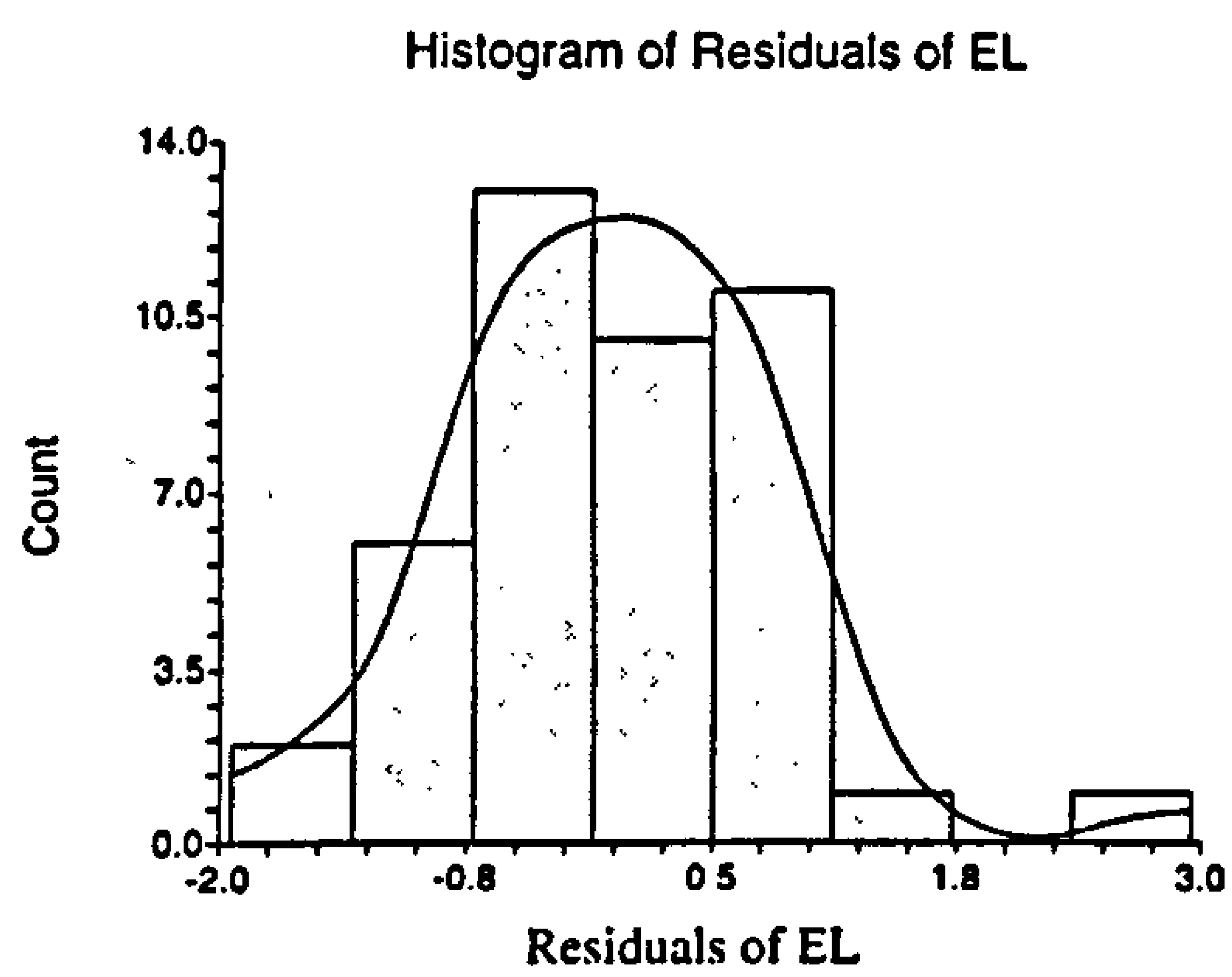
Independent Variable	**Variance Inflation**	**R-Squared Vs Other X's**	**Tolerance**	**Diagonal of X'X Inverse**
AvNO	1.006330	0.006291	0.993709	2.036659
AvHOURS	1.032698	0.031663	0.968337	7.212632E-04
PRORATE	1.033574	0.032483	0.967517	5.399778E-04

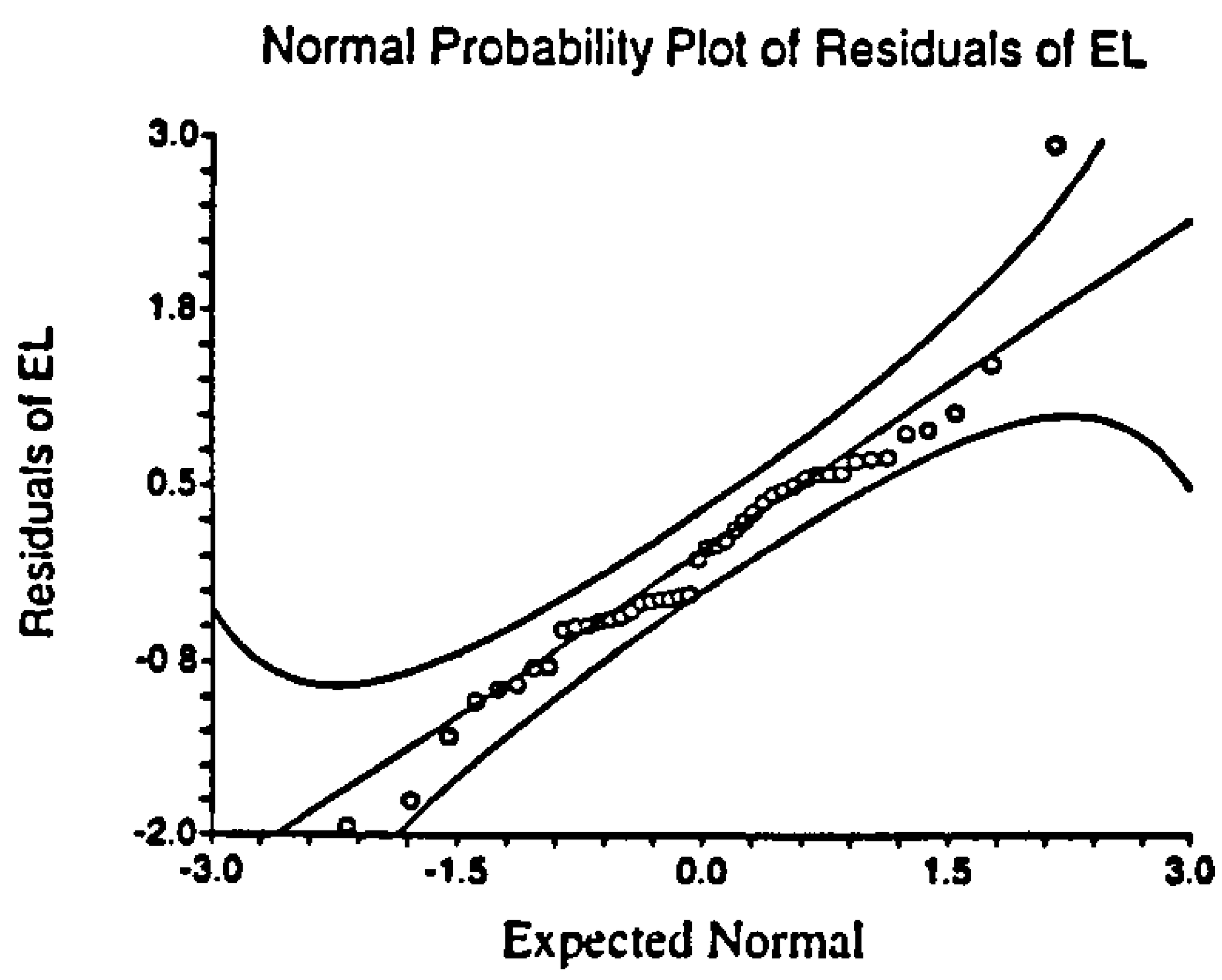
Eigenvalues of Centered Correlations

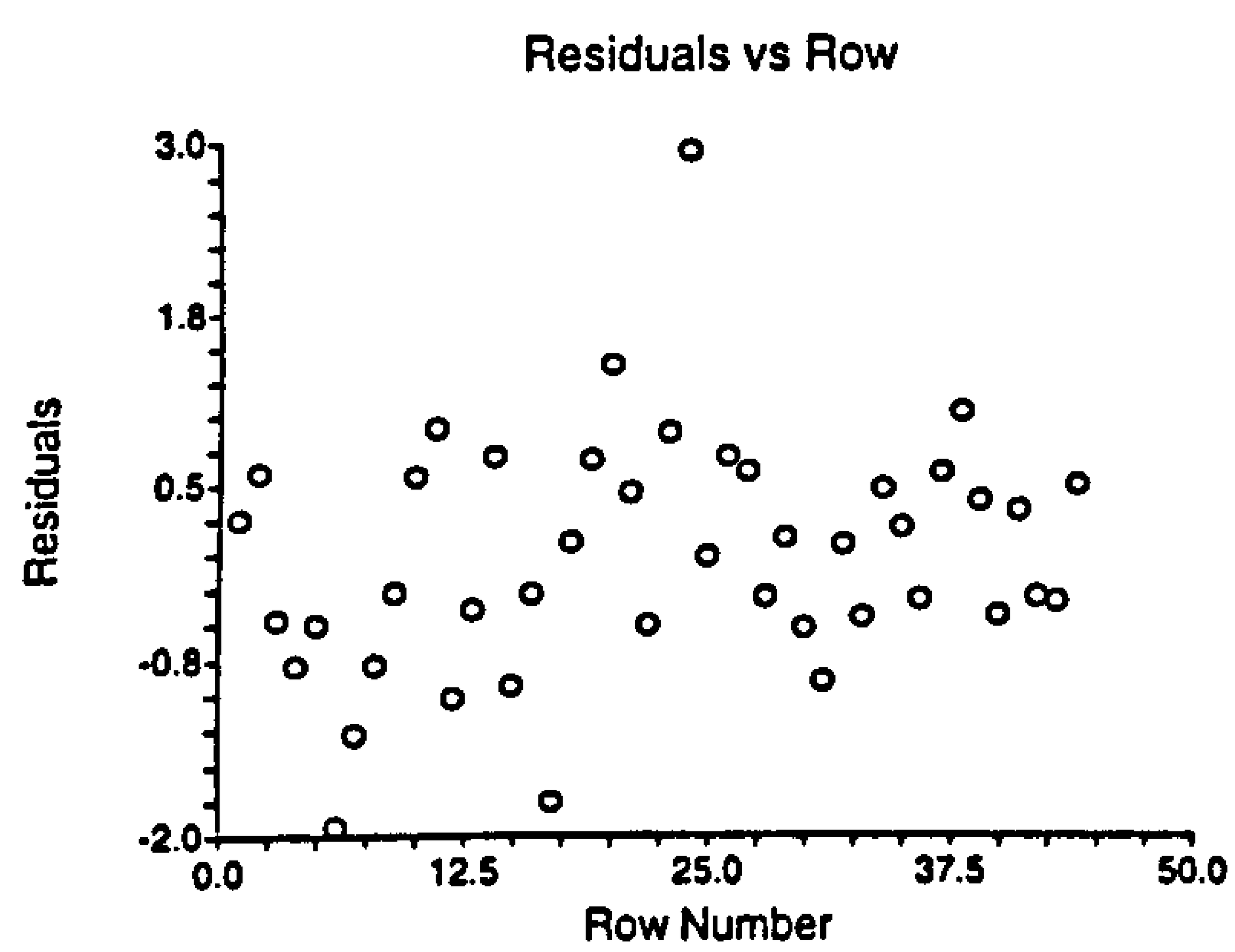
No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.207259	40.24	40.24	1.00
2	0.964748	32.16	72.40	1.25
3	0.827993	27.60	100.00	1.46

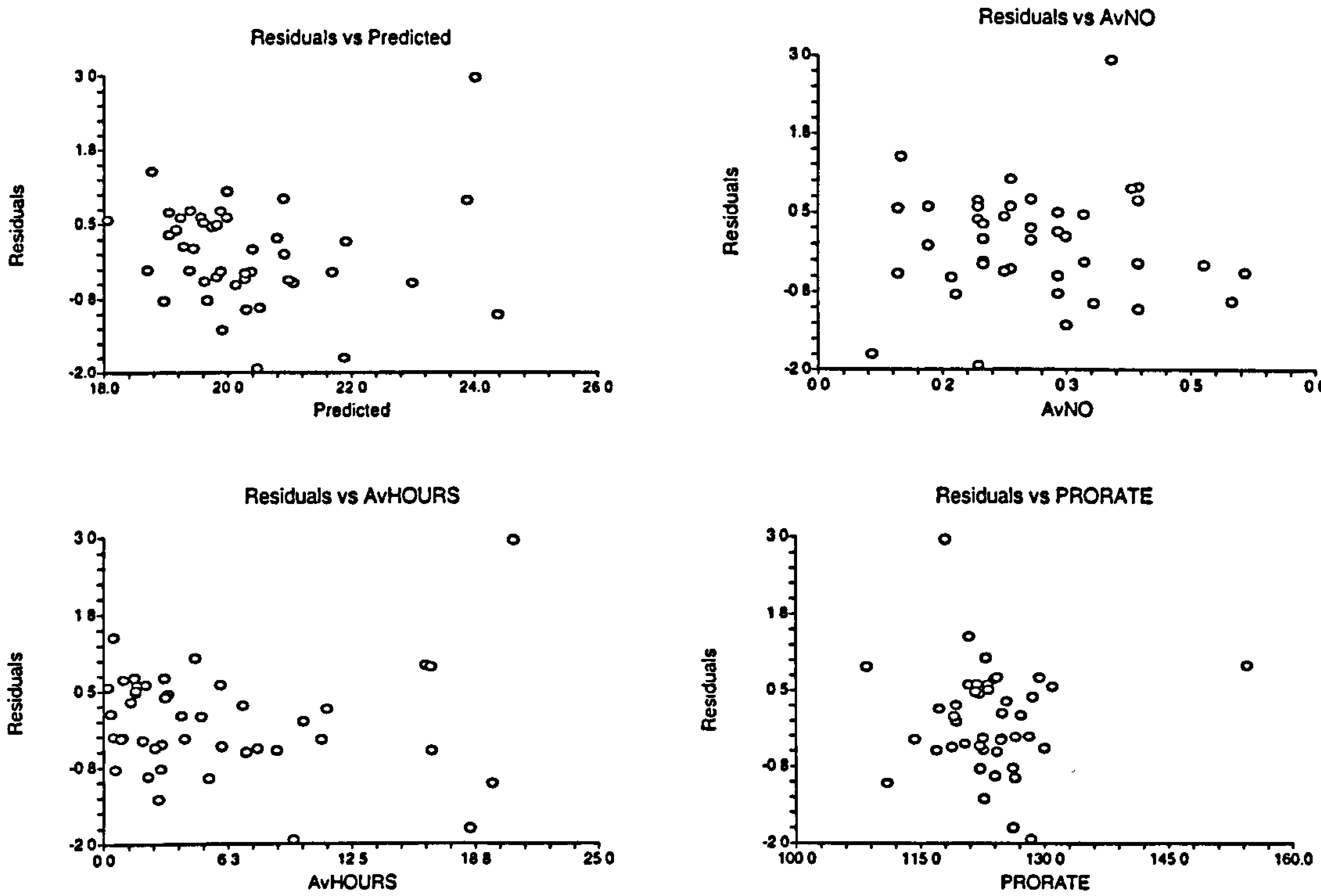
All Condition Numbers less than 100. Multicollinearity is NOT a problem.

Plots Section









2) Ridge Regression

The ridge regression is designed to fit a regression model if there is a suspicion that multicollinearity is a problem. Even multicollinearity may not be a problem; in this case, we apply ridge procedure to the full model and to the model of the three selected variables for the sake of comparison and to get another indicator about the multicollinearity problem. Tables (7.10c) and (7.10d) report the computer output of this procedure. It seems reasonable from this output that:

a) The R-squared value of the full model is 0.7273 while for the model with three independent variables it is 0.7224, which means that this model explains about 72% of the variability in EL.

b) The signs of the estimated coefficients agree with the expectations of the researcher, i.e. both AvNO, and AvHOURS have an increasing relationship with EL but AVL has a decreasing one.

c) The F-test in the ANOVA section indicates that the fitted model is significant.

d) All the assumptions on the model are met except possibly the existence of outliers. A minor problem in normality of the errors is seen in the full model but that is not the case in the model with three variables as it is observed from the histogram of the errors.

Table (7.10 c): Ridge Regression Report of *Full Model*

Least Squares Multicollinearity Section

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance
AvNO	1.1352	0.1191	0.8809
AvHOURS	1.0338	0.0327	0.9673
PRORATE	1.0529	0.0502	0.9498
AVL	1.1431	0.1252	0.8748

Since all VIF's are less than 10, multicollinearity is not a problem.

Eigenvalues of Correlations

No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.329820	33.25	33.25	1.00
2	1.204565	30.11	63.36	1.10
3	0.834443	20.86	84.22	1.59
4	0.631171	15.78	100.00	2.11

All Condition Numbers less than 100. Multicollinearity is NOT a problem.

Ridge Regression Coefficient Section for k = 0.005000

Independent Variable	**Regression Coefficient**	**Standard Error**	**Stand'zed Regression Coefficient**	**VIF**
Intercept	28.76593			
AvNO	3.592767	1.353768	0.2349	1.1208
AvHOURS	0.1995978	2.402415E-02	0.7026	1.0228
PRORATE	-6.679998E-02	2.096459E-02	-0.2719	1.0413
AVL	-2.472465E-02	2.937908E-02	-0.0747	1.1285

Model

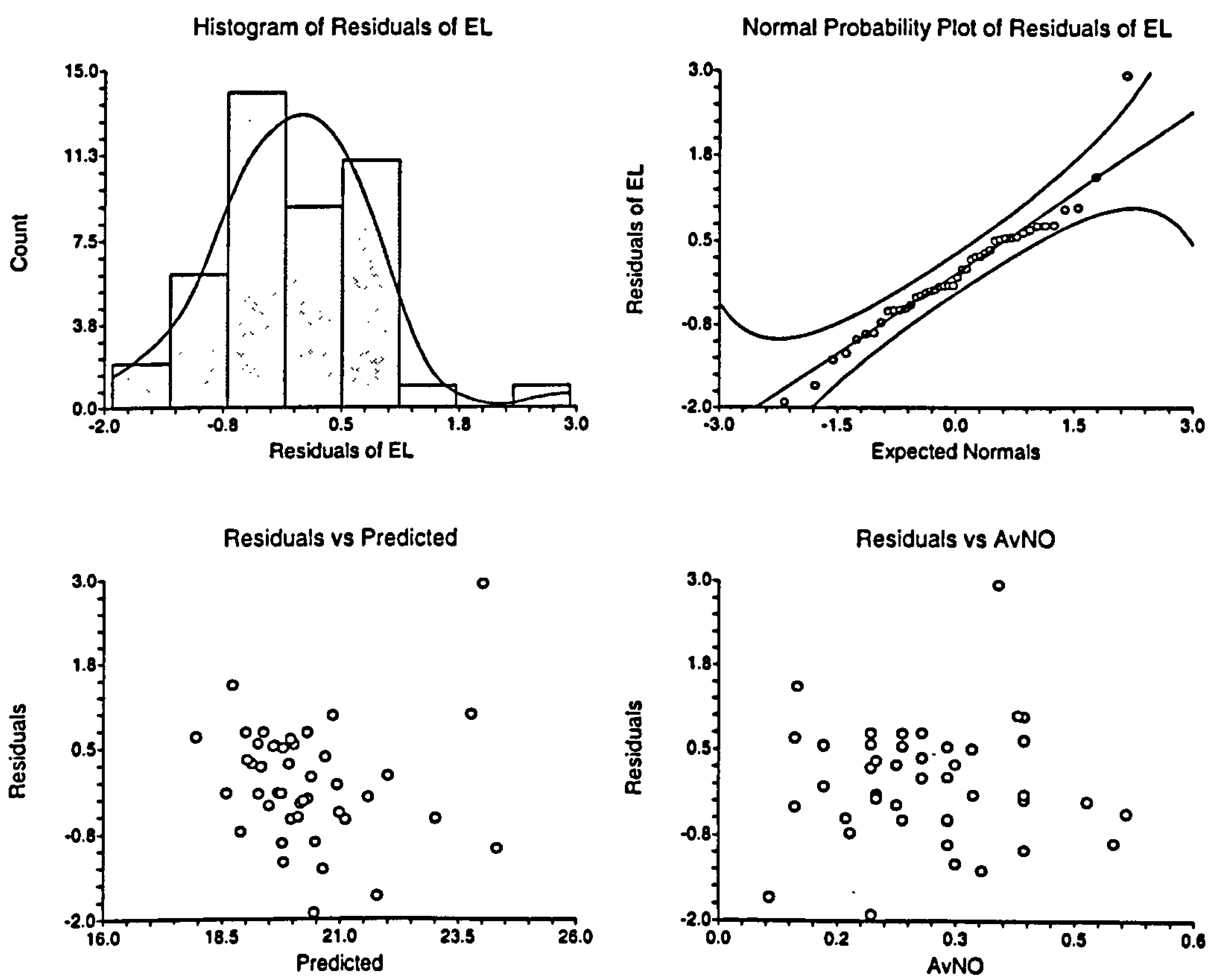
EI=28.76593+ 3.592767*AvNO+ .1995978*AvHOURS-6.679998E-02*PRORATE-2.472465E-02*AVL

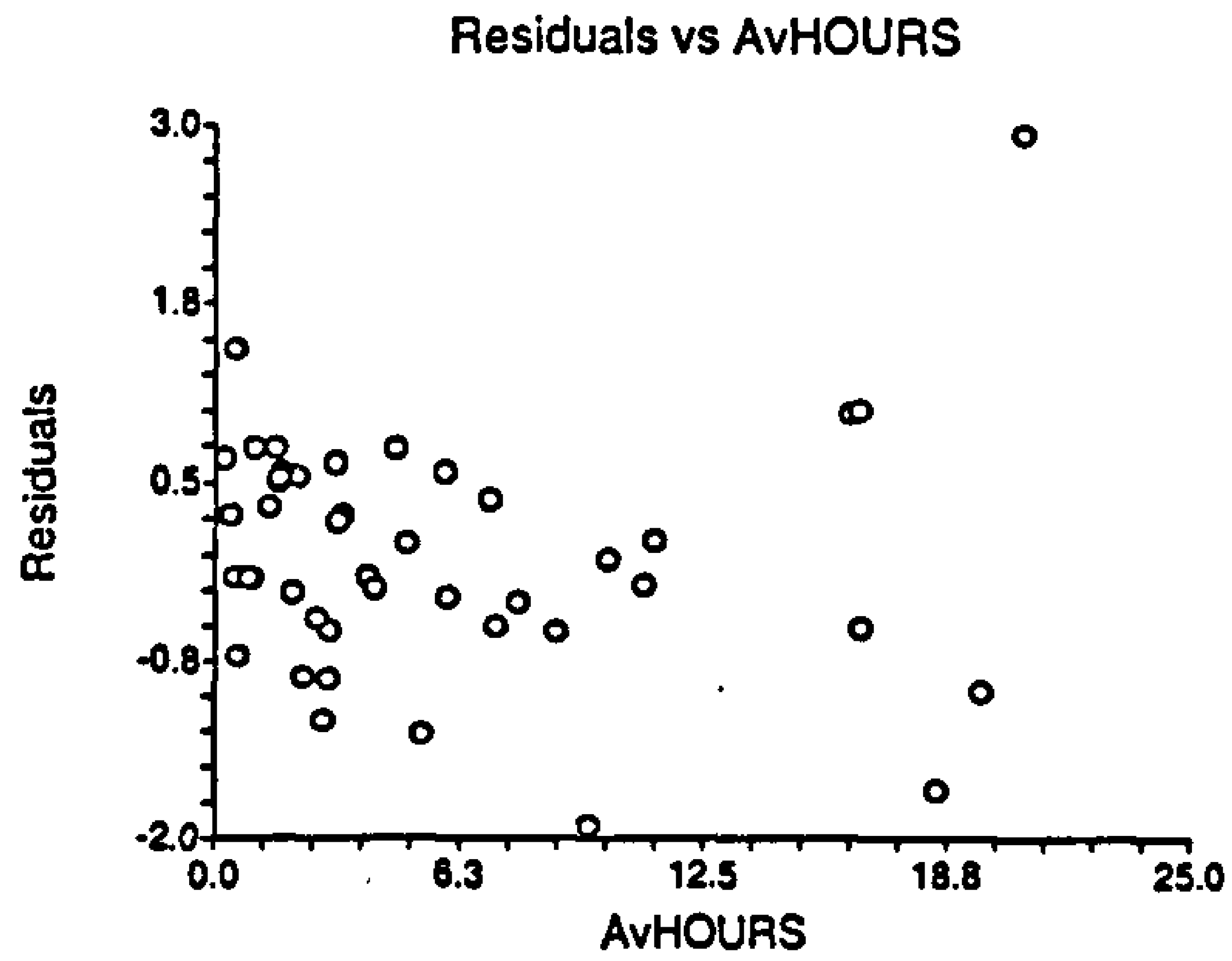
Analysis of Variance Section for k = 0.005000

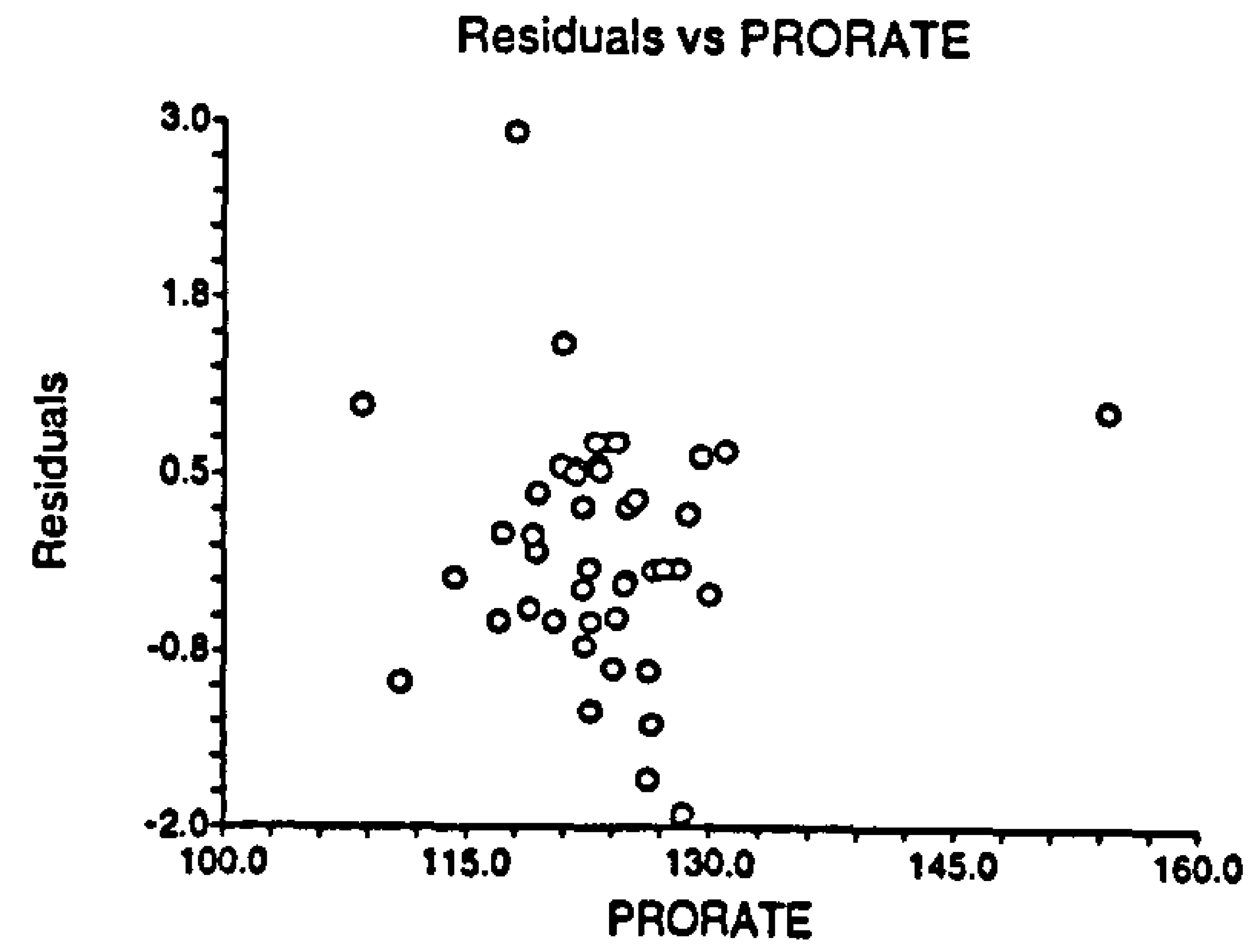
Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level
Intercept	1	18178.63	18178.63		
Model	4	84.05661	21.01415	26.0101	0.000000
Error	39	31.50902	0.8079236		
Total(Adjusted)	43	115.5656	2.687573		

Mean of Dependent	**20.32611**
Root Mean Square Error	**0.8988457**
R-Squared	**0.7273**
Coefficient of Variation	**4.422124E-02**

Residual Plots Section







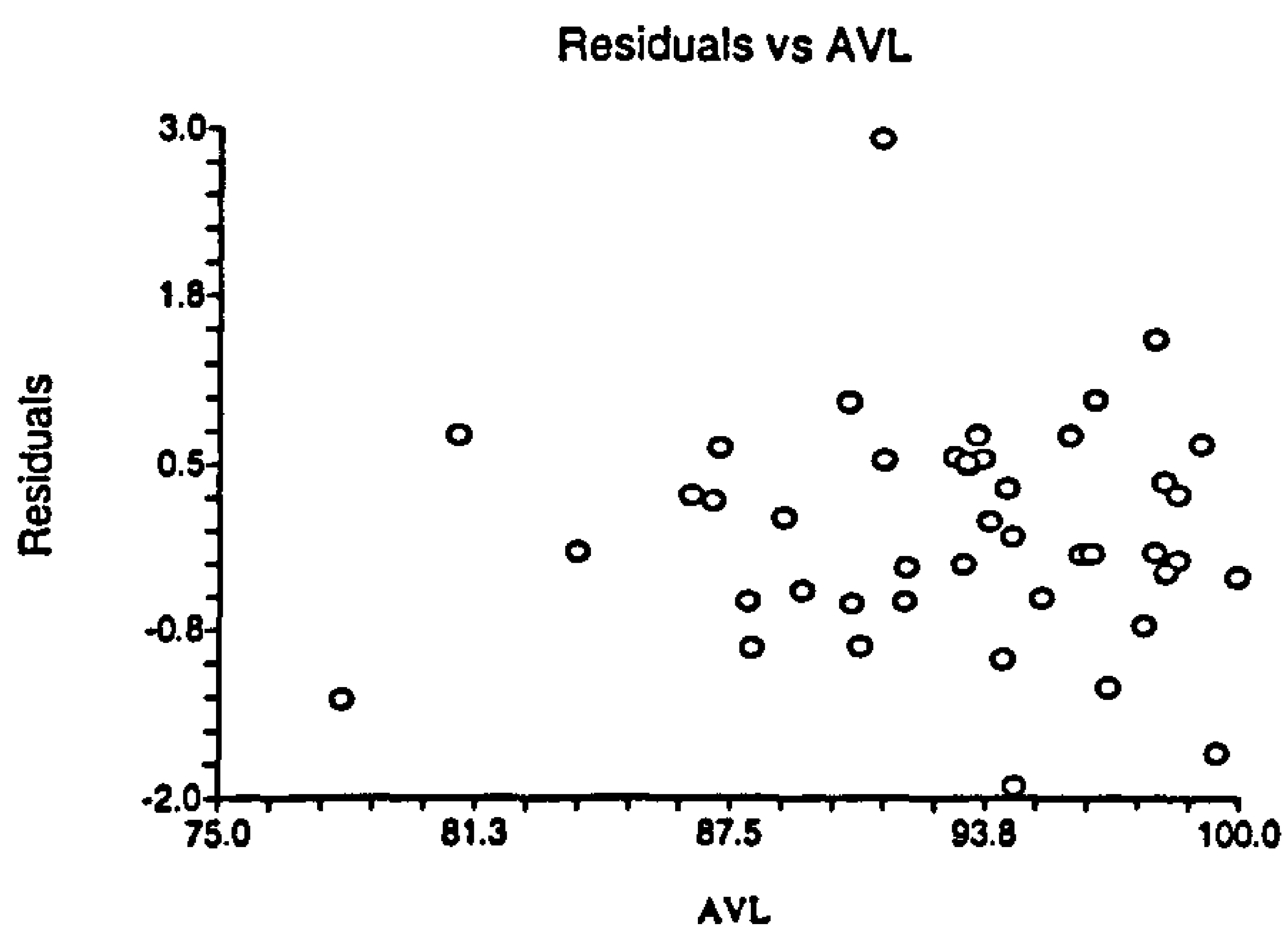


Table (7.10 d): Ridge Regression Report for the *Three Selected Variables*

Least Squares Multicollinearity Section

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance
AvNO	1.0063	0.0063	0.9937
AvHOURS	1.0327	0.0317	0.9683
PRORATE	1.0336	0.0325	0.9675

Since all VIF's are less than 10, multicollinearity is not a problem.

Eigenvalues of Correlations

No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.207259	40.24	40.24	1.00
2	0.964748	32.16	72.40	1.25
3	0.827993	27.60	100.00	1.46

All Condition Numbers less than 100. Multicollinearity is NOT a problem.

Ridge Regression Coefficient Section for k = 0.005000

Independent Variable	Regression Coefficient	Standard Error	Stand'zed Regression Coefficient	VIF
Intercept	26.0783			
AvNO	3.974617	1.271565	0.2599	0.9962
AvHOURS	0.1989508	2.392285E-02	0.7003	1.0218
PRORATE	-6.442608E-02	2.069905E-02	-0.2622	1.0226

Model

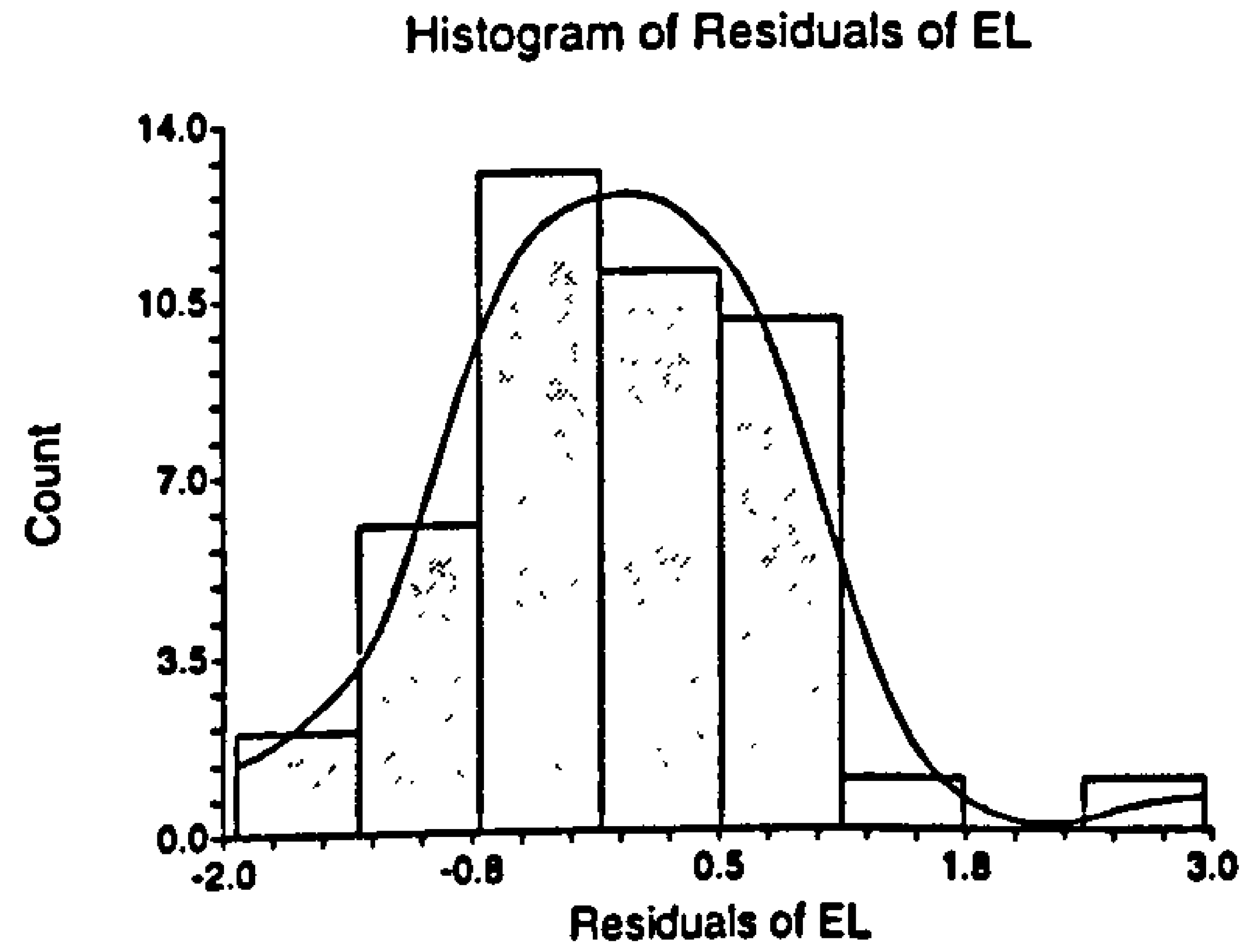
EL= 26.0783+ 3.974617*AvNO+ .1989508*AvHOURS-6.442608E-02*PRORATE

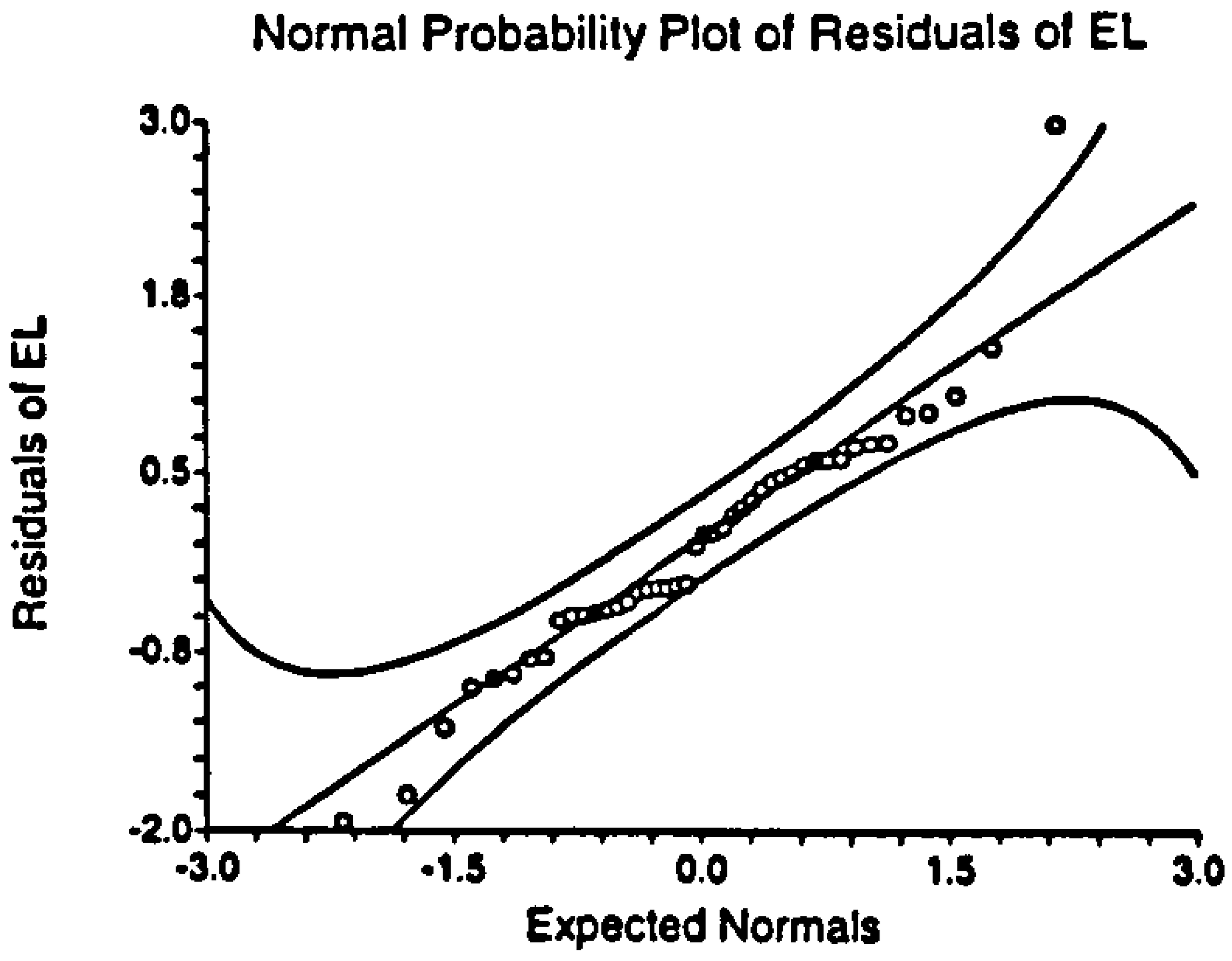
Analysis of Variance Section for k = 0.005000

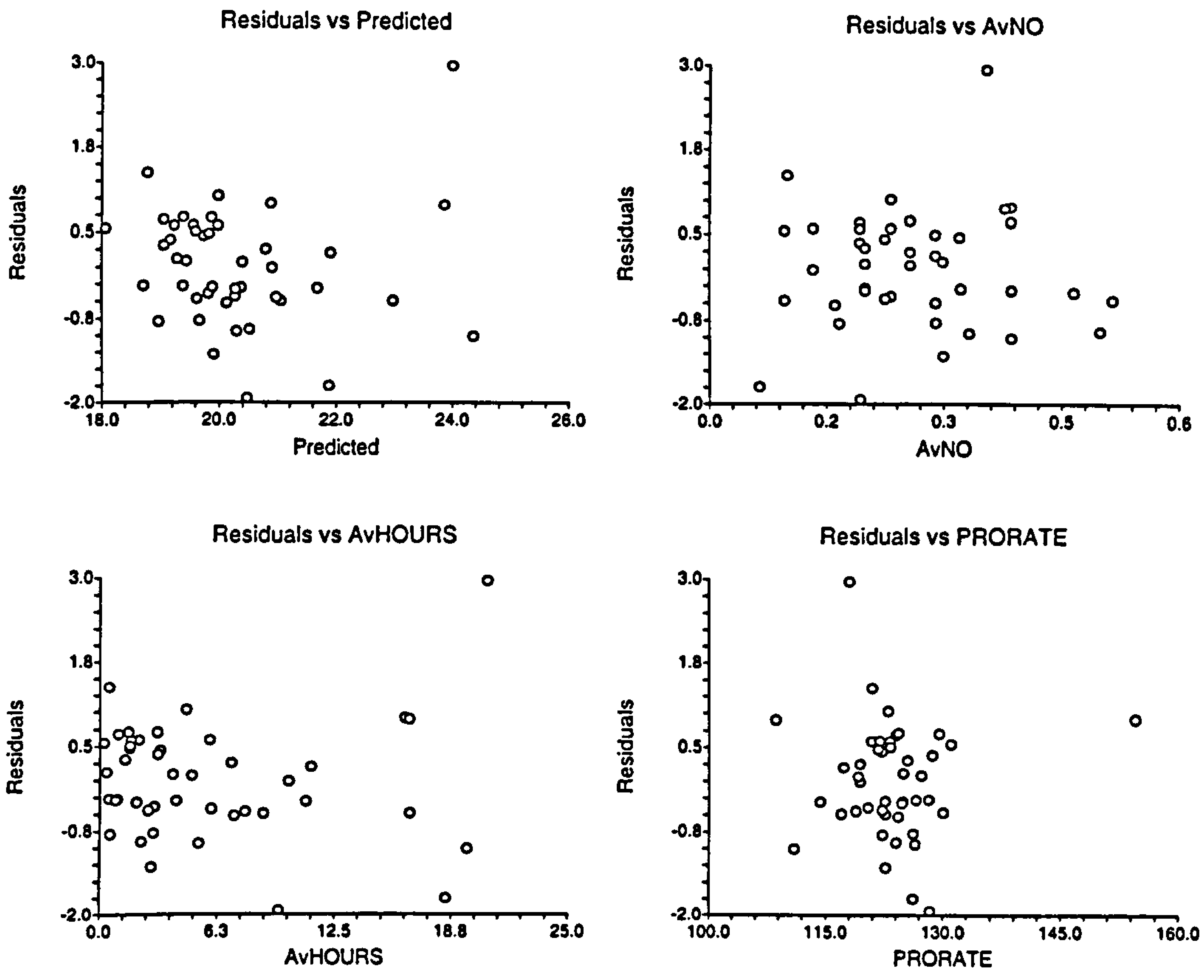
Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level
Intercept	1	18178.63	18178.63		
Model	3	83.48807	27.82936	34.7026	0.000000
Error	40	32.07756	0.801939		
Total(Adjusted)	43	115.5656	2.687573		

Mean of Dependent	**20.32611**
Root Mean Square Error	**0.8955104**
R-Squared	**0.7224**
Coefficient of Variation	**4.405716E-02**

Residual Plots Section







It is known that ridge regression is designed to be used in cases of the existence of multicollinearity problem. For our data, it was observed that multicollinearity is not a problem, eventhough we have considered a ridge procedure for the sake of comparison. It is interesting to note that the estimated full multiple regression and full ridge models are

EL=28.789+3.605AvNO+0.200AvHOURS-0.067PRORATE-0.025AVL,

EL=28.766+3.593AvNO+0.199AvHOURS-0.067PRORATE-0.025AVL, respectively.

Moreover, the R-squared values are 0.73 and 0.73 respectively. It is clear that the results of both models are almost the same. This confirms our previous result about milticollinearity. Hence, from now on we will not consider the ridge regression if multicollinearity is not a problem.

3) Robust Regression

Robust regression is designed for situation in which normality is a problem and when the data has some outliers. In this section we apply this procedure. This procedure applies three weight (influence) functions, Andrew's Sine, Tukey's Biweight, and Least Absolute Deviation.

1- Andrew's Sine

Table (7.10e) reports the computer output for the full model. From this output it seems reasonable that

a) The influential independent variables are AvHOURS, PRORATE and AVL.

b) The signs of the coefficients of the estimated parameters agrees with the expectations of the researcher in the sense that AvHOURS has an increasing relationship with EL but each of PRORATE and AVL has a decreasing relationship with EL.

c) The R-squared value is 0.888155, which is a very high percentage of determination.

d) The F-test in the ANOVA table indicates that the model is significant.

Table (7.10 e): Robust Regression Report for the Full Model Using Andrew's Sine

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	43.06403	2.764981	15.5748	0.000000	Reject Ho	1.000000
AvNO	0.5494547	0.6313672	0.8703	0.389769	Accept Ho	0.135535
AvHOURS	8.653179E-02	1.471458E-02	5.8807	0.000001	Reject Ho	0.999916
PRORATE	-0.1565388	1.840285E-02	-8.5062	0.000000	Reject Ho	1.000000
AVL	-4.524881E-02	0.0139723	-3.2385	0.002539	Reject Ho	0.883639
R-Squared	0.888155					

Model

EL=43.06403+ .5494547*AvNO+ 8.653179E-02*AvHOURS-.1565388*PRORATE-4.524881E-02*AVL

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	43.06403	2.764981	37.46164	48.66641	0.000000
AvNO	0.5494547	0.6313672	-0.7298167	1.828726	0.054303
AvHOURS	8.653179E-02	1.471458E-02	5.671721E-02	0.1163464	0.411662
PRORATE	-0.1565388	1.840285E-02	-0.1938265	-0.1192511	-0.617225
AVL	-4.524881E-02	0.0139723	-7.355937E-02	-1.693824E-02	-0.185138
T-Critical	2.026192				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	12519.26	12519.26			
Model	4	32.14933	8.037333	73.4536	0.000000	1.000000
Error	37	4.048558	0.1094205			
Total(Adjusted)	41	36.19789	0.8828753			
Root Mean Square Error		**0.3307877**	**R-Squared**	**0.888155**		
Mean of Dependent Variable		**20.20533**	**Adj R-Squared**	**0.876063**		
Coefficient of Variation		**1.637131E-02**				

Table (7.10f) reports the computer output for the model with the three selected independent variables. From this output it seems reasonable that

a) All the three independent variables in the model are influential as it was expected.

b) The signs of the coefficients of the estimated parameters agrees with the researcher expectations

c) The R-squared value is 0.886643, which is a very high percentage of determination. It is interesting to note that this value of R-squared is almost the same as that of the full model.

d) The F-test in the ANOVA table indicates that the model is significant.

Table (7.10 f): Robust Regression Report For the Selected Three Variables Using Andrew's Sine

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	44.05896	2.322016	18.9744	0.000000	Reject Ho	1.000000
AvHOURS	8.288898E-02	0.0138275	5.9945	0.000001	Reject Ho	0.999947
PRORATE	-0.1616058	1.661115E-02	-9.7288	0.000000	Reject Ho	1.000000
AVL	-4.751976E-02	0.0132433	-3.5882	0.000938	Reject Ho	0.937752
R-Squared	0.886643					

Model

EL=44.05896+ 8.288898E-02*AvHOURS-.1616058*PRORATE-4.751976E-02*AVL

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	44.05896	2.322016	39.35828	48.75964	0.000000
AvHOURS	8.288898E-02	0.0138275	5.489666E-02	0.1108813	0.394973
PRORATE	-0.1616058	1.661115E-02	-0.1952333	-0.1279782	-0.640502
AVL	-4.751976E-02	0.0132433	-7.432941E-02	-2.071011E-02	-0.197095
T-Critical	2.024394				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	12494.95	12494.95			
Model	3	31.3945	10.46483	99.0750	0.000000	1.000000
Error	38	4.013763	0.1056253			
Total(Adjusted)	41	35.40826	0.8636162			

Root Mean Square Error	**0.3250005**	**R-Squared**	**0.886643**
Mean of Dependent Variable	**20.19455**	**Adj R-Squared**	**0.877694**
Coefficient of Variation	**1.609348E-02**		

2- Tukey's Biweight

The following is the computer output for the full model. From this output it seems reasonable that:

a) The influential independent variables are AvHOURS, and PRORATE

b) The signs of the coefficients of the estimated parameters agrees with the expectations of the researcher in the sense that AvHOURS has an increasing relationship with EL but PRORATE has a decreasing relationship with EL.

c) The R-squared value 0.818959, which is a very high percentage of determination.

d) The F-test in the ANOVA table indicates that the model is significant.

Table (7.10 g): Robust Regression Report for the Full Model Using Tukey's Biweight

Regression Equation Section

Independent Variable (5%)	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power
Intercept	39.49982	3.57583	11.0463	0.000000	Reject Ho	1.000000
AvNO	1.003856	0.8157428	1.2306	0.226237	Accept Ho	0.224072
AvHOURS	0.100239	1.984953E-02	5.0499	0.000012	Reject Ho	0.998438
PRORATE	-0.1382195	2.322478E-02	-5.9514	0.000001	Reject Ho	0.999936
AVL	-0.0337513	1.681047E-02	-2.0078	0.052014	Accept Ho	0.498276
R-Squared	0.818959					

Model
EL= 39.49982+ 1.003856*AvNO+ .100239*AvHOURS-.1382195*PRORATE-.0337513*AVL

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	39.49982	3.57583	32.25449	46.74514	0.000000
AvNO	1.003856	0.8157428	-0.6489961	2.656708	0.097228
AvHOURS	0.100239	1.984953E-02	6.002006E-02	0.140458	0.446968
PRORATE	-0.1382195	2.322478E-02	-0.1852774	-9.116165E-02	-0.544249
AVL	-0.0337513	1.681047E-02	-6.781255E-02	3.099498E-04	-0.148521
T-Critical	2.026192				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	14716.2	14716.2			
Model	4	35.66846	8.917115	41.8434	0.000000	1.000000
Error	37	7.884947	0.2131067			
Total(Adjusted)	41	43.55341	1.062278			

Root Mean Square Error	**0.4616348**	**R-Squared**	**0.818959**
Mean of Dependent Variable	**20.12343**	**Adj R-Squared**	**0.799387**
Coefficient of Variation	**2.294016E-02**		

Since the full model contains some nonsignificant independent variables, we run the same procedure using only the significant independent variables. The resulting model will be called the final model. Table (7.10h) provides the computer output for the model with the two selected independent variables. From this output it seems reasonable that:

a) The influential independent variables are AvHOURS, and PRORATE. The signs of the coefficients of the estimated parameters agree with the expectations of the researcher.

b) The R-squared value is 0.733573, which is a very high percentage of determination, but it is smaller than that of the full model.

c) The F-test in the ANOVA table indicates that the model is significant.

Table (7.10 h): Robust Regression Report for the Model with the Two Selected Variables Using Tukey's Biweight

Regression Equation Section

Independent Variable	**Regression Coefficient**	**Standard Error**	**T-Value (Ho: B=0)**	**Prob Level**	**Decision (5%)**	**Power (5%)**
Intercept	24.96581	1.930966	12.9292	0.000000	Reject Ho	1.000000
AvHOURS	0.1835575	0.0198095	9.2661	0.000000	Reject Ho	1.000000
PRORATE	-4.614591E-02	1.545111E-02	-2.9866	0.004799	Reject Ho	0.830010
R-Squared	0.733573					

Model

EL= 24.96581+ .1835575*AvHOURS-4.614591E-02*PRORATE

Regression Coefficient Section

Independent Variable	**Regression Coefficient**	**Standard Error**	**Lower 95% C.L.**	**Upper 95% C.L.**	**Standardized Coefficient**
Intercept	24.96581	1.930966	21.06318	28.86844	0.000000
AvHOURS	0.1835575	0.0198095	0.143521	0.223594	0.771636
PRORATE	-4.614591E-02	1.545111E-02	-7.737377E-02	-1.491805E-02	-0.248706
T-Critical	2.021075				

Analysis of Variance Section

Source	**DF**	**Sum of Squares**	**Mean Square**	**F-Ratio**	**Prob Level**	**Power (5%)**
Intercept	1	15150.59	15150.59			
Model	2	36.50659	18.25329	55.0675	0.000000	1.000000
Error	40	13.25886	0.3314715			
Total(Adjusted)	42	49.76545	1.184892			

Root Mean Square Error	**0.5757356**	**R-Squared**	**0.733573**
Mean of Dependent Variable	**20.16928**	**Adj R-Squared**	**0.720252**
Coefficient of Variation	**2.854518E-02**		

3- Least Absolute Deviation

Table (7.10i) provides the computer output for the full model. From this output it seems reasonable that:

a) The influential independent variables are AvHOURS, PRORATE, and AVL.

b) The signs of the coefficients of the estimated parameters agree with the expectations of the researcher in the sense that AvHOURS has an increasing relationship with EL but each of PRORATE and AVL has a decreasing relationship with EL.

c) The R-squared value 0.845562, which is a very high percentage of determination.

d) The F-test in the ANOVA table indicates that the model is significant.

Table (7.10 i): Robust Report for the Full Model Using Least Absolute Deviation

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	45.47372	3.399487	13.3766	0.000000	Reject Ho	1.000000
AvNO	0.5171685	0.7955345	0.6501	0.519446	Accept Ho	0.097202
AvHOURS	6.699534E-02	0.0140952	4.7531	0.000027	Reject Ho	0.996243
PRORATE	-0.175245	2.228668E-02	-7.8632	0.000000	Reject Ho	1.000000
AVL	-4.544906E-02	1.712835E-02	-2.6534	0.011468	Reject Ho	0.734888
R-Squared	0.845562					

Model
EL= 45.47372+ .5171685*AvNO+ 6.699534E-02*AvHOURS-.175245*PRORATE-4.544906E
-02*AVL

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	45.47372	3.399487	38.59761	52.34983	0.000000
AvNO	0.5171685	0.7955345	-1.091952	2.126289	0.046255
AvHOURS	6.699534E-02	0.0140952	3.848511E-02	9.550557E-02	0.359283
PRORATE	-0.175245	2.228668E-02	-0.2203241	-0.130166	-0.625611
AVL	-4.544906E-02	1.712835E-02	-8.009441E-02	-0.0108037	-0.176924
T-Critical	2.022691				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	12786.21	12786.21			
Model	4	35.04299	8.760747	53.3820	0.000000	1.000000
Error	39	6.400451	0.1641141			
Total(Adjusted)	43	41.44344	0.9638009			

Root Mean Square Error	0.40511	R-Squared	0.845562
Mean of Dependent Variable	20.21327	Adj R-Squared	0.829722
Coefficient of Variation	2.004179E-02		

Table (7.10j) reports the computer output for the model with the three selected independent variables. From this output it seems reasonable that:

i. The influential independent variables are AvHOURS, PRORATE, and AVL.

ii. The signs of the coefficients of the estimated parameters agree with the expectations of the researcher in the sense that AvHOURS has an increasing relationship with EL but each of PRORATE and AVL has a decreasing relationship with EL.

iii. The R-Squared value 0.831353 which is a very high percentage of determination and it is about the same as that of the full model.

iv. The F-test in the ANOVA table indicates that the model is significant.

Table (7.10 j): Robust Report for the Model with the Three Selected Variables Using Least Absolute Deviation

Regression Equation Section

Independent		Regression	Standard	T-Value	Prob	Decision
	Power					
Variable	Coefficient	Error	(Ho: B=0)	Level	(5%)	(5%)
Intercept	46.11743	2.955879	15.6019	0.000000	Reject Ho	1.000000
AvHOURS	6.564681E-02	1.428782E-02	4.5946	0.000043	Reject Ho	0.994163
PRORATE	-0.1789688	2.106788E-02	-8.4949	0.000000	Reject Ho	1.000000
AVL	-0.0459222	1.633069E-02	-2.8120	0.007592	Reject Ho	0.783473
R-Squared	0.831353					

Model

EL=46.11743+ 6.564681E-02*AvHOURS-.1789688*PRORATE-.0459222*AVL

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	46.11743	2.955879	40.14337	52.09148	0.000000
AvHOURS	6.564681E-02	1.428782E-02	3.677005E-02	9.452357E-02	0.347894
PRORATE	-0.1789688	2.106788E-02	-0.2215486	-0.1363891	-0.644566
AVL	-0.0459222	1.633069E-02	-7.892776E-02	-1.291665E-02	-0.183438
T-Critical	2.021075				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	13461.84	13461.84			
Model	3	36.23626	12.07875	65.7271	0.000000	1.000000
Error	40	7.350854	0.1837714			
Total(Adjusted)	43	43.58711	1.013654			

Root Mean Square Error	0.4286856	R-Squared	0.831353
Mean of Dependent Variable	20.20392	Adj R-Squared	0.818704
Coefficient of Variation	2.121795E-02		

To sum up we observe from the robust regression models that

a) The estimated coefficients of the independent variables seem to be different. This is due to changes in the method of weighing outliers according to the specified truncation factors used by these three procedures.

b) The final model contains the same three variables (AvHOURS, PRORATE, and AVL) except in the case of using the Tukey's method, which has used only two independent variables (AvHOURS and PRORATE).

c) In the case of the full model, the R-squared values obtained by robust regression vary between 0.82 and 0.89 according to the weighting process. However, in the case of final model the R-squared values vary between 0.73 and 0.89. In both cases, the maximum R-squared values occur when Andrew's sine method is used.

Therefore, we recommend using the Andrew's method.

Continuation of Table (7.10 j): Comparison between initial and final robust models

	Intercept	AvNO	AvHOUR	PRORATE	AVL	R-Squared
Robust Full Model						
1-Least Abs. Dev. 1.0	33.14937	1.743778	.144374	-.08863285	-.03480079	0.845562
2-Tukey's Biweight 6.0	39.49982	1.003856	.100239	-.1382195	-.0337513	0.818959
3- Andrew's Sine 2.1	43.06403	.5494547	.08653179	-.1565388	-.04524881	0.888155
Robust Final Model						
1-Least Abs. Dev. 1.0	37.89022		.1259492	-.1130916	-.04742426	0.83133
2-Tukey's Biweight 6.0	24.96581		.1835575	-.04614591		0.733573
3-Andrew's Sine 2.1	44.05896		.0 8288898	-.1616058	-.04751976	0.886643

4) Nonlinear Regression Models

Based on the exploration of the data and the transformations of the variables, there is no clear indication of the possibility of suggesting suitable nonlinear regression models. However for the sake of comparison with linear regression models and to detect unexpected nonlinear models we consider in this section several nonlinear models.

1) **Multiplicative Model:**

All the above models are additive. One may also try to explore multiplicative models. In this section we run a multiplicative model, which is a sort of nonlinear regression models. Table (7.10k) reports the computer output of this full model.

It seems reasonable from this model that

a) The influential variables are AvNO, AvHours, and AVL.

b) The R-Squared value is 0.430624, which is not large enough as a determination coefficient.

c) The F-test in the ANOVA table accepts the model.

Table (7.10 k): Nonlinear Regression Report of the Multiplicative Full Model

Parameter Test Section

Parameter Name	Variable Name	Parameter Estimate	Asymptotic Standard Error	T-Value (Ho: Bi=0)	Prob Level
B1	AvNO	7.443619E-02	2.155465E-02	3.4534	0.001323
B2	AvHOURS	0.0496583	8.431031E-03	5.8899	0.000001
B3	PRORATE	4.919687E-02	0.1141516	0.4310	0.668798
B4	AVL	0.6220154	0.1219932	5.0988	0.000009

Model Estimation Section

Parameter Name	Variable Name	Parameter Estimate	Asymptotic Standard Error	Lower 95.0% C.L.	Upper 95.0% C.L.
B1	AvNO	7.443619E-02	2.155465E-02	3.087262E-02	0.1179998
B2	AvHOURS	0.0496583	8.431031E-03	3.261855E-02	6.669805E-02
B3	PRORATE	4.919687E-02	0.1141516	-0.1815121	0.2799059
B4	AVL	0.6220154	0.1219932	0.3754579	0.868573

R-Squared 0.430624
Iterations 17

Model
EL=AvNO^(B1) *AvHOURS^(B2) *PRORATE^(B3) *AVL^(B4)

Estimated Model
EL=AvNO^(7.443619E-02) *AvHOURS^(0.0496583) *PRORATE^(4.919687E-02) *AVL^(0.6220154)

Analysis of Variance Table

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level
Mean	1	18178.63	18178.63		
Model	4	18228.39	4557.098		
Model (Adjusted)	3	49.76529	16.58843	10.0841	0.000045
Error	40	65.80033	1.645008		
Total (Adjusted)	43	115.5656			
Total	44	18294.19			

Table (7.10l) reports the computer output of this model with the selected three variables. It seems reasonable from this model that:

a) All the variables in the model are influential variables as expected.

b) The R-Squared value is 0.428102, which is not large enough as a determination coefficient. This value is almost the same as that of the full model.

c) The F-test in the ANOVA table accepts the model.

Table (7.10 l): Nonlinear Regression Report of the Multiplicative Model with the Selected Three Variables

Parameter Test Section

Parameter Name	Variable Name	Parameter Estimate	Asymptotic Standard Error	T-Value (Ho: Bi=0)	Prob Level
B1	AvNO	7.584204E-02	2.109673E-02	3.5950	0.000863
B2	AvHOURS	4.953545E-02	8.329256E-03	5.9472	0.000001
B3	AVL	0.6747255	7.36358E-03	91.6301	0.000000

Model Estimation Section

Parameter Name	Variable Name	Parameter Estimate	Asymptotic Standard Error	Lower 95.0% C.L.	Upper 95.0% C.L.
B1	AvNO	7.584204E-02	2.109673E-02	3.323634E-02	0.1184477
B2	AvHOURS	4.953545E-02	8.329256E-03	3.271417E-02	6.635673E-02
B3	AVL	0.6747255	7.36358E-03	0.6598544	0.6895965

R-Squared 0.428102
Iterations 10
Model
EL=AvNO^(B1) *AvHOURS^(B2) *AVL^(B3)

Estimated Model
EL=AvNO^(7.584204E-02) *AvHOURS^(4.953545E-02) *AVL^(0.6747255)

Analysis of Variance Table

Source	**DF**	**Sum of Squares**	**Mean Square**	**F-Ratio**	**Prob Level**
Mean	1	18178.63	18178.63		
Model	3	18228.1	6076.034		
Model (Adjusted)	2	49.47385	24.73692	15.3455	0.000011
Error	41	66.09178	1.611995		
Total (Adjusted)	43	115.5656			
Total	44	18294.19			

Comparing the R-squared value of the multiplicative model with that of the additive model, we conclude that the multiplicative model is not useful with respect to the additive one.

2) Quadratic Regression Model

One may fit a quadratic regression model. The following Table (7.10m) reports the computer output of the fitted model. It seems reasonable from this output that:

The R-squared value is 0.807448 which is higher than that of the additive model.

This result is expected since four quadratic terms are added to the model.

a) Instead of the t-test, the NCSS2000 Package provides confidence intervals for the parameters of the model. At the 0.05 level of significance, the confidence intervals that contain the number zero show the significant terms. These significant terms are AvNO, AvHOURS, PRORATE, and the squares of AvNO, AvHOURS, and AVL.

Table (7.10 m): Nonlinear Regression Report

Model Estimation Section

Parameter Name	Parameter Estimate	Asymptotic Standard Error	Lower 95% C.L.	Upper 95% C.L.
A	47.33813	42.87769	-39.70821	134.3845
B	4.784168	5.529918	-6.442163	16.0105
C	-4.385167E-02	8.349159E-02	-0.2133486	0.1256453
D	-0.8719588	0.3863441	-1.656279	-8.763855E-02
E	0.7775392	0.7158151	-0.6756427	2.230721
F	-6.31932	9.018538	-24.62793	11.98929
G	0.010863	4.171871E-03	2.393649E-03	1.933234E-02
H	3.071575E-03	1.468859E-03	8.963299E-05	6.053517E-03
I	-4.593988E-03	3.964326E-03	-0.012642	3.454022E-03

Model
EL=A+B*AVNO+C*AVHOURS+D*PRORATE+E*AVL+F*AVNO^2+G*AVHOURS^2
+H*PRORATE^2+I*AVL^2
R-Squared 0.807448
Iterations 10
Estimated Model
EL=(47.33813)+(4.784168)*(AVNO)+(-4.385167E-02)*(AVHOURS)+(-
.8719588)*(PRORATE)+(.7775392)*(AVL)+(-
6.31932)*(AVNO)^2+(.010863)*(AVHOURS)^2+(3.071575E-03)*(PRORATE)^2+(-
4.593988E-03)*(AVL)^2

It seems reasonable that this model has a higher value of R-squared than that of the linear multiple regression but it is too complicated model. On the other hand still the robust regressio model has higher R-squared value than this model. So, we prefer robust regression model.

3) Polynomial Regression Model with Interaction

One may fit a quadratic regression model with interactions. Table (7.10n) reports the computer output of Response-Surface Regression procedure, which searches for the optimal model including up to cubic terms with interactions. It seems reasonable from this output that:

b) The R-squared value for the fitted model is 0.825879, which is very high.

c) At the 0.05 level of significance, the F-tests show that the model is useful. Moreover, the linear, quadratic and linear by linear interaction terms that are involved in the model are significant.

d) The only significant linear terms are AvNO, AvHOURS, and PRORATE.

e) The estimated model is EL=43.02719-45.53971*AvNO-9.654256E-02*AvHOURS-.1899082*PRORATE+ 6.711594E-03*AvHOURS^2+ .583281*AvNO*AvHOURS+ .3677993*AvNO*PRORATE

f) At the 0.05 level of significance the only significant terms are PRORATE and the interaction between AvNO and AvHOURS.

Table (7.10 n): Response-Surface Regression (Polynomial Regression Model with Interaction)

Hierarchical Model Summary Section

Number of Terms Removed	3
Number of Terms Remaining	6
R-Squared Cutoff Value	0.010000
R-Squared of Final Model	0.825879

Sequential ANOVA Section

Source	Df	Sequential Sum-Squares	Mean Square	F-Ratio	Prob Level	Incremental R-Squared
Regression	6	95.44321	15.9072	29.25	0.000000	0.825879
Linear	3	83.85179	27.9506	51.39	0.000000	0.725577
Quadratic	1	2.921314	2.921314	5.37	0.026104	0.025278
Lin x Lin	2	8.670107	4.335053	7.97	0.001323	0.075023
Total Error	37	20.12242	0.5438491			0.174121

ANOVA Section

Factor	Df	Last Sum-Squares	Mean Square	F-Ratio	Prob Level	Term R-Squared
AvNO	3	14.90091	4.966969	9.13	0.000118	0.128939
AvHOURS	3	63.79533	21.26511	39.10	0.000000	0.552027
PRORATE	2	9.478274	4.739137	8.71	0.000792	0.082016
Total Error	37	20.12242	0.5438491		0.174121	

Based on these comments we do not recommend using this model.

Table (7.10 o): Multiple Logarithmic Regression Report

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	4.546106	1.170647	3.8834	0.000387	Reject Ho	0.966087
LNNO	4.160706E-02	1.882103E-02	2.2107	0.032992	Reject Ho	0.577740
LNHR	3.755876E-02	7.213486E-03	5.2067	0.000007	Reject Ho	0.999080
LNPR	-0.4139806	0.1607194	-2.5758	0.013905	Reject Ho	0.709523
LNAV	0.1037136	0.1617489	0.6412	0.525142	Accept Ho	0.095894
R-Squared	0.569801					

Model
EL= 4.546106+ 4.160706E-02*LNAvNO+ 3.755876E-02*LNAvHR-
.4139806*LNPRORATE+ .1037136*LNAVL

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	4.546106	1.170647	2.178249	6.913962	0.0000
LNAvNO	4.160706E-02	1.882103E-02	3.537935E-03	0.0796762	0.2540
LNAvHOUR	3.755876E-02	7.213486E-03	2.296811E-02	5.214941E-02	0.5752
LNORORATE	-0.4139806	0.1607194	-0.7390663	-8.889486E-02	-0.2852
LNAVL	0.1037136	0.1617489	-0.2234544	0.4308817	0.0756
T-Critical	2.022691				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	398.3815	398.3815			
Model	4	0.139129	3.478226E-02	12.9139	0.000001	0.999967
Error	39	0.1050422	2.693389E-03			
Total (Adjusted)	43	0.2441712	0.0056784			

Root Mean Square Error	**5.189787E-02**	**R-Squared**	**0.5698**
Mean of Dependent	**3.009007**	**Adj R-Squared**	**0.5257**
Coefficient of Variation	**1.724751E-02**	**Press Value**	**0.2001552**
Sum lPress Residualsl	**1.880178**	**Press R-Squared**	**0.1803**

Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	2.9956	0.002739	Rejected
Kurtosis	2.9894	0.002795	Rejected
Omnibus	17.9100	0.000129	Rejected

Serial-Correlation Section

Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.397321	9	0.081451	17	-0.283406
2	0.139991	10	0.113021	18	-0.180410
3	0.075922	11	-0.061317	19	-0.159186
4	-0.050655	12	0.039941	20	-0.200698
5	-0.105880	13	0.102086	21	-0.116587
6	-0.130806	14	0.083540	22	-0.025187
7	-0.207086	15	0.032069	23	-0.007960
8	-0.118190	16	-0.290586	24	0.135239

Above serial correlations significant if their absolute values are greater than 0.301511

Durbin-Watson Value 1.2029

Multicollinearity Section

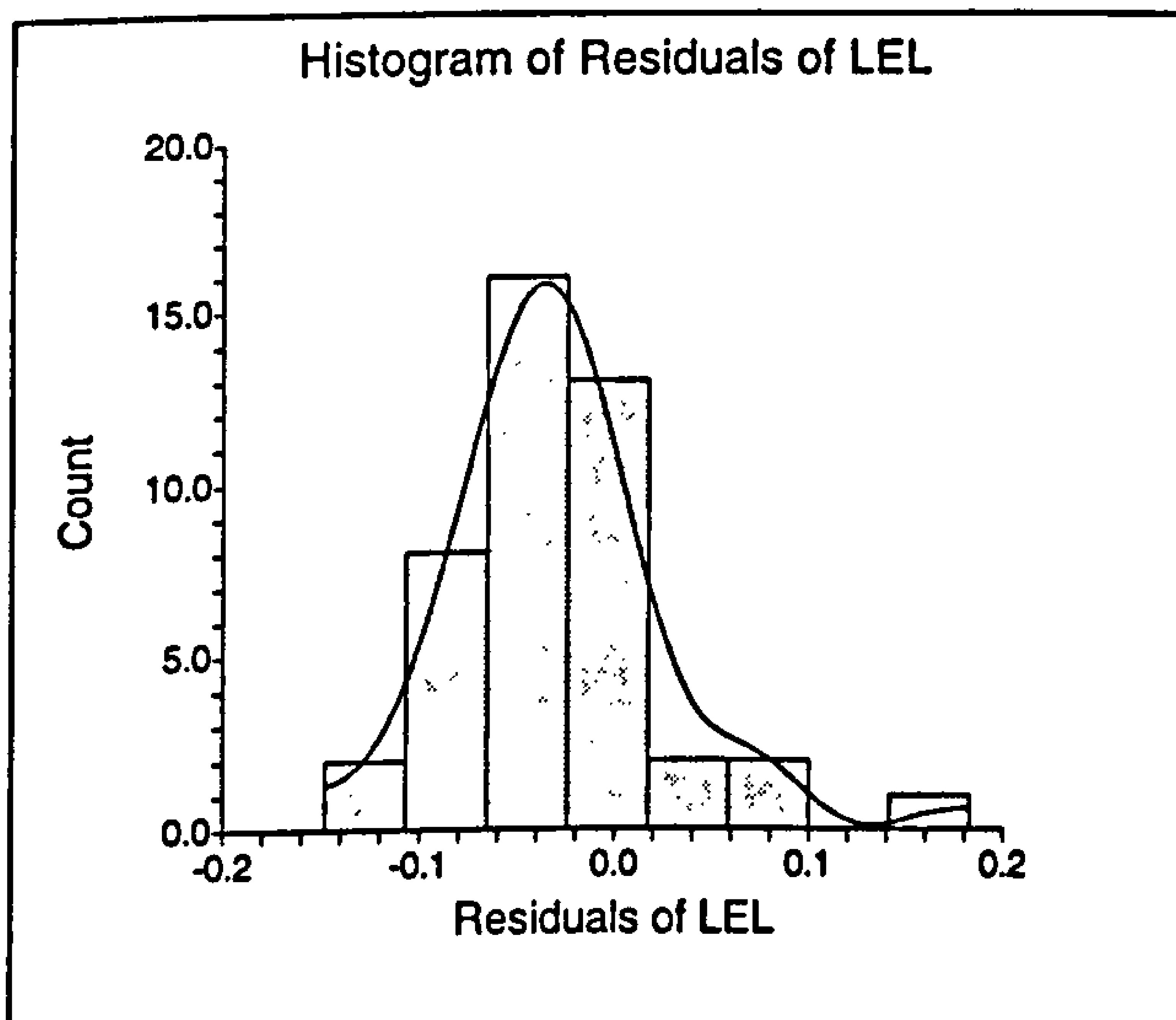
Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance	Diagonal of X'X Inverse
LNAvNO	1.196896	0.164505	0.835495	0.1315188
LNAvHOUR	1.106508	0.096256	0.903744	0.0193193
LNPRORATE	1.111728	0.100500	0.899500	9.59042
LNAVL	1.258824	0.205608	0.794392	9.713676

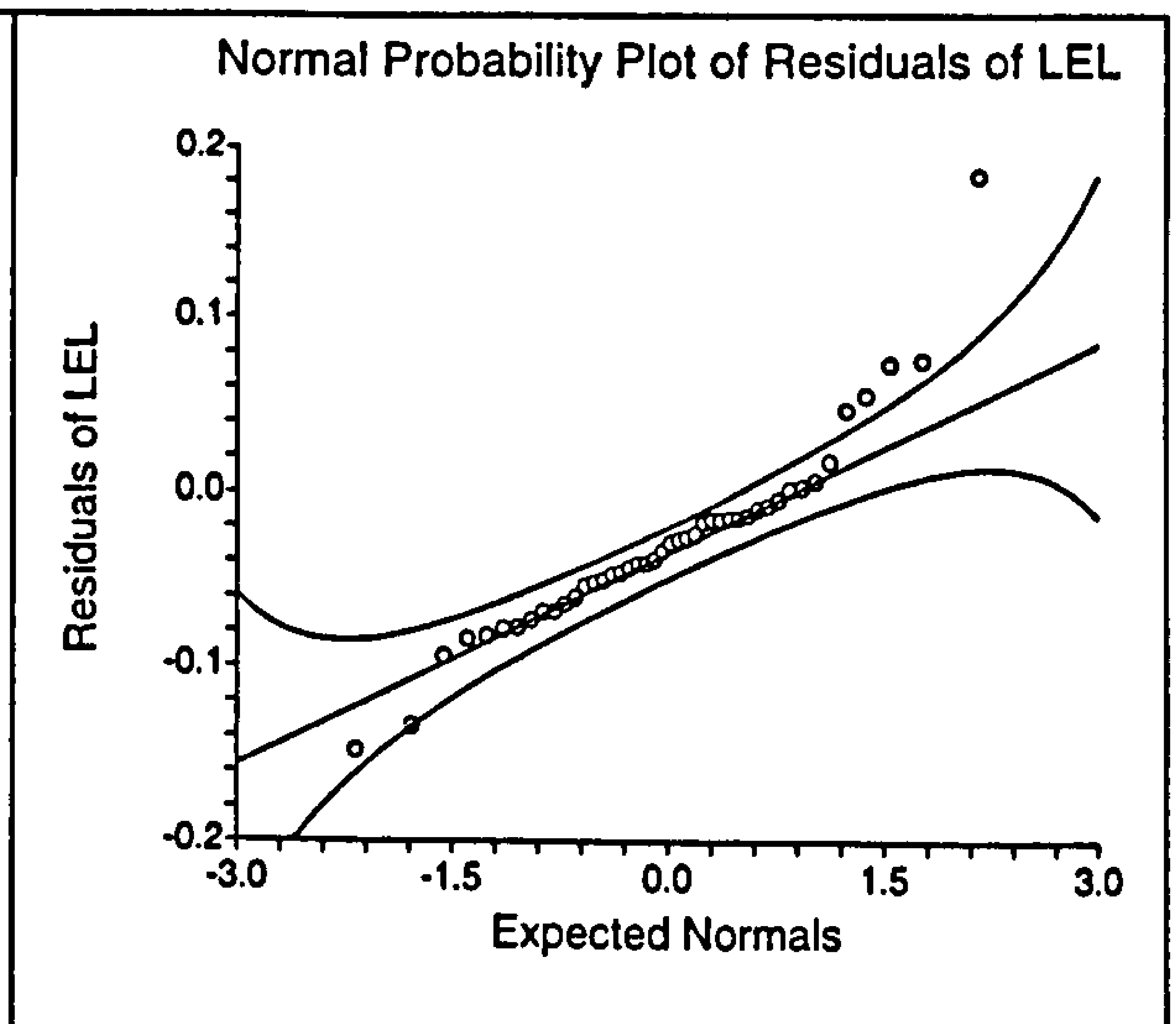
Eigenvalues of Centered Correlations

No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.479792	36.99	36.99	1.00
2	1.158454	28.96	65.96	1.28
3	0.851807	21.30	87.25	1.74
4	0.509946	12.75	100.00	2.90

All Condition Numbers less than 100. Multicollinearity is NOT a problem.

Plots Section





Comments and Conclusions

To sum up, we collect the main results in Tables (7.11)-(7.13).

1) Table (7.11) aims at comparing the full models, which involve all the available independent variables. It seems reasonable from this table that:

A. In the all-possible regression models, we enter the most effective variable in the first step in the sense that it gives the minimum Cp value. In the second step the most two effective variables are entered in the model and so on. This is called sequentially entering process. The coefficients of the estimated models by sequentially entering one variable at a time are almost the same. This indicates a type of consistency of the obtained results.

B. Comparing the full multiple regression-estimated coefficients with the corresponding coefficients of the ridge regression, we observe that they are almost the same in the sense that the estimated coefficients and the R-squared values are almost the same. This means that the multicollinearity could not be a problem since ridge regression is designed to deal with multicollinearity if it exists. This result confirms the results of testing multicollinearity, which was given in a previous section.

C. The estimated coefficients of robust regression models are different because the three robust procedures use different truncation methods in weighing the error terms. However, the largest R-squared is obtained by the Andrew's sine method. Moreover, the R-squared values that are given by this procedure are higher than those given by other procedures. This is natural in our case since the data has some outliers and there are some minor problems with the normality assumption. Because the robust regression is designed to handle such problems we recommend using the results of the robust regression.

D. The coefficients of each of the independent variables have the same signs in all models. The AvNO and AvHOURS variables have positive signs, which means that the dependent variable increases as each of these independent variables increases. However, The PRORATE, and AVL variables have negative signs, this means that the dependent variable decreases as each of these variables increases. This observation agrees with the researcher expectations as it was stated in section 7.4.2.

E. The relative importance of the independent variables can be seen from the all possible regression as well as from the stepwise procedure. This importance may be arranged according to the relative increase in R-squared value when that variable enters the model. It seems reasonable that the most important variable is AvHOURS, which is the first variable to enter the model with an R-squared value about 0.58. The second important variable is AvNO, which raised R-squared value to about 0.66, i.e. its own contribution to R-Squared is about 0.08. The third important variable is PRORATE which increased R-squared value to about 0.73, i.e. its contribution to R-squared value is about 0.07. Finally, the fourth important variable is AVL, which increased R-squared value to about 0.73, i.e. its contribution to the R-squared value is about zero. So, it is very clear that this last variable has non-significant influence on the model.

F. Comparing the R-squared values of multiple regression, ridge regression, stepwise regression, and the three robust regression procedures, we observe that this value ranges between 0.73 and 0.89. The maximum is that of the Andrew sine robust model and the minimum is shared by stepwise, multiple and ridge regression models. This is another reason to recommend using the robust regression model.

Table (7.11) Comparison of Full Models

Model	Intercep	AvNO	AvHOURS	PRORATE	AVL	R-Squared
All Possible						
Size 1	19.0700		.2169614			0.583198
Size 2	17.9827	4.209713	.2124972			0.658721
Size 3	26.0888	3.99093	.1998955	-.06459102		0.725577
Size 4 (Multiple)	28.7895	3.605562	.2005477	-.06698804	-.02482429	0.730505
Stepwise						
1- Forward	26.0888	3.99093	.1998955	-.06459102		0.725577
2- Step	26.0888	3.99093	.1998955	-.06459102		0.725577
3-Backward	26.0888	3.99093	.1998955	-.06459102		0.725577
4-Min MSE	26.0888	3.99093	.1998955	-.06459102		0.725577
Ridge	28.7659	3.592767	.1995978	-.06679998	-.02472465	0.7273
Robust						
1-Least Abs. Dev. 1.0	33.1493	1.743778	.144374	-.08863285	-.03480079	0.845562
2-Tukey's Biweight 6.0	39.4998	1.003856	.100239	-.1382195	-.0337513	0.818959
3- Andrew's Sine 2.1	43.0640	.5494547	.08653179	-.1565388	-.04524881	0.888155

2) Since there are independent variables that are included in the full regression models even though they are not significant, one should fit the models another time after deleting these variables from the models. The following table summarizes the results of these models with the selected variables. This will be called the final run results. It seems reasonable from table (7.12) that the same conclusions given above are still valid. Moreover, the robust regression models excluded AvNO and included AVL. On the contrary the situation is reversed for other regression models. This may be due to outliers.

Table (7.12) Comparison of Models with Significant Variables (Final Run)
A blank cell means that the variable is not included in the model

Procedure	Intercept	AvNO	AvHOURS	PRORATE	AVL	R-Squared
All Possible						
Size 1	19.07001		.2169614			0.583198
Size 2	17.98276	4.209713	.2124972			0.658721
Size 3	26.08888	3.99093	.1998955	-.06459102		0.725577
Size 4 (Multiple)	26.08888	3.99093	.1998955	-.06459102		0.725577
Stepwise						
1- Forward	26.08888	3.99093	.1998955	-.06459102		0.725577
2- Step	26.08888	3.99093	.1998955	-.06459102		0.725577
3-Backward	26.08888	3.99093	.1998955	-.06459102		0.725577
4-Min MSE	26.08888	3.99093	.1998955	-.06459102		0.725577
Ridge	26.0783	3.974617	.1989508	-.06459102		0.7224
Robust						
1-Least Abs. Dev. 1.(	37.89022		.1259492	-.1130916	-.04742426	0.83133
2-Tukey's Biweight (	24.96581		.1835575	-.04614591		0.733573
3-Andrew's Sine 2.1	44.05896		.0 8288898	-.1616058	-.04751976	0.886643

3) Table (7.13) provides the R-squared values of the nonlinear regression models. It seems reasonable from this table that the values of R-squared ranges between about 0.43 and 0.83. The minimum value corresponds to the Multiplicative model and the maximum value is that of the polynomial regression with interaction terms. It is natural to have a higher value of R-squared each time you enter another term to the model. This explains why the value of R-squared is so high in the case of polynomial regression with interaction terms. Moreover, if we compare the R-squared value of the polynomial regression with interaction terms with that of the robust regression, we still observe that the R-squared value of robust regression is larger than that of the polynomial regression with interaction terms. For this reason and the fact that the polynomial regression with interaction terms is a more complicated model, we recommend using robust regression.

Table (7.13) Comparison of Nonlinear Models

Model	R-Squared
Multiplicative	0.430624
Quadratic	0.807448
Polynomial with interactions	0.825879
Logarithmic	0.569801

Final Conclusions

As a final statement in this direction we say that, there are three reasons to recommend the robust regression model.

i) Robust regression is sensitive to outliers to some extent but not as sensitive as ordinary least square regression. Moreover robust regression is not too sensitive to normality assumption.

ii) The other linear models have smaller R-squared values than that of the robust model.

iii) The nonlinear models are more complicated than the robust model and they have R-squared value smaller than that of the robust model. Moreover, the signs of estimated coefficients of the independent variables are consistent with the researcher expectations, in the case of robust regression, but some of those signs may disagree with those expectations in some nonlinear models.

Recommendation

Based on the above observations we recommend using the Andrew's sine robust estimated regression model, which is given by

El=44.05896+0.08288898AvHOURS-0.1616058PRORATE-0.04751976AVL.

This model explains about 89% of the variability in EL.

Notes:

a) Since the Andrew method has the largest R-squared values among all robust models, we will use only this robust procedure for the other kilns.

b) Since the nonlinear models are complicated and have R-squared values smaller than those of the final robust regression models, we will give their results on the other kilns in Appendix(20).

c) Since multicollinearity is not a problem in our data we will not treate the ridge regression for other kilns.

7.5.3.2 Fuel For Kiln 1

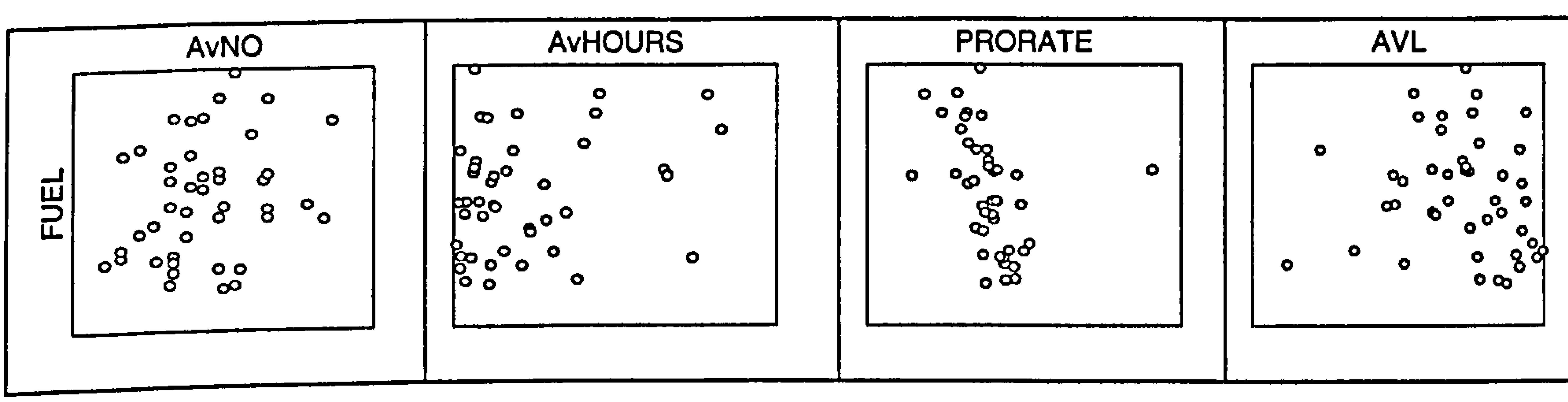
7.5.3.2.1 Exploring Data

I) **Screening Process:**

As a first step in analyzing the FUEL data of Kiln1, we start exploring the available data. This exploration includes:

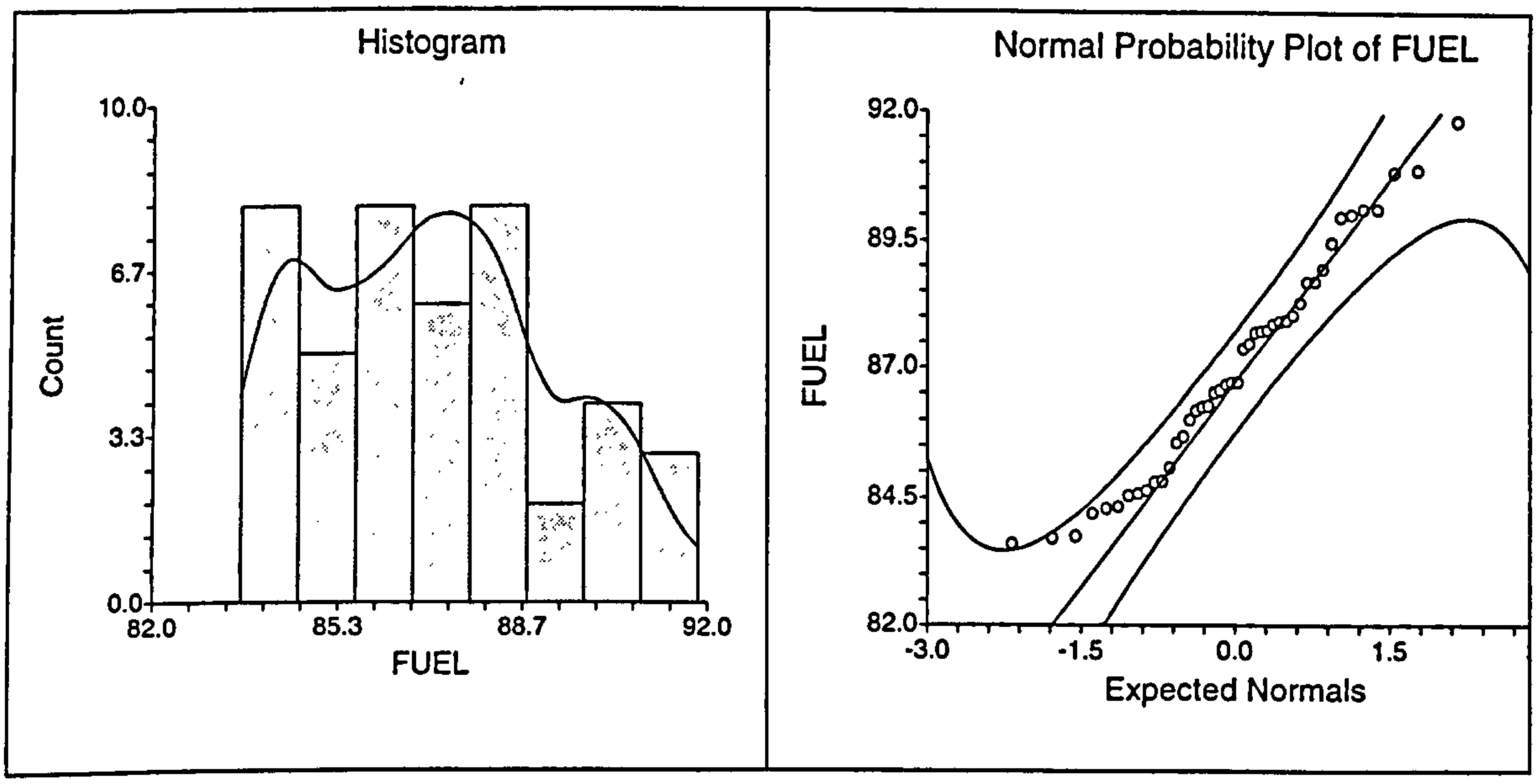
1) The *scatter plots* of the dependent variable FUEL against each of the independent variables (AvNO, AvHOURS, PRORATE, and AVL) are shown in Table (7.9a/FL). It seems reasonable from these plots that there is a linear relationship between FUEL and PRORATE if we disregard an existing outliear. Moreover, there is an indication of possible outliers.

Table (7.9a/FL) Scatter Plots of FUEL vs. Independent Variables



2) The *histogram and the normal probability plot* of FUEL are given in Table (7.9b/FL). It seems reasonable from the normal probability plot that normality could not be a problem. However, the histogram shows that the distribution of FUEL is multimodal and not symmetric. These inconsistencies may be due to the existence of possible outliers in the data.

Table (7.9b/FL) Histogram and Normal Probability Plot of FUEL



3) *Normality test*, which is given in Table (7.9c/FL), accepted the normality of FUEL.

So, transformations are not needed.

Table (7.9c/FL) Normality Tests of Transformations of EL

Normality Tests Section

	------ Skewness Test -----			----- Kurtosis Test ---------			Omnibus Test -		Variable
Variable	**Value**	**Z**	**Prob**	**Value**	**Z**	**Prob**	**K2**	**Prob**	**Normal?**
FUEL	0.27	0.82	0.4147	2.17	-1.44	0.1487	2.75	0.2528	Yes

4) A search for the *outliers* is performed. Table (7.9d/FL) shows that there is only one outlier in the data, namely 34[th] observation.

Table (7.9d/FL) Outliers in EL

Row	Value	T2 Prob	T2 Outlier?
1	0.03	0.8743	
2	0.38	0.5407	
3	1.03	0.3161	
4	1.18	0.2839	
5	0.12	0.7317	
6	2.22	0.1437	
7	2.43	0.1260	

8	1.50	0.2270	
9	1.66	0.2050	
10	0.79	0.3788	
11	0.17	0.6814	
12	3.03	0.0887	
13	0.04	0.8399	
14	0.05	0.8321	
15	1.55	0.2202	
16	2.28	0.1387	
17	1.28	0.2649	
18	1.01	0.3211	
19	0.02	0.8818	
20	0.59	0.4479	
21	0.06	0.8128	
22	0.11	0.7467	
23	0.09	0.7631	
24	1.25	0.2698	
25	0.77	0.3837	
26	0.14	0.7078	
27	1.87	0.1787	
28	0.16	0.6885	
29	0.03	0.8696	
30	0.22	0.6420	
31	0.15	0.6997	
32	2.00	0.1647	
33	1.81	0.1858	
34	4.84	0.0332	Yes
35	3.10	0.0853	
36	2.00	0.1643	
37	0.33	0.5702	
38	0.58	0.4500	
39	0.10	0.7530	
40	0.45	0.5057	
41	1.24	0.2720	
42	0.13	0.7243	
43	0.02	0.8861	
44	0.21	0.6467	

5) To check the validity of our observations from the scatter plot, we calculate the *correlation coefficients* between FUEL and each of the independent variables. Table (7.9e/FL) reports these correlations together with their p-values. It seems reasonable from these results that, at the 0.05 level of significance, there is a linear relationship between FUEL and each of AvNO and PRORATE since their p-values are less than 0.05. This result confirms the observation from the scatter plots.

Table (7.9e/FL): Correlations between EL and Independent Variables

	FUEL	AvNO	AvHOURS	PRORATE	AVL
FUEL	1.000000	0.297749	0.260236	-0.607530	-0.140044
	0.000000	0.049651	0.087998	0.000012	0.364574

6) To explore the possibility of ***multicollinearity***, the correlations between the independent variables that are given in Table (7.9f/FL). As it was pointed out in the previous section, mulicollinearity may occur if there is a very high correlation between some of the independent variables. The threshold, which is recommended here, is 0.75. It seems reasonable that all the obtained correlations are less than 0.75 since the maximum absolute value of these correlations is 0.507610. So, multicollinearity may not be a problem.

Table (7.9f/FL): Correlations Between Independent Variables

	AvNO	AvHOURS	PRORATE	AVL
AvNO	1.000000	0.182399	-0.331517	-0.356613
	0.000000	0.236002	0.027927	0.017496
AvHOURS	0.182399	1.000000	-0.507610	-0.083451
	0.236002	0.000000	0.000436	0.590192
PRORATE	-0.331517	-0.507610	1.000000	-0.047850
	0.027927	0.000436	0.000000	0.757747
AVL	-0.356613	-0.083451	-0.047850	1.000000
	0.017496	0.590192	0.757747	0.000000

7) Based on the above exploration, it is worth mentioning that the ordinary least square method to estimate the coefficients of the multiple regression will not be completely reliable because of the existence of outliers and the possibility that FUEL is not normally distributed as it seen from the histogram of FUEL. As a remedy of this situation, one should use the robust regression. However, a closer look at the residuals of the models to be fitted will give a more mature judgment to this situation. Moreover, the results from the scatter plots and the correlations between FUEL and the independent variables raise the following question: Which variables should be included in the regression models to be fitted to the given data? To answer this question, the following steps were implemented.

(II) Variables Selection

As a first selection procedure, we applied the ***All-Possible Regression Procedure***. The results of this procedure are given in Table (7.10a). Based on the values Cp criterion, It seems reasonable that the best model is that with only one independent variable, namely, PRORATE. This result confirms the result obtained from the correlation matrix.

Table (7.10a/FL): All Possible Results Section

Model Size	R-Squared	Root MSE	Cp	Model
1	0.369092	1.76251	1.357890	C (PRORATE)
1	0.088654	2.118314	19.741430	A (AvNO)
1	0.067723	2.142502	21.113547	B (AvHOURS)
1	0.019612	2.197089	24.267344	D (AVL)

Variables in Best Model
PRORATE

2	0.397758	1.742879	1.478793	CD
2	0.379520	1.769071	2.674304	AC
2	0.372216	1.779454	3.153145	BC
2	0.132520	2.091757	18.865919	AB
2	0.089968	2.142445	21.655309	AD

7) Based on the above exploration, it is worth mentioning that the ordinary least square method to estimate the coefficients of the multiple regression will not be completely reliable because of the existence of outliers and the possibility that FUEL is not normally distributed as it seen from the histogram of FUEL. As a remedy of this situation, one should use the robust regression. However, a closer look at the residuals of the models to be fitted will give a more mature judgment to this situation. Moreover, the results from the scatter plots and the correlations between FUEL and the independent variables raise the following question: Which variables should be included in the regression models to be fitted to the given data? To answer this question, the following steps were implemented.

(II) Variables Selection

As a first selection procedure, we applied the ***All-Possible Regression Procedure***. The results of this procedure are given in Table (7.10a). Based on the values Cp criterion, It seems reasonable that the best model is that with only one independent variable, namely, PRORATE. This result confirms the result obtained from the correlation matrix.

Table (7.10a/FL): All Possible Results Section

Model Size	R-Squared	Root MSE	Cp	Model
1	0.369092	1.76251	1.357890	C (PRORATE)
1	0.088654	2.118314	19.741430	A (AvNO)
1	0.067723	2.142502	21.113547	B (AvHOURS)
1	0.019612	2.197089	24.267344	D (AVL)
Variables in Best Model PRORATE				
2	0.397758	1.742879	1.478793	CD
2	0.379520	1.769071	2.674304	AC
2	0.372216	1.779454	3.153145	BC
2	0.132520	2.091757	18.865919	AB
2	0.089968	2.142445	21.655309	AD

2	0.081822	2.152013	22.189283	BD

Variables in Best Model
PRORATE, AVL

3	0.403794	1.755665	3.083088	BCD
3	0.399226	1.762378	3.382545	ACD
3	0.382846	1.786242	4.456291	ABC
3	0.133547	2.116488	20.798558	ABD

Variables in Best Model
AvHOURS, PRORATE, AVL

4	0.405061	1.77614	5.000000	ABCD

Variables in Best Model
AvNO, AvHOURS, PRORATE, AVL

8) As another selection procedure we have applied the four criteria in ***The Stepwise Regression***, and obtained the results in Table (7.10b/FL). It seems reasonable from these results that the only selected variable to be included in the model is PRORATE, except in the case of using backward method where PRORATE and AVL were selected. They are almost the same as those obtained in the above discussions.

Table (7.10b/FL): Stepwise Regression Report

Forward

Iter. MaxR-Squared

No.	Action	Variable	R-Squared	Sqrt (MSE)	Other X's
0	Unchanged		0.000000	2.193003	0.000000
1	Added	PRORATE	0.369092	1.76251	0.000000
2	Unchanged		0.369092	1.76251	0.000000

List of Variables Selected

PRORATE

Backward

Iter. Max R-Squared

No.	Action	Variable	R-Squared	Sqrt (MSE)	Other X's
0	Unchanged		0.405061	1.77614	0.351655
1	Removed	AvNO	0.403794	1.755665	0.269302
2	Removed	AvHOURS	0.397758	1.742879	0.002290
3	Unchanged		0.397758	1.742879	0.002290

List of Variables Selected

PRORATE, AVL

Stepwise

Iter. Max R-Squared

No.	Action	Variable	R-Squared	Sqrt (MSE)	Other X's
0	Unchanged		0.000000	2.193003	0.000000
1	Added	PRORATE	0.369092	1.76251	0.000000
2	Unchanged		0.369092	1.76251	0.000000

List of Variables Selected

PRORATE

Min MSE

Iter. Max R-Squared

No.	Action	Variable	R-Squared	Sqrt (MSE)	Other X's
0	Unchanged		0.000000	2.193003	0.000000
1	Added	PRORATE	0.369092	1.76251	0.000000
2	Unchanged		0.369092	1.76251	0.000000

List of Variables Selected

PRORATE

9) A third selection procedure depends on the *t-test* of the coefficients of the full multiple regression model. Table (7.10c/FL) provides the computer output of the Multiple Regression Equation Section. It seems reasonable from this output that the estimated coefficient of PRORATE is the only significant coefficient in the model. This is clear from the p-values corresponding to the tests of the coefficients. However, R-squared is only about 40%, which is a low value.

Table (7.10c/FL): t-tests for Coefficients of Multiple Regression Model

Regression Equation Section

Independent Variable	**Regression Coefficient**	**Standard Error**	**T-Value (Ho: B=0)**	**Prob Level**	**Decision (5%)**	**Power (5%)**
Intercept	131.2914	11.98047	10.9588	0.000000	Reject Ho	1.000000
AvNO	0.8061154	2.796582	0.2883	0.774682	Accept Ho	0.059108
AvHOURS	-3.088235E-02	4.993085E-02	-0.6185	0.539840	Accept Ho	0.092643
PRORATE	-0.3071651	7.280807E-02	-4.2188	0.000142	Reject Ho	0.984337
AVL	-7.071456E-02	5.859827E-02	-1.2068	0.234790	Accept Ho	0.217730
R-Squared	0.405061					

Model
FUEL= 131.2914+.8061154*AvNO-3.088235E-02*AvHOURS-.3071651*PRORATE-7.071456E-02*AVL

(III) Residual Exploration

10) The next step is to do an error analysis of the residuals of the above model. This analysis includes a plot section, which reports the histogram, and the normal probability plot of the residuals. It also gives the scatter plots of residuals against each of row number, predicted values, and each of the independent variables. These plots may be used to explore the normality, the constant variance, the zero means, the outliers, and the randomness assumptions of the errors. It seems reasonable from the

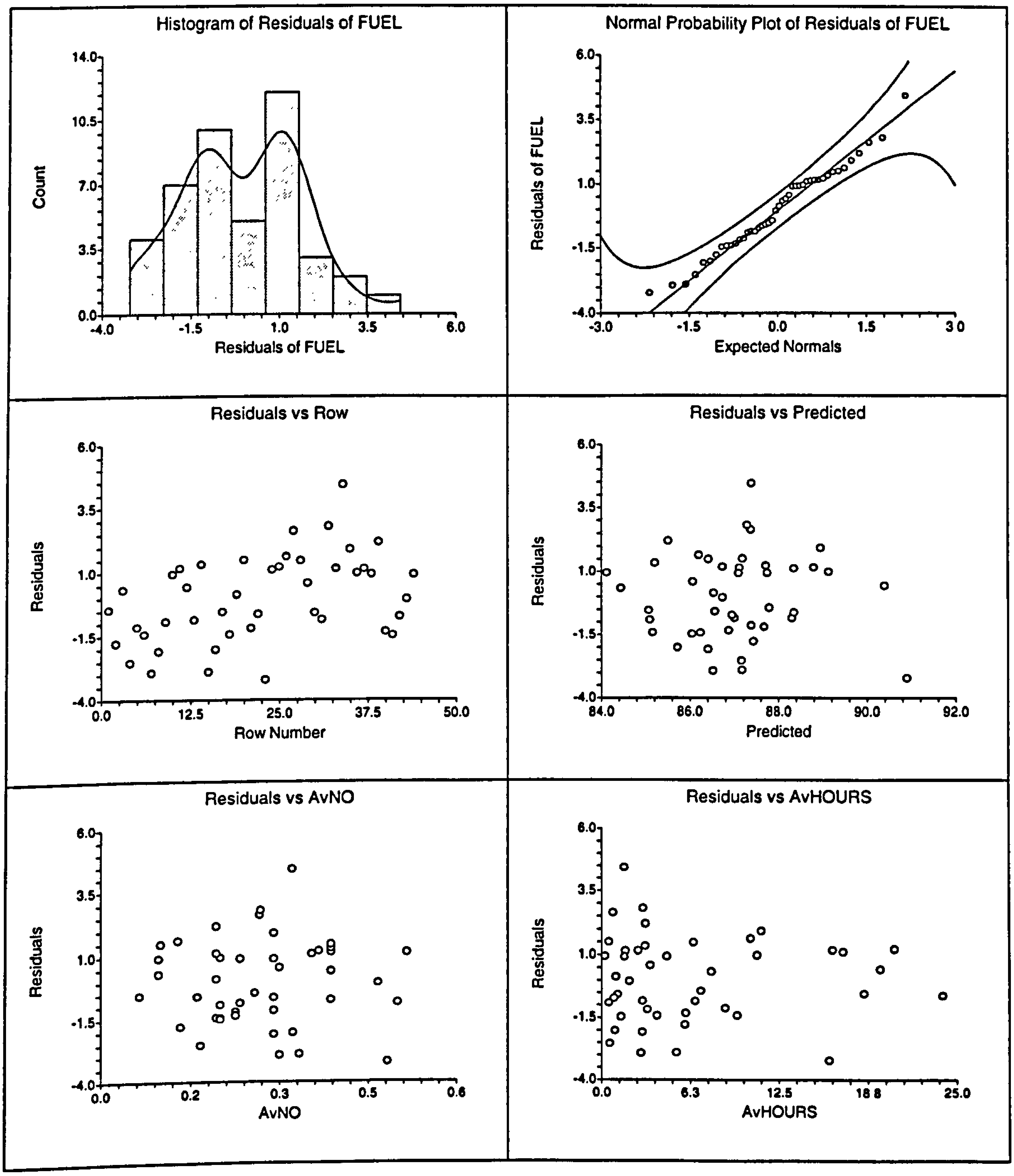
plots given in table (7.11/FL) that:

a) Since all points fall inside the confidence band of the normal probability plot. This means that there is a probability of 95% that the errors are normally distributed. The histogram shows that the distribution of the errors is bimodal and it lacks symmetry. This observation throws some shadow on the possibility to accept the normality assumption.

b) The scatter plot of the residuals against the row numbers shows that almost all the errors are negative for the first half of the data while they are positive for the last half. This may indicate that the errors are not random. Moreover, the errors range between -4 and 6, which is not symmetric about zero.

c) The scatter plots of the residuals against the predicted values show that the errors are high in the middle but they are low at the two extremes. The same remark is clear from the scatter plots of errors against AvNO and PRORATES. From the scatter plot of errors against AvHOURS, the errors are high for low values of AvHOURs and the errors are low for high values of AvHOURS. The reverse situation is seen in the scatter plot of errors against AVL. This indicates that the assumption that the errors have constant variance is not accepted.

Table (7.11/FL): Plot of Residuals



As a quantitative procedure, one should run a test for normality of the residuals. The following is the output of the normality section. It seems reasonable from this output that the three tests accepted the normality of the residuals. However, one can't assume a constant variance of the residuals.

Table (7.12FL): Normality Tests of Residuals

Assumption	Value	Probability	Decision(5%)
Skewness	0.4510	0.651989	Accepted
Kurtosis	-0.0824	0.934367	Accepted
Omnibus	0.2102	0.900242	Accepted

11) To have a quantitative test of the randomness (independence) of the residuals, we run the serial correlation (autocorrelation) procedure, which reports the serial correlations between the residuals together with the Durbin-Watson test. Table (7.13/FL) gives the computer output of this section. It seems reasonable from this output that the first two autocorrelations at lag one and at lag two are greater than the critical value 0.301511, which means that the errors are dependent. The same result is obtained based on the Durbin-Watson Value, which is 1.1826.

Table (7.13/FL): Serial-Correlation of Residuals

Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.404474	9	0.019641	17	-0.051785
2	0.317495	10	0.101855	18	-0.098527
3	0.088620	11	-0.054749	19	-0.130885
4	-0.045157	12	-0.055260	20	-0.166720
5	0.131361	13	-0.049385	21	-0.152852
6	0.093694	14	0.071947	22	-0.079243
7	0.204954	15	0.065890	23	-0.065535
8	0.225340	16	-0.076353	24	-0.031376

Above serial correlations significant if their absolute values are greater than 0.301511
Durbin-Watson Value 1.1826

(IV) Multicollinearity

12) The final screening process is to explore the multicollinearity in a quantitative way. It was observed on a previous paragraph that multicollinearity might not be a problem based on correlations between independent variables. Table (7.14) provides the computer out put of the multicollinearity section of the fitted model. It seems reasonable from this output that:

a) Since the R-squared vs. other X's which is the R-squared of regressing each independent variable on the remaining other independent variables and the maximum R-squared value is 0.351655 which is low, multicollinearity could not a problem.

b) The maximum value of the variance inflation factor is 1.542388, which is much less than 10, and then multicollinearity could not be a problem.

c) The same conclusion is obtained from eigenvalues and condition numbers.

Table (7.14): Multicollinearity Problem

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance	Diagonal of X'X Inverse
AvNO	1.333143	0.249893	0.750107	2.479138
AvHOURS	1.370257	0.270210	0.729790	7.902849E-04
PRORATE	1.542388	0.351655	0.648345	1.68037E-03
AVL	1.207923	0.172132	0.827868	1.088467E-03

Eigenvalues of Centered Correlations

No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.751481	43.79	43.79	1.00
2	1.182813	29.57	73.36	1.48
3	0.674164	16.85	90.21	2.60
4	0.391542	9.79	100.00	4.47

All Condition Numbers less than 100. Multicollinearity is NOT a problem.

7.5.3.2.2 Regression Analysis:

According to the above screening work, it seems reasonable to fit a regression model to the available data. So, the question is which is the most suitable procedure to fit that model. In this section we will use multiple regression, ridge regression, robust regression. Moreover, we will fit some nonlinear regression models.Then we select the model with highest R-squared value and the sensibility of the estimated coefficients of the model.

1) Multiple Regression

In the above screening process we reported a part of the output of the multiple regression procedure. Here we concentrate on both the full model and the model including only the influential independent variables that were selected in the screening process. Table (7.15a) reports the computer output of this procedure. It seems reasonable from this output that:

a) The R-squared value for the full model is 0.4051 while that of the model containing the only selected independent variable (PRORATE) is 0.369092. The difference between these two values is about 0.03. So, one may claim that the other three independent variables do not contribute to the variability in FUEL provided that the model is significant. However, due to the above discussion about the violation of some of the underlying assumptions, we prefer not to accept these results and search for some transformations of the data and/or apply other techniques of regression analysis other than the ordinary least square, method.

b) The sign of PRORATE which is the only significant independent variable in the models agrees with the researcher expectations.

c) The F-test in the ANOVA section indicates that the fitted model is significant.

d) The PRESS R-squared value is 0.2391, which means that this model can predict about 24% of the variability in FUEL. This percentage seems to be not useful to some extent.

e) Some of the assumptions on the model are not met. The violated assumptions are the existence of outliers, and the randomness of errors.

Table (7.15 a): Multiple Regression Equation Section of Full Model

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	131.2914	11.98047	107.0586	155.5242	0.0000
AvNO	0.8061154	2.796582	-4.850505	6.462736	0.0411
AvHOURS	-3.088235E-02	4.993085E-02	-0.131877	7.011233E-02	-0.0894
PRORATE	-0.3071651	7.280807E-02	-0.4544334	-0.1598969	-0.6471
AVL	-7.071456E-02	5.859827E-02	-0.1892407	4.781162E-02	-0.1638
T-Critical	2.022691				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	333120.3	333120.3			
Model	4	83.76599	20.9415	6.6382	0.000356	0.984430
Error	39	123.0322	3.154672			
Total(Adjusted)	43	206.7982	4.809261			

Root Mean Square Error	1.77614	R-Squared	0.4051
Mean of Dependent	87.01101	Adj R-Squared	0.3440
Coefficient of Variation	2.041282E-02	Press Value	157.3504
Sum |Press Residuals|	70.01398	Press R-Squared	0.2391

Table (7.15b) provides the output of the model with only one independent variable. The conclusions from this table are similar to these obtained from the full model.

Table (7.15 b): Multiple Regression Equation Section of the Single Selected Variables

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	122.4343	7.151221	17.1208	0.000000	Reject Ho	1.000000
PRORATE	-0.2883676	5.817511E-02	-4.9569	0.000012	Reject Ho	0.998017
R-Squared	0.369092					

Model
FUEL=122.4343-.2883676*PRORATE

2) Ridge Regression

The ridge regression is designed to fit a regression model if there is a suspicion that multicollinearity is a problem. In this section we apply this procedure to the full model.

Table (7.16) provides the computer output of this procedure. It seems reasonable from this output that:

a) All the assumptions on the model are met except possibly the existence of outliers. A minor problem in normality of the errors is seen in the full model. This is due to the existence of outliers. Comparing this situation with that of multiple regression, one observes that this ridge model is more useful than the multiple regression model.

b) The R-squared value of the full model is 0.4028 which is almost the same as that of the multiple regressions, which was 0.4051.

c) The signs of the estimated coefficients agree with the expectations of the researcher except that of AvHOURS. This may be due to the fact that the coefficient of AvHOURS is nearly zero.

d) The F-test in the ANOVA section indicates that the fitted model is significant.

Table (7.16): Ridge Regression Report of Full Model

Least Squares Multicollinearity Section

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance
AvNO	1.3331	0.2499	0.7501
AvHOURS	1.3703	0.2702	0.7298
PRORATE	1.5424	0.3517	0.6483
AVL	1.2079	0.1721	0.8279

Since all VIF's are less than 10, multicollinearity is not a problem.

Eigenvalues of Correlations

No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.751481	43.79	43.79	1.00
2	1.182813	29.57	73.36	1.48
3	0.674164	16.85	90.21	2.60
4	0.391542	9.79	100.00	4.47

All Condition Numbers less than 100. Multicollinearity is NOT a problem.

Ridge Regression Coefficient Section for k = 0.005000

Independent Variable	Regression Coefficient	Standard Error	Stand'zed Regression Coefficient	VIF
Intercept	130.8831			
AvNO	0.8397473	2.778252	0.0428	1.3108
AvHOURS	-2.983807E-02	0.0495877	-0.0864	1.3464
PRORATE	-0.3045999	7.219131E-02	-0.6417	1.5106
AVL	-6.988057E-02	5.826792E-02	-0.1619	1.1898

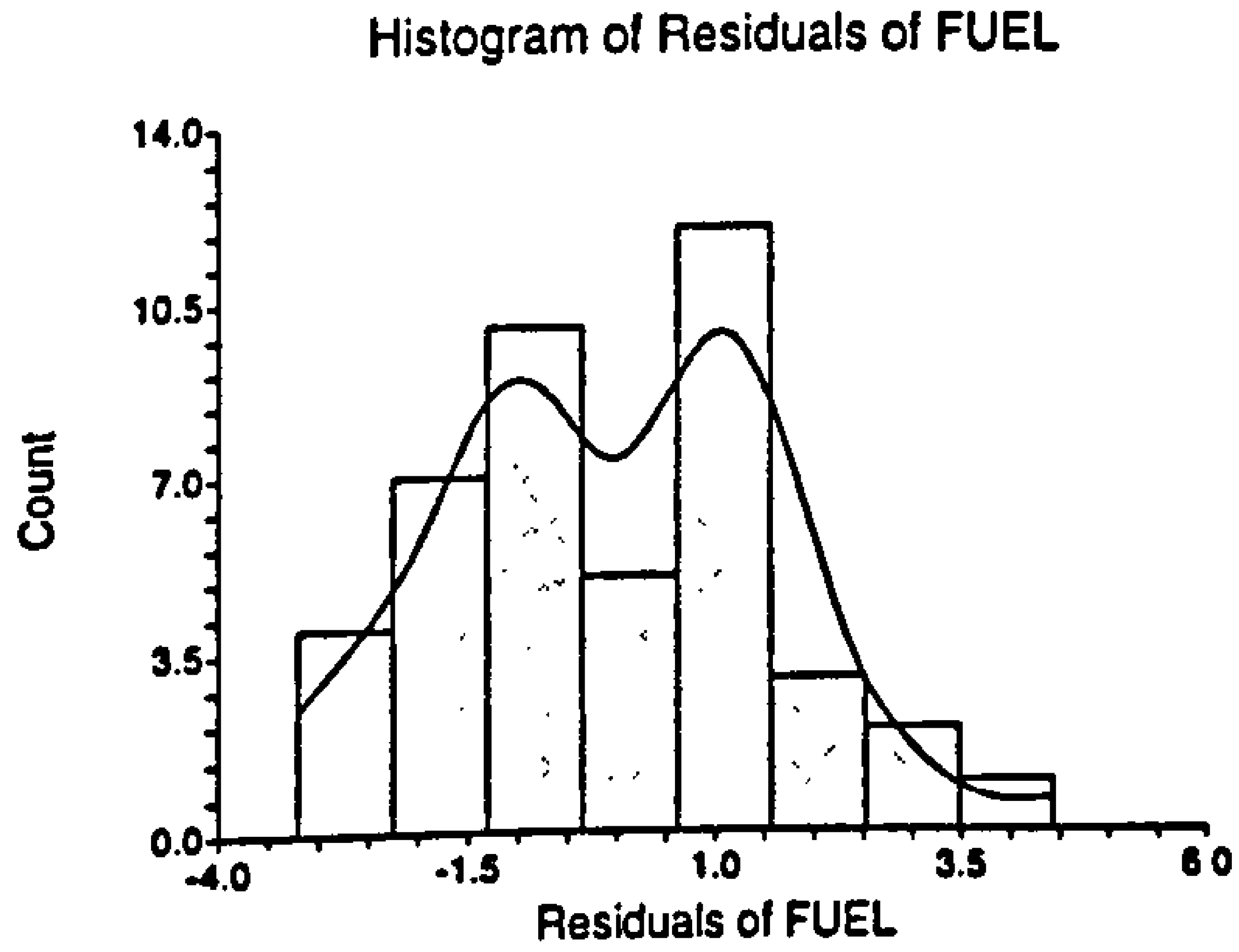
Model
FUEL=130.8831+.8397473*AvNO-2.983807E-02*AvHOURS-.3045999*PRORATE-6.988057E-02*AVL

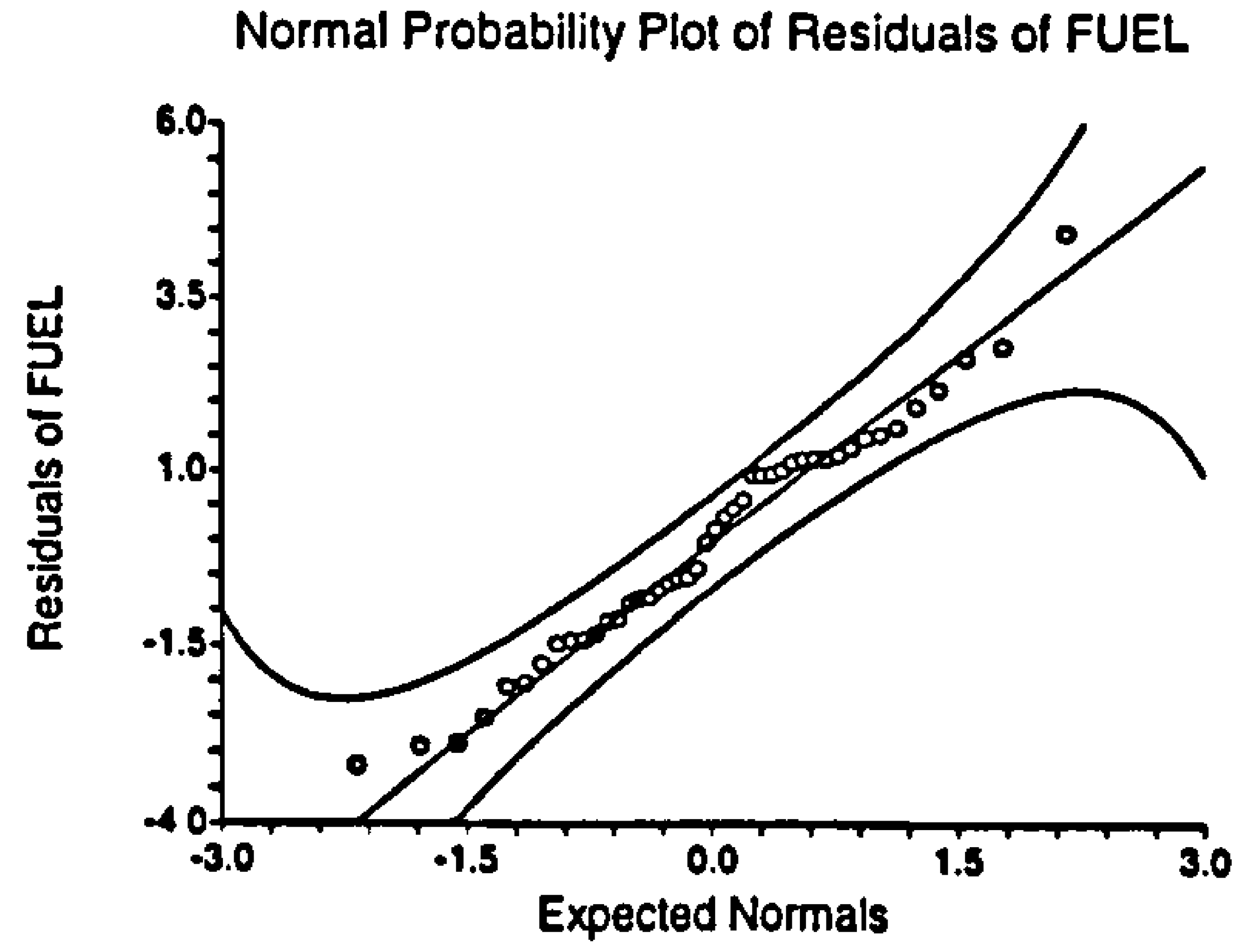
Analysis of Variance Section for k = 0.005000

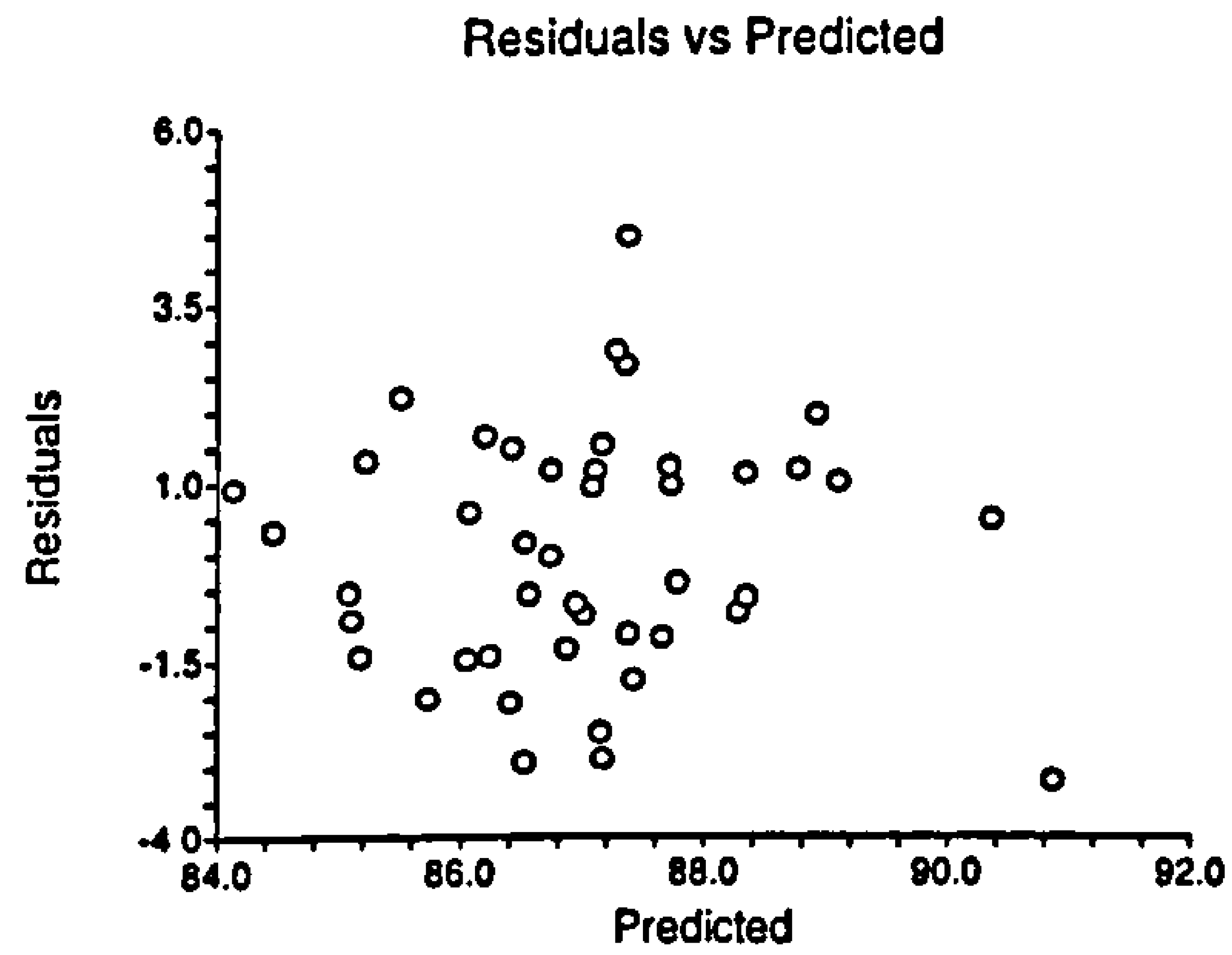
Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level
Intercept	1	333120.3	333120.3		
Model	4	83.29937	20.82484	6.5763	0.000381
Error	39	123.4988	3.166637		
Total(Adjusted)	43	206.7982	4.809261		

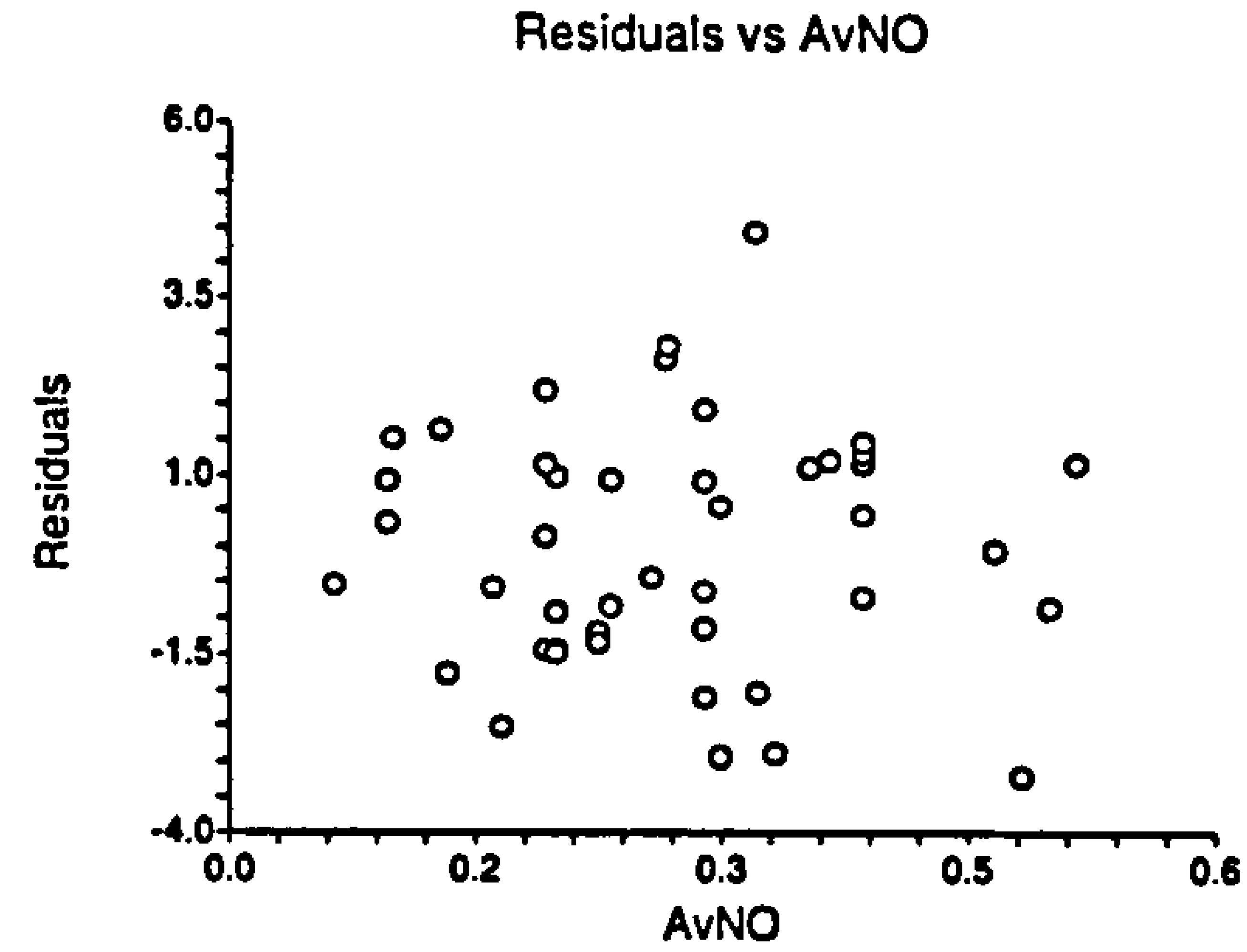
Mean of Dependent	87.01101
Root Mean Square Error	1.779505
R-Squared	0.4028
Coefficient of Variation	2.045149E-02

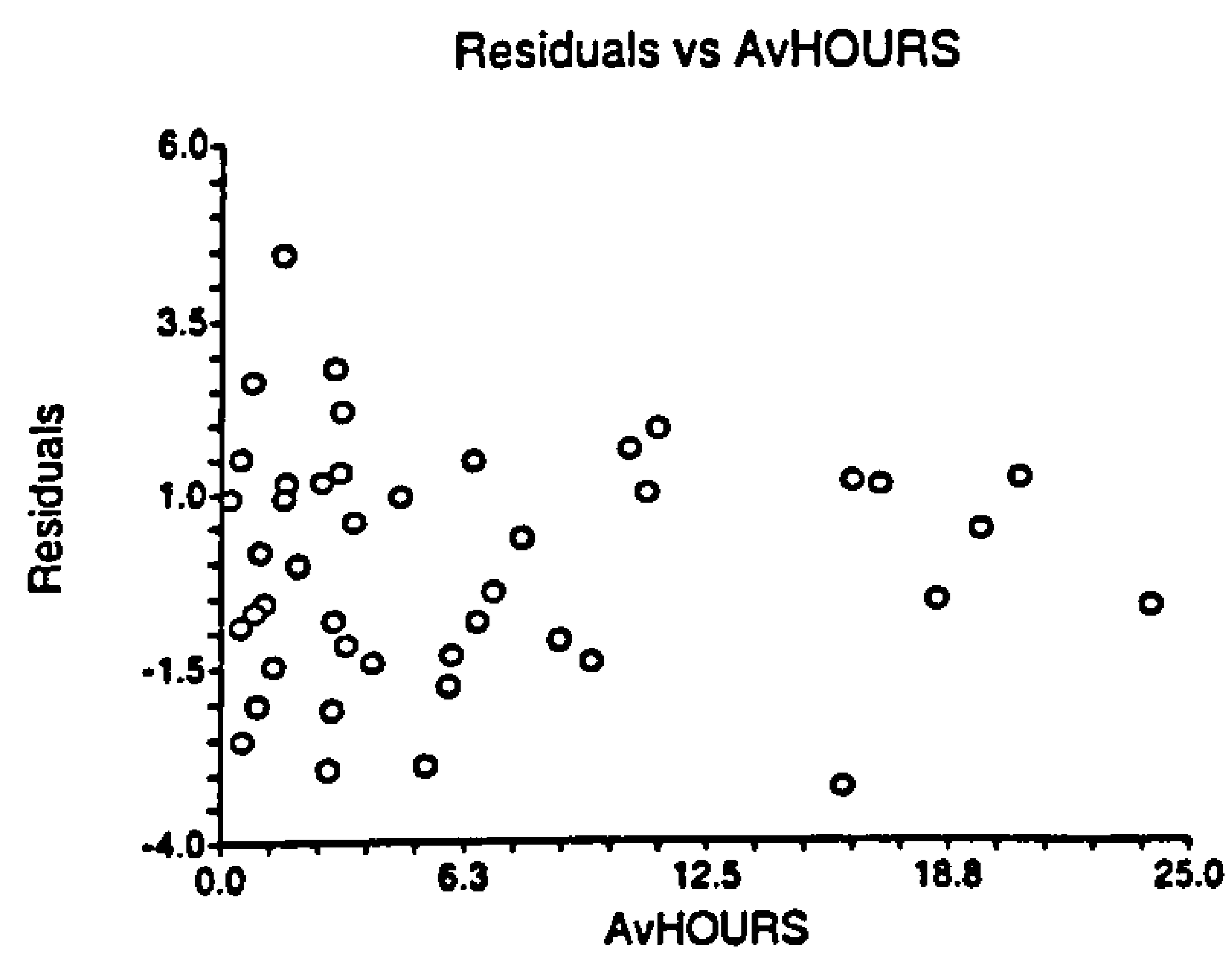
Residual Plots Section

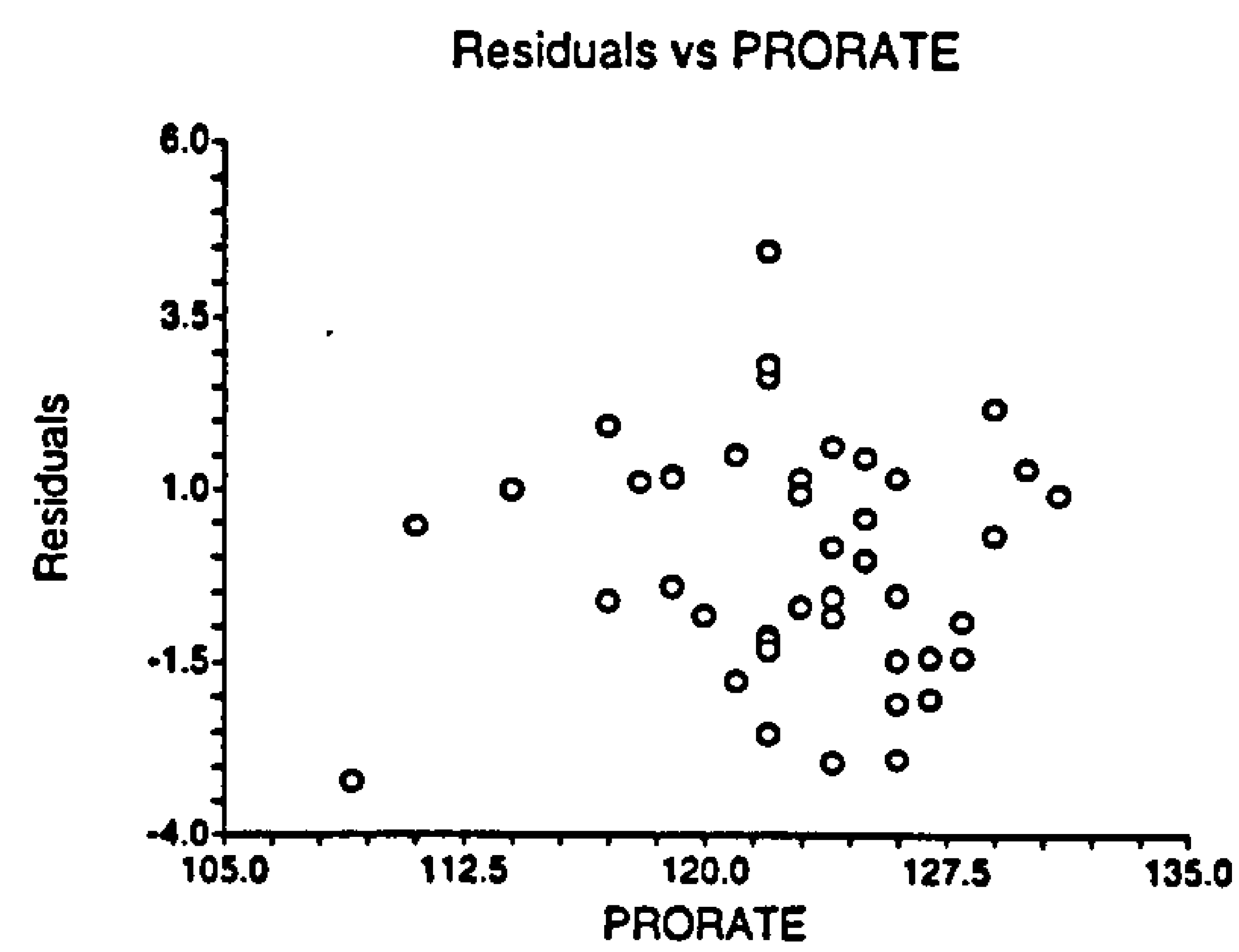


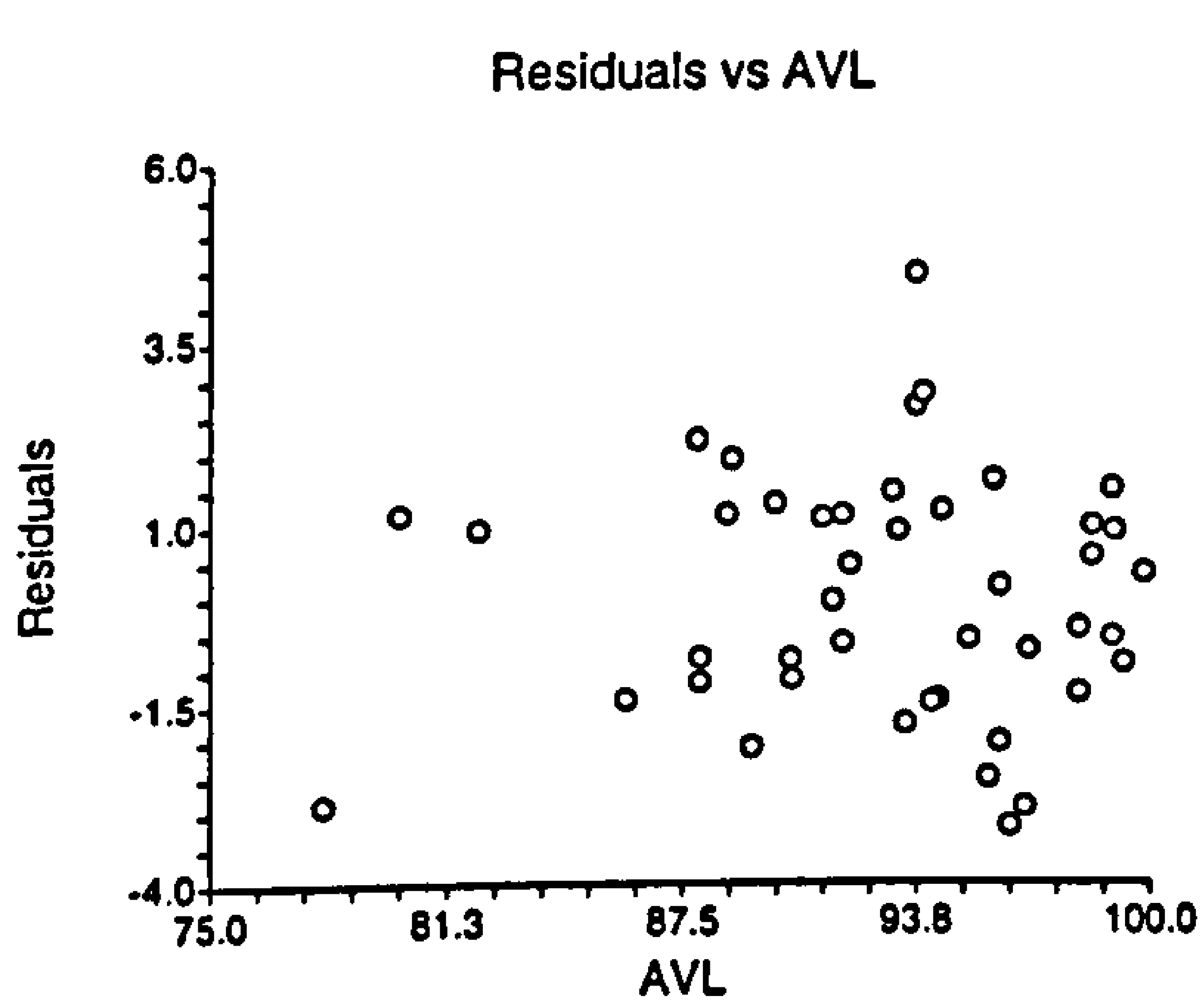



Residuals vs Predicted
6.0
3.5
1.0
-1.5
-4.0
84.0
86.0
88.0
90.0
92.0
Predicted
Residuals


Residuals vs AvNO
6.0
3.5
1.0
-1.5
-4.0
0.0
0.2
0.3
0.5
0.6
AvNO
Residuals


Residuals vs AvHOURS
6.0
3.5
1.0
-1.5
-4.0
0.0
6.3
12.5
18.8
25.0
AvHOURS
Residuals


Residuals vs PRORATE
6.0
3.5
1.0
-1.5
-4.0
105.0
112.5
120.0
127.5
135.0
PRORATE
Residuals


Residuals vs AVL
6.0
3.5
1.0
-1.5
-4.0
75.0
81.3
87.5
93.8
100.0
AVL
Residuals

3) Robust Regression

Robust regression is designed for situation in which normality is a problem and when the data has some outliers. In this section we apply this procedure. This procedure applies three weight functions, Andrew's Sine, Tukey's Biweight, and Least Absolute Deviation.

1- Andrew's Sine

Table (7.17a) provides the computer output for the full model. From this output It seems reasonable that:

a) The only influential independent variable is PRORATE.

b) The sign of the estimated coefficient of PRORATE agrees with the expectations of the researcher.

c) The R-squared value is 0.592504, which is higher than that of both the multiple regression model and that of the ridge regression model.

d) The F-test in the ANOVA table indicates that the model is significant.

Table (7.17a): Robust Regression Report for the Full Model Using Andrew's Sine

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	133.3905	8.905165	14.9790	0.000000	Reject Ho	1.000000
AvNO	1.999712	2.096263	0.9539	0.345989	Accept Ho	0.153576
AvHOURS	-1.0394E-02	3.639963E-02	-0.2856	0.776726	Accept Ho	0.058938
PRORATE	-0.3304487	5.623593E-02	-5.8761	0.000001	Reject Ho	0.999917
AVL	-6.732384E-02		4.633835E-02		-1.4529	0.154258
	Accept Ho	0.294007				
R-Squared	0.592504					

Model

FUEL=133.3905+1.999712*AvNO-1.039429E-02*AvHOURS-.3304487*PRORATE-6.732384E-02*AVL

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	133.3905	8.905165	115.3781	151.4029	0.000000
AvNO	1.999712	2.096263	-2.240381	6.239804	0.111813
AvHOURS	-1.039429E-02	3.639963E-02	-8.401949E-02	6.323091E-02	-0.033581
PRORATE	-0.3304487	5.623593E-02	-0.4441966	-0.2167008	-0.711767
AVL	-6.732384E-02	4.633835E-02	-0.161052	2.640432E-02	-0.165054
T-Critical	2.022691				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	252376.2	252376.2			
Model	4	80.15228	20.03807	14.1766	0.000000	0.999992
Error	39	55.12497	1.413461			
Total(Adjusted)	43	135.2773	3.145983			

Root Mean Square Error	**1.188891**	**R-Squared**	**0.592504**
Mean of Dependent Variable	**86.95525**	**Adj R-Squared**	**0.550709**
Coefficient of Variation	**1.367244E-02**		

Table (7.17b) reports the computer output for the model with the selected independent variable. From this output It seems reasonable that

a) The sign of the estimated coefficients of PRORATE agrees with the researcher expectations
b) The R-squared value is 0.569486. The difference between this value and that of the full model is about 0.02, which really means that the other three variables do not contribute much to the variability in FUEL.
c) The F-test in the ANOVA table indicates that the model is significant.

Table (7.17b): Robust Regression Report For the only Selected Variable Using Andrew's Sine

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	130.7643	5.892294	22.1924	0.000000	Reject Ho	1.000000
PRORATE	-0.3568497	4.787545E-02	-7.4537	0.000000	Reject Ho	1.000000
R-Squared	0.569486					

Model
FUEL= 130.7643-.3568497*PRORATE

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	130.7643	5.892294	118.8732	142.6554	0.000000
PRORATE	-0.3568497	4.787545E-02	-0.4534663	-0.2602332	-0.754643
T-Critical	2.018082				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	254291.7	254291.7			
Model	1	79.4075	79.4075	55.5578	0.000000	1.000000
Error	42	60.02966	1.429278			
Total(Adjusted)	43	139.4372	3.242725			

Root Mean Square Error	**1.195524**	**R-Squared**	**0.569486**
Mean of Dependent Variable	**86.87169**	**Adj R-Squared**	**0.559236**
Coefficient of Variation	**1.376195E-02**		

2- Tukey's Biweight

Table (17.7c) provides the computer output for the full model. From this output It seems reasonable that:

a) The only influential independent variables is PRORATE.

The sign of the estimated coefficients of PRORATE agrees with the expectations of the researcher.

b) The R-squared value 0.472987 which is still higher than that of the multiple regression model.

c) The F-test in the ANOVA table indicates that the model is significant.

Table (7.17c): Robust Regression Report for the Full Model Using Tukey's Biweight

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	131.6169	10.59781	12.4193	0.000000	Reject Ho	1.000000
AvNO	0.934507	2.479209	0.3769	0.708263	Accept Ho	0.065631
AvHOURS	-1.748374E-02	4.401718E-02	-0.3972	0.693384	Accept Ho	0.067373
PRORATE	-0.3084482	6.523722E-02	-4.7281	0.000029	Reject Ho	0.995962
AVL	-7.412268E-02	5.300346E-02	-1.3984	0.169880	Accept Ho	0.276080
R-Squared	0.472987					

Model

FUEL= 131.6169+ .934507*AvNO-1.748374E-02*AvHOURS-.3084482*PRORATE-7.412268E-02*AVL

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	131.6169	10.59781	110.1808	153.053	0.000000
AvNO	0.934507	2.479209	-4.080167	5.949181	0.050387
AvHOURS	-1.748374E-02	4.401718E-02	-0.1065169	7.154942E-02	-0.053774
PRORATE	-0.3084482	6.523722E-02	-0.440403	-0.1764935	-0.670319
AVL	-7.412268E-02	5.300346E-02	-0.1813323	3.308694E-02	-0.179034
T-Critical	2.022691				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	297596.1	297596.1			
Model	4	79.85985	19.96496	8.7505	0.000038	0.997755
Error	39	88.98173	2.281583			
Total(Adjusted)	43	168.8416	3.926548			

Root Mean Square Error	1.510491	R-Squared	0.472987
Mean of Dependent Variable	86.94908	Adj R-Squared	0.418934
Coefficient of Variation	1.737213E-02		

Table (7.17d) gives the computer output for the model with the selected independent variable. From this output It seems reasonable that:

a) The influential independent variable is PRORATE. The sign of the coefficient of the estimated parameter of this influential variable agrees with the expectations of the researcher.

b) The R-squared value 0.428325, which differs from that of the full model by about 0.04. This is still leading to the same conclusion that the other three independent variables do not contribute much to the variability in FUEL.

c) The F-test in the ANOVA table indicates that the model is significant.

Table (7.17d): Robust Regression Report for the Model with the Only Selected Variable Using Tukey's Biweight

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	124.1656	6.640089	18.6994	0.000000	Reject Ho	1.000000
PRORATE	-0.3028438	5.398608E-02	-5.6097	0.000001	Reject Ho	0.999782
R-Squared	0.428325					

Model
FUEL=124.1656-.3028438*PRORATE

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	124.1656	6.640089	110.7654	137.5659	0.000000
PRORATE	-0.3028438	5.398608E-02	-0.4117922	-0.1938955	-0.654466
T-Critical	2.018082				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	301774.7	301774.7			
Model	1	74.31618	74.31618	31.4683	0.000001	0.999782
Error	42	99.18793	2.361617			
Total(Adjusted)	43	173.5041	4.034979			

Root Mean Square Error	1.536755	R-Squared	0.428325
Mean of Dependent Variable	86.94196	Adj R-Squared	0.414714
Coefficient of Variation	1.767565E-02		

3- Least Absolute Deviation

Table (7.17e) provides the computer output for the full model. From this output It seems reasonable that:

a) The only influential independent variable is PRORATE.

b) The sign of the estimated coefficient of PRORATE agrees with the expectations of the researcher.

c) The R-squared value is 0.535717, which is higher than that of the multiple regression models.

d) The F-test in the ANOVA table indicates that the model is significant.

Table (7.17e): Robust Regression Report for the Full Model Using Least Absolute Deviation

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	131.6548	9.51674	13.8340	0.000000	Reject Ho	1.000000
AvNO	1.565402	2.232305	0.7012	0.487311	Accept Ho	0.105100
AvHOURS	-1.61358E-02	3.901833E-02	-0.4135	0.681472	Accept Ho	0.068847
PRORATE	-0.3102915	5.929102E-02	-5.2334	0.000006	Reject Ho	0.999158
AVL	-0.0736572	4.866323E-02	-1.5136	0.138187	Accept Ho	0.314646
R-Squared	0.535717					

Model
FUEL=131.6548+1.565402*AvNO-1.613586E-02*AvHOURS-.3102915*PRORATE .0736572*AVL

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	131.6548	9.51674	112.4054	150.9042	0.000000
AvNO	1.565402	2.232305	-2.949861	6.080665	0.088135
AvHOURS	-1.61358E-02	3.901833E-02	-9.50578E-02	6.278616E-02	0.052294
PRORATE	-0.3102915	5.929102E-02	-0.4302189	-0.1903641	-0.688856
AVL	-0.0736572	4.866323E-02	-0.1720879	2.477347E-02	-0.182533
T-Critical	2.022691				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	276875.8	276875.8			
Model	4	79.11857	19.77964	11.2501	0.000004	0.999813
Error	39	68.56873	1.758173			
Total(Adjusted)	43	147.6873	3.434588			

Root Mean Square Error	**1.325961**	**R-Squared**	**0.535717**
Mean of Dependent Variable	**86.97614**	**Adj R-Squared**	**0.488098**
Coefficient of Variation	**1.524511E-02**		

Table (7.17f) provides the computer output for the model with the selected independent variable. From this output It seems reasonable that:

a) The influential independent variable is PRORATE.

b) The sign of the coefficient of the estimated parameter agrees with the expectations of the researcher.

c) The R-squared value 0.501982 which differs from that of the full model by 0.03, i.e. the other three independent variables do not contribute much to the variability in FUEL.

d) The F-test in the ANOVA table indicates that the model is significant.

Table (7.17f): Robust Regression Report for Model with the Only Selected Variable Using Least Absolute Deviation

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	126.9424	6.15933	20.6098	0.000000	Reject Ho	1.000000
PRORATE	-0.325762	5.006734E-02	-6.5065	0.000000	Reject Ho	0.999994
R-Squared	0.501982					

Model
FUEL= 126.9424-.325762*PRORATE

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	126.9424	6.15933	114.5124	139.3724	0.000000
PRORATE	-0.325762	5.006734E-02	-0.426802	-0.224722	-0.708507
T-Critical	2.018082				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	277486	277486			
Model	1	76.26131	76.26131	42.3342	0.000000	0.999994
Error	42	75.65919	1.801409			
Total(Adjusted)	43	151.9205	3.533035			

Root Mean Square Error	1.342166	R-Squared	0.501982
Mean of Dependent Variable	86.89274	Adj R-Squared	0.490124
Coefficient of Variation	1.544624E-02		

4) Nonlinear Regression Models

1) Quadratic Regression Model

One may fit a quadratic regression model. Table (7.18a) provides the computer output of the fitted model. It seems reasonable from this output that:

a) The R-squared for the fitted model is 0.433526, which differs from that of the linear multiple regression models by about 0.03. This increase in R-squared value does not worth complicating the model.

b) At the 0.05 level of significance, the confidence intervals that contain the number zero show the significant terms. All the variables are significant. However, the signs of the estimated coefficients contradict the results of other models as well as the researcher expectations. So, we cannot recommend this model.

Table (7.18a): Nonlinear Regression Report

Model Estimation Section

Parameter Name	Parameter Estimate	Asymptotic Standard Error	Lower 95% C.L.	Upper 95% C.L.
A	-39.52533	149.3095	-342.6397	263.5891
B	-4.925167E-02	12.38361	-25.18931	25.09081
C	-3.076755E-02	0.1644164	-0.3645505	0.3030154
D	1.399583	2.303035	-3.275826	6.074992
E	1.40041	1.566206	-1.779156	4.579976
F	0.8881106	20.26169	-40.2453	42.02153
G	8.753253E-04	7.480506E-03	-1.431091E-02	1.606156E-02
H	-6.961847E-03	9.484698E-03	-2.621681E-02	1.229311E-02
I	-8.113177E-03	8.683505E-03	-2.574163E-02	9.515278E-03

Model FUEL =
A+B*AVNO+C*AVHOURS+D*PRORATE+E*AVL+F*AVNO^2+G*AVHOURS^2
+H*PRORATE^2+I*AVL^2
R-Squared 0.433526
Iterations 12
Estimated Model
FUEL=(-39.52533)+(-4.925167E-02)*(AVNO)+(-3.076755E-02)*(AVHOURS)+(1.399583)*(PRORATE)+(1.40041)*(AVL)+(.8881106)*(AVNO)^2
+(8.753253E-04)*(AVHOURS)^2+(-6.961847E-03)*(PRORATE)^2+(-8.113177E-03)*(AVL)^2

Analysis of Variance Table

Source	DF	Sum of Squares	Mean Square

Mean	1	333120.3	333120.3
Model	9	333210	37023.33
Model (Adjusted)	8	89.65248	11.20656
Error	35	117.1457	3.347021
Total (Adjusted)	43	206.7982	
Total	44	333327.1	

2) <u>Polynomial Regression Model with Interaction</u>

One may fit a quadratic regression model with interactions. Table (7.18b) reports the computer output of **Response-Surface Regression** procedure, which searches for the optimal model including up to cubic terms with interactions. It seems reasonable from this output that:

i. The R-squared for the fitted model is 0.484707, which is higher than that of the linear multiple regression model.

ii. At the 0.05 level of significance, the F-tests show that the model is useful. Moreover, the linear, and linear by linear interaction terms that are involved in the model are significant. However, the quadratic term is not significant. This may explain the contradictory result, which was obtained by the quadratic regression in the previous paragraph.

iii. At the 0.05 level of significance, the only significant linear term is PRORATE. However, at 0.10 level of significance the PRORAT and AVL are the two linear significant terms.

iv. The estimated model is

v. FUEL = -62.77656+ 173.7997*AvNO-4.638486E-02*AvHOURS

.3348687*PRORATE+ 3.767953*AVL-29.77582*AvNO^2-1.864908E-02*AVL^2-

1.679568*AvNO*AVL

e) The one-degree of freedom breakdown ANOVA shows that, at the 0.05 level of significance the only significant terms are AvNO, PRORATE, AVL, and the interaction between AvNO and AVL.

Table (7.18b) : Response-Surface Regression Report

Model Summary Section

Number of Terms Removed	7
Number of Terms Remaining	7
R-Squared Cutoff Value	0.010000
R-Squared of Final Model	0.484707

Sequential ANOVA Section

Source	Df	Sequential Sum-Squares	Mean Square	F-Ratio	Prob Level	Incremental R-Squared
Regression	7	100.2366	14.31952	4.84	0.000644	0.484707
Linear	4	83.76599	20.9415	7.07	0.000261	0.405061
Quadratic	2	4.026907	2.013453	0.68	0.512901	0.019473
Lin x Lin	1	12.44374	12.44374	4.20	0.047667	0.060173
Total Error	36	106.5616	2.960044			0.515293

ANOVA Section

Factor	Df	Last Sum-Squares	Mean Square	F-Ratio	Prob Level	Term R-Squared
AvNO	3	12.49244	4.164148	1.41	0.256642	0.060409
AvHOURS	1	2.473184	2.473184	0.84	0.366762	0.011959
PRORATE	1	57.82119	57.82119	19.53	0.000087	0.279602
AVL	3	19.99782	6.66594	2.25	0.098955	0.096702
Total Error	36	106.5616	2.960044			0.515293

Estimation Section

Parameter	df	Regression Coefficient	Standard Error	T-Ratio	Prob Level	Last R-Squared
Intercept	1	-62.77656				
AvNO	1	173.7997	84.82743	2.05	0.047819	0.060086
AvHOURS	1	-4.638486E-02	5.074545E-02	-0.91	0.366762	0.011959
PRORATE	1	-0.3348687	7.576702E-02	-4.42	0.000087	0.279602
AVL	1	3.767953	1.799447	2.09	0.043368	0.062760
AvNO^2	1	-29.77582	22.92197	-1.30	0.202200	0.024153
AVL^2	1	-1.864908E-02	9.276585E-03	-2.01	0.051933	0.057848
AvNO*AVL	1	-1.679568	0.8191649	-2.05	0.047667	0.060173

Model
FUEL=-62.77656+ 173.7997*AvNO-4.638486E-02*AvHOURS-.3348687*PRORATE+ 3.767953*AVL-29.77582*AvNO^2-1.864908E-02*AVL^2-1.679568*AvNO*AVL

3) <u>Logarithmic Model</u>

The following output is the result of multiple regression of Ln (FUEL) as a function of the logarithm of each of the independent variables. It seems reasonable from this output that:

a) The only significant independent variable is ln (PRORATE).

b) The R-squared value is 0.400893, which is almost the same as that of the linear multiple regression models.

c) The F-test accepts the model.

d) The PRESS R-squared value is 0.2249, which is too low as a predictability percentage.

e) Multicolinearity could not be a problem.

f) There are some outliers in the data.

g) It is interesting to notice that this model has almost the same properties of the linear multiple regression models. It suffers from the violation of some of the underlying assumptions.

Table (7.18c): Multiple Regression Report of Logarithmic Model

Regression Equation Section

Independent		Regression		Standard		T-Value
	Prob	Decision	Power			
Variable	Coefficient	Error	(Ho: B=0)	Level	(5%)	(5%)
Intercept	6.85005	0.6376491	10.7427	0.000000	Reject Ho	1.000000
LNNO	2.658619E-03	7.834291E-03	0.3394	0.736162	Accept Ho	0.062649
LNHR	-2.227411E-03	3.142852E-03	-0.7087	0.482712	Accept Ho	0.106308
LNPR	-0.4213629	9.993252E-02	-4.2165	0.000143	Reject Ho	0.984246
LNAV	-7.745749E-02	6.487729E-02	-1.1939	0.239726	Accept Ho	0.214102
R-Squared		0.400893				

Model
FUEL= 6.85005+ 2.658619E-03*LNNO-2.227411E-03*LNHR-.4213629*LNPR-7.745749E-02*LNAV

Regression Coefficient Section
Independent Regression Standard Lower Upper
Standardized

Variable	Coefficient	Error	95% C.L.	95% C.L.	Coefficient
Intercept	6.85005	0.6376491	5.560283	8.139817	0.0000
LNNO	2.658619E-03	7.834291E-03	-1.318773E-02	1.850497E-02	0.0500
LNHR	-2.227411E-03	3.142852E-03	-8.584429E-03	4.129607E-03	-0.1034
LNPR	-0.4213629	9.993252E-02	-0.6234955	-0.2192303	-0.6426
LNAV	-7.745749E-02	6.487729E-02	-0.2086842	5.376923E-02	-0.1738
T-Critical	2.022691				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	877.479	877.479			
Model	4	1.088824E-02	2.72206E-03	6.5242	0.000404	0.982805
Error	39	1.627169E-02	4.172229E-04			
Total(Adjusted)	43	2.715993E-02	6.316263E-04			

Root Mean Square Error	2.042603E-02	R-Squared	0.4009
Mean of Dependent	4.465725	Adj R-Squared	0.3394
Coefficient of Variation	4.573957E-03	Press Value	2.105044E-02
Sum |Press Residuals|	0.8007173	Press R-Squared	0.2249

Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	0.3030	0.761882	Accepted
Kurtosis	-0.1272	0.898793	Accepted
Omnibus	0.1080	0.947436	Accepted

Serial-Correlation Section

Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.423070	9	0.028385	17	-0.072905
2	0.322494	10	0.084870	18	-0.104944
3	0.113148	11	-0.043942	19	-0.130811
4	-0.017411	12	-0.052771	20	-0.176798
5	0.145143	13	-0.051074	21	-0.157434
6	0.130028	14	0.064046	22	-0.105097
7	0.227618	15	0.050701	23	-0.083922
8	0.243795	16	-0.097843	24	-0.058195

Above serial correlations significant if their absolute values are greater than 0.301511

Durbin-Watson Value 1.1460

Multicollinearity Section

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance	Diagonal of X'X Inverse
LNNO	1.411099	0.291333	0.708667	0.1471063
LNHR	1.384896	0.277924	0.722076	2.367444E-02
LNPR	1.511786	0.338531	0.661469	23.93567
LNAV	1.380231	0.275483	0.724517	10.08828

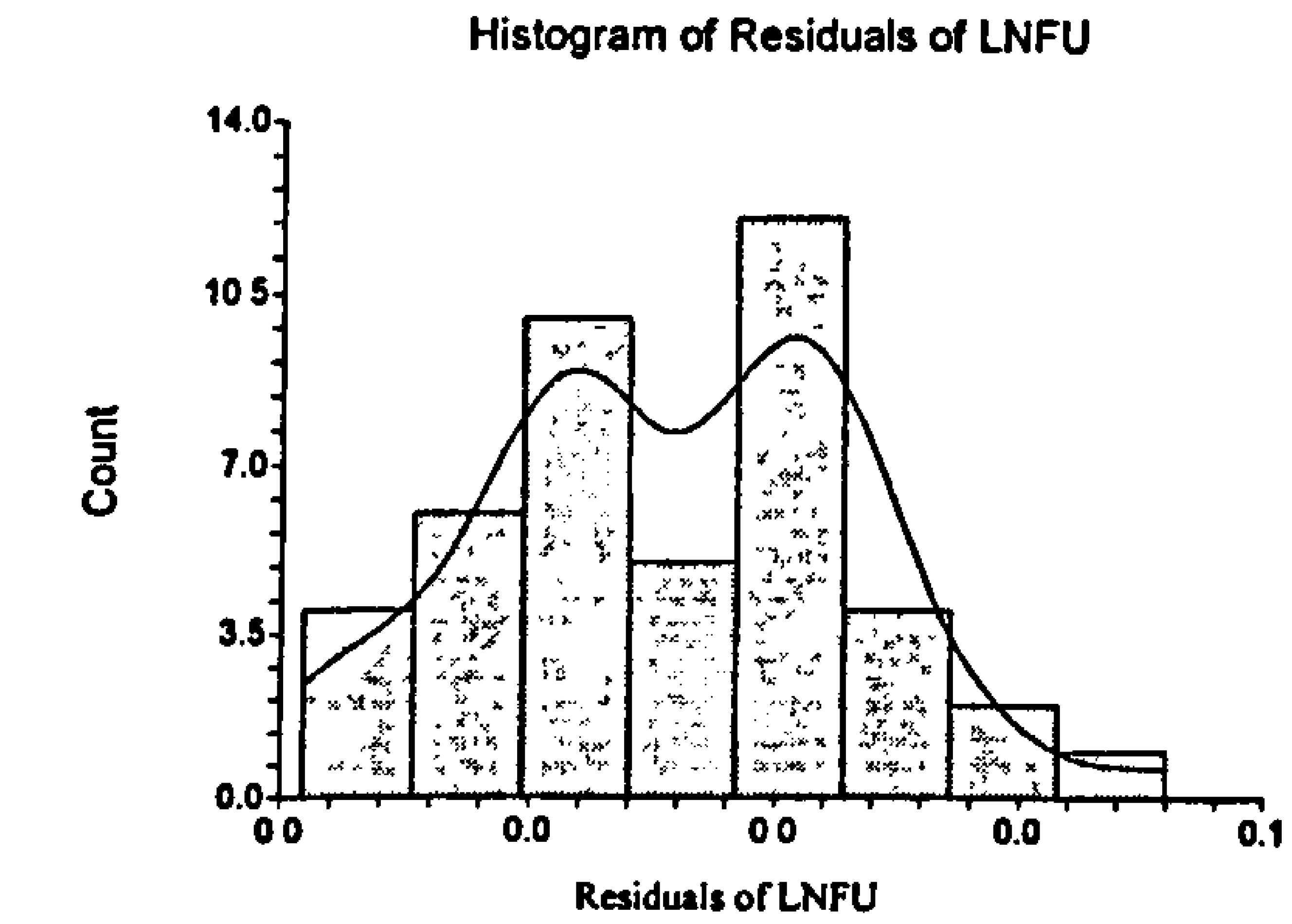
Eigenvalues of Centered Correlations

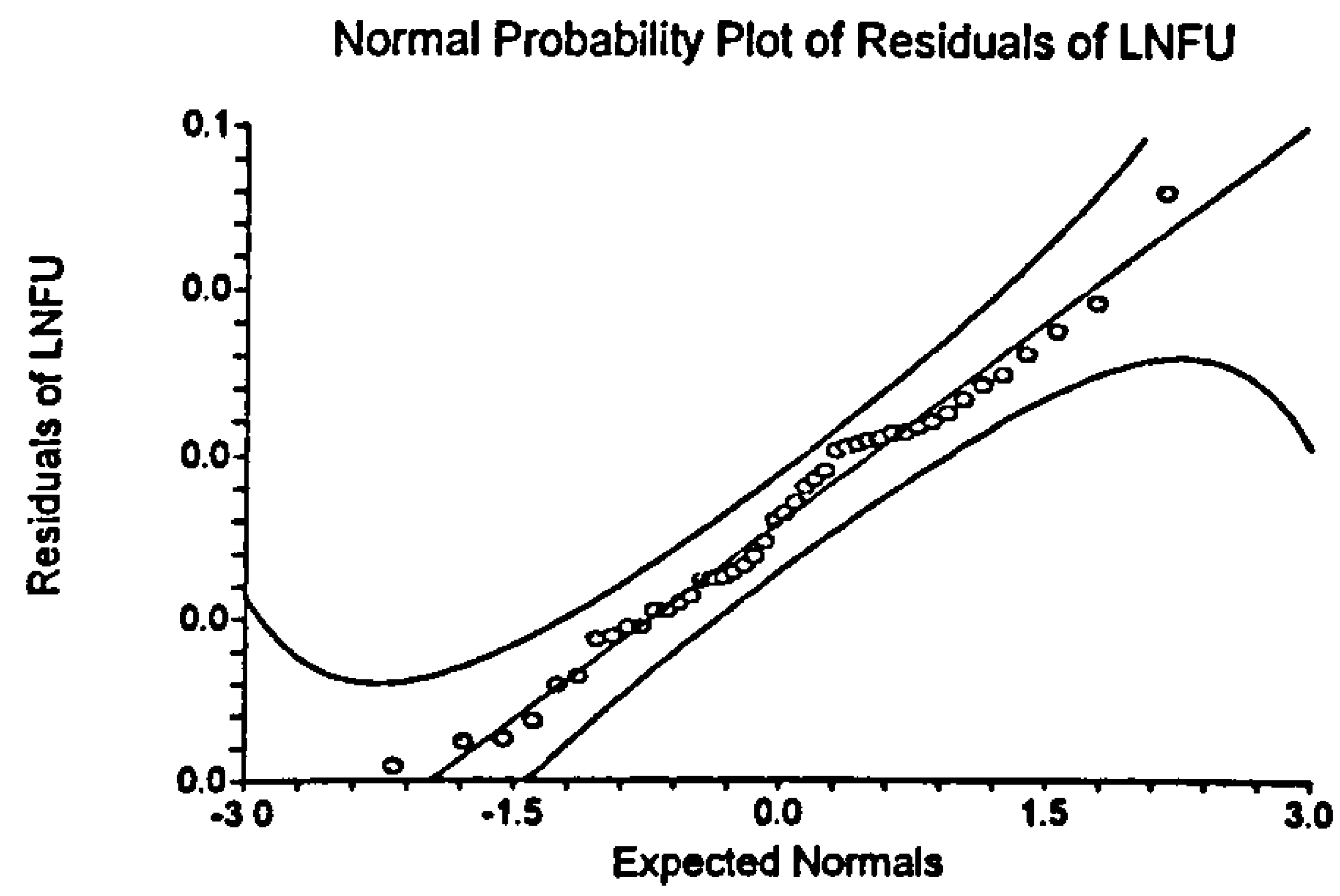
No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number

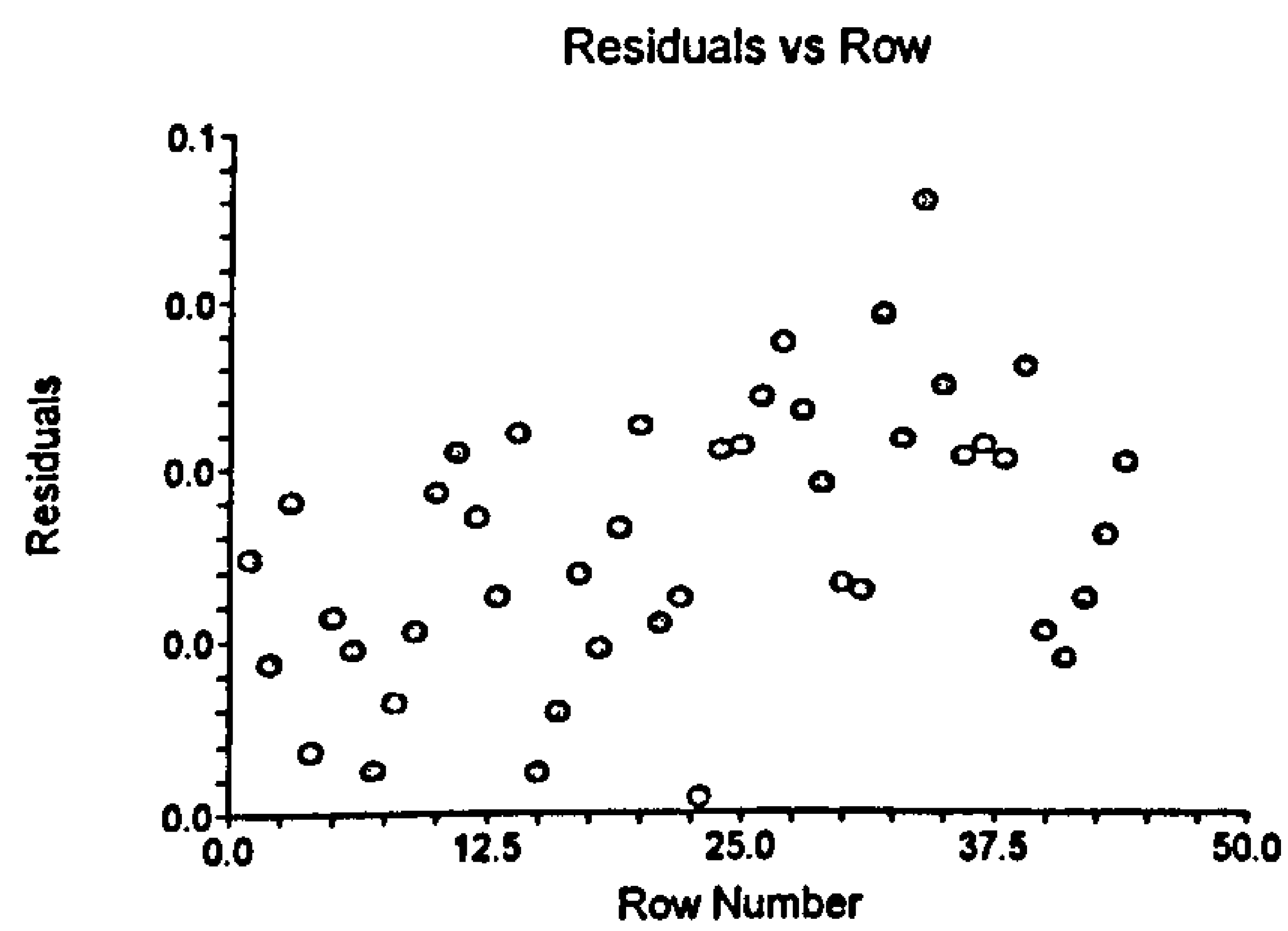
1	1.811494	45.29	45.29	1.00
2	1.111304	27.78	73.07	1.63
3	0.727816	18.20	91.27	2.49
4	0.349386	8.73	100.00	5.18

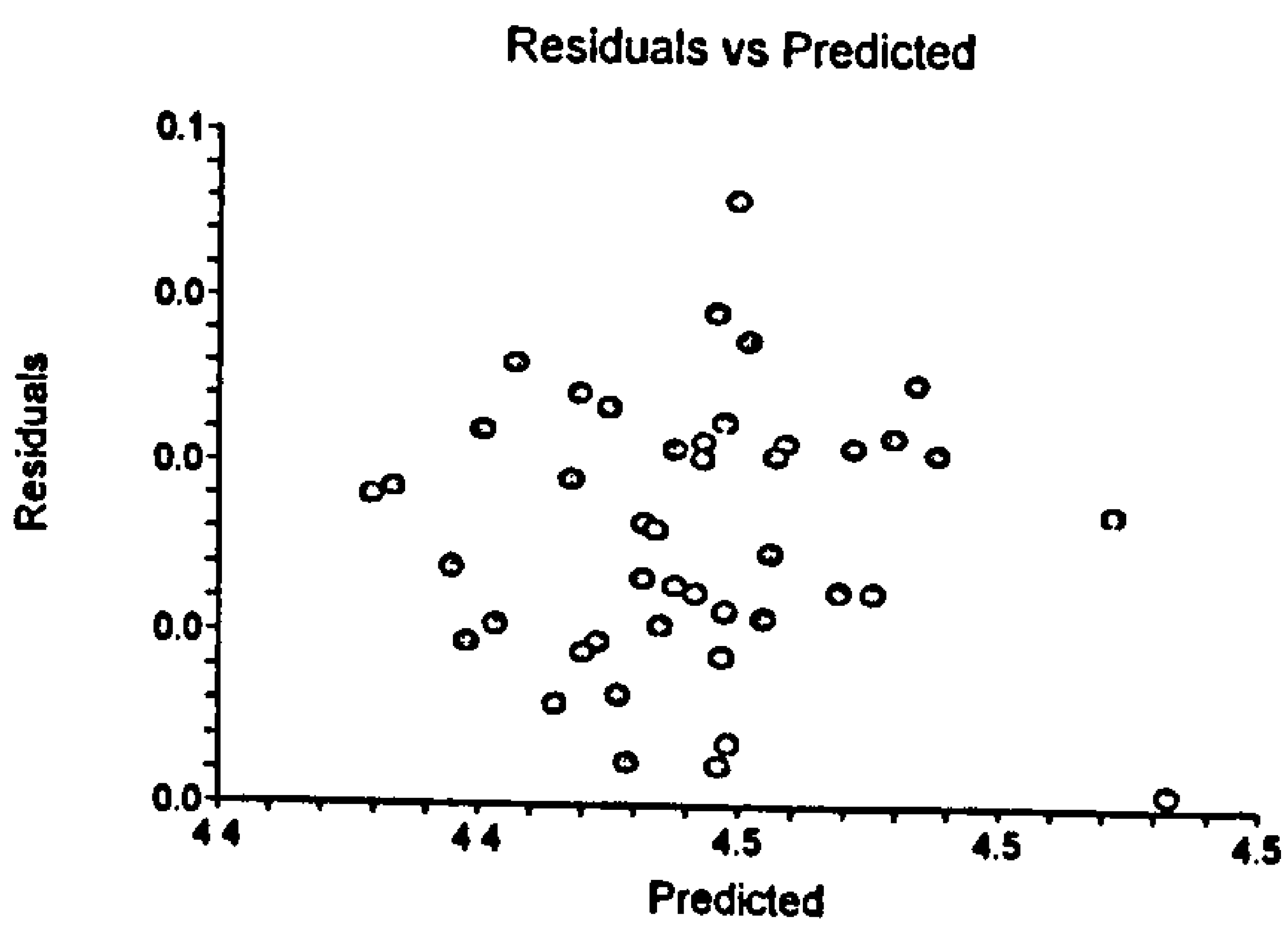
All Condition Numbers less than 100. Multicollinearity is NOT a problem.

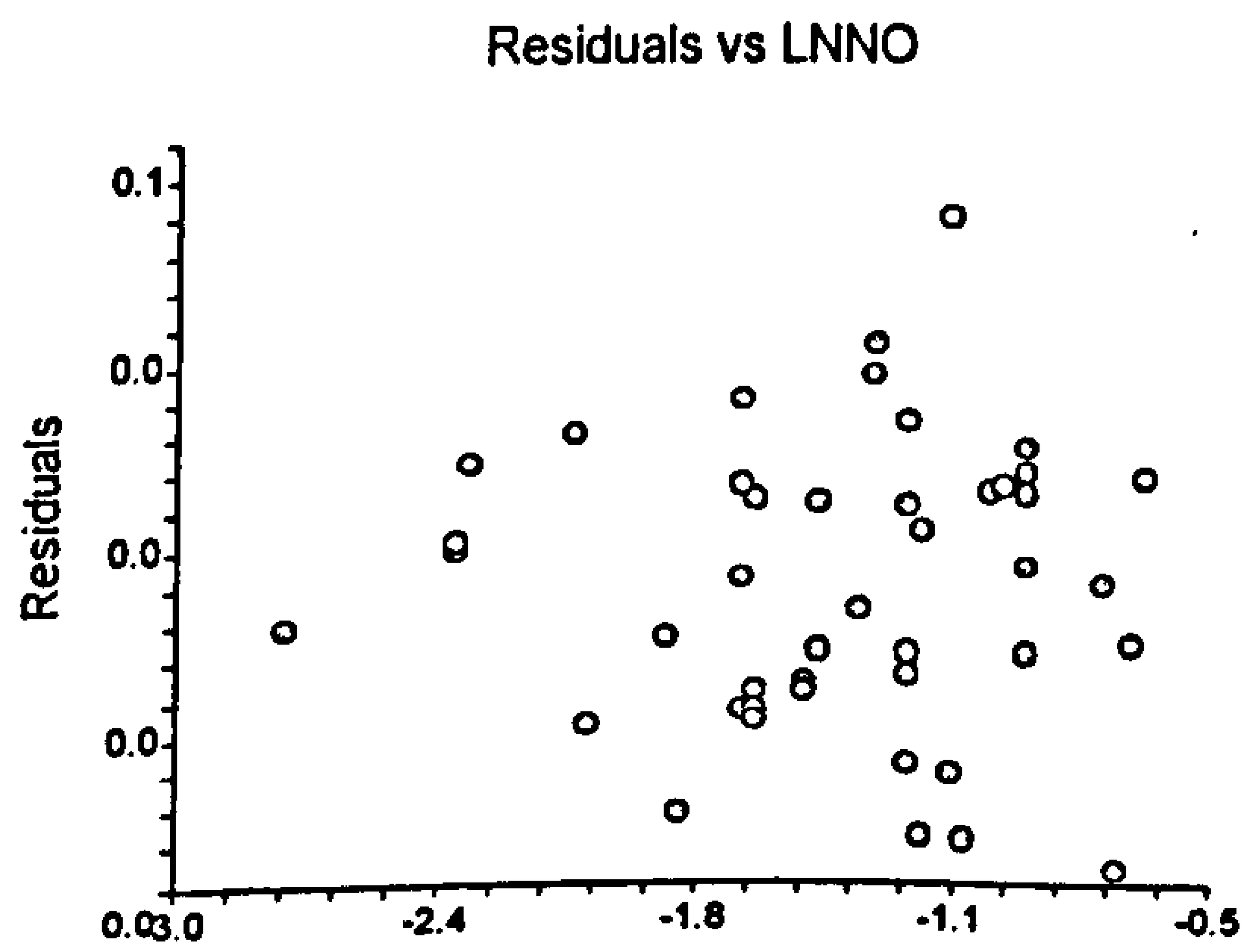
Plots Section

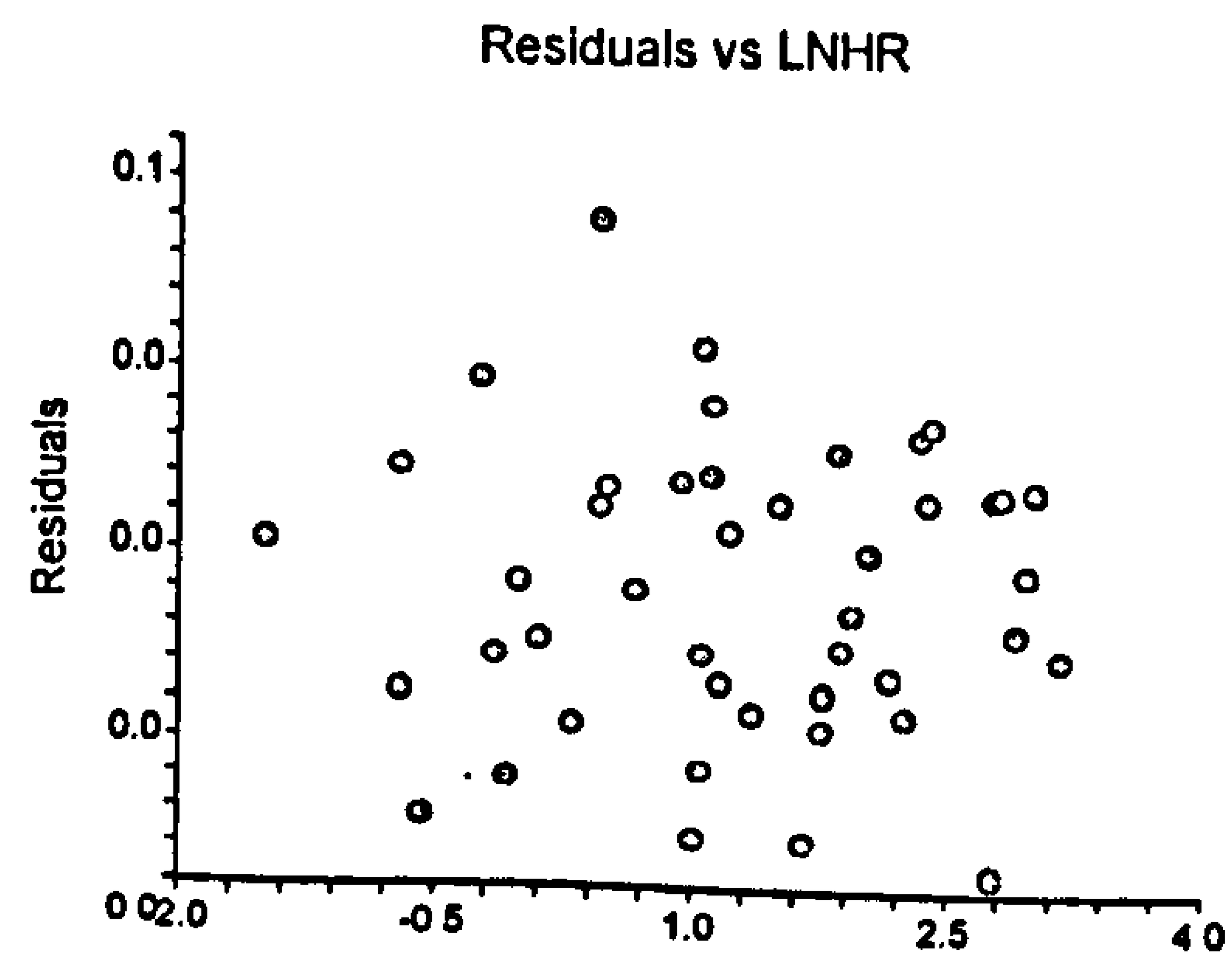


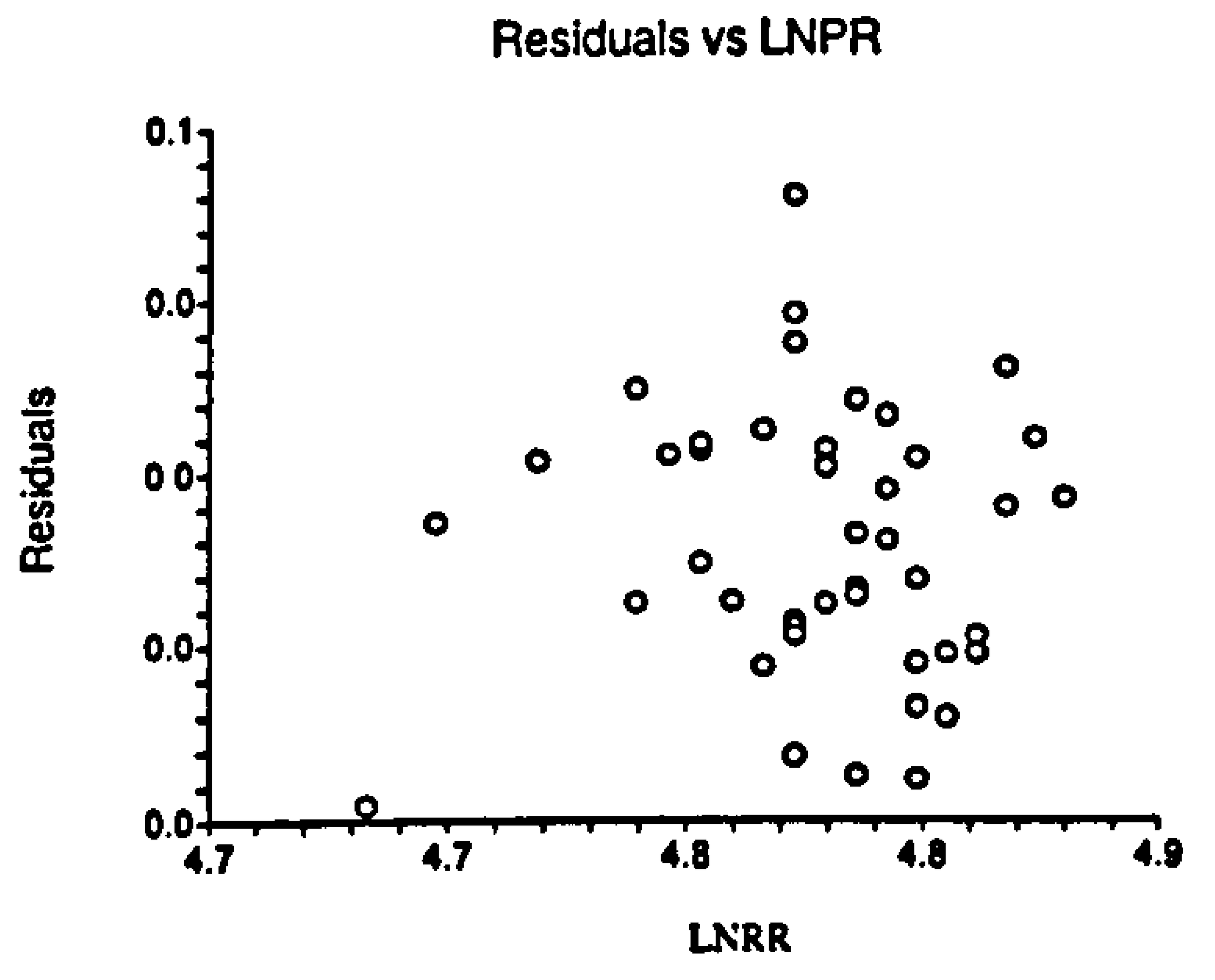


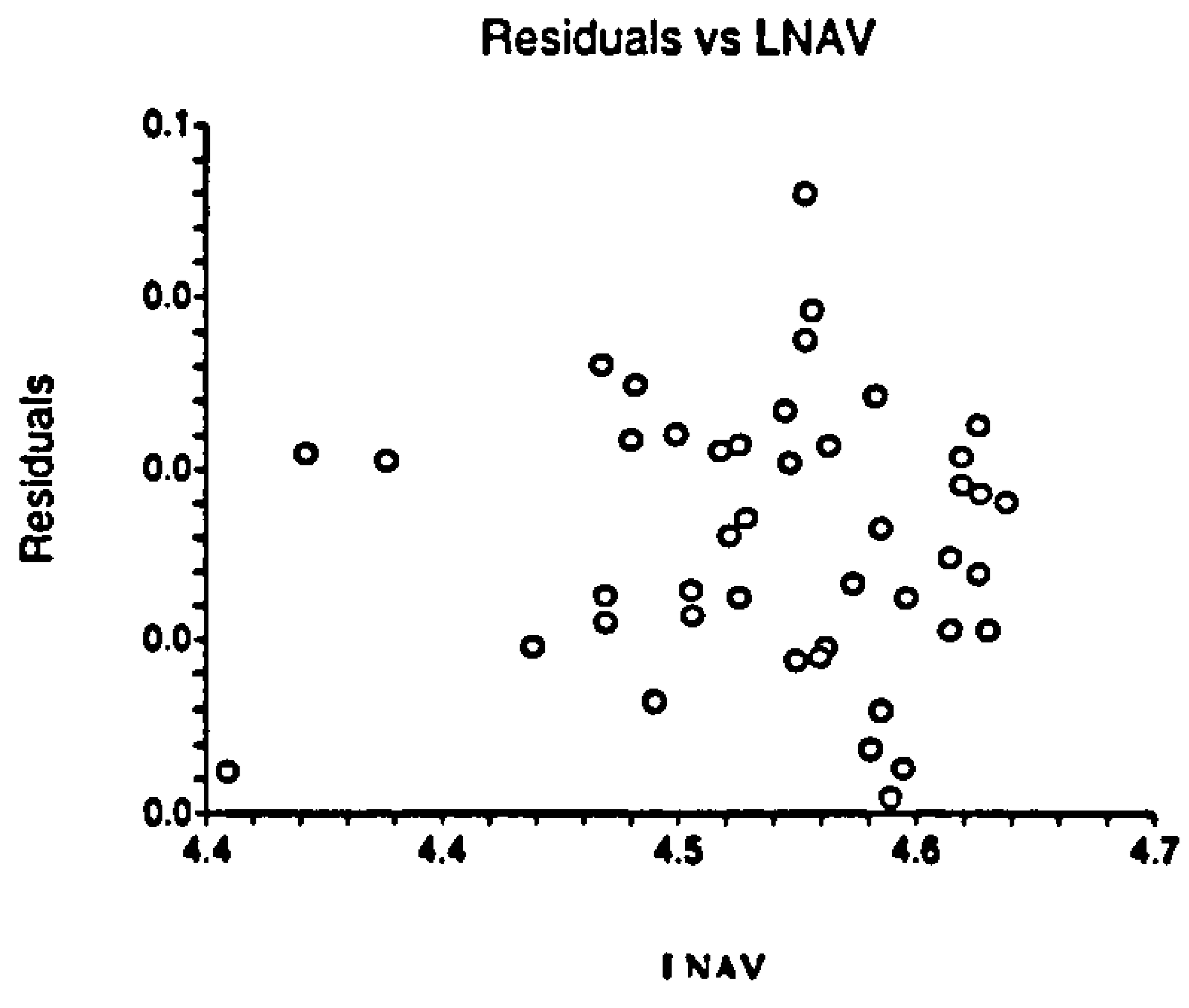












Comments and Conclusions

To sum up, we collect the main results in Tables (7.19a)-(7.19c).

1) Table (7.19a) aims at comparing the full models, which involve all the available independent variables. It seems reasonable from this table that.

a) The coefficients of the estimated models by sequentially entering one variable at a time are almost the same.

b) Comparing the full multiple regression-estimated coefficients with the corresponding coefficients of the ridge regression, we observe that they are almost the same. This means that the multicollinearity may not a problem since ridge regression is designed to deal with multicollinearity if it exists.

c) The robust regression models gave slightly different results because they use different truncation methods in weighing the error terms. However, the largest R-squared is obtained by the Andrew's sine method. Moreover, the R-squared values that are given by this procedure are higher than those given by other procedures. This is natural in our case since the data has some outliers. Because the robust regression is designed to handle such problems we recommend using the results of the robust regression.

d) The coefficients of each of the independent variables have the same signs in all models. These signs agree with the researcher expectations as it was stated in the data section. There is one exception with the signs of AvHOURS, but this is not a problem because this variable is not significant.

e) The relative importance of the independent variables can be seen from the all possible regression as well as from the stepwise procedure. This importance may be arranged according to the relative increase in R-squared value when that variable enters the model. It seems reasonable that the most important variable is PRORATE which is the first variable to enter the model with an R-squared value about 0.37. The second important variable is AVL, which raised R-squared value to about 0.40, i.e. its own contribution to R-Squared is about 0.03. The third important variable is AvHOURS, which has not increased the value of R-squared up to two decimals. Finally, the fourth important variable is AvNO, which increased R-squared value to about 0.41, i.e. its contribution to the R-squared value is about 0.01. So, it is very clear that these last two variables have a non-significant influence on the model.

f) Comparing the R-squared values of multiple regression, ridge regression, stepwise regression, and the three robust regression procedures, we observe that the values range between 0.37 and 0.59. The maximum is that of the Andrew sine robust model and the minimum is that of stepwise regression model. This is another reason to recommend using the robust regression procedure.

Table (7.19a) Comparison of Full Models
A blank cell means that the variable is not included in the model

Procedure	Intercept	AvNO	AvHOURS	PRORATE	AVL	R-Squared
All Possible						
Size 1	122.4343			-.2883676		0.369092
Size 2	129.706			-.2922173	-0.07317076	0.397758
Size 3	133.0156		-0.03138953	-.3143395	-0.07740773	0.403794
Size 4 (Multiple)	131.2914	.8061154	-0.03088235	-.3071651	-0.07071456	0.405061
Stepwise						
1- Forward	122.4343			-.2883676		0.369092
2- Step	129.706			-.2922173	-0.07317076	0.397758
3-Backward	122.4343			-.2883676		0.369092
4-Min MSE	122.4343			-.2883676		0.369092
Ridge	130.8831	.8397473	-0.02983807	-.3045999	-0.06988057	0.4028
Robust						
1-Least Abs. Dev. 1.0	131.6548	1.565402	-0.01613586	-.3102915	-0.0736572	0.535717
2-Tukey's Biweight 6.0	131.6169	.934507	-0.01748374	-.3084482	-0.07412268	0.472987
3-Andrew's Sine 2.1	133.3905	1.999712	-0.01039429	-.3304487	-0.06732384	0.592504

2) Since there are independent variables that are included in the full regression models even though they are not significant, one should fit the models another time after deleting these variables from the models. Table (7.19b) summarizes the results of these models with the selected variable. This will be called the final run results. It seems reasonable from this table that the same conclusions given above are still valid. However, these are simple linear regression models.

Table (7.19b) Comparison of Models with Significant Variables (Final Run)
A blank cell means that the variable is not included in the model

Procedure	Intercept	AvNO	AvHOURS	PRORATE	AVL	R-Squared
Multiple	122.4343			-.2883676		0.369092
Ridge	122.2581			-.2869329		0.3673
Robust						
Least Abs. Dev. 1.0	126.9424			-.325762		0.501982
Tukey's Biweight 6.0	124.1656			-.3028438		0.428325
Andrew's Sine 2.1	130.7643			-.3568497		0.569486

3) Table (7.19c) provides the R-squared values of the nonlinear regression models. It seems reasonable from this table that the values of R-squared ranges from about 0.40 to 0.48. The minimum value corresponds to the logarithmic model and the maximum value is that of the polynomial regression with interaction terms. Moreover, if we compare the R-squared value of the polynomial regression with interaction terms with that of the robust regression, we still observe that the R-squared value of robust regression is larger than that of the polynomial regression with interaction terms. For this reason and the fact that the polynomial regression with interaction terms is a more complicated model, we recommend using robust regression.

Table (7.19c) Comparison of Nonlinear Models

Model	R-Squared
Quadratic	0.433526
Polynomial with interactions	0.484707
Logarithmic	0.4009

As a final statement in this direction we say that, there are three reasons to recommend the robust regression model.

i) It is not sensitive to outliers which form a problem in our data.

ii) The other linear models have smaller R-squared values than that of the robust model.

iii) The nonlinear models are more complicated than the robust model and they have R-squared value smaller than that of the robust model. Moreover, the signs of the estimated coefficients of some of the no liners models contradict the researcher expect tations.

Recommendation:

Based on the above observations, we recommend using the final robust estimated regression mode, which is given by

FUEL= 130.7643 - 0.3568497 PRORATE

This model explains about 57% of the variability in FUEL.

7.5.4 Presentation of Regression Analysis of Kiln 2

7.5.4.1 EL for Kiln 2

From the full output reports, which are given in Appendix (03), we obtain the following summary of the results.

1- Table (7.20) provides the results of *checking the underlying assumptions*. It seems reasonable from this table that we have only problems with the existence of outliers. Other assumptions are satisfied. This suggests using the robust regression procedure.

Table (7.20) Check of the Regression Assumptions

Assumption	Tools to check assumptions	Results
1. Linearity	1. Scatter Plots	Only with AvNO, AvHOURS, PRORATE
	2. Correlation Matrix	Only with AvNO, AvHOURS, PRORATE
2. Normality of Residuals of Multiple Regression Model	1. Skewness Tests	Accepted
	2. Kurtosis Test	Accepted
	3. Omnibus Test	Accepted
	4. Normal Probability Plot	Almost normal
	5. Histogram	Almost normal
3. Constant Error Variance	1. Scatter Plots	No patterns
4. Independent Errors	1. Serial Correlations	Uncorrelated
	2. Durbin-Watson Test	Uncorrelated
5. Multicollinearity	1. Correlation Matrix	Absolute maximum 0.604090
	2. Condition Numbers	All are less than 100
	3. Variance Inflation Factors	All are less than 10
	4. Eigenvalues	No values are close to zero
6. Outliers	1. Histogram	Outliers Exist
	2. Normal Probability Plot	Outliers Exist
	3. Scatter Plots	Outliers Exist

2. Table (7.21) reports a summary of the results of *all possible regression models*. It seems reasonable from this table that based on the Cp criterion the best model is the one with only three independent variables, namely AvNO, AvHOURS, and AVL. Moreover, the difference in R-squared between the full model with four variables and the model with these three variables is about 0.002 which is not worth complicating the model.

Table (7.21) Results of All Possible Regression Procedure

Model Size	R-Squared	Root MSE	Cp	Variables in Model	Best Model
1	0.493395	1.078017	25.984241	AvHOURS	
2	0.645813	0.9117998	7.530435	AvNO, AvHOURS	
3	0.692469	0.8596796	3.269285	AvNO, AvHOURS, AVL	This one
4	0.694476	0.867257	5.000000	AvNO, AvHOURS, PRORATE, AVL	

3. Table (7.22) reports a summary of the obtained ***initial regression models*** using different procedures together with their R-squared values. It seems reasonable from this table that

i) The coefficients of the estimated models by sequentially entering one variable at a time are almost the same. This is seen from all possible regression models section. This indicates a type of consistancy of the results.

ii) The robust regression models gave slightly different results because they use different influence functions and different truncation methods in weighing the error terms. However, the largest R-squared is obtained by the Andrew's sine method.

iii) The coefficients of the significant (influential) variables (AvNO, AvHOURS, and AVL) have the same signs in all models. The AvNO and AvHOURS variables have positive signs which means that the dependent variable increases as each of these independent variables increases. However, the AVL variable has negative sign, this means that the dependent variable decreases as each of these variables increases. These signs agree with the researcher expectations, which are stated earlier.

iv) Since the data has some outliers, we recommend using the robust regression model using Andrew's sine influence function after deleting the independent variables, which do not contribute to the dependent variable. This will be done in what is called the final run section together with a final run of multiple regression model after deleting PRORATE which was not influential.

Table (7.22) Comparison of the estimated models using different procedures

Procedure	Intercept	AvNO	AvHOURS	PRORATE	AVL	R-Squared
All Possible:						
Size 1	18.1658		.1932887			0.493395
Size 2	16.77679	4.844143	.2047747			0.645813
Size 3	21.38579	4.555209	.2051574		-4.967873E-0	0.692469
Size 4 (Multiple)	23.66305	4.368386	.1947135	-1.886802E-02	-4.835395E-0	0.694476
Stepwise:						
1- Forward	21.38579	4.555209	.2051574		-4.967873E-0	0.692469
2- Step	21.38579	4.555209	.2051574		-4.967873E-0	0.692469
3-Backward	21.38579	4.555209	.2051574		-4.967873E-0	0.692469
4-Min MSE	21.38579	4.555209	.2051574		-4.967873E-0	0.692469
Robust						
1- Least Abs. Dev. 1.0	22.71663	4.184063	.1984934	-8.909876E-03	-5.159259E-0	0.803048
2- Tukey'sBiweigh	21.05387	4.53777	.2057262	5.928104E-03	-5.467566E-0	0.788249
3- Andrew's Sine 2	24.03158	3.806809	.1843329	-.023665	-4.407377E-0	0.821279

4) It seems reasonable from the *final multiple regression* report(see Table (7.23a)) that

i) R-squared value is 0.692469. This means that this model explains about 69% of the variability in the independent variable. This is a useful coefficient of determination.

ii) The PRESS R-squared is 0.6398, which means that this model can predict to good accuracy about 64% of the values of Y. This is also a useful indicator of predictability.

iii) Normality of the residuals is not accepted.

iv) Based on serial correlations and the Durbin-Watson test, the residuals are uncorrelated.

v) Multicollinearity may not be a problem.

vi) From the plot section, one observes that the data has some outliers.

Table (7.23a) Final Multiple Regression

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	21.38579	1.861768	11.4868	0.000000	Reject Ho	1.000000
AvNO	4.555209	1.067893	4.2656	0.000111	Reject Ho	0.986316
AvHOURS	0.2051574	0.0236813	8.6633	0.000000	Reject Ho	1.000000
AVL	-4.967873E-02	1.968032E-02	-2.5243	0.015461	Reject Ho	0.693668
R-Squared	0.692469					

Model
EL= 21.38579+ 4.555209*AvNO+ .2051574*AvHOURS-4.967873E-02*AVL

Regression Coefficient Section

Independent Variable	Regression Standardized Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Coefficient
Intercept	21.38579	1.861768	17.62859	25.14299	0.0000
AvNO	4.555209	1.067893	2.400113	6.710305	0.3692
AvHOURS	0.2051574	0.0236813	0.1573666	0.2529482	0.7456
AVL	-4.967873E-02	1.968032E-02	-8.939523E-02	-9.962227E-03	-0.2173
T-Critical	2.018082				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	17363.04	17363.04			
Model	3	69.8932	23.29773	31.5239	0.000000	1.000000
Error	42	31.04006	0.7390491			
Total(Adjusted)	45	100.9333	2.242961			

Root Mean Square Error	0.8596796	R-Squared	0.6925
Mean of Dependent	19.42826	Adj R-Squared	0.6705
Coefficient of Variation	4.424892E-02	Press Value	36.35623
Sum |Press Residuals|	32.08179	Press R-Squared	0.6398

Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	1.7111	0.087059	Accepted
Kurtosis	1.5114	0.130676	Accepted
Omnibus	5.2124	0.073815	Accepted

Serial-Correlation Section

Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.184714	9	-0.127136	17	-0.072700
2	0.253302	10	0.097496	18	-0.224156
3	0.145195	11	-0.146517	19	-0.133218
4	-0.064974	12	0.116362	20	-0.073672
5	0.209841	13	-0.121545	21	-0.204071

6	-0.200744	14	-0.030768	22	-0.016221
7	-0.069457	15	0.024553	23	-0.096134
8	-0.160934	16	-0.271734	24	0.086054

Above serial correlations significant if their absolute values are greater than 0.294884
Durbin-Watson Value 1.5563

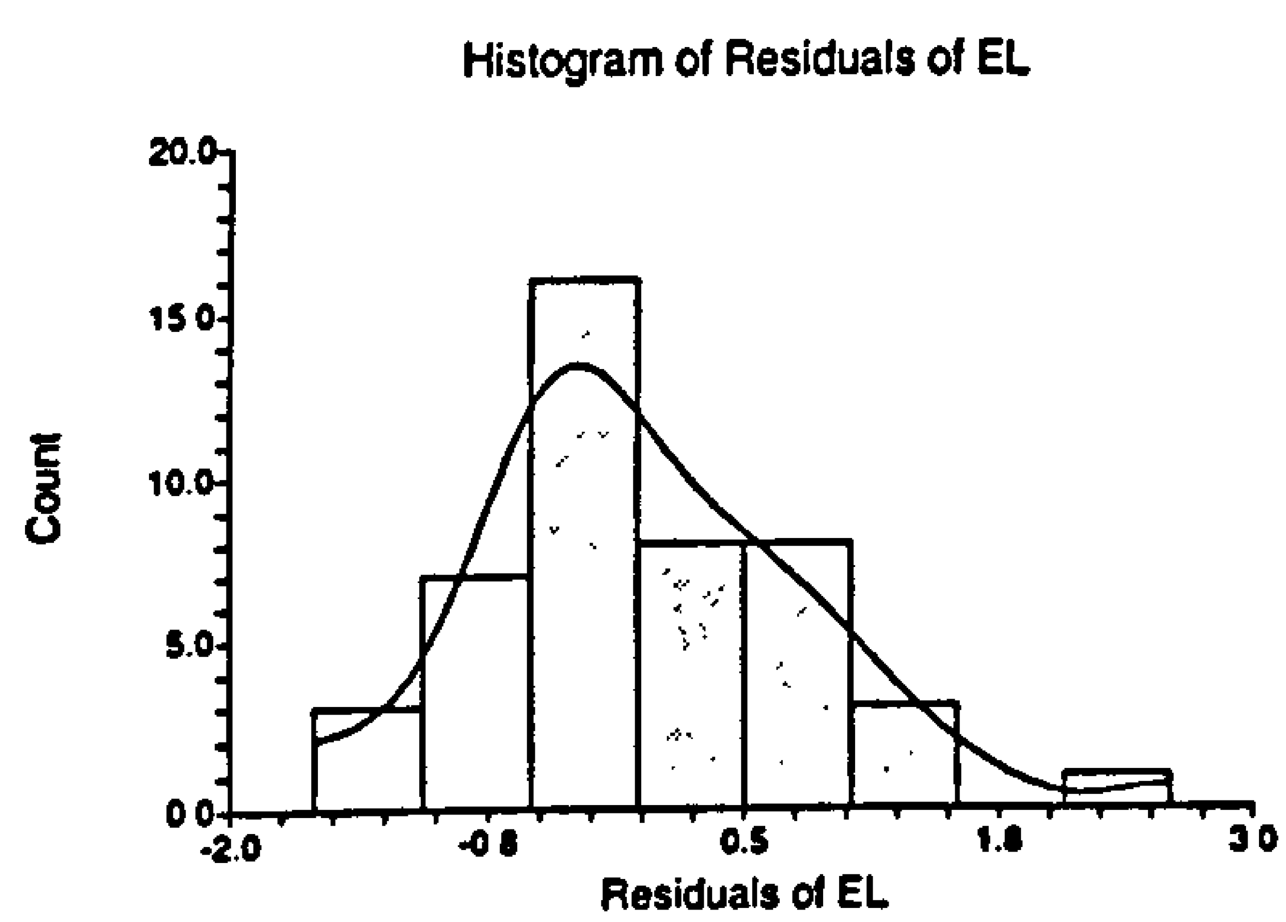
Multicollinearity Section				
Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance	Diagonal of X'X Inverse
AvNO	1.023186	0.022661	0.977339	1.543059
AvHOURS	1.011472	0.011342	0.988658	7.588185E-04
AVL	1.011945	0.011804	0.988196	5.240723E-04

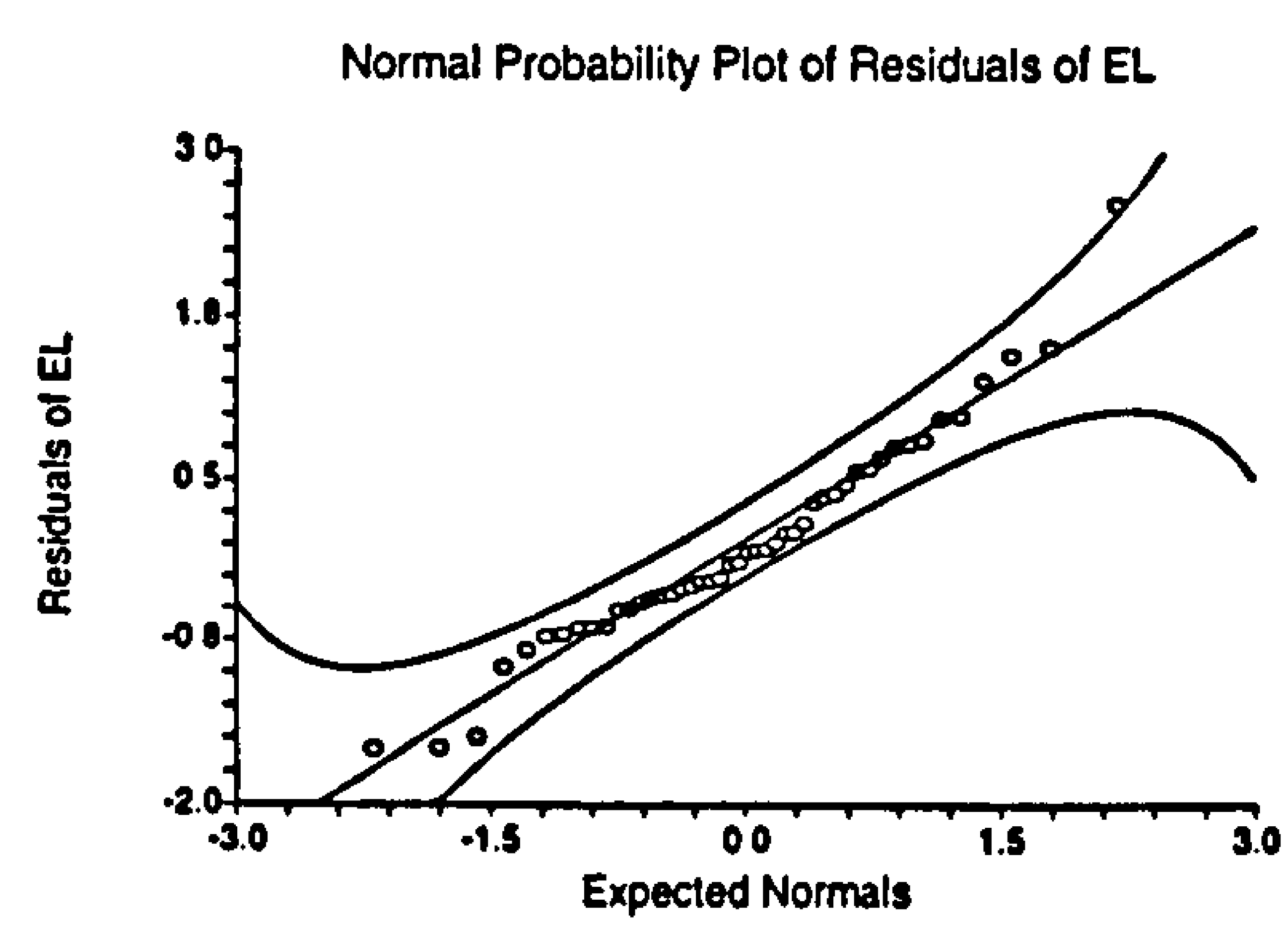
Eigenvalues of Centered Correlations

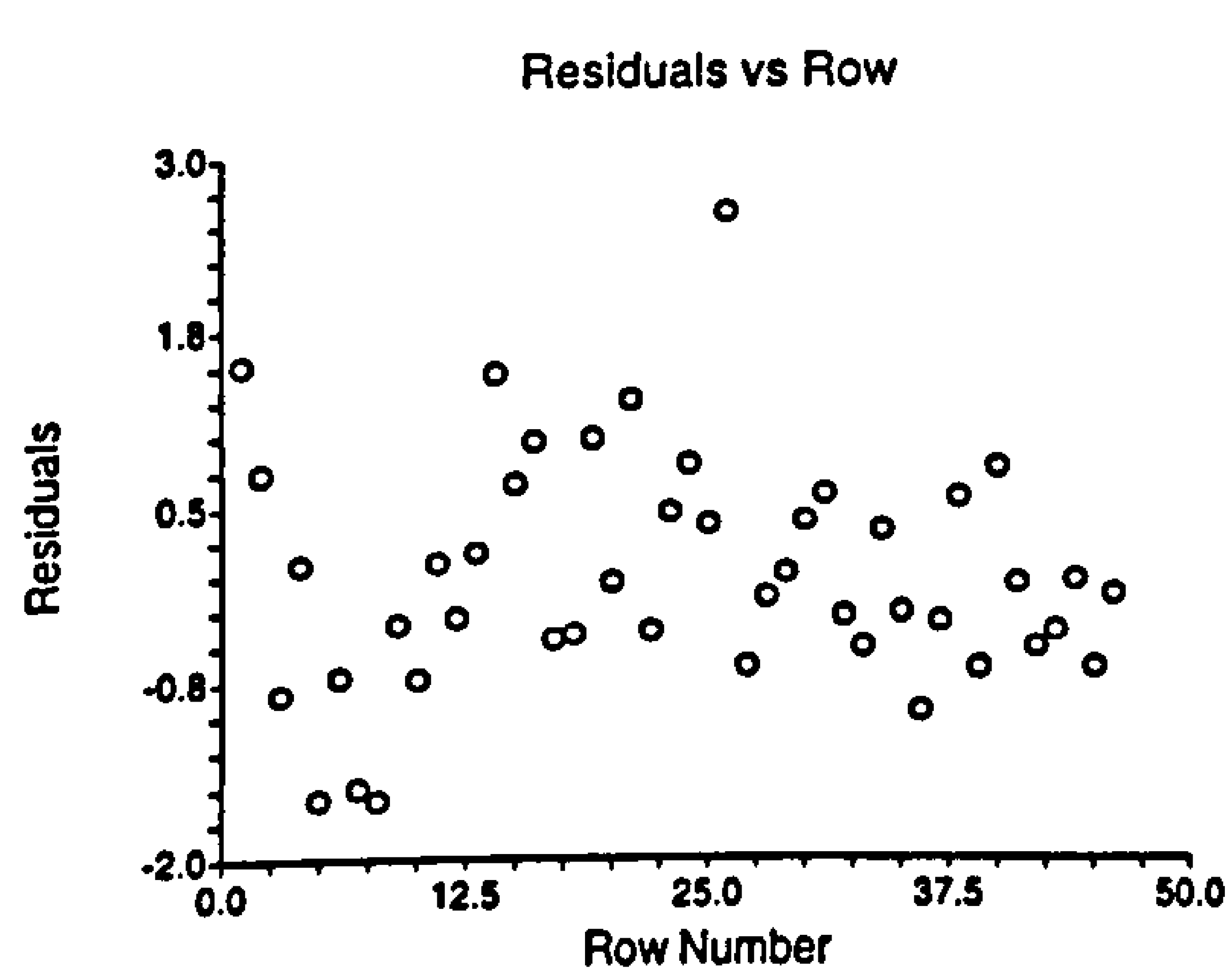
No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.161062	38.70	38.70	1.00
2	0.982145	32.74	71.44	1.18
3	0.856793	28.56	100.00	1.36

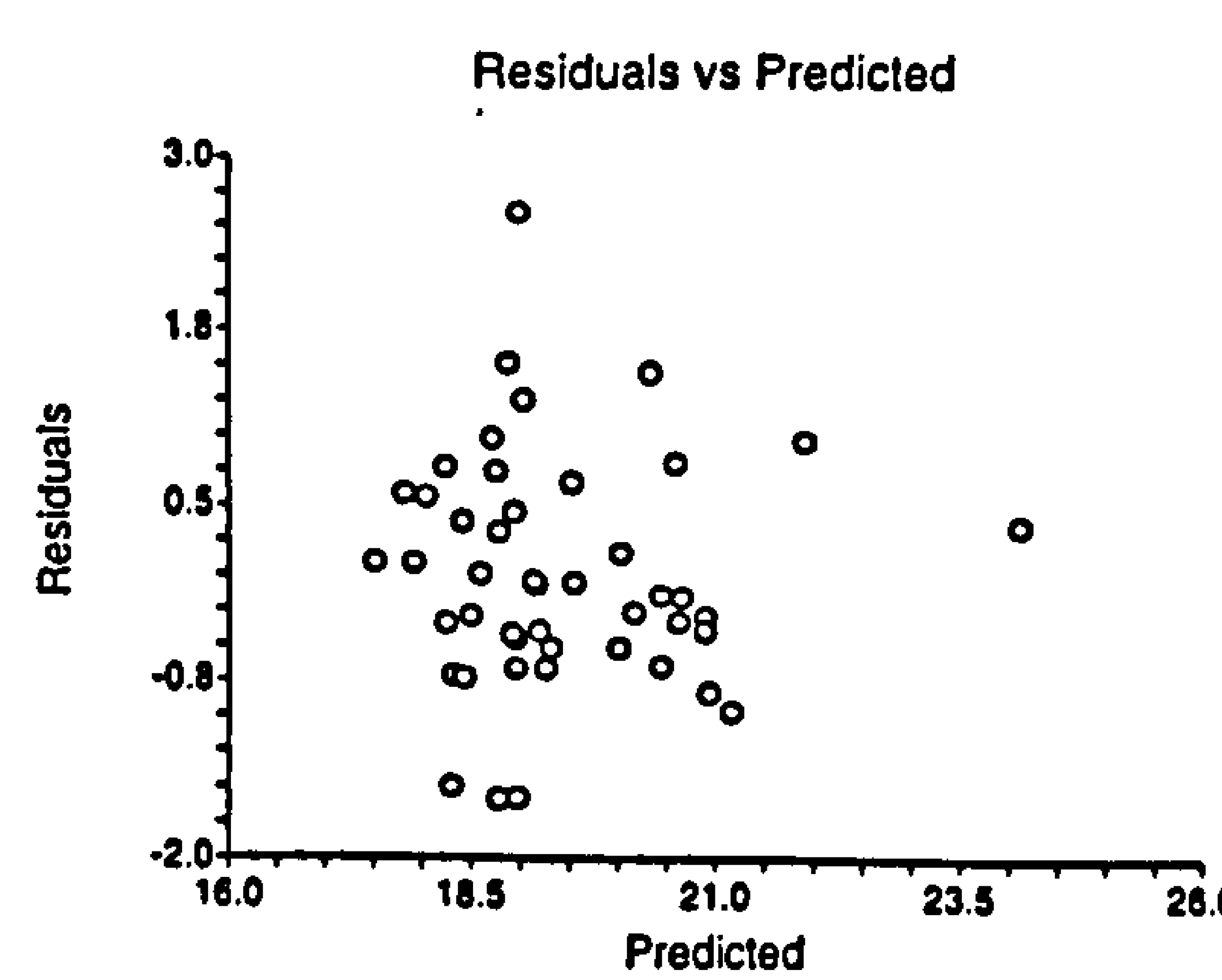
All Condition Numbers less than 100. Multicollinearity is NOT a problem.

Plots Section

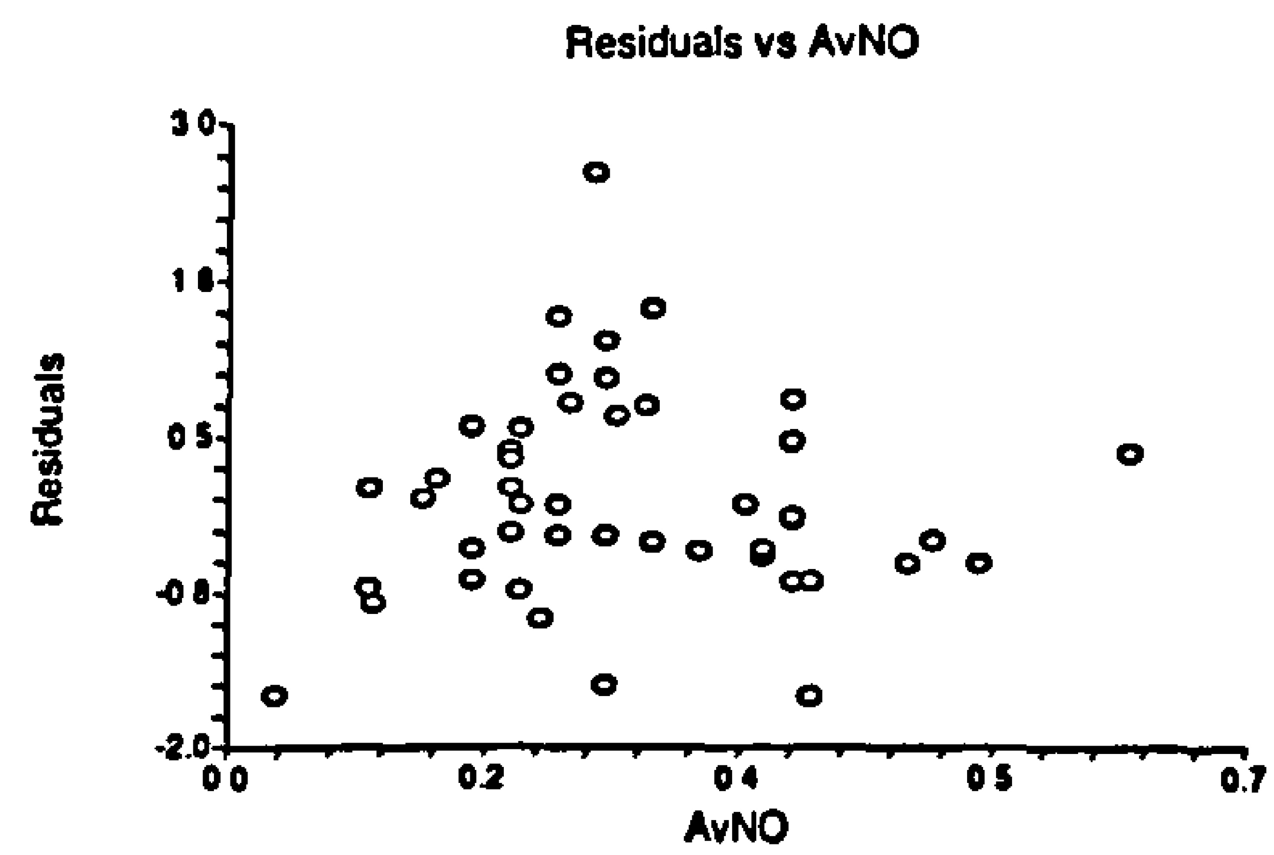


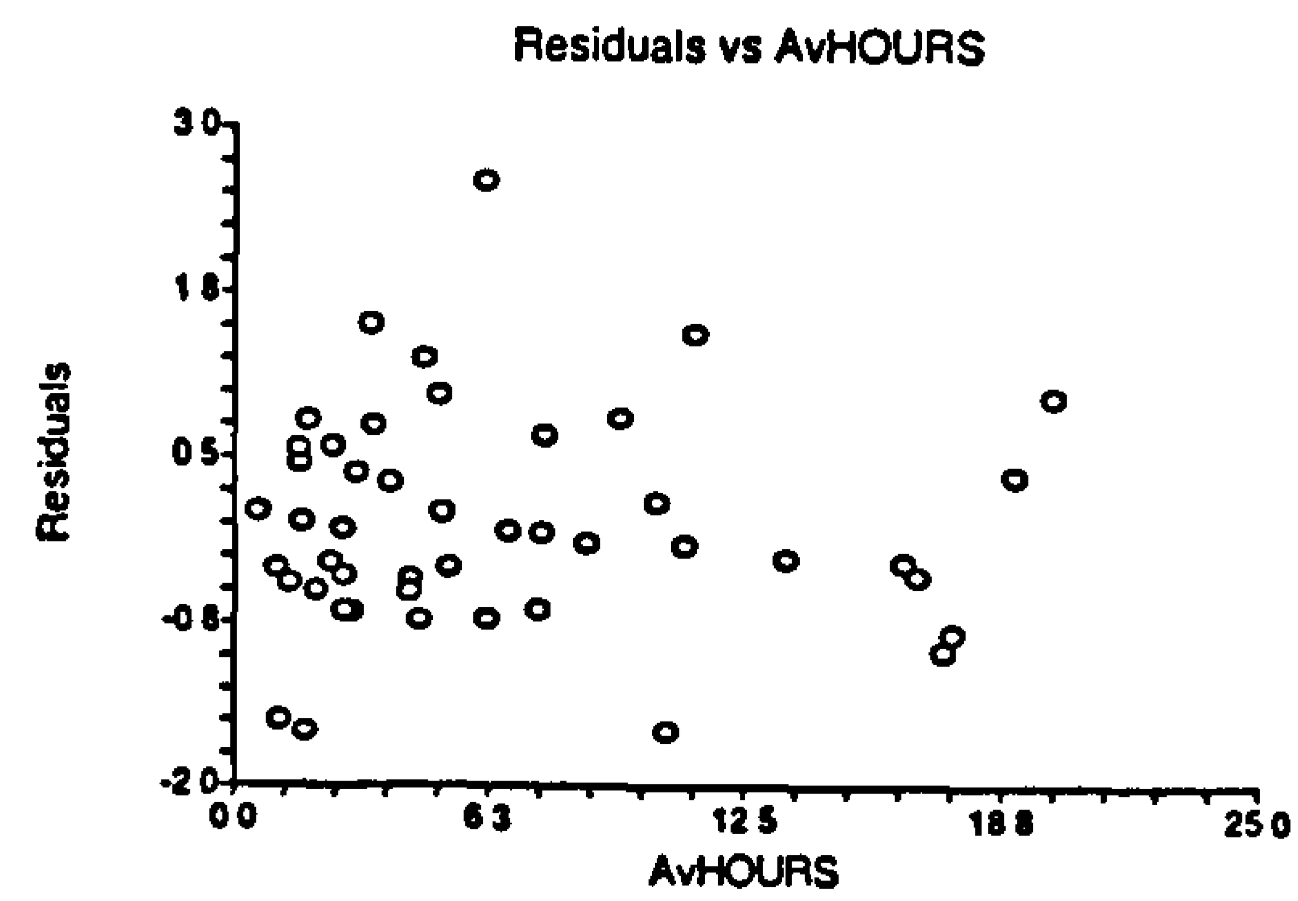


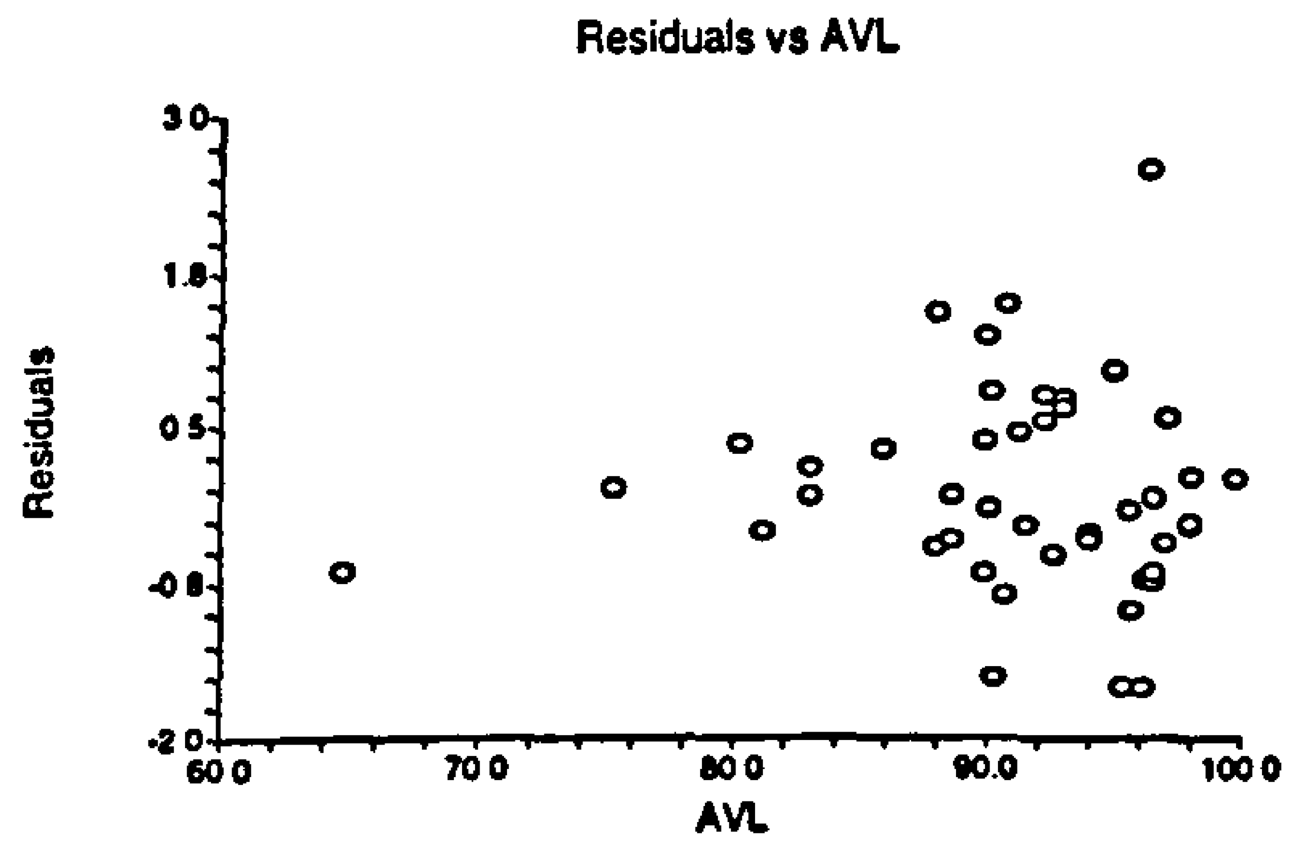




Multiple Regression Report







5) It seems reasonable from the ***final robust regression*** report(see Table (7.23b)) that R-squared value is 0.816465. This means that this model explains about 82% of the variability in the independent variable.

Table (7.23b) Final Run of Robust Regression

Robust Regression Report

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	21.28163	1.104166	19.2739	0.000000	Reject Ho	1.000000
AvNO	4.159125	0.6790985	6.1245	0.000000	Reject Ho	0.999971
AvHOURS	0.1893761	1.561334E-02	12.1291	0.000000	Reject Ho	1.000000
AVL	-4.711448E-02	1.177016E-02	-4.0029	0.000249	Reject Ho	0.974401
R-Squared	0.816465					

Model
EL= 21.28163+ 4.159125*AvNO+ .1893761*AvHOURS-4.711448E-02*AVL

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	21.28163	1.104166	19.05334	23.50993	0.000000
AvNO	4.159125	0.6790985	2.788649	5.529602	0.407890
AvHOURS	0.1893761	1.561334E-02	0.1578671	0.2208851	0.806568
AVL	-4.711448E-02	1.177016E-02	-7.086761E-02	-2.336134E-02	-0.265633
T-Critical	2.018082				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	12530.4	12530.4			
Model	3	41.7846	13.9282	62.2798	0.000000	1.000000
Error	42	9.392836	0.223639			
Total(Adjusted)	45	51.17743	1.137276			
Root Mean Square Error		0.4729048	R-Squared	0.816465		
Mean of Dependent Variable		19.33854	Adj R-Squared	0.803356		
Coefficient of Variation		2.445401E-02				

6. Conclusions:

Comparing the final runs of the OLS multiple regression model and the final run of the robust regression, we observe that

1) The existance of outliers is the only problem with the OLS multiple regression. However, this problem does not affect the robust regression.
2) The final multiple regression model and the robust regression model have only the same three significant independent variable, namely, AvNo, AvHOURS, and AVL.
3) For multiple regression R^2=0.69, while for robust regression R^2=0.82. This means that the robust regression explains much more of the variability in the dependent variable than the multiple regression does.
4) Based on the above observations, we recommend using the final robust estimated regression mode, which is given by

   EL= 21.28163+ 4.159125*AvNO+ .1893761*AvHOURS-4.711448E-02*AVL

7.5.4.2 FUEL for Kiln 2

From the full output reports, which are given in Appendix (03), we obtain the following summary of the results.

1- Table (7.24) provides the results of ***checking the underlying assumptions***. It seems reasonable from this table that we have only problems with the normality assumption and the existence of outliers. Other assumptions are satisfied. This suggests using robust regression procedure.

Table (7.24) Check of the Regression Assumptions

	Tools	Result
1. Linearity	1. Scatter Plots	Only with AvHOURS, PRORATE
	2. Correlation Matrix	Only with AvHOURS, PRORATE
2. Normality of Residuals	1. Skewness Tests	Rejected
	2. Kurtosis Test	Rejected
	3. Omnibus Test	Rejected
	4. Normal Probability Plot	Almost normal
	5. Histogram	Almost normal
3. Constant Error Variance	1. Scatter Plots	No patterns
4. Independent Errors	1. Serial Correlations	Uncorrelated
	2. Durbin-Watson Test	Uncorrelated
5. Multicollinearity	1. Correlation Matrix	Absolute maximum 0.604090
	2. Condition Numbers	All are less than 100
	3. Variance Inflation Factors	All are less than 10
	4. Eigene Values	No values are close to zero
6. Outliers	1. Histogram	Outliers Exist
	2. Normal Probability Plot	Outliers Exist
	3. Scatter Plots	Outliers Exist

2. Table (7.25) reports a summary of the results of ***all possible regression models***. It seems reasonable from this table that based on the Cp criterion the best model is the one with only PRORATE as an independent variable. Moreover, the difference in R-squared between the full model with four variables and the model with this variable is about 0.03 which is not worth complicating the model.

Table (7.25) Results of All Possible Regression Procedure

Model Size	R-Squared	Root MSE	Cp	Variables in Model	Best Model
1	0.316854	3.406805	1.362116	PRORATE	This one
2	0.340390	3.386307	1.868210	AvHOURS, PRORATE	
3	0.354068	3.390672	3.000009	AvNO, AvHOURS, PRORAT	
4	0.354068	3.431772	5.000000	AvNO, AvHOURS, PRORAT AVL	

3. Table (7.26) reports a summary of the obtained ***initial regression models*** using different procedures together with their R-squared values. It seems reasonable from this table that:

i) The coefficients of the estimated models by sequentially entering one variable at a time are slightly different. This is seen from all possible regression models section.

ii) The robust regression models gave slightly different results because they use different truncation methods in weighing the error terms. However, the largest R-squared is obtained by the Andrew's sine method.

iii) The coefficients of the influential variable (PRORATE) have the same sign in all models. The PRORATE, variable has negative sign, this means that the dependent variable decreases as this variable increases. The signs of the other independent variables are not important since their coefficients are not significantly different from zeros. This means that they do not have significant influence on the dependent variable.

iv) Since the data has some outliers and normality is not completely satisfied, we recommend using the robust regression model .

Table (7.26) Comparison of the estimated models using different procedures

Procedure	Intercept	AvNO	AvHOURS	PRORATE	AVL	R-Squared
All Possible						
Size 1	146.2044			-.4817984		0.316854
Size 2	133.2387		.1441649	-.3822605		0.340390
Size 3	126.6222	4.180379	.177123	-.3386572		0.354068
Size 4 (Multiple)	126.637	4.179579	.1771547	-.3386031	-0.0233412	0.354068

Stepwise						
1- Forward	146.2044			-.4817984		0.316854
2- Step	146.2044			-.4817984		0.316854
3-Backward	146.2044			-.4817984		0.316854
4-Min MSE	146.2044			-.4817984		0.316854
Robust						
1- Least Abs. Dev.	117.8641	3.739973	.1114986	-.2617749	-0.0406019	0.462703
2-Tukey's Biweigh	113.4744	2.419786	.0987749	-.2085551	-0.0246739	0.402501
3- Andrew's Sine 2	112.3856	3.32656	0.05018888	-.2016648	-0.021949	0.471992

4) It seems reasonable from the final run of multiple regression (which is in fact simple linear

regression) (see Table (7.27)) that :

i) R-squared value is 0.316854. This means that this model explains about 31% of the variability in the independent variable. This is a low coefficient of determination.

ii) The PRESS R-squared is 0.2227, which means that this model can predict to good accuracy about 22% of the values of Y. This is also too low, and hence such a model is not useful.

iii) Normality of the residuals is rejected. This is why the above model is not useful, because ordinary least square method which is used to fit the multiple model is very sensitive to the normality assumption and to the existence of outliers.Based on serial correlations and the Durbin-Watson test, the residuals are uncorrelated.

iv) Multicollinearity should not be a problem.

v) From the plot section, one observes that the data has some outliers.

Table (7.27) Final Multiple Regression

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	146.2044	12.89324	11.3396	0.000000	Reject Ho	1.000000
PRORATE	-0.4817984	0.1066513	-4.5175	0.000047	Reject Ho	0.992994
R-Squared	0.316854					

Model
FUEL= 146.2044-.4817984*PRORATE

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	146.2044	12.89324	120.2198	172.1891	0.0000
PRORATE	-0.4817984	0.1066513	-0.6967399	-0.2668568	-0.5629
T-Critical	2.015368				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	356250.7	356250.7			
Model	1	236.8607	236.8607	20.4079	0.000047	0.992994
Error	44	510.6782	11.60632			
Total(Adjusted)	45	747.5389	16.61197			

Root Mean Square Error	3.406805	R-Squared	0.3169
Mean of Dependent	88.0033	Adj R-Squared	0.3013
Coefficient of Variation	3.871224E-02	Press Value	581.0776
Sum |Press Residuals|	102.6675	Press R-Squared	0.2227

Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	5.0558	0.000000	Rejected
Kurtosis	4.4514	0.000009	Rejected
Omnibus	45.3761	0.000000	Rejected

Serial-Correlation Section

Lag	Correlation	Lag	Correlation	Lag	Correlation
1	-0.124772	9	-0.110814	17	-0.007351
2	0.118337	10	0.143999	18	-0.174962
3	-0.118181	11	0.023560	19	-0.057019
4	-0.104916	12	0.120127	20	-0.007713
5	0.002227	13	-0.006958	21	-0.069326
6	-0.218503	14	0.061535	22	0.243033
7	-0.012810	15	0.074733	23	-0.061986
8	0.078833	16	-0.113359	24	0.052205

Above serial correlations significant if their absolute values are greater than 0.294884
Durbin-Watson Value 2.2392

Multicollinearity Section

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance	Diagonal of X'X Inverse
PRORATE	1.000000	0.000000	1.000000	9.800259E-04

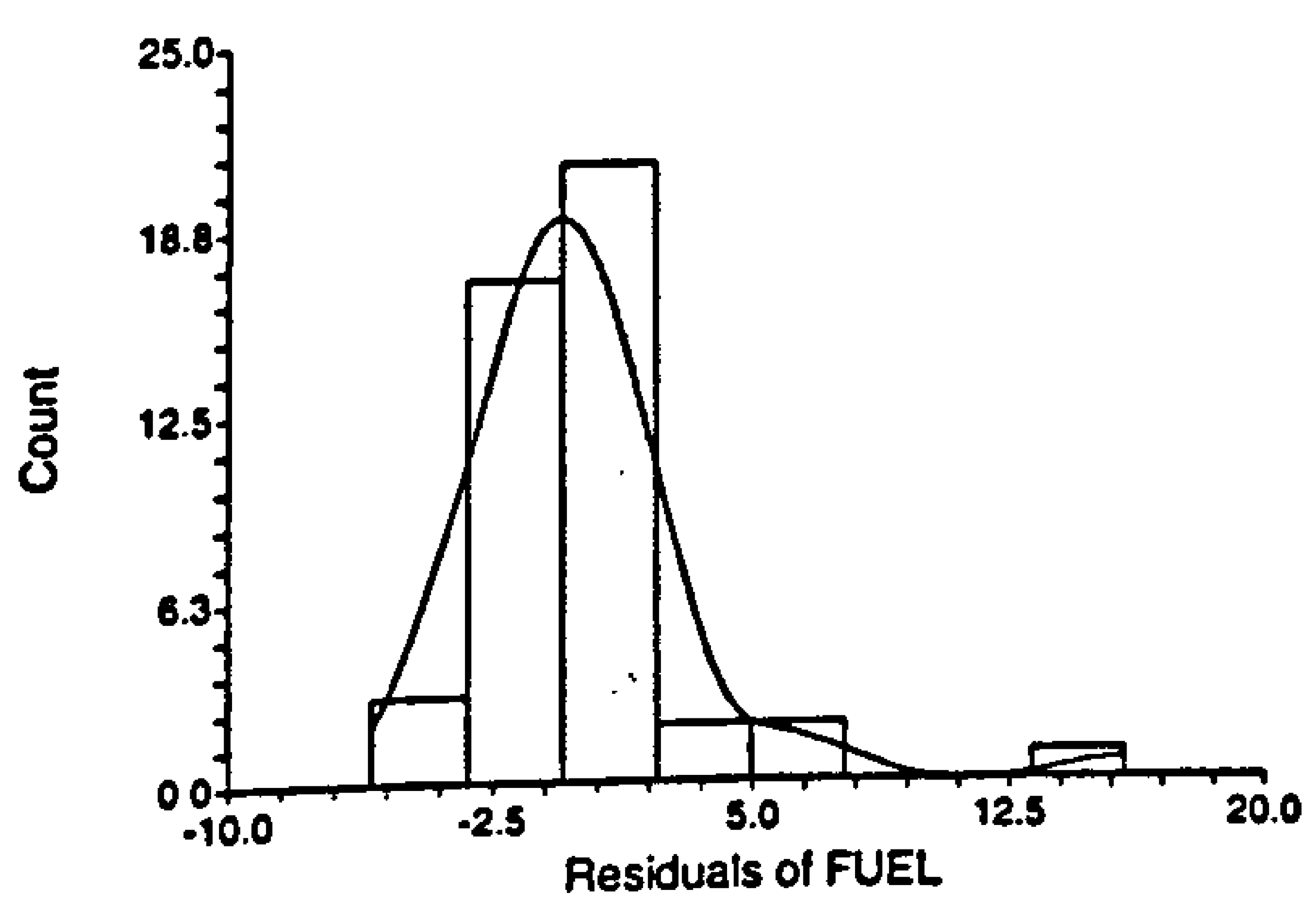
Eigenvalues of Centered Correlations

No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.000000	100.00	100.00	1.00

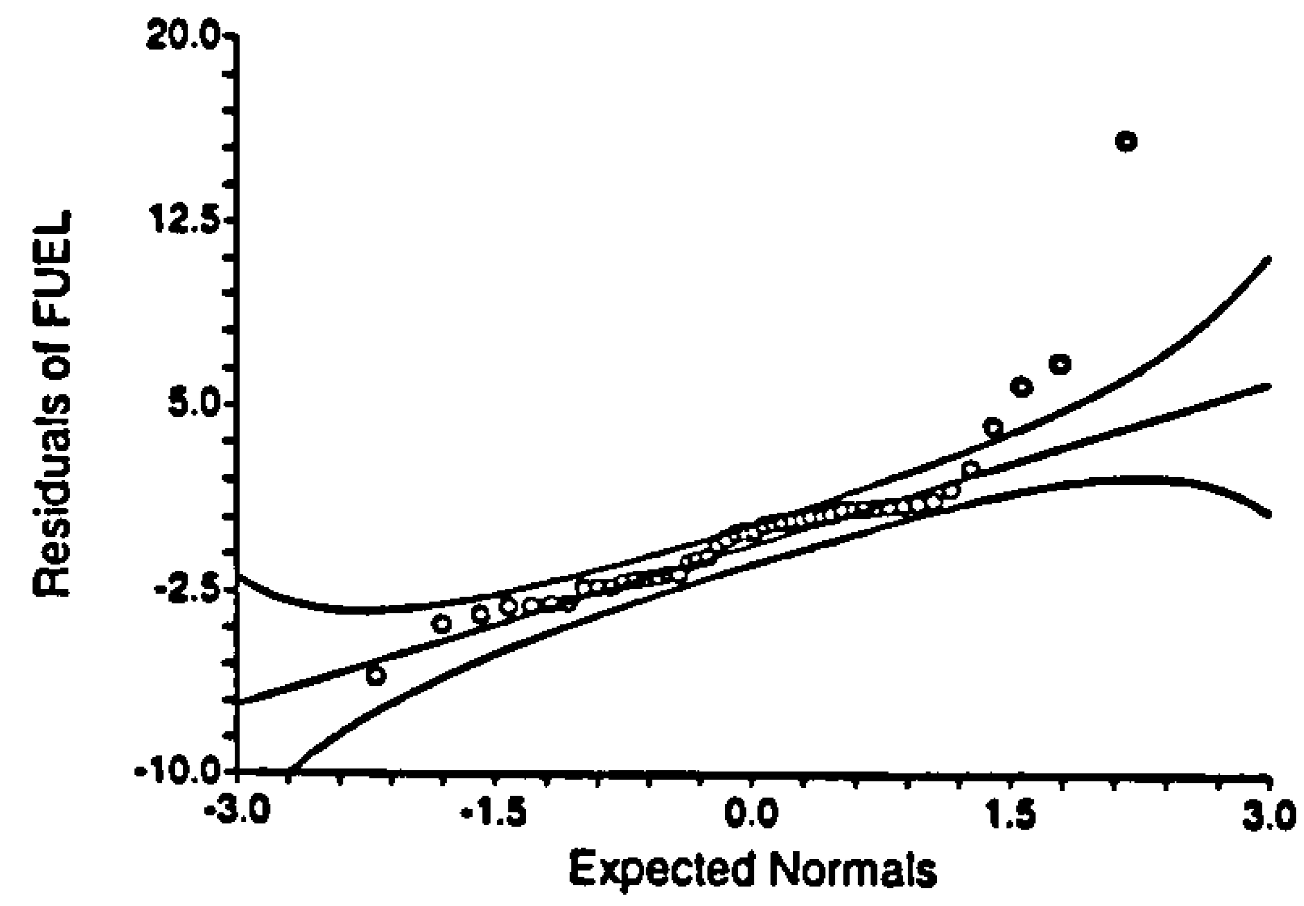
All Condition Numbers less than 100. Multicollinearity is NOT a problem.

Plots Section

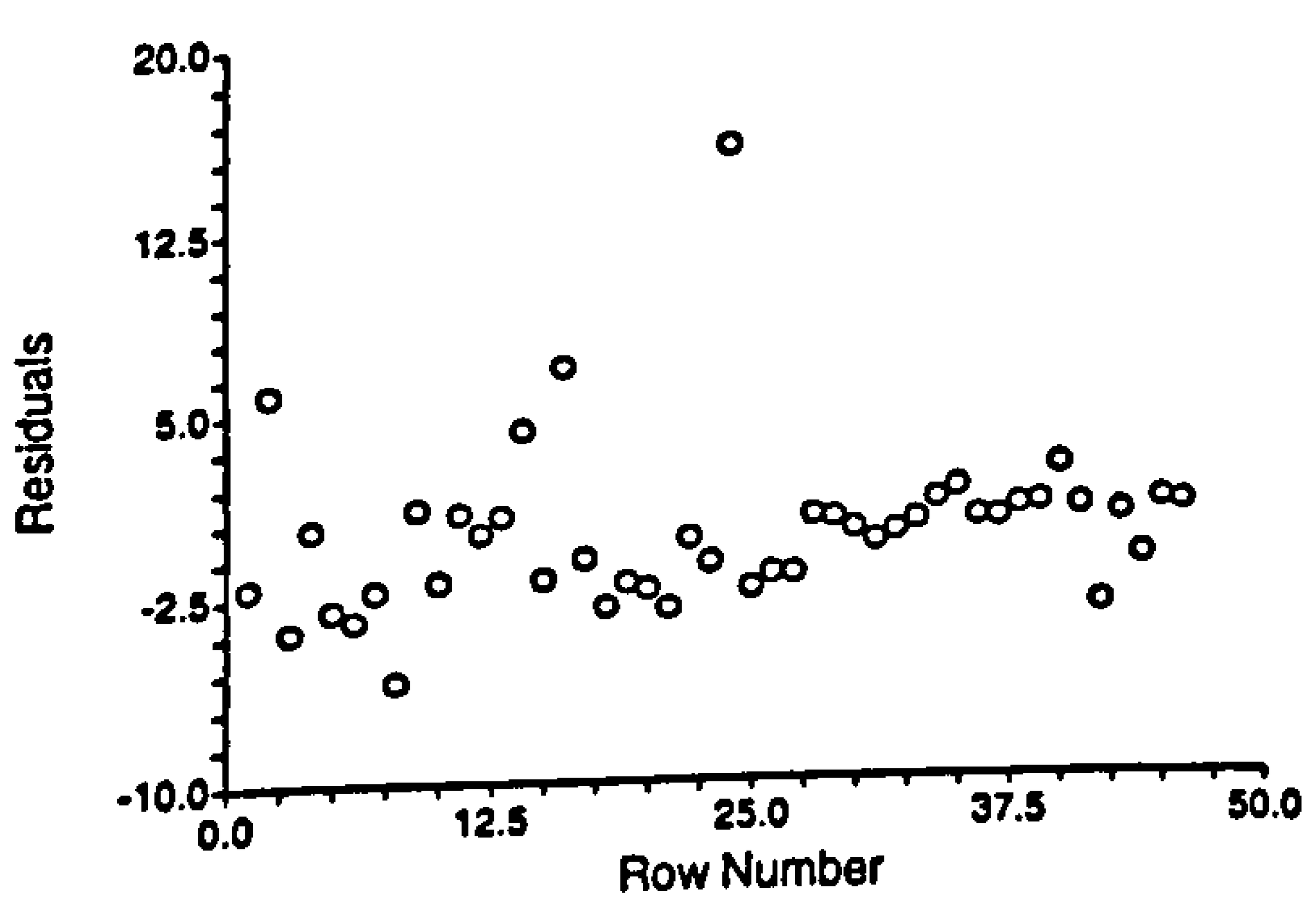
Histogram of Residuals of FUEL



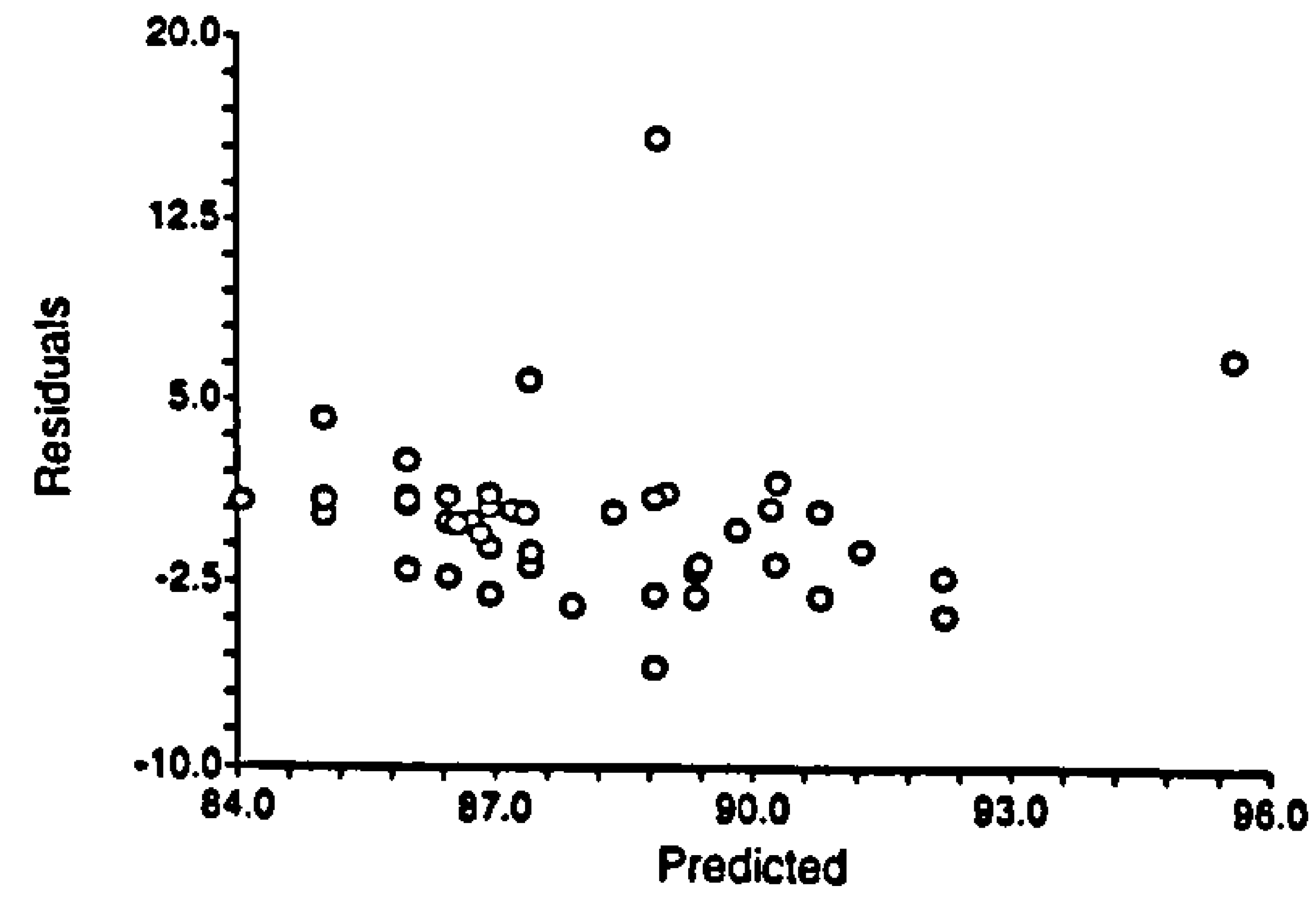
Normal Probability Plot of Residuals of FUEL

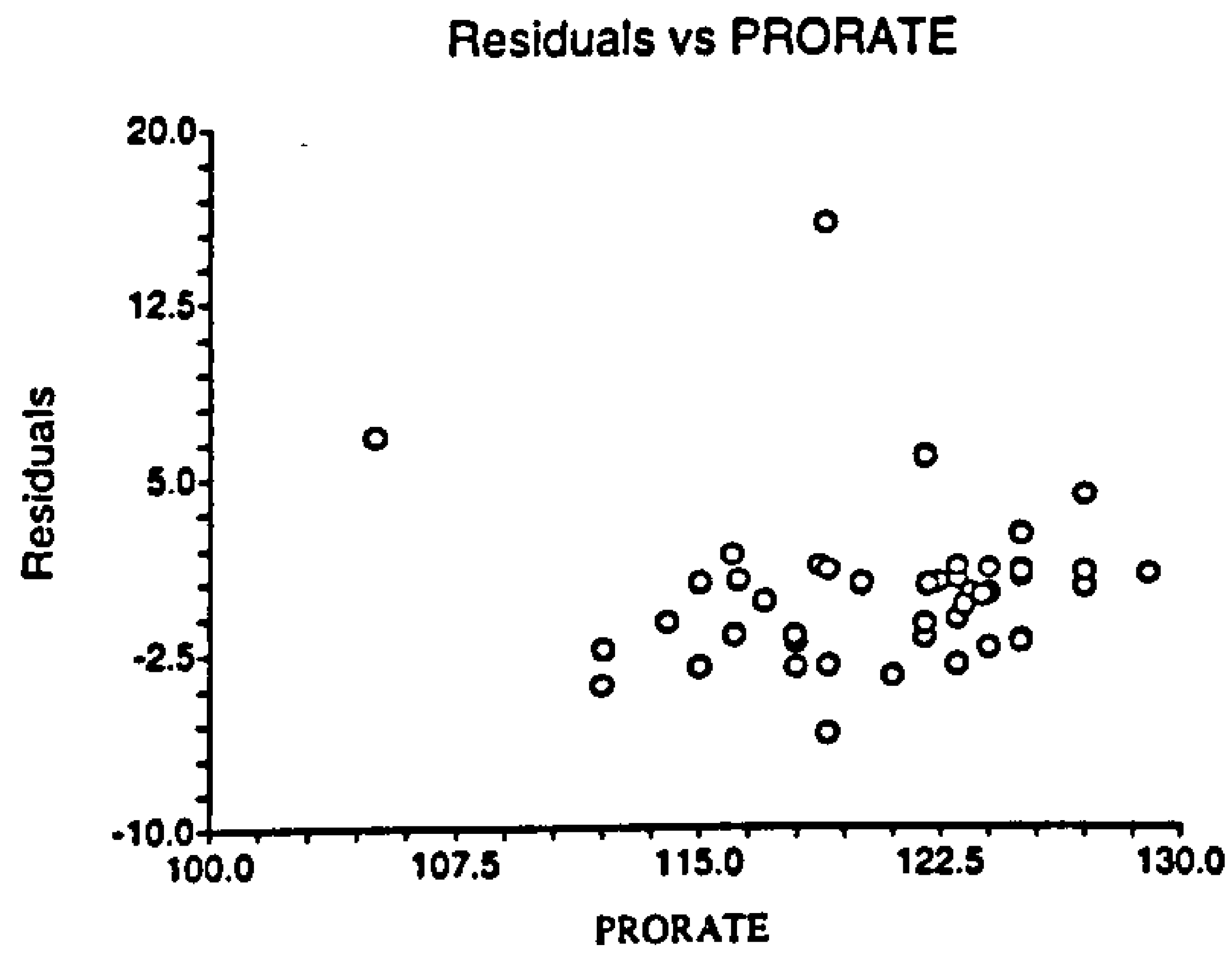


Residuals vs Row



Residuals vs Predicted





5) It seems reasonable from the ***final robust regression*** report (see Table (7.28)) that R-squared value is 0.679696. This means that this model explains about 68% of the variability in the independent variable. This is a reasonable coefficient of determination.

Table (7.28) Final Run of Andrew's sine Robust Regression

Robust Regression Report

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	42.42273	2.398746	17.6854	0.000000		Reject
Ho	1.000000					
PRORATE	-0.1915063	1.981893E-02	-9.6628	0.000000		Reject
Ho	1.000000					
R-Squared	0.679696					

Model

FUEL=42.42273-.1915063*PRORATE

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	42.42273	2.398746	37.58838	47.25708	0.000000
PRORATE	-0.1915063	1.981893E-02	-0.2314487	-0.1515639	-0.824437
T-Critical	2.015368				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	11586.74	11586.74			
Model	1	25.46377	25.46377	93.3697	0.000000	1.000000
Error	44	11.99967	0.2727199			
Total(Adjusted)	45	37.46344	0.8325209			

Root Mean Square Error	**0.5222259**	**R-Squared**	**0.679696**
Mean of Dependent Variable	**19.26173**	**Adj R-Squared**	**0.672417**
Coefficient of Variation	**2.711209E-02**		

6. Conclusions:

Comparing the final runs of the OLS multiple regression model and the final run of the robust regression, we observe that

1) Some of the assumptions of the OLS multiple regression are not valid, namely, the normality assumption and the existance of some outliers. However, this problem does not affect the robust regression.
2) The final multiple regression model and the robust regression model have only the same single significant independent variable, namely, PRORATE.
3) For multiple regression R^2=0.32, while for robust regression R^2=0.68. This means that the robust regression explains much more of the variability in the dependent variable than the multiple regression does.
4) Based on the above observations, we recommend using the final robust estimated regression mode, which is given by

   FUEL=42.42273-.1915063*PRORATE

7.5.5 Presentation of Regression Analysis of Kiln 4

7.5.5.1 EL for Kiln 4

From the full output reports, which are given in Appendices (03)-(08), we obtain the following summary of the results.

1- Table (7.29) provides the results of *checking the underlying assumptions*. It seems reasonable from this table that we have only problems with the normality assumption and the existence of outliers. Other assumptions are satisfied. This suggests using robust regression procedure.

Table (7.29) Check of the Regression Assumptions

	Tools	Result
1. Linearity	1. Scatter Plots	Only with AvHOURS , PRORATE
	2. Correlation Matrix	Only with AvHOURS , PRORATE
2. Normality of Residuals	1. Skewness Tests	Rejected
	2. Kurtosis Test	Rejected
	3. Omnibus Test	Rejected
	4. Normal Probability Plot	Almost normal
	5. Histogram	Almost normal
3. Constant Error Variance	1. Scatter Plots	No patterns
4. Independent Errors	1. Serial Correlations	Uncorrelated
	2. Durbin-Watson Test	Uncorrelated
5. Multicollinearity	1. Correlation Matrix	Absolute maximum 0.674955
	2. Condition Numbers	All are less than 100
	3. Variance Inflation Factors	All are less than 10
	4. Eigenvalues	No values are close to zero
6. Outliers	1. Histogram	Outliers Exist
	2. Normal Probability Plot	Outliers Exist
	3. Scatter Plots	Outliers Exist

2. Table (7.30) reports a summary of the results of *all possible regression models*. It seems reasonable from this table that based on the Cp criterion the best model is the one with only two independent variables, namely AvHOURS, and PRORATE. This result agrees with those obtained by checking the assumptions of the regression model. Moreover, the difference in R-squared between the full model with seven variables and the model with these two variables is about 0.03 which is not worth complicating the model.

Table (7.30) Results of All Possible Regression Procedure

Model Size	R-Squared	Root MSE	Cp	Variables in Model	Best Model
1	0.478692	2.335238	17.435451	AvHOURS	
2	0.642689	1.954412	1.565417	AvHOURS, PRORATE	This One
3	0.675762	1.940209	2.056766	AvHOURS, PRORATE, Sratio	
4	0.689051	1.924871	2.537989	AvNO, AvHOURS, PRORATE, Sratio	
5	0.692094	1.941139	4.190139	AvNO, AvHOURS, PRORATE, AVL, Sratio	
6	0.693707	1.962753	6.005784	AvNO, AvHOURS, PRORATE, AVL, Aratio, Sratio	
7	0.693758	1.990431	8.000000	AvNO, AvHOURS, PRORATE, AVL, Aratio, Sratio, LimeSF	

3. Table (7.31) reports a summary of the ***initial regression models*** obtained using different procedures together with their R-squared values. It seems reasonable from this table that:

i) The coefficients of the estimated models by sequentially entering one variable at a time are almost the same. This is seen from all possible regression models section. This indicates a type of consistancy of the results.

ii) The robust regression models gave slightly different results because they use different truncation methods in weighing the error terms. However, the largest R-squared is obtained by the Andrew's sine method.

iii) The coefficients of each of the influential variables (AvHOURS and PRORATE) have the same signs in all models. The AvHOURS has positive sign which means that the dependent variable increases as AvHOURS increases. However, The PRORATE has a negative sign, this means that the dependent variable decreases as PRORATE increases. The signs of the other independent variables are not important since their coefficients are not significantly different from zeros. This means that they do not have significant influence on the dependent variable.

iv) Since the data has some outliers and normality is not completely satisfied, we recommend using the robust regression model after deleting the independent variables which do not contribute to the dependent variable. This will be done in the final run section.

Table (7.31) Comparison of the estimated models using different procedures

Procedure	Intercept	AvNO	AvHOUR	PRORAT	AVL	Aratio	Sratio	LimeSF	R-Squar
All Possible									
Size 1	26.40998		.6012123						0.478692
Size 2	45.48269		.5393302	-.7251238					0.642689
Size 3	64.60806		.5574555	-.7333947			-7.7567		0.675762
Size 4	64.88363	3.65526	.5558688	-.7394745			-8.171618		0.689051
Size 5	60.94692	5.29567	.5555202	-.7434682	0.0567557		-8.833318		0.692094
Size 6	66.21109	5.07463	.5501702	-.7764478	0.0449554	.439819	-10.40267		0.693707
Size 7 (Multiple)	69.28985	5.10581	.5499542	-.778922	0.0460286	.430390	-10.27637	-.035973	0.693758
Stepwise									
1- Forward	45.48269		.5393302	-.7251238					0.642689
2- Step	45.48269		.5393302	-.7251238					0.642689
3-Backward	45.48269		.5393302	-.7251238					0.642689
4-Min MSE	45.48269		.5393302	-.7251238					0.642689
Robust									
1-Andrew's Sine	40.66003	-1.31434	.5713608	-.6175287	-0.111305	.586734	-15.0595	.5151706	0.889237
2-Tukey's Biweight	71.68401	1.11419	.545378	-.7378607	-0.090268	1.10584	-15.27019	.192758	0.814294
3-Least Abs Dev.	53.29206	.900403	.5724383	-.6781944	-0.066995	.569365	-13.90032	.3188286	0.856053

4) It seems reasonable from the ***final multiple regression*** report(see Table (7.32)) that

i) R-squared value is 0.642689. This means that this model explains about 64% of the variability in the independent variable. This is a reasonable coefficient of determination.

ii) The PRESS R-squared is 0.5961, which means that this model can predict to good accuracy about 60% of the values of Y. This is a reasonable predictability percentage.

iii) There seems to be a problem in the normality of the residuals because all three tests rejected normality assumption.

iv) Based on serial correlations and the Durbin-Watson test, the residuals are uncorrelated.

v) Multicollinearity could not be a problem.

vi) From the plot section, one observes that the data has some outliers.

Table (7.32) Final Run of Multiple Regression

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	45.48269	4.421327	10.2871	0.000000	Reject Ho	1.000000
AvHOURS	0.5393302	8.236534E-02	6.5480	0.000000	Reject Ho	0.999995
PRORATE	-0.7251238	0.1671578	-4.3380	0.000091	Reject Ho	0.988527
R-Squared	0.642689					

Model
EL=45.48269+0.5393302 AvHOURS-0.7251238 PRORATE

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	45.48269	4.421327	36.55363	54.41174	0.0000
AvHOURS	0.5393302	8.236534E-02	0.37299	0.7056704	0.6207
PRORATE	-0.7251238	0.1671578	-1.062706	-0.3875418	-0.4112
T-Critical	2.019541				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	37062.02	37062.02			
Model	2	301.4063	150.7032	36.8729	0.000000	1.000000
Error	41	167.5709	4.087096			
Total(Adjusted)	43	468.9773	10.90645			

Root Mean Square Error	**2.021657**	**R-Squared**	**0.6427**
Mean of Dependent	**29.02273**	**Adj R-Squared**	**0.6253**
Coefficient of Variation	**6.965771E-02**	**Press Value**	**189.4116**
Sum |Press Residuals|	**69.39913**	**Press R-Squared**	**0.5961**

Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	-1.9290	0.053728	Accepted
Kurtosis	2.0655	0.038879	Rejected
Omnibus	7.9873	0.018432	Rejected

Serial-Correlation Section

Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.153278	9	-0.096050	17	-0.142750
2	0.037799	10	-0.023122	18	-0.202290
3	-0.089417	11	0.063234	19	-0.132233
4	-0.123005	12	0.093254	20	-0.014997
5	-0.185412	13	0.167895	21	-0.010158
6	-0.031247	14	0.096943	22	0.015079
7	-0.220749	15	-0.053078	23	0.065212
8	-0.020809	16	-0.025421	24	0.049925

Above serial correlations significant if their absolute values are greater than 0.301511

Durbin-Watson Value 1.6926

Multicollinearity Section

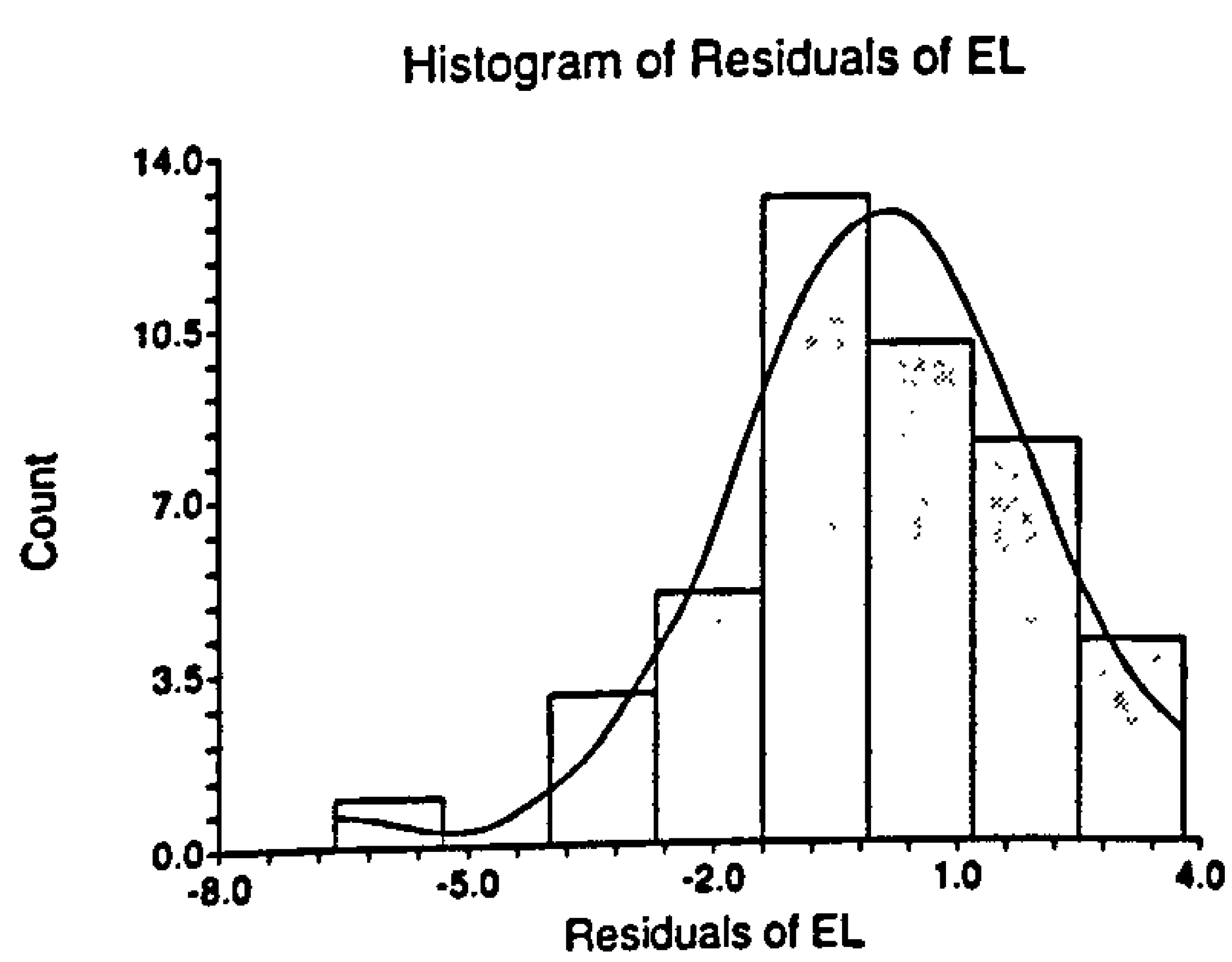
Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance	Diagonal of X'X Inverse
AvHOURS	1.030924	0.029996	0.970004	1.65987E-03
PRORATE	1.030924	0.029996	0.970004	6.836573E-03

Eigenvalues of Centered Correlations

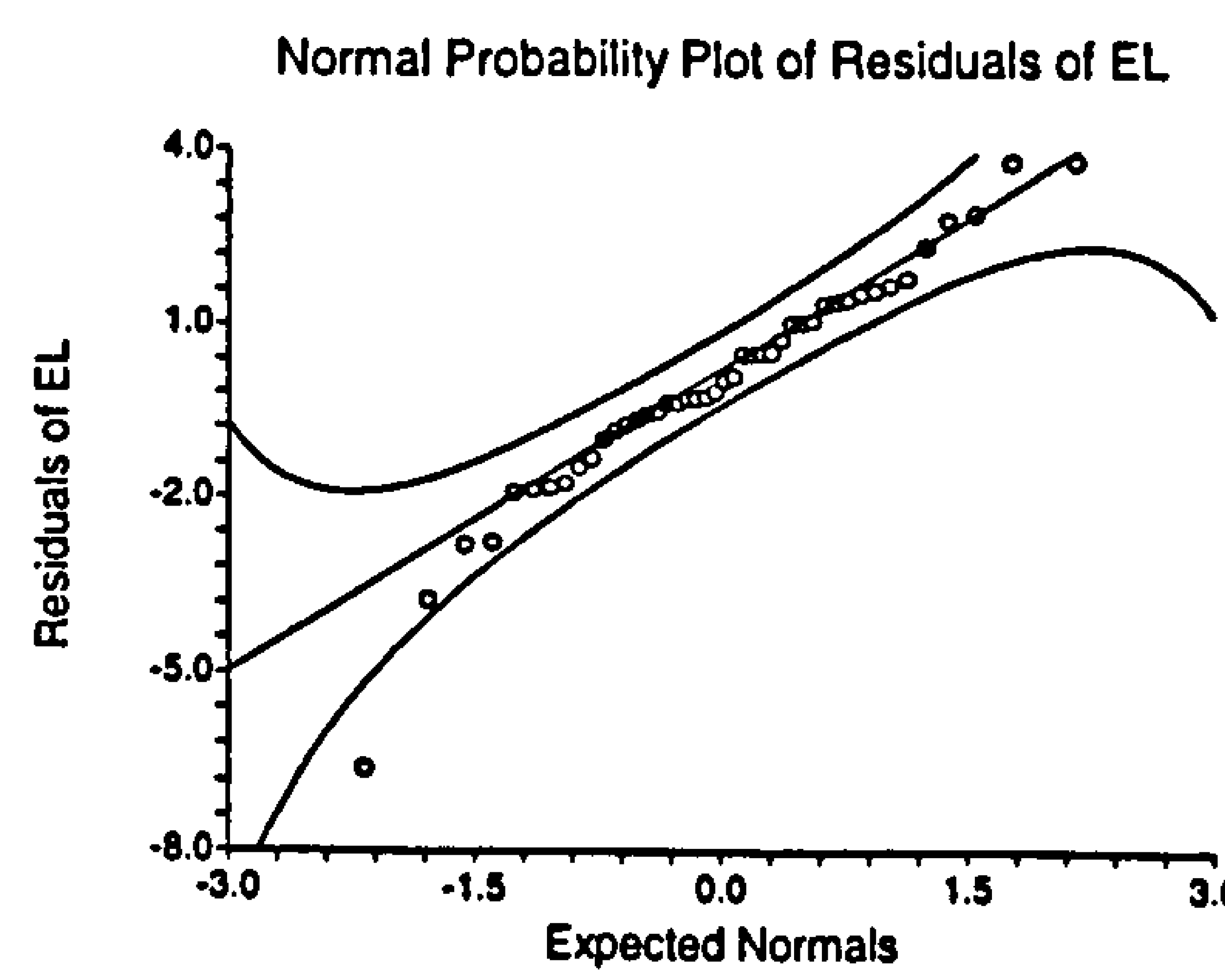
No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.173195	58.66	58.66	1.00
2	0.826805	41.34	100.00	1.42

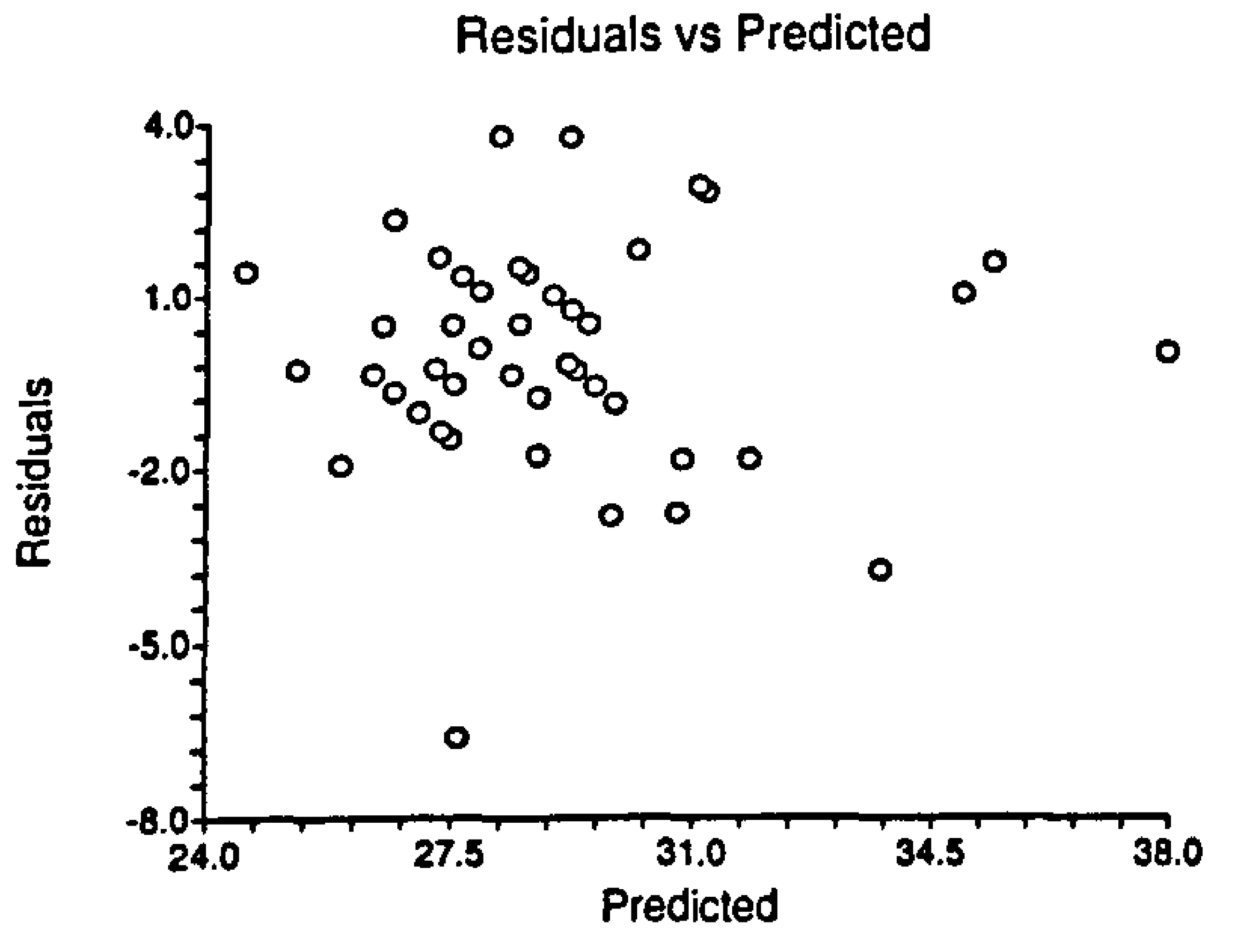
All Condition Numbers less than 100. Multicollinearity is NOT a problem.

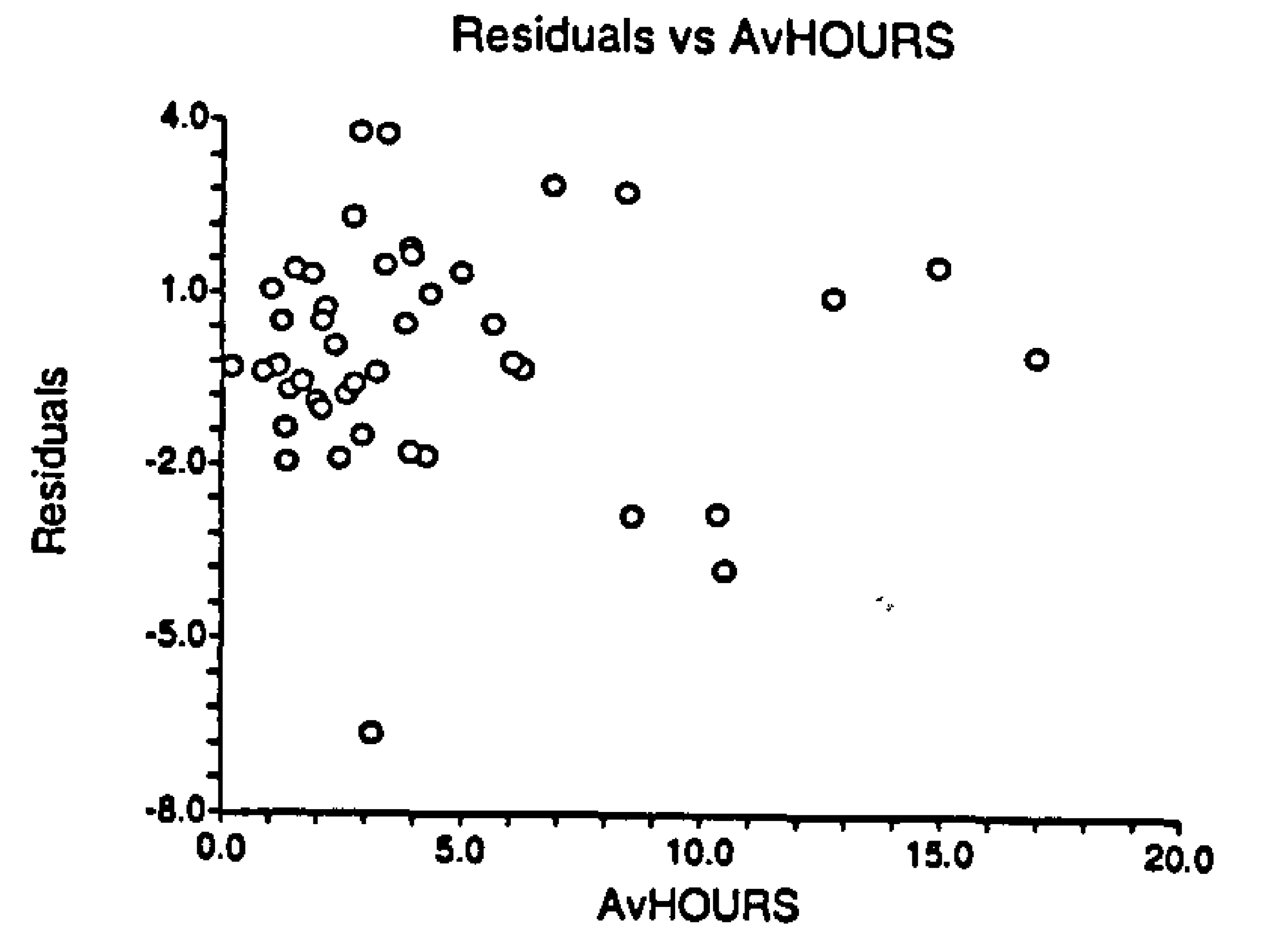
Plots Section

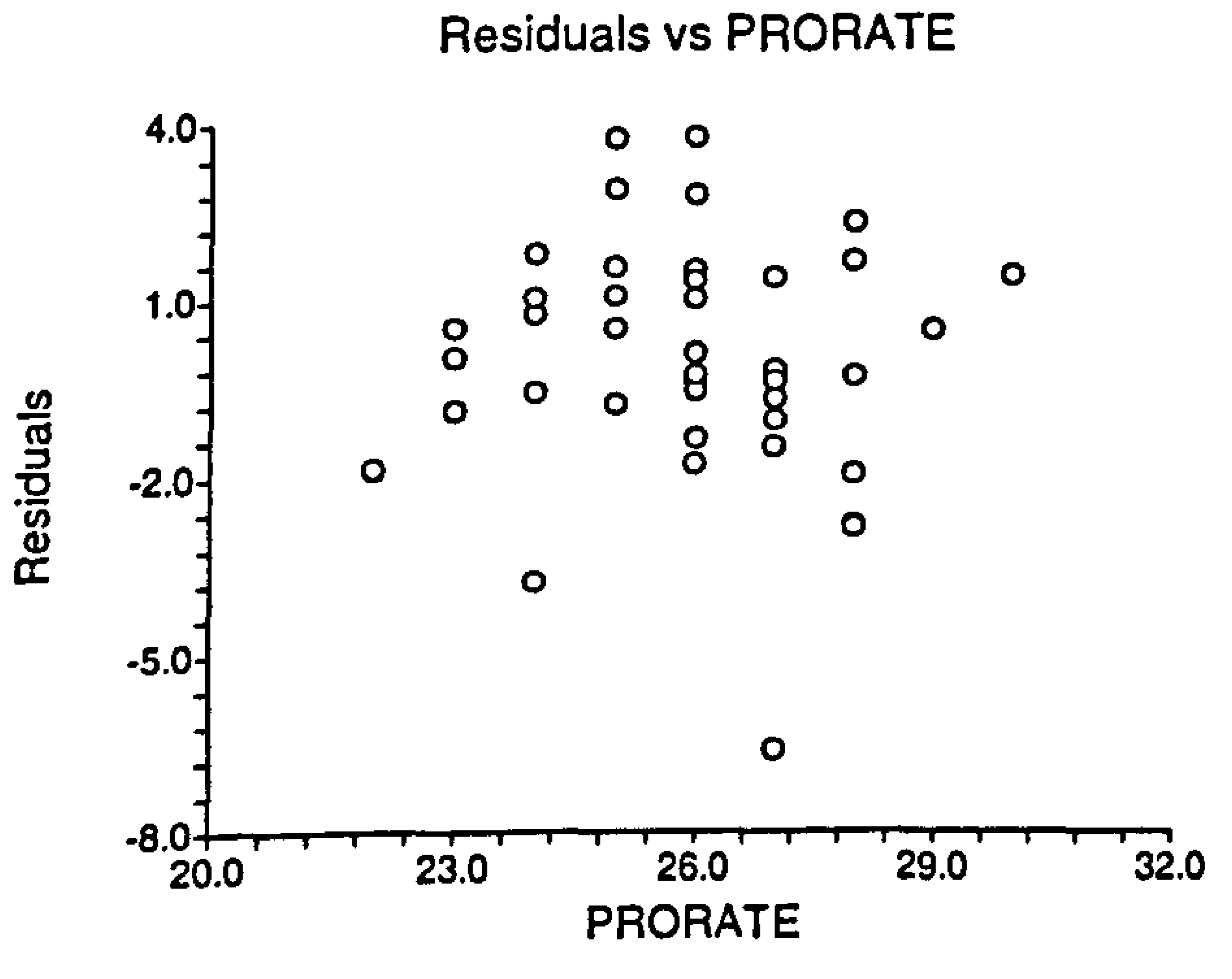


415









5. Table (7.33) provides the ***final robust regression*** report. It seems reasonable from this report that

i) the estimated regression model is

EL = 40.70766+ .6352955*AvHOURS-.5509918*PRORATE

ii) All the independent variables in the model contribute to the dependent variable.

iii) The R-squared value is 0.861101, which means that this model explains about 86% of the variability in EL by the independent variables in the model.

Table (7.33) Final run of robust regression

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	40.70766	2.561289	15.8934	0.000000	Reject Ho	1.000000
AvHOURS	0.6352955	4.827703E-02	13.1594	0.000000	Reject Ho	1.000000
PRORATE	-0.5509918	9.671085E-02	-5.6973	0.000001	Reject Ho	0.999841

Model

EL = 40.70766+ .6352955*AvHOURS-.5509918*PRORATE

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	27209.55	27209.55			
Model	2	267.8359	133.9179	127.0888	0.000000	1.000000
Error	41	43.20314	1.053735			
Total(Adjusted)	43	311.039	7.233465			
Root Mean Square Error		1.026516	R-Squared	0.861101		
Mean of Dependent Variable		29.01167	Adj R-Squared	0.854325		
Coefficient of Variation		3.538286E-02				

6. Conclusions:

Comparing the final runs of the OLS multiple regression model and the final run of the robust regression, we observe that

1) Some of the assumptions of the OLS multiple regression are not valid, namely, the normality assumption and the existance of some outliers. However, this problem does not affect the robust regression.

2) The final multiple regression model and the robust regression model have only the same two significant independent variables, namely, AvHOURS, PRORATE.

3) For multiple regression R^2=0.64, while for robust regression R^2=0.86. This means that the robust regression explains much more of the variability in the dependent variable than the multiple regression does.

4) Based on the above observations, we recommend using the final robust estimated regression mode, which is given by

   EL = 40.70766+ .6352955*AvHOURS-.5509918*PRORATE

7.5.5.2 FUEL for Kiln 4

From the full output reports, which are given in Appendices (03)-(08), we obtain the following summary of the results.

1- Table (7.34) provides the results of ***checking the underlying assumptions***. It seems reasonable from this table that we have problems with the existence of outliers and residuals being correlated. Other assumptions are satisfied. This suggests using robust regression procedure.

Table (7.34) Check of the Regression Assumptions

	Tools	Result
1. Linearity	1. Scatter Plots	Only with PRORATE, Aratio
	2. Correlation Matrix	Only with PRORATE, Aratio
2. Normality of Residuals	1. Skewness Tests	Accepted
	2. Kurtosis Test	Accepted
	3. Omnibus Test	Accepted
	2. Normal Probability Plot	Almost normal
	3. Histogram	Almost normal
3. Constant Error Variance	1. Scatter Plots	No patterns
4. Independent Errors	1. Serial Correlations	Correlated
	2. Durbin-Watson Test	Correlated
5. Multicollinearity	1. Correlation Matrix	Absolute maximum 0.674955
	2. Condition Numbers	All are less than 100
	3. Variance Inflation Factors	All are less than 10
	4. Eigenvalues	No values are close to zero
6. Outliers	1. Histogram	Outliers Exist
	2. Normal Probability Plot	Outliers Exist
	3. Scatter Plots	Outliers Exist

2. Table (7.35) reports a summary of the results of ***all possible regression models***. It seems reasonable from this table that based on the Cp criterion the best model is the one with only four independent variables, namely AvNO, PRORATE, AVL, and Aratio. Moreover, the difference in R-squared between the full model with seven variables and the model with these four variables is about 0.04 which is not worth complicating the model.

Table (7.35) Results of All Possible Regression Procedure

Model Size	R-Squared	Root MSE	Cp	Variables in Model	Best Model
1	0.306427	4.099166	13.509869	PRORATE	
2	0.409871	3.898711	9.341296	PRORATE, Aratio	
3	0.464826	3.760042	7.025814	AvNO, PRORATE, Aratio	
4	0.515544	3.624206	5.043059	AvNO, PRORATE, AVL, Arati(	This One
5	0.536422	3.592839	5.403515	AvNO, AvHOURS, PRORATE, AVL, Aratio	
6	0.550562	3.586419	6.293168	AvNO, AvHOURS, PRORATE, AVL, Aratio, Sratio	
7	0.554295	3.622155	8.000000	AvNO, AvHOURS, PRORATE, AVL, Aratio, Sratio, LimeSF	

3. Table (7.36) reports a summary of the obtained ***initial regression models*** using different procedures together with their R-squared values. It seems reasonable from this table that

i) The coefficients of the estimated models by sequentially entering one variable at a time are almost the same. This is seen from all possible regression models section. This indicates a type of consistancy of the results.

ii) The robust regression models gave slightly different results because they use different truncation methods in weighing the error terms. However, the largest R-squared is obtained by the Andrew's sine method.

iii) The coefficients of each of the influential variables (PRORATE, Aratio, and AvNO) have the same signs in all models. These signs agree with the researcher expectations. It should be mentioned here that coefficient of AVL is not significantly different from zero based on multiple regression. However, based on the Cp criterion, AVL was among the selected variables in the best model together with the above three variables. For this reason we will provide to final multiple regression models, one with AVL and the other without AVL.

iv) Since the data has some outliers and residuals are correlated, we recommend using the robust regression model after deleting the independent variables, which do not contribute to the dependent variable. This will be done in the final run section.

Table (7.36) Comparison of the estimated models using different procedures

Procedure	Intercept	AvNO	AvHOUR	PRORAT	AVL	Aratio	Sratio	LimeSF	R-Squar
All Possible									
Size 1	131.3087			-1.46676					0.306427
Size 2	132.7447			-1.31143		-3.74550			0.409871
Size 3	130.222	11.20576		-1.32948		-3.60287			0.464826
Size 4	94.1445	21.42626		-1.27890	.363221	-4.62388			0.515544
Size 5	97.51344	21.40608	-.1924092	-1.36774	.355185	-4.25253			0.536422
Size 6	61.19806	20.82551	-.193908	-1.24554	.353236	-5.22403	14.1708		0.550562
Size 7 (Multiple)	21.31061	20.42155	-.1911103	-1.21348	.339332	- 5.10187	12.5344	.4660547	0.554295
Stepwise									
1- Forward	132.7447			-1.31143		-3.74550			0.409871
2- Step	132.7447			-1.31143		-3.74550			0.409871
3-Backward	94.1445	21.42626		-1.27890	.363221	-4.62388			0.515544
4-Min MSE	94.1445	21.42626		-1.27890	.363221	-4.62388			0.515544
Robust									
1-Least Abs. Dev. 1.0	101.6708	28.22709	-.208469	-1.29606	.531040	-3.61594	-5.5598	-.1171465	0.787937
2-Tukey's Biweight 6.0	137.756	36.67571	-.238045	-1.11186	.583440	-3.34833	-11.596	-.4696321	0.802041
3-Andrew's Sine 2.1	173.7851	31.05059	-.1714061	-1.46392	.586872	-1.96470	-15.905	-.6520774	0.871107

4) It seems reasonable from the ***final multiple regression*** report with (see Table (7.37)) that:

i) The AvNO variable which was significant in the full model, is no longer significant in this final run of the model with three independent variables. On the other hand, AVL is not significant in the final run with four variables. This type of inconsistancy of the results concerning AVL and AvNO may be due to the lack of validity of some assumptions of the OLS multiple regression model.

ii) R-squared value based on the model with three variables is 0.464826. This means that this model explains about 46% of the variability in the independent variable. This is a relatively low coefficient of determination.

iii) The PRESS R-squared is 0.3336, which means that this model can predict to good accuracy about 33% of the values of Y. This is a low percentage of predictability.

iv) The normality of the residuals is acceptable.

v) Based on serial correlations and the Durbin-Watson test, the residuals may be correlated since the serial correlation at lag one is larger than the specified critical point.

vi) Multicollinearity should not be a problem.

vii) From the plot section, one observes that the data has some outliers. Moreover, from the plot of residuals versus predicted values, It seems reasonable that the residuals are high in the middle and low at the the two extreems, i.e. the constant variance assumption is not valid.

Based on the above results we do not recommend using this model because the residuals do not satisfy some of the required assumptions, and the R-squared value is low. Since two of the stepwise selection procedures selected three variables and the other two procedures selected four variables out of seven variables we will run two final multiple regression models, one with three variables and the other in four variables. The results of these two runs are given in Tables (7.37a) and (7.37b).

Table (7.37a) Final Run of Multiple Regression with Three (effectively two) Variables

Regression Equation Section

Independent Variable	Coefficient	Regression Error	Standard (Ho: B=0)	T-Value Level	Prob (5%)	Decision (5%)	Power
Intercept	130.222	8.090147	16.0964	0.000000	Reject Ho	1.000000	
AvNO	11.20576	5.599541	2.0012	0.052365	Accept Ho	0.496795	
PRORATE	-1.32948	0.3181363	-4.1790	0.000160	Reject Ho	0.982745	
Aratio	-3.60287	1.567088	-2.2991	0.026942	Reject Ho	0.611140	

R-Squared 0.464826

Model

FU=130.222+11.20576 AvNO-1.32948 PRORATE-3.60287 Aratio

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	130.222	8.090147	113.8581	146.5858	0.0000
AvNO	11.20576	5.599541	-0.1203811	22.5319	0.2347
PRORATE	-1.32948	0.3181363	-1.972972	-0.685989	-0.5071
Aratio	-3.60287	1.567088	-6.772605	-0.4331351	-0.2792
T-Critical	2.022691				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	375075.7	375075.7			
Model	3	478.9002	159.6334	11.2912	0.000018	0.998581
Error	39	551.3788	14.13792			
Total(Adjusted)	42	1030.279	24.53045			

Root Mean Square Error	**3.760042**	**R-Squared**	**0.4648**
Mean of Dependent	**93.39535**	**Adj R-Squared**	**0.4237**
Coefficient of Variation	**4.025942E-02**	**Press Value**	**686.6133**
Sum |Press Residuals|	**144.1882**	**Press R-Squared**	**0.3336**

Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	0.4515	0.651619	Accepted
Kurtosis	-0.9924	0.320993	Accepted
Omnibus	1.1888	0.551905	Accepted

Serial-Correlation Section

Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.364623	9	0.140202	17	0.050611
2	-0.002239	10	0.232979	18	-0.124319
3	-0.041616	11	0.006610	19	-0.075969
4	-0.030170	12	-0.029141	20	0.082483
5	-0.287814	13	-0.030732	21	-0.010742
6	-0.070967	14	-0.131439	22	-0.187132
7	0.016814	15	-0.241367	23	-0.020244
8	0.050530	16	0.000853	24	0.026821

Above serial correlations significant if their absolute values are greater than 0.304997

Durbin-Watson Value 1.1594

Multicollinearity Section

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance	Diagonal of X'X Inverse
AvNO	1.002366	0.002360	0.997640	2.217785
PRORATE	1.073156	0.068169	0.931831	7.158811E-03
Aratio	1.074517	0.069349	0.930651	0.1737006

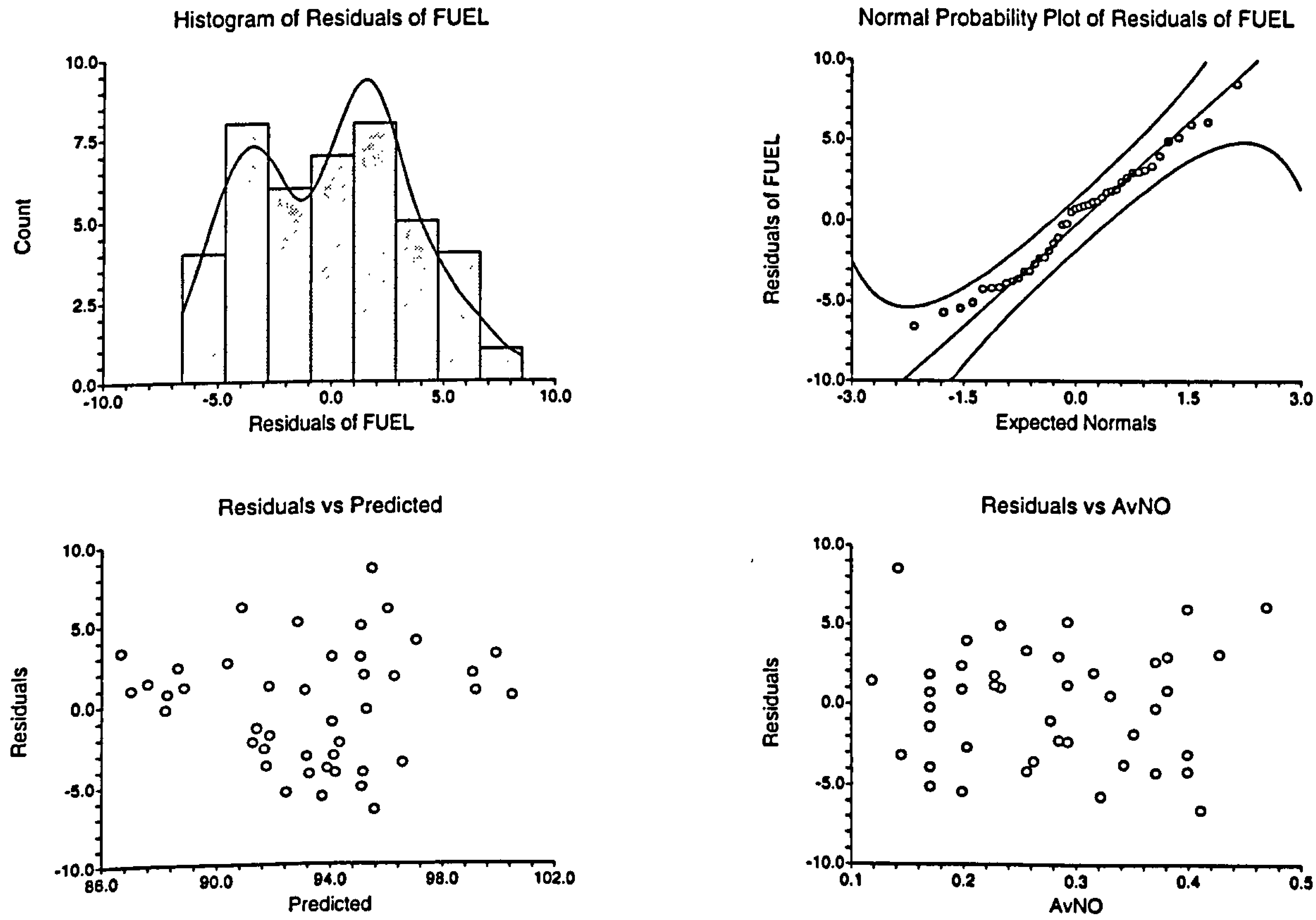
Eigenvalues of Centered Correlations

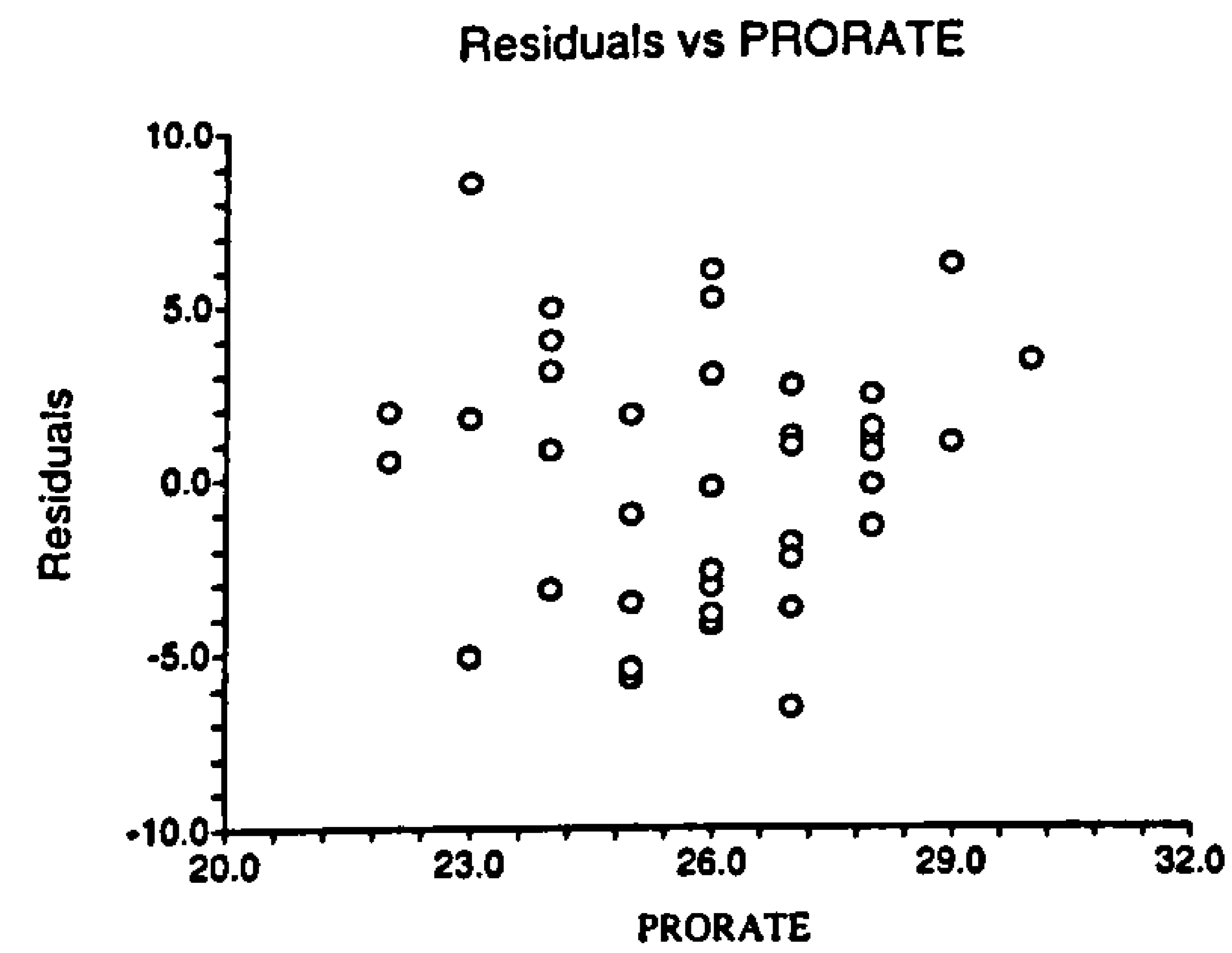
No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.260626	42.02	42.02	1.00
2	1.005061	33.50	75.52	1.25
3	0.734313	24.48	100.00	1.72

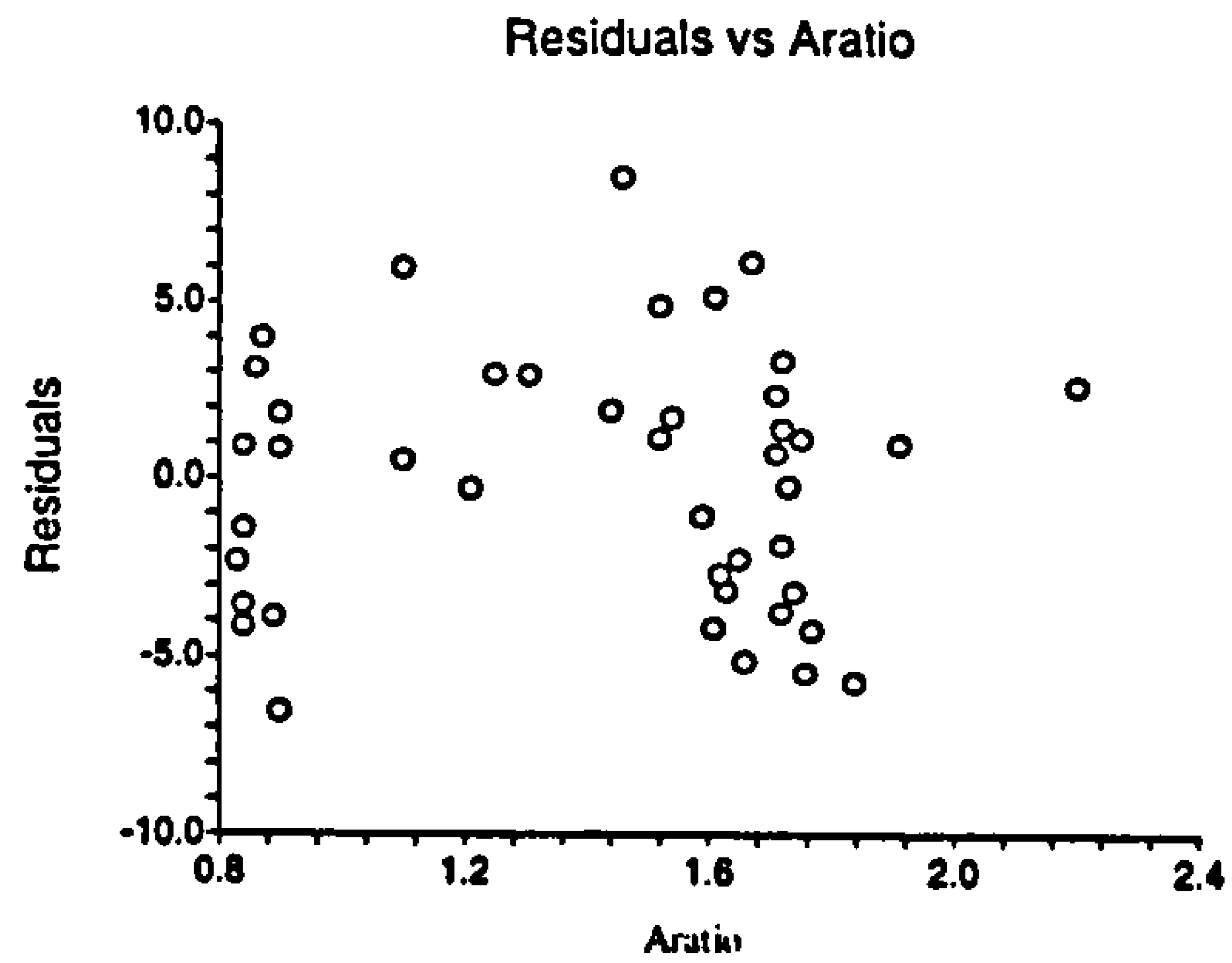
All Condition Numbers less than 100. Multicollinearity is NOT a problem.

Plots Section

423







<u>**Table (7.37b) Final Run of Multiple Regression with Four (effectively 3) Variables**</u>

Independent Variable	Regression Coefficient (5%)	Standard Error (5%)	T-Value	Prob	Decision (Ho: B=0)	Power Level
Intercept	94.1445	19.69726	4.7796	0.000026	Reject Ho	0.996489
AvNO	21.4262	7.442296	2.8790	0.006517	Reject Ho	0.801107
PRORATE	-1.27890	0.30768	-4.1565	0.000177	Reject Ho	0.981663
AVL	0.3632	0.18210	1.9946	0.053301	Accept Ho	0.493697
Aratio	-4.6238	1.59485	-2.8992	0.006183	Reject Ho	0.806565
R-Squared	0.515544					

Model
FUEL= 94.1445+ 21.42626*AvNO-1.278909*PRORATE+ .3632214*AVL-4.62388*Aratio

5. Table (7.38) provides the <u>*final robust regression*</u> report. It seems reasonable from this report that:

i) the estimated regression model is FUEL = 81.24993+ 24.31024*AvNO-1.345305*PRORATE+ .5033563*AVL-3.914322*Aratio

ii) All the independent variables in the model contribute to the dependent variable.

iii) The R-squared value is 0.768843, which means that this model explains about 77% of the variability in FUEL by the independent variables in the model.

Table (7.38) Final run of robust regression

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	81.24993	12.07103	6.7310	0.000000	Reject Ho	0.999998
AvNO	24.31024	5.043628	4.8200	0.000023	Reject Ho	0.996882
PRORATE	-1.345305	0.21277	-6.3228	0.000000	Reject Ho	0.999986
AVL	0.5033563	0.1037308	4.8525	0.000021	Reject Ho	0.997169
Aratio	-3.914322	1.000345	-3.9130	0.000365	Reject Ho	0.968011

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	259078.8	259078.8			
Model	4	461.9742	115.4935	31.5976	0.000000	1.000000
Error	38	138.8954	3.655142			
Total(Adjusted)	42	600.8696	14.30642			

Root Mean Square Error	**1.911842**	**R-Squared**	**0.768843**
Mean of Dependent Variable	**93.09084**	**Adj R-Squared**	**0.744510**
Coefficient of Variation	**2.053739E-02**		

6. Conclusions:

Comparing the final runs of the OLS multiple regression model and the final run of the robust regression, we observe that

1) The only problem with the OLS multiple regression is the existance of some outliers. However, this problem does not affect the robust regression.
2) The final multiple regression model has only three significant independent variables, namely, AvNO, and Aratio. On the other hand the final robust regresion model has four significant independent variables, namely, AvNO, PRORATE, AVL, and Aratio.
3) For multiple regression R^2=0.52, while for robust regression R^2=0.77. This means that the robust regression explains much more of the variability in the dependent variable than the multiple regression does.

4) Based on the above observations, we recommend using the final robust estimated regression mode, which is given by

FUEL = 81.24993+ 24.31024*AvNO-1.345305*PRORATE+ .5033563*AVL -3.914322*Aratio

7.5.6 Presentation of Regression Analysis Analysis of Kiln 5

7.5.6.1 EL for Kiln 5

From the full output reports, which are given in Appendices (03)-(08), we obtain the following summary of the results.

1- Table (7.39) provides the results of ***checking the underlying assumptions***. It seems reasonable from this table that we have problems with the existence of outliers and residuals being correlated. This suggests using the robust regression procedure.

Table (7.39) Check of the Regression Assumptions

	Tools	Result
1. Linearity	1. Scatter Plots	Only with AvNO, AvHOURS, PRORATE, AVL, Aratio
	2. Correlation Matrix	Only with AvNO, AvHOURS, PRORATE AVL, Aratio
2. Normality of Residuals	1. Skewness Tests	Accepted
	2. Kurtosis Test	Accepted
	3. Omnibus Test	Accepted
	4. Normal Probability Plot	Almost normal
	5. Histogram	Almost normal
3. Constant Error Variance	1. Scatter Plots	No patterns
4. Independent Errors	1. Serial Correlations	Correlated
	2. Durbin-Watson Test	Correlated
5. Multicollinearity	1. Correlation Matrix	Absolute maximum 0.585609
	2. Condition Numbers	All are less than 100
	3. Variance Inflation Factors	All are less than 10
	4. Eigenvalues	No values are close to zero
6. Outliers	1. Histogram	Outliers Exist
	2. Normal Probability Plot	Outliers Exist
	3. Scatter Plots	Outliers Exist

2. Table (7.40) reports a summary of the results of ***all possible regression models***. It seems reasonable from this table that based on the Cp criterion the best model is the one with only four independent variables, namely AvNO, AvHOURS, PRORATE, AVL, Aratio, and Sratio. This result partially agrees with those obtained by checking the assumptions of the

regression model. Moreover, the difference in R-squared between the full model with seven variables and the model with these six variables is about 0.01 which is not worth complicating the model.

Table (7.40) Results of All Possible Regression Procedure

Model Size	R-Squared	Root MSE	Cp	Variables in Model	Best Model
1	0.332877	2.640648	87.616588	AVL	
2	0.554627	2.183748	47.197225	PRORATE,AVL	
3	0.690792	1.842162	23.149672	AvHOURS, PRORATE,AVL	
4	0.741333	1.706359	15.481444	AvHOURS, PRORATE,AVL,Arati	
5	0.792597	1.547918	7.674917	AvHOURS, PRORATE,AVL, Aratio,Sratio	
6	0.805211	1.520247	7.262050	AvNO,AvHOURS, PRORATE, AVL,Aratio,Sratio	This One
7	0.811808	1.514892	8.000000	AvNO, AvHOURS, PRORATE, AVL, Aratio, Sratio, LimeSF	

3. Table (7.41) reports a summary of the obtained ***initial regression models*** using different procedures together with their R-squared values. It seems reasonable from this table that

i) The coefficients of the estimated models by sequentially entering one variable at a time are almost the same. This is seen from all possible regression models section. This indicates a type of consistancy of the results.

ii) The robust regression models gave slightly different results because they use different truncation methods in weighing the error terms. However, the largest R-squared is obtained by the Andrew's sine method.

iii) The coefficients of each of the influential variables (PRORATE, Aratio, Sratio, AvNO, AvHOURS, and AVL) have the same signs in all models. The AvNO and AvHOURS variables have positive signs which means that the dependent variable increases as each of these independent variables increases. However, The PRORATE, AVL, Aratio, and Sratio variables have negative signs, this means that the dependent variable decreases as each of these variables increases. The sign of the other independent variable is not important since its coefficient is not significant. This means that it does not have significant influence on the dependent variable.

iv) Since the data has some outliers, we recommend using the robust regression model after deleting the independent variables which do not contribute to the dependent variable. This will be done in the final run section.

<u>Table (7.41) Comparison of the estimated models using different procedures</u>

Procedure	Intercept	AvNO	AvHOUR	PRORAT	AVL	Aratio	Sratio	LimeSF	R-Squa
All Possible									
Size 1	62.96053				-.323661				0.33287
Size 2	77.67101			-.299970	-.266049				0.55462
Size 3	70.36381		.3496429	-.245929					0.69079
Size 4	78.821878.82		.3610218	-.215552	-.246305	-7.306294			0.74133
Size 5	170.2305		.3277374	-.240229	-.236396	-9.95722	-35.096		0.79259
Size 6	173.2238	2.21819	.3305127	-.235694	-.189791	-9.661064	-38.731		0.80521
Size 7 (Multiple)	121.0618	2.15843	.345149	-.232833	-.183895	-8.629664	-33.664	.391420	0.81180
Stepwise									
1- Forward	170.2305		.3277374	-.240229	-.236396	-9.95722	-35.096		0.792597
2- Step	70.36381		.3496429	-.245929	-.266049				0.690792
3-Backward	173.2238	2.21819	.3305127	-.235694	-.189791	-9.661064	-38.731		0.805211
4-Min MSE	173.2238	2.21819	.3305127	-.235694	-.189791	-9.661064	-38.731		0.805211
Robust									
1-Least Abs Dev. 1.0	97.29821	1.28628	.3104494	-.212922	-.197207	-6.999886	-32.948	.598294	0.862344
2-Tukey's Biweight 6.0	110.254	1.82941	.336508	-.223910	-.191452	-7.712717	-32.582	.464017	0.824929
3-Andrew's Sine 2.1	90.29779	.775301	.2616146	-.196950	-.162685	-8.020124	-31.268	.60622	0.906371

4) It seems reasonable from the ***<u>final multiple regression</u>*** report (see Table (7.42)) that :

i) R-squared value is 0.792597. This means that this model explains about 79% of the variability in the independent variable. This is a high coefficient of determination.

ii) The PRESS R-squared is 0.7199, which means that this model can predict to good accuracy about 72% of the values of Y. This is also a reasonable predictability indicator.

iii) Normality of the residuals is accepted.

iv) Based on serial correlations and the Durbin-Watson test, the residuals may be correlated.

v) Multicollinearity could not be a problem.

vi) From the plot section, one observes that the data has some outliers.

Table (7.42) Final Run of Multiple Regression

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	170.2305	30.31622	5.6152	0.000002	Reject Ho	0.999775
AvHOURS	0.3277374	7.093763E-02	4.6201	0.000043	Reject Ho	0.994466
PRORATE	-0.240229	4.996301E-02	-4.8081	0.000024	Reject Ho	0.996771
AVL	-0.2363961	4.268832E-02	-5.5377	0.000002	Reject Ho	0.999702
Aratio	-9.95722	2.552045	-3.9017	0.000378	Reject Ho	0.967213
Sratio	-35.09634	11.45171	-3.0647	0.003996	Reject Ho	0.847663
R-Squared	0.792597					

Model

EL=170.2305+0.3277374 AvHOURS-0.240229 PRORATE-0.2363961 AVL-9.95722 Aratio -35.09634 Sratio

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	170.2305	30.31622	108.8586	231.6025	0.0000
AvHOURS	0.3277374	7.093763E-02	0.1841317	0.4713431	0.3590
PRORATE	-0.240229	4.996301E-02	-0.3413738	-0.1390842	-0.3797
AVL	-0.2363961	4.268832E-02	-0.322814	-0.1499781	-0.4214
Aratio	-9.95722	2.552045	-15.12357	-4.790876	-0.3173
Sratio	-35.09634	11.45171	-58.27912	-11.91356	-0.2484
T-Critical	2.024394				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	49379	49379			
Model	5	347.9501	69.59003	29.0437	0.000000	1.000000
Error	38	91.04987	2.396049			
Total(Adjusted)	43	439	10.2093			

Root Mean Square Error	1.547918	R-Squared	0.7926
Mean of Dependent	33.5	Adj R-Squared	0.7653
Coefficient of Variation	0.0462065	Press Value	122.9501
Sum |Press Residuals|	60.95128	Press R-Squared	0.7199

Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	1.0314	0.302335	Accepted
Kurtosis	-0.8401	0.400842	Accepted
Omnibus	1.7697	0.412784	Accepted

Serial-Correlation Section

Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.400400	9	0.088349	17	-0.134995
2	0.285875	10	0.078795	18	-0.202207
3	0.144053	11	-0.107417	19	-0.220969
4	0.293665	12	-0.255053	20	-0.161037
5	0.124336	13	-0.142543	21	-0.221713
6	0.131367	14	-0.029195	22	-0.255162
7	0.124174	15	-0.051084	23	-0.156053
8	0.099114	16	-0.213589	24	-0.101301

Above serial correlations significant if their absolute values are greater than 0.301511

Durbin-Watson Value 1.1742

Multicollinearity Section

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance	Diagonal of X'X Inverse
AvHOURS	1.106202	0.096006	0.903994	2.100185E-03
PRORATE	1.142521	0.124742	0.875258	1.041841E-03
AVL	1.060937	0.057437	0.942563	7.605405E-04
Aratio	1.211945	0.174880	0.825120	2.718197
Sratio	1.203617	0.169171	0.830829	54.73249

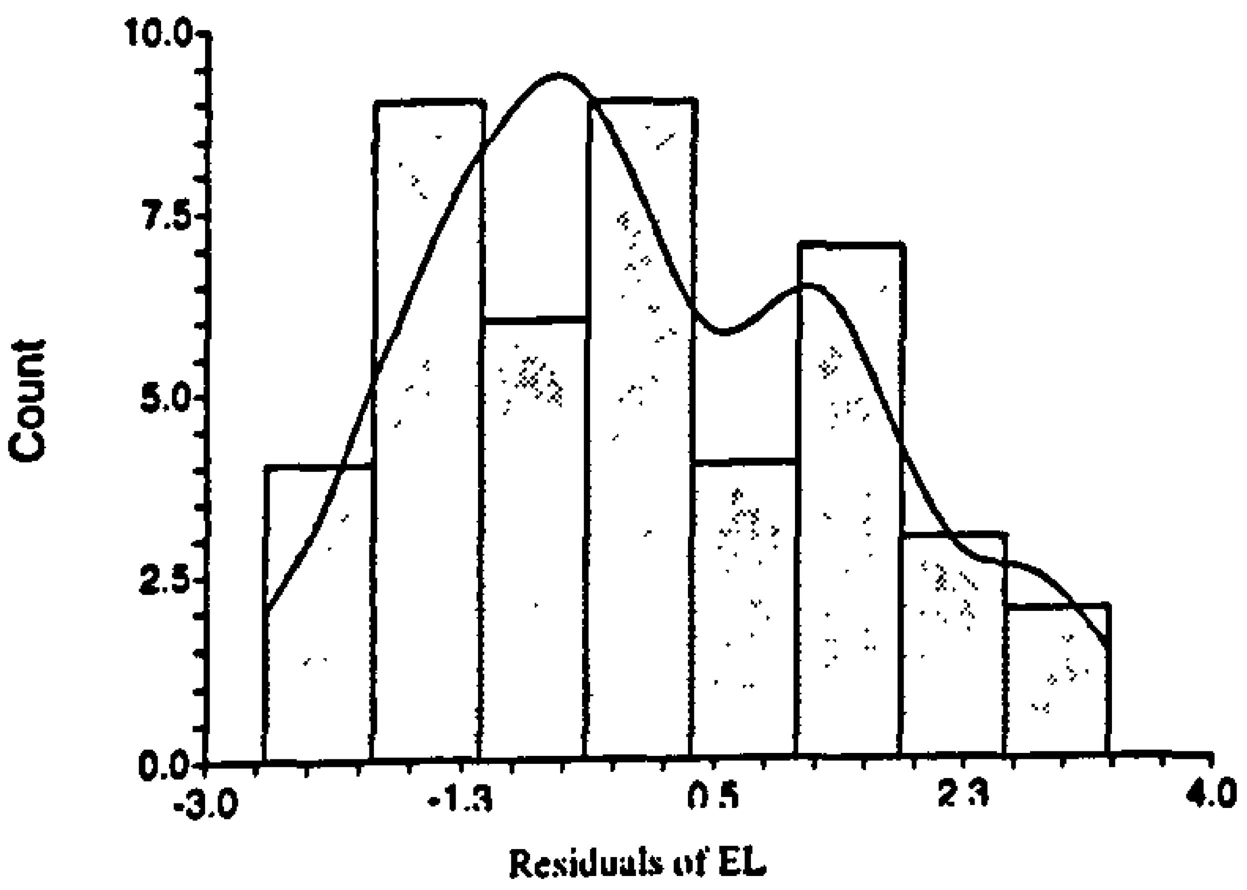
Eigenvalues of Centered Correlations

No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.577810	31.56	31.56	1.00
2	1.248592	24.97	56.53	1.26
3	0.918559	18.37	74.90	1.72
4	0.670063	13.40	88.30	2.35
5	0.584977	11.70	100.00	2.70

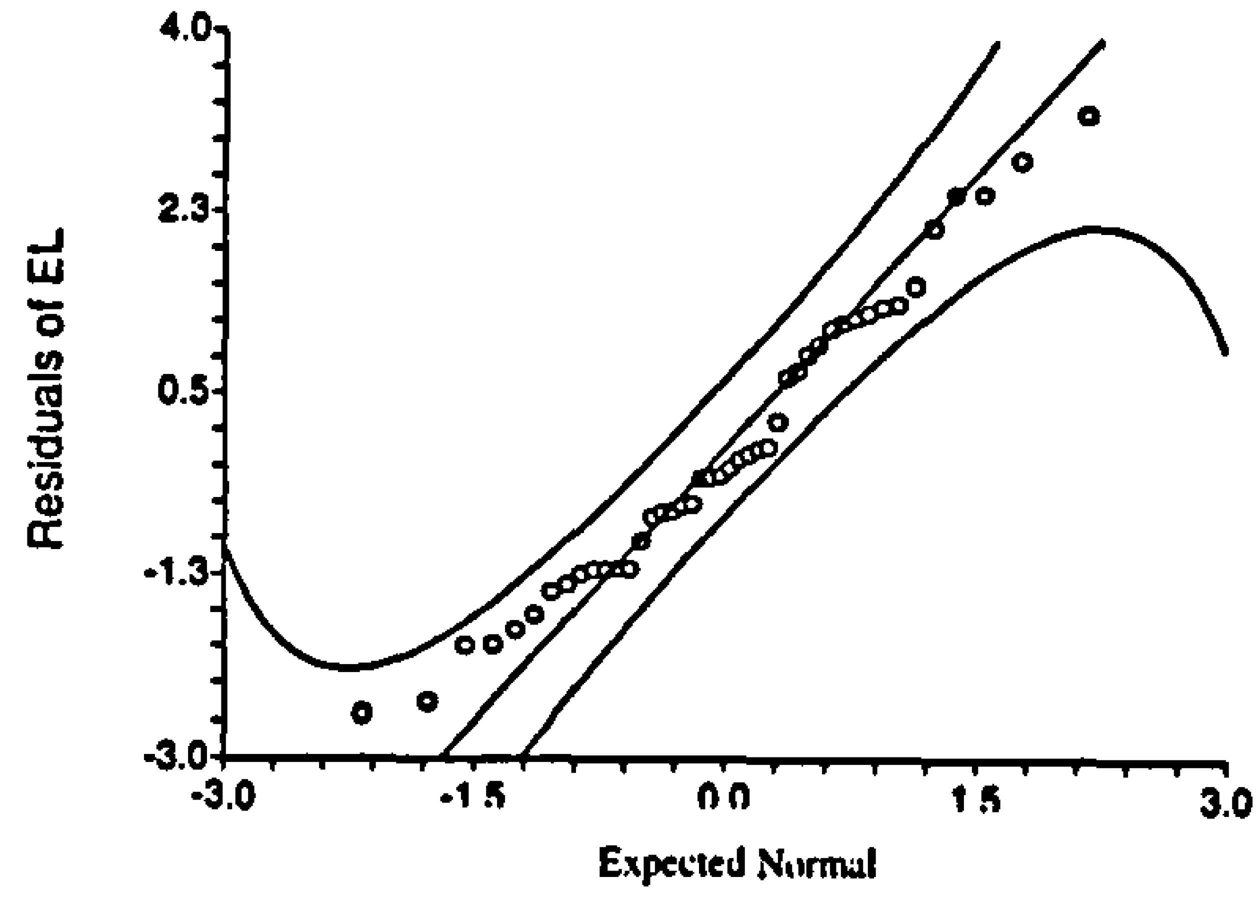
All Condition Numbers less than 100. Multicollinearity is NOT a problem.

Plots Section

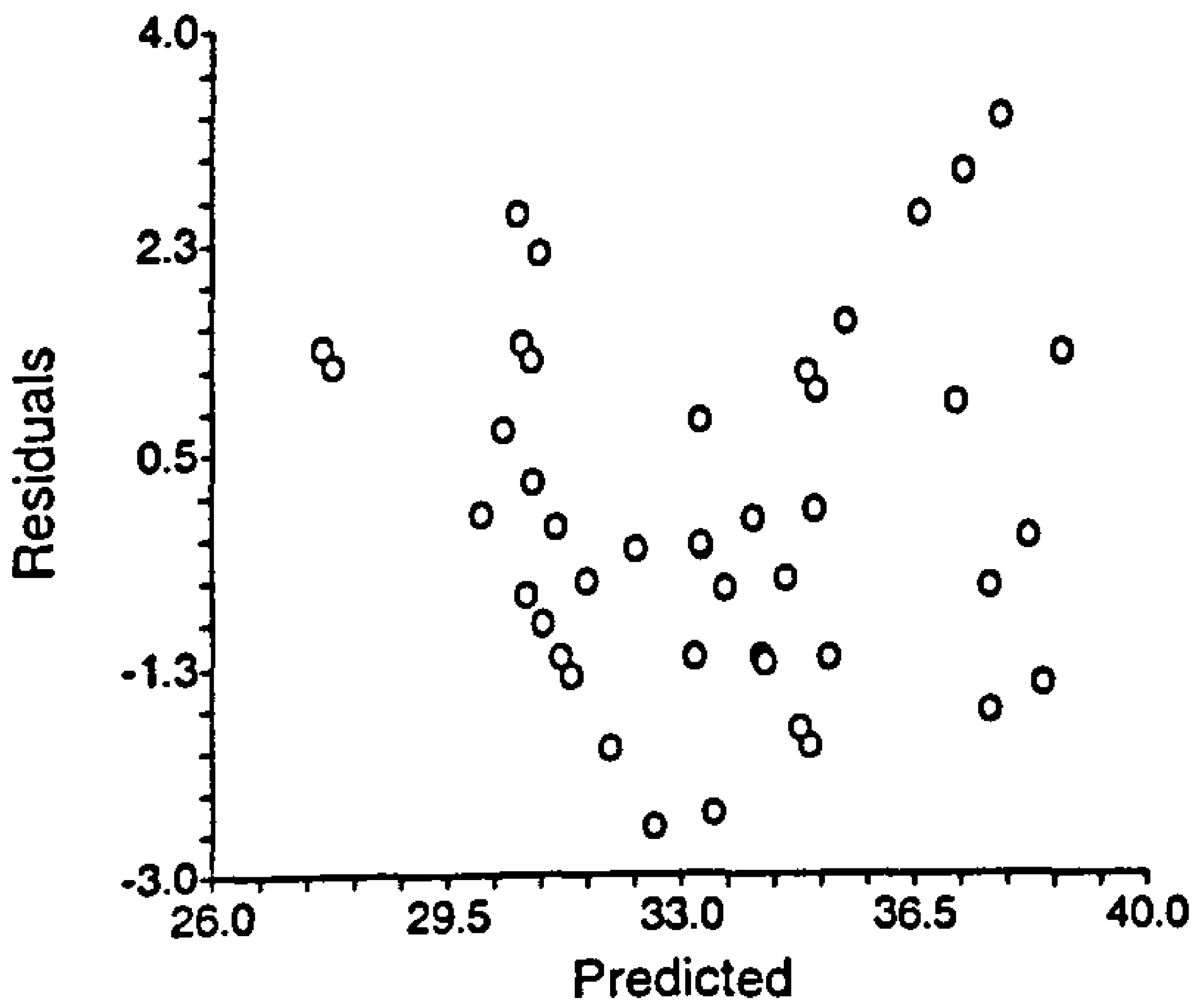
Histogram of Residuals of EL



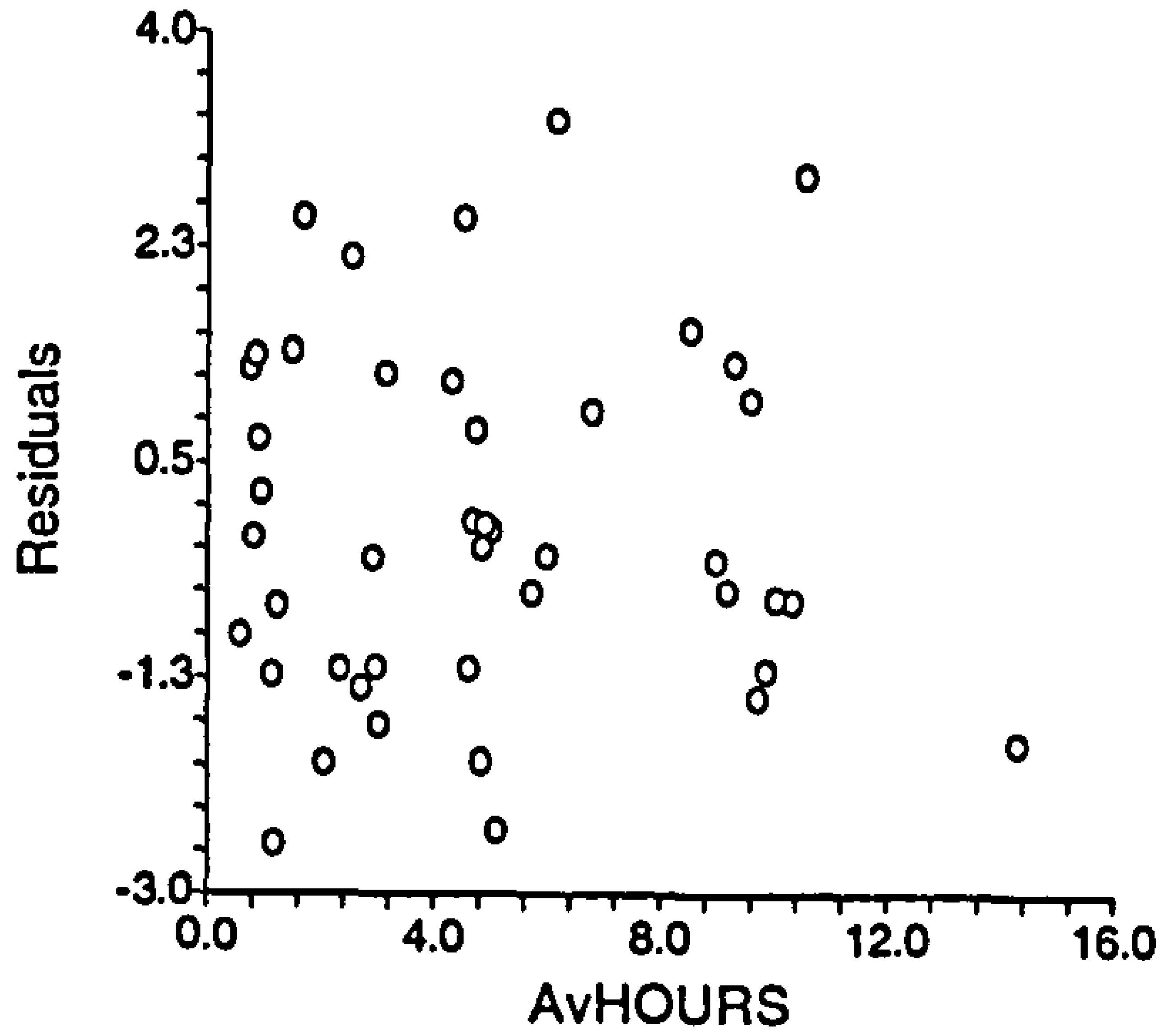
Normal Probability Plot of Residuals of EL



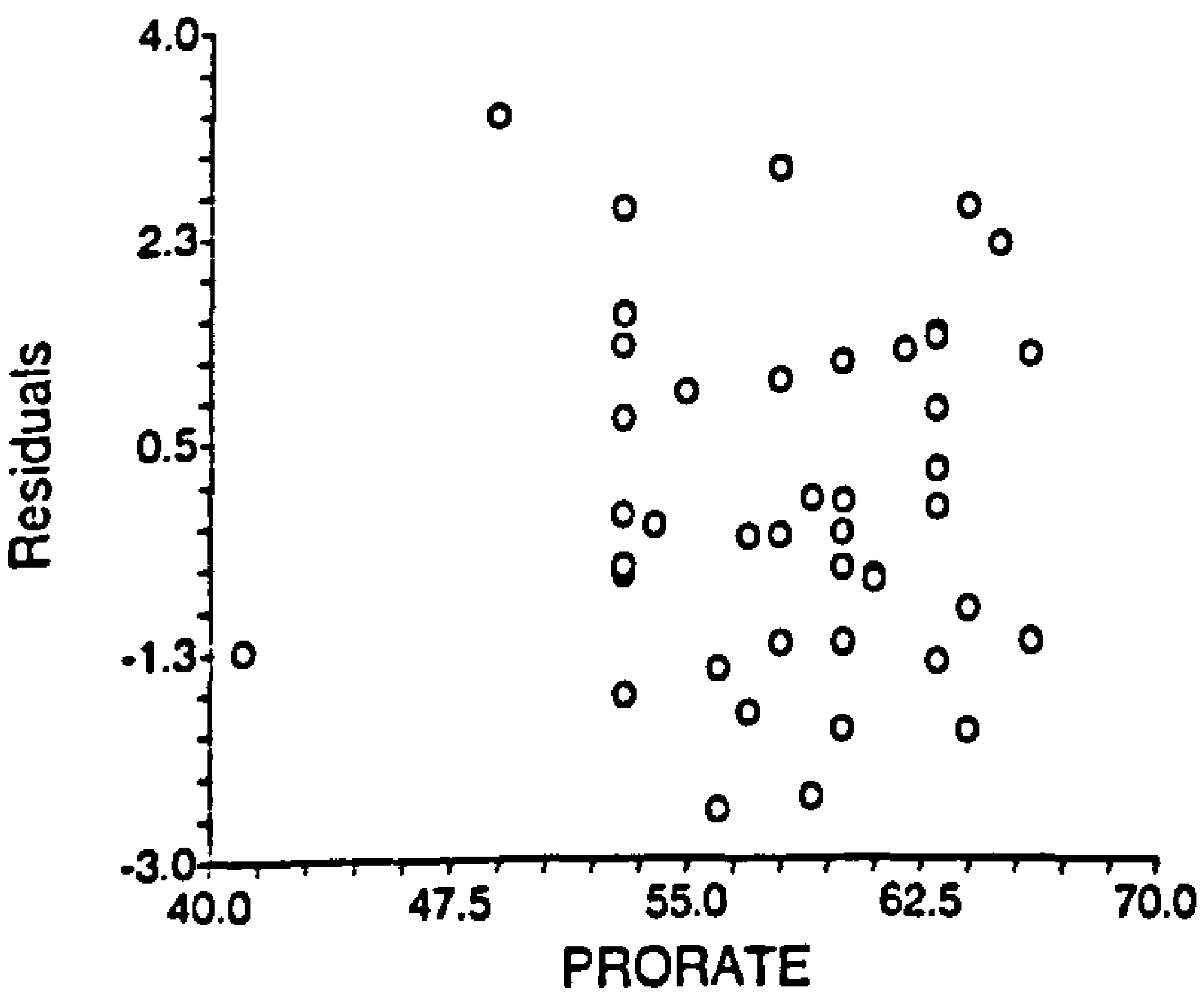
Residuals vs Predicted



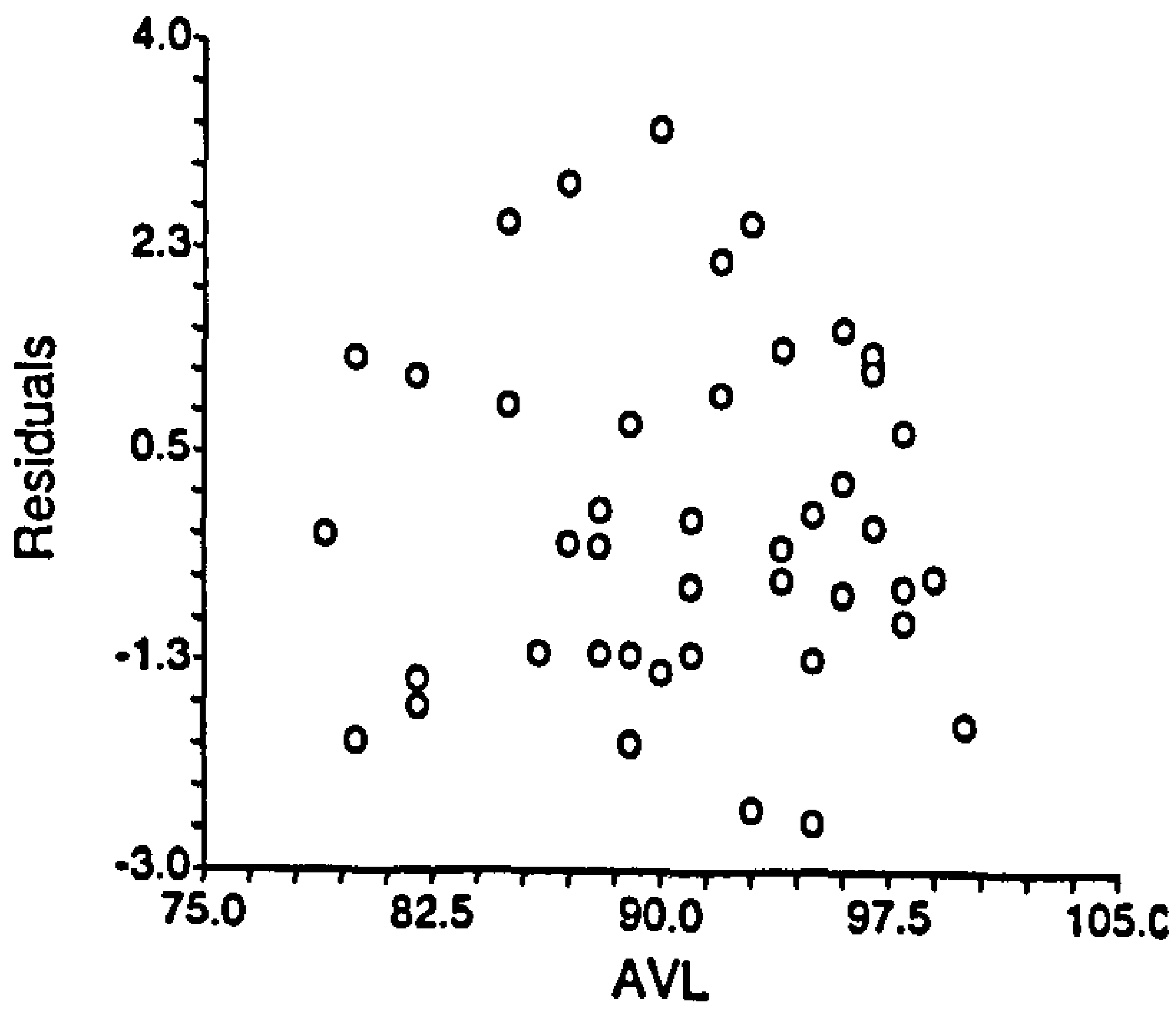
Residuals vs AvHOURS

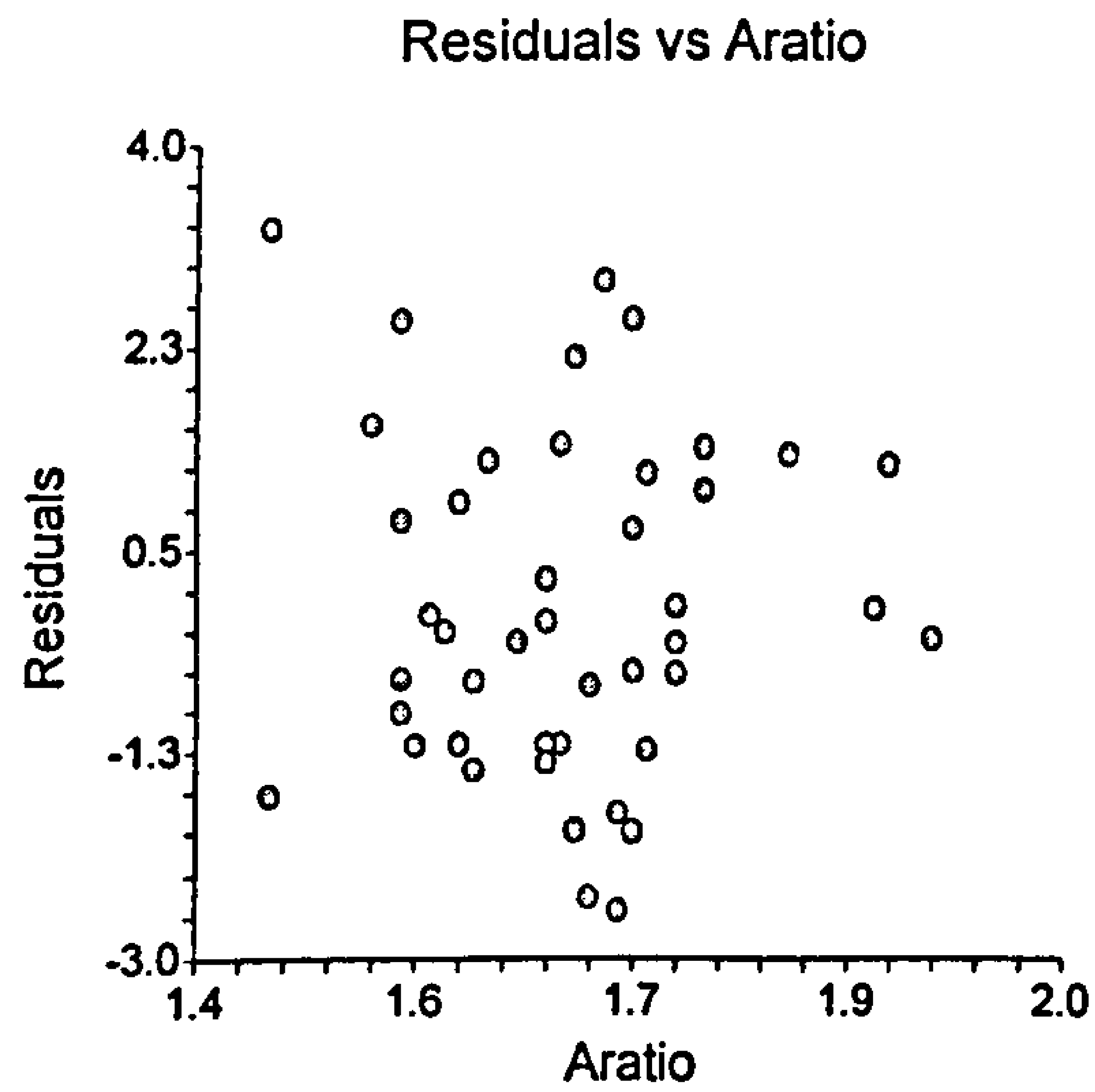


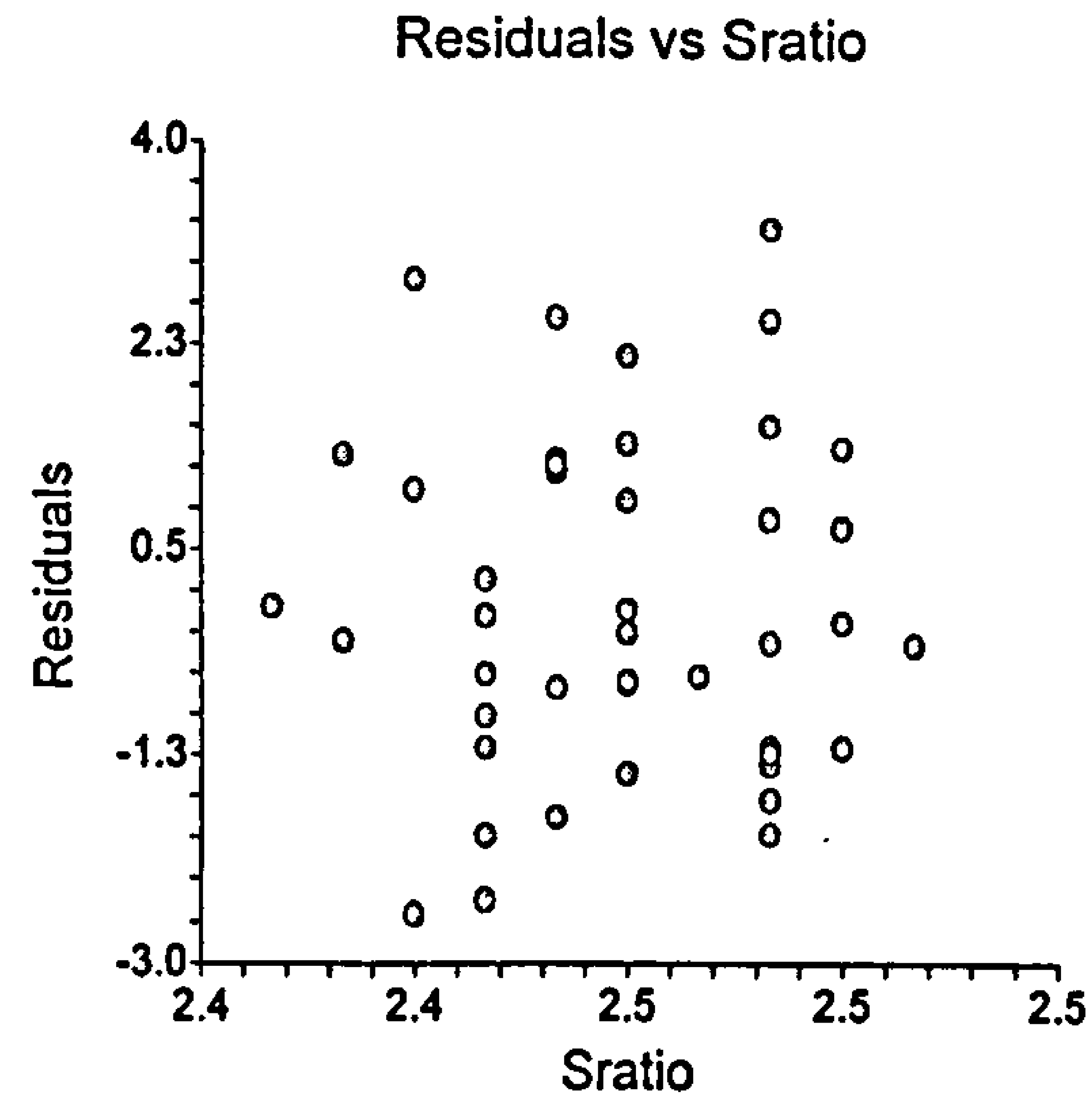
Residuals vs PRORATE



Residuals vs AVL







5. Table (7.43) provides the ***final robust regression*** report. It seems reasonable from this report that:

   i) the estimated regression model is EL= 90.27061+ .2561762*AvHOURS-.196848*PRORATE-.1774019*AVL-7.993319*Aratio-30.18439*Sratio+.5956308*LimeSF.

   ii) All the independent variables in the model contribute to the dependent variable.

   iii) The R-squared value is 0.909245, which means that this model explains about 91% of the variability in EL by the independent variables in the model.

Table (7.43) Final run of robust regression

Regression Equation Section

Independent Variable	**Regression Coefficient**	**Standard Error**	**T-Value (Ho: B=0)**	**Prob Level**	**Decision (5%)**	**Power (5%)**
Intercept	90.27061	30.05543	3.0035	0.004904	Reject Ho	0.831515
AvHOURS	0.2561762	3.579722E-02	7.1563	0.000000	Reject Ho	1.000000
PRORATE	-0.196848	2.501416E-02	-7.8695	0.000000	Reject Ho	1.000000
AVL	-0.1774019	0.0232406	-7.6333	0.000000	Reject Ho	1.000000
Aratio	-7.993319	1.380129	-5.7917	0.000001	Reject Ho	0.999877
Sratio	-30.18439	6.586811	-4.5825	0.000056	Reject Ho	0.993678
LimeSF	0.5956308	0.1808475	3.2936	0.002269	Reject Ho	0.89

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	29854.31	29854.31			
Model	6	147.6527	24.60879	58.4422	0.000000	1.000000
Error	35	14.73778	0.4210793			
Total(Adjusted)	41	162.3905	3.960744			

Root Mean Square Error	**0.6489062**	**R-Squared**	**0.909245**
Mean of Dependent Variable	**32.82941**	**Adj R-Squared**	**0.893687**
Coefficient of Variation	**0.019766**		

6. Conclusions:

Comparing the final runs of the OLS multiple regression model and the final run of the robust regression, we observe that

1) Some of the assumptions on OLS multiple regression are not satisfied since the data has some outliers, and the residuals may be correlated. However, these problems do not affect the robust regression.

2) The final multiple regression model has only five significant independent variables, namely, AvHOURS, PRORATE, AVL, Aratio, and Sraio. On the other hand the final robust regresion model has six significant independent variables, namely, AvHOURS, PRORATE, AVL, Aratio, Sraio, and LimeSF.

3) For multiple regression R^2=0.79, while for robust regression R^2=0.91. This means that the robust regression explains much more of the variability in the dependent variable than the multiple regression does.

4) Based on the above observations, we recommend using the final robust estimated regression mode, which is given by

$$EL= 90.27061+ .2561762*AvHOURS-.196848*PRORATE-.1774019*AVL$$
$$-7.993319*Aratio-30.18439*Sratio+ .5956308*LimeSF.$$

7.5.6.2 FUEL for Kiln 5

From the full output reports, which are given in Appendices (03)-(08), we obtain the following summary of the results.

1- Table (7.44) provides the results of ***checking the underlying assumptions***. It seems reasonable from this table that we have problems with the normality assumption and the existence of outliers. Other assumptions are satisfied. This suggests using the robust regression procedure.

Table (7.44) Check of the Regression Assumptions

	Tools	Result
1. Linearity	1. Scatter Plots	Only with AvNO, AvHOURS, PRORATE, AVL, Aratio
	2. Correlation Matrix	Only with AvNO, PRORATE, AVL, Aratio
2. Normality of Residuals	1. Skewness Tests	Rejected
	2. Kurtosis Test	Rejected
	3. Omnibus Test	Rejected
	4. Normal Probability Plot	Almost normal
	5. Histogram	Almost normal
3. Constant Error Variance	1. Scatter Plots	No patterns
4. Independent Errors	1. Serial Correlations	Correlated
	2. Durbin-Watson Test	Correlated
5. Multicollinearity	1. Correlation Matrix	Absolute maximum 0.585609
	2. Condition Numbers	All are less than 100
	3. Variance Inflation Factors	All are less than 10
	4. Eigenvalues	No values are close to zero
6. Outliers	1. Histogram	Outliers Exist
	2. Normal Probability Plot	Outliers Exist
	3. Scatter Plots	Outliers Exist

2. Table (7.45) reports a summary of the results of ***all possible regression models***. It seems reasonable from this table that based on the Cp criterion the best model is the one with only four independent variables, namely AvNO, PRORATE, AVL, and Aratio. This result partially agrees with those obtained by checking the assumptions of

the regression model. Moreover, the difference in R-squared between the full model with seven variables and the model with these two variables is about 0.04 which is not worth complicating the model.

Table (7.45) Results of All Possible Regression Procedure

Model Size	R-Squared	Root MSE	Cp	Variables in Model	Best Model
1	0.312032	7.347728	14.041407	PRORATE	
2	0.471748	6.516621	3.495401	AvNO, PRORATE	
3	0.500316	6.416697	3.251318	AvNO, PRORATE, AVL	This one
4	0.523736	6.344322	3.411603	AvNO, PRORATE, Aratio, Sratio	
5	0.540117	6.315758	4.124832	AvNO, PRORATE, AVL, Aratio, Sra	
6	0.541197	6.393019	6.040011	AvNO, AvHOURS, PRORATE , AVL, Aratio, Sratio	
7	0.541706	6.477604	8.000000	AvNO, AvHOURS, PRORATE, AVL, Aratio, Sratio, LimeSF	

3. Table (7.46) reports a summary of the obtained *initial regression models* using different procedures together with their R-squared values. It seems reasonable from this table that

i) The coefficients of the estimated models by sequentially entering one variable at a time are almost the same. This is seen from all possible regression models section. This indicates a type of consistancy of the results.

ii) Amonge the three robust regression models, there are slightly different results because they use different truncation methods in weighing the error terms. However, the largest R-squared is obtained by the Andrew's sine method.

iii) The coefficients of each of the influential variables (PRORATE, Aratio, AvNO, and AVL) have the same signs in all models. The AvNO variable has positive signs which means that the dependent variable increases as this independent variable increases. However, The PRORATE, Aratio, and AVL variables have negative signs, this means that the dependent variable decreases as each of these variables increases. The signs of the other independent variables are not important since their coefficients are not significantly different from zeros. This means that they do not have significant influence on the dependent variable.

iv) Since the data has some outliers and the normality assumption is rejected, we recommend using the robust regression model after deleting the independent variables which do not contribute to the dependent variable. This will be done in the final run section.

Table (7.46) Comparison of the estimated models using different procedures

Procedure	Intercep	AvNO	AvHOUR	PRORATE	AVL	Aratio	Sratio	LimeSF	R-Squ
All Possible									
Size 1	154.564			-.9684158					0.3120
Size 2	140.879	17.0365		-.857336					0.4717
Size 3	301.018	18.2789		-.9026544			-64.2466		0.5003
Size 4	375.928	17.1298		-.8623834		-15.31227	-85.1647		0.5237
Size 5	370.818	13.0837		-.8520316	-.2478964	-14.17906	-74.2258		0.5401
Size 6	418.538	13.1469		-.8522946	-.2522745	-15.14272	-78.6441		0.5411
Size 7 (Multiple)	403.450	13.1750	.07534107	-.838965	-.2451752	-14.95106	-76.0228	-.29800	0.5417
Stepwise									
1- Forward	140.879	17.0365		-.857336					0.4717
2- Step	301.018	18.2789		-.9026544			-64.2466		0.4974
3-Backward	140.879	17.0365		-.857336					0.4717
4-Min MSE	140.879	17.0365		-.857336					0.4717
Robust									
1-Least Abs Dev. 1.0	197.833	7.68672	.8233584	-.128106	-.4697443	-19.30831	-50.3535	1.0358	0.7501
2-Tukey's Biweight 6.0	232.114	4.94323	.817874	-0.0375113	-.4739273	-25.07257	-60.2469	.98874	0.7401
3-Andrew's Sin 2.1	-42.885	4.41736	1.266695	0.06395721	-.4595352	-14.93584	-19.1537	2.5513	0.8998

4) It seems reasonable from the ***final multiple regression*** report (see Table (7.47)) that:

i) R-squared value is 0.471748. This means that this model explains about 47% of the variability in the independent variable. This is somehow a low coefficient of determination.

ii) The PRESS R-squared is 0.3212, which means that this model can predict to good accuracy about 32% of the values of Y. This is also a low indicator of predictability.

iii) Normality of the residuals is accepted.

iv) Based on serial correlations and the Durbin-Watson test, the residuals may be correlated.

v) Multicollinearity could not be a problem.

vi) From the plots section, one observes that the data has some outliers.

Table (7.47) Final Run of Multiple Regression

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	140.8791	12.17315	11.5729	0.000000	Reject Ho	1.00000C
AvNO	17.03651	4.838781	3.5208	0.001070	Reject Ho	0.930197
PRORATE	-0.857336	0.1992976	-4.3018	0.000102	Reject Ho	0.987424
R-Squared	0.471748					

Model

FU=140.8791+17.03651 AvNO-0.857336 PRORATE

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	140.8791	12.17315	116.2949	165.4632	0.0000
AvNO	17.03651	4.838781	7.264395	26.80863	0.4047
PRORATE	-0.857336	0.1992976	-1.259826	-0.4548463	-0.4945
T-Critical	2.019541				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	422576	422576			
Model	2	1554.88	777.4399	18.3072	0.000002	0.999774
Error	41	1741.12	42.46634			
Total(Adjusted)	43	3296	76.65116			

Root Mean Square Error	**6.516621**	**R-Squared**	**0.4717**
Mean of Dependent	**98**	**Adj R-Squared**	**0.4460**
Coefficient of Variation	**6.649613E-02**	**Press Value**	**2237.431**
Sum |Press Residuals|	**254.0656**	**Press R-Squared**	**0.3212**

Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	0.2513	0.801590	Accepted
Kurtosis	-0.1900	0.849316	Accepted
Omnibus	0.0992	0.951589	Accepted

Serial-Correlation Section

Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.428654	9	-0.044404	17	-0.042068
2	0.132532	10	0.013303	18	0.000093
3	0.123800	11	0.241067	19	-0.046453
4	0.137751	12	0.174620	20	-0.260852
5	-0.019480	13	0.145370	21	-0.279847
6	-0.015291	14	0.145383	22	-0.208235
7	-0.047488	15	0.230729	23	-0.102927
8	-0.077639	16	-0.016076	24	-0.101988

Above serial correlations significant if their absolute values are greater than 0.301511

Durbin-Watson Value 1.1135

Multicollinearity Section

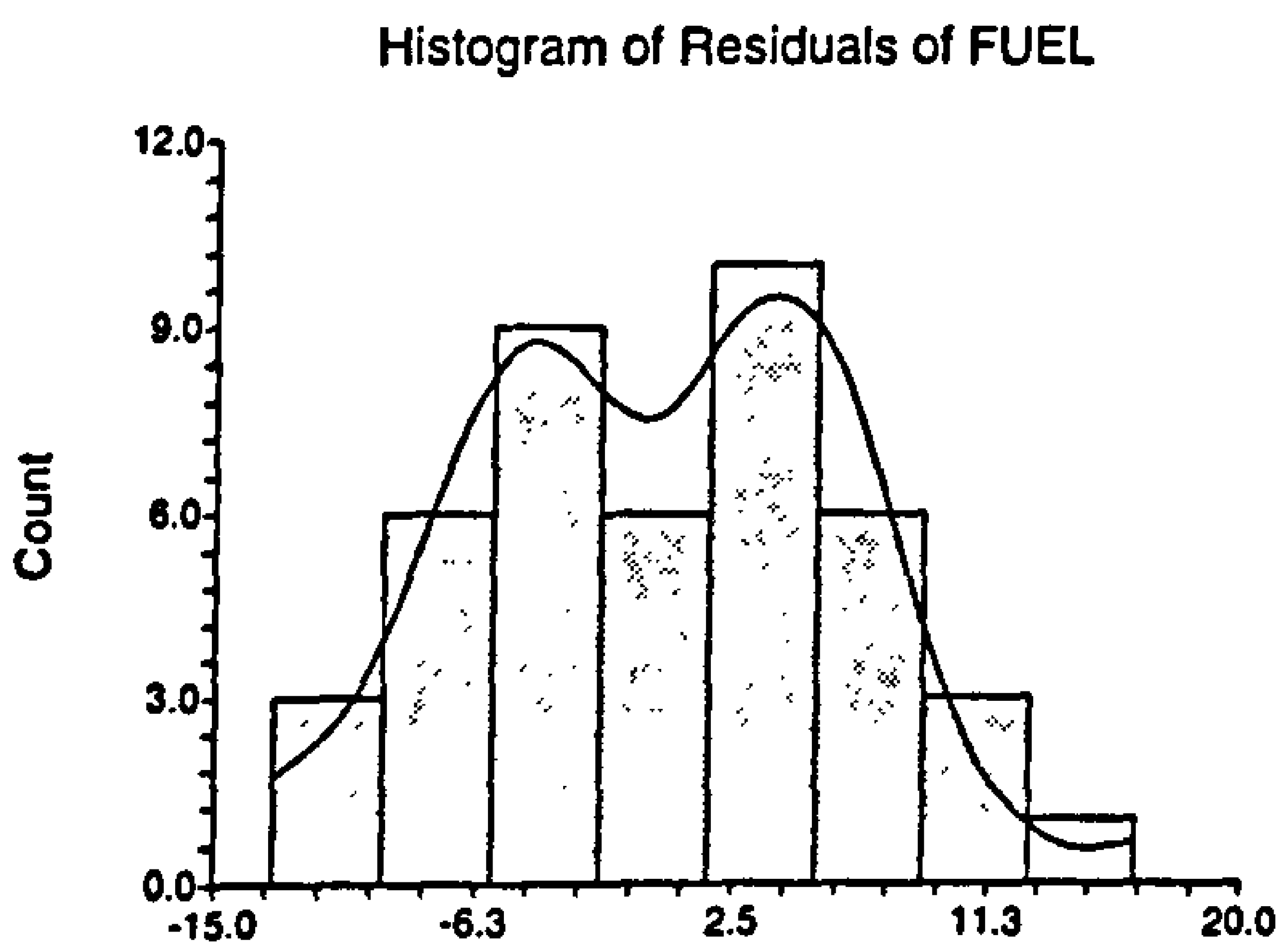
Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance	Diagonal of X'X Inverse
AvNO	1.025704	0.025060	0.974940	0.5513497
PRORATE	1.025704	0.025060	0.974940	9.353181E-04

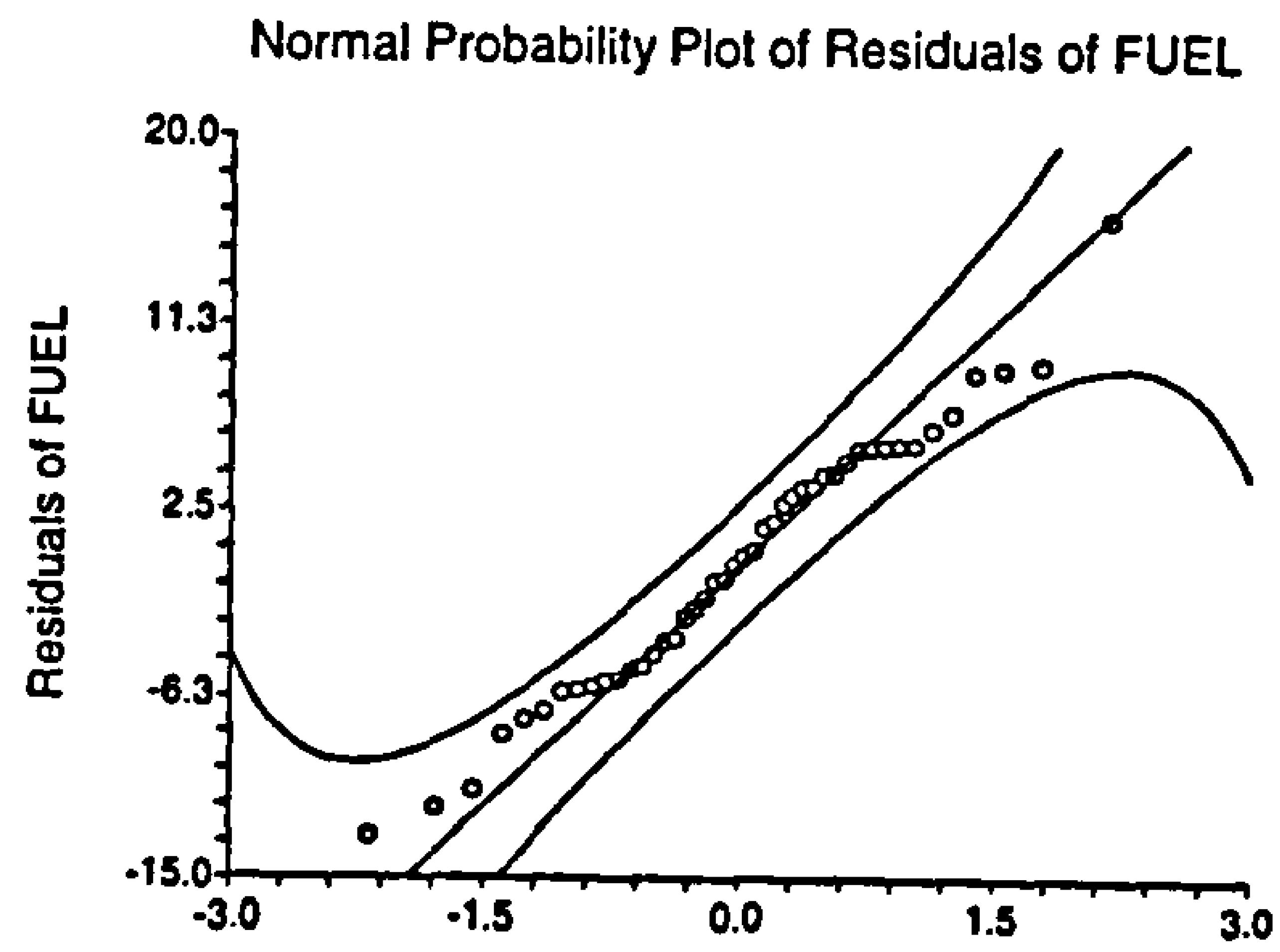
Eigenvalues of Centered Correlations

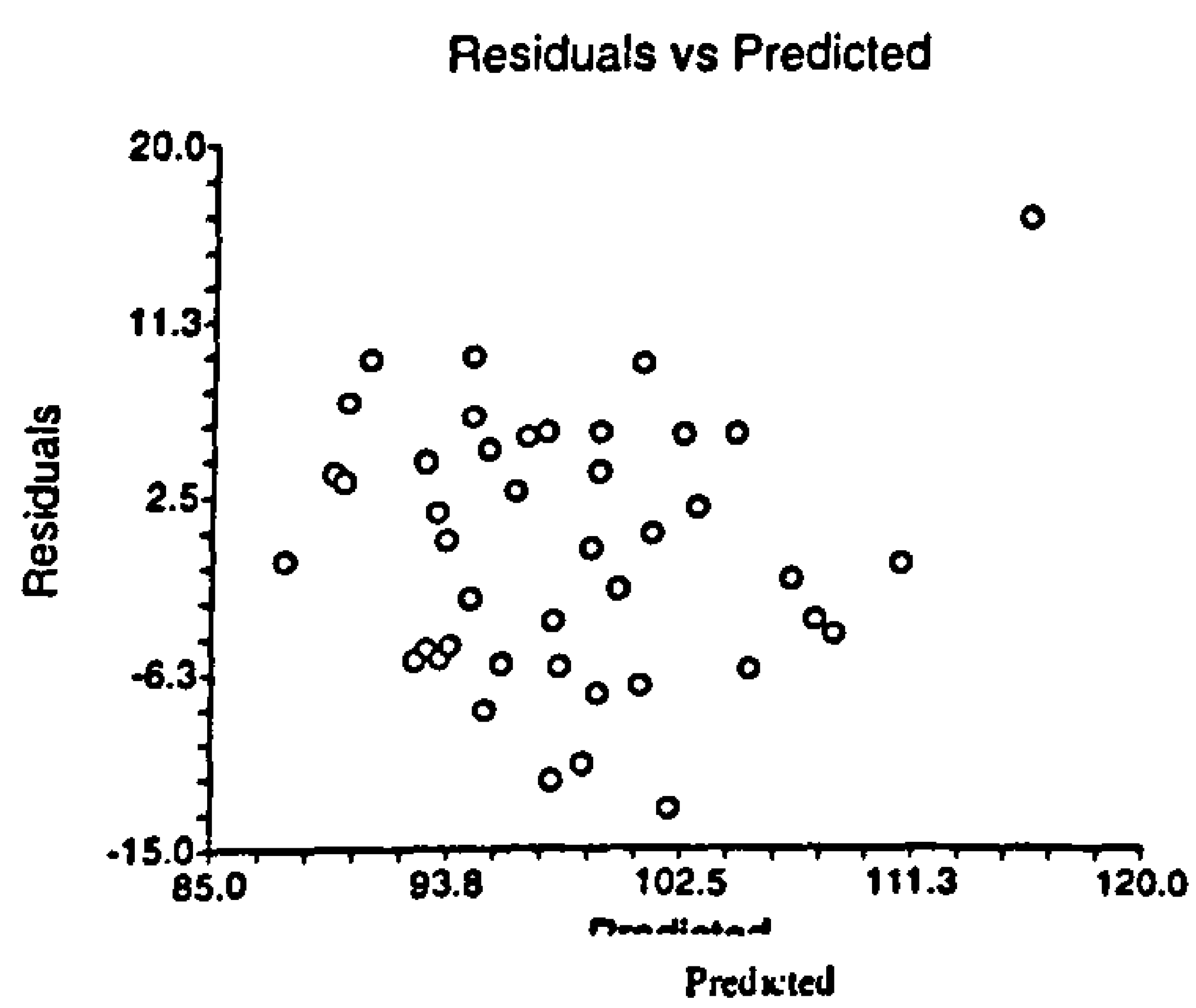
No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.158303	57.92	57.92	1.00
2	0.841697	42.08	100.00	1.38

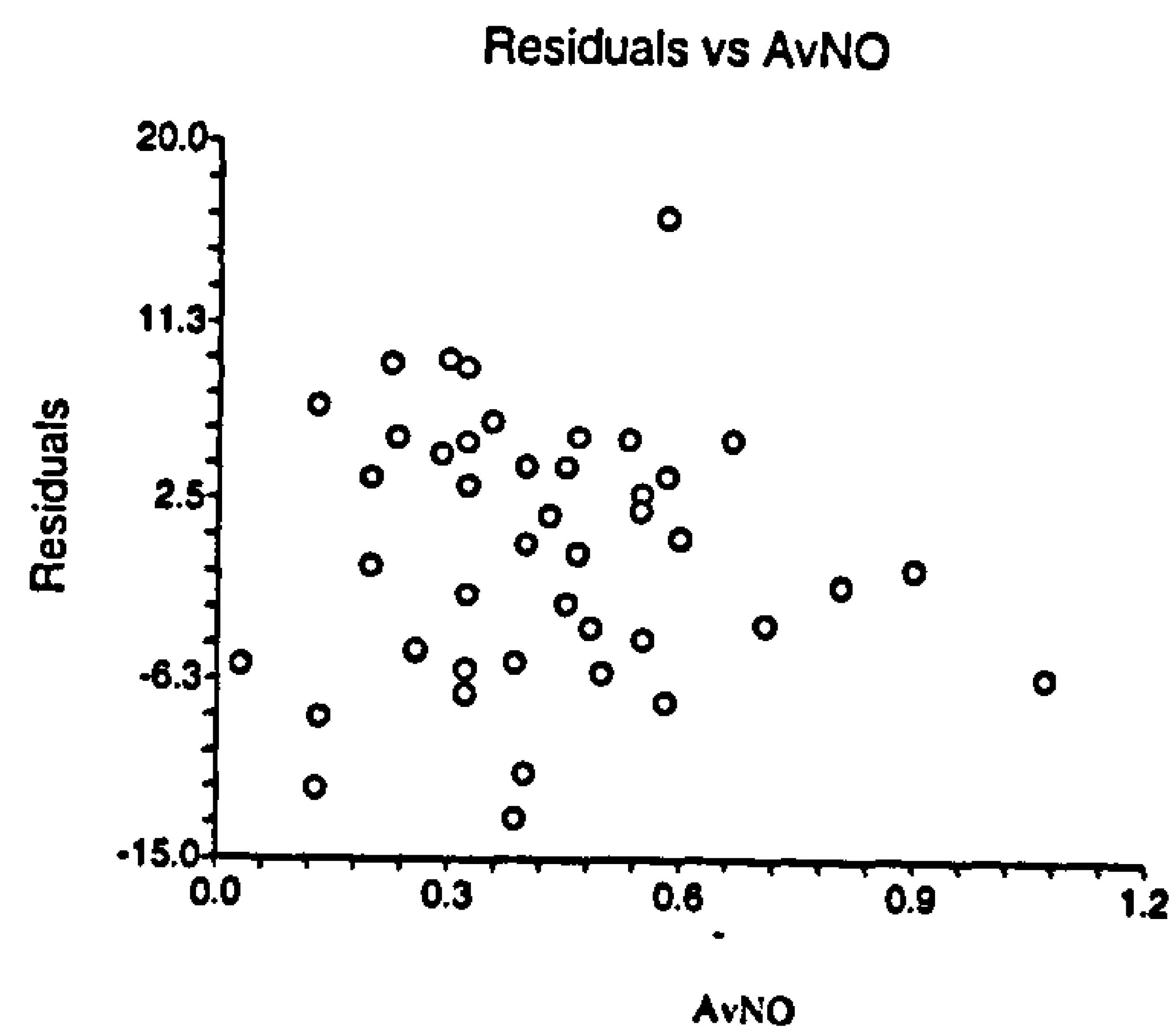
All Condition Numbers less than 100. Multicollinearity is NOT a problem.

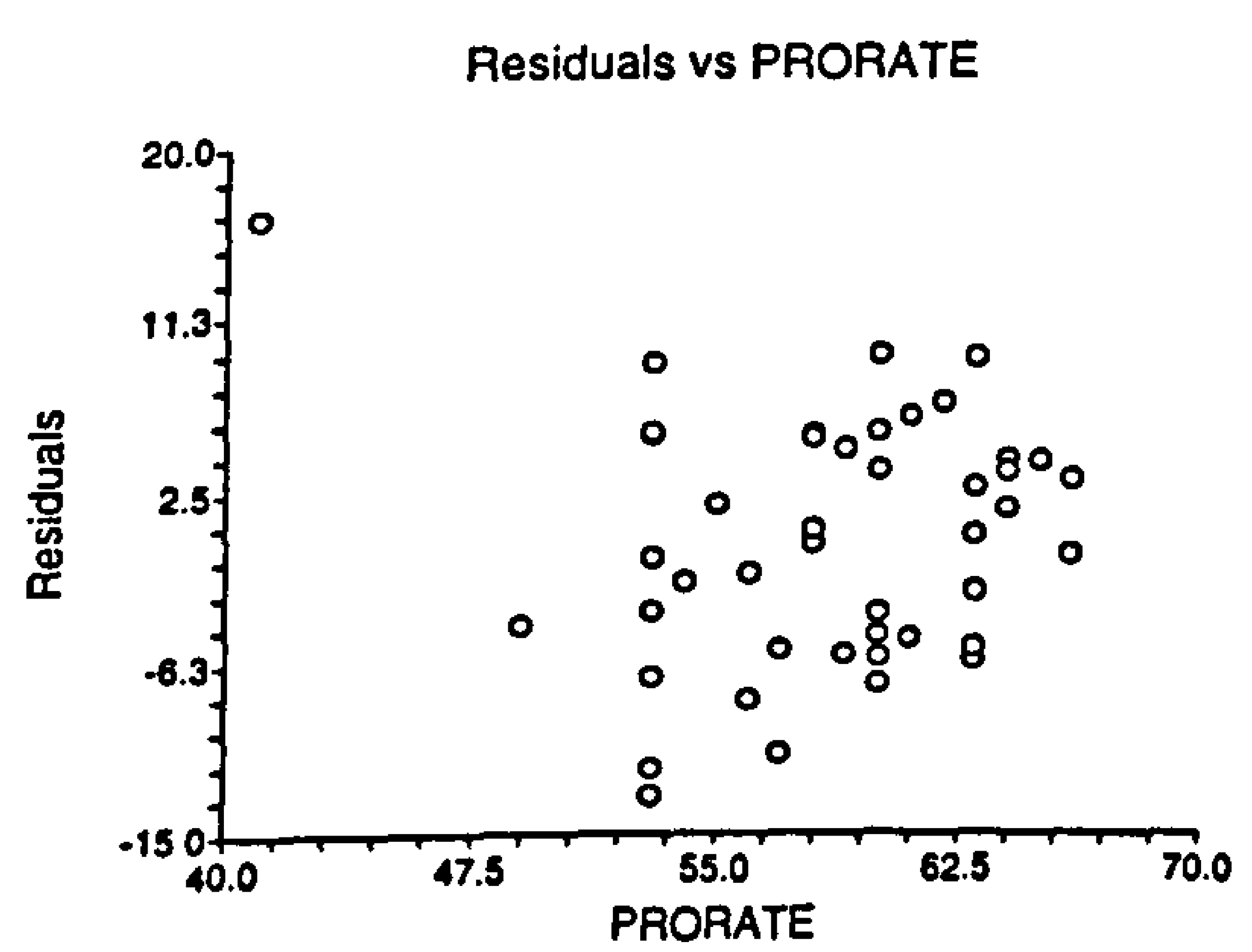
Plots Section











5.Table (7.48) provides the ***final robust regression*** report. It seems reasonable from this report that:

i) the estimated regression model is FUEL = -99.77511+ 1.300123*AvHOURS-.5187489*AVL-12.60869*Aratio+ 2.727895*LimeSF.

ii) All the independent variables in the model contribute to the dependent variable.

iii) The R-squared value is 0.891453, which means that this model explains about 89% of the variability in FUEL by the independent variables in the model.

Table (7.48) Final run of robust regression

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	-99.77511	55.96642	-1.7828	0.083059	Accept Ho	0.411191
AvHOURS	1.300123	0.1187442	10.9489	0.000000	Reject Ho	1.000000
AVL	-0.5187489	6.959292E-02	-7.4540	0.000000	Reject Ho	1.000000
Aratio	-12.60869	3.878477	-3.2509	0.002499	Reject Ho	0.885533
LimeSF	2.727895	0.5544551	4.9200	0.000019	Reject Ho	0.997637

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	278455.1	278455.1			
Model	4	1211.04	302.76	73.9137	0.000000	1.000000
Error	36	147.4607	4.096129			
Total(Adjusted)	40	1358.501	33.96252			

Root Mean Square Error	**2.02389**	**R-Squared**	**0.891453**
Mean of Dependent Variable	**97.67168**	**Adj R-Squared**	**0.879393**
Coefficient of Variation	**2.072136E-02**		

6. Conclusions:

Comparing the final runs of the OLS multiple regression model and the final run of the robust regression, we observe that

1) Some of the assumptions on OLS multiple regression are not satisfied since the data has some outliers, and the residuals may be correlated. However, these problems do not affect the robust regression.

2) The final multiple regression model has only two significant independent variables, namely, AvNO, and PRORATE. On the other hand the final robust regresion model has four significant independent variables, namely, AvHOURS, AVL, Aratio, and LimeSF.

3) For multiple regression R^2=0.47, while for robust regression R^2=0.89. This means that the robust regression explains much more of the variability in the dependent variable than the multiple regression does.

4) Based on the above observations, we recommend using the final robust estimated regression mode, which is given by

FUEL = -99.77511+ 1.300123*AvHOURS-.5187489*AVL-12.60869*Aratio

+ 2.727895*LimeSF

7.5.7 Presentation of Regression Analysis of Kiln 6

7.5.7.1 EL for Kiln 6

From the full output reports, which are given in Appendices (03)-(08), we obtain the following summary of the results.

1- Table (7.49) provides the results of ***checking the underlying assumptions***. It seems reasonable from this table that we have problems with the normality assumption and the existence of outliers. Other assumptions are satisfied. This suggests using the robust regression procedure.

Table (7.49) Check of the Regression Assumptions

	Tools	Result
1. Linearity	1. Scatter Plots	Only with AvNO,AvHOURS, PRORATE, AVL,Aratio
	2. Correlation Matrix	Only with AvHOURS, PRORATE, Aratio
2. Normality of Residuals	1. Skewness Tests	Accepted
	2. Kurtosis Test	Rejected
	3. Omnibus Test	Rejected
	4. Normal Probability Plot	Almost normal
	5. Histogram	Almost normal
3. Constant Error Variance	1. Scatter Plots	No patterns
4. Independent Errors	1. Serial Correlations	Uncorrelated
	2. Durbin-Watson Test	Uncorrelated
5. Multicollinearity	1. Correlation Matrix	Absolute maximum 0.510319
	2. Condition Numbers	All are less than 100
	3. Variance Inflation Factors	All are less than 10
	4. Eigenvalues	No values are close to zero
6. Outliers	1. Histogram	Outliers Exist
	2. Normal Probability Plot	Outliers Exist
	3. Scatter Plots	Outliers Exist

2. Table (7.50) reports a summary of the results of ***all possible regression models***. It seems reasonable from this table that based on the Cp criterion the best model is the one with only four independent variables, namely AvNO, AvHOURS, PRORATE, AVL, and

LimeSF. Moreover, the difference in R-squared between the full model with seven variables and the model with these two variables is about 0.01 which is not worth complicating the model.

Table (7.50) Results of All Possible Regression Procedure

Model Size	R-Squared	Root MSE	Cp	Variables in Model	Best Model
1	0.442506	3.21917	21.261272	AvHOURS	
2	0.538271	2.94781	12.145395	AvHOURS, PRORATE	
3	0.569587	2.86362	10.167774	AvNO, AvHOURS, PRORATE,	
4	0.617841	2.72586	6.787007	AvNO, AvHOURS, PRORATE, AVL	
5	0.65564	2.62714	4.914823	AvNO, AvHOURS, PRORATE, AVL, LimeSF	This One
6	0.65990	2.64589	6.458217	AvNO, AvHOURS, PRORATE, AVL, Sratio, Lim	
7	0.66418	2.66548	8.000000	AvNO, AvHOURS, PRORATE, AVL, Aratio, Sratio, LimeSF	

3. Table (7.51) reports a summary of the obtained ***initial regression models*** using different procedures together with their R-squared values. It seems reasonable from this table that

i) The coefficients of the estimated models by sequentially entering one variable at a time are almost the same. This is seen from all possible regression models section. This indicates a type of consistancy of the results.

ii) The robust regression models gave slightly different results because they use different truncation methods in weighing the error terms. However, the largest R-squared is obtained by the Andrew's sine method.

iii) The coefficients of each of the influential variables (PRORATE, Aratio, Sratio, AvNO, AvHOURS, and AVL) have the same signs in all models. These signs agree wuth the researcher expectations.

iv) Since the data has some outliers and the normality assumption is rejected by two out of three tests, we recommend using the robust regression model after deleting the independent variables, which do not contribute to the dependent variable. This will be done in the final run section.

Table (7.51) Comparison of the estimated models using different procedures

Procedure	Intercept	AvNO	AvHOUR	PRORAT	AVL	Aratio	Sratio	LimeSF	R-Squar
All Possible									
Size 1	25.6384		.5810547						0.442506
Size 2	49.4980		.4675554	-.203022					0.538271
Size 3	53.9358	-5.65279	.467125	-.225917					0.569587
Size 4	73.7504	-10.0027	.417057	-.248120	-.171401				0.617841
Size 5	2.18503	-10.8793	.3365999	-.261517	-.235706			.849189	0.655646
Size 6	-23.5788	-10.7515	.3275611	-.242970	-.233886		4.68660	.990393	0.659905
Size 7 (Multiple)	-12.3201	-10.4219	.3354227	-.252337	-.233462	-3.598366	5.11546	.896529	0.664180
Robust:									
1-Least Abs. Dev. 1.0	-40.4476	-7.34004	.3617401	-.203679	-.184512	.6159818	7.34926	.970364	0.876003
2-Tukey's Biweight 6.0	-31.1893	-6.51252	.3234505	-.224457	-.184017	1.825766	4.97393	.933554	0.875090
3-Andrew's Sine 2.1	-59.4090	-6.03497	.3517462	-.191332	-.174246	2.031039	8.29473	1.09326	0.944461
Stepwise									
1- Forward	49.49805		.4675554	-.203022					0.538271
2- Step	49.49805		.4675554	-.203022					0.538271
3-Backward	2.185037	-10.8793	.3365999	-.261517	-.235706			.849189	0.655646
4-Min MSE	2.185037	-10.8793	.3365999	-.261517	-.235706			.849189	0.655646

4) It seems reasonable from the ***final multiple regression model*** (see Table (7.52)) that

i) R-squared value is 0.617841. This means that this model explains about 62% of the variability in the independent variable. This is a reasonable coefficient of determination.

ii) The PRESS R-squared is 0.4658, which means that this model can predict to good accuracy about 47% of the values of Y. this seams to be a low indicator of predictability.

iii) There may be a problem with the normality of the residuals since the skewness test accepts normality while kurtosis test rejects normality.

iv) Based on serial correlations and the Durbin-Watson test, the residuals may be correlated.

v) Multicollinearity could not be a problem.

vi) From the plot section, one observes that the data has some outliers.

Table (7.52) Final Run of Multiple Regression

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	73.75043	11.87232	6.2120	0.000000	Reject Ho	0.999979
AvNO	-10.0027	3.673549	-2.7229	0.009542	Reject Ho	0.757118
AvHOURS	0.417057	9.524225E-02	4.3789	0.000084	Reject Ho	0.989601
PRORATE	-0.2481208	6.610345E-02	-3.7535	0.000554	Reject Ho	0.955623
AVL	-0.1714019	7.626729E-02	-2.2474	0.030199	Reject Ho	0.592233
R-Squared	0.617841					

Model

EL=73.75043-10.0027 AvNO+0.417057 AvHOURS-0.2481208 PRORATE-0.1714019 AVL

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	73.75043	11.87232	49.75557	97.74529	0.0000
AvNO	-10.0027	3.673549	-17.42722	-2.578184	-0.3202
AvHOURS	0.417057	9.524225E-02	0.2245653	0.6095487	0.4775
PRORATE	-0.2481208	6.610345E-02	-0.3817208	-0.1145207	-0.4102
AVL	-0.1714019	7.626729E-02	-0.3255438	-1.725992E-02	-0.2641
T-Critical	2.021075				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	37903.02	37903.02			
Model	4	486.2272	121.5568	16.1671	0.000000	0.999999
Error	40	300.7506	7.518764			
Total(Adjusted)	44	786.9778	17.88586			

Root Mean Square Error	2.742037	R-Squared	0.6178
Mean of Dependent	29.02222	Adj R-Squared	0.5796
Coefficient of Variation	9.448058E-02	Press Value	420.3882
Sum |Press Residuals|	86.7097	Press R-Squared	0.4658

Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	1.3732	0.169702	Accepted
Kurtosis	3.2499	0.001155	Rejected
Omnibus	12.4472	0.001982	Rejected

Serial-Correlation Section

Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.319093	9	-0.035898	17	-0.086472
2	-0.034896	10	0.041271	18	-0.214139
3	0.068922	11	-0.095213	19	-0.003152
4	-0.061756	12	0.034110	20	0.125739
5	-0.231138	13	0.114604	21	0.042776
6	-0.083574	14	0.106381	22	-0.115366
7	0.111430	15	-0.078313	23	-0.120678
8	-0.136814	16	-0.051624	24	-0.057294

Above serial correlations significant if their absolute values are greater than 0.298142

Durbin-Watson Value 1.3426

Multicollinearity Section

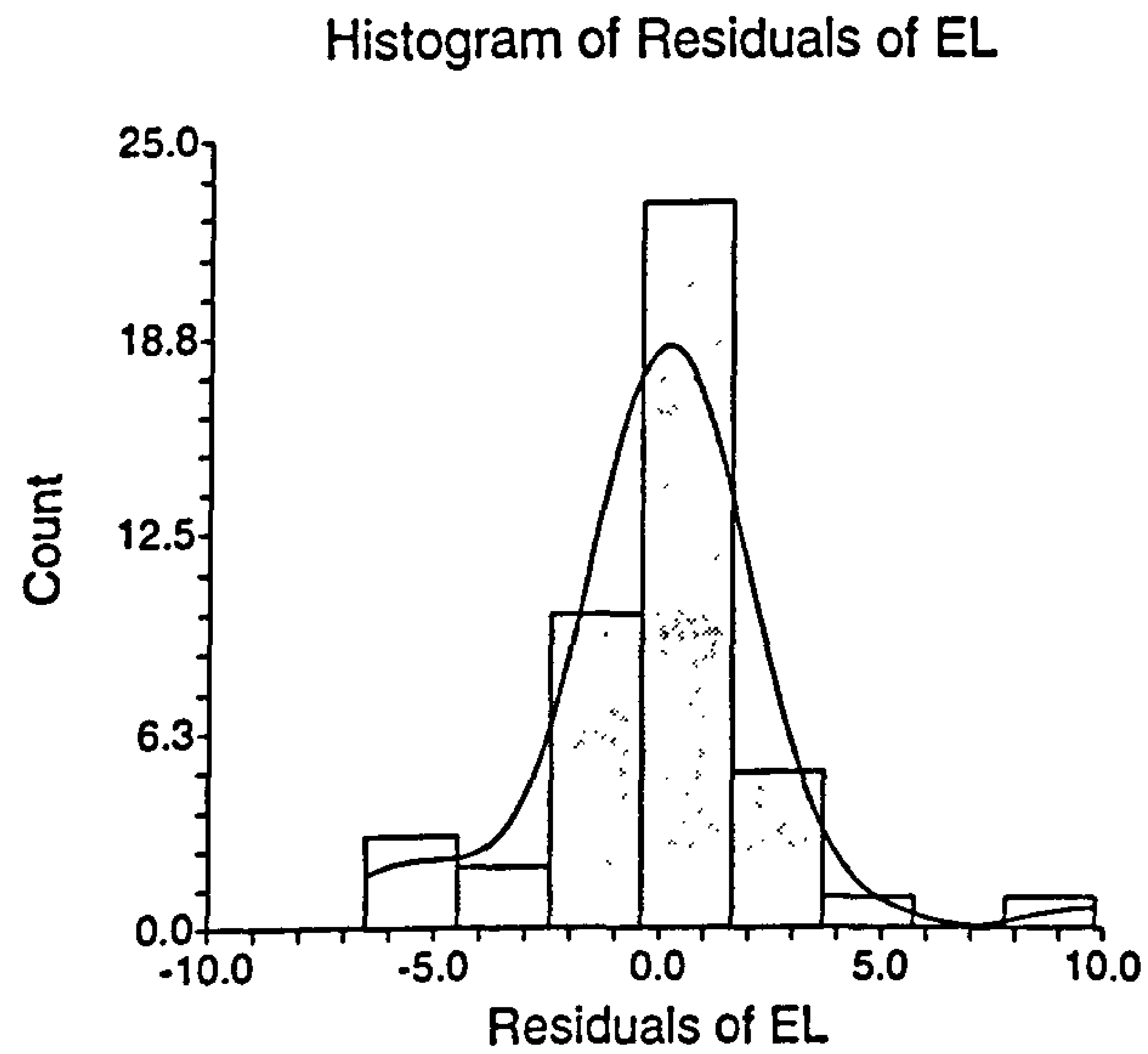
Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance	Diagonal of X'X Inverse
AvNO	1.446997	0.308913	0.691087	1.794837
AvHOURS	1.244402	0.196401	0.803599	1.20646E-03
PRORATE	1.249976	0.199985	0.800015	5.81168E-04
AVL	1.445749	0.308317	0.691683	7.736244E-04

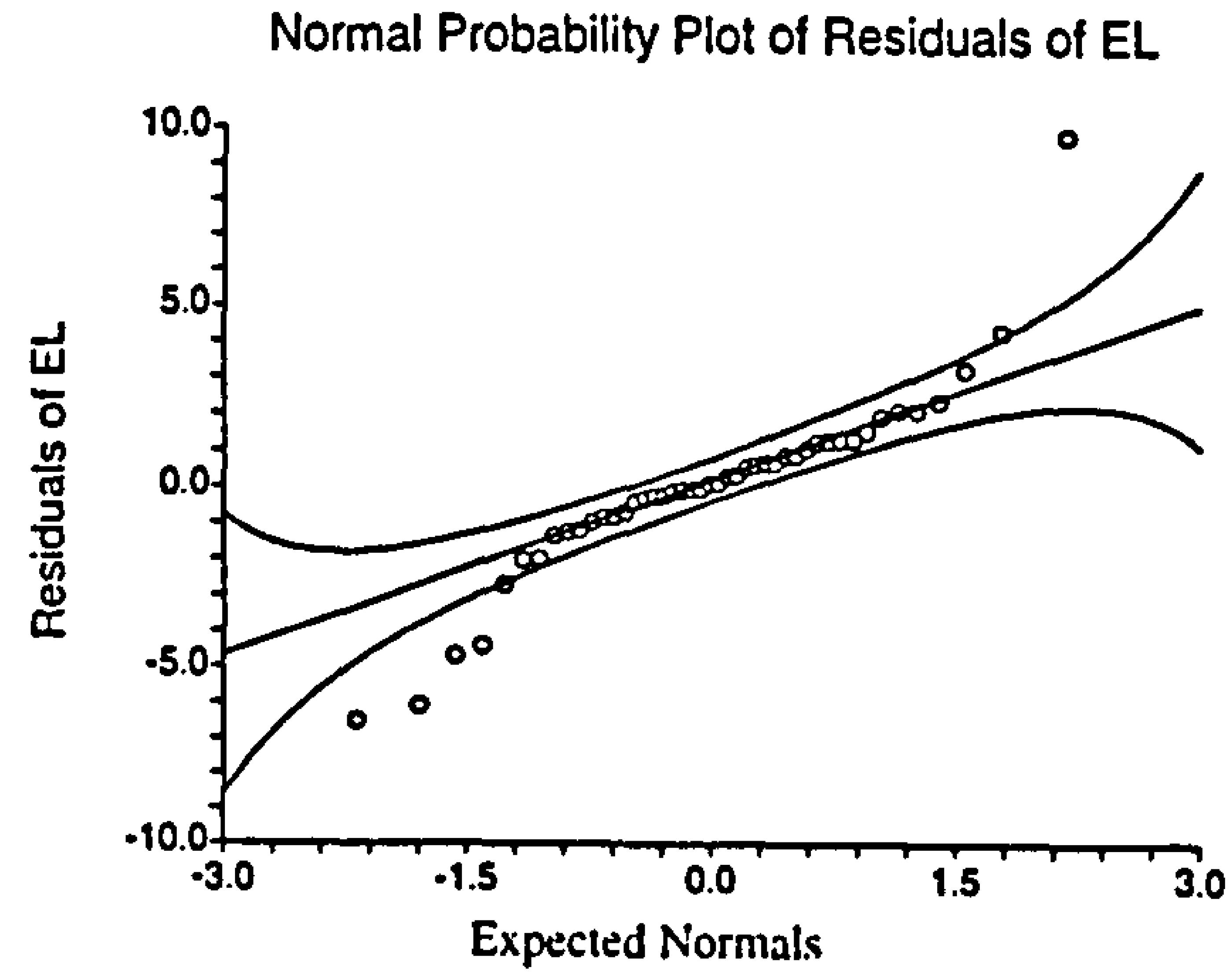
Eigenvalues of Centered Correlations

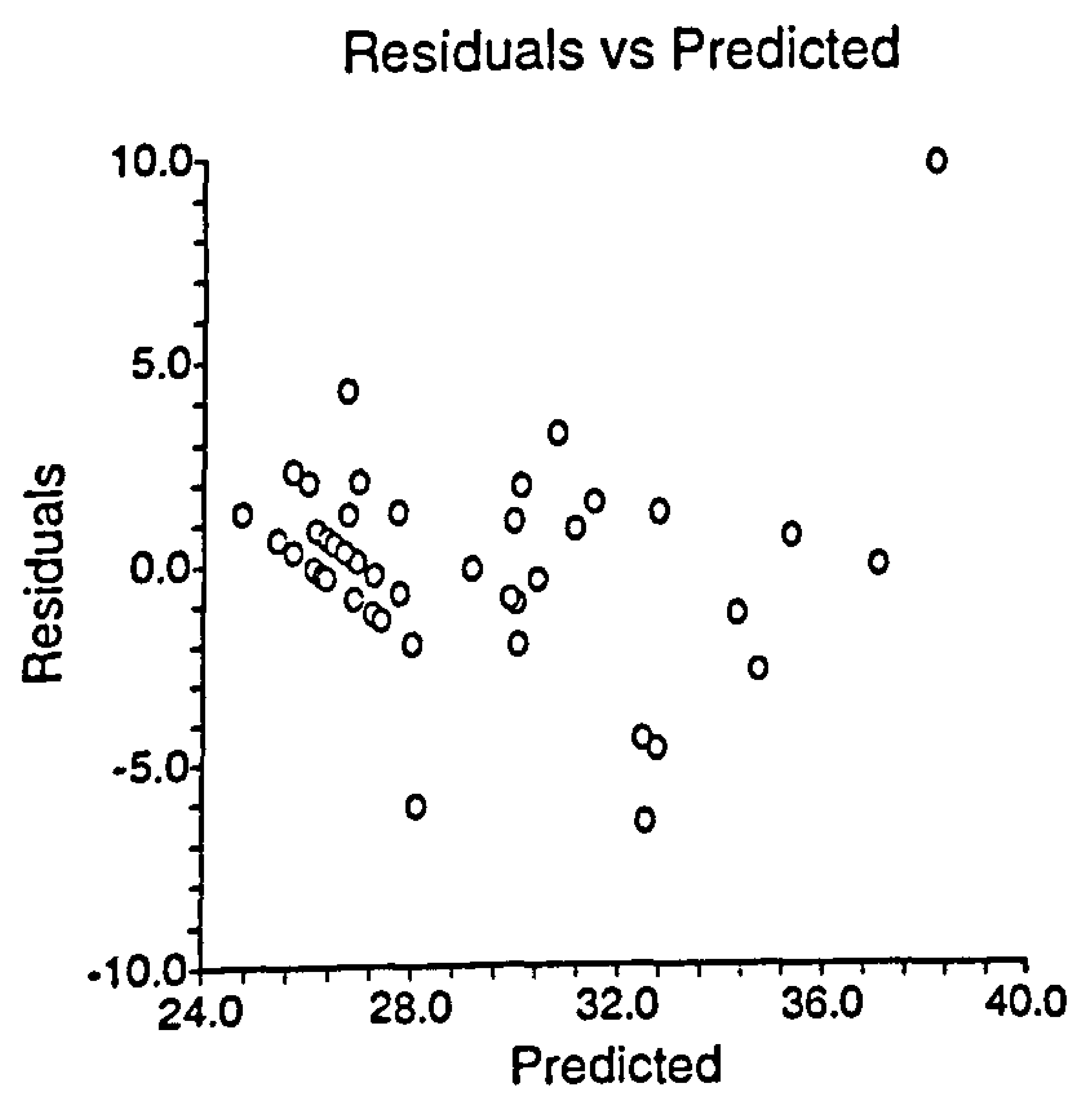
No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.731496	43.29	43.29	1.00
2	1.171218	29.28	72.57	1.48
3	0.702333	17.56	90.13	2.47
4	0.394952	9.87	100.00	4.38

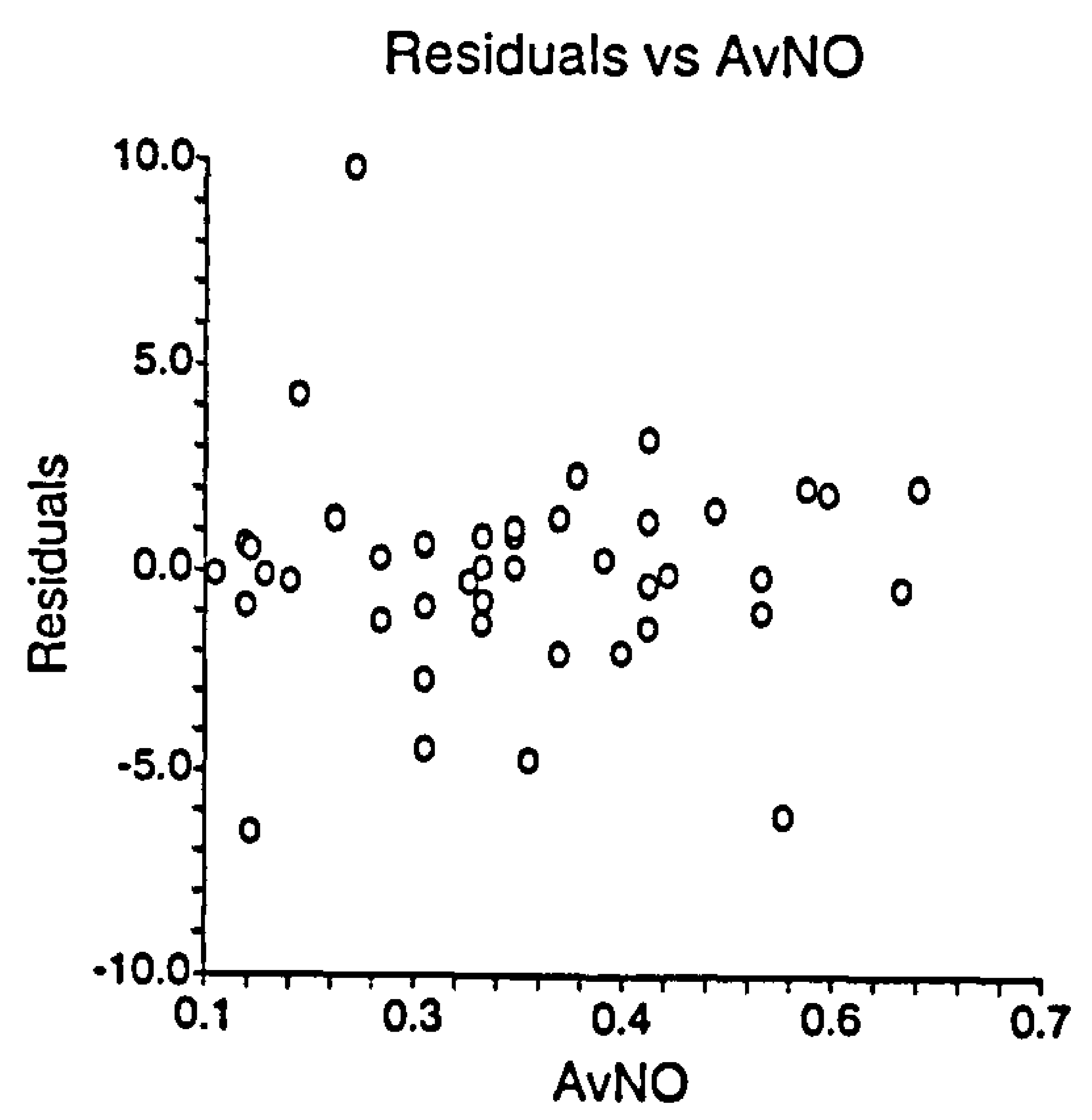
All Condition Numbers less than 100. Multicollinearity is NOT a problem.

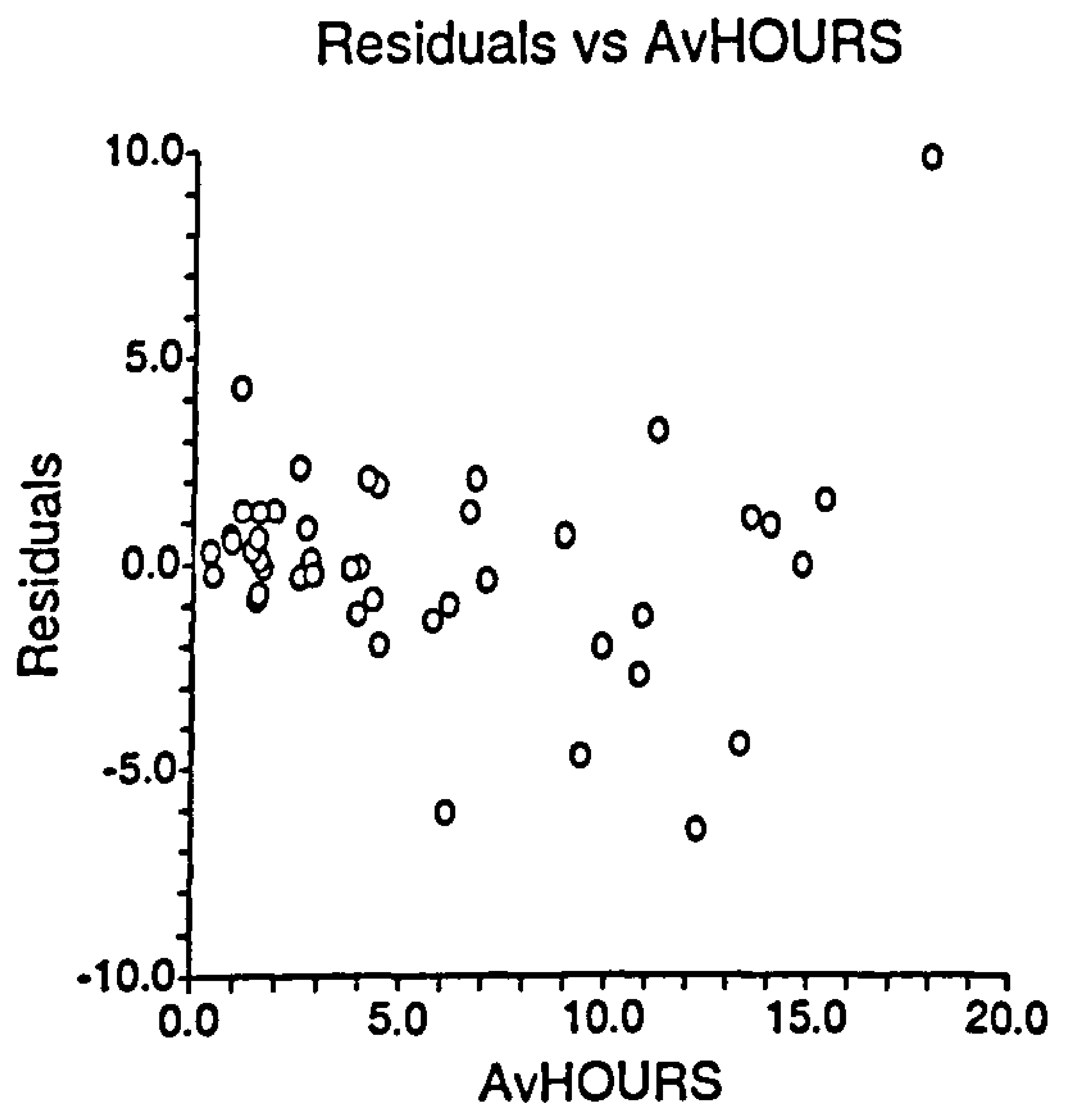
Plots Section

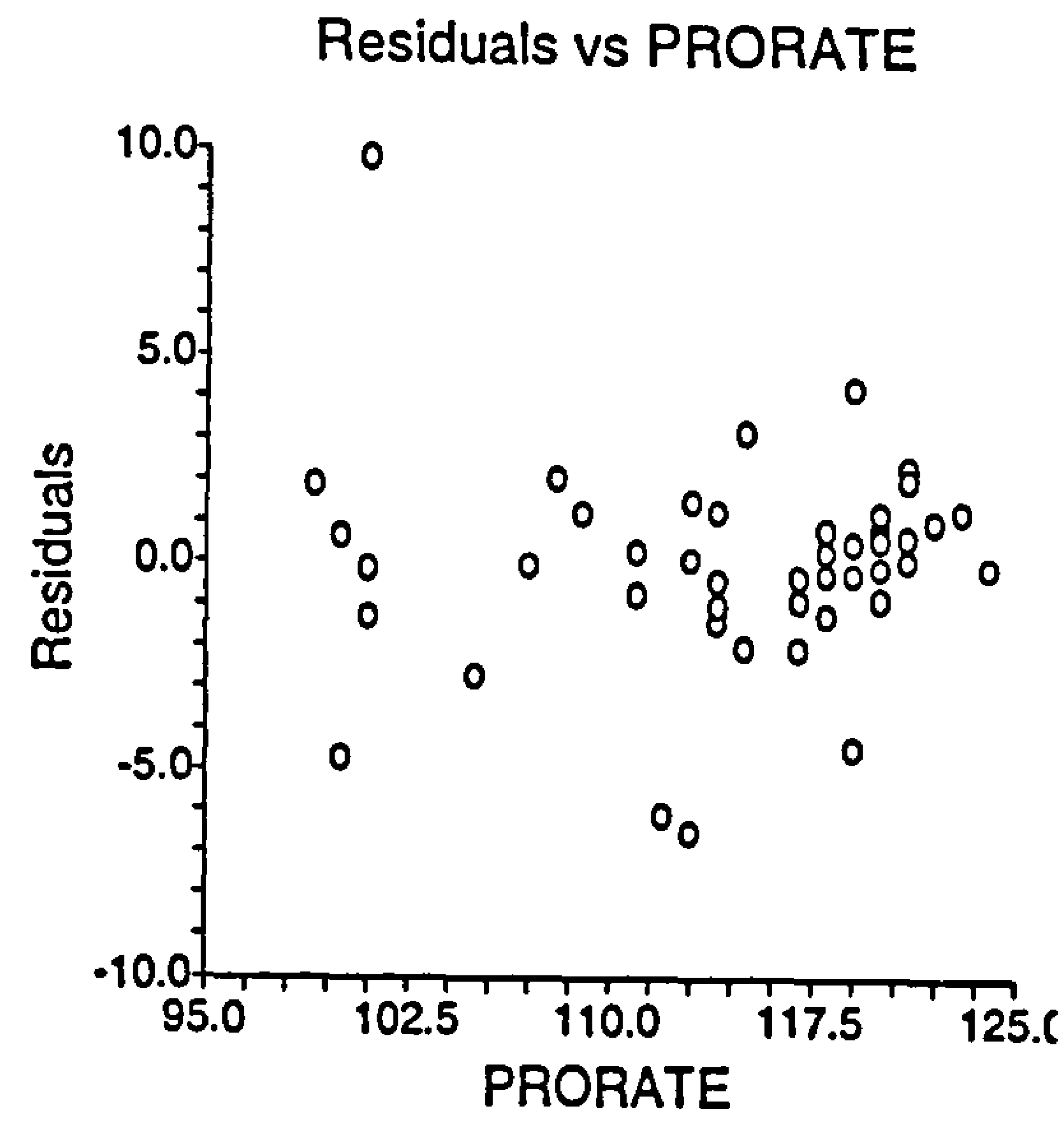


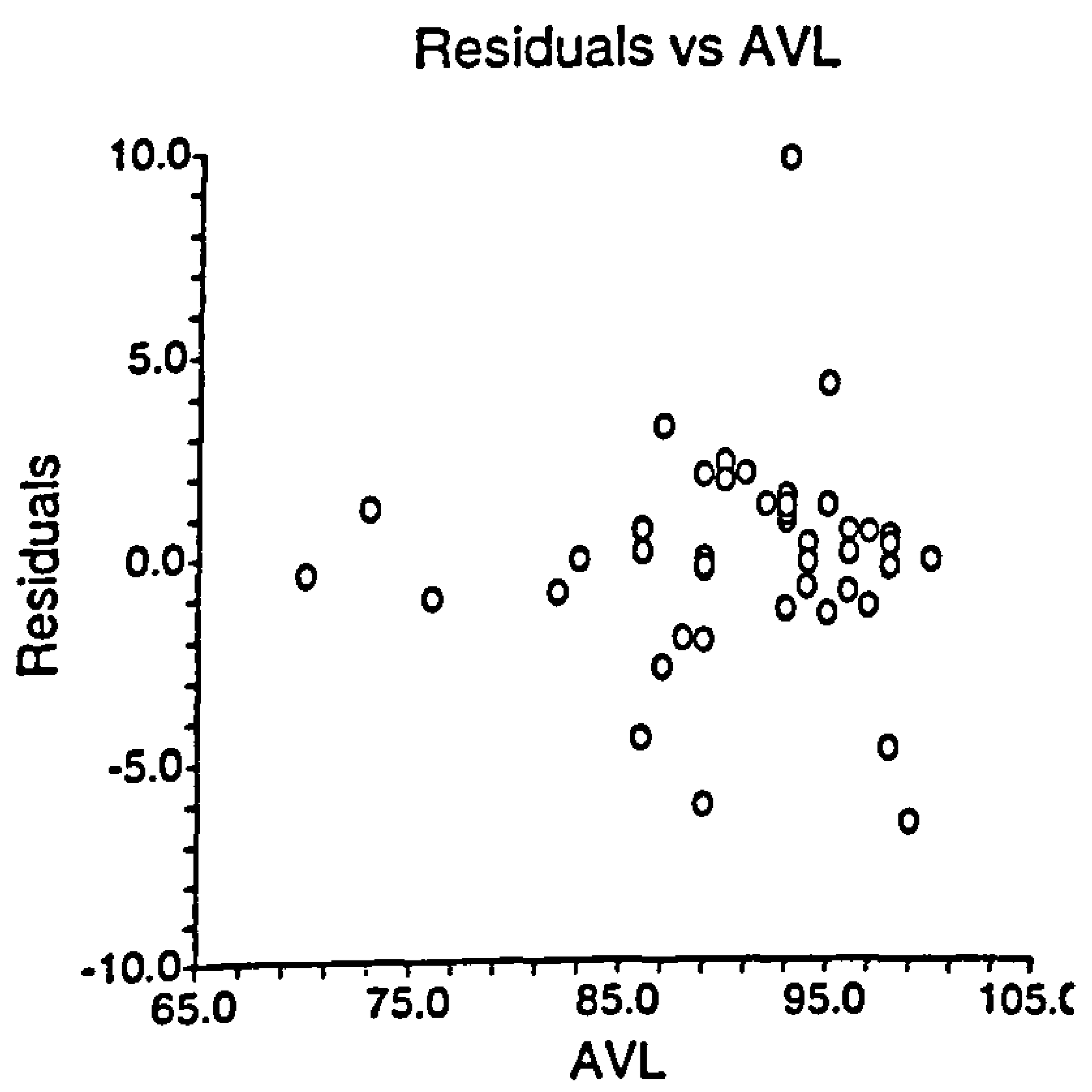












5. Table (7.53) provides the *final robust regression* report. It seems reasonable from this report that:

i) the estimated regression model is EL= -46.12966-6.216772*AvNO+ .3714466*AvHOURS-.1846394*PRORATE-.164175*AVL+8.667938*Sratio+ .962311*LimeSF

ii) The R-squared value is 0.944030, which means that this model explains about 94% of the variability in EL by the independent variables in the model.

Table (7.53) Final run of robust regression

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	-46.12966	17.3371	-2.6607	0.011945	Reject Ho	0.733258
AvNO	-6.216772	1.078716	-5.7631	0.000002	Reject Ho	0.999858
AvHOURS	0.3714466	3.013762E-02	12.3250	0.000000	Reject Ho	1.000000
PRORATE	-0.1846394	1.962513E-02	-9.4083	0.000000	Reject Ho	1.000000
AVL	-0.164175	2.603588E-02	-6.3057	0.000000	Reject Ho	0.999984
Sratio	8.667938	1.888174	4.5906	0.000061	Reject Ho	0.993675
LimeSF	0.962311	0.1646284	5.8454	0.000002	Reject Ho	0.999896

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	22874.89	22874.89			
Model	6	218.5204	36.42007	92.7676	0.000000	1.000000
Error	33	12.95562	0.3925946			
Total(Adjusted)	39	231.476	5.935282			

Root Mean Square Error	**0.6265737**	**R-Squared**	**0.944030**
Mean of Dependent Variable	**28.66953**	**Adj R-Squared**	**0.933854**
Coefficient of Variation	**2.185504E-02**		

6. Conclusions:

Comparing the final runs of the OLS multiple regression model and the final run of the robust regression, we observe that

1) Some of the assumptions on OLS multiple regression are not satisfied since the data has some outliers, and the normality assumption is rejected. However, these problems do not affect the robust regression.

2) The final multiple regression model has only four significant independent variables, namely, AvNO, AvHOURS, PRORATE, and AVL. On the other hand the final robust regresion model has six significant independent variables, namely,AvNO, AvHOURS, PRORATE, AVL, Sraio, and LimeSF.

3) For multiple regression R^2=0.62, while for robust regression R^2=0.93. This means that the robust regression explains much more of the variability in the dependent variable than the multiple regression does.

4) Based on the above observations, we recommend using the final robust estimated regression mode, which is given by

EL=-46.12966-6.216772*AvNO+.3714466*AvHOURS-.1846394*PRORATE

-.164175*AVL+8.667938*Sratio+ .962311*LimeSF

7.5.7.2 FUEL for Kiln 6

From the full output reports, which are given in Appendices (03)-(08), we obtain the following summary of the results.

1- Table (7.54) provides the results of *checking the underlying assumptions*. It seems reasonable from this table that we have problems with the existence of outliers and residuals being correlated. Other assumptions are satisfied. This suggests using the robust regression procedure.

Table (7.54) Check of the Regression Assumptions

	Tools	Result
1. Linearity	1. Scatter Plots	Only with AvNO, AvHOURS, PRORAT AVL, Aratio
	2. Correlation Matrix	Only with AvHOURS, PRORATE, AVL, Aratio
2. Normality of Residuals	1. Skewness Tests	Accepted
	2. Kurtosis Test	Accepted
	3. Omnibus Test	Accepted
	2. Normal Probability Plot	Almost normal
	3. Histogram	Almost normal
3. Constant Error Variance	1. Scatter Plots	No patterns
4. Independent Errors	1. Serial Correlations	Correlated
	2. Durbin-Watson Test	Correlated
5. Multicollinearity	1. Correlation Matrix	Absolute maximum 0.510319
	2. Condition Numbers	All are less than 100
	3. Variance Inflation Factors	All are less than 10
	4. Eigenvalues	No values are close to zero
6. Outliers	1. Histogram	Outliers Exist
	2. Normal Probability Plot	Outliers Exist
	3. Scatter Plots	Outliers Exist

2. Table (7.55) reports a summary of the results of *all possible regression models*. It seems reasonable from this table that based on the Cp criterion the best model is the one with only three independent variables, namely PRORATE,AVL, and Aratio. Moreover, the difference in R-squared between the full model with seven variables and the model with these two variables is about 0.01 which is not worth complicating the model.

Table (7.55) Results of All Possible Regression Procedure

Model Size	R-Squared	Root MSE	Cp	Variables in Model	Best Model
1	0.509426	4.285896	13.841560	PRORATE	
2	0.635339	3.739961	2.022394	PRORATE, AVL	
3	0.666537	3.620827	0.598268	PRORATE, AVL, Aratio	This One
4	0.669816	3.648884	2.238462	AvNO, PRORATE, AVL, Aratio	
5	0.671408	3.687658	4.063674	AvNO, PRORATE, AVL, Aratio, LimeSF	
6	0.671977	3.733925	6.001287	AvNO, AvHOURS, PRORATE, AVL, Aratio, LimeSF	
7	0.671988	3.785362	8.000000	AvNO, AvHOURS, PRORATE, AVL, Aratio, Sratio, LimeSF	

3. Table (7.56) reports a summary of the obtained ***initial regression models*** using different procedures together with their R-squared values. It seems reasonable from this table that

i) The coefficients of the estimated models by sequentially entering one variable at a time are almost the same. This is seen from all possible regression models section. This indicates a type of consistancy of the results.

ii) The robust regression models gave slightly different results because they use different truncation methods in weighing the error terms. However, the largest R-squared is obtained by the Andrew's sine method.

iii) The coefficients of each of the influential variables (PRORATE, and AVL) have the same signs in all models. These signs agree with the resarcher expectations.

iv) Since the data has some outliers and normality is not completely satisfied, we recommend using the robust regression model after deleting the independent variables, which do not contribute to the dependent variable. This will be done in the final run section.

Table (7.56) Comparison of the estimated models using different procedures
A blank means that the variable is not included in the model

Procedure	Intercep	AvNO	AvHOURS	PRORAT	AVL	Aratio	Sratio	LimeSF	R-Squar
All ssible									
Size 1	157.328			-.610394					0.509426
Size 2	185.169			-.593225	-.327421				0.635339
Size 3	203.106			-.525554	-.359668	-13.0040			0.666537
Size 4	207.590	-3.04266		-.538473	-.390159	-12.5804			0.669816
Size 5	183.198	-3.14442		-.540267	-.404586	-11.5819		.2580884	0.671408
Size 6	180.171	-3.35914	-.0353482	-.550825	-.413874	-11.3410		.3106465	0.671977
Size 7 (Multiple)	182.029	-3.37099	-0.0034723	-.551661	-.414057	-11.3164	-.3548891	.3006852	0.671988
Robust									
1-Least Abs Dev. 1.0	162.218	-4.10220	-.0974346	-.571128	-.446674	-7.64951	-2.406976	.5569151	0.743333
2-Tukey's Biweight 6.0	163.322	-5.51061	-0.0305202	-.579876	-.449869	-6.98323	-.6270468	.5031241	0.717617
3-Andrew's Sine 2.1	186.451	-3.56597	-.2126591	-.578979	-.461809	-5.82560	-6.663563	.4008127	0.767835
Stepwise									
1- Forward	185.167			-0.59322	-0.32742				0.569331
2- Step	185.167			-0.59322	-0.32742				0.569331
3-Backward	203.106			-.525554	-.359668	-13.0040			0.666537
4-Min MSE	203.106			-.525554	-.359668	-13.0040			0.666537

4) It seems reasonable from the ***final multiple regression*** report (see Table (7.57) that :

i) R-squared value is 0.626062. This means that this model explains about 63% of the variability in the independent variable. This seams to be a reasonable coefficient of determination.

ii) The PRESS R-squared is 0.5606, which means that this model can predict to good accuracy about 64% of the values of Y. This is useful indicator of predictability.

iii) Normality of the residuals is accepted.

iv) Based on serial correlations and the Durbin-Watson test, the residuals may be correlated..

v) Multicollinearity could not be a problem.

vi) From the plots section, one observes that the data has some outliers.

Table (7.57) Final Run of Multiple Regression

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	184.435	11.86818	15.5403	0.000000	Reject Ho	1.000000
PRORATE	-0.5952328	8.112888E-02	-7.3369	0.000000	Reject Ho	1.000000
AVL	-0.31791	8.703502E-02	-3.6527	0.000714	Reject Ho	0.946080
R-Squared	0.626062					

Model:

FU=184.435-0.5952328 PRORATE-0.31791 AVL

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	184.435	11.86818	160.484	208.386	0.0000
PRORATE	-0.5952328	8.112888E-02	-0.7589576	-0.4315082	-0.6933
AVL	-0.31791	8.703502E-02	-0.4935537	-0.1422662	-0.3452
T-Critical	2.018082				

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	344618.8	344618.8			
Model	2	992.4615	496.2308	35.1591	0.000000	1.000000
Error	42	592.7829	14.11388			
Total(Adjusted)	44	1585.244	36.02828			

Root Mean Square Error	3.756844	R-Squared	0.6261
Mean of Dependent	87.51111	Adj R-Squared	0.6083
Coefficient of Variation	4.292991E-02	Press Value	696.6278
Sum |Press Residuals|	140.0677	Press R-Squared	0.5606

Normality Tests Section

Assumption	Value	Probability	Decision(5%)
Skewness	0.3932	0.694203	Accepted
Kurtosis	0.2883	0.773102	Accepted
Omnibus	0.2377	0.887940	Accepted

Serial-Correlation Section

Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.456203	9	-0.027808	17	-0.069375
2	0.292773	10	-0.051608	18	0.151216
3	0.123676	11	0.000120	19	0.023730
4	-0.044545	12	0.001552	20	-0.011114
5	-0.119775	13	0.031890	21	-0.015427
6	-0.222414	14	-0.007450	22	-0.068404
7	-0.346507	15	-0.010834	23	0.060498
8	-0.201209	16	-0.099242	24	0.058135

Above serial correlations significant if their absolute values are greater than 0.298142
Durbin-Watson Value 1.0420

Multicollinearity Section

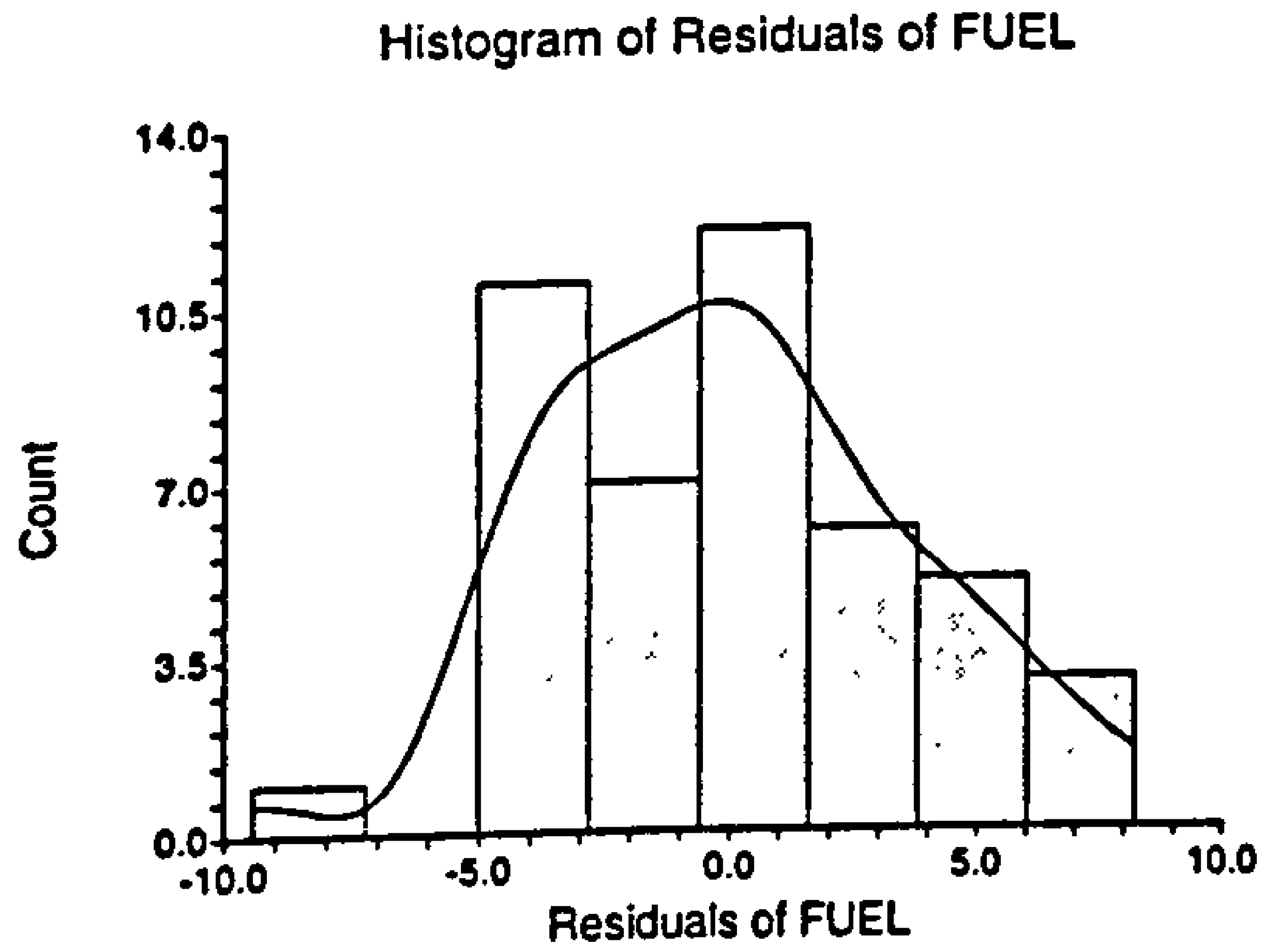
Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance	Diagonal of X'X Inverse
PRORATE	1.003008	0.002999	0.997001	4.663421E-04
AVL	1.003008	0.002999	0.997001	5.367126E-04

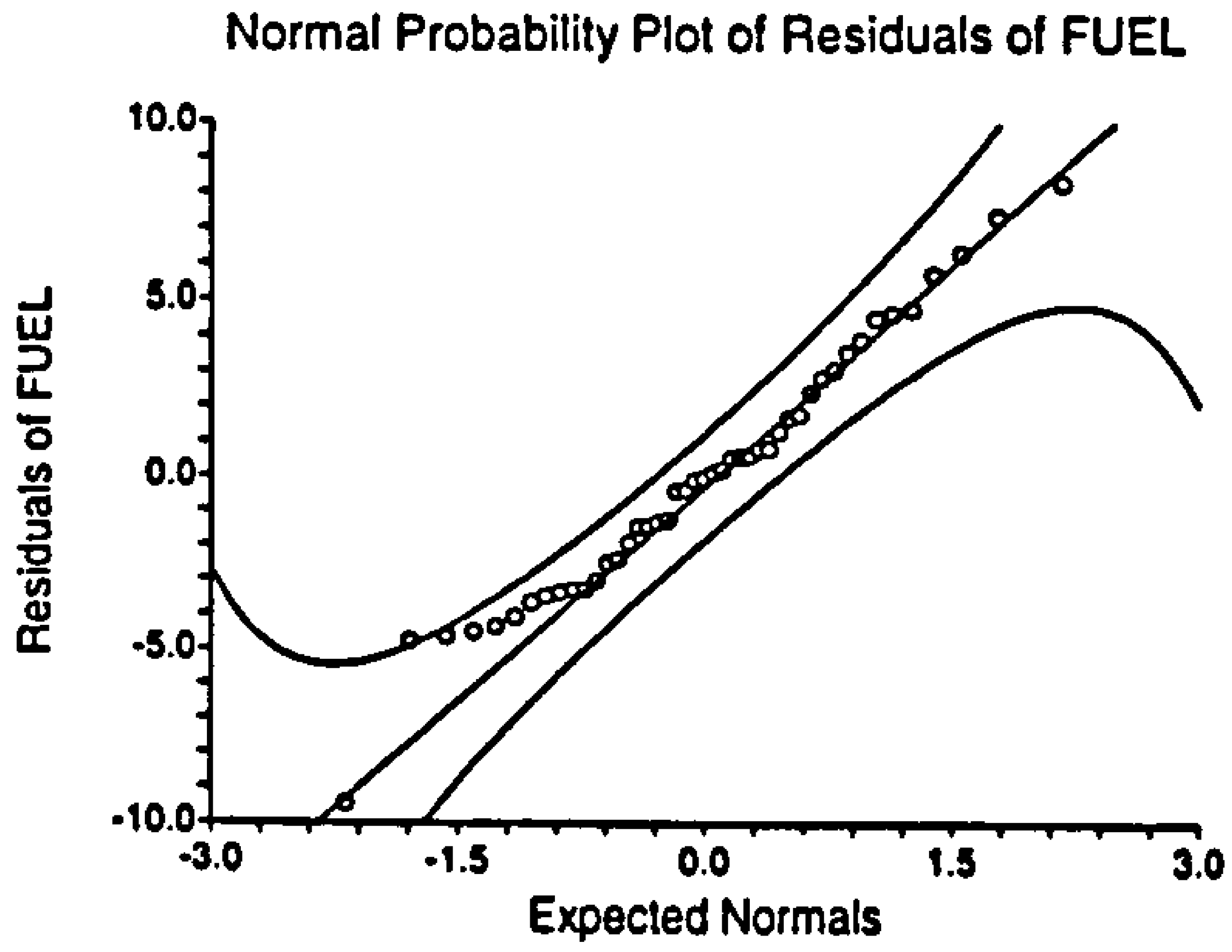
Eigenvalues of Centered Correlations

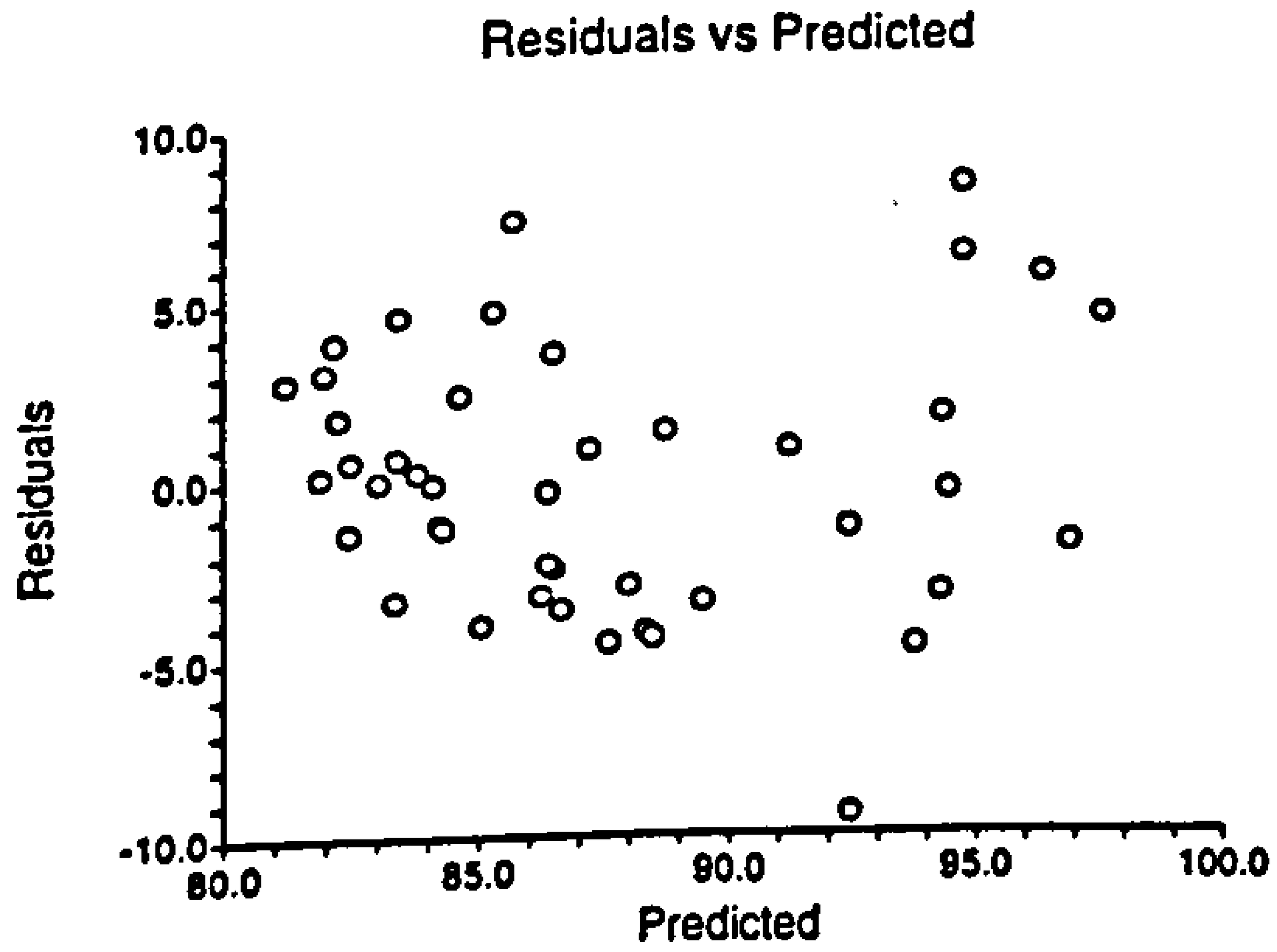
No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.054767	52.74	52.74	1.00
2	0.945233	47.26	100.00	1.12

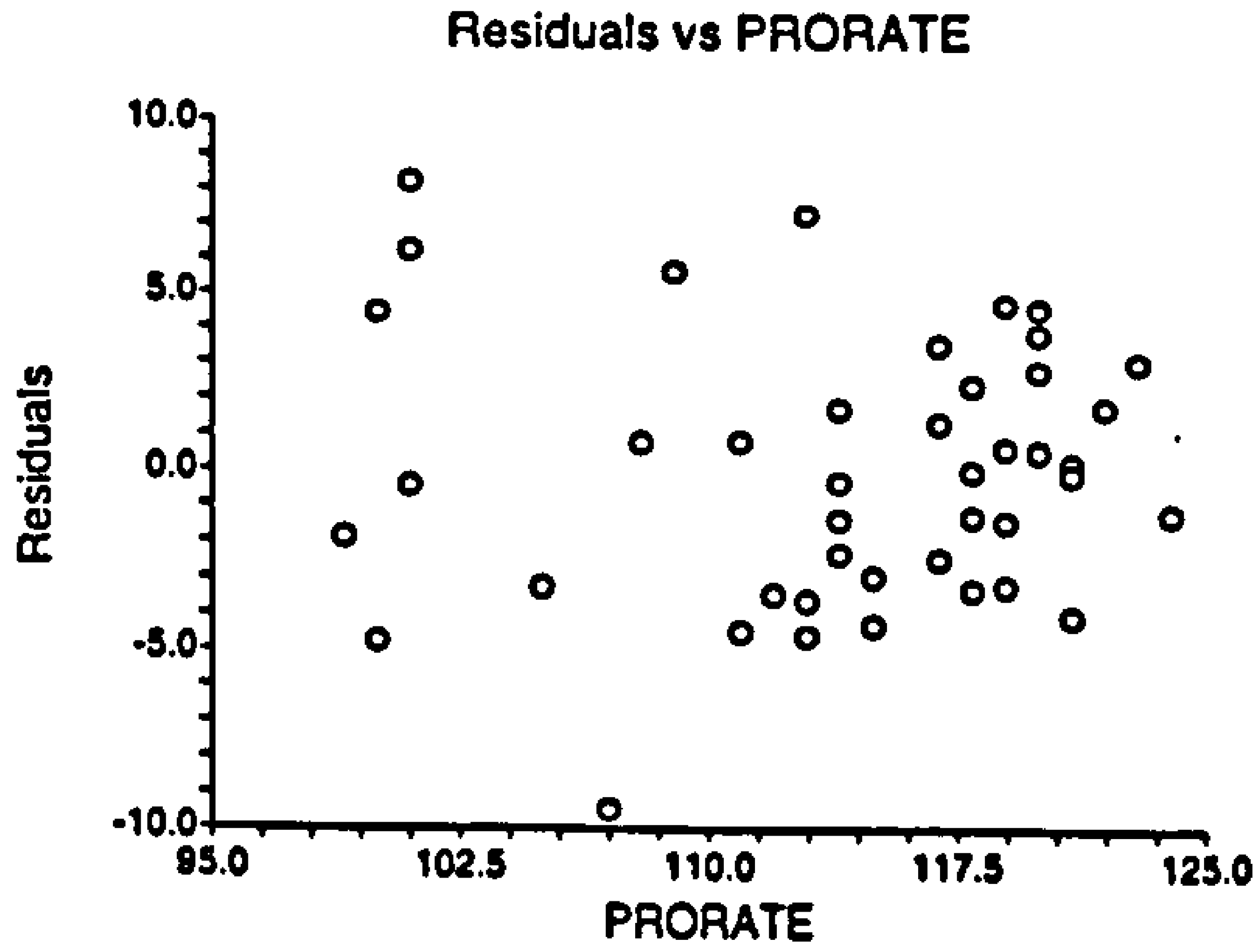
All Condition Numbers less than 100. Multicollinearity is NOT a problem.

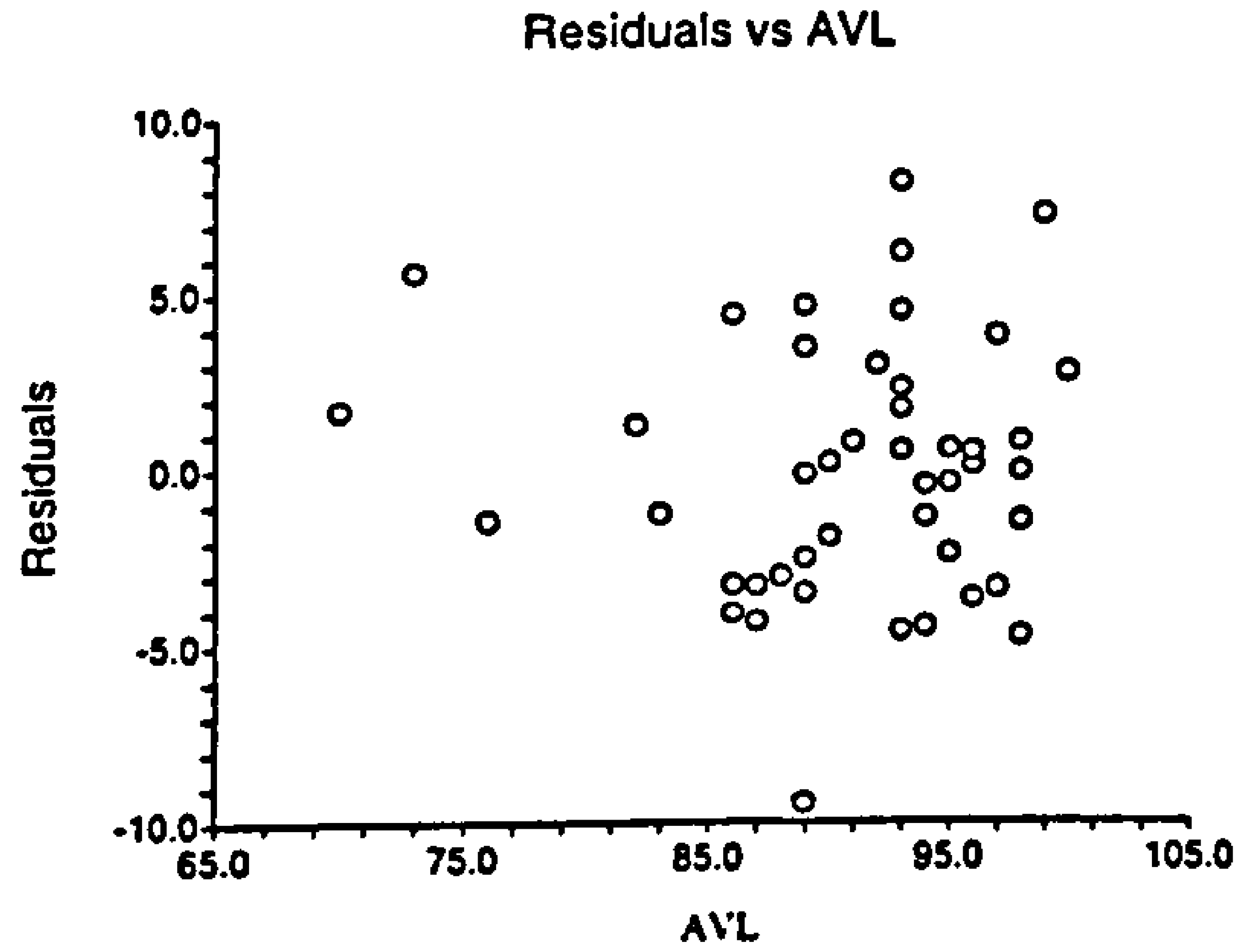
Plots Section











5. Table (7.58) provides the ***final robust regression*** report. It seems reasonable from this report that

i) the estimated regression model is FUEL = 170.4404-. 4811263*PRORATE-.314158*AVL

ii) The R-squared value is 0.682036, which means that this model explains about 68% of the variability in EL by the independent variables in the model.

Table (7.58) Final run of robust regression

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	170.4404	8.931788	19.0824	0.000000	Reject Ho	1.000000
PRORATE	-0.4811263	6.042021E-02	-7.9630	0.000000	Reject Ho	1.000000
AVL	-0.314158	6.217327E-02	-5.0529	0.000009	Reject Ho	0.998533

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	250868.4	250868.4			
Model	2	456.5449	228.2724	45.0452	0.000000	1.000000
Error	42	212.8407	5.067635			
Total(Adjusted)	44	669.3856	15.21331			

Root Mean Square Error	2.251141	R-Squared	0.682036
Mean of Dependent Variable	86.28694	Adj R-Squared	0.666895
Coefficient of Variation	2.608901E-02		

6. Conclusions

Comparing the final runs of the OLS multiple regression model and the final run of the robust regression, we observe that

1) The existance of some outliers is the only problem with the OLS multiple regression. However, this problem does not affect the robust regression.

2) The final multiple regression model and the final robust regresion model have the same two significant independent variables, namely, PRORATE, and AVL.

3) For multiple regression R^2=0.62, while for robust regression R^2=0.68. This means that the robust regression explains slightly higher proportion of the variability in the dependent variable than the multiple regression does.

4) Based on the above observations, we recommend using the final robust estimated regression mode, which is given by

   FUEL = 170.4404-. 4811263*PRORATE-.314158*AVL

7.6 Summary of the Statistical Results

7.6.1 Results of the Kilns

Results about the validity of the assumptions

It seems reasonable from the produced computer output, which is given in the Appendices (0-21) that:

1) There is a limited number of outliers in all variables in the data for all kilns. This is a problem for the usual least square method. So, one may not depend on estimated multiple regression models. A solution to this problem is to use robust regression method, which tries to minimize the effects of outliers by re-weighting them.
2) There is a minor problem in the normality assumption of errors in both the electricity models and the fuel models for almost all kilns. Transformations of the raw data were not useful to make the errors normally distributed. This may be due to the existence of outliers. Again, this is a limitation of multiple regression but it is not a problem for robust regression because robust regression relaxes the normality assumption. So, this is an additional reason to use robust regression.
3) There is no problem about the randomness assumption of the error almost in all models.
4) For almost all models, the residuals (errors) are uncorrelated.
5) For almost all models, the residuals have constant variance for each given value of the independent variables.
6) The results of the ridge regression on kiln 1 are the same as those of the multiple regression. And since multicollinearity is not a problem in all kilns in both EL and FUEL models, we have not reported any results about ridge regression other than kiln 1.

So, based on the above observations, one can apply robust regression to the available data.

Results about significant independent variables

1) There are some independent variables that do not influence the dependent variable, i.e. their coefficients in the regression equations are not significantly different from zero. So these variables should be removed from the final estimated models. Moreover, these variables are kiln dependent, i.e. they vary from one kiln to another. Moreover, for each kiln, they may vary from the electricity model to the fuel model.

2) The following table summarizes the R-squared values of the three final robust procedures

Table (7.59) Comparison of R-squared of Final Robust Regression Models for EL

EL	K1	K2	K4	K5	K6
Andrew	0.886643	0.816465	0.861101	0.909245	0.944030
Tukey	0.733573	0.788249	0.775002	0.824840	0.897732
Abs Dev.	0.824531	0.803048	0.808749	0.854955	0.877640

It is clear from Table (7.59) that for the El models;

1. For each kiln, the smallest value of R-square is for the Tukey procedure, and the maximum value of R-square is for the Andrew procedure with only one exception in kiln 6 where R-squared value for absolute deviation is slightly smaller than that of the Tukey procedure. Therefore, we recommend using results based on the Andrew procedure. So, from now on, the term "final robust model" will mean the final robust model obtained by the Andrew procedure.
2. Regardless of the procedure, the minimum R-squared value is 0.73 and the maximum one is 0.94. All these values indicate that the obtained models are significant.

3) Tables (7.60) and (7.61) show that, (for EL and FUEL models), the robust model introduces more significant independent variables than the corresponding multiple regression model and it contains the same significant independent variables that appear in the corresponding final multiple regression model with one except for the AvNO in kiln 1 which is significant in EL multiple model but not significant in robust model, and two exceptions in the FUEL models where AvNO and PRORATE are significant in multiple model for Kiln 5 but not in robust model.

The signs of the estimated coefficients of the independent variables almost agree with the researcher expectations.

Table (7.60): Signs of estimated coefficients in final EL models using different procedures.

	AvNO					AvHOURS					PRORATE					AVL					Aratio			Sratio			LimeSF		
Kiln No.	4	5	6	1	2	4	5	6	1	2	4	5	6	1	2	4	5	6	1	2	4	5	6	4	5	6	4	5	6
Multiple			-	+	+	+	+	+	+	+	-	-	-	-			-	-		-		-			-				
Robust			-		+	+	+	+	-	+	-	-	-	-			-	-	-	-		-			-	+		+	+

Table (7.61): Signs of estimated coefficients in final FUEL models using different procedures.

	AvNO					AvHOURS					PRORATE					AVL					Aratio			Sratio			LimeSF		
Kiln No.	4	5	6	1	2	4	5	6	1	2	4	5	6	1	2	4	5	6	1	2	4	5	6	4	5	6	4	5	6
Multiple	+	+									-	-	-	-	-			-			-								
Robust	+						+				-		-	-	-	+	-	-			-	-						+	

Results of Final Multiple EL Models

Table (7.62) Final Multiple Regression Models for EL

Kiln	Intercept	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	Lime SF	R-Squared
K 1	26.0888	3.99093	.1998955	-0.64591		NA	NA	NA	0.725577
K 2	21.3857	4.55520	.2051574		-.04967	NA	NA	NA	0.692469
K 4	45.4826		0.539330	-0.72512					0.642689
K 5	170.230		0.327737	-0.24022	-0.2363	-9.957	-35.096		0.792597
K 6	73.7504	-10.0027	0.417057	-0.24812	-0.1714				0.617841

It is clear from Table (7.62) and the computer output in the appendices that (for EL models);

1) The R-squared values range from 0.62 to 0.79, which means that the models can explain at least from 62% to 79% of the variability in EL using the independent variables in the models. This indicates that the obtained models are useful to some extent.

2) Each of the seven independent variables is significant in at least on kiln with an exception of LimeSF. This indicates that the EL consumption may be explained by these six variables.

3) The signs of the estimated coefficients of all independent variables agree with the researcher expectations in almost all cases. The exceptions which disagree with the researcher expectations are

    a) AvNO only in kiln 6, and

    b) Sratio only in kiln 5.

4) The signs of PRORATE, AVL, and Aratio are negative. This means that as these independent variables increase the EL decreases.

5) The signs of AvHOURS and AvNO are positive. This means that as these independent variables increase the EL increases. However, there is an exception with AvNO which is only negative in kiln 6

6) The p-values for the F-test are all zeros. This means that all models are significant.

7) The coefficient of AvHOURS is significant in all kilns. The coefficients of PRORATE are significant in four kilns out of five. The coefficients of AvNO are significant in three kilns out of five. However, The coefficients of Aratio and Sratio are significant in only one kiln out of

three kilns. The LimeSF is not significant in any kiln. These observations reflect the relative importance of these independent variables in explaining the variability in EL.

Results of Final FUEL Multiple Models

Table (7.63) Final Multiple Regression Models for FUEL

Kiln	Intercept	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	Lime SF	R-Squared
K 1	122.434			-.2883676		NA	NA	NA	0.369092
K 2	146.204			-.4817984		NA	NA	NA	0.316854
K 4	130.222	11.20576		-1.32948		-3.60287			0.464826
	94.144	21.42626		-1.27890	.36322	-4.6238			0.5155
K 5	140.879	17.03651		-0.857336					0.471748
K 6	184.435			-0.595232	-0.3179				0.626062

It is clear from Table (7.63) and the computer output in the appendices that (for FUEL models);

1) The R-squared values range from 0.32 to 0.63, which means that the models can explain at least from 32% to 63% of the variability in FUEL using the independent variables in the models. These percentages are low.

2) Only four out of the seven independent variables are significant in at least one kiln. This indicates that the FUEL consumption may be explained by these four variables.

3) The signs of the estimated coefficients of all independent variables agrees with the researcher expectations in all cases.

4) The signs of PRORATE, and Aratio are negative. This means that as these independent variables increase the FUEL decreases.

5) The signs of AvNo, AvHOURS, Sratio and LimeSF are positive. This means that as these independent variables increase the FUEL increases.

6) The p-values for the F-test are all zeros. This means that all models are significant.

7) The coefficient of PRORATE is significant in all kilns. The coefficients of AvNO are significant in two kilns out of five. Aratio is significant in two kilns out of three. The

coefficient of AVL is significant in one kiln out of five. The coefficients of AvHOURS, Sratio, and LimeSF are not significant in any kiln. These observations reflect the relative importance of these independent variables in explaining the variability in FUEL.

Results of Final Robust EL Models

Table (7.64) Final Robust Regression Models for EL
A blank means the coefficient is not significantly different from zero.

Kiln	Intercept	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF	R-Squared
K 1	44.0589		.082888	-.161605	-.047519	NA	NA	NA	0.886643
K 2	21.2816	4.15912	.189376		-.047114	NA	NA	NA	0.816465
K 4	40.7076		.635295	-.550991					0.861101
K 5	90.2706		.256176	-.196848	-.177401	-7.99331	-30.1843	.59563	0.909245
K 6	-46.1296	-6.21677	.371446	-.184639	-.164175		8.66793	.96231	0.944030

It is clear from Table (7.64) and the computer output in the appendices that (for EL models);

1) The R-squared values range from 0.82 to 0.94, which means that the models can explain at least from 82% to 94% of the variability in EL using the independent variables in the models. This indicates that the obtained models are significant.

2) Each of the seven independent variables is significant in at least one kiln. This indicates that the EL consumption may be explained by these variables.

3) The signs of the estimated coefficients of all independent variables agree with the researcher expectations in almost all cases. The exceptions which disagree with the researcher expectations are

   a. AvNO only in kiln 6, and

   b. Sratio only in kiln 5.

4) The signs of PRORATE, AVL, and Aratio are negative. This means that as these independent variables increase the EL decreases.

5) The signs of AvHOURS, AvNO, Sratio and LimeSF are positive with an exception of one case of each of AvNO and Sratio. This means that as these independent variables increase the EL increases.

6) The p-values for the F-test are all zeros. This means that all models are significant.

7) The coefficient of AvHOURS is significant in all kilns. The coefficients of PRORATE and AVL are significant in four kilns out of five. The coefficients of AvNO are significant in two kilns out of five. The Sratio, and LimeSF are significant in two out of three kilns. However, The coefficient of Aratio is significant in only one kiln out of three. These observations reflect the relative importance of these independent variables in explaining the variability in EL.

8) The power of the explanation in the above model (81-94%) considered to be high. This could be explained due to the fact that the model is handling the electricity consumption where the factors affecting this consumption considered to be relatively limited because most of the electrical machines are constant speed machines. In case of unsteady operational condition the production rate will be decreased and this resulted in increase of the average electricity consumption, also in case of repeated emergency stoppages this will decrease in availability of the production line and resulted in additional energy consumption at no load of the kiln or during shutdown which will increase the average electricity consumption of the product

Results of Final Robust FUEL Models

Table (7.65) Final Robust Regression Models for FUEL

A blank means the coefficient is not significantly different from zero.

Kiln	Intercept	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF	R-Squared
K 1	130.764			-.356849		NA	NA	NA	0.569486
K 2	42.4227			-.191506		NA	NA	NA	0.679696
K 4	81.2499	24.3102		-1.34530	.503356	-3.91432			0.768843
K 5	-99.7751		1.30012		-.518748	-12.6086		2.7278	0.891453
K 6	170.440			-. 48112	-.314158				0.682036

It is clear from Table (7.65) and the computer output in the appendices that (for FUEL models);

1) The R-squared values range from 0.57 to 0.89, which means that the models can explain at least from 57% to 89% of the variability in FUEL using the independent variables in the models.

2) Each of the seven independent variables except the Sratio is significant in at least one kiln. This indicates that the FUEL consumption may be explained by these six variables.

3) The signs of the estimated coefficients of all independent variables agree with the researcher expectations in almost all cases. The only exception is AVL in kiln 4.

4) The intercept is significant for all kilns except for kiln 5. However, it is not easy to interpret the negative value of the intercept in case of kiln 5.

5) The signs of PRORATE, and Aratio are negative. This means that as these independent variables increase the FUEL decreases.

6) The signs of AvNo, AvHOURS and LimeSF are positive. This means that as these independent variables increase the FUEL increases.

7) The sign of AVL is negative, except kiln 4 which is positive.

8) The p-values for the all F-test are zeros. This means that all models are significant.

9) The coefficient of AvHOURS is significant in all kilns. The coefficients of PRORATE are significant in four kilns out of five. The coefficients of AVL are significant in three kilns out of five. The coefficient of Aratio is significant in only two kilns out of three .The coefficients of AvNO, and AvHOURS, are significant in only one kiln out of five. LimeSF is significant in one out of three kilns. However, the coefficient of Sratio is not significant in any of the three kilns. These observations reflect the relative importance of these independent variables in explaining the variability in FUEL.

10) The explanation power of the above fuel model (57-89%) is considered to be acceptable but we noticed that it is lower than the power explanation of the electrical model. This can be explained due to the fact that the burning process in the kiln is a very complicated process performed in

several stages with several operating conditions (temperature, pressure, chemical reactions, cooling and heating process, variation of the quality of the raw materials ...etc, which means that the operational conditions can be affected by several factors and needs very special skills to keep it in steady conditions. The variation in operational conditions will mainly affect the production rate of the kiln causing a serious variation on the fuel consumption. Emergency breakdowns resulted in lower availability will cause additional fuel consumption in case the kiln at no load condition. Also it will cause additional fuel consumption in case of heating up the kiln uptill reaching steady state operational conditions. To improve the explanation power of the model it is recommended to investigate additional variables, which may affect the fuel consumption, this can be done in further research.

Results on Nonlinear Regression for Kiln 1

Tables (7.66a) and (7.66b) provides the R-squared values for some nonlinear model for kiln1. Comparing these values with the corresponding ones in the final robust models, we observed that these nonlinear models were dominated by the final robust models. Futhere more, nonlinear models are complicated this is why we have deferred all results about these models to the appendix (20).

Table (7.66a) Comparison of Nonlinear Models for EL of kiln 1

Model	R-Squared
Multiplicative	0.430624
Quadratic	0.807448
Polynomial with interactio	0.825879
Logarithmic	0.569801

Table (7.66b) Comparison of Nonlinear Models for FUEL of Kiln 1

Model	R-Squared
Quadratic	0.433526
Polynomial with interactions	0.484707
Logarithmic	0.400893

Effect of Rounding the Data on Regression

It was stated that the some of the data is rounded for some variables. These were rounded to the nearest integer. The data for other variables are not rounded. To study the effect of rounding the data on the obtained initial and final multiple and robust regression models, we have considered the rounded variables in Kiln 6. Since the variables were rounded to the nearest integer then the rounding error will be between 0 and 0.5. Moreover, there are no records of the unrounded data. So, we have used Mathematica 4.1 Package to generate uniform random values from the interval (0,0.5) and to add them to the values of the rounded data. The new "unrounded" variables(EL, FUEL, PRORATE and AVL) were labeled REL, RFU, RPR, and RAV respectivelly. This seams to be a reasonable procedure to obtain data similar to the unrounded data. After this process, we have run both initial and robust multiple and robust regression models for both EL and FUEL. The computer output is reported in Appendix (21). A summary of the results is given in Tables (7.66c)-(7.66f). It is clear from these tables that, in all cases,

1) The R-squared values of both models based on rounded or unrounded data are almost the same.
2) The estimated coefficients of the independent parameters in both models are almost the same in magnitude and sign.
3) The significant independent variables in both models are the same.

So, one may claim that rounding the data on some variables has very minor effects on the results.

Table(7. 66c) Comparison of Multiple EL Regression Models
Based on Rounded and Unrounded Data

Initial	Intercept	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF	R-Squared
El	-12.32015	-10.42198	.3354227	-.2523372	-.2334629	-3.598366	5.115469	.8965291	0.664180
REL	-12.452	-10.798	.32368	-.23941	-.23398	-2.9712	4.1948	.92892	0.65946
Final									
EL	73.75043	-10.0027	0.417057	-0.24812	-0.17140				0.61784
REL	74.4247	-10.258	0.40721	-0.2511	-0.1695				0.61266

Table(7.66d) Comparison of Robust EL Regression Models
Based on Rounded and Unrounded Data

Initial	Intercept	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF	R-Squared
El	-59.40908	-6.034978	.3517462	-.1913321	-.1742463	2.031039	8.29473	1.093266	0.944
REL	-53.0900	-5.87408	.348381	-.193613	-.154948	2.71095	7.00200	1.0336	0.938
Final									
EL	-46.1296	-6.21677	.3714466	-.184639	-.164175		8.66793	.962311	0.944
REL	-47.1702	-6.31719	.367193	-.182856	-.152915		8.38989	.97193	0.944

Table(7.66e) Comparison of Multiple FUEL Regression Models
Based on Rounded and Unrounded Data

Initial	Intercept	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF	R-Squared
FUEL	182.0291	-3.370992	-0.003472324	-.5516616	-.4140579	-11.31641	-.3548891	.3006852	0.671988
RFU	180.92	-3.50333	-.03568044	-.554481	-.420630	-11.08196	-.81855	.33620	0.67898
Final									
FUEL	184.435			-0.5952328	-0.31791				0.626062
RFU	185.244			-.5959902	-.3200226				0.63141

Table(7.66f) Comparison of Robust FUEL Regression Models
Based on Rounded and Unrounded Data

Initial	Intercept	AvNO	AvHOURS	PRORATE	AVL	Aratio	Sratio	LimeSF	R-Squared
FUEL	186.4518	-3.565976	-.2126591	-.5789794	-.4618095	-5.825606	-6.663563	.4008127	0.767835
RFU	187.045	-3.569283	-.1936044	-.5785337	-.4619228	-5.033528	-7.2385	.39825	0.771165
Final									
FUEL	170.440			-.4811263	-.314158				0.68203
RFU	170.7			-.484849	-.307915				0.6823

7.6.2 Results of Mills

As previously explained, the main consumers of energy are the kilns, cement mills and the raw mills. The previous analysis was done on kilns to verify the effects of the independent variables on the energy consumption including electricity and fuel.

The raw and cement mills are major consumers of electricity consumption and the energy consumed by these mills represents around 35% of the total energy consumed in cement manufacturing. According to the empirical analysis established in Chapter 6, the independent variables, production rate, availability, and number of stoppages and duration of stoppages were selected to study their effect on electrical energy consumption for the cement and raw mills. A similar statistical analysis to that of the kilns was carried out for the raw and cement mills. The variables (production rate and availability) were defined before, but for the number of stoppages and duration of stoppages are defined as follows:

Duration of stoppages (HOURS):

This represents the total stoppages hours for the equipment or the production unit during the month including all types of stoppages (down time + programmed stoppages)

Number of stoppages (NO):

This represents the total number of stoppages of the production line per a month. This number of stoppages includes all types of stoppages (break down: mechanical, electrical, production and programmed maintenance).

A complete listing of computer output is given in Appendices (14)- (18). It should be noted that the Appendices provide the computer output of all regression procedures: all possible regression, stepwise, multiple, and robust regression. Moreover, exploration of the data is provided in the Appendix (14).

From the exploration of the computer output of the available data about mills, we observe that:

1) The data has some outliers.

2) There are linear relationships between EL and some of the independent variables, but such a relationship is not clear with some other variables.

3) The normality of residual distribution is maintained in almost all cases.

4) Randomness assumption of errors is satisfied, in almost all cases.

5) Independence of errors is maintained in almost all regression models.

6) It can be seen from Table (7.67a) that

   a) NO and HOURS has very strong correlation in 10 mills out of 12, the exceptional two cases are mills Cm5 and Rm5.

   b) The significant correlations that are at least 0.75 occurred in Cm6, Cm7, R-cm2, and R-cm3. In these cases multicollinearity may be a problem.

   c) So, one may recommend using only one of them in the regression model. He may choose the variable with larger correlation with the independent variable EL. If this is done then one will avoid any suspicion of multicollinearity.

Table (7.67a): Absolute Maximum Correlation between Independent Variables

Mill	Max. Correlation Between given variables	Mill	Max. Correlation Between given variables
Cm4	0.637440 ((NO,HOURS)	Rm4	0.577019 (NO, HOURS)
Cm5	0.594690 (NO,AVL)	Rm5	0.262090 (NO,PRORATE)
Cm6	0.773979 (NO,HOURS)	Rm6	0.529388 (NO,HOURS)
Cm7	0.824325 (NO,HOURS)	R-cm2	0.824925 (NO,HOURS)
R-rm1	0.512549 (NO,HOURS)	R-cm3	0.834379 (NO,HOURS)
R-rm2	0.704405 (NO,HOURS)	R-cm4	0.529388 (NO,HOURS)

7) It can be seen from Table (7.67b) that there are 5 mills out of 12 in which both NO and HOURS have significant correlation with EL.

8) No multicollinearity exists in all the regression models even though NO and HOURS were included in the models. Table (7.67c) shows that there are five cases in which both are significant. However, in one of these mils, namely, Rm5, the correlation between NO and HOURS is less that 0.75.

Table (7.67b): Independent variables with significant correlation with the EL

Mill	Variables	Mill	Variables
Cm4	NO,PRORATE,AVL	Rm4	HOURS,PRORATE
Cm5	NO,PRORATE	Rm5	PRORATE
Cm6	HOURS,PRORATE	Rm6	NO,HOURS
Cm7	HOURS,PRORATE,AVL	R-cm2	NO,HOURS,AVL
R-rm1	HOURS	R-cm3	NO,HOURS,PRORATE,AVL
R-rm2	NO,HOURS	R-cm4	NO,HOURS

Table (7.67c): Significant variables in multiple regression model of EL

Mill	Variables	Mill	Variables
Cm4	PRORATE,AVL	Rm4	PRORATE
Cm5	PRORATE	Rm5	NO,HOURS,PRORATE
Cm6	NO,HOURS,PRORATE	Rm6	NO,HOURS
Cm7	HOURS,PRORATE,AVL	R-cm2	HOURS,PRORATE
R-rm1	NO,HOURS	R-cm3	HOURS
R-rm2	HOURS, PRORATE	R-cm4	NO,HOURS

9) It seems reasonable from Table (7.67d) that the R-Squared values for the robust regression models of all mills are greater than .75 except for R-rm2. This indicates that the regression models have useful capability to explain the variability in EL using the available independent variables. The significant variables in the models are given in Table (7.67e).

Table (7.67d): Full Robust regression models for raw mills and cement mills.

Mill	Intercept	NO	HOURS	PRORATE	AVL	R-Squared
Cm4	90.53497	-0.04979696	-0.003311579	-.8052055	-0.1623507	0.812713
Cm5	101.2321	-0.0793353	0.01370153	-.387904	-0.0865309	0.866787
Cm6	95.17316	0.1105267	0.0233406	-.3366229	0.05994438	0.942969
Cm7	113.0717	-0.009459768	0.01718058	-.347304	-0.05360653	0.967995
R-rm1	29.90324	0.02779012	0.01174738	-.031937	-0.03668998	0.782735
R-rm2	42.30705	0.01991421	0.008826557	-.06835705	-0.08741473	0.640407
Rm4	44.00504	-0.0216251	0.003414498	-.3041281	-0.02896447	0.953502
Rm5	36.12789	0.1099125	0.01589599	-.1468642	0.0273312	0.787710
Rm6	43.328	-0.005750638	0.00270675	-.09397328	-0.04355091	0.858078
R-cm2	100.6041	-0.1147548	0.008798144	-.6338901	0.01835155	0.897508
R-cm3	63.75738	-0.06033308	0.01677138	-.294803	0.09251468	0.871942
R-cm4	60.11334	-0.1143857	0.006440092	-.1012563	0.01759519	0.813413

Table (7.67e): Significant variables in robust regression model of EL

Mill	Max. Correlation	Mill	Max. Correlation
Cm4	NO,PRORATE,AVL	Rm4	HOURS,PRORATE,AVL
Cm5	NO,HOURS,PRORATE,AVL	Rm5	NO,HOURS,PRORATE,AVL
Cm6	NO,HOURS,PRORATE	Rm6	PRORATE
Cm7	HOURS,PRORATE,AVL	R-cm2	NO,HOURS,PRORATE
R-rm1	NO,HOURS	R-cm3	NO,HOURS,PRORATE,AVL
R-rm2	HOURS, PRORATE,AVL	R-cm4	NO,HOURS

10) From tables (7.68) we conclude that:

11) The sign of the number of stoppages variable disagrees with the researcher expectations in 5 mils out of 12. We checked the possible causes for this phenomena and the answer which we received from the operational people that some of the operators were not precise in recording the stoppages of the mills during the peak load period because as they wrongly claim that these stoppages are routine stoppages and need not to be recorded.

12) The sign of the duration of stoppages variable is always positive which agrees with the researcher expectations.

13) The sign of the production rate variable is always negative which agrees with the researcher expectations.

14) The sign of the availability variable is negative ,(as expected), except in Rm5 and R-cm3.

Table (7.68): Signs of influential variables in robust models of EL.
Blank means that the variable is not significant

	NO	HOURS	PRORATE	AVL
Cm4	-		-	-
Cm5	-	+	-	-
Cm6	+	+	-	
Cm7		+	-	-
R-rm1	+	+		
R-rm2		+	-	-
Rm4		+	-	-
Rm5	+	+	-	+
Rm6			-	
R-cm2	-	+	-	
R-cm3	-	+	-	+
R-cm4	-	+		

7.6.3 Comparison with Industry Experience

In order to verify the results of the statistical models it is necessary to compare them with available statistical data in the cement industry. Exhaustive effort of the researcher to find such information/data with the most reputable cement manufacturers/associations such as British Cement Association, Arab Cement Union, international cement manufacturers such as Lafarge, Holder Bank, and energy associations such as Energy Efficiency Office etc., revealed that there is no such data available and there is a real lack of statistical analysis approach in the industry. The only available data of relevance is a compilation report on main equipment data and operating statistics of one of the leading international cement manufacturers.

The report contains statistical data on fuel consumption, number of stoppages, actual kiln operating time, production rate, availability and other variables. We chose one type of kilns, which are distributed in different factories all over the world, and we ran the same types of regression analysis.

It should be noted that the available data from that reputable manufacturer's report is different from our data. In the sense that the data represent 30 kilns in 30 countries, while our data represent data for five individual kilns in Jordan over a time period. It is hoped that the results of individual kilns in Jordan will provide similar results to the average behavior of the reputable manufacturer's data. If that is the case, then this will support our analysis.

A complete computer run appears in Appendix (19). From the full output reports, which are given in Appendix (19), we obtain the following summary of the results.

1) The only independent variables, which have significant correlations with the dependent variable FUEL, are PRORATE and NO. So these are the variables that one expects to be included in the model.

2) The absolute maximum correlation between the independent variables is 0.844989 between AVL and HOURS.

3) The data has some outliers.

4) Table (7.69) provides the results of checking the underling assumptions. It seems reasonable from this table that we have only problems with the existence of some outliers. Other assumptions are satisfied. This suggests using robust regression procedure.

Table (7.69) Check of the Regression Assumptions

	Tools	Result
1. Linearity	1. Scatter Plots	Only with PRORATE, NO
	2. Correlation Matrix	Only with PRORATE, NO
2. Normality of Residuals	1. Skewness Tests	Accepted
	2. Kurtosis Test	Accepted
	3. Omnibus Test	Accepted
	4. Normal Probability Plot	Normal
	5. Histogram	Normal
3. Constant Error Variance	1. Scatter Plots	No patterns
4. Independent Errors	1. Serial Correlations	Uncorrelated
	2. Durbin-Watson Test	Uncorrelated
5. Multicollinearity	1. Correlation Matrix	Absolute maximum 0.844989 between AVL & HOURS. However, the regression analysis showed that there is no problem of multicollinearity.
	2. Condition Numbers	All are less than 100. No problem
	3. Variance Inflation Factors	All are less than 10
	4. Eigenvalues	No values are close to zero. No problem

6. Outliers	1. Histogram	No Outliers
	2. Normal Probability Plot	No Outliers
	3. Scatter Plots	Outliers Exist

5) Table (7.70) reports a summary of the results of all possible regression models. It seems reasonable from this table that based on the Cp criterion the best model is the one with all four independent variables.

Table (7.70) Results of All Possible Regression Procedure

Model Size	R-Squared	Root MSE	Cp	Variables in Model	Best Model
1	0.370809	199.633	25.731869	NO	
2	0.544714	173.1809	13.986271	PRORATE, NO	
3	0.638920	157.4069	8.540098	PRORATE, NO,AVL	
4	0.709012	144.345	5.000000	PRORATE,NO,AVL , HOURS	This one

6) Table (7.71) reports a summary of the obtained regression models using different procedures together with their R-squared values. It seems reasonable from this table that

a) The coefficients of the estimated models by sequentially entering one variable at a time are almost the same. This is seen from all possible regression models section.

b) The robust regression models gave slightly different results because they use different truncation methods in weighing the error terms. However, the largest R-squared is obtained by the Andrew's sine method.

c) The coefficients of the independent variables agree with the researcher's expectations to some extent.

Table (7.71) Comparison of the estimated models using different procedures

Procedure	Intercept	NO	HOURS	PRORATE	AVL	R-Squared
All Possible						
Size 1	3460.706	3460.706				0.370809
Size 2	3803.208	-.2197309		1.566703		0.544714
Size 3	4472.333	1.45332		-.2401529	-7.711369	0.638920
Size 4 (Multiple)	4935.27	1.012281	0.1104942	-.3096707	-20.87037	0.709012
Stepwise						
1- Forward	4472.333	1.45332		-.2401529	-7.711369	0.638920
2- Step	4472.333	1.45332		-.2401529	-7.711369	0.638920
3-Backward	4472.333	1.45332		-.2401529	-7.711369	0.638920
4-Min MSE	4472.333	1.45332		-.2401529	-7.711369	0.638920
Robust						
1-Least Abs. Dev. 1	4853.686	0.8637481	0.1196288	-.2782551	-20.94327	0.809344
2-Tukey's Biweight (	4533.963	1.159865	0.08188538	-.2255317	-15.34315	0.755562
3-Andrew's Sine 2.1	5103.305	0.8119035	0.1396625	-.2936428	-25.25007	0.843503

The original data given by the reputable manufacture contained as independent variable the annual working hours and the above run used this variable along with the other 3 variables. In our original analysis, the duration of stoppages was used in all cases and not the total annual working hours as in this case. So an additional statistical run was carried after calculating duration of stoppages by deducting the total annual working hours from the total hours in one year. The results of this run gave almost the same R^2 value, and all the variable showed almost the same effect.

It is worth noticing that the findings are very close to the research results concerning the fuel consumption, which supports the findings of the research that the variables included are significant.

7.7 Discussion of the Statistical Findings

In the foregoing part of this chapter, the findings of this research were presented. In this part, the major trends and patterns of relationships are analysed. Therefore, the main focus is to track the direction of the relationship between electricity and fuel consumption on the one hand and the independent variables, namely, average number of stoppages, average duration of stoppages, production rate, availability, Alumina Ratio (Aratio), Silica Ratio (Sratio), and Lime Saturation Factor (LimeSF) on the other hand. The implications of these results can then be identified.

It may be useful to recall that the researcher expects that AvNO, AvHOURS, Sratio and LimeSF have increasing relationships with both EL and FUEL, i.e. as these independent variables increase, the dependent variable increases. On the contrary he has the expectation that the other three independent variables (PRORATE, AVL, and Aratio) have decreasing relationship with each of the dependent variables (EL and FUEL), i.e. as each of these variables increase the dependent variable decreases.

It is worth noting that different Kilns studied had slightly different energy consumption models, mainly due to the effects of the independent variables based on their different values, which are related to the operational efficiency, the type of the kiln, its age, design and maintenance.

Operational efficiency is an important factor, which can affect the level of energy consumption. If we have well-trained operators this will help in achieving the objective of continuity and steadiness of operation and also it will help to achieve the optimum operational and production level, thus achieve the optimum energy consumption.

The type of the kiln is an important factor, which can affect the energy consumption, if the kiln is wet or dry process. According to Saxena, J.P. et al (1995)., the specific heat consumption can vary between 1500 kcal/kg of clinker for the wet process to around 700 kcal/kg of clinker for the six stages preheater plus calciner plus high efficiency cooler. Also the electrical energy consumption varies in accordance with the kiln type and the type of the mills used.

The design features are important factors, which affect the energy consumption in general, for example a reliable automatic control system can assure the continuity and steadiness of operation, which is of great importance of energy consumption in cement kiln.

The age of the kiln can affect the energy consumption. Usually aging factors affect the availability and reliability of the equipment, which may cause increase in the number and duration of stoppages, which in turn will increase energy consumption. Continuous upgrading and modification to cope with technological advance will be needed for the old kiln.

Proper maintenance management is of great importance to avoid overloading of machines, decrease friction between parts, avoid drifts in conveyor belts and in general to enhance the stability of the operational process to avoid the stoppages and the loss of production which will cause an increase in energy consumption.

Based on our statistical analysis of the available data, we concluded that the most suitable regression procedure for the available data is the robust regression, that is due to the following reasons;

i) It is not sensitive to outliers or to normality assumption where both form minor problems in some cases of our data.

ii) The other linear models have smaller R-squared values than that of the robust model.

iii) The nonlinear models are more complicated than the robust model and they have R-squared value smaller than that of the robust model.

iv) Moreover, the signs of estimated coefficients of the independent variables are consistent with the researcher expectations, in the case of robust regression, but some limited number of those signs may disagree with those expectations when we have used other models.

So our following conclusions will be based on the final runs of the robust regression using Andrew's sine weight function.

First, All the obtained robust regression models for all kilns are useful since each of them has a significant F-value at the 0.05 level of significance. So, these models are useful. The forms of the estimated models for EL and FUEL are given in Tables (7.64) and (7.65).

Second, There are some of the independent variables that do not influence the dependent variable, i.e. their coefficients in the regression equations are not significantly different from zeros. They vary from one kiln to another. Moreover, for each kiln, they may vary from the electricity model to the fuel model.

In the case of estimated electricity robust regression models, (see Table (7.60)), the AvHOURS has a significant coefficient in all kilns. PRORATE and AVL have significant coefficients in four kilns out of five. Aratio has a significant coefficient in one kiln out of three. AvNO is significant in two out of five kilns. Sratio, and LimSF have significant coefficients in only two kilns out of three. These observations reflect the relative importance of these independent variables in explaining the variability in EL.

In the case of estimated fuel robust regression model, (see Table (7.61)), PRORATE has a significant coefficients in four out of five kilns. AVL has significant coefficients in three kilns out of five. AvNO, AvHOURS, have significant coefficients in one out of five kilns. Aratio has significant coefficients in two kilns out of three. However, LimeSF has a significant coefficient in only one kiln out of three. Sratio is not significant in any kiln. These observations reflect the relative importance of these independent variables in explaining the variability in FUEL.

Third, Almost all the signs of the estimated coefficients of an independent variable, which are significant in both EL and FUEL robust models, are the same in both models for almost all kilns, (see Tables (7.60) and (7.61)). These results agree with the expectations of the researcher. The robust regression models are given in Table (7.65).

Fourth, It is interesting to note that, in the case of FUEL the stated researcher's expectations were confirmed for all variables and for all kilns except for the sign of AVL in kiln 4. However, in the case of EL these expectations were confirmed for all variables except the signs of AvNO and Sratio in kiln 6 and AvHOURS for kiln 1, (see Tables (7.60) and (7.61)).

Fifth, This research demonstrates in a statistical analysis, the significance of energy consumption (fuel and electricity) as a function of energy management, as well as of production. If the plant is subject to frequent stoppages, interruptions and start-ups, its energy consumption will increase significantly. The findings of the statistical analysis revealed that the relationship between energy consumption and stoppages is non-negative except in kiln6-EL. This means that, as both average number and duration of stoppages is increased; the consumption of both electricity and fuel will increase in most cases.

Sixth, The statistical evidence revealed by the regression analysis showed that the relationship between production rate and energy consumption per unit is negative whenever it is significant. This confirms the researcher's expectation that at higher levels of production rate, consumption of both electricity and fuel decreases per unit of cement output.

Production rate has exerted a negative impact on both electricity consumption and fuel consumption in almost all models fitted. This result is expected, since as we are approaching full utilisation of the kiln's resources, the optimal state of consumption of both electrical power and fuel is achieved, hence an optimal state of productivity is maintained. Another explanation of this result might be as follows: when production rate is high, this indicates that number of stoppages is low, hence saving those times of heating-up to restart again where consumption of both electricity and fuel are wasted for non-productive operations. The period of heating-up is normally 24 hours of continuous consumption of electrical power and fuel at zero production level.

Seventh, The relationship between energy consumption and availability is negative whenever it is significant, i.e. as availability increases energy consumption decreases. This result, also, confirms the researcher expectations.

Availability, as measured in this research, maintained in most of the cases a negative and significant impact on both electricity and fuel consumption. This relationship is also expected, since as percentage of actually utilised production capacity is increased because of higher availability, the levels of consumption of both electric power and fuel are decreased. This result is extremely important because it clearly indicates to executives the need to maintain the highest level of availability in order to accomplish the goal of reducing the consumption of electrical power and fuel.

Eighth, The statistical evidence revealed by the regression analysis showed that the relationship between Aratio and energy consumption per unit is negative whenever it is significant. This confirms the researcher's expectation that at higher levels of Aratio, consumption of both electricity and fuel decreases per unit of cement output.

Experiments of the production and quality control departments have proven that increasing alumina ratio gives better burning conditions inside kilns, due to the creation of a greater liquid phase with the least possible temperature.

Ninth, The statistical evidence revealed by the regression analysis showed that the relationship between LimeSF and energy consumption per unit is positive whenever it is significant. This confirms the researcher's expectation that at higher levels of LimeSF, consumption of both electricity and fuel increases per unit of cement output.

The higher the Lime SF the more heat is needed for clinker burning because it becomes hard to burn and forms dusty clinker containing free lime. Therefore, it is important to have an optimum value for this factor according to specifications and operating conditions.

Tenth, The statistical evidence revealed by the regression analysis showed that the signs of Sratio were positive or not significant except for kiln 5. This support the researcher's expectation that at higher levels of silica ratio the burnability of the clinker becomes harder because of reducing the liquid phase content and tendency toward formation of coating in the kiln. An increased silica ratio also causes a slow setting and hardening of the cement while a decreased ratio increases the content of liquid phase thus improving the burnability of the clinker.

Eleventh, the R-squared values of electricity models are all very high, which means that the models can explain from 82% to 94% of the variability in EL using the independent variables in the models. On the other hand, the R-squared values of fuel models indicate that the models can explain from 57% to 89% of the variability in FUEL using the independent variables in the models, (see e.g. Tables (7.64) and (7.65)). It seems reasonable that the R-Square values for the FUEL models are smaller than the corresponding values of the EL models. This difference can be explained by the fact that the burning process in the kiln is very complicated process affected by many factors while the factors affecting electricity consumption are limited mainly due to the factors related to start up or shut down to the electric motors.

Twelfth, The statistical analysis of raw mills and cement mills demonstrate the usefulness of the independent variables in explaining the variability in electrical energy consumption as demonstrated by the R^2 results that ranges from 0.64 to 0.97. Moreover, the R-Squared values for all mills range between 0.78 and 0.97 except for R-rm2, (see Table (7.67d)). This indicates that the regression models have useful capability to explain the variability in EL using the available independent variables.

The results of the mills confirm the expectations of the research; i.e. in almost all cases, the increase in number of stoppages and duration of stoppages affect positively the electricity consumption, whereas production rate and availability affect electricity consumption negatively. That is to say, the production rate and availability showed the same decreasing relationship with electricity consumption. That is if these variables increase, the energy consumption will decrease. On the other hand, the duration of stoppages and number of stoppages showed the same increasing relationship with electricity consumption. i.e. if these variables increase then electricity energy consumption will increase.

Thirteenth, in general, the statistical models revealed that there is a strong relationship between energy consumption and the following independent variables: average number of stoppages, average duration of stoppages, production rate, availability, Alumina Ratio (Aratio), Silica Ratio (Sratio), and Lime Saturation Factor (LimeSF). This strong relation is shown in the values of R^2 of the models. On the other hand, the p-values for the F-test show that the models are significant. This demonstrates that the models are useful and the selection of the independent variables was successful. This encourages further research to include other independent variables to reach more powerful models.

Fourteenth, To create benchmarking with reputable cement manufacturers, the results of the statistical analysis, which was carried out for around 30 kilns distributed all over the world, revealed values of R^2 close to those obtained from the Jordanian data. Moreover, the signs of coefficients of the independent variables in the estimated regression model agree with the researcher's expectations, which have almost been confirmed in the Jordanian data. It is also worth noticing that the findings of the reputable manufacturer's data are very close to our research results concerning the fuel consumption.

This means that our selected control variables have comparable effect on the cement industry in general, which asserts positively the choice of the model and its use in the Industry. Moreover, this supports the findings of the research that the variables included are significant.

Fifteenth, In spite the fact that we checked statistically for any effect of rounding some limited parts of the data (as recorder by the factory operators) and the obtained findings that indicated there is a negligible effect of rounding on the obtained results, we recommend that,

i) The Clerks who report the data should have training sessions.

ii) The data on all variables should be reported daily to as much accurate digits as possible.

iii) In case a suspicious value is reported it should be checked for possible error in reporting it. If not, a remark should be given which explains the reason(s) of this value being out of range.

iv) There is a need to report data about all variables that are included in the manufacturing process.

Sixteenth, The power of the explanation (R2) in the kiln electrical consumption models ranges between 81%-94% which is considered to be high. This could be explained due to the fact that the model is handling the electricity consumption where the factors affecting this consumption considered to be relatively limited because most of the electrical machines are constant speed machines. In case of unsteady operational condition the production rate will be decreased and this resulted in increase of the average electricity consumption, also in case of repeated emergency stoppages this will decrease the availability of the production line and resulted in additional energy consumption at no load of the kiln, or during shutdown which will increase the average electricity consumption of the product

Seventeenth, The explanation power (R2) in the kiln fuel consumption models ranges between 57%-89% which is considered to be strong and acceptable but we noticed that it is lower than the power of explanation of the electrical model. This can be explained due to the fact that the burning process in the kiln is a very complicated process performed in several stages with several operating conditions (temperature, pressure, chemical reactions, cooling and heating process, variation of the quality of the raw materials ...etc, which means that the operational conditions can be affected by several factors and needs very special skills to keep it in steady conditions. The variation in operational conditions will mainly affect the production rate of the kiln causing a serious variation on the fuel consumption. Emergency breakdowns resulted in lower availability will cause increase of the average fuel consumption in case the kiln at no load condition. Also it will cause additional fuel consumption in case of heating up the kiln uptill reaching steady state operational conditions. To improve the explanation power of the model it is recommended to investigate additional variables, which may affect the fuel consumption, this can be done in further research.

7.8 Economic Modelling and Analysis

7.8.1 Prelude

The previous statistical analysis has established a statistically significant set of models for the relationships between energy consumption on one hand (measured in kWh/tonne of product or kilogram of fuel/tonne of product) and control variables on the other hand, at the Rashadiya and Fuhais plants in Jordan Cement Factories.

The control variables (independent variables) that have been considered are availability, production rate, average number of stoppages, and average duration of stoppage, lime saturation factor, silica ratio and Alumina ratio.

Using the statistical analysis, the researcher has identified the various sources of operational inefficiencies that can cause a decreased return to scale and an increase in operating cost. Thus, improving the values of these variables will lead to high production volumes (i.e., q according to section 7.1.1) which stabilising the returns to scale. These inefficiencies will cause an increase in the energy consumption and the production cost in general.

Based on the results of previous analysis, performed using the data collected from the factory during the period 1990-1993, several management initiatives were launched to improve energy consumption through controlling (i.e., improving) the values of the control variables. For examples, effective maintenance management and effective planning in particular have been used to reduce machine breakdowns, duration of breakdown time, and consequently increase production line availability. However, the improvement of these variables has its own associated costs, manifested in intensive technical and management staff training, inventory of spare parts, tighter quality control, and the like.

7.8.2 Cost of Control Factors Improvement

As mentioned above there is a cost involved for the improvement of the control factors (independent variables). For the sake of establishing a preliminary economical modelling and analysis, availability effect will be discussed as an example as it is similar in treatment to the effect of other control variables. High availability implies sufficient control of raw material preparation, high quality equipment operation and maintenance, and good house keeping. Since, in cement factories, production is performed in a series of sequential and interrelated steps, sophisticated management techniques are needed to ensure high levels of availability.

Planning is the first step towards availability improvement. Planning includes the preparation of production line setup manuals, production manuals, maintenance manuals, managing inventory of spare parts, record keeping and data analysis, etc. The implementation of the production operations together with effective maintenance of equipment should follow strict and preplanned schedules. During the implementation phase, one should not underestimate the importance of quality control functions on every single step that needs to be done in the factory. Quality control includes inspecting all work procedures and work instructions, as well as auditing the realisations of these procedures. Quality control also includes the quality of raw materials and semi-finished materials moving between production stations. These are just some of the factors that management has considered to improve availability. These plans and actions require two types of effort; one that is cost-free, and the other based on dedication of more resources. Cost free initiatives are those associated with good house keeping and improving employee loyalty. On the other hand, quality initiatives may require more human resources or physical resources such as additional inspection and testing equipment and the like.

7.8.3 Objective of the Economic Analysis

Since energy consumption improvement has incurred additional costs associated with improving availability, as an example, the objective of the following preliminary rational economic analysis is to determine the extent or the feasibility of the introduced management principles on the overall economic performance of the Jordan Cement Factories. That is, the objective of the preliminary economic analysis presented in the following discussion is to demonstrate the economic implications of improving the values of the selected control variables through developing a relationship between the returns of energy consumption saving (i.e., savings in JD/tonne) and the cost of improving availability, as an example. To develop such a relationship, the author has capitalised on the statistical relationships that were derived earlier. This effort will provide an estimate of the per unit savings (in JD/tonne) in energy consumption resulting from the per unit improvement in the availability of the production line. The fundamental question of the impact of the combined selected control variables on the net JD value of energy savings requires extensive economic analysis and modelling. However, the author is seeking to outline a preliminary methodology for economic analysis aimed at demonstrating the role of the statistically derived relationships between energy consumption and the control variables. The statistical model is used as a basis since it has explained about 80% of the variations in electrical energy consumption and about 80% of the fuel energy consumption. The methodology will be illustrated using estimated data collected from the production lines under study.

7.8.4 Preliminary Investigation

The new management initiatives were launched based on the assumption that energy consumption can be reduced through maximising plant availability and production rates while decreasing the number, duration of stoppages and optimising the quality control factors. Emphasis was placed on quality planning, employee training and empowerment, accurate costing, statistical analysis techniques, and establishment of quantitative performance indicators, etc. To provide a preliminary estimate of the improvement associated with availability and other independent variables improvement, data were collected for the years 1995 to 1998 for the cost of electrical energy per tonne cement, and availability. Table (7.72) shows the trends of improvement at the Rashadiya plant.

As shown in table (7.72) one can observe the relationship between availability and energy consumption. As availability increases the energy consumption per tonne of cement decreases and consequently the cost of energy per tonne decreases too. This result is not unexpected as the statistical models developed earlier demonstrate an inverse linear relationship between availability and energy consumption. The data given in the table even suggested a strong relationship between availability and energy cost. Thus, an economic model can be derived between the cost of energy consumption rate (JD/tonne) and availability to achieve one of the objectives of the research mentioned in section 7.2.1 which is to build a procedural method, which may help in planning the expected budget needed for the expenses of the dependent variables, based on the levels of the independent variables. That is to find a regression model that expresses each of EL and FUEL as a function of the influential independent variables that may affect the cost of the consumption of energy.

Table (7.72)

Relationship between Availability and Energy Consumption and Cost

Year	Energy consumption (kwh/tonne)	Energy cost (JD/tonne) *	Availability (%)
1995	111.22	4.516	89
1996	106.56	3.548	90
1997	105.24	3.440	91
1998	102.00	3.130	95

Assuming fixed prices for unit energy for applicable consumption ranges.
Source: Jordan Cement Factories (1999)

Having developed the objective and background of the preliminary economic analysis in this research, we present in the next sections the assumptions underlying the economic modelling process, the modelling methodology, and finally the obtained results.

7.8.5 Economic Model Assumptions

As explained before, the economic analysis is based on the statistical models that were developed earlier. In this context, the following assumptions are made:

1. The equations developed in the statistical analysis are directly used as a basis for evaluating the economic implications on energy consumption caused by the selected control variables. That is the effects of control variables are independent and are additive.
2. The overall economic effect on energy consumption is the sum of independent individual effects of the selected variables.
3. The model is based on evaluating the incremental (i.e., per unit) savings in energy usage resulting from the incremental improvements in the values of the selected variables.

4. The selected variables that have a statistically significant effect on energy consumption are Availability (AVL), Production Rate (PRORATE), Average no of Stoppages (AvNO), Average Duration of Stoppage (AvHOURS), Lime saturation factor (LimeSF), Alumina ratio (Aratio), Silica ratio (Sratio). These are the only variables that are considered in the model. These variables explain about 80% of the electrical energy variation and 80% of the fuel energy variations.

5. The model considers the benefits gained as a result of improving control variables without considering the restored opportunity that would have been lost if availability were not improved.

6. The economic effect for the first four variables will be studied, since they are performance indicators, where LimeSF, Aratio and Sratio are process parameters and can be changed without any cost involved.

7.8.6 Economic Model Formulation

The general form of the energy consumption equation as given earlier is as follows:

A- Electric Energy:

$$EL = b_{0e} + b_{1e}AvNO + b_{2e}AvHOURS + b_{3e}PRORATE + b_{4e}AVL$$

B- Fuel Energy:

$$FUEL = b_{0f} + b_{1f}AvNO + b_{2f}AvHOURS + b_{3f}PRORATE + b_{4f}AVL$$

Model A is used for electrical energy and model B is used for fuel energy, and the coefficients in the equations represent the rate of change of the energy consumption with respect to the associated variables. For example, the rate of change of electrical energy consumption per unit of production with respect to production rate is

$$(\Delta EL / \Delta \; PRORATE)$$

And this equals (b_{3e}). (Since the relationship between energy consumption and production rate is strong and negative as shown by the statistical derivation, then (b_{3e}) will be a negative coefficient). This implies that:

$$\Delta \; EL \text{ (with respect to PRORATE)} = \; b_{3e} . \; \Delta \; PRORATE$$

If this concept of incremental or marginal change is applied to the above two equations, one can obtain the following total change in energy consumption (for the addressed production lines given earlier):

$$\Delta \; EL \quad = b_{1e} . \; \Delta \; AvNO + b_{2e} . \; \Delta \; AvHOURS \; + b_{3e} . \; \Delta \; PRORATE + b_{4e} . \; \Delta \; AVL \text{ (kwh/tonne)}$$

$$\Delta \; FUEL = b_{1f} . \; \Delta \; AvNO + b_{2f} . \; \Delta \; AvHOURS + b_{3f} . \; \Delta \; PRORATE + b_{4f} . \; \Delta \; AVL \text{ (kg fuel/tonne)}$$

These equations provide a tool for evaluating the economic implications of additional investment. Using the economic concept of the net return (i.e., profit or savings), one can construct a simple cost function based on the estimates of the per unit return of energy savings and cost of the quantities on right-hand side of the above equations. The net savings concept implies that net profit = total return - total expenditure. In the context of this problem at hand, the total income is the realisable income as a result of savings in the use of energy, while total expenditures represent the additional cost for improving the control variables.

If we denote the total return of electrical energy management by (R_{TE}), the total return of fuel energy management by (R_{TF}), the total cost of improvement in the control variables (Cimprove), the cost of improvement in the electrical and fuel energy management by (C_{TE}) and (C_{TF}), respectively and the net return (or Savings) by (S), then the following relationships are valid for the electrical and fuel energy savings, respectively.

$S_E = R_{TE} - C_{TE}$ (units are JD)..............................(1)

$S_F = R_{TF} - C_{TF}$ (units are JD)..............................(2)

To express the above equations in terms of the model parameters, one would need to define the following cost multipliers and quantities:

C_{AvNO} = cost of reducing AvNO by one unit in JD

$C_{AvHOURS}$ = cost of reducing AvHOURS by one unit in JD

$C_{PRORATE}$ = cost of increasing PRORATE by one unit in JD

C_{avl} = cost of increasing AVL by one unit in JD

C_{elect} = cost of one unit electrical energy in JD

C_{fuel} = cost of one unit of fuel energy in JD

Cimprove = total cost of improvement for all variables

P = annual production of cement in tonnes of a specific production line

Cimprove = $C_{TE} + C_{TF}$

For this preliminary derivation of the formulas, we shall assume that the cost of improvement is divided equally between C_{TE} and C_{TF}

Thus the total improvement cost in electric energy can be expressed using the following equation (3):

$$C_{TE} = 0.5 \; [\Delta AvNO.C_{AvNO} + \Delta AvHOURS.C_{AvHOURS} + \Delta PRORATE.C_{PRORATE} + \Delta AVL.C_{avl}]$$

The total improvement cost of the fuel energy can be expressed as follows in equation (4):

$$C_{TF} = 0.5 \,.\, [\Delta AvNO.C_{AvNO} + \Delta AvHOURS.C_{AvHOURS} + \Delta PRORATE.C_{PRORATE} + \Delta AVL.C_{avl}]$$

Now the total annual return resulting from saving in electrical energy consumption can be expressed as follows in equation (5):

$$R_{TE} = [b_{1e}.\, \Delta AvNO + b_{2e}.\, \Delta AvHOURS + b_{3e}.\, \Delta PRORATE + b_{4e}.\, \Delta AVL].C_{elect}\,.P$$

While total annual return in fuel energy is expressed as follows in equation (6):

$$R_{TF} = [b_{1f}.\, \Delta AvNO + b_{2f}.\, \Delta AvHOURS + b_{3f}.\, \Delta PRORATE + b_{4f}.\, \Delta AVL].C_{fuel}\,.P$$

Therefore, the net annual saving from electrical energy can be expressed as follows equation (7):

$$S_E = \{[b_{1e}.\ \Delta\ AvNO + b_{2e}.\ \Delta\ AvHOURS + b_{3e}.\ \Delta\ PRORATE + b_{4e}.\ \Delta\ AVL]\ .\ C_{elect}\ .P\ -\ 0.5[\ \Delta\ AvNO.C_{AvNO} + \Delta\ AvHOURS.C_{AvHOURS} + \Delta\ PRORATE.C_{PRORATE} + \Delta\ AVL.C_{avl}]\}$$

The net annual savings resulting from fuel energy can be expressed as follows in equation (8):

$$S_F = \{[b_{1f}.\ \Delta\ AvNO + b_{2f}.\ \Delta\ AvHOURS + b_{3f}.\ \Delta\ PRORATE + b_{4f}.\ \Delta\ AVL]\ .C_{fuel}\ .\ P\ -\ 0.5\ .\ [\ \Delta\ AvNO.C_{AvNO} + \Delta\ AvHOURS.C_{AvHOURS} + \Delta\ PRORATE.C_{PRORATE} + \Delta\ AVL.C_{avl}]\}$$

Once the cost parameters are estimated, one can compute the expected energy savings resulting from any level of improvement in the values of the control variables.

7.8.7 Methodology of Economic Analysis

The ultimate objective of the economic analysis is to provide deeper insight and understanding of the economic implications of management decisions. The application of the economic model requires a structured methodology that can provide accurate and effective economic information to enable the decision making process. The proposed methodology in this research consists of the logical phases of the management process. These phases are as follows:

Phase I: Data Acquisition

To implement the economic model the following data must be collected from the appropriate resources.

1. Determine the production line (or machine) that will be considered in the economic analysis process and whose statistical model has been derived.
2. Determine the actual cost of improvement in the operational efficiency of the selected production line. This data can be collected, for instance, from maintenance records.
3. Determine the level of improvement in the operational efficiency in terms of the selected control variables such as production rate or availability.
4. Assertion the market prices of the electrical and fuel energy.

Phase II: Computational Analysis

In this phase several computational steps are required to produce the following information:

1. The cost multipliers (C_{AvNO}, $C_{AvHOURS}$, $C_{PRORATE}$, C_{avl})
2. C_{TE}, C_{TF} and Cimprove
3. R_{TE} and R_{TF} (using C_{elect} and C_{fuel})
4. S_E and S_F

In essence, the model equations and parameters are used in this phase to compute the net savings resulting from improvements.

Phase III: Evaluation

Based on the results of the previous phase, management would evaluate the worth of any improvement effort in terms of the net savings. The evaluation process can be conducted to assess the economic impact of an already implemented improvement program, and/or to assess whether or not a proposed set of improvements is economically feasible prior to implementation.

7.8.8 Limitations of the Model

The model explained and the associated implementation methodology can provide a deeper understanding of the mechanics of the improvement process. Using this model, one can determine the extent to which management should seek further improvement in the production processes. The incremental per unit cost of improvement in each of the selected control variables depends on the values (i.e., ranges) of these variables. For example, at values of availability that are relatively low (e.g., 85%), the per unit cost of improvement in availability will be less than in the case when availability is relatively high (e.g., 95%). This implies that the incremental savings resulting of any improvement effort may not be economically justified, especially when the value of availability becomes closer to 100%. Therefore, the optimal level of improvement can be determined using an advanced economic analysis.

7.8.9 Verification and analysis of statistical and economical modelling

7.8.9.1 Statistical Model Verification

Before testing the preliminary economic model to arrive at potential savings in energy cost resulting from improvement in plant performance measured by the independent variables, the regression equations derived from the statistical analysis are to be verified and tested for its prediction power using more recent data. The regression equations were derived from data for the period 1990-1993. If one obtains acceptable results upon using more recent data for the independent variables to predict the dependent variables using the regression equations, then confidence in the regression equations increases, and this will be checked and demonstrated through case studies 1 and 2. These cases are presented in the following paragraphs.

Case study no. 1:

Verification of the Prediction Power of the Robust Model for Fuel of Kiln 5

To check the validity of the robust model for fuel for kiln no. 5 which its derived formula is the following:

$$Fuel = -99.77511 + 1.300123 AvHOURS - 0.5187489 AVL - 12.60869 Aratio + 2.727895\ LimeSF$$

this model (which has been derived using 1990-1993 data) was used to predict the fuel consumption of kiln 5 based on 1994 data that is given in Table (7.73). First we have tested for possible outliers in the 1994 data. The following list provides the results of this test. It shows that observation number 6 is an outlier in fuel data and observation number 5 is an outlier in AvHOURS data. After investigating the data, it was discovered that the fuel consumption (observation no. 6) was increased to 999 kcal/kg clinker while the average for the whole year was 942. For the average hours it revealed from the data that there were stoppages hours equal 358 (observation no. 5) while the average stoppages hours for the whole year was 115.8 hr.

Data Screening Report of Kiln 5
Outlier Section of FUEL

Row	T2 Value	T2 Prob	Outlier?
1	0.19	0.6687	
2	0.12	0.7317	
3	0.02	0.8974	
4	1.02	0.3336	
5	0.38	0.5507	
6	6.50	0.0270	Yes
7	0.07	0.7968	
8	0.02	0.8974	
9	2.23	0.1631	
10	0.09	0.7640	
11	0.33	0.5791	
12	0.02	0.8974	

Outlier Section of AvHOURS

Row	T2 Value	T2 Prob	Outlier?
1	0.95	0.3518	
2	0.05	0.8294	
3	0.11	0.7511	
4	0.25	0.6272	
5	5.75	0.0354	Yes
6	0.10	0.7525	
7	0.64	0.4406	
8	0.51	0.4904	
9	0.48	0.5012	
10	0.52	0.4863	
11	1.07	0.3229	
12	0.58	0.4631	

It should be mentioned that the fuel in table (7.73) is reported in (kcal/ kg clinker), to convert it to (kg fuel / tonne clinker) each value in the fuel column was divided by the convergence factor 9.65, which represents the calorific value for the fuel (9650 kcal/kg fuel).

Table(7.73)

The 1994 Data of Kiln 5

FUEL	AvNO	AvHOURS	PRORATE	AVAL	Aratio	Sratio	LimSF
952	0.290323	6.93548	59.96	94	1.61	2.45	95.8
935	0.25	3.07143	61.61	96	1.58	2.45	96.2
940	0.451613	4.83871	59.8	80	1.55	2.48	95.4
965	0.666667	2.16667	56.1	91	1.58	2.47	96.6
929	0.193548	11.5484	59.22	99	1.63	2.45	95.9
999	0.733333	4.83333	58.84	80	1.6	2.47	96.7
937	0.516129	1.19355	59.5	95	1.59	2.46	96.7
940	0.387097	6.09677	57.17	95	1.54	2.48	96.4
910	0.566667	1.53333	58.9	94	1.56	2.49	96.4
936	0.387097	1.45161	56.47	94	1.49	2.49	97.5
930	0.4	0.433333	57.35	98	1.61	2.46	96.6
940	0.516129	1.32258	56.68	96	1.57	2.46	97.2

Source: Fuhais plant-Planning department-Annual report 1994

Table (7.74) reports the results of predicted and the observed values of fuel based on the final robust model together with percentage relative errors in predictions, standard errors in point predicted values, and a 95% confidence prediction intervals for the fuel of each of the 12 months of 1994.

It is seen from Table (7.74) that

1) The observed mean of fuel is 97.69 while the mean of the predicted fuel is 100.312.

2) The relative percentage error in predicting the mean fuel consumption for the year 1994 is -2.67944%, which seems to be reasonable percentage error.

3) The limits of the 95% prediction confidence intervals (band) of the predicted values are close enough to the observed values. This means that our model is good to predict the fuel consumption.

4) If the two outliers are ignored, there are 7 negative percentage relative errors and 3 positive percentage relative errors, which seems to be a reasonable distribution of errors around zero. But this may indicates that the kiln in 1994 was operating slightly better than in the period 1990-1993.

5) If the two outliers are ignored, there is only one observed value of fuel outside the obtained prediction intervals. This can be seen from the plot given bellow, which presents the observed fuel values in red and the predicted fuel values in blue.

6) In general we can say that the model seems to be useful to do predictions.

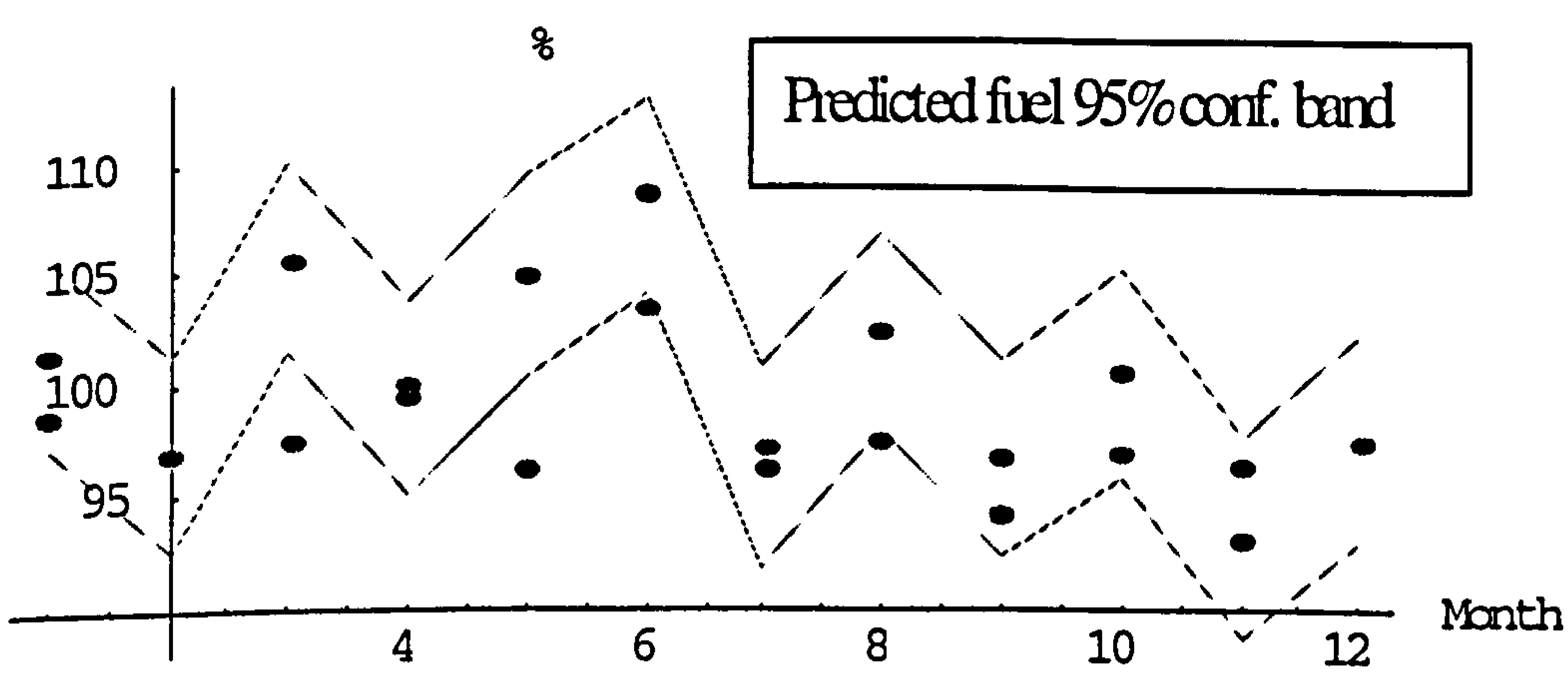
Table(7.74)

Predicted Fuel consumption of kiln 5

Observed mean of fuel is 97.6943
Predicted mean of fuel is 100.312
Percentage relative error in predicting the fuel mean is -2.67944

Predicted fuel consumption of kiln 5 based on our the robust model

Observed	Predicted	% error	Standard Error	95% Prediction Intervals Lower Limit	Upper Limit
98.6528	101.512	-2.89802	4.22759	97.2842	105.739
96.8912	96.92	-0.0297319	4.30401	92.616	101.224
97.4093	105.714	-8.52514	4.3385	101.375	110.052
100.	99.4286	0.5714	4.33306	95.0955	103.762
96.2694	104.936	-9.00245	4.5383	100.398	109.474
103.523	108.622	-4.92559	4.44582	104.177	113.068
97.0984	96.2351	0.88911	4.41175	91.8234	100.647
97.4093	102.422	-5.14599	4.36579	98.0562	106.788
94.3005	96.7555	-2.6034	4.33776	92.4178	101.093
96.9948	100.533	-3.64737	4.68168	95.8509	105.214
96.3731	93.1656	3.32822	4.45707	88.7085	97.6226
97.4093	97.5003	-0.0933559	4.5764	92.9239	102.077



Case study 2: Verification of the Prediction Power of the Robust Model for Fuel of Kiln 6

To check the validity of the robust model for FUEL of kiln 6 which its formula is the following

Fuel = 170.4401-0.4811263 PRORATE-0.314158 AVL

this model (which has been derived using 1990-1993 data) was used to predict the fuel consumption of kiln 6 based on 1994 data that is given in Table (7.75). First we have tested for possible outliers in the 1994 data. The following list provides the results of this test. It shows that observation number 2 is an outlier in each of FUEL, AvHOURS, and LimeSF data, while observation number 10 is an outlier in AvNO data.

Data Screening Report of Kiln 6

Outlier Section of FUEL

Row	T2 Value	T2 Prob.	Outlier?
1	0.89	0.3650	
2	7.62	0.0185	Yes
3	0.07	0.8005	
4	0.00	0.9554	
5	0.01	0.9188	
6	0.26	0.6216	
7	0.22	0.6493	
8	0.40	0.5423	
9	0.40	0.5423	
10	0.69	0.4239	
11	0.05	0.8259	
12	0.40	0.5423	

Outlier Section of AvHOURS

Row	T2 Value	T2 Prob.	Outlier?
1	0.00	0.9853	
2	7.69	0.0181	Yes
3	0.01	0.9116	

4	0.09	0.7712
5	0.24	0.6342
6	0.19	0.6684
7	0.27	0.6113
8	0.33	0.5760
9	1.41	0.2599
10	0.24	0.6339
11	0.30	0.5922
12	0.21	0.6541

Outlier Section of LimeSF

Row	T2 Value	T2 Prob.	Outlier?
1	0.00	0.9848	
2	5.81	0.0346	Yes
3	0.60	0.4534	
4	0.24	0.6366	
5	0.07	0.7906	
6	1.31	0.2759	
7	0.94	0.3521	
8	1.06	0.3252	
9	0.32	0.5844	
10	0.20	0.6636	
11	0.01	0.9394	
12	0.44	0.5224	

So, observation 2 is outlier.

Table (7.75)

Kiln 6

The 1994 data of kiln 6

Fuel	AvNO	AvHOURS	PRORATE	AVL	Sratio	Aratio	LimeSF
879.	0.354839	4.03226	116.98	91	1.67	2.43	93.4
924.	0.214286	20.2857	112.89	97	1.69	2.4	97.5
862.	0.612903	3.48065	113.65	87	1.63	2.45	94.7
857.	0.566667	2.40667	112.48	90	1.65	2.46	94.2
853.	0.258065	1.29355	111.92	95	1.69	2.45	92.9
843.	0.3	1.58	114.43	96	1.71	2.48	91.4
844.	0.322581	1.09677	115.5	96	1.75	2.43	91.7
840.	0.290323	0.787097	116.17	97	1.74	2.43	91.6
840.	0.366667	11.0567	116.89	90	1.72	2.45	92.4
835.	0.322581	1.29032	114.82	95	1.68	2.46	92.6
850.	0.466667	0.93	113.58	96	1.75	2.4	93.5
840.	0.193548	1.46129	114.53	94	1.72	2.48	94.5

It should be mentioned that the fuel in table (7.75) is reported in (kcal/ kg clinker), to convert it to (kg fuel / tonne clinker) each value in the fuel column was divided by the convergence factor 9.65, which represents the calorific value for the fuel (9650 kcal/kg fuel).

Table (7.76) reports the results of predicted and the observed values of fuel based on our final robust model together with percentage relative errors in predictions, standard errors in point predicted values, and a 95% confidence prediction intervals for the fuel of each of the 12 months of 1994.

It is seen from Table (7.76) that

1) Observed mean of fuel is 88.6615, the mean of the predicted fuel is 85.9314

2) The relative percentage error in predicting the mean fuel consumption for the year 1994 is 3.0792%, which seems to be reasonable percentage error.

3) The limits of the 95% prediction confidence intervals (band) of the predicted values are close enough to the observed values. This means that our model is good enough to predict the fuel consumption.

4) All percentage relative errors are positive, which means that the model under predicts the fuel for 1994. However, the maximum percentage relative error in prediction is 6.06% provided that the outlier in the fuel data is ignored. This may suggest that the kiln 6 was operating less well than in the ealier period and may suggest less effective operation during 1994. To check this finding we compare the average fuel consumption of the period 1990-1993 which was 831 kcal/kg clinker, with the average consumption in the 1993 which was 850 kcal/kg clinker.

5) If the outlier in the fuel data is ignored, there is only one observed values of fuel outside the obtained prediction intervals. This can be seen from the plot given

bellow, which presents the observed fuel values in red and the predicted fuel values in blue.

6) In general we can say that the model seems to be useful to do predictions.

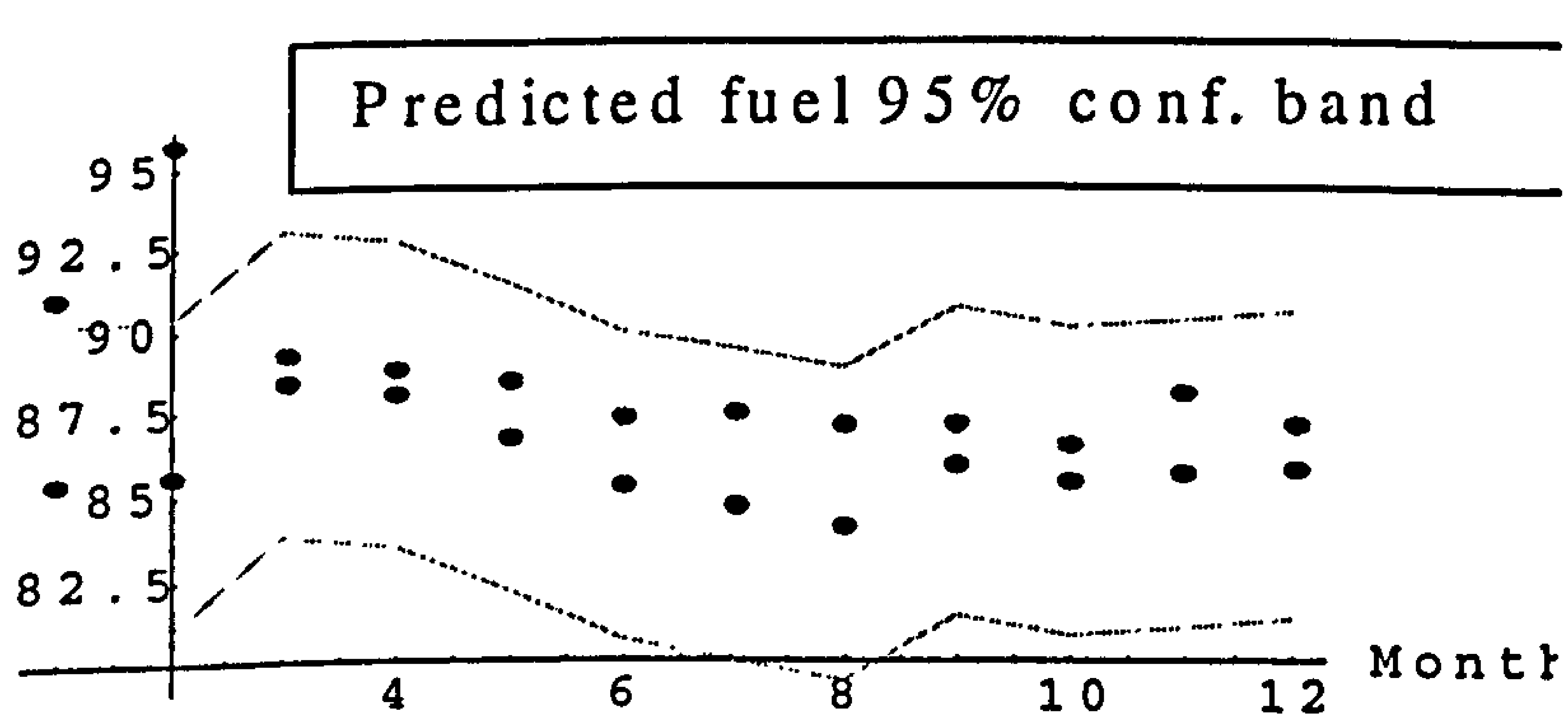
Table(7.76)

Predicted Fuel consumption of kiln 6

Observed mean of fuel is 88.6615
Predicted mean of fuel is 85.9314
Percentage relative error in predicting the fuel mean is 3.0792

Predicted fuel consumption of kiln 6 based on our the robust model

				95% Prediction Intervals	
Observed	Predicted	% error	Standard Error	Lower Limit	Upper Limit
91.0881	85.5696	6.05844	4.59771	80.9719	90.1673
95.7513	85.6524	10.547	4.63714	81.0153	90.2896
89.3264	88.4284	1.00539	4.60873	83.8196	93.0371
88.8083	88.0488	0.855209	4.59421	83.4546	92.643
88.3938	86.7474	1.86252	4.61681	82.1306	91.3642
87.3575	85.2256	2.44039	4.62078	80.6049	89.8464
87.4611	84.7108	3.14459	4.62167	80.0892	89.3325
87.0466	84.0743	3.41461	4.63651	79.4378	88.7108
87.0466	85.927	1.28621	4.59857	81.3285	90.5256
86.5285	85.3522	1.35947	4.60995	80.7422	89.9621
88.0829	85.6346	2.77953	4.62177	81.0128	90.2564
87.0466	85.8059	1.42542	4.60132	81.2045	90.4072



7.8.9.2 Results of the economic analysis

In the following case study no. 3, the preliminary economic model, which is essentially, built on the regression equations, is used to estimate the savings in energy cost. However, these savings are calculated for two cases: the first is to compare with the least energy consumption within a given year, the second is to compare with the optimal energy consumption. That is to say that in both cases actual consumption is compared with "would be" consumption, either the least consumption in a given year or the optimal consumption.

Case study 3 Total Potential Savings in Energy costs using the economical model

This case study assesses the possible value of total savings in energy costs for all Kilns under two theoretical cases;

1) Taking the minimum monthly electricity and fuel consumption (during 1990-1993), as the average possible achievable consumption for all months of the year 1993.
2) Taking an optimum electricity and fuel consumption, as the average possible achievable consumption of the year 1993.

The savings would result from comparing the actual electricity and fuel consumption (using 4 variables) by the values of the two cases mentioned above. Moreover, the following considerations are taken:

- Fuel cost = 70 JD/tonne
- Electricity cost = 0.044 JD/kWh

- Production capacity is as follows

  kiln 4 : 250,000 tonne/yr.

  Kiln 5 : 500,000 tonne/yr.

  Kilns 1,2, 6: 900,000 tonne/yr.

Table (7.77) shows the total possible savings resulting from this case study.

Table (7.77)
Total Savings in Energy Cost Resulting from Minimising and Optimising Consumption

	Kiln 4		Kiln 5		Kiln 6		Kiln 1		Kiln 2	
	FU	EL	FU	EL	FU	EL	FU	EL	FU	EL
Actual	93.75	28.25	99.33	34	91	29.18	88.25	20.33	87.67	19.21
Minimum	87.0	21.0	87.0	29.0	80.0	22.0	81.0	18.2	83.0	16.8
Difference	6.75	7.25	12.	5.0	11.0	7.18	7.25	2.13	4.67	2.41
JD/Yr.	118125	79750	431667	110000	693000	284400	456750	84480	294000	95370
Total saving	2,647,542 JD									
Optimum	78.0	20.	78.0	20.0	78.0	20.0	78.0	15.0	78.0	15.0
Difference	15.76	8.25	21.33	14.0	13.0	9.18	10.26	5.33	9.67	4.21
JD/Yr.	275625	90750	746667	308000	819000	363600	645750	211200	609000	16665
Total saving	4,236,242 JD									

- Actual values for each kiln (FUEL, EL) were taken from planning department records.
- Minimum value is taken as the minimum value of the variables during (1990-1993).
- Optimum value is the optimum value it can be got within available modern technology, but it may need additional investment and/or technical modification.
- Difference = Actual – minimum or (optimum)
- Saving fuel (JD/Yr) = difference (Kg fuel/tonne clinker) * production capacity (tonne clinker /Yr)*0.07 (JD/Kg fuel)

- Saving electricity (JD/Yr) = difference (kWh/tonne clinker) * production capacity (tonne clinker /Yr)*0.044 (JD/kWh)

Table (7.77) shows that the total possible savings in the cost of energy consumption, in one year, reaches about JD 2.6 million if one substitutes the minimum consumption value during the period 1990-1993 in place of the actual value. On the other hand, if one substitutes the optimal consumption in place of the actual value one gets a possible total saving of about JD 4.2 million, but it may need additional investment and/or technical modification. These figures show the extent of energy savings possible in JCF if a proper energy management system is put in place.

7.8.9.3 Direct Application of the Economic Model

The application of the model is presented here for illustrative purposes and is not meant to be comprehensive or to cover the economic impact of all control variables at JCF facilities. For this purpose, the economic aspect of the effect of availability on electric energy savings at Kiln 6 of the Fuhais plant has been taken as an illustrative vehicle.

At the Fuhais plant Kiln6, the electrical energy consumption equation:

EL = -46.12966 - 6.216772*AvNO + .3714466*AvHOURS-1846394

*PRORATE - 0.164175*AVL + 8.667938*Sratio+ .962311*LimeSF

As mentioned earlier, the estimate of per unit costs of improvement of the control variables requires detailed and carefully planned and gathered data from accounting files. This data include the overall cost of improvement and the achieved levels of improvement in terms of the values of the control variables. Furthermore, one would need to allocate the total improvement cost to the control variables (according to a selected criteria) in order to compute the cost of per unit improvement for each control variable.

Thus, since at the present time no sufficient historical data are available to serve this purpose, we will resort to estimating these cost coefficients based on the experience and the data available for the years 1997-1998. In the year 1998, about 100,000 JD was spent to improve the overall operational efficiency of the plants in support of already undertaken housekeeping measures. This improvement scheme involved a diversified set of activities and work aspects including testing and inspection equipment and procedures, employee training, incentive plans, spare parts, etc.

The estimated levels of improvement of the various factors affecting energy consumption measured as a difference between 1997 and 1998 are as follows. Availability changed from 91% to %95, Stoppage Hours from 3257.0 to 2825.0, Number of Stoppages from (251) to (178); and finally Production Rate improved from 130.1 tonne/hour in 1997 to 135.7 tonne/hour in 1998.

The activities and procedures implemented to improve operational efficiency have collectively resulted in improving the four control variables under consideration. Thus, the accurate allocation of the improvement cost to certain activities and procedures associated with improving any individual control variable is difficult at this stage, as the data required to support these computations do not exist. Because of this difficulty, we will assume the totality of activities and procedures that were implemented in 1998 has actually improved 'availability'. Therefore, the cost per unit improvement in availability can be computed as total cost of improvement divided by the total change in availability.

The percentage of the overall cost spent for improving availability at Kiln6 was estimated at 5%, which is equivalent to 5,000 JD. The 5% was roughly approximated using process flow analysis that allocates general expenditures to workstations according to production time of each process, number of processes and human resources involved. Thus, using the data from 1997 and 1998 for Kiln6, the per unit cost of availability (measured in percent) becomes 5,000 JD/(95%-91%) = 1250 JD, implying that C_{avl} equals 1250.00 JD for Kiln6 at Fuhais plant. Similar calculations can be made for the other control variables once their cost allocation percentages become known.

Using equation 7 for S_E and assuming that availability is the only control variable under consideration, one can compute the net annual savings of electrical energy for Kiln6 during the year 1998. Using C_{elect} at a value of (0.044) JD per kWh and an annual volume of production of Kiln6 at 900,000 tonnes, one can apply equation 7 to compute the net annual 1998 savings. Using equation 5 for R_{TE}, the value of the overall return becomes 0.164x4x0.044x900,000 which is 25,977 JD per year. According to equation 3, the electricity improvement cost becomes (5,000.00/2)= JD2500 per year. This implies a net saving (return-cost) of about 23,477 JD per year, for Kiln 6. Similarly, the cost of improving other control variables can be computed thus leading to compute the total savings using the economic model.

It should be noted that the allocation of the overall improvement cost to individual activities resulting in improving certain control variables, and its impact on the economic model is recommended for future research, for which data collection procedures would be tailored to meet this objective.

7.8.10 Remarks on the Economic Model

The previous section has discussed the structure and the results of economic evaluation of implementing a management improvement scheme at the Jordan Cement Factories. The management improvement scheme utilises a set of control factors as key indicators. The preliminary economic analysis was intended to demonstrate a basic management principle, that is, to measure the performance of the management system quantitatively using indicators that make business sense, for instance, the net savings achieved as a result of management decisions. However, the economic analysis did not address the various aspects of the problem in terms of the economic and financial circumstances of JCF. That would require a detailed and extensive basic research that will be recommended for future studies. In general, the following summary remarks can be made on the economic modelling scheme.

1. The model provides the JCF management with a proper tool for approximating the economic impact of any intended (or existing) improvement plan.
2. The model is based on the incremental cost of improvement vs. the incremental savings resulting from improvement. This approach requires careful data collection and analysis. The factors C_{AvNO}, $C_{AvHOURS}$, $C_{PRORATE}$, and C_{avI} should be computed first prior to economic appraisal. These factors are assumed to be constant in a range of values of the associated variables. Even if these factors are not constant for the full range of the related variables, they can be assumed constant during piecewise ranges of the variables.

3. Any intended improvement scheme will not only improve the values of the selected control variables, but also will improve other control variables as well. This implies that the net savings realisable from any improvement effort will exceed the theoretically calculated values using the given model equations. This is because the control variables explain about 80 % of the energy variations. Other control factors that explain about 20% of the energy variations will become improved while their effect is not included in the economic model. There are even intangible benefits of any improvement process that will add more value to the operations effectiveness, for instance, the increase in the level of confidence and loyalty among employees and supervisors. The model, however, provides basic information to the decision-maker on the effectiveness of any course of action.

4. The model did not address the issue of lost opportunity due to losses in production as a result of stoppages, reduced availability and low production rate. The improvement of these factors will have huge additional financial gains, which may exceed by far the savings in electricity and fuel cost. This issue is recommended for further research.

References

Abraham, B. and Box, G.E.P. (1978), Linear models and spurious observations, Appl. Statist., 27, 131-138.

Andrews, D. F. (1947), A robust method for multiple linear regression, Technometrics, 16, 523-531.

Anscombe, F.J. (1973), Graphics in statistical analysis, American Statistician, 27, 17-21.

Atkinson, A.C . (1985), Plots, Transformations, and Regression, Oxford University Press, Oxford.

Atkinson, A.C., and Riani, M. (2000), Robust Diagnostic Regression Analysis, Springer.

Bhattacharyya, G. K. and Johnson, R. A.(1977), Statistical concepts and methods, Wiley.

Belsley, Kuh, and Welsch . (1980), Regression Diagnostics, John Wiley & Sons. New York.

Bendel, R.B. and Afifi, A.A. (1977), Comparison of stopping rules in forward stepwise regression, J. Amer. Statist. Assoc., 72, 46-53.

Berenson, M. L., Levine, D. M., and Goldstein, M. (1983), Intermediate Statistical Methods and Applications: A Computer Package Approach, Prentice-Hall, New Jersey.

Chatterjee and Price (1979), Regression Analysis by Example, John Wiley & Sons. New York.

Cochran, W. G. (1938), The omission or addition of an independent variate in multiple linear regression, J. Roy. Statist. Soc., B-5,171-176.

Cook, D. and Weisberg, S . (1982), Residuals and Influence in Regression, Chapman and Hall. New York.

D'Agostino, R.B., Belanger, A., D'Agostino, R.B. Jr . (1990), A Suggestion for Using Powerful and Informative Tests of Normality, The American Statistician , November 1990, Volume 44 Number 4, pages 316-321.

Daniel, C. and Wood, F.S. (1980), Fitting equations to data: Computer analysis of multifactor data, 2nd ed., Wiley, N.Y.

Draper, N.R. and Smith, H . (1981), _Applied Regression Analysis - Second Edition_, John Wiley & Sons. New York, NY.

Farrell, M.J (1957), _The Measurement of Productive Efficiency_, Journal of Royal Statistical Society, Vol. 120, No. 3

Ferguson, T.S. (1961), _On the rejection of outliers_, Proc. Fourth Berkeley Symp., 1, 253-287.

Freesman, D., Pisani, R., Purues, R. and Adhikari, A. (1991), _Statistics, Second Edition_, Norton & Company, New York, U.S.A.

Glantz, S A and Slinker, B K, (1990), _Primer of Applied Regression and Analysis of Variance_, McGraw-Hill

Goldgerger, A.S. (1961), _Stepwise least-squares: residual analysis and specification_, J. Am. Statist. Assoc., 56, 998-1000.

Graybill, Franklin (1961), _An Introduction to Linear Statistical Models_, McGraw-Hill. New York, New York.

Hamilton, L. (1991), _Regression with Graphics: A Second Course in Applied Statistics_, Brooks/Cole Publishing Company. Pacific Grove, California.

Hawkins, D.M. (1973), _On the investigation of alternative regressions by principle component analysis_, Appl. Statist., 22, 275-286.

Hawkins, D.M., Bradu, D., and Kass, G.V. (1984), _Location of Several Outliers in Multiple Regression Data Using Elemental Sets_, Technomitrics, 26, 197-208.

Hocking, R.R. (1972), _Criteria for selection of a subset regression: which one should be used?_, Technometrics, 14, 967-970.

Horel, A. E. (1962), _Application of ridge analysis to regression problems_, Chem. Eng. Prog., 58, 54-59.

Hoerl, A.E. and Kennard, R.W. (1970), _Ridge Regression: Biased estimation for nonorthogonal problems_, Technometrics 12, 55-82

Hoerl, A.E. and Kennard R.W. (1976), _Ridge regression: Iterative estimation of the biasing parameter_, Communications in Statistics A5, 77-88

Kailash C. Kapuv and Leonard R. Lamberson (1996), _Mechanical design handbook_, McGraw-Hill

O'Leary, D.P. (1990), *Robust regression computation using iteratively reweighted least squares*, SIAM J. of Matrix Analysis and Applications, 11, 466-480.

Mallows, C. L., (1964), *Choosing variables in a linear regression: A graphical aid*, Presented at the Central Regional Meeting of the Institute of Mathematical Statistics, Manhattan, Kansas, May 7-9, 1964.

Mallows, C. L., (1966), *Choosing a subset regression*, presented at annual meeting of the American Statistical Association, Los Angeles, August 15-19, 1966.

Mallows, C. L. (1973), *Some comments on Cp*, Technometrics,15, 661-675.

Mikel P. Groover, *Automation, Production systems, and Computeach integrated manufacturing*, Prentice-Hall International.USA, 1987.

Montgomery, Douglas and Peck. (1992), *Introduction to Linear Regression Analysis*

Mosteller, F. and Tukey, J.W. (1977), *Data Analysis and Regression*, Addison-Wesley. Menlo Park, California.

Mukhopadhyay, P. (1996), *Mathematical Statistics*, New Centeral Book Agency (P) Ltd, Calcutta, India.

Myers, R.H. (1990), *Classical and Modern Regression with Applications*, PWS-Kent Publishing Company. Boston, Massachusetts.

NCSS, WWW.NCSS.COM, NC SS2000.

Netter, J. and Wasserman, W. (1974), *Applied Linear Statistical Models*, Richard D. Irwin, Inc. Homewood, Illinois, U.S.A.

Netter, J., Wasserman, W. and Kutner, M. (1985), *Applied Linear Statistical Models*, Second Edition, Irwin, Illinois, U.S.A.

Ostle, B. and Mensing, R. (1979), *Statistics in Research*, Fifth Edition, The Iowa State University Press, Ames, Iowa, U.S.A.

Ott and Lyman (1984), *An Introduction to Statistical Methods and Data Analysis*, Second Edition, Duxbury Press, Boston, U.S.A.

Ratkowsky, David A. (1989), *Handbook of Nonlinear Regression Models*, Marcel Dekker. New York.

Rawlings John O. (1988), *Applied Regression Analysis: A Research Tool*, Wadsworth. Belmont, California.

Rousseeuw, P.J. and Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, Wiley, N.Y.,USA

Ryan, Thomas P. (1997), *Modern Regression Methods*, John Wiley & Sons. New York.

Sakamoto,Y., (1986), *Akaike Information Criterion Statistics*, KTK, Scientific Publishers

Saxena, J.P. et al. (1995), *Energy Efficiency Through Technological Improvements*, *World Cement*, June 1995.

Sanders, D. (1995), *Statistics: A First Course*, Fifth Edition, McGraw- Hill, New York, U.S.A.

Schmee and Hahn. (November, 1979), *A Simple Method for Regression Analysis*, Technometrics, Volume 21, Number 4, pages 417-432

Younger, M.S. (1972), *Handbook for Linear Regression*, Duxbury press, Massachusetts.

Chapter Eight

Energy Management Modelling in the Cement Industry in Jordan

Contents

Chapter Eight

Energy Management Modelling in the Cement Industry in Jordan

8.1 Introduction

The preliminary empirical results obtained in chapter six and the qualitative and quantitative evidence resulting from the case studies illustrate, beyond doubt, the vital need for establishing a detailed Energy Management System (EMS), which will be presented in this Chapter. The modeling and analysis of chapter seven indicate areas of good practice or of optimum production conditions. This chapter addresses the means of implementation of good practice, and of energy conservation and efficiency measures, in a practical, day-to-day sense as well as in a long-term strategic sense.

As we mentioned in chapter four, one of the objectives and steps of the research is to develop an Energy Management System to reduce energy consumption while maximising the throughput of the industry. This system includes an organisational structure, job descriptions and work instructions for activities related to energy management. It also exploits the relationship between energy management and total quality management. In other words, it integrates the energy management system with the operational strategy of the cement plants.

The following sections are devoted to achieving this objective through concentrating on the managerial aspects of energy management in the cement industry, and therefore explores the importance of organisational/procedural aspects of energy management, whereas Chapter six addressed actual cases of energy management (engineering activities). The two chapters together address both aspects of energy management, i.e. technical (scientific/engineering) and managerial (organisational/ structural).

Management is defined as the judicious use of means to achieve an end; it is a process, which includes several key functions. Management scientists view these functions very widely. While Fayol singled out planning, organisation, command, co-ordination and control, Gulik developed these functions into a set known by the acronym, "POSDCORB", which stands for Planning, Organising, Staffing, Directing, Co-coordinating, Reporting and Budgeting (Huse, 1979) . In this study we shall focus on planning, execution, and control as key management functions with the understanding that "organising" is an overall function embedded in each of the other functions.

The purpose of management in any organisation is to achieve specific desired end results or goals within definite time spans. Inputs to a process are various types of resources such as raw materials, energy, capital equipment, technology and people. Management is also concerned with the people in a particular organisation who carry out this process. After determining the goals, resources needed, and resource allocations, managers determine specific function plans. These plans are then executed, after obtaining necessary approvals from top management. One of the major steps in the execution of plans is control, which is considered a basic management function. This involves a comparison between resource inputs, including energy actually used, and those inputs, which are planned or budgeted (Henry, 1980). In this case, if there are any deviations, then managers should question the issue, seeking an explanation for them, and act to modify procedures where appropriate

Energy Management is defined as the strategy of adjusting and optimising energy utilisation in order to reduce energy requirements per unit of output, while holding constant, or reducing, total costs of the system for the same output. An Energy Management System is a system encompassing procedures, resources and organisational aspects, all designed to make energy management activities as effective as possible.

Energy management has not received widespread special attention until recently, mainly due to the relatively low cost of energy. The increase in energy cost and its consumption, especially in the cement industry, gave rise to awareness of a significant potential for the reduction of production cost through effective energy management systems. Companies thus began establishing energy management systems in their plants in order to audit, control and save energy.

Energy management involves two main activities:

* Energy conservation measures, which represent the core of the EMS and the activity by which energy consumption per unit output and cost, can be reduced.

* Energy auditing which is the process by which the effectiveness of conservation can be gauged. In other words, auditing determines how effective (or cost-effective) the conservation measures are. In most instances the two activities are continuous, and their order is reversible, such that auditing leads to more conservation and further conservation requires more auditing.

Reay (1977) recommends that energy management should apply the same basic techniques, which are applied in assessing the relative merits of, for example, the options for capital expenditure as part of a plant expansion programme. As he points out, such decisions are normally taken at boardroom level, which is where the ultimate responsibility for, and direction of an energy management programme should also lie. Top-level commitment to energy management must be demonstrated; without it, the task of those responsible for implementing an energy management programme at plant level would be more difficult.

Of the various ways in which such a programme can be organised, the most common is to begin by appointing an energy manager, who should be given sufficient authority to enlist the support of personnel as required at most staff levels within the company. Reay (1977) points out that in a large organisation with factories located in several parts of the country (and overseas), the energy manager will probably advise on the appointment of assistants at each factory, and will co-ordinate and direct their activities, reporting as necessary to the board when significant capital expenditure on equipment is required. Ideally, the manager should be allocated financial resources, which he can use for good housekeeping, and other tasks as recommended by his assistants in the field. His budget should be sufficient to enable him, if necessary, to make use of outside resources, such as consultants who can assist in audits and more specialised system design.

The energy used by the various items of equipment will often be difficult to quantify at plant level. The total costs of inputs such as electricity, gas and water, will be known, but information will not be available on the particular efficiency of individual plant items. Such data must be collected and analysed as part of the energy audit, a process that may take several months. For example computer based building management systems (BMS) can be installed to sample temperature, occupancy levels, lighting etc of buildings and hence provide a control mechanism. Extension of this approach to industrial processes is gradually being applied to new installations, but retrofitting onto old installations is often not cost effective. The advantage of having a continuous monitoring facility such as a BMS is that data can be accessed and analysed in real time and controlling actions effected immediately.

Energy conservation and load management, with regard to electrical energy, are two important activities within any EMS. This is particularly true for electrical utilities, the contribution of which still cannot be quantified and valued. Nordel (1987) comments that utilities are increasingly focusing on conservation and load management rather than construction of ever more costly electric generating facilities. However, while methods for the evaluation of the economics of new plant construction are well established in the industry, the economics of investment in, and promotion of, conservation and load management actions are less well understood. (not so true now)

Nordel (1987) defines conservation as a demand-side load shape modification strategy in which the primary objective is the reduction in total energy consumption. Secondary effects will include changes to the load patterns, which usually result in changes to contributions to system peak.

He defines load management as a demand-side load shape modification strategy in which the primary objective is the shift of energy consumption pattern to reduce contribution to system peak.

Secondary effects will result in either increases or decreases in total energy consumption.

Naturally, the definitions and strategies given by Nordel are also applicable to other forms of energy with some modifications. Moreover, what is applicable for an electricity utility is also applicable for a large energy-intensive industry.

Energy management is actually the joint responsibility of the energy consumer and the energy supplier. This is expressed clearly for managing the electrical load by Salehfar and Patton (1991), who point out that technical, economic, environmental, political, and social considerations have in recent years prevented the electricity utilities from encouraging and serving uncontrolled growth. Increasingly, they are looking for ways to improve their operating efficiency/utilisation to reduce the amount of investment required for additional generation, transmission, and distribution. These methods form the set of functions known as load management (LM). There are three basic components of LM- direct load control (DLC), indirect load control, and thermal energy storage, which may be used individually or in combination to influence a utility's customer load and so fulfil the objectives of its LM programme. The most frequent objectives of LM programme, the authors claim, are to reduce system peak load, improve load factor, reduce system reserve requirements, reduce production and fuel costs, and improves generation system reliability performance.

Venkatesh and Chankong (1995), demonstrated the importance of developing models and tools for optimal energy management in an industrial /commercial entity, and argue that co-generation plants can be beneficial to industries having a continuous demand of electric power and steam. Alteration of the design and operation of an electric power station to co-generation (the production of both useful heat and work), they argued, improves energy utilisation.

The issue of energy management has been, since the oil shocks of the seventies and eighties, taken to the extreme in the case of the electric utilities. This is expressed by Ilic et al. (1993), who propose a model that attempts to quantify trade-off between economics and system security. This issue is very important for large consumers, such as the cement industry, which have their own in-plant generation and are aware of the economics of energy management.

Co-operation and interaction among electric utilities and large consumers are advocated and encouraged. Rahman and Baba (1989) designed a simulator that integrates load forecasting and load management functions whereby the large industrial/commercial consumers receive advance notice on electricity tariff, which is dynamically adjusted by the utility based on the forecast.

Another aspect that has received attention recently in the design and operation of EMS is the computational capabilities of computer-based EMS. Kulseth et al (1993) note that increasing CPU power in mainframes, led to demand and provision of additional functionality, thus increasing the role of EMS. They predict that in the next few years, advances in the computer industry and changes in the electric utility business environment will shift the focus in EMS development to the software engineering and integration aspect of these systems. However, increasing functionality will be accompanied by a focus on engineering of systems in such a way as to protect a utility's investment in an EMS by providing for lower maintenance, ease of use, portability, incremental additions, etc.

Electric utilities are moving more and more towards the integration of EMS into the overall business management. This is stated by Murphy and Wu (1993), who note the pressure for changes in the traditional design process resulting from such trends as the move by electricity utilities to integrate power system operation and control with other corporate functions, such as accounting, customer services, and management information systems. This is done with a view to increase overall corporate efficiency and enable the utility to provide a higher level service to its customers.

Another study by Wilker et al (1993) assessed the potential for energy efficiency in electric end-use technologies. The authors distinguish among the following types of energy saving potential:

- Maximum Technical Potential: This concept measures the impact on electricity consumption assuming a 100 % penetration of all-applicable energy-efficient measures and technologies.
- Economic Potential: This concept measures the impact on electricity consumption if all the efficient technologies that are cost-effective to customers are adopted.
- Achievable Potential: This concept measures the impact on electricity consumption due to the adoption of efficient technologies in response to utility programmes. Though technologies are not expected to be adopted unless they are economic or cost-effective, there are several reasons why even cost-effective technologies will not be adopted, including lack of information, transaction costs, and other market imperfections.
- Naturally Occurring Potential: This concept measures the impact on electricity consumption due to the replacement of inefficient technologies in response to the normal workings of the energy marketplace and mandated by standards (p135).

An investigation, through mail survey, of large commercial and institutional buildings was undertaken by Buchanan and Taylor (1989), to determine how EMS affects electricity consumption.

The study concluded that the installation of an energy management system is one action which commercial and institutional utility customers can take to try to reduce their electricity bills. In order for the EMS to work effectively in reducing demand and energy consumption, the building in which it is installed must be conducive to load management. Several attributes were found to contribute to this kind of environment:

- Informed building manager;
- Building function tolerance to temperature fluctuations;

- Large open or common areas (p 218).

Although the research and findings mentioned above concern EMS in commercial buildings, many parallel conclusions can be drawn for the case of a large energy-intensive industry.

Of particular interest is the parallel between the "informed building manager" and an energy conservation manager in the industry (as is explained in detail in this chapter).

According to Ellerbrock (1994), the important ongoing tasks for efficient operation of cement works include not only quality management, personnel management, and maintenance and production management, but also energy management. Successful energy management cannot be achieved without continuous availability of all the necessary information. This is not difficult to achieve nowadays with the aid of computer-assisted systems for acquisition and processing of the measured data. Application of these systems makes it necessary to set up a continuous energy management organisation with suitable working methods and motivated, energy aware, employees. The work of this organisation is aimed at certain energy-productivity targets at different cost centre levels. Important areas of action in energy management include the efficient use of energy, and the use of secondary materials.

Ellerbrock argues that efficient use of energy covers optimisation of operational plants, use of new technology, and optimisation of contractual arrangements for the use of electricity. He claims that these last-named areas are usually very effective and quickly lead to measurable success. On the other hand, particular attention has to be paid to the requirements of product quality during manufacture when using secondary materials.

Optimisation of plant systems starts with the installation or modernisation of the power distribution system. Only by the use of modern switchgear, high-efficiency motors and motor monitoring devices is it possible to achieve any significant energy saving. From the continuous display of measured data, it is also possible to work out strategies for maintenance measures and production planning involving the least possible consumption of energy. Continuous recording of electrical power consumption can, for example, contribute to reducing peaks in the power consumption by avoiding unnecessary no-load operation of machinery, or by appropriate load changes, or by suitable plant changeover. Practical examples have shown that it is usually more cost effective overall to change the production plan and reduce energy costs in this way (Ellerbrock, 1994, p 297).

8.2 Rationale for Energy Management Model

The previous chapter has focused on statistical and economical modelling and analysis, where a set of important management factors (i.e., independent variables) has been identified for the purpose of controlling energy consumption. According to the statistical findings presented in chapter seven ,these factors have explained around 80% of the total variations in electrical and fuel energy consumption in most of the models, thus, justifying the use of these factors for exercising proper levels of energy planning and control at the Jordan Cement Factories (JCF). This in turn requires the development of a coherent management model whose objective is to establish transformational relationships that can translate the process of managing these factors into daily operations and controls. This management model should incorporate an integrated and transparent set of cause-and-effect relationships between factors that control energy consumption, on the one hand, and pre-determined daily actions on the other hand. The transformation process of factors controlling energy use into daily actions will further require organisational development and function structuring. This will necessitate the establishment of an Energy Management System.

While the researcher was trying to establish the concept of an integrated energy management system he contacted many cement manufacturers and other industrial bodies to check the availability of such a system without reaching a positive results. This search was done in the same time when he was trying to check if statistical analysis of the independent factors affecting the energy consumption is used in the industry, the result of this search was stated in chapter 3.

In a study conducted by Worrell et al. (1995), the energy efficiency in those production processes that consume more than 1% of all primary energy in the European Union (EU) have been analysed, among them, the cement industry. The study found that energy savings can be achieved through implementation of energy management systems in the kiln, without giving any specific example of an actual building up or implementation of such a system.

Also through our search of other highly intensive industries it was found that there is a real interest of the issue of energy management and conservation but the effort concentrated on a specific energy saving item and not on establishing an integrated energy management system. For example, according to Energy Information Adminstration of US Department of Energy, the Industry Analysis Brief (2000) showed that about 61% of the steel industry population reported engaging in at least one energy management activity. This showed the relative importance of energy as an input in the manufacturing in terms of its cost as a percentage of total production cost. If these reporting industries consumes nearly 94% of the total steel industry energy, then one concludes that energy management is a major concern for the whole steel industry in the U.S.A. This also indicated that the energy management handling still approaching certain items for energy saving potentioal and still not concentrating on an integrated energy management system approach.

In a paper titled "Reducing the electrical energy costs and consumption in a Portuguese cement plant" by C. Barreiro and colleagues, we found that "Holderbank" management and consulting Ltd was involved in 1986 to assist Cecil cement company in implementing an energy management system at their Outao cement plant. They declare the objective is to increase the plants' profitability by reducing power cost and consumption. The energy management system was implemented in two steps:

- Electrical energy cost reduction (cost/Kwh) and
- Electric energy consumption reduction (Kwh/tonne).

The basic system concept is composed of three distinct, but closely interrelated management tasks, which are supported by software and hardware information system:

1. Energy planning, PROPLAN

   Using the minimum amount of energy cost while meeting production and shipping needs requires careful scheduling of mail equipment. The PROPLAN (production planning) model utilises linear optimisation software to identify the most energy cost efficient production plan.

2. Energy control- ELCON

   Once a production plan is established the operation of the equipment must be constantly monitored and controlled to make sure that energy usage stays within target limits and that costs are minimised. The Elcon (Electric load control) module keeps track of usage (online) and indicates where and when corrective measures are necessary.

3. Energy analysis (EDI)

The EDI (Energy data and information) module provides constant feed back on the complete energy picture. It supplies up-to-date information to:

- Check the impact of energy saving measures
- Pinpoint week spots energy use
- Create an easily accessible database of energy use
- Revise the PROPLAN production schedule and
- Establish new set points for ELCON operation control

The above-mentioned system covers the electrical energy consumption only ignoring the thermal energy consumption, which is a major part of the total energy consumption.

Also during the search for similar EMS model building we found in the last stage of our research a very useful good practice guides consisting of many several disciplines energy issues prepared by the Department of Environment in UK. It is a detailed energy materials designed to offer energy management guidance technical advice and information to build up energy management information system, monitoring and targeting procedures and other important energy related issues to help the industry to set their best practice in energy management.

This chapter will present the development of a management model that uses quantitative data analysis as its foundation, and ultimately produces as an output an integrated energy management system including a set of organisational and procedural actions. These procedures can be simplified further to facilitate application and implementation by trained staff at JCF.

8.3 Objectives of the Energy Management Model

To use and capitalise on the results of the quantitative treatment of energy consumption factors in chapter seven, a management model will be developed. The objectives of the management model are summarised as follows:

1. Establish and strengthen a quantitative basis for the decision making process at all management levels.

   Establish clearly identified positive relationships between individual energy management factors (e.g., production rate) on one front of the model and the corresponding requirements of timely and procedural actions, on the other. This will be achieved through a functional mapping process of the various functions at the JCF.

2. Integrate all procedural requirements into a unified platform using an organisational structure that supports these procedures at the JCF.

3. Use model methodology as an illustrative vehicle to demonstrate its applicability to other areas of improvement at JCF.

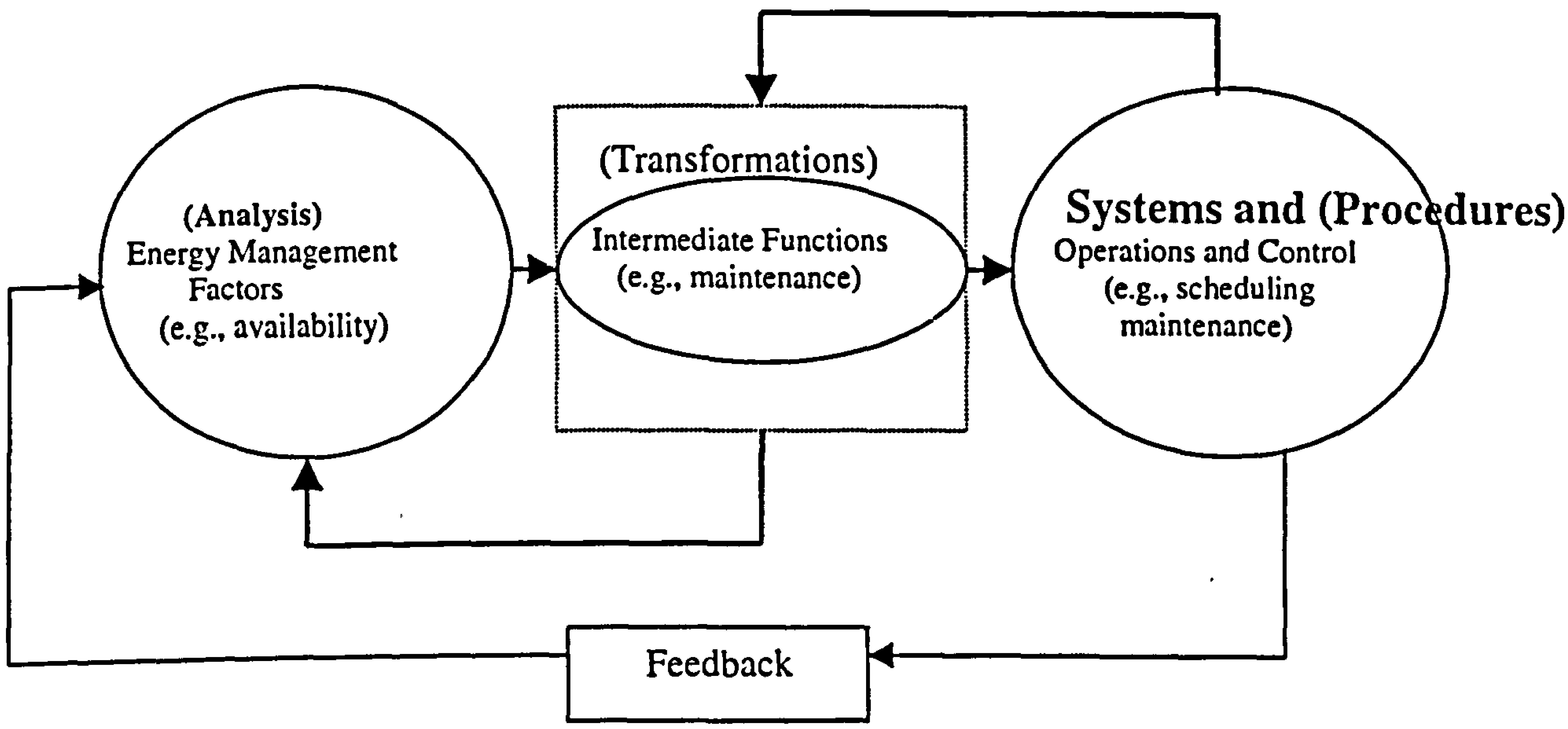
8.4 General Framework of the Energy Management Model

The concept of the model is founded on establishing a set of mapping functions between energy management factors and procedural actions. *Energy management factors* are defined as the factors that affect the use and consumption rates of energy. At the JCF, according to the statistical analysis of Chapter 7, these factors include availability, production rate, average number of stoppages, and average duration of stoppages. In the context of the presented methodology, these management factors are called high-level factors. In management theory, high-level management factors (i.e., independent variables) should actually reflect the strategic objectives of the organisation. Therefore, the management model presented here is actually a tool for *aligning* operations with strategies. Management factors, whether high- or low-level, can be determined through analysing data of the system as well as experience and judgement of its managers.

Thus, the mapping process aims at developing multi-level relationships between high-level management factors and lower-level management factors. To align high-level management factors with lower-level factors, one may require two or more relationships to carry out the transformation from the inputs of the model to the outputs. Here, we are defining the inputs as the energy management factors such as availability and production rate, and the final outputs as any necessary procedural actions whose implementation will directly affect the levels of the inputs. The intermediate functions represent internal factors that must be defined to facilitate establishing the relations between inputs and outputs.

For example, a mapping process that will translate a high-level management factor into operational controls can be explained as followed. Availability, as explained Chapter 7, is an important management factor. The maintenance function and its related control factors can be considered as intermediate functions. The maintenance function in turn is affected by lower-level factors such as technical skills of maintenance staff, activity recording system, and maintenance scheduling. The implementation of these activities requires tailored organisational structures and related operational procedures. These operations may include preventive maintenance procedure, corrective maintenance, procedure for analysing maintenance reports, and so on. Thus, availability as a high-level management factor is aligned with daily operations and procedures, such as maintenance scheduling and staff training. The number of necessary transformations (number of levels) depends on the complexity of operations and the required level of details of the operations control. A schematic representation of this process is shown in figure (8.1).

Figure (8.1)
Conceptual Framework of the Management Model



The mapping process among variables of the system, affecting strategic objectives (best use of energy) assists in identifying how the ordinary operator can ultimately perform certain actions and measures on the factory floor to achieve these objectives. This process can be summarised in three basic phases:

1. Analysis Phase: This phase basically entails the necessary data collection and identification of relationships amongst certain management factors and lower-level performance factors. In the context of the present discussion, Chapters 6 and 7 have actually dealt with this phase. The outcomes of this phase have been a set of energy management factors that included availability, production rate, and average number of stoppages and average duration of stoppages that altogether affect energy consumption rates. These key management factors reflect one strategic objective of the organisation; in the case of the JCF, this objective is the rational consumption of energy.

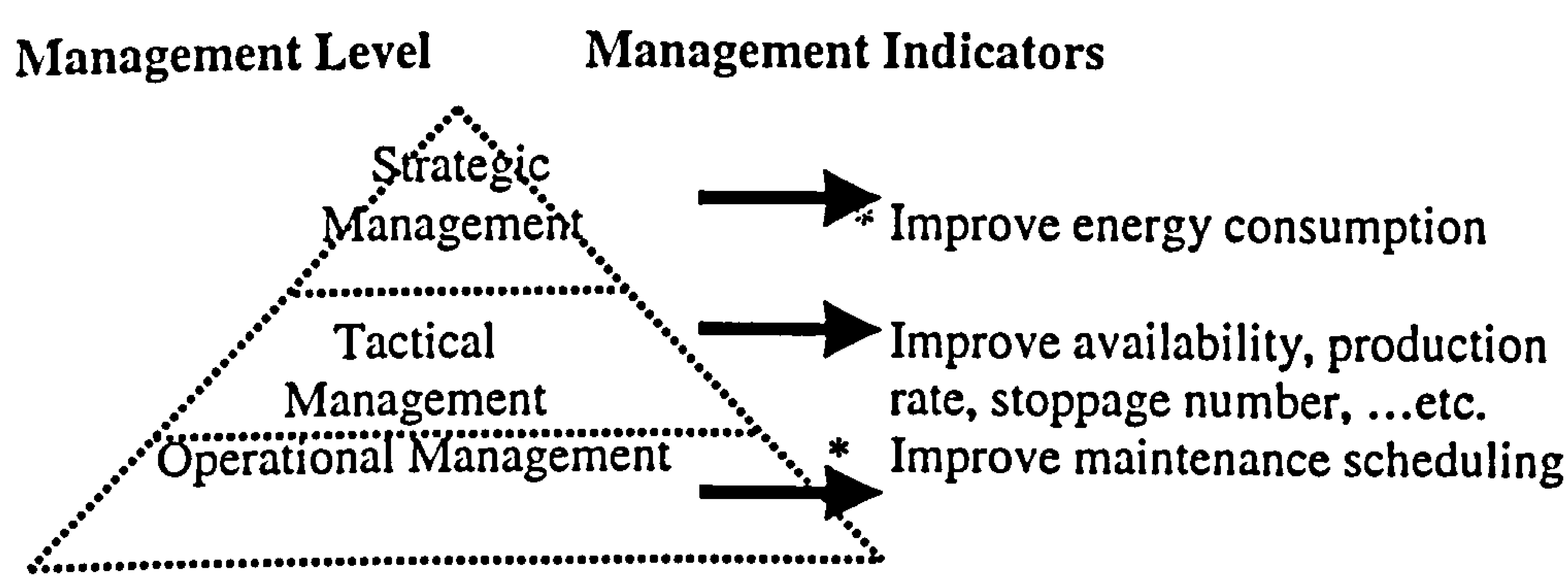
2. Transformation Phase: During this phase, efforts must be focused on establishing intermediary factors that are directly or indirectly related to the energy management factors on one hand, and to other lower-level factors on the other hand. The latter factors must lead to easy-to-understand and implement actions.

3. Energy Management System Development Phase: During this phase, a system should be developed. This system includes organisational and procedural aspects. Such organisational structure and procedures have to be developed in a systematic and standardised manner. Organisation chart, job descriptions and data recording forms must also accompany such related procedures.

Indeed, the prescribed management framework is virtually streamlined with the conventional management model that is, in turn, manifested in the nature of the management activities involved in each of the above phases. Figure (8.2) shows the pyramid of typical management layers, strategic, tactical, and operational, and the corresponding management factors and functions. For instance, strategic management would set performance indicators for the organisation that measure the achievement of the strategic objectives, while tactical management would transform these indicators into manageable intermediary control factors. The identification of the intermediary control factors requires analysis of the data generated at the operational levels. Tactical and operational management would then translate these manageable control factors into operational actions and procedures. Operational management would actually involve in day-to-day transactions, and measure performance and *align* measurements with intermediary control factors.

At the JCF, one strategic objective is to reduce the rate of energy consumption; at the tactical level, middle management would actually study and improve factors that affect energy consumption. Availability is such an important factor. Operational management assumes the responsibility of, for instance, planning and controlling maintenance activities that will improve availability, as shown in figure (8.2). This breaking down of management activity at all three levels, strategic, tactical, and operational, is actually inherited in the framework of the proposed management model.

Figure (8.2)
Conventional Management Model



The concept of the management model used in this research has been developed using principles and techniques of Total Quality Management (TQM) which is defined according to Barrie G. Dale as the mutual co-operation of everyone in an organisation and associated business processes to produce products and services which meet and, hopefully, exceed the needs and expectations of customers. TQM is both a philosophy and a set of guiding principles for managing an organisation. According to United Glass Co. statement in the good practice guide no. 169 prepared by UK Department of Environment , "We needed a vehicle for preaching the world in energy efficiency. It made sense to utilise the company initiatives on quality to achieve our aim of managing energy better." TQM was suggested by the best practice programme initiated by the Department of the Environment in the UK to be used as a vehicle to implement energy conservation practices in a detailed good practice guide 169 prepared by UK Department of Environment, called "Putting energy into total quality, a guide for energy managers". The utilisation of TQM in the suggested energy management system will be discussed in the following sections.

8.5 Development of Energy Management at Jordan Cement Factories

The Energy Management System (EMS), which is currently being used at the JCF, is the result of the original research work of the author in collaboration with the factory managers and the energy section heads. They were responsible for the idea as well as the design and supervision of the implementation.

Applying energy modelling concept to the particular case of the Jordan Cement Factories with the aim of building an integrated Energy Management System (EMS) requires top management commitment and the creation of a certain organisation culture in order to guarantee effective implementation of the EMS. This EMS is designed to achieve the strategic objective of reducing energy consumption through the integration and interaction of strategic, tactical and operational management.

This integration and interaction require the accomplishment of the following: determining the main goals, preparing necessary plans, supervising the implementation and execution of these plans, and defining adequate methods for auditing and controlling the energy consumption. The objective of such a system is to ensure a minimum cost of producing cement by lowering the cost of utilised energy through implementing the following measures:

1. Formulating a policy for energy management, a coherent and effective corporate policy statement provides the foundation to the planning, according to good practice guide 186 prepared by Department of the Environment 1996 good successful policy can be recognised as having five key attributes as follows: (Thrust, Commitment, Applicability, Implementation and Review)
2. Preparing an annual energy saving plan.
3. Determining the annual, monthly, and daily energy cost and consumption.
4. Determining the form and the duties of the energy management staff. According to good practice guide 85 prepared by Department of the Environment 1998 standards needed for managing energy so as to define the role of individual responsible for managing energy within an organisation. Managing energy is a complex and challenging task, which need specialised energy staff to ensure the effective management of energy resources to meet the organisation objectives.
5. Following-up the performance of cement production units to identify obstacles and problems and methods to increase efficiency. According to good practice guide 119 prepared by Department of the Environment 1997 it is important and essential to define levels of energy use in the organisation and to be able to determine energy saving in order to gain an appreciation of energy issues you need to draw up a profile of energy use which answer the following questions: (why? when? how much? and cost).
6. Preparing necessary studies to overcome problems and upgrade present equipment in order to lower the specific energy consumption.

7 Determining the requirements for executing the energy plans and programmes and supervising implementation.

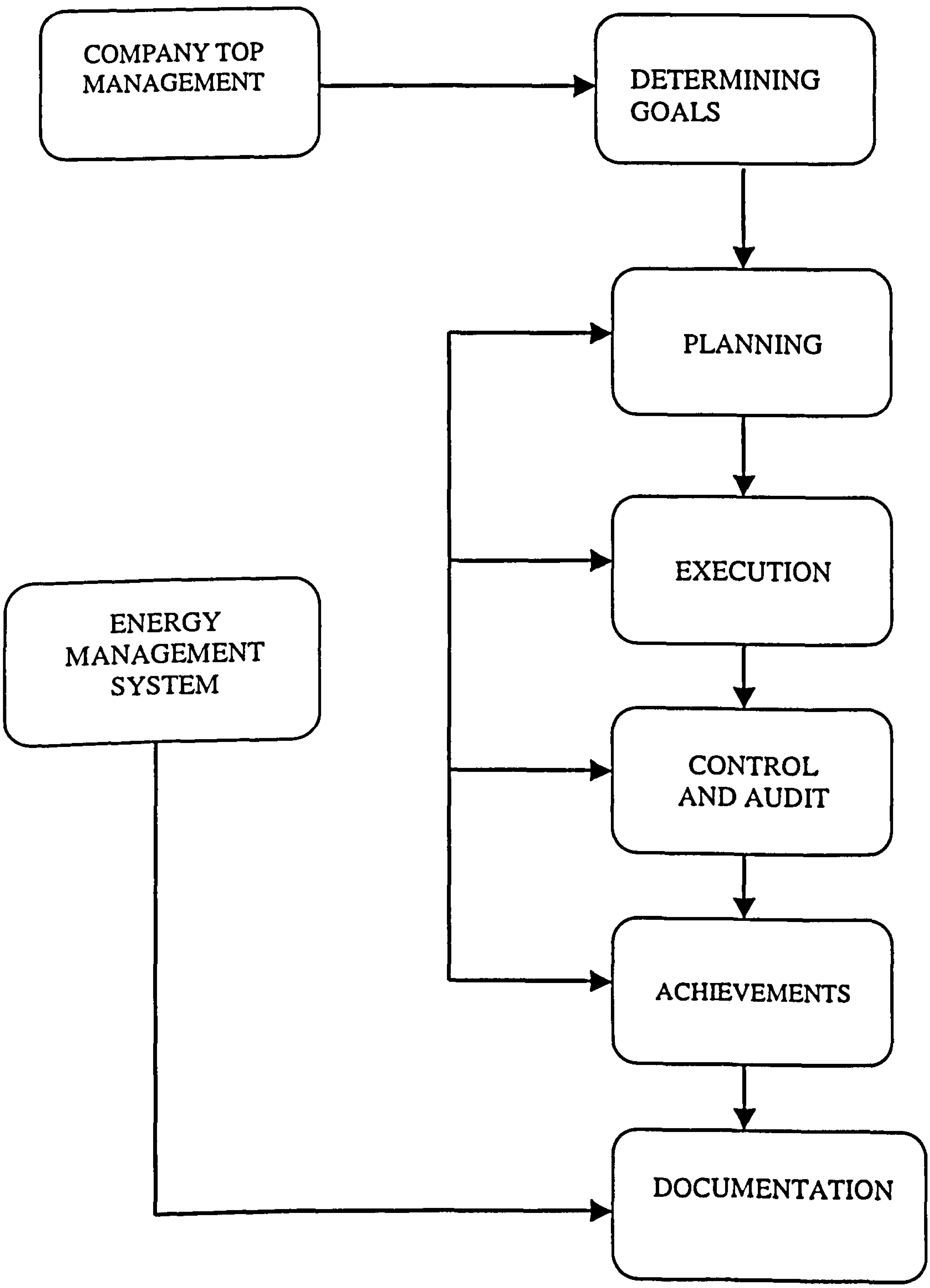
8. Identifying suitable methods to audit and control energy consumption.

9. Following-up worldwide developments in the field of energy in the cement industry, which can serve in increasing efficiency and modernising equipment.

10. Increasing the cement company's awareness of the importance of energy management. According to good practice guide 85 prepared by Department of the Environment 2000, awareness training should be pitched at the heart, it needs to provide reasons why should be more energy efficient, also it should give everyone in the organisation the incentive, motivation and convention that being energy efficient is the correct approach.

11. Keeping top management involved and informed on energy issues to help secure the necessary support for energy management plans and programmes. According to good practice guide 119 prepared by Department of the Environment 1997, to gain creditability and high level management attention and commitment a case for an energy management program must incorporate important elements which includes (timing, benefit, pathway, choice, resources). Figure (8.3) presents the main functions and procedures of the Energy Management System developed by the researcher. It is clear from figure (8.3) that the involvement of top management is essential for the success of the EMS; indeed, it is built into the system, such that without it the system collapses. Moreover, the personal involvement of the author and the top management from the beginning was the main contributor to its adoption and institutionalisation and to its successful implementation.

Figure (8.3)
ENERGY MANAGEMENT FUNCTIONS & PROCEDURES IN THE JORDAN CEMENT FACTORIES



8.6 Elements of The Energy Management System

8.6.1 Overview

Energy is considered as a resource to be managed along with land, labour, capital and raw materials, which suggests that the key elements of energy management system will be the same as the elements involved in the management of other resources. The elements that must be present in a successful energy management programme are grouped under four basic functions of management, namely:

1. Planning
2. Execution
3. Control
4. Organising

However, as organisation or the "organising" function is an overall function embedded in each of the other three, emphasis will be focused only on the first three functions. The most important factor in implementing an energy management system is the support of top management and the realisation that energy can be managed, and that its management will contribute to minimising production cost. The person in charge must have enough experience in this field and be delegated adequate authority to be able to prepare plans, follow-up the execution, and carry out the necessary audit and control.

The first step of energy management procedures is the determination of the goals and purpose of the management system. The main goal is the reduction of production cost by minimising energy consumption per unit of production. This goal must be acknowledged by all company employees, from top management down to workers and labourers. This will place the responsibility for realising the objectives on all employees of the company; success can only be achieved through the support and participation of everyone.

As we illustrated in chapter five, in the cement industry, the cost of energy in producing cement is relatively high in comparison with other types of industry. Therefore, establishment of energy management systems is important and, indeed, essential. The elements of the developed system are outlined in the next sections.

8.6.2 *Basic Elements of the EMS*

A) Planning

According to good practice guide no119 prepared by Department of Environment 1997 the first ingredient in action planning is to provide clear written objectives. Targets to achieve these objectives must be agreed in consultation with those who are interested to achieve them. After realistic objectives, targets and goals have been set, the planning process begins with the realisation that energy is a valuable resource that must be managed. Moreover, there must be a decision to see that proper management is really applied (Henry et al, 1980).

In the cement industry, planning for energy saving processes is essential. Based on our practical experience and continuous internal discussion about energy issues and also based on practical implementation of energy management procedures (supported by literature review and learning from others experience) planning can be done by following the planning flow sheet, figure (8.4) procedures and the planning procedures in table (8.1):

1. At the end of every year, an annual energy saving plan is prepared for the following year by the energy manager and energy section heads.

2. Collection of necessary data to prepare the plan, including;
   - The estimated annual production for the following year;
   - The estimated thermal and electrical energy required to produce the planned quantity;
   - Calculation of the expected annual energy cost;
   - Preparation of the annual maintenance plan for all production units;
3. Analysis of the data to set the goals for the following year.
4. Making proposals and generating ideas and solutions to identified problems and implementing necessary modifications that would result in minimising energy consumption.
5. Determination of the necessary budget and resources of materials, engineering, studies and labour, and then the preparation of a schedule for execution.
6. Presentation and discussion of suggested measures with the managing director and plant works manager for approval.
7. Preparation of the plan document and its distribution to concerned departments.
8. Plan implementation according to the established time schedule.
9. Preparation of monthly reports on the achievements and obstacles in plan implementation to the managing director and the plant works manager. Preparation of an annual report at year's end to present the annual achievements with regard to the implemented plan and to show the benefits.
10. Documentation of the plan, reports, achievements and obstacles to provide results and recommendations that are useful for the preparation of the annual plan for the following year.

The expected percentage of energy saving must be estimated and compared against the actual percentage saved as a result of the implementation of the plan as calculated by the end of the year. In addition, the top management must be informed of the progress in plan implementation and related achievements.

Information on the success of any plan in realising the set goals and overcoming the problems and obstacles is essential in the preparation of the annual plan for the following year.

Figure (8.4)

PLANNING FLOW SHEET PROCEDURE

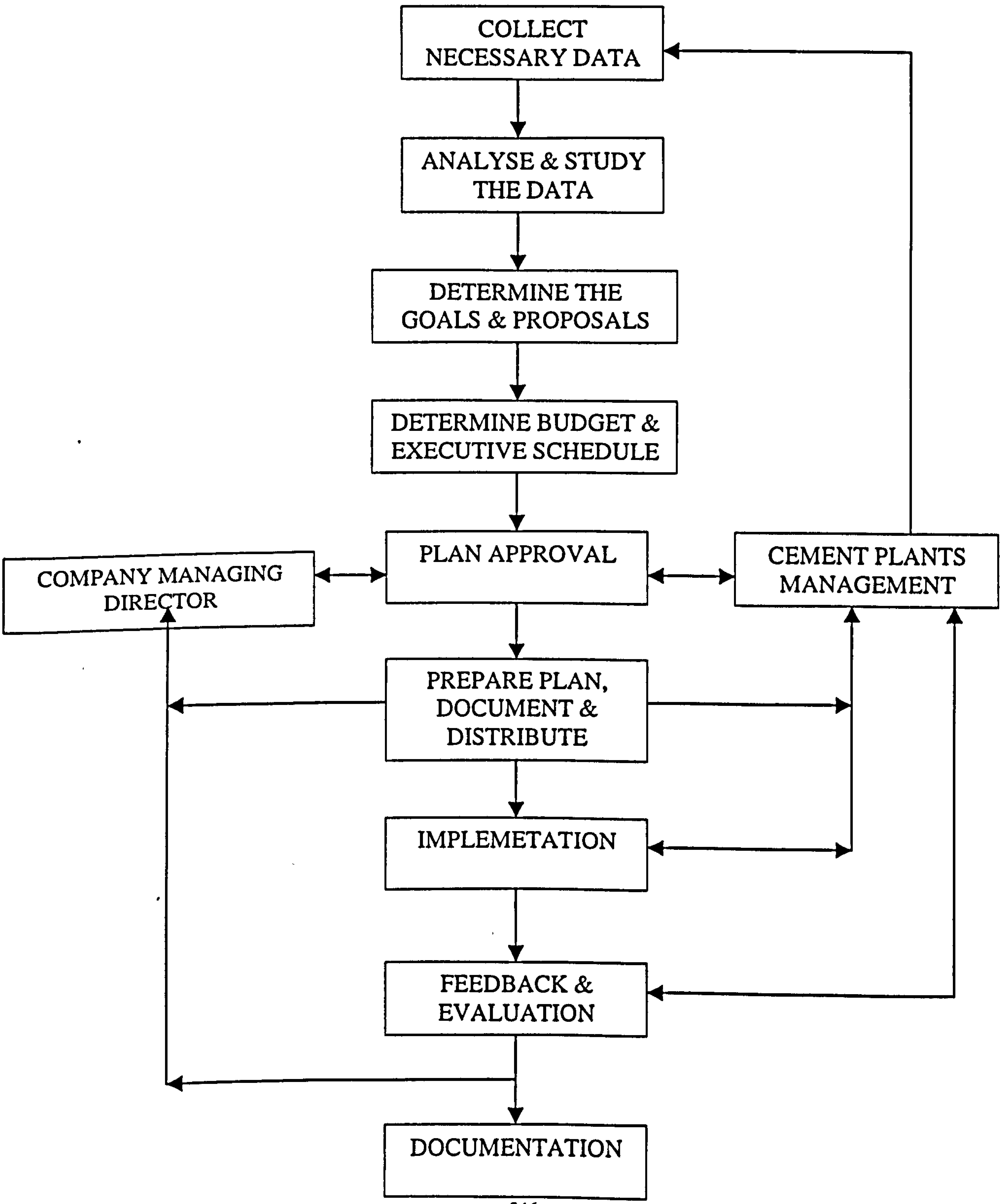


Table (8.1)
Planning Procedure

Procedure No.	Procedure Name	Responsible Person	Reference
1.	Collect necessary data	Energy saving manager and energy plants staff	• Annual production plans • Annual energy Requirements • Annual maintenance plan • Problems and obstacles
2.	Analyse and study the data	Energy saving manager and plants energy section heads	Job instructions how to analyse and study the data
3.	Determine the goals and proposals	Energy saving manager	
4.	Determine budget & execution schedule	Energy saving manager	
5.	Plan approval	Energy saving manager with managing director and plant works manager	Job instruction on how to get the necessary approval
6.	Prepare plan, document and distribute	Energy saving manager and plants section heads	
7.	Implementation of the plan in cooperation with all necessary departments	Plants section head and all necessary staff especially energy quality circle members	Energy annual plan
8.	Obtaining the necessary feed-back to evaluate	Energy saving manager and plants section heads	Following-up reports
9.	Documentation	Energy saving manager and plants section heads	

B) Execution

B.1) Commitment

The execution of energy plans and programmes requires management commitment. This commitment must stem from the highest level of management and must include the allocation of the necessary personnel, time and funds to carry out the programme or plan, in order to achieve satisfactory results concerning energy conservation. According to good practice guide 119 prepared by Department of Environment 1997 it is essential to gain the commitment of the most senior members of the management team as individuals and part of the corporate body which require a convening case for action or resources, clear action pathways and merits that fit the strategic company goals and fulfil the needs of stack holders.

To achieve this target, the organisational structure of the energy management system must be suitable and sufficient to ensure that all plans and programmes are completely implemented. To ensure commitment to energy issues from all necessary personnel, a committee as well as quality circles must be formed, especially in the production plants where energy use is considerable.

The experience of the Jordan Cement Factories regarding organisation structure and the forming of the committee and quality circle and the interaction among them lead us to suggest the following organisation structure:

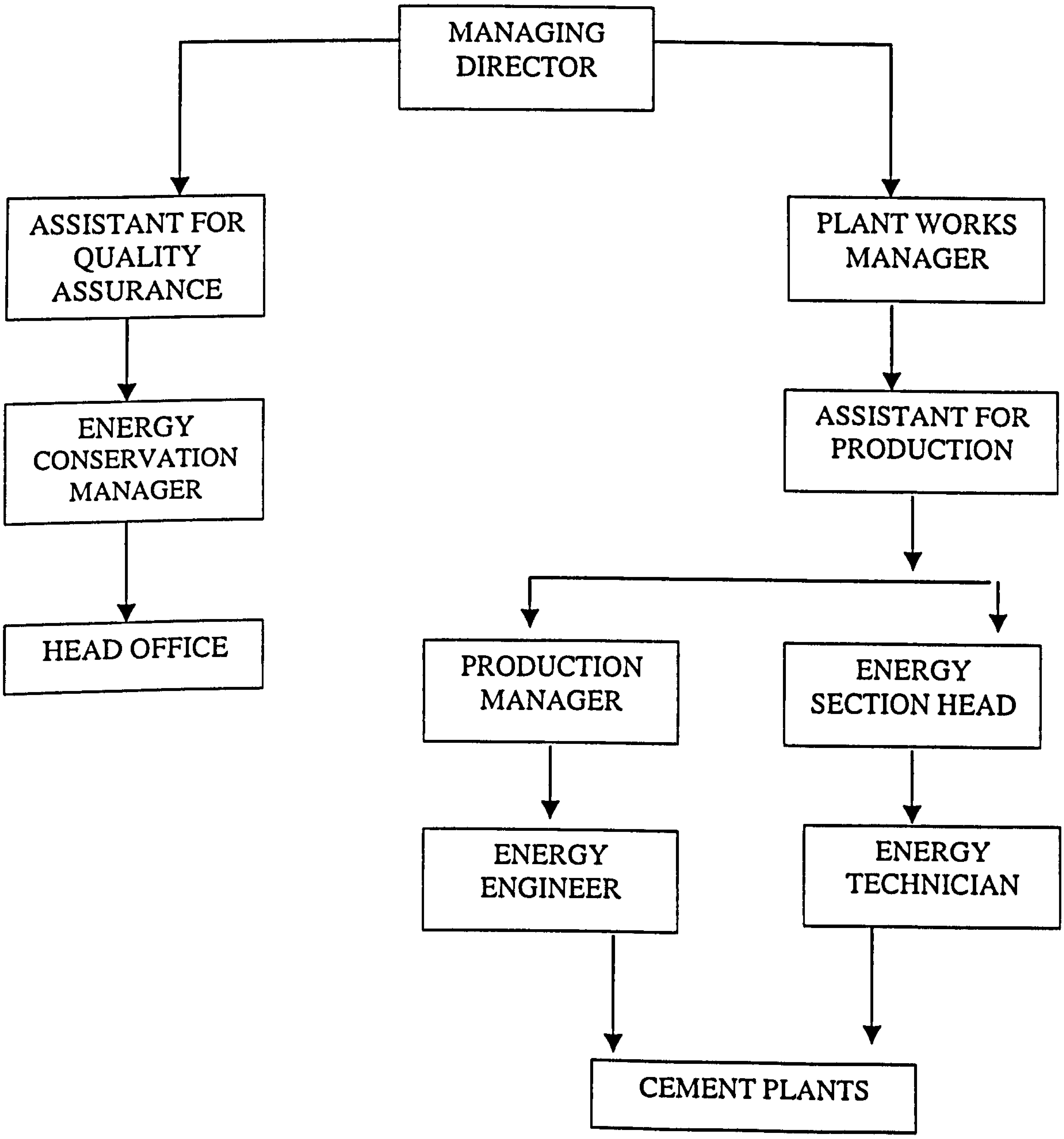
B.2) Organisation Structure

According to good practice guide 119 prepared by Department of Environment 1997, it is essential to identify the key players in energy management so that their respective input can be sought in sitting and achieving improvement targets. It is important to appoint energy manager in a managerial role to assess in driving forward a programme for continual improvement in energy efficiency.

Looking at the organisational chart (figure 8.5) concerning the execution function of energy management, the following observations can be made:

- The organisation diagram consists of one energy conservation manager in the head office and one energy section head in addition to one engineer and one technician in each plant.
- The energy conservation manager reports directly to the assistant managing director for quality assurance affairs. Therefore, he can make direct contact with the managing director if necessary to provide explanations, reports...etc.
- The energy section heads in the plants are connected and report to the works manager's assistant for production affairs. The assistant will attempt to produce the highest possible quality of cement at a minimum cost of energy, while the energy section head controls and audits the energy consumption in the production department. Most energy in the cement industry is used by the operation on production units. It is easy to implement any proposals and instructions concerning energy issues via the assistant.
- The energy conservation manager prepares energy conservation plans, and programmes in co-operation with plant section heads while the implementation is carried out by and under the supervision of plant energy section staff.
- While the energy conservation manager audits and controls energy consumption for the whole company, the energy section head audits and controls energy consumption in his plant only.
- While the energy conservation manager prepares annual plans and programmes for the whole company, the section head helps in the preparation of the annual plans and programmes for his own plant and carries out their implementation.
- While the energy conservation manager sets the goals and objectives of the whole company, the section head determines the goals for his own plant only.
- While the energy conservation manager prepares necessary studies and proposals for the whole company, the section head prepares them for his own plant only.

Figure (8.5)
ORGANISATIONAL CHART FOR EXECUTION



B.3. Duties and Objectives of Energy Management Staff

To implement successfully the energy management system and fulfil the established goals (according to good practice guide no 186 prepared by Department of Environment 1997 and organization corporate goals are a summary statement of its commitment to improve its energy performance), the duties and objectives of the energy department and energy sections in the plants must be clearly defined. The sited objectives need to reflect long-term corporate goals and are usually expressed as a percentage of improvement in the energy performance. The job description of each employee in the management system must also be established.

The example given below shows the duties and objectives of the energy department in the head office and in the section in cement plants, as well as job descriptions of all energy management employees.

B.3.1 Objectives and Duties of Energy Conservation Manager (Head Office).

- Objectives:

Auditing and controlling the plants' energy consumption, preparing plans studies and programmes to conserve energy in both plants to ensure that the energy consumption is at a minimum level.

- Duties:

1. Prepare an annual energy plan in co-operation with plants' energy section.
2. Follow-up and organise daily and monthly energy consumption data and highlight any deviations.
3. Prepare monthly reports regarding energy issues (consumption, cost, studies, researches, instructions and suggestions).
4. Follow-up the activities of the plants' energy committee and energy quality circles.

5. Follow-up the performance of the plants' production units regarding energy consumption.
6. Prepare the necessary technical studies in co-operation and co-ordination with plants' energy section heads.
7. Participate in the implementation of studies and results of research to ensure that they actually lead to minimisation of energy consumption and conform to the goal of conservation of energy.
8. Make frequent plant visits to ensure that suggestions and instructions concerning energy issues are implemented. Also follow-up the progress in the implementation of plans and programmes.
9. Review the plant energy reports (weekly and monthly studies and reports) and make notes.
10. Send brochures, leaflets and instructions on energy issues to plants to enhance awareness, knowledge and experience.
11. Correspond with specialised companies, where needed, to resolve encountered problems and obstacles and introduce developments in energy issues.
12. Document all work related to the energy conservation programme, including reports, studies and projects.
12. Follow-up assignments requested by the managing director or his assistant.

B.3.2 Objectives and Duties of the Energy Section (Plants):

- Objectives:

To control and conserve thermal and electrical energy in the plant through preparation of plans and programmes, follow-up of the production process, audit of energy consumption, suggesting of appropriate solutions to problems, and ensuring that all production units operate at a high production rate, and at minimum energy cots.

- Duties:

The main duties of energy section are as the follows:

1. Preparation of an Annual Energy Plan:

   At the end of each year, an annual energy plan is prepared in co-operation with the energy conservation manager through the following procedures:

   - Collect necessary data required for preparation of the plan from the production and maintenance annual plans;
   - Study and analyse data;
   - Propose suitable ideas and solutions to control and save energy to all production units including quarries, crushers, raw mills, kilns, cement mills and packing plant;
   - Discuss the proposed ideas and solutions with works manager, plant department managers and energy manager to obtain approval;
   - Prepare an annual energy plan to execute the proposals and allocate the necessary budget and manpower and define an execution schedule;
   - Distribute the plan to all concerned departments;
   - Follow-up the implementation of the plan;
   - Prepare weekly and monthly reports on the achievements and on the evaluation of the execution;
   - Document all reports and the results of the execution in special files and on computer.

Figure (8.6) is a flow chart, which summarises the procedure for the preparation of the annual energy plan in cement plants. Moreover, table (8.2) presents the information in a tabular form.

Figure (8.6)
PLANT ANNUAL ENERGY PLAN FLOW CHART

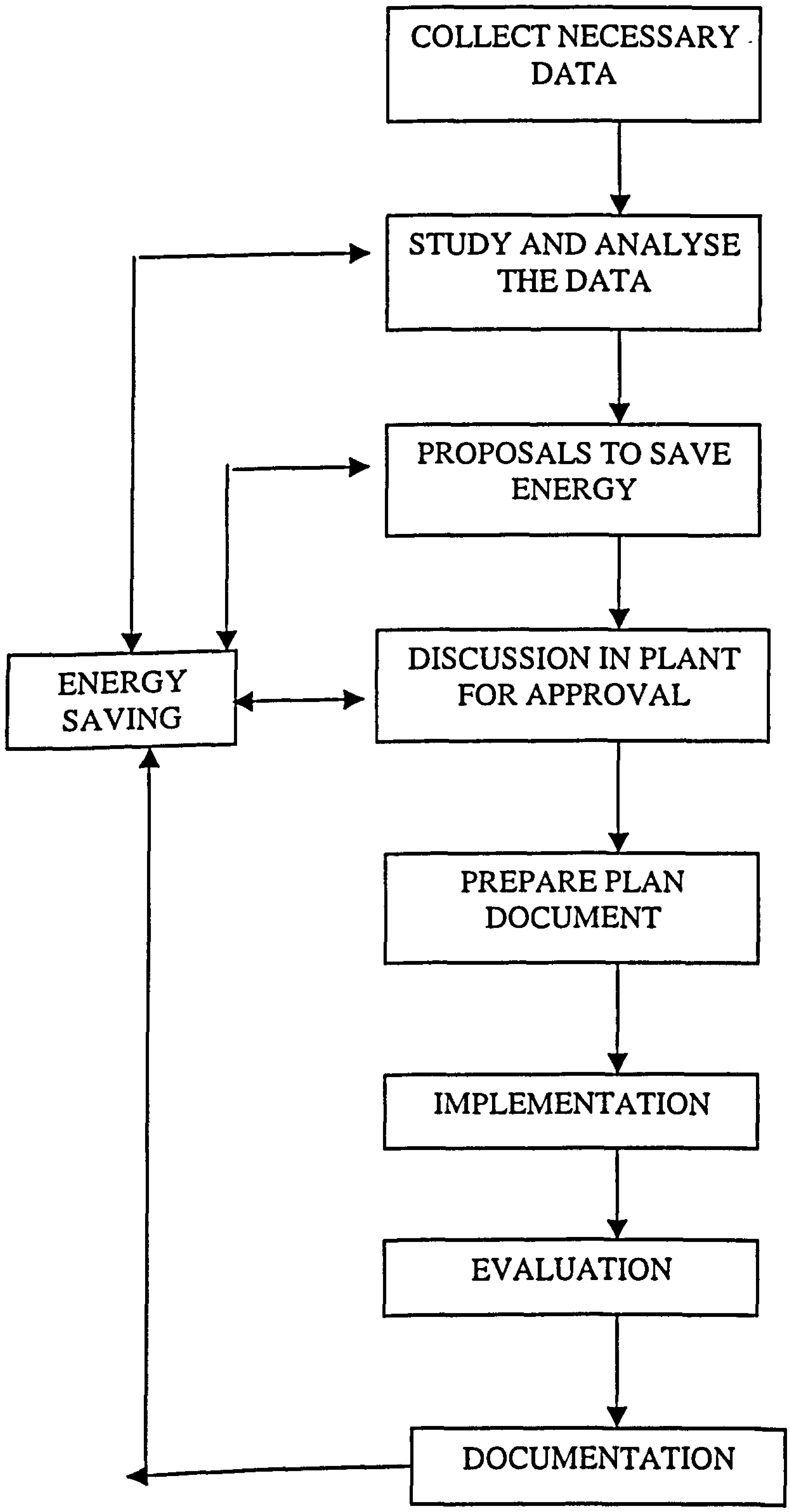


Table (8.2)
ANNUAL ENERGY PLAN PROCEDURE

Procedure Name	Responsible person	Reference
Collect necessary data	Energy Engineer and Energy and Technician	• Annual Production and quarry plans. • Studies' results. • Recommendation
Study & analyse the data	• Energy Manager • Section Head	Job instructions: how to analyse and study the data.
Proposals to save energy	• Energy Manager • Section Head	
Discussion in Plant for approval	All concerned departments	
Prepare Plan document	Energy Section Staff	Job instruction: how to prepare the plan
Follow-up the implementation	Energy Section Staff Manager	
Evaluate the progress	Energy Manager and Section Head	Results of implementation and reports
Documentation	Energy Section Staff	Job instruction on how to document

4. Energy Studies:

The aim of conducting energy studies is to identify and resolve problems related to energy demand and thus increase operational efficiency and performance, by modernising and developing the equipment. This is done through the following steps:

- Identify obstacles and problems through the review of daily production reports, daily plant performance record, and energy plant committee or energy quality circle recommendations.
- Carefully study obstacles and problems and collect data needed to develop solutions to overcome the problems.
- Prepare the necessary studies on equipment modernisation.
- Discuss results of the studies at a meeting of the energy plant committee or the energy quality circle to receive recommendations.

- Implement the recommendations of the studies. Prepare reports on achievements for review at the next meeting of the energy plant committee.
- Document studies, research and results in special files. Send a copy to the energy manager.

All studies, research and recommendations must be made under the supervision and control of the energy conservation manager. Figure (8.7) and table (8.3) show the details of this task.

Figure (8.7)
ENERGY STUDIES PROCEDURE IN THE PLANT

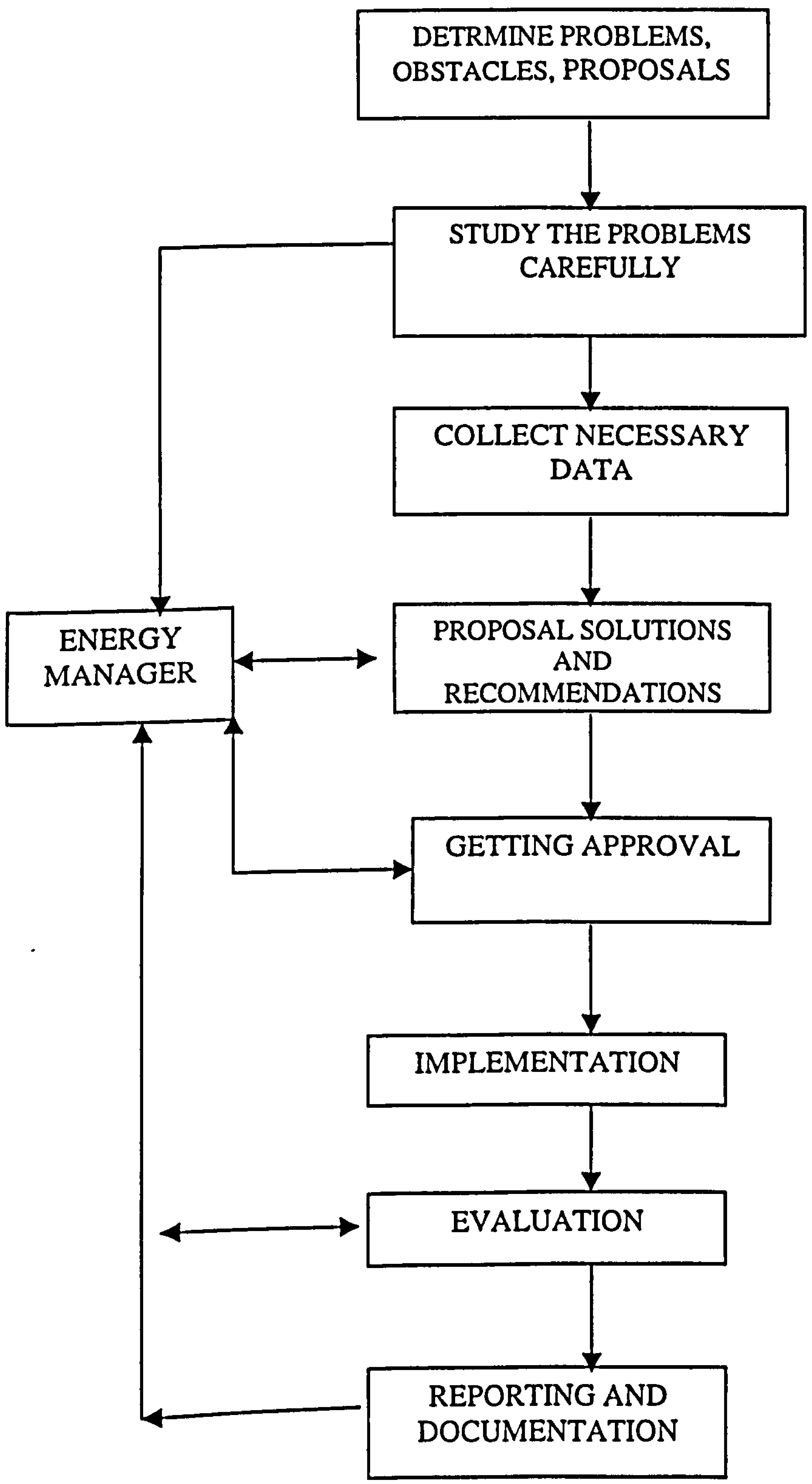


Table (8.3) : ENERGY STUDIES PROCEDURE

Procedure Name	Responsible Person	References
Determine problems, obstacles, suggestions, deviations	• Section Head • Engineer	• Reports • Recommendation of meetings
Study carefully the above mentioned items	Energy Manager and Section Head	
Collecting data	• Section Head • Engineer • Technician	
Propose suitable solutions after discussion and correspondence with special companies	Energy Manager and Section Head	
Write report and send to works manager to get the necessary approval	Section Head	Studies reports file
Implement the recommendations	• Section Head • Engineer • Technician	
Write a report about the results and achievements of the study (evaluation)	Section Head	
Documentation	• Section Head • Engineer	File of results and achievements of the tech. studies.

4. Follow-up of Performance of Production Units:

The objective of this task is to control and monitor the performance of production units and to participate in maintenance work to resolve problems related to energy. The procedures for this task are as follows:

- Read daily operation reports and clarify deviations.
- Control the actual operation conditions of units and equipment in the field and make observations.
- Evaluate the commitment of concerned departments regarding the recommendations of the plant energy committee and the energy quality circle.

- Maintenance plan: Participate in maintenance work to ensure that all tasks related to energy are performed well. For example, avoidance of air infiltration and the time of start and stop of the equipment on no-load test.
- Production plans: through monitoring production and quarry plans, it is possible to propose ideas concerning energy such as, the increase of night utilisation of power, prevention of operation during peak hours, and controlling the volume of raw material stock piles etc.
- Modifications and up grading: Through follow-up of activities related to modification and up grade of main equipment or the increase in operation efficiency, it is possible to evaluate the effect of these activities on energy conservation through preparation of useful reports.
- All the above-mentioned steps must be discussed at the meetings of the plant energy committee and the energy quality circle and then should be documented in special files.

A copy of the reports must be sent to the energy manager to keep him well informed of the actual performance of production units.

B3.3 Job Description of Plant Energy Section Staff

The job descriptions of the energy section staff in the cement plants of the following posts will be presented in appendix no (23):

1. Energy Section Head
2. Energy Engineer
3. Energy Technician

B.4 Energy Committee And Quality Circle In the Plant

It is important for a company to establish a special committee or assign a consultant to assist the energy management in executing its plans and programmes, and to guarantee the involvement of all concerned employees in energy management. Such a committee should contribute in the following ways:

- Clarify goals and objectives to all employees, especially those related to energy issues.
- Assist in, and facilitate in the execution and implementation of plans, programmes and instructions concerning energy aspects.
- Explain to employees the necessity and importance of conservation of energy and that it is the responsibility of every one in the plant.
- Increase self-control on energy consumption by involving the people who are using energy in a committee or a quality circle.

The forming of a committee or quality circle is the second most important factor in ensuring successful execution of plans and programmes. The experience of Jordan Cement Factories (JCF) can be used as an example to demonstrate the importance of these committees.

The top management of JCF has formed the following committees in the plants:

1. Plant energy committee.
2. Energy quality circle.

The objectives and duties of these committees are explained in the following:

<u>Plant Energy Committee</u>

The aim of this committee is to control and conserve energy in the plant. The following set-up is used:

The committee is comprised of the works manager, production, quarry, electrical, mechanical maintenance managers and the energy section heads. Naturally, the energy section head of each plant reports to the energy manager. So although the committee is at the level of the plant, global (company) decisions are derived from its deliberations and recommendations.

The committee holds a monthly meeting or upon the request of its chairman, who is the works manager.

The energy section head usually prepares the agenda of the monthly meeting.

The agenda contains the following items:

- Review of recommendations of the previous meeting and reports on the achievements.
- Review of results and recommendations of studies related to energy.
- Discussion of energy consumption of production units from quarry to packing plant and analysis of deviations to identify causes.
- Annual energy plan: Discussion of achievements and the deviations.
- Discussion of possible modifications, suggestions and ideas concerning conservation and control of energy consumption.
- Discussion of reports and achievements of the quality circle weekly meetings.

- The energy section head will set the date, time and place of the meeting.
- He will be responsible for the preparation of the report on the proceedings and outcomes of the meeting and send a copy to the energy manager.
- He will document the reports of the committee.

Figure (8.8) and table (8.4) depict the functions and tasks of the plant energy committee.

Figure (8.8)
PLANT ENERGY COMMITTEE FLOW SHEET

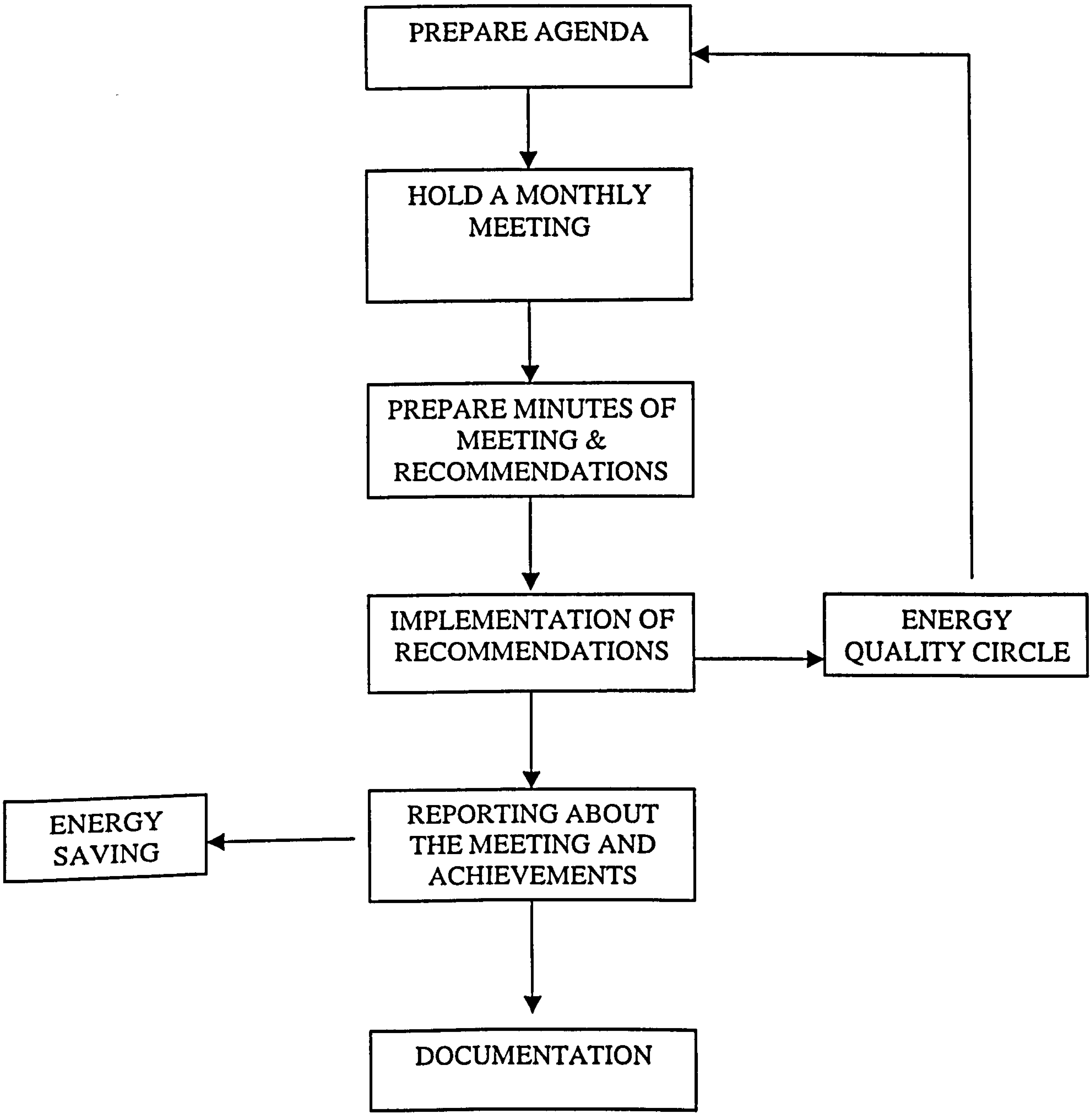


Table (8.4)
PLANT ENERGY COMMITTEE PROCEDURE

Procedure Name	Responsible Person	References
Prepare agenda	Section Head	Weekly, monthly reports, previous recommendations
Hold a monthly meeting and discuss the agenda	Committee Chairman and Section Head	
Prepare minutes of meeting and recommendations	Section Head	
Follow-up to implement the recommendations	Section Head, Engineer and the Technician	Minutes of meeting
Write report about the meeting and the achievements, send one copy to energy manager	Section Head	Studies reports file
Documentation	Section Head	Plant energy committee file

Our experience of this committee shows that the advantages include the participation of all the main managers in the factory in setting targets for the energy consumption and systemically observing and following up the actual results toward decreasing the energy costs. Also it creates the incentives for them and for their staff to achieve the best results in minimising energy consumption issue, because through the committee evaluation process their success will be demonstrated will be reflected positively on them and they will be accounted for their failure. As a result of the work of this committee as a part of energy management system there was gradual improvement (decrease in the energy consumption in the Jordan Cement Factories), which will be demonstrated in the following sections.

As for the obstacles which we faced in running this committee it revealed that the managers are not accustomed to work together as one team to handle a certain issue, also they were not accustomed to participate in disciplined systematic meetings and to follow a very strict recording and archiving and auditing systems. It was a positive outcome to let them adapt themselves to this teamwork style of management.

- Energy Quality Circle (EQC)

According to guide no 169 of the good practice program prepared by Department of Environment 1995 (a quality organization means reducing unnecessary waste and making efficient use of energy and resources. Quality circles are a TQM tool to achieve quality organisation through establishing teamwork approach in all its forms. The team is usually consisting of multi disciplinary specialist where their combined experience and capabilities can form a breakthrough in analysing and resolving business problems. Team problem solving is a potent way of focusing on quality issues and resolving them permanently. The same is true for energy issues.

The quality circle is established to follow-up and implements the recommendations of the energy committee. It discusses the actual status of energy consumption, and develops suitable solutions to problems and obstacles, to make sure that energy consumption is within the optimum range. The energy quality circle consists of the following:

- Energy engineer (chief/chair);
- Representative of operations from the production department;
- One member each from electrical, mechanical and quarry department;
- The circle holds a weekly meeting.
- The meeting agenda is prepared by the energy engineer. It undertakes the review of the recommendations of plant energy committee meeting, the recommendations of the previous meeting of the quality circle, the actual energy consumption of the main production units and any ideas or proposals, which may help, control and conserve energy in the plant.

- The energy engineer is in charge of writing and preparing minutes of meetings. A copy is distributed to each member, the energy section head and the chairman of the energy committee.
- Reports and results are documented in special files containing the account on activities of the circle.

The advantages of this committee was obvious, it reflect the positive and active participation of the middle management and factory staff in implementing the plans and the targets stetted by the higher management to achieve the target of decreasing the energy consumption in the factory, also it create team work style of handling the problem in the plant and also implementing the plans. It was instrumental in achieving the sited targets. As for the obstacles the middle management of the factory was not accustomed to accept the participation of the technical staff in implementing plans and discussing problems and getting the credit in participation in achieving the targets. The main reasons behind the success of this committee were the top management commitment to the energy conservation issues, the teamwork spirit and the accountability of the success and failure.

Figure (8.9) and table (8.5) summarise the functions and activities of the EQC.

Figure (8.9)
Energy Quality Circle FLOW SHEET

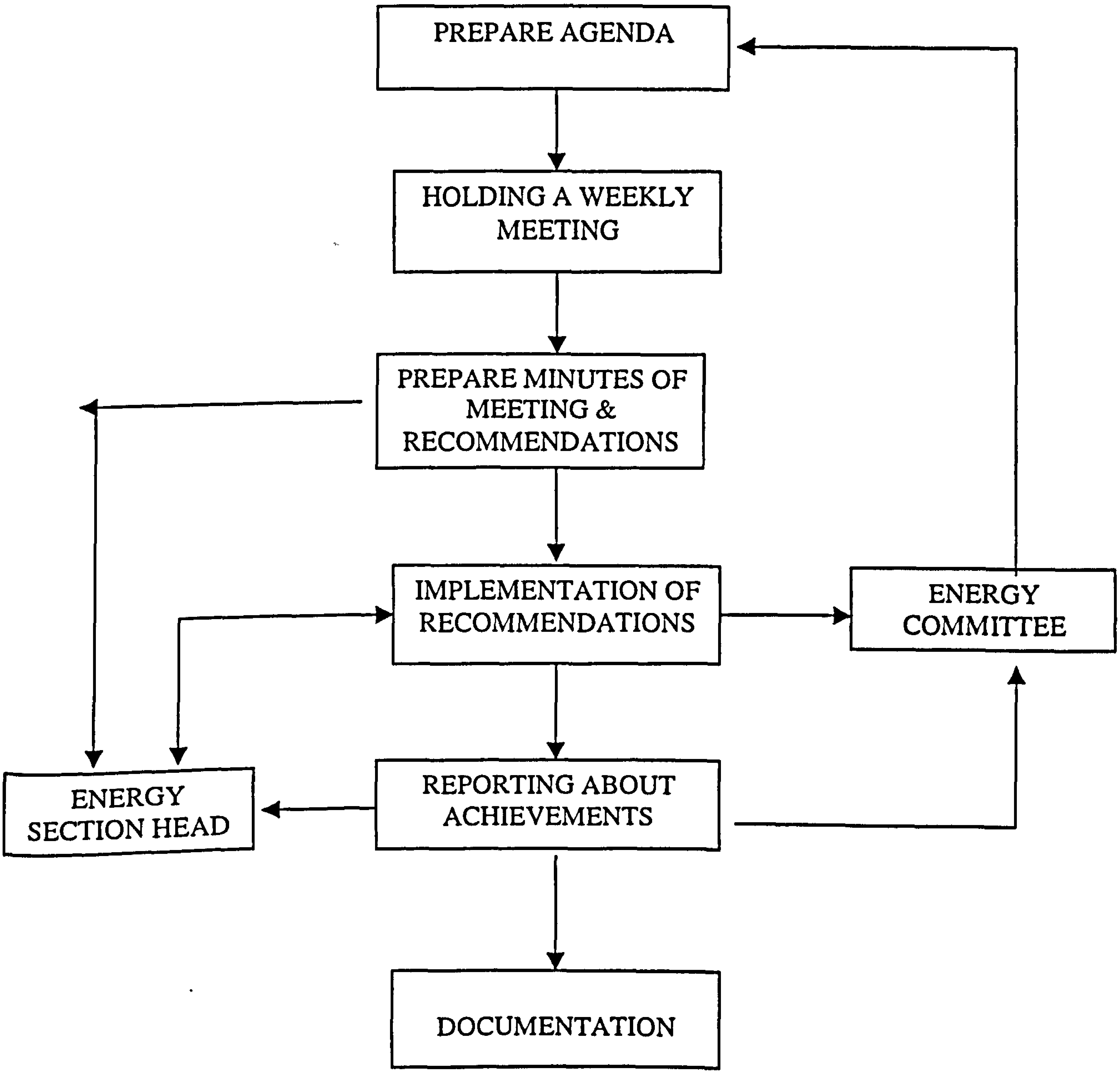


Table (8.5)
ENERGY QUALITY CIRCLE PROCEDURE

Procedure Name	Responsible Person	References
Prepare Agenda	Energy Engineer	• Minutes of previous meeting • Recommendations of energy committee • Notes about operation, round and check-up • Results of studies • Reports on energy consumption
Holding weekly meeting	Energy Engineer & the members	The weekly data of the meeting
Prepare minutes of meeting and recommendations	Energy Engineer	
Follow-up to implement the recommendations	Engineer and Tech.	Recommendations of the meeting
Write a report and distribute one copy to energy committee	Energy Engineer	
Documentation	Energy Engineer and Tech.	Energy quality circle file.

C) Control

C.1 Introduction

Control is the third element of energy management, after determining the goals, their planning and execution. The actual results should be measured and evaluated by comparing them to the goals and corrective action taken if necessary.

The control procedure must cover all the work and activities during planning and executing the energy plans and programmes.

The enclosed energy control procedure table (8.6) illustrates the efficient control system, which is an important part of the energy management system. It begins at controlling the energy consumption for all production units. The procedure includes studying and analysing the consumption figures, determining the deviations and their causes, suggestions, solutions, and recommending their implementation, with approval from the energy committee, by the members of the energy quality circle. Finally, achievements are reported to the energy manager and work manager. If any problems or obstacles arise, the necessary measures are taken to resolve them. The implementations of goals are controlled by energy management staff and by the concerned people via the energy committee. Reporting ensures that the highest level of management is always aware of the current situation related to energy in the company.

Considering the experience of Jordan Cement Factories regarding energy control system, it can be indicated that energy is controlled by the following:

1. The energy manager in the head office who is able to control the energy of the entire company.
2. The energy section staff in the plants who are able to control the energy within their own plants.

3. The plant's energy committee, who are able to control the implementation of plans, programmes and necessary adjustments within their own plants.
4. Daily, weekly and monthly energy reports that include consumption, cost, problems, obstacles, recommendations and achievements.

C.2 Energy Control Procedure

The following tables (tables 8.6 and 8.7) represent the energy control procedure in the JCF.

Table (8.6)
ENERGY CONTROL PROCEDURE

No.	Procedure Name	Responsible Person	References
1.	Preparation of energy consumption data and annual production plan	Production & quarry department	Production forms Quarry forms Annual production plan brochure.
2.	Preparation of energy annual plan	Energy section head	Job instructions No.2
3.	Collection of energy consumption data	Energy technician	Job instructions No.3
4.	Data insertion into computer and statistical analysis	Energy Engineer	Job instructions No.4
5.	Data saving	Energy Engineer	Job instructions No.5
6.	Implementation of recommendations after data analysis	Section Head	Job instructions No.6
7.	Plant energy committee & energy quality circle meeting	Section Head Engineer Technician	Job instructions No.7
8.	Preparation of weekly & monthly reports	Section Head Engineer	Job instructions No.8
9.	Energy technical studies	Section Head Engineer Technician	Job instructions No.9
10.	Instructions & suggestions leading to control & energy conservation	Energy Section Head	Job instructions No.10
11.	Documentation	Section Head Engineer Technician	Job instruction No.11

Table (8.7)
ENERGY CONTROL PROCEDURE JOB INSTRUCTIONS

No.	Procedure Name	Instructions
1.	Preparation of energy consumption data and annual production plan	Data & plan are prepared by production & quarry department
2.	Preparation of energy annual plan	Energy Engineer requests from production & quarry department the necessary data to prepare the plan. Data analysis. Suggestion of suitable measures to control & conserve energy. Discussion of suggestions with Work Manager & concerned Departments. Preparation of the annual energy plan. Printing & distribution of the plan. Implementation of the plan.
3.	Collection of the energy consumption data	The energy technician collects the data daily from production & quarry department and delivers them to the Energy Engineer.
4.	Data insertion into computer & statistical analysis	The data are input by the Energy Engineer. The section head & the engineer perform the data analysis. Determination of the deviation & the causes. Suggestion of suitable recommendations. Using control chart & Pareto diagram.
5.	Data Saving	After data analysis, the information is saved on the computer to be used in report writing, and studies preparation, etc.
6.	Implementation of suitable recommendations after data analysis	After data analysis, suitable recommendations are suggested to conserve energy. Energy Engineer & Technician implement the recommendations by taking necessary action.
7.	Meeting of plant energy committee & energy quality circle	The Energy Section Head participates in the monthly meeting of energy committee after agenda preparation. The Energy Engineer holds a weekly meeting for the members of quality circle.
8.	Preparation of weekly & monthly reports	The monthly report is prepared by the section head and the weekly report is prepared by the Energy Engineer
9.	Energy technical studies	The energy Section Head & the Engineer prepare the necessary studies concerning the energy issues from analysing the data, recommendation of energy committee & quality circle meetings, & also from the control of the production process, etc.
10.	Instruction & suggestions leading to control & energy conservation	After the implementation of the suggestions & instructions concerning the energy issue, the energy consumption & cost will decrease which indicates that energy use is under control.
11.	Documentation	All the energy section staffs document all reports, studies, researches, and minutes of meetings in order to get the necessary benefits in the future.

8.7 Energy Auditing in Cement Industry

8.7.1 Overview

According to good practice guide 186 prepared by Department of Environment 1997 organisation needs to audit energy performance. The pressures to do so come usually from two different quarters:

- The growth of in contracting out the management of energy consumption
- The rise of environmental management and reporting.

It is stated that there is no legal requirement for an audit of energy performance, but there are strong voluntary pressures to move in this direction arising both from BS 7750 and EC's Eco-Management and audit schemes (EMAS).

Energy audit is the process by which the effectiveness of conservation can be gauged. This can be accomplished by the energy audit is to quantify the energy used and the available possibilities for energy saving.

The availability of energy information data base is an important factor for any energy management plans or programmes, while the energy audit is the most important element of a successful energy management plan or programme since it determines the point and direction of progress for the whole programme.

The energy audit consists of two principal phases, as follows:

1. Auditing the quantity and cost of energy consumption for each production unit.
2. Auditing by observing the performance of production units from an energy-use perspective and identifying energy saving possibilities (plant audit).

The explanation of the importance of each energy-auditing phase is as follows:

A) Auditing Energy Quantities and Cost:

The first step is to collect data and perform analysis based on available energy consumption and cost records. In the cement industry, this audit is very important for the following reasons:

1. Determination of the total quantities of energy used (electrical and thermal), kWh and Kcal.
2. Determination of the total cost of electrical and thermal energy used.
3. Determination of the specific heat consumption (Kcal/kg clinker).
4. Determination of the specific power consumption (kWh / tonne)

B) Plant Audit:

After completing the audit of energy consumption and cost, it is necessary to make a plant audit. In the cement industry, this audit is important and necessary to collect information about energy consumption for every production unit in the plant.

The purposes of plant audit in any cement plant are the following:

B.1 Allocation of the energy consumption for each production unit:

- Energy consumption for quarry department including diesel requirement for trucks, explosive materials, etc.
- Energy consumption for crushers (kWh / tonne).
- Energy consumption for raw mill department (kWh / tonne).
- Energy consumption for burning department (Kcal /kg and kWh / tonne).
- Energy consumption for cement grinding department (kWh/ tonne).
- Energy consumption for packing plant department (kWh / tonne).
- Total specific power consumption for the whole plant (kWh / tonne).

- Total specific heat and power consumption for each plant (kcal / kg + kWh / tonne

B.2 Observing and follow-up on the performance of production units from an energy-use perspective:

It is important to follow-up on the performance of each production unit in the cement plant, and to evaluate the consumption of these units in order to propose solutions, which minimise energy consumption.

Observing the performance of each production unit consists of the following:

- Determination of daily, monthly and yearly production.
- Determination of running hours.
- Determination of production rate and comparison with design capacity.
- Determination of daily, monthly and yearly specific heat and power consumption.

B.3 Identification of energy conservation possibilities in the plant:

After determining the cost and quantities of energy consumption for each production unit, the allocation of each department, following-up, and observing the performance of the production units, it is possible to measure the effectiveness of energy saving measures already carried out, or to identify the opportunities for potential energy saving measures. In the cement industry, all efforts to save energy must be concentrated on the higher energy consuming equipment.

The following procedure indicates the possible means in identifying the areas where energy can be conserved.

- Collection of daily and monthly energy consumption for all plant production units.
- Analysis of the consumption figures.

- Determination of deviations in comparison with design capacity.
- With the use of a Pareto-diagram, determination of the production unit or machine that has the highest consumption of energy.

Proposals for suggestions, instructions, and modifications to reduce the energy consumption, starting with highest energy consumption among all machines and equipment.

Follow-up on implementation of these proposals.

Identification of achievements and reporting to plant manager and energy manager.

B.4 Measuring plant energy balance:

In the cement industry it is important to plan execute energy measurement from time to time, in order to have energy balance (heat and mass balance) to obtain a detailed picture of the process, determine the energy flows and to highlight the weak points, and heat losses.

The requirements, which must be fulfilled in order to attain correct measurement results, include:

- Use of reliable instruments.
- Use of appropriate methods.
- Selection of a time when a representative measurement result may be obtained. The plant should be running undisturbed for a long enough time to be in equilibrium before measurement activation begins.
- Correct processing of the results.

The measurement results should be studied carefully. Countermeasures should be taken to improve the energy use and to reduce the energy losses.

Figure (8.10) and table (8.8) show the procedure for measuring the energy balance of the plant.

Figure (8.10)
Procedure of Measuring Plant Energy Balance

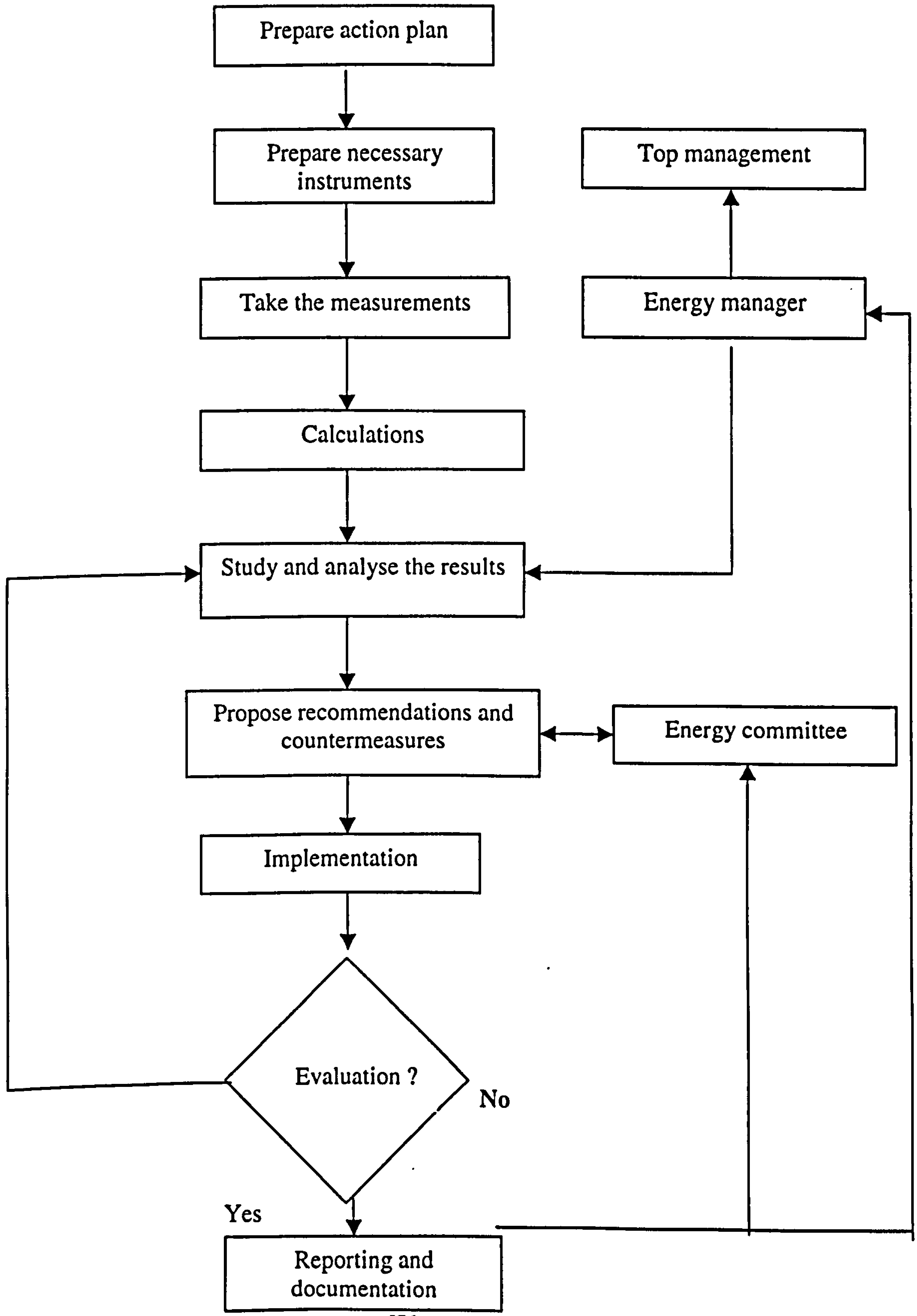


Table (8.8)
Procedure of measuring plant energy balance

Serial No	Procedure Name	Responsible person	References
1	Prepare action plan	Energy engineer and technician	- List and required instruments - Work procedure
2	Prepare necessary instrument	Energy engineer technician	
3	Take the measurement	Energy section head Energy engineer	
4	Calculations	Energy section head Energy engineer	
5	Study and analyse the results	Energy section head Energy engineer	
6	Propose recommendations and countermeasures	Energy engineer Energy section head Energy committee	
7	Implementation	Energy engineer Energy section head Energy committee Energy quality circle	
8	Evaluation	Energy section head	- Targets of the recommendation
9	Reporting and documentation	Energy engineer	

8.7.2 *Energy Auditing Procedure in Cement Industry*

According to Kodak Limited statement "We appoint an independent energy auditor from within the company to visit parts of a process and calculate how much energy we ought to use, then measure how much energy we do use, then we work out what to do to match the two" which was reported in good practice guide 169 prepared by Department of Environment 1995. The previous statement provides a simplistic approach to the audit procedures.

In the proposed EMS, the energy manager and energy section head must prepare all the audit reports concerning energy issues at weekly and monthly intervals with the following procedures:

1- Daily production reports received from production and quarry departments.

2- Daily and monthly thermal and electrical energy consumption received from production, quarry and electrical department.

3- Analysis of these reports and comparison with the annual plans to demonstrate the deviations.

4- Preparation of a weekly report consisting of the following:

- The deviations on the energy consumption.
- The recommendations and achievements of quality circle meetings.
- Graphs to demonstrate the actual energy consumption for all the production units.
- Recommendations concerning control and energy conservation.

5. Preparation of a monthly report consisting of the following:

- Heat energy consumption for all production units and deviations thereof.
- Power energy consumption including calculations of the specific power consumption, consumption on maximum, daily demand and night utilisation.
- Total cost of the energy consumption.

- Average cost per tonne.
- Graphs to demonstrate the performance of all the production units.
- Achievements of energy section for saving and controlling energy consumption.
- Plant energy committee minutes of meeting.
- Reports and achievements of energy quality circle.
- General recommendations to overcome all the problems that minimise energy consumption.

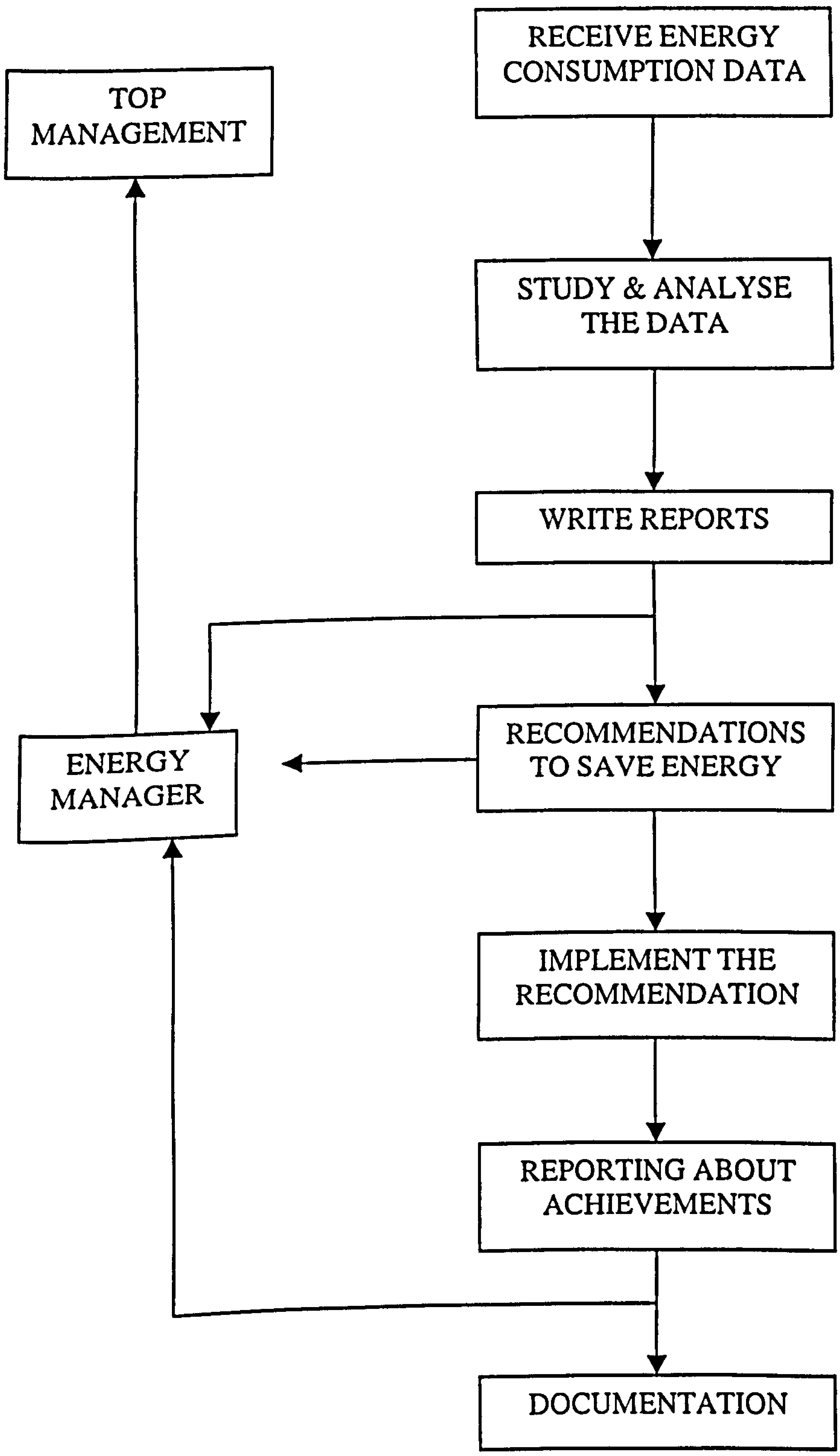
6. One copy of these reports to be sent to the works manager, one to the energy manager, and one copy of the document to be filed with other reports.

Table (8.9) and figure (8.11) represent the functions and activities of the energy-auditing task.

Table (8.9)
ENERGY AUDITING PROCEDURE

S.No.	Procedure Name	Responsible Person	References
1	Daily and monthly reports on energy	Energy Technician	Quarry and production report
2	Analysis of the report figures	☐ Energy Manager ☐ Section Head ☐ The Engineer	Commissioning figures
3	Weekly, monthly reports and clear the deviations.	☐ Energy Manager ☐ Section Head ☐ The Engineer	
4	Printing and distribution of the report including recommendations to control and save energy	Energy Engineer	
5	Implementation of the recommendations of the reports	Energy Engineer and Technician	Weekly and monthly reports.
6	Achievements report	Energy Engineer	Results of Implementation
7	Documentation	☐ Energy Manager ☐ Section Head ☐ The Engineer ☐ Technicians	

Figure (8.11)
ENERGY AUDITING FLOW SHEET



8.8 *Implementing Quality Assurance Systems and Total Quality Management*

The management model used and the basic theme of the present research is quality-oriented, using principles and techniques of Total Quality Management (TQM). The EMS developed in this thesis is integrated within the overall concept of TQM. Therefore, the following two sections describe the quality assurance and total quality management systems. Quality systems have been implemented around the world for some years now and are therefore only described in outline here, with specific issues arising from the Jordanian cement industry highlighted. The reader is referred to Oakland (2000), and Lindsay and Petrick (1997) for further information on TQM and Organization Development.

According to statement in good practice guide no. 119 by Hepworth Refractories " Energy management really is no different from the normal practice of the management, the principles of TQM simply highlight this". Another statement in the same guide by Arjo Wiggins "we managed energy like any other business resource - labour, downtime, material, there is really no difference in our approach." Another statement from the same good practice guide no. 119 prepared by UK Department of Environment "TQM places energy management fairly and squarely in the business arena, and energy managers in a TQ organisation will find themselves discussing business issues as will as technical ones". According to United Glass, they stated in the above guide " we needed a vehicle for preaching the world on energy efficiency. It made sense to utilise company initiatives on quality to achieve our aim of managing energy better."

All these previously mentioned statements proved without any doubt that there is a trend in the industry to utilise the total quality management approach in implementing energy management systems and practices. We were utilising the TQM concept in implementing the energy management and we used it in this research without knowing it is used in other places uptill we found the very useful documents (good practice guides which was prepared by Department of Environment in UK which use in some part of it similar approach to our energy management model.

8.8.1 Quality Assurance System ISO 9000

1. Introduction:

   ISO 9000 is a written standard that defines the basic elements of a system which companies use to ensure that their products and services meet or exceed customer expectations. It is brief and simple.

   ISO 9000 aims at creating and continuing organisational success by providing mechanisms for determining and fulfilling customer needs, preventing errors where possible and correcting any such errors in such a way that the process is improved.

   This system can be integrated with the energy management system and towards achieving the main objectives of reducing the energy consumption.

2. Quality management system structure:

   This structure allows company personnel to find their way around relevant procedures easily. (Sadgrove, 1994). This system consists of the quality policy, quality manual, procedures, work instructions, reports, forms and records.

3. Quality Audit:

   Quality audit is a systematic and independent examination to determine whether quality activities and related results comply with planned arrangements and whether these arrangements are implemented effectively and are suitable to achieve objectives.

4. Consequences of implementing an ISO 9000 quality system:

A quality system establishes and enforces consistent methods and quality controls throughout the organisation.

5. Process of continuous improvement:

Implementation of ISO 9002 quality assurance system results in a continuous process of maintaining and improving quality. It is for all responsible persons of the company to create an environment for quality improvement.

6. More And More Satisfied Customers:

Dependability and reliability, high and consistent quality of products meeting the latest standards, professional handling of customer inquiries and orders, customers satisfaction, growth in confidence, improved and better customer supplier relationship.

7. Aim For Zero Defect Product:

A benefit of zero defect production is that there is less waste of raw materials, less wastage of energy, and a reduction in waste disposal needs. All of these factors lead to improved performance indicators.

8. Effect of Implementing Effect of Energy Management on JCF as Part of the Quality System (TQM)

- Effect of implementing quality assurance system in conjunction with EMS on the variable cost of Rashadiya cement plant including energy cost.

The following table shows the reduction of variable cost /tonne of cement production in Rashadiya plant for the last three years after implementing the EMS, which is one of the most direct benefits for any company.

Year	Variable cost JD/tonne
1996	18.44
1997	17.77
1998	15.82

Source J.C.F Rashadiya plant

□ Effect of implementing ISO 9002 in conjunction with energy management system.

The effect of implementing ISO 9002 on quality control is illustrated by the compressive strength of cement produced at Rashadya (as per British Standard BS for 28 days).

The mean value of compressive strength for two years is considered, 1996 prior to implementing the ISO 9000 system, and 1999 after the ISO 9002 system has been implemented and tested.

The mean value of strength (N/mm^2) decreased from 55.2 in 1996 to 48.2 in 1999both within the required specification. This has a direct bearing on the cost especially for electrical energy cost, since after the implementation of the procedures and instruction the cement was produced with less energy. And this proves the integrated relation between the energy management system and the quality management system.

8.8.2 Total Quality Management

1. Quality in JCF

   Quality has been a natural aspect of JCF operations since 1990, and it is integrated into everything that is done to keep business development. Quality is important to JCF competitiveness with others, especially when the company will loose monopoly of the local cement market in 2001.

2. Commitment of Top Management

There is a clear commitment of JCF top management to quality. According to the good practice guide no 169 prepared by UK Department of Environment (1995) ," Top management team work aim to reduce unproductive conflict and increase overall commitment to business strategy and goals. The leadership given by top management and their style of management, is one of make or break components of a TQM program". JCF organisation shows that top management is truly concerned with developing the business by introducing into the structure activities for inspection, control, planning, assessments, improvements and development, and training.

3. Culture change

JCF top management lead a process for culture change. It is obvious that they have succeeded in this and that JCF staff are now more concerned about internal as well as external customer requirements, which they intend to meet. In keeping with many other practitioners of quality systems we have found that the culture change has been very marked and has beneficially influenced all aspects of the operation of the manufacturing plants.

4. Improvement

Best results in improving work at JCF are achieved by implementing one improvement project after another, i.e. (quality awareness, review of organisation, ISO 9000, ISO 14000, upgrading of plants, three years technical plan ...etc.).

5. Quality circles in JCF

Teams or circles for quality in JCF are widely used to solve problems or improve activities. According to the good practice guide no 169 prepared by UK Department of

Environment, "Team problem-solving is a potent way of focusing on quality issues and resolving them permanently. The same is true of energy issues. Again, the ability to use systematic problem-solving techniques is a must for the manager of the future."

Quality teams or circles in JCF are distinguished by:

- ☐ Using TQM methodology for improvement.
- ☐ Using group techniques to generate ideas for improvements (Brainstorming).
- ☐ Wide involvement of staff in decision-making.
- ☐ Using a systematic approach to solve problems:

- Detecting problems by product inspection feed back using statistical control chart.
- Diagnosis by set priorities (Pareto chart) and root - cause analysis (cause – effect diagram).
- Experiment (i.e. confirm or reject hypotheses).
- Evaluating possible solutions for improvement purposes (i.e. cost, impact, feasibility and interactions). According to the good practice guide no 169 prepared by UK Department of Environment (1995), "In improvement phase you can establish energy as an area of waste which involves everyone. There are opportunities to link energy management into your improvement program".
- Choosing and implementing solutions. For an example of such a solution and According to the good practice guide no 169 prepared by UK Department of Environment (1995), "Improving reliability had an impact on energy efficiency because it reduced the energy required to keep restarting our more heavy machinery.
- Track results
- Proceduralise.
- Transfer learning.

- Using seven tools (Pareto Chart, cause and effect diagram, check sheet, control chart, histogram, scatter diagram and operation chart).

The EMS as an integral part of the TQM will be instrumental in achieving the strategic objectives of JCF related to energy consumption. The energy management factors (availability, production rate, and number and duration of stoppages), which are called high level factors, are be monitored through the EMS towards achieving the strategic objective, which is to reduce energy consumption. From the other angle the implementation of TQM enhances the effectiveness of EMS as it helps in improving the above mentioned energy management factors.

TQM incorporating EMS in the cement industry as described here has taken a great deal of time, money and commitment. The decision to introduce it was not fully endorsed at the outset. Indeed there was substantial resistance to the concept. It was seen by some as a deflection for the main purpose of the company. It would save finances and people's time. Using the declaration of commitment from the top, and the internal communications suggested by the quality standard TQM was introduced from 1995. Communication with staff and the views of staff is probably the most important aspect of the successful application of the quality system.

Respect for people is important: Roberts and Sergesketter (1993) claim that quality is personal, placing the responsibility on the individual. The experience at JCF on the other hand is much more in agreement with Deeming's (1993) view that quality problems are the fault of the system 85 to 90% of the time. Staffs are valued for their contribution to the whole operation of the company. Attitudes within the workforce have noticeably changed, although we did not monitor this explicitly the atmosphere and culture of the company has improved and with it the productivity of the manufacturing facilities. It is the teamwork enshrined in the TQM system that both gains the credit and carries the blame for events connected to production.

By incorporating the energy management function within the TQM we have been able to utilise the management structure of control, communication, feedback, improvement, record keeping and auditing etc to target and monitor energy performance improvements. Measurement is critical. We have found that measurements of factors such as energy consumption by individual items of plant, raw material used, water consumption etc provides the key information for decision making at local level. This was also the conclusion reached by the authors of Guide 169 (1995) who report that BP chemical have 450 data points across one of their major sites.

Despite the early reservations of some employees, the TQM combined with the EMS have been successful. Apart from better internal and external relations, the product is now more closely monitored and controlled and the consumption of energy per unit output has been reduced, giving both financial and environmental rewards.

8.9 Case Studies in Energy Management

The following two case studies show the role of the EMS and indicate its importance and influence on the productivity and profitability of the Jordan Cement Factories. It should be noted that the case studies here differ from the ones in Chapter six in that they concentrate more on management aspects rather than engineering aspects (with which Chapter 6 was concerned).

The examples of Rashadiya and Fuhais plants are presented to illustrate the role of energy management systems in energy conservation in the cement industry. In the Rashadiya plant, an optimisation plan was implemented where a decrease in specific power consumption (consumption per unit of cement produced) was achieved through the increase of production rate. In the Fuhais plant, an optimisation measure was introduced to the cooler to decrease the specific heat consumption by improving the cooler efficiency and heat recovery, which led to an increase in the kiln production rate. The two examples show that efforts must be concentrated on fields where the energy consumption is high relative to other functions.

An important aspect of energy management is, in co-ordination with the firm's top management, setting goals and objectives in an optimal manner. The elements of energy management are restated as: determination of goals, planning, execution and control to achieve desired results, as exercised in building EMS. The following examples show the role and EMS of energy management in the Jordan Cement Factories.

8.9.1 The Role of Energy Management in Reducing Energy Consumption

In the Rashadiya plant there are 4 identical cement mills. Specifications of these mills are shown in table (8.10) below.

Table (8.10)

CEMENT MILLS SPECIFICATIONS

No.	Item	Unit	Description
1	Production Rate	T/h	85
2	Absorbed Power	KWh	3200
3	Specific Power Consumption	KWh/T	37.64 for main drive only
4	Fineness	Cm^2/g	3200
5	Materials	%	Clinker 95% + Gypsum 5 %
6	Steel balls weight	Tonne	- First Compartment 64 Tonne Second Comp.166 tonne
7	Shell Liners		-First Comp: Step type - Second comp.: classifying type, boltless segregation lining
8	Diaphragm		-12 segments slit diaphragms first comp. Slit width 13 mm slit opening 30% second comp. Slit width 10 mm slit opening 30%
9	Effective length	Meter	First comp. 3.613 Second comp. 8.515

Grinding of cement consumes approximately 45% of the total power required for the cement industry. Effort was thus applied to this process to minimise energy consumption by decreasing specific power consumption (kWh/tonne).

The first step was to determine the goal. The energy committee in the plant set the goal after the necessary studies and research had been conducted by the plant energy section, through correspondence with specialised companies. The decision was made to optimise the mills by making internal modifications only. Table (8.11) shows the optimisation parameters. The internal modifications included those measures that were applicable to the plant and were carried out for short periods of time for the purpose of the investigations. They required little expenditure of capital, if any. These modifications were termed optimisation parameters since it was intended that they would optimise production and energy consumption simultaneously.

Table (8.11)

THE OPTIMISATION PARAMETERS

No.	Item	Unit	Present Operation Figures	Target Figures
1	Production rate	Tonne/Hour	85	100
2	Absorbed power	KWh	3122	3250
3	Specific power consumption	KWh/tonne	36.7	32.50
4	Fineness	Cm2/g	3200	3200

It can be seen from table (8.9) that the production rate would have to be increased by approximately 17% while achieving a decrease in specific energy consumption by approximately 13%.

Planning: After determining the goal, the second step was planning. This process began by demonstrating to top management the benefits that could be attained through realising the established goal. A feasibility study on the project was prepared and the payback period was determined. The payback period is defined as the number of years that a firm can recover certain investment outlay from expected revenues and can simply be calculated as follows (Weston and Brigham, 1978):

$$\text{Pay back period} = \frac{\text{Investment Cost}}{\text{Expected Annual Revenue}}$$

In other words the pay back period is the number of years (or months) when the expected revenues become equal to the investment cost. The shorter the payback period, the more attractive the project is.

After receiving approval from top management, the necessary tender documents were prepared. These included all necessary requirements and special conditions, specifications of the existing cement mill, drawings, erection and supervision, guarantee, and the cost of the optimisation (quotation). The tender documents were prepared in November 1994. After offers had been received, a special committee selected the most suitable one.

The selected offer covered the following measures:

1. Replace the shell liners of the first compartment with a new lifting type.
2. Replace the whole intermediate diaphragm with a new type (flow control diaphragm).
3. Replace the outlet diaphragm with a new slot design.
4. Enlarge the first compartment by 1/2 meters.

The winners guaranteed to increase the mill output to 100 tonne/Hour and reduce the specific power consumption to 32 kWh/tonne while keeping the fineness unchanged at 3200 cm^2/g. After determination of the winner, a plan was prepared to determine the exact time schedule, the description of the work, and requirements of manpower, tools and machines.

Execution: After the plan had been prepared, and the execution date set, and all spare parts on the plant secured, a committee from all concerned departments was formed to follow-up on the optimisation and maintenance of the mill. The maintenance and upgrading of the mill was done according to the plan in three weeks during September of 1995.

The hardest and longest job was the one performed inside the mill. It involved changing the shell liners (the life-time of the old liners was 33000 hours) of the first compartment and removing the old diaphragm including the frame to install a new segment type (flow control diaphragm), enlarging the length of the first compartment by 1/2 a meter, removing the outlet diaphragm to install the new type, and design of diaphragm. Work inside the mill was carried out 24 hours/day utilising eight fitters and four welders. The execution of this task was done under the control and supervision of works manager, production and mechanical managers and with the presence of an expert sent by the suppliers.

Control: Each phase of this optimisation from goal determination, convincing top management, planning for goal achievement, execution, and finally the evaluation process was carried out by or under the supervision of energy management staff with the energy conservation manager in the head office coordinating and cooperating with plant energy section staff. This was done by writing and preparing the suitable reports, studies and recommendations in each phase of this optimisation.

Evaluation: After completing the optimisation execution and restart-up the mill with the new modification, an evaluation process was undertaken, in order to assess the results of the optimisation. The following table (8.12) shows the results.

Table (8.12)

ACTUAL RESULTS

	Item	Unit	Operation Data
1.	Production rate	Tonne/Hour	98
2.	Absorbed power	kWh	3214
3.	Specific power consumption	KWh/tone	32.79
4.	Fineness	cm^2/g	3100-3292

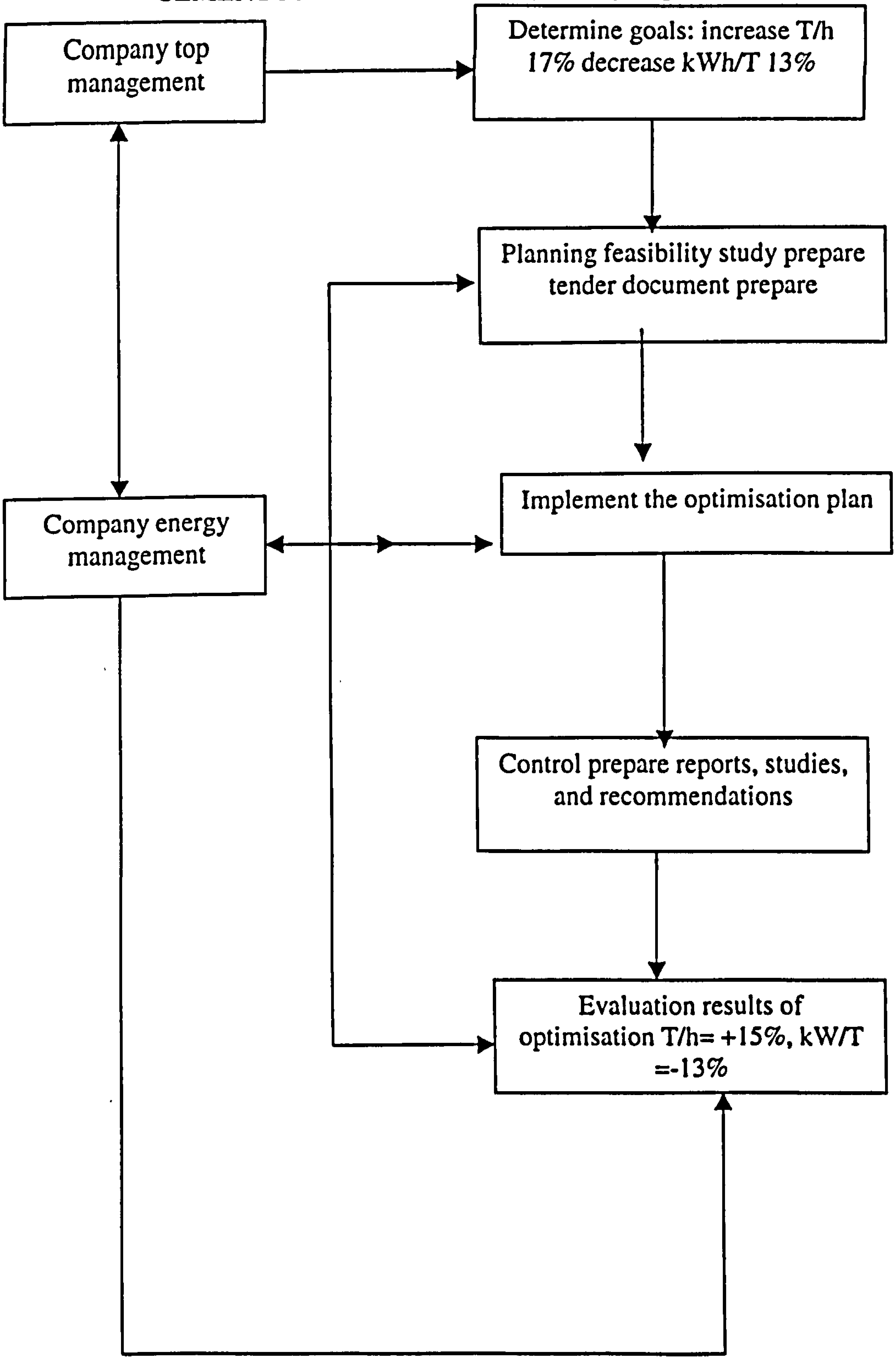
From table (8.12) it can be seen that the production rate did not reach 100 ton/h because there were still some problems to be resolved in order to obtain the demand rate. Nonetheless, it can be seen that the specific energy consumption was reduced by approximately 13%, while the production rate increased by 15%. These were considered as good and acceptable results. Yearly benefits from conservation of energy according to the 1996 plan equal approximately 96000 JD/year in comparison with the case of grinding the same quantity of cement in the mill before the optimisation. The payback period according to the investment cost and the gross earning was approximately equal to 2.5 years.

There have been other benefits of this optimisation, as follows:

1. The lifetime of the mill shell liners was almost finished and it had been intended to buy a new set of liners. This project saved on the cost of new liners.
2. The new type of first compartment diaphragm is a segment type. In case of wear or cracks, it is possible to change the broken segment only. Again this saved the cost of two new sets for the first compartment and second compartment diaphragm.
3. This optimisation reduced the energy power consumption of the cyclone separator, because with the new design the speed of the separation was reduced.
4. This optimisation reduced the need to operate the mill in the day time and peak hours.

After this successful project had been carried out as a part of the requirements of the company's energy management system, it was decided to optimise another mill during 1996. This means another goal was determined from the evaluation of the previous optimisation. Therefore, the interaction of the three elements of management is evident. As the implementation of the programme proceeds and results are evaluated, plans for saving energy are often modified and new goals may be set. A successful energy management programme or plan requires continual planning, execution and control as shown in figure (8.12).

Figure (8.12)
CEMENT MILL OPTIMISATION FLOW SHEET



8.9.2 Management Procedures and Countermeasures Taken to Reduce Energy Consumption

This case study analyses the cause of heat and power energy consumption exceeding the design and commissioning specifications in JCF plant, the cause of which could very well be that management initiatives and measures were not implemented previously. In other words, to find the cure for an ailment, it is necessary to determine its cause.

In the cement industry, two main factors affect energy use:

1. Production rate: must be maximised.
2. Availability: must also be maximised through elimination of emergency and other stoppages.

A number of management initiatives and countermeasures were applied to achieve certain set targets for the above mentioned two factors. It should be mentioned that the results of adopting and applying such management initiatives and measures would only be apparent after a few years of operation. However, from the experience of other cement industries, the effectiveness of such systems is a well-established fact.

1. Increasing maintenance and equipment efficiencies:

   From sections (6.5 and 6.6) in Chapter six, it can be seen that the decrease in maintenance and equipment efficiency is responsible a 22% increase in heat consumption and a 15% increase in power consumption because it leads to a decrease in production rate and more frequent equipment stoppages.

   To minimise the impact of the identified cause, a two fold approach was followed. The first concentrated on improving maintenance, and the second aimed to increase equipment efficiency. Improving the maintenance practice was accomplished through the following measures:

- Concentration on training including field training, lectures and participation in training courses inside and outside Jordan for the maintenance staff.
- Implementation of a new maintenance management system.
- To improve maintenance work, a quality committee was created to follow-up on the implementation of the correct maintenance procedures and instructions and thus ensure that the equipment are maintained in the proper way.
- Preparation of a job standard for each maintenance job to help the technician perform maintenance in the correct manner.
- Preparation of a quality manual for each section and department (required for ISO certificate).

To increase the equipment efficiency the following steps taken:

- Conducting needed studies and research to analyse causes of lower efficiency and implementation of recommendations of these studies.
- Making essential modifications on important equipment to increase efficiency, such as the cooler in Fuhais plant.
- Changing low efficiency equipment with new and more developed substitutes, such as the root blowers of Rashadiya plant.
- Preparing a plan for bringing experts from manufacturers to inspect and check-up the main equipment and evaluate its current condition.

The two approaches were adopted as preliminary investigations showed that they are most effective. Moreover, these two measures must be carried out anyway.

2. Increasing operation efficiency:

The operator plays an essential role in maintaining maximum production rate and in minimising stoppages through taking the correct action at the right time. Such responsibilities require qualified operators.

According to the analysis in sections (6.5 and 6.6), this aspect is accountable for 18% and 12% of the reasons for heat energy and power consumption respectively. It is possible to overcome the influence of this factor at relatively low cost compared with the previous item.

The following procedures were taken:

- Training operators as follows:

  On field training, purchasing simulators, training inside and outside the country including organising visits to similar plants, holding lectures on operation procedures and problems, distributing brochures and books, and preparing an annual training plan for 1996.
- Motivating operators through the introduction of an incentive system.
- Increasing the control system on the operators and on the production process by creating an auditing system.
- Purchasing new instruments to help the operator control the system such as the kiln scanner.

3. Planning:

To increase the availability in the cement plant by eliminating emergency stoppages, it is important to prepare annual plans for production, maintenance, and spare parts supply.

The allocation of missing efficient planning is responsible for approximately 30% of the reasons for the variation in power energy consumption because any stoppages for any equipment will create a necessity to operate in the daytime period and sometimes during peak hours. In addition, especially for the kiln, the equipment will run for a long time unloaded. For heat energy consumption, this percentage is approximately 9%. Therefore, to eliminate the influence of improper planning, the following procedures were taken:

- Create a central planning department, responsible for preparing annual plans for production, maintenance, spare parts, projects and follow-up the implementation of these plans, control and auditing the performance of the production units.
- Study carefully the reasons of each stoppage, propose suitable solutions.

4. Quality Control:

The lack of quality control has a great impact on both heat and power consumption. It is responsible for 12% and 2% of the reasons affecting heat and power consumption respectively. The main problem arising from the lack of control on the raw mixture is related to the combustibility process. Lack of control over the raw mixture may lead to the disruption of the entire production process and, thus, cause time delays. To compensate for the lack of quality control, the following has been done:

- Introduction of a new data management system to record and present quality and process information.
- Training of quality control staff.
- Purchase of new apparatus and instruments to increase the effectiveness of the control system and minimise the standard deviation.
- Improvement of the sampling system by increasing the number of samples/day and collecting samples from new points along the production process.

5. Air Infiltration:

Air infiltration is an important factor for both heat and power consumption. The presence of false air in the system requires more heat energy to maintain the desired temperature, especially within the burning department. Consequently, increased power energy is required to sustain the increased load on the fans. This, in turn, leads to a reduction in the production rate. The contribution of air infiltration to heat and power energy consumption is about 5% and 5% respectively.

The following has been done to minimise the influence of air infiltration:

- Involvement of technicians measuring the flow in each part of the plant and determining the places where air infiltration is present.
- Purchase of special instruments to aid the technicians.
- In each maintenance schedule, holes and cracks on the outlet shell were welded. Manholes were closed tightly to prevent any infiltration.
- Daily auditing on fans power consumption to ensure a normal load and corrective measures taken as necessary.

6. Measuring Instruments:

From the analysis in sections (6.5 and 6.6) in Chapter 6, it can be inferred that errors due to the measuring instruments are responsible for the causes of the increase in heat and power energy consumption by 10% and 3% respectively. These instruments and their readings are vital if the operator is to perform his task correctly.

The following countermeasures have been taken to minimise the effect of this factor:

- A quality system manual prepared, which includes job procedures and instructions.
- Continued calibration and maintenance of the instruments according to a schedule.

- Purchase of new instruments e.g. the kiln scanner to facilitate the control system. The new stationary kiln instrument (which controls the kiln shell temperature) is more efficient than the old movable system that constantly needed part replacement.

7. Spare Parts:

To minimise the emergency and non-scheduled stoppages, spare parts must be available of good quality and in sufficient quantity to achieve maintenance at the proper time. The contribution of this factor is 6% and 2% for heat and power consumption respectively. To minimise the effect of lack of spare parts, the following steps have been taken:

- An annual plan prepared for spare parts according to the estimated maintenance plan.
- A plan developed to order and buy the spare parts according to their type, quality and installation area.
- The quality of the purchased spare parts, and specifically the lifetime, determined. A file was created for each company containing the name, quantity, lifetime, and other important factors.

8. Blending Silo Level (stock):

The effect of this factor on heat energy consumption is 3%. In the event of the raw mill being at halt for an extended period of time, the stock in the blending silo decreases. In turn, this leads to a decrease in the kiln feed rate and perhaps shut down of the production process. Therefore, it is important to have adequate stock to sustain a consistent operation. The countermeasure to minimise the influence of this factor is to ensure that the production rate of raw mill is always high and the availability is high.

9. Grinding hot or poor clinker quality:

To ensure that the cement mills are not fed with hot or poor clinker quality, which affects the mills operation (4% of power energy) the following have been introduced:

- Increased control of laboratory on the clinker production and extraction to mills.
- Adequate availability of clinker.
- A clinker transport system to the clinker store, especially transport over the store itself, so eliminating the need to the clinker in an outside field.

10. Effect of Management:

This factor is important in determining the performance in the company. The elimination of all management obstacles will lead to the production of cement at low costs and with a high productivity level. In this regard the following measures have been taken:

- Motivation: a new motivation system was created in the company to minimise carelessness among all employees, especially operators, technicians, and labourers, and to increase their loyalty.
- Determination of the organisational structure of all the departments and the job descriptions for all the personnel.
- Preparation works for obtaining the ISO 9002 certificate, and to develop the company management system.
- Delegation of power and authorities in order to create non-centralised management.

The above case study proves the importance of using certain statistical tolls such as cause and effect diagram (fish-bone) etc to analyse the factors affecting energy consumption and the function necessary to remedy their effects.

8.10 Cost Benefit Analysis of Energy Management System

The experience in Jordan factories (Rashadiya plant) shows that implementing this management system for the last few years reduced the energy cost (electrical and thermal). It is also important to say that the commitment of top management helped in achieving these objectives.

The following examples show the effect of energy management system in reducing energy cost:

8.10.1 Specific power consumption for the whole plant

Table (8.13) shows the specific power consumption for the Rashadiya plant for the last 4 years from 1995 to 1998 and 8 months of 1999.

Table (8.13)

Evolution of Specific Power Consumption

Year	Specific power consumption kWh/tonne	Cost JD/tonne Cement *
1995	111.22	4.516
1996	106.56	4.326
1997	105.24	4.272
1998	102.02	4.142
First (8) months of 1999	101.20	4.108

*Electrical tariff 0.045 JD/KWH for day & 0.023 JD/KWH for night tariff. (1995) average (0.0406) JD/KWH

Table (8.13) shows that the specific power consumption decreased by about 10% during a period of four years, which is considered a significant reduction. It is the result of implementing energy management system in Rashadiya plant.

8.10.2 Specific heat consumption

Table (8.14) shows the specific heat consumption for the years 1995-1998 and the first 8 months of 1999 for the Rashadiya plant.

Table (8.14)

Evolution of Specific Heat Consumption

Year	Specific Heat Consumption Kcal/kg	Cost JD/ton Clinker *
1995	888	5.89
1996	855	5.72
1997	858	5.74
1998	873	5.85
First (8) months of 1999	832	5.57

* Fuel oil price varied between 65 JD/tonne to 72.5 JD/tonne during the period

Table (8.14) shows that there is a continuous improvement in specific heat consumption especially for the last months, as a result of implementing energy management system, which means reducing cost of production especially the energy cost.

Finally table (8.15) shows the energy cost (electrical and thermal) for the years 1995 –1998 and first 8 months of 1999.

Table (8.15)
Total Energy Costs

Year	1995	1996	1997	1998	1999 (first 8 months)	Cost difference
Cost JD/tonne	10.406	10.046	10.012	9.992	9.678	0.728
Cost JD/tonne adjusted for inflation	10.406	9.477	9.158	8.87	8.341	2.065

The figures in tables (13, 14 and 15) show that after implementing energy management system in Rashadiya Cement Plant, the plant performance has been improved. As manifested by a reduction in the energy cost by about 7.5% which is equivalent to 1.45 MJD, for full production of the Rashadiya plant with 2 M tonne production capacity. Therefore, the annual savings for JCF production capacity of 3.5 M tonne will be 3.5 M tonne X 0.728 JD/tonne = 2.548 MJD = 3.588 MUS$(calculations based on constant energy prices). After adjusting for inflation the total annual saving will be 3.5 M tonne X 2.065 JD/tonne = 7.228 MJD = 10.325 MUS$ (this is a theoretical calculation because energy prices in Jordan is not following the free market prices and it is managed and partially subsidised by the government. Thus, implementing energy management system in any firm or plant decreases the energy cost, and hence the total cost, making the company competitive with other companies as a result of an increase in profits.

The above mentioned important results confirm the clear benefit of the EMS at JCF. When the management model was being developed the high-level management factors (availability, production rate, and number and duration of stoppages) were the main input in the EMS. The EMS was designed to control these factors towards achieving the strategic objective of reducing energy consumption through some intermediate functions and low level measures. To demonstrate the functionality, effectiveness and strength of the EMS table (8.16) relates energy consumption and cost to the high-level management factors as applicable to the Rashadiya plant.

Table (8.16)
Relationship Between Energy Consumption and Cost and High level Management Factors

		1995	1996	1997	1998	1999 (8 months)
KWh/tonne		111.22	105.56	105.24	102.02	101.20
Kcal/kg		888	855	858	873	832
Total cost JD/Tonne		10.406	10.046	10.012	9.992	9.678
Availability 2 kilns	1	89	90	91	95	97
	2	90	90	93	91	94
Loading rates T/D	1	2816	2932	3073	3021	3096
	2	2883	2982	2.985	3010	3090
No. of stoppages	1	62	107	96	71	14
	2	72	97	110	87	15
Duration of stoppages	1	740	712	688	261	180
	2	735	730	750	270	236

It is clear from the above table that the improvement of the independent variables reflected positively on the dependent variables, which are the specific electricity consumption and specific fuel consumption. This strongly confirm and support the findings of the statistical analysis and demonstrate and confirm the importance of implementing an energy management system.

8.11 Conclusions

This chapter has dealt with the details of developing a coherent energy management model whose objective is to establish transformational management process of certain high-level management factors into daily operations and controls. The high-level management factors are the same factors used as independent variables in the statistical and economic models of Chapter 7, which statistically proved to be the major factors affecting the energy consumption at JCF.

The chapter has presented a detailed analysis of the organisational and procedural aspects of energy management with concentration on management functions; especially planning, controlling, executing, organising, and auditing. A detailed mapping and analysis of these functions as the main components of the EMS resulted in establishing job descriptions, organisational charts, work instructions and procedures for all important functions of the EMS.

The chapter demonstrates that the EMS was built as an integral part of the TQM at JCF. This TQM system proved to be effective in achieving good results, which lead to the proper and effective implementation of EMS in achieving its objectives to reduce energy consumption.

The cost benefit analysis presented in this chapter proved, beyond any doubt, the importance of implementing the EMS in JCF. Without the comprehensive implementation of the EMS, JCF could have achieved isolated and remote successes with little effect and value. In other words EMS has created the proper environment for the success of JCF efforts in energy management and conservation.

References

Barrie G. Dale (2001), *Managing quality*, Blackwell, UK.

Buchanon, S. and Taylor, R. (1989), *The Electricity Consumption Impacts of Commercial Energy Management Systems*, *IEEE Transactions on Power Systems*, Vol. 4, No. 1, February, p. 218.

C. Barreiro and colleagues, *Reducing the energy costs and consumption in a Portuguese cement plant*, Zement-Kalk-Gibs, 7/1990.

Deeming, W.E (1993), *The New Economies*, MIT, Cambridge, Mass, USA

Ellerbrock H.G. (1994), *Combination Technology and Energy Management*, Translation Z KG, No. 11/94, p. 297.

EEBPP Energy efficiency best practice programs, Good practice guide 169, *Putting Energy into Total Quality: A Guide for Managers*, Department of Environment UK.1995.

EEBPP Energy efficiency best practice programs, Good practice guide 85, *Energy efficiency training and development*, Department of Environment UK.1998.

EEBPP Energy efficiency best practice programs, Good practice guide 186 , *Developing an effective energy policy*, Department of Environment UK.1996.

EEBPP Energy efficiency best practice programs, Good practice guide 119, *Organising energy management*, Department of Environment UK.1996.

Ernst Worrell et al(2001),*Policy Scenarios for Energy Efficiency Improvement in Industry*, Vol 29 Energy Policy November,2001.

Henry, H. Henry et al. (1980), *Energy Management, Theory and Practice*, Marcel Dekker, Inc. New York, U.S.A.

Huse, E.F. (1979), *The Modern Manager*, West Publishing Co., Minnesota, U.S.A.

Ilic, M.D. et al. (1993) " Short -term Economic Energy Management in a Competitive Utility Environment", *IEEE Transactions on Power Systems*, Vol. 8, No.1, February.

Kulseth, C.R., Roethke, E., and Bhatnager, R. (1993) " User Interface Issues for Evolving Open Systems", *IEEE Transactions on Power Systems* Vol. 8, No. 3, August, p. 1039.

Murphy, L., and Wu, F. (1993) " An Open Design Approach for Distributed Energy Management Systems", *IEEE Transactions on Power Systems*, Vol. 8, No. 3, August 1172.

Nordel, D.E. (1987) "An Approach to Conservation and load Management Economic Analysis", IEEE Transactions on Power Systems, Vol. PWRS-2, No.1, February, p. 108.

Rahaman, S. and Baba, M. (1989), An Integrated Load Forecasting - Load Management Simulator: Its Design and Performance, IEEE Transactions on Power Systems, Vol. 4, No. 1, February.

Reay, D. (1977), Industrial Energy Conservation, Pergamon Press, U.K., p. 306.

Sadgrove, K. (1994), ISO 9000/BS 5750 Made Easy-A Practical Guide to Quality, Kogan Page Ltd, U.K., p. 65.

Salehfar, H. and Patton A.D. (1991), A Production Costing Methodology for Evaluation of Direct Load Control, IEEE Transactions on Power Systems, Vol. 6, No.1, February, p. 278.

Venkatesh, B. and Chankong, V. (1995), Decision Models for Management of Cogeneration Plants, IEEE Transactions on Power Systems, Vol. 10, No. 3, August, p. 1250.

Weston, J. and Brigham, E. (1978), Managerial Finance, Sixth Edition, the Dryden Press, Hinsdale, Ill, U.S.A.

Wilker, G., Faruqui, A., Gellings, C. and Seiden, K. (1993), The Potential for Energy Efficiency in Electric End-Use Technologies, IEEE Transactions on Power Systems, Vol. 8, No. 3, August, p. 1351.

Oakland, J (2000), TQM text with cases, Butterworth and Heinemann, Oxford

Lindsay, W.M and Petrick, J.A (1997), Total Quality and Organisation Development, St. Lucie Press, Florida.

Roberts, H.V and Sergesketter, B.F (1993), Quality is personal: A Foundation for Total Quality Management, New York Free Press pp1-27.

US Department of Energy, Energy Information Adminstration (2000), Industry Analysis Brief, (www.doe.gov).

Chapter Nine

Discussion and Conclusions

Contents

Chapter Nine

Discussion and Conclusions

9.1 Summary

Energy, economy and industry are three strongly interrelated issues that have appealed to researchers as excellent research topics for the last twenty, or so, years. The interrelation and interaction of each issue with the other two is much more important and worthy of research and analysis than the individual issue itself. It is well known that energy is an essential input to all economic activities, including industrialisation. Energy is also very important for the well being and comfortable standard of living of people who, in turn, operate these economic sectors.

At the same time, industry is a measure of development of any nation. It is the largest contributor to national economies of industrialised and newly industrialised countries. Developing countries are still trying to enhance the share of the industrial sector to their respective national economies, which falls behind that of other sectors (e.g. agriculture, tourism, construction, etc).

Similarly, economic strength and healthy economic development lead to more industrial production and exploration of energy resources. Moreover, increased industrialisation results in an increase in energy consumption. However, most energy resources are both limited and, to some extent, cause environmental pollution.

Therefore, two strong trends have recently emerged. The first, which stems from environmental awareness, is directed towards protecting the environment with new technologies and processes that reduce the harm to the environment, or penalise the polluter to encourage a reduction in pollution, and/or to finance future research endeavours.

The second is a trend to reduce the consumption of energy for the same industrial output and thus achieve the same, or even better, economic return. This trend is called: increasing the efficiency of energy use or rational use of energy. It involves, mainly, energy management and conservation.

This present research has been concerned with this latter trend, namely energy management and efficient use of energy. The research has focused on a particular industry in a developing country and presented activities related to energy conservation measures within the context of an energy management system adopted for this particular industry. The Jordan Cement Factories (JCF) were analysed from the perspective of energy management and conservation.

Towards that end, it was first necessary to show the significance of energy cost with respect to the overall manufacturing costs. The analysis of production costs, that demonstrated the significance of energy costs, was followed by the investigation and examination of the basic management factors that have direct impact on energy consumption at the JCF. Among these factors are, for instance, production line availability, production rate, average number of stoppages and the average duration of stoppage. These factors were selected using preliminary data analysis, the practical factories experience and the experience and technical knowledge of the concerned parties.

The statistical analysis was established and it proved the existence of strong relationships between energy consumption as dependent variables and management factors (like availability, production rate, average number of stoppages and the average duration of stoppage) as independent variables. Several models were developed for a set of selected production lines (i.e., processes) in the JFC at Fuhais and Rashadiya plants. These statistical models were generated using actual data for electrical energy, fuel energy consumption and the independent variables. The derived models have demonstrated the existence of strong relationship between energy consumption and management control factors; for instance, the values of R^2 range from 60% to 85%. This implies that an equivalent percent of the variations in energy consumption can be attributed to the selected management factors.

The economical model developed in this research is concerned with demonstrating that effective management practices associated with proper maintenance and housekeeping can result in highly significant savings in energy usage. Although a preliminary simplistic methodology was used to evaluate the economic impact of any improvement program, the preliminary economic treatment showed that the cost of improvement is actually negligible compared to the realisable savings in energy usage.

Finally, the research presented the organisational and institutional aspects of the energy management system, together with the result of the practical experience, which illustrates the contribution of the system in reducing cost, improving productivity and enhancing the overall performance of staff and company as a whole. The results of the case studies or "success stories" are translated through a simple cost benefit analysis into an evidence of the effectiveness of the measures taken towards energy management and conservation.

9.2 Discussion & Findings

The research topic is of vital importance to the national economy of Jordan, the energy sector of Jordan and the Jordan Cement Factories. Many scholars and researchers have studied the issue of energy saving (or conservation) in all industrial sectors including the cement industry for many countries in the world (both developed and developing), but these studies were concentrated on separate and isolated energy conservation issues and not trying to have a complimentary integrated approach.

All previous research points to the need for energy management and conservation measures in industry in general and the cement industry in particular. This is especially true since the cement industry is an energy-intensive industry and, therefore, could cause environmental pollution.

The research work combines the implementation of empirical analysis, statistical modelling and analysis, economical modelling and analysis, and verification analysis supported with actual testing of case studies concerning energy management and conservation procedures.

One can raise several questions so as to initiate serious discussion concerning the research subject; it is worthwhile to discuss the following questions:

First) Was it important to select (energy management) as a main general subject of the research? And as an answer to this important question, the research demonstrate in the introductory chapters that the energy is a key factor in the policy planning and development of national economies since it interacts with a wide spectrum of economic and fiscal issues such as balance of payments, inflation, employment, investment and trade. Also energy is a clear example of limited world resources whose usage can have major impact on the natural environment. These factors confirm the importance of tackling the energy management issues.

Second) Was it important to concentrate on the industrial sector in general and one of the energy intensive industries, which is the cement industry in particular, taking JCF as a case study? And as an answer to this question, and according to the world energy council (1995) " The industrial sector is the largest primary consumer accounting for 43% of the total world consumption in 1992 ". For the selection of cement as an example of the importance of energy management in the highly energy intensive industries, energy cost represents around 35% of the total cost of producing cement, and between 60-70% of the variable cost. This demonstrates the importance of the selection of the cement as a case study of this research. Also from the environmental point of view the carbon reduction in the cement industry is specifically important because according to Ernst Worrell et al (2001), in 1997 industrial energy use and process emissions from the cement manufacture accounted for 33% of the total US CO_2 emissions, and this is due to the calcinations of the limestone for the production of clinker i.e. process CO_2 emissions from calcinations are added to the cement sub sector energy related CO_2 emissions. This also demonstrates the importance of the selection of the cement industry from the environmental point of view, taking in consideration that for some time attention has been paid to the so called greenhouse gases which are being blamed for changing the earth's climate. CO_2, which is product of fossil fuel combustion, is classed as one of the main greenhouse gases, (Rose D. et al 1991)

The main theme of the research (which is to conserve energy and manage it rationally) would be particularly useful in making a significant contribution towards achieving Kyoto targets. Kyoto Protocols defined the allowable greenhouse gas emissions for each industrialised country in terms of assigned amounts for the commitment period 2008-2012. The commitments add up to a reduction of 5.2% below 1990 levels, besides other specified commitment for other countries. (Grubb et al (1999). The industrial sector will be called on to make a major contribution to meeting these targets. So the concentration of the research on the industrial sector and specifically on one of the highly energy intensive industry is clearly justified, and the contribution of its approach which concentrate on energy management and conservation issues expected to be instrumental in helping to achieve Kyoto targets. The research is expected to be significant issue in the potential application of the research approach to other high-energy intensive industries.

Third) Was there a real chance to improve energy efficiency use in the industries in general and on the cement industry in particular? And as an answer to this question, the author noticed through his research that there was decoupling of energy demand growth and economic growth after the oil price shocks in 1973, particularly in high-income countries, which were able to reduce wasteful energy consumption or improve the efficiency use by adopting energy efficient technologies. As a result, the energy intensity, which is a measure of efficiency of the energy use, has declined by 1-2% per annum during the past two decades. It is revealed through the researcher literature review in general and concentrating in the highly energy intensive industries with special concentration on cement in particular that there is a considerable efficiency gap between the optimum designed energy use efficiency and the actual achieved energy use efficiency which necessitate the importance of investigating the ways and means to improve energy efficiency.

Fourth) Was it important to concentrate on statistical analysis to establish the relation between the energy consumption on one hand, and management functions and practices as given by the variables on the other hand, which will help to demonstrate, mathematically, the significance of energy management in controlling the energy consumption and cost? And as an answer to this question, the researcher through the literature review could not identify or locate a serious effort to use statistical analysis to establish this relation neither in energy intensive industries in general nor in cement industry in particular. It reveals from the analysis that the main factors affecting the energy consumption are: production rate, average number of stoppages, average duration of stoppages, availability, lime saturation factor, silica ratio and alumina ratio. According to the statistical analysis these factors can explain between 60-90% of energy consumption variation of the production unit concerned. The power of explanation of the energy management factors revealed from the statistical analysis proved the importance of using the statistical analysis and modelling approach.

Fifth) Was it important to concentrate on establishing an integrated energy management system? And as an answer to this question, through his condensed literature review the researcher could not locate serious efforts concentrating on the institutional aspects of energy management irrespective that there was several and separate energy management initiatives. For example the researcher was managing one of the ministries in Jordan (ministry of water and irrigation), the cost of energy was representing 84% of the variable cost of the water, and the total number of staff was around 7000 employees and non of them was allocated to address the energy management issues. In spite the fact that the water sector is mainly managed by an international management company and many international aid programmes were available, but none of them address the energy management issues. This clarifies in a very clear way the great necessity to institutionalise the energy management issues not only at Jordan level but also at global level, because of its great economical and environmental implications and the importance of establishing an integrated energy management system.

The following points are worth highlighting to demonstrate that the research follows the methodology, which aimed and established to achieve the objectives stated in the thesis:

1. The research clearly highlights the fact that energy, as a commodity in Jordan, is not only scarce as an indigenous resource, but also a heavy burden on the national economy, as represented by the high-energy import bill. Even at the sectoral level, the data that were analysed regarding the industrial sector's energy consumption (including cement) proved the point previously made at the national level. The Jordanian Government's measures for the promotion of rational use of energy and the obstacles to energy conservation in Jordan were discussed and presented. This covers the first objective of the research (To assess the current energy consumption patterns in the industrial sector in Jordan, with special reference to the cement industry).

2. The literature review revealed the following:

   2.1 The research also shows that JCF is an energy-intensive enterprise, much like other cement industries throughout the world. A detailed review of related previous research and literature was presented, including; Technical Measures relating to Energy, Fiscal Instruments applied to Energy, and Previous investigations of using statistical techniques and energy management systems, in the cement industry and other highly energy intensive industries.

   Through the review the researcher could not identify a detailed empirical and analytical approach for the main factors *affecting the energy consumption in the* energy intensive industries. Also, he could not find any significant attempt to develop a statistical model using these factors for the cement or other highly energy intensive industry. Further, he could not be able to locate a detailed integrated energy management system, which has been developed or used, but found some serious effort to establish limited energy management activities or limited energy management systems. These important findings confirm the importance and the urgent need for depth investigation to the research topics, and also it demonstrates its originality.

2.2 It revealed from the literature review in chapter 3 that international cement manufacturers are coming under increased legislative pressure and, for example, in the UK cement manufacture is subject to the recently introduced IPPC (Integrated Pollution Prevention and Control). The cement industry is aware of its responsibility and obliged to adopt good environmental practice, which indeed incorporates good energy management practice. This emphasises the real need for the main theme of this thesis.

2.3 The literature review confirmed the importance of investigating the research problem as it revealed that it exists at a global level and its economical implication is of great importance.

The above three points covers the literature survey included in the research.

3. The cost analysis revealed the following:

3.1 The research analyses the total production cost breakdown. From this analysis, it is clear that energy consumption cost is high (represent around 35% of the total cost and around 70% of the total variable cost) as a percentage of total production cost. This is not a unique fact to JCF but rather, a universal fact that is strongly related to the high-energy intensity of cement manufacturing. This fulfils the objective of analysing production costing of cement manufacturing.

3.2 Energy cost analysis revealed that economy of scale concept is playing an important role in deciding the total cost of cement. The most common source of economies of scale is the spreading of fixed costs over an increasing level of output. In other words, economies of scale relate the cost of production to the quantity of production. Naturally the bigger the plant was, and the more the production volume was, the lower per unit cost of production will be. Conversely, when a plant is small, its per unit cost will be higher.

It is worth to mention here that energy intensive industries including cement are highly capital-intensive projects i.e. the capital needed to establish such a projects is extremely high. For example as a rule of thumb in the industry 1 million ton production line of cement cost $200 million. Similar patterns exist in other energy intensive industries. As a matter of fact, steel and aluminium industries are much more energy intensive than the cement industry; this situation necessitates a careful feasibility study before taking final decision to establish such an energy intensive industry. If the energy resources are not available in the country and if the local market is not consuming the major amount of production one should be very careful in taking the decision to build an energy intensive industry, taking into consideration that for such type of industry the cost per job creation is extremely high. Not to mention again the serious environmental limitation to such industries and the high cost of transport because these industries are bulky in its nature.

The research revealed that energy costing analysis for the cement industry is an essential step in evaluating the cost-effectiveness of energy management and to demonstrate the importance and validity of the research topics.

The previous two points covers the objectives of analysing production costing of cement manufacturing.

4. The research highlights the practical experience of JCF in rational use of energy, including energy conservation, energy auditing and energy management practices and procedures. The energy management issue was dealt with from different aspects, among them the Philosophy of the Energy Management; Factors and means of energy consumption control including the stability of the operational process, which affects certain control factors and therefore affects the level of energy consumption.

   It is revealed from the preliminary *detailed empirical analysis* of the main factors affecting energy consumption that these factors include availability, production rate, average number and duration of stoppages and factors affecting quality of produced cement.

   The preliminary empirical results and the qualitative and quantitative evidence resulting from the case studies illustrated, beyond doubt, the vital need for establishing a detailed statistical modelling and analysis for the factors affecting the energy consumption.

   This covers the specific areas of the objective of the research of improving and conserving energy consumption.

5. The statistical analysis and modelling revealed the following:

   5.1 The research demonstrates through using statistical analysis, the significance of energy consumption (fuel and electricity) as a function of energy management, as well as of production. If the plant is subject to frequent stoppages, interruptions and start-ups, its specific energy consumption will increase significantly. The findings of the statistical analysis revealed that the relationship between energy consumption and emergency stoppages is almost non-negative. This means that, as both number and duration of stoppages are increased; the consumption of both electricity and fuel will increase.

5.2 The research demonstrates through using statistical analysis that the relationship between production rate and energy consumption per unit is negative whenever it is significant. This indicates that at higher levels of production rate, consumption of both electricity and fuel decreases per unit of cement output.

Production rate has exerted a negative impact on both specific electricity consumption and specific fuel consumption in almost all model applications. This result is expected, since as we are approaching full utilisation of the kiln's resources, the optimal state of consumption of both electrical power and fuel is achieved, hence an optimal state of productivity is maintained. Another explanation of this result might be as follows: when production rate is high, this indicates that number of stoppages is low, hence saving those times of heating-up to restart again where consumption of both electricity and fuel are wasted for non-productive operations. The period of heating-up is normally 24 hours of continuous consumption of electrical power and fuel at zero production level.

5.3 The research demonstrates through using statistical analysis that the relationship between energy consumption and availability is negative whenever it is significant, i.e. as availability increases energy consumption decreases. This result, also, confirms the author expectations or predictions. Availability, as measured in this research, maintained a negative impact on both electricity and fuel consumption. This relationship is also expected, since as percentage of actually utilised production capacity is increased because of higher availability, the levels of consumption of both electric power and fuel are decreased. This result is important because it clearly indicates to executives the need to maintain the highest level of availability in order to accomplish the goal of reducing the consumption of electrical power and fuel.

5.4 The research demonstrates through using statistical analysis that the relationship between Aratio and energy consumption per unit is negative whenever it is significant. This confirms the author's expectation that at higher levels of Aratio, consumption of both electricity and fuel decreases per unit of cement output.

5.5 The research demonstrates through using statistical analysis that the relationship between LimeSF and energy consumption per unit is positive whenever it is significant. Again this confirms the author's expectation that at higher levels of LimeSF, consumption of both electricity and fuel increases per unit of cement output.

5.6 The research demonstrates through using statistical analysis that the signs of Sratio depend on the kiln. They may be negative as well as being positive. This inconsistency may be due to the existence of some outliers in the data.

5.7 The research demonstrates through using statistical analysis that the R-squared values of electricity models are all very high, which means that the models can explain from 82% to 94% of the variability in EL using the independent variables in the models. On the other hand, the R-squared values of fuel models indicate that the models can explain from 57% to 89% of the variability in FUEL using the independent variables in the models. It seems reasonable that the R-Square values for the FUEL models are smaller than the corresponding values of the EL models. This difference can be explained by the fact that the burning process in the kiln is very complicated process affected by many factors while the factors affecting electricity consumption are limited to the factors related to start up or shut down to the electric motors. This result indicates the need to study the possibilities to add another factor to the statistical analysis to increase the power of explanation of the fuel consumption by the model.

5.8 The research demonstrates through using statistical analysis of the factors affecting the raw mills and cement mills the usefulness of the independent variables in explaining the variability in electrical energy consumption as demonstrated by the R^2 results that ranges from 0.64 to 0.97. Moreover, the R-Squared values for all mills range between 0.78 and 0.97 except for R-rm2. This indicates that the regression models have useful capability to explain the variability in EL using the available independent variables.

The results of the mills confirm the expectations of the researcher; i.e. in almost all cases, the increase in number of stoppages and duration of stoppages affect positively the electricity consumption, whereas production rate and availability affect electricity consumption negatively. That is to say, the production rate and availability showed the same decreasing relationship with electricity consumption. That is if these variables increase, the energy consumption will decrease. On the other hand, the duration of stoppages and number of stoppages showed the same increasing relationship with electricity consumption i.e. if these variables increase then electricity energy consumption will increase.

5.9 In general, the research through the analysis of the statistical models revealed that there is a strong relationship between energy consumption and the following independent variables: average number of stoppages, average duration of stoppages, production rate, availability, Alumina Ratio (Aratio), Silica Ratio (Sratio), and Lime Saturation Factor (LimeSF). This strong relation is shown in the values of R^2

of the models. On the other hand, the p-values for the F-test, which were less than 0.05 shows that the models are significant. This encourages further research to include other independent variables to reach more powerful models.

5.10 To create benchmarking with reputable cement manufacturers, the results of the statistical analysis, which was carried out for around 30 kilns distributed all over the world, revealed values of R^2 close to those obtained from the Jordanian data. Moreover, the signs of coefficients of the independent variables in the estimated regression model almost agree with the researcher's expectations. This means that the selected control variables can have comparable effect on the cement industry in general, which asserts positively the choice of the model and its use in the industry. Moreover, this supports the findings of the research that the variables included are significant.

5.11 It is worth to mention here that we faced some problem in collecting the information and data from the factories. Usually the data collection in the factories does not take into consideration the possibility that these data may be used in a scientific academic research and because of that we faced the following problems:

i) The data that is used in this study is reported monthly and we found that some of the data was rounded by the clerks who reported them.

ii) The data has some outliers, their number is very limited and we managed to handle them using exploration techniques that identify the outliers. Moreover, robust regression analysis is used to fit regression models to the data since this is the suitable procedure the existence of the outliers in the data.

iii)The number of factors reported could be enhanced to include another factors, which is expected to improve the power of explanation of the model in the future. This work hopefully will lead to another serious effort for further future research in this direction.

iv) The effect of rounding the Data on the regression model was studied and it was found that it has very minor effects on the results, R-squared values of both models based on rounded or unrounded data are almost the same and the estimated coefficients of the independent parameters in both models are almost the same in magnitude and sign. v) In developing the statistical model we started with the first four independent variables, which are: PRORATE, AVL, AvNO, and AvHOURS. Through the research work it was found that additional variables needed to be added to support the explanation power of the model. The variables affecting the quality of the product were added to the analysis to support the model as mentioned above. Further improvements of the model by adding additional variables are still possible and could be a subject for further research.

The above mentioned points (5.1 up to 5.11) covers the research objective regarding establishing statistical and economical models that relates energy consumption with management practices and factors, and their related economic impact on energy and production cost.

6. The development of an energy management system revealed the following:

6.1 The research has dealt with the details of developing a coherent energy management model whose objective is to establish transformational management process of certain high-level management factors into daily operations and controls. The high-level management factors are the same factors used as independent variables in the statistical and economic models, which statistically proved to be the major factors affecting the energy consumption at JCF.

6.2 The research also has presented a detailed analysis of the organisational and procedural aspects of energy management with concentration on management functions, especially planning, controlling, executing and organising, and auditing. A detailed mapping and analysis of these functions as the main components of the EMS resulted in establishing job descriptions, organisational charts, work instructions and procedures for all important functions of the EMS. This covers the research objective related to establishing an EMS based on modern management techniques, which is integrated within an overall Total Quality Management (TQM) concept.

7. A detailed empirical analysis for the causes of high-energy consumption at JCF revealed certain factors, which contributed to this situation. To support this analysis several case studies were carried out to demonstrate the potential for energy saving and to create success stories in order to gain management commitment and support and create institutional awareness regarding energy management. This covers the research objective related to investigating the causes of high-energy consumption and demonstration of energy management potential and importance of energy saving through important lessons learned through implementing practical case studies.

8. A very important issue, which was only briefly mentioned in the research, is the impact of JCF on the surrounding environment. This is a research topic on its own and deserves to be addressed separately in another independent research. However, it is worth mentioning that rational use of energy and environmental protection has common objectives. Moreover, a measure to alleviate a problem in the one area (energy) will

result in improving the other issue (environment) as well. For example, it is a known fact that frequent cement plant stoppages and consequent restarting result in more particulate emissions in the atmosphere. So, not only will the energy consumption increase as a result of frequent stoppages, but also more environmental harm is done as well. Similarly, if any measure is successful in reducing energy consumption, then, it will automatically lead to a reduction in harm to the environment.

9. From an overall perspective, the research follows the methodology, which aimed and established to achieve the objectives stated in the thesis. However, there is still room for further research that will hopefully complement this research.

9.3 Conclusions

The research work presented in this thesis is a result of the extensive experience of engineering and senior management functions within both the power station and cement manufacturing sectors, of developing energy management systems and of modelling energy management and hence formulating a statistical approach capable of facilitating a detailed analysis of the underlying factors dictating the energy requirements for cement manufacture, the author was convinced of the huge potential for energy conservation at JCF and in all industrial activities in general. Now, as a result of the research work, major steps were taken to realise this potential, thus leading not only to reduced energy consumption and cost thereof, but also to higher profitability.

The exhaustive detailed review of related previous research and literature included; Technical Measures relating to Energy, Fiscal Instruments applied to Energy, and Previous investigations of using statistical techniques and energy management systems, in the cement and other highly energy intensive industries. As mentioned before, it was not possible to identify any existing detailed empirical and analytical approach for the main factors affecting the energy consumption in the energy intensive industries. Also, the search did not reveal any significant attempts to develop a statistical model using these factors for the cement or other highly energy intensive industry. Further, it has been impossible to locate a detailed integrated energy management system, which has been developed or used. The work presented here is a novel and original approach to energy management, which has been validated by experience in the field.

Based on this search, building up several statistical models to represent the relationships between energy consumption on one hand, and management functions and practices as given by the variables on the other hand, with detailed statistical analysis of the factors affecting the consumption of energy in the cement is an original research attempt, which will help to demonstrate, mathematically, the significance of energy management in controlling the energy consumption and cost. This represents an original method for rational use of energy and for the prediction of energy consumption.

Also the research is original as major part of it which is the statistical modelling, was employed as a basis for developing economic models and translating the impact of sound management practices and tools into quantifiable economic variables. Managers can immediately use such variables for developing "best practices" policies and decisions. Economic variables such as the net monetary savings and the rate of return on investment were used for this purpose. The statistical model was transformed into a practical economic model to relate the saving of energy consumption as a function of the cost of improving the independent variables affecting energy consumption. The objective of building the economic model was to verify whether there would be financial gains from improving the control variables as predicted by the statistical model or not.

As we mentioned above, this research is original in that it proves in a mathematical manner, using statistical and economical analyses, the *significance* of energy consumption (fuel and electricity) as a function of a group of energy management factors, as well as of production. The corollary is that if the plant is subject to more frequent stoppages, interruptions and start-ups its energy consumption will increase significantly.

The findings of the statistical and economical modelling and analysis were strongly supported by results of the cost benefit analysis, which proved, beyond any doubt, the importance of the statistical analysis findings. Also it proves that without the comprehensive implementation of the EMS, JCF could have achieved isolated and remote successes with little effect and value. In other words EMS has created the proper environment for the success of JCF efforts in energy management and conservation.

Energy conservation in the cement industry is not only a function of technology and management commitment as well as the conscious involvement of the work force, but also of the comprehension of the value of energy conservation and containment of environmental pollution.

The following points highlight the main conclusions of the research work:

1. *General*

1.1 Energy is essential to economic prosperity and quality of life, however as it revealed from the research, energy resources are scarce in Jordan thus its use impose economic and financial burdens on the national economy.

1.2 The concept of rational use of energy has two embedded principles, the first is increased energy use efficiency and the second is environmental protection.

1.3 According to the world energy council 1995 the industrial sector is the largest primary energy consumer accounting for about 43% of the total world consumption in 1992. This justify the research work to address the energy management and conservation issues in one of the highly energy intensive industries which is the cement industry.

1.4 Policies should be guided by the need to conserve energy for future generations and by concern for environmental impact of burning fossil fuels. We cannot afford to continue in irresponsible consumption behaviour when we know the economic and environmental implications of such behaviour.

1.5 It is revealed through the researcher literature review in general and concentrating in the highly energy intensive industries with special concentration on cement in particular that there is a considerable efficiency gap between the optimum designed energy use efficiency and the actual achieved energy use efficiency which necessitate the importance of investigating the ways and means to improve energy efficiency.

2. *Energy Management Issues*

2.1 Institutionalisation of energy management is of great importance and should be imposed in all the industrial activities and other activities. Currently it is almost partially existed or not existed. Energy control and management system must be among the priority issues for the cement sectors and for the other highly energy intensive industries. The potential for energy efficiency improvement for mature cement industries expected to be around 10% while for the rest of the industry between 20-25% and the similar potential expected for other highly energy intensive industries and this confirm the importance of the research work and findings.

2.2 JCF, being an energy-intensive industry, is a good candidate for measures aimed at improving energy use efficiency. Among these measures are the following:

a - management commitment and understanding;

b - implementing management functions;

c - energy conservation exercises;

d - energy auditing exercises.

In other words, an Energy Management System (EMS) within the organisational structure of the company, and as an institutional unit need to be established to guarantee efficient energy use in the cement *industries and the same* is applied to the other highly energy intensive industries.

2.3 Practical experience with EMS at JCF suggested that the system worked effectively. Management commitment is manifested in the allocation of competent staff and resources to facilitate the functions and activities of the EMS. Moreover, top management at JCF allocated their time, participation and involvement in EMS functions and activities. In short, the EMS has become an essential element in the overall management of JCF.

3. *Costing Issues*

3.1 The analysis of the production cost at JCF shows that electricity and fuel cost constitute a sizable portion of the cost of the production (around 70% of the variable cost, and around 35% of the total cost).Energy cost in other highly energy intensive industries represent major part of the total cost of the industries production costs. This confirms the importance of investigating the research problem; as it revealed that it exist at a global level and its economical implications of great importance.

3.2 Monitoring energy cost, building statistical and economical models and implementing energy management system in cement plants proved to be cost effective and useful in improving the utilization of energy at JCF.

4. *Statistical Analysis*

4.1 The research confirms the importance of investigating and analysing the factors affecting the energy consumption. It reveals from the analysis that the main factors affecting the energy consumption are: production rate, average number of stoppages, average duration of stoppages, availability, lime saturation factor, silica ratio and Alumina ratio. According to the statistical analysis these factors can explain between 60-90% of energy consumption variation of the production unit concerned which indicates the need more investigation for the factors affecting the energy consumption.

4.2 It is expected, based on the analysis and the findings of the research, that the factors related to the mode and stability of operation (availability, average number of stoppages, average duration of stoppages, production rate etc) expected to have similar important effect on energy consumption in other highly intensive industries, which confirms the importance of the research work and its possible applicability in other highly energy intensive industries.

4.3 Statistical analysis of actual data and practical cases proves that energy consumption is strongly related to mode of operation and housekeeping practices of the plants. This fact not only confirms the importance of having an EMS with all its functions and activities at JCF and other energy intensive industries, but also sets a precedent of a statistical tool to be used continuously to monitor energy consumption and relate it to other independent variables.

4.4 The statistical model established by the research will be useful not only in similar future studies but, for instance, it could be used to define the relationship between environmental pollution and operation.

4.5 In an attempt to verify the results of the statistical analysis, a similar model was established using data from a reputable international cement manufacturer. The results were in agreement with the results of the research, which indicates that the model construction was sound and the selection of variables was appropriate.

5. Several practical cases, as discussed throughout the thesis, demonstrate the importance and cost-effectiveness of the energy conservation measures suggested by the EMS. This can be verified through the application of the economic model, which translates the statistical relationships between dependent and independent variables into financial gains. The developed economic model was preliminary model and briefly tested but more data is needed to extend the actual testing of the derived economic model and to improve it. This is beyond the scope of the research and could be candidate for further research.

6. EMS

6.1 In arriving at the final EMS, several important and useful steps have been undertaken. The thesis suggests and describes in detail the organisational structure, job descriptions, and the responsibilities and duties of personnel and committees. These efforts will be valuable for other company endeavours. ISO 9000, Total Quality Management, and environmental protection are all candidate activities, which have benefited from the efforts, exerted so far to bring about the final EMS.

6.2 An added benefit from the efforts of the research work leading to the EMS at JCF is the systematic and orderly manner of doing things. The creation of quality circles and energy committees has brought this about. The staff of JCF can use the same methodology and procedures to tackle other multi-disciplinary issues and difficult tasks.

6.3 The efforts of the research work have increased the level of awareness of even the shop-floor worker with respect to rational use of energy. The almost total involvement of all staff in monitoring, costing, and effecting measurers dealing with energy consumption has created an atmosphere of energy awareness and involvement.

6.4 As demonstrated in the research, the annual savings (in case of full production) as a result of improving availability, conserving energy, introducing EMS and employing management techniques at JCF amount to US$ 3.5 million. The reduction in the cost of energy brought about by all measures undertaken reaches about 7.5% of total energy consumption, which is quite an achievement.

7. The methodology of the research blends theoretical work with practical evidence. In other words, the methodology includes production-costing, analysis of the factors affecting energy consumption, statistical and economic models in addition to certain case studies of energy conservation. Moreover, the thesis also includes a description of an energy management system including organisational structure and job descriptions. This blend of theory and practice provides convincing evidence of the workability of the models derived and the ideas and concepts promoted.

9.4 Further Research

Generally it is recommended that cement industries and other highly energy intensive industries continue to give high priority to energy management issues. On the academic scene, however, the following issues need further research by others because they have not yet been investigated and they will contribute to the subject of energy management and rational use of energy:

1. Issues Related to Model Development and Improvement

1.1 Detailed study and analysis of the types of stoppages mainly break down stoppages and programmed stoppages and the causes of these stoppages and how to minimise and avoid them if possible. It revealed from the research that stoppages are one of the main factors, which affect the energy consumption, and caused by many reasons most important of it is poor maintenance.

1.2.Develop the statistical model that correlates energy consumption with other independent variables in addition to the ones that we used in the research so as to improve the power of explanation of the model.

1.3 Developing, testing and verifying the suitability of the derived preliminary economic model by assessing the need for more accurate improvement cost data to be included in a more comprehensive approach to economical modelling and evaluation. This specifically covers the related cost of improving the energy management variables, such as availability, production rate, average number and duration of stoppages etc.

2. Environmental and Global Issues

2.1 Building a statistical model, which formulates the mathematical relationship between the independent variables used in the research and the air pollution. Air pollution can be measured in terms of concentration of particulate in stack emissions as well as particulate grain size and distance of coverage (distance from plant).

2.2 Kyoto Protocols defined the allowable greenhouse gas emissions for each industrialised country in terms of assigned amounts for the commitment period 2008-2012. Based on the important findings of this research it is recommended that a detailed research to be established to study the contribution of energy conservation and management as an instrumental vehicle towards achieving Koyoto targets.

2.3 The potential of energy conservation and the possibilities of improving environmental protection in the other highly energy intensive industry is still high. Other detailed research for these industries is recommended and expected to be of great importance.

2.4 Developing an econometric model, which calculates the environmental cost of cement production and predicts environmental damage as a function of cement production.

2.5 Assessing the real need for institutionalisation of energy management issues and assessing the impact and effectiveness of applying similar statistical and economical modelling and analysis and extrapolating the EMS on other energy-intensive industries.

References:

1. Rose D. et al, *Ecological and Economic aspects of cement production when using waste-derived fuels*, ZKG, 11/1991.

2. Grubb et al (1999) *The Kyoto Protocol: A guide and Assessment*, The Brookings Institution, Published in Association with Royal Institute of International Affairs, USA.

3. World Energy Council, *Energy Efficiency Improvement Utilizing High Technology-An Assessment of Energy Use in Industry and Buildings*, Report and Case Studies, 1995.

4. Ernst Worrell et al(2001),*Policy Scenarios for Energy Efficiency Improvement in Industry*, Vol 29 Energy Policy November,2001.