CHAPTER I

INTRODUCTION

AND

SUMMARY

CHAPTER I

INTRODUCTION AND SUMMARY

1.1 Chapterwise Summary:

This dissertation is devoted to the problem of subset selection in regression analysis. It consists of five chapters. First, we present chapterwise summary of the dissertation.

Chapter I contains some basic concepts of regression analysis and some important results which are relevant for in later disscussion. Section (1.2) gives an introduction about linear regression model. In section (1.3) we discuss different methods of estimation of parameters and its properties. Residual, types of residuals, its properties and some results based on residuals are discussed in section (1.4). Distribution properties and ANOVA table which are discussed in section (1.5). Finally in last section, we discuss the problem and need of subset selection.

Chapter II is devoted to various selection criteria for subset selection. In section (2.2), we discuss coefficient of determination of R^2. Also in this section, we discuss, Aitkin's (1974) test procedure for selecting subset by using R^2-statistic and in the next section we give another criterion based on

adjusted R^2-statistic. Lastly we give Haitvosky (1969) test procedure for selection of subset by using adjusted R^2-statistic. Mean square error of prediction criterion is discussed in Section (2.4). In Section (2.5), Mallows C_p-statistic is discussed in detail. All the above techniques are suitably illustrated with real life data as well as simulated data.

It is difficult to select a subset by using above selection criteria when the number of independent variable is too large. In this situation several other suitable methods of subset selection are avilable and these are discussed in Chapter III.

We have discussed a recent method based on principal compontents in the Chapter IV. This method is proposed by Bonesh and Meditra (1994).

Chapter V deals with effect of subset selection on estimates of parameters, estimates of error variance . After studying the effect of dropping variables, the question arises as to how many variables should be included in the linear regression equation which is discussed in the section (5.4). In the last section of this Chapter, we discuss the bias reduction methods, namely,

1: Jackknife statistic

2: Bootstrap method

and compare estimators value by generating random samples.

1.2 Introduction and notation.

For various types of data, various techniques have been developed for extracting relevant information from that for determining what the data "mean". One common type of data occur when the value of several variables are measured for each of several units. For example, in medical study, age, weight, height and blood pressure of a group of 100 subjects might be recorded. Here observations are taken on four variables correponding to each of 100 subjects recorded.

Linear regression analysis is commonly used statistical technique for dealing with such data. This technique is frequently used in almost all fields. In the linear regression technique a variable of main interest Y is called the criterion variable (also called the dependent variable, regressond, responce variable, the predictant) and the set of other variable $X_1, X_2, \ldots, X_k$ are called the explanatory variables (also called the regressors, predictors, independent variables)

1.2.1 Linear regression model :

Before we discuss the linear regression model in detail, consider the following example:

Suppose Y is the concentration of Vitamin B_2 in a plant called Turnip plant.and X_1, X_2 and X_3 are respectively the

sunlight, soil mixture measurement and air temperature. Let the data $(y_i, x_{i1}, x_{i2}, x_{i3})$, $i = 1,2,\ldots,n$ on (Y, X_1, X_2, X_3) be available from n plants. Here one can easily see that the variable Y is dependent on other variables. So variable Y can be called the 'dependent variable' or 'responce variable'. As opposed to that the variables X_1, X_2, X_3 can be called 'independent variables'.

1.2.2 Formulation of the model :

In above example, the concentration of Vitamin B_2 in the leaves of a turnip plant (Y) is, approximately a fuction of sunlight (X_1), soil mixture (X_2) and temperature (X_3). In a real life processes, the Y is a function of X_1, X_2, X_3 but not exactly mathematical function of X_1, X_2, X_3 and hence we can write a model as

$$Y = F(X_1, X_2, X_3) + \varepsilon$$

where F is a suitable function and ε is a random error.

Specifically, in linear regression model, the F is taken as linear function and it is expressed as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

Note that the form of the function is specified to be linear but the coefficents β_0, β_1, β_2, β_3 are not specified. They are unknown parameters and these are called regression coefficents or

regression parameters.

1.2.3 General case of K- variables

Suppose data consisting of n observations on a response variable Y and k explanatory variables are available. The regression model is,

$$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + \varepsilon \qquad (1.2.1)$$

It can be expressed in terms of observed data,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} + \varepsilon_i \qquad (1.2.2)$$

$$i = 1,2,\ldots,n$$

where β_0 is called intercept and $\beta_1, \beta_2, \beta_3, \ldots, \beta_k$ are called regression coefficients or regression paramaters. Further writing these equations in the matrix form, we get

$$\underline{Y} = X \underline{\beta} + \underline{\varepsilon} , \qquad (1.2.3)$$

where

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \ldots & X_{1k} \\ 1 & X_{21} & X_{22} & \ldots & X_{2K} \\ . & & & & \\ . & & & & \\ . & & & & \\ 1 & X_{n1} & X_{n2} & \ldots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_k \end{bmatrix}$$

We note that the model (1.2.2) is a very general one. For example, setting $X_{ij} = X_i^j$, we have the polynomial,

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \quad + \beta_k X_i^k + \varepsilon_i$$

which is also a special case. The essential aspect of (1.2.2) is that it is linear in the unkown parameter β_j, (j = 0 1 2 k) and precisely for this reason, it is called a linear model.

In contrast,

$$Y_i = \beta_0 + \beta_1 \exp(-\beta_2 X_i) + \varepsilon_i \qquad (1.2.4)$$

is a non-linear model, being non-linear in β_2. The following assumptions about linear model are normally made:

Assumption (1.2.1) :

1) $E(\underline{\varepsilon}) = 0$ and $cov(\underline{\varepsilon}) = \sigma^2 I_n$

and further analysis,

2) $\underline{\varepsilon}$ is distributed as $N(0, \sigma^2 I_n)$

1.3 Estimation of parameters :

In this section, we discuss a method of estimation of parameters for the model (1.2.3).

1.3.1 Least squares estimation :

A widely used method of obtaining an estimator of $\underline{\beta}$ is the 'Least square method'. The method is as follows:

Consider the model

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \qquad (1.3.1)$$

Then,

$$\underline{\varepsilon} = \underline{Y} - X\underline{\beta},$$

and

$$\underline{\varepsilon}'\underline{\varepsilon} = (\underline{Y} - X\,\underline{\beta}\,)'(\underline{Y} - X\underline{\beta}),$$

$$= \underline{Y}'\underline{Y} - 2\,\underline{\beta}'\,X'\underline{Y} + \underline{\beta}'\,X'X\,\underline{\beta}, \qquad (1.3.2)$$

Differentiating (1.3.2) with respect to $\underline{\beta}$ and equating it to zero, we get

$$-\,2\,X'\underline{Y} + X'X\,\underline{\beta} = 0$$

which implies that

$$X'X\,\underline{\beta} = X'\underline{Y} \qquad (1.3.3)$$

The equations (1.3.3) are known as normal equations. If $(X'X)^{-1}$ exists then (1.3.3) have the unique solution for $\underline{\beta}$ and is given by

$$\underline{b} = (X'X)^{-1}X'\underline{Y} \qquad (1.3.4)$$

If $(X'X)^{-1}$ does not exist, then there is no unique solution for $\underline{\beta}$. In this case, generalized inverse of (X'X) is used to obtain a solution for $\underline{\beta}$.

Remark (1.3.1): For the sake of completenes, we give another method known as "maximum likelihood" method of estimation.

Suppose ε_i ; i = 1, 2, . . ., n are independantly and identically distributed normal variates having $N(0,\sigma^2)$ distribution, so that we can write the likelihood $L(\underline{\beta}:\underline{\varepsilon})$ based on n observation as

$$L(\underline{\beta}:\underline{\varepsilon}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{(\underline{Y}-X\underline{\beta})'(\underline{Y}-X\underline{\beta})}{2\,\sigma^2} \qquad (1.3.5)$$

Taking partial derivatives on both sides of equation (1.3.5) with

respect to $\underline{\beta}$ and equating it to zero, we get the normal equation

$$- (X'X)\, b + X'\underline{Y} = 0 \qquad (1.3.6)$$

This implies,

$$b = \begin{cases} (X'X)^{-1}X'\underline{Y}\ ; & \text{if } (X'X) \text{ is non-singular} \\ (X'X)^{+}X'\underline{Y}; & \text{if } X'X \text{ is singular} \end{cases}$$

where $(X'X)^{+}$ is a generalised inverse of $(X'X)$.

1.3.2 Mean and variance of LSE of $\underline{\beta}$:

Now, we obtain the mean and dispersion matrix of $\underline{b}$.

Property 1.3.1 : $\underline{b}$ is an unbiased estimator of $\underline{\beta}$.

Proof : We have

$$\underline{b} = (X'X)^{-1}X'\underline{Y}$$

Hence

$$E[\underline{b}] = E\,[\,(X'X)^{-1}X'\underline{Y}\,]$$

$$= (X'X)^{-1}X'E\,[\underline{Y}]$$

From the expression (1.2.3) we get,

$$E[\underline{Y}] = X\underline{\beta}$$

Thus,

$$E[b] = (X'X)^{-1}X'X\,\underline{\beta}$$

$$= \underline{\beta}$$

Property 1.3.2 : The variance covariance matrix of $\underline{b}$ is given by

$$Var(\underline{b}) = (X'X)^{-1}\sigma^2$$

Proof : Observe that,

$$\begin{aligned} Var[\underline{b}] &= Var[(X'X)^{-1}X'\underline{Y}\ ], \\ &= (X'X)^{-1}X'Var[\underline{Y}]\ X\ (X'X)^{-1}, \\ &= (X'X)^{-1}X'[\sigma^2 I_n]X\ (X'X)^{-1}, \\ &= (X'X)^{-1}X'X\ (X'X)^{-1}\sigma^2 \\ &= (X'X)^{-1}\sigma^2 \end{aligned}$$

Further here we note that $\underline{b}$ is the best linear unbiased estimator (BLUE) of $\underline{\beta}$. Moreover, if the underlying distribution is normal, then $\underline{b}$ is also M.L.E. of $\underline{\beta}$.

1.4 Residuals

In the Section (1.2) and (1.3), we developed multiple linear regression model and obtained the estimates of unkown parameters. In regression analysis, the major objactive is to predict the future observations and this is given by

$$\hat{Y}_i = x_i' b \qquad (1.4.1)$$

where x_i is a given row vector of independent variables.

Quite obviously, if the model is a good fit to the data, the difference between the actual value and the predicted value i.e.$Y_i - \hat{Y}_i$ where Y_i is the observed value, must be 'small'. On the other hand, if this difference is significantly 'large', there is enough scope to suspect the adequacy of the proposed model and one must search for alternative models. Thus, the difference $Y_i - \hat{Y}_i$ which is called 'Residual', plays an important

role in determining adequacy of the model.

In this section we discuss the residual and its various properties, which are useful in the later discussion.

Examining the residual is one of the most important tasks in any regression analysis. It involves the careful inspection of the difference between the observed and the predicted values of the responce variable after a regression equation is fitted to the data . Many-time simple graphical tools are used to study residuals for observing pattern in the data. First, formally we define residuals.

Definition (4.1.1) Residual: Let the model be

$$\underline{Y} = X\,\underline{\beta} + \underline{\varepsilon} \quad \text{and} \quad \hat{\underline{Y}} = X\,\underline{b}.$$

Then the i^{th} residual is definied as,

$$e_i = Y_i - \hat{Y}_i, \qquad i=1,2,\ldots n. \tag{1.4.2}$$

The residual vector is given by

$$\underline{e} = \underline{Y} - \hat{\underline{Y}} \tag{1.4.3}$$

Using the experssion (1.4.1), we have

$$\underline{e} = \underline{Y} - X\,\underline{b}$$

this yields ,

$$\underline{e} = \underline{Y} - X\,(X'X)^{-1}\,X'\,\underline{Y},$$

$$= \underline{Y}(I - H)\,\underline{Y}, \tag{1.4.4}$$

where $H = X\,(X'X)^{-1}\,X'$

Now in order to use residuals in checking the adequacy of the

model, we need the following relationship between 'Residual vector' and 'error vector'.

Result (1.4.1): $\underline{e} = (I - H)\ \underline{\varepsilon}$

Proof : We have

$$\underline{e} = (I - H)\ (X\ \underline{\beta}+\underline{\varepsilon}),$$

$$= X\ \underline{\beta}- HX\ \underline{\beta}+ \underline{\varepsilon}- H\ \underline{\varepsilon},$$

since $H= X\ (X'X)^{-1}\ X'$ we have

$$\underline{e} = X\ \underline{\beta}- X\ (X'X)^{-1}\ X'X\ \underline{\beta}+\underline{\varepsilon} - H\underline{\varepsilon},$$

$$=\underline{\varepsilon} - H\ \underline{\varepsilon},$$

Thus,

$$=(I - H)\ \underline{\varepsilon}. \qquad (1.4.5)$$

Hence the result.

From the experssion (1.4.5), it is clear that relationship between $\underline{e}$ and $\underline{\varepsilon}$ depend only on the matrix H. Using this result, we have the i^{th} residual as,

$$e_i = \varepsilon_i - \sum_{j=1}^{n} H_{ij}\varepsilon_j \quad \text{for} \quad i=1,2,..n. \qquad (1.4.6)$$

The experssion (1.4.6) shows that if H_{ij} is sufficintly small then the residual vector 'e' is very 'close' to the error vector ε.

Now for further study, we need to discuss the behavior of matrix H. The matrix H is also known as 'hat matrix '. It has the following properties.

1: H is symmetric matrix.

2: H is an idempotent matrix , since

$$H^2=H\,H = X\,(X'X)^{-1}\,X'\,X\,(X'X)^{-1}\,X' = X\,(X'X)^{-1}\,X' = H.$$

Now below we give ordinary residuals-

1:Ordinary residuals

The ordinary residual vector is given by

$$\underline{e} = \underline{Y} - \hat{\underline{Y}}.$$

From the result (1.4.1), the vector $\underline{e}$ can be experssed as

$$\underline{e} = (I - H)\,\underline{\varepsilon}.$$

This equation implies that the ordinary residuals are useful for checking the assumptions.

Result (1.4.2): 1 : $E(\underline{e}) = \underline{0}$

2 : $V(\underline{e}) = \sigma^2\,(I - H)$

Proof : From the assumptions (1.2.1), we have

$$E(\underline{\varepsilon}) = \underline{0}, \qquad V(\underline{\varepsilon}) = \sigma^2\,I,$$

From the result (1.4.5) we have,

$$\underline{e} = (I - H)\,\underline{\varepsilon}$$

Therefore,

$$E(\underline{e}) = E\left[(I - H)\,\underline{\varepsilon}\right]$$
$$= (I - H)\,E(\underline{\varepsilon})$$
$$= (I - H)\,\underline{0}$$
$$= \underline{0}$$

Now,

$$V(\underline{e}) = V\left[(I - H)\,\underline{\varepsilon}\right]$$

$$= (I - H)V(\underline{\varepsilon})(I - H)$$

$$= (I - H)\,\sigma^2 I_n\,(I - H)$$

$$= (I - H)\,\sigma^2$$

since $(I - H)$ is an idempotent matrix .

Thus the ordinary residuals have zero mean and they are having unequal variances.

In particular, the variance of the i^{th} residual is,

$$V(e_i) = (1 - H_{ii})\,\sigma^2, \qquad i = 1, 2 \ldots n$$

Where H_{ii} is the i^{th} diagonal element of H and the covariance between e_i and e_j $(i = j)$ is,

$$cov(e_i, e_j) = -H_{ij}\,\sigma^2$$

Result (1.4.3): Rank (H) = k+1

Proof : Since matrix H is an idempotent matrix, we get

$$rank(H) = trace\,(H),$$

$$= trace\left[X\,(X'X)^{-1}\,X'\right],$$

$$= trace\left[(X'X)^{-1}\,X'X\right],$$

$$= trace\,(I_{k+1}),$$

$$= k+1,$$

1.4.2 Sum of squares :

In this section, we discuss various sum of squares and associated results, which are needed for later discussion.

1 : Residual sum of squares :

The sum of squares of deviation of the observed Y_i's from their estimated expected values is usually known as residuals or error sum of squares and it is denoted by RSS.

Thus,

$$RSS = \sum_{i=1}^{n} (y_i - \hat{Y}_i)^2,$$

$$= (\underline{Y} - \hat{\underline{Y}})'(\underline{Y}-\hat{\underline{Y}}),$$

$$= \left[(I-H)\ \underline{Y}\right]' \left[(I-H)\ \underline{Y} \right],$$

$$= \underline{Y}'(I-H)\underline{Y}$$

since (I-H) is symmetric and idempotent matrix, we get

$$RSS = \underline{Y}'(I - X\ (X'X)^{-1}\ X')\underline{Y},$$

$$= \underline{Y}'\underline{Y} - \underline{Y}'X\ \underline{b},$$

This is a convenient form for computing the RSS.

Result (1.4.4): $\hat{\sigma}^2$ = RSS / (n-k-1) is an unbiased estimate of σ^2.

Proof : We have

$$RSS = \underline{Y}'\ (I - H\ )\underline{Y}$$

Thus,

$$E(RSS) = E\left[\underline{Y}'\ (I - H\ )\underline{Y}\right]$$

By using the results of matrix theory, we have

$$E(RSS) = tr\left[(I-H)\ \sigma^2\ I + \underline{b}'\ X\ (I-H)\ X\underline{b}\right]$$

where H is an idempotent matrix with rank (k+1) and also (I-H) is

an idempotent matrix with rank (n-k-1).

Consider the matrix X'(I - H) X ,

$$X'(I - H) X = X'\left[I - X (X'X)^{-1} X' \right] X,$$

$$= X'X - X'X (X'X)^{-1} X'X,$$

$$= X'X - X'X,$$

$$= 0$$

Using the above results, we get,

$$E(RSS) = (n-k-1)\,\sigma^2,$$

which yields ,

$$E(RSS / (n-k-1)) = \sigma^2,$$

that is ,

$$E(\hat{\sigma}^2) = \sigma^2,$$

Thus, $\hat{\sigma}^2$ is an unbiased estimate of σ^2.

2 :Total sum of squares (TSS):

The sum of squares of observed Y_i's is usually known as total sum of squares and denoted by TSS. Hence,

$$TSS = \underline{Y}'\underline{Y},$$

3 : Regression sum of squares (SSR):

The regression sum of squares is denoted by RSS .and it is given by,

$$SSR = TSS - RSS$$

$$= \underline{Y}'\underline{Y} - (\underline{Y}'\underline{Y} - \underline{b}'X'\underline{Y})$$

$$= \underline{b}'X'\underline{Y}.$$

.5 Goodness of Fit of a model:

Normally, the analysis of variance table is prepared for esting whether the proposed regression model is a good fit or ot. This is summarised below:

et the proposed model be $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ with the assumption $\underline{\varepsilon} = N(\underline{0}, \sigma^2 I_n)$. Thus the ANOVA table is given below:

Table 1.5.1

ource of variation	d.f.	S.S	M.S.	F-statistic
egression	k+1	SSR	MSR=SSR/(k+1)	F(R)=MSR/MSE
esidual	n-k-1	RSS	MSE=RSS/(n-k-1)	
otal	n	TSS		

able 1.5.2 ANOVA table (Showing a term mean)

urce of ariation	d.f.	S.S	M.S.	F-statistic
ean	1	SSM	MSS=SSM/1	F(M)=MSM / MSE
egression	k	SSR_m	$MSR=SSR_m/k$	$F(R)=MSR_m/MSE$
esidual	n-k-1	RSS	MSE=RSS/(n-k-1)	
otal	n	TSS		

Table (1.5.3) ANOVA Table (Corrected for the mean)

Source of variation	d.f.	S.S	M.S.	F-statistic
Regression	k	SSR_m	$MSR=SSR_m/k$	$F(R)=MSR_m/MSE$
Residual	n-k-1	RSS	MSE=RSS/(n-k-1)	
Total	n-1	TSS		

where SSR_m denotes the corrected regression sum of squares.

The table (1.5.1), (1.5.2), (1.5.3) all are summarizing the same thing as the table (1.5.3) is simply an abbrevited version of table (1.5.2) with SSM is removed from the body of the table and substracted from TSS to give TSS = TSS - SSM = $\underline{Y}'\underline{Y} - n\,\bar{\underline{Y}}^2$, the corrected sum of squares of the $\underline{Y}$ observations.

The general linear hypothesis: Some times, the following hypothesis are of interest :

$$H = C'\underline{\beta} = \underline{m} ,$$

where $\underline{\beta}$ is the (k+1) order vector of parameters of the model, C' is any matrix of order s × (k+1) and $\underline{m}$ is vector of order s of specified constants. There is only one limitation on C' that it has full row rank.

The F-statistic for testing the above hypothesis,

$$F(H)=\left[(C'\underline{b} - \underline{m})' \left[C' (X'X)^{-1} C \right]^{-1} (C'\underline{b}-\underline{m}) \right] \Big/ s\,\hat{\sigma}^2$$

with s and n-k-1 d.f.

The particular cases of general hypothesis are given below:

1 : H : $\underline{\beta}=\underline{0}$, the hypothesis that all elements in β are zero

2 : H : $\underline{\beta}=\underline{\beta}_o$, the hypothesis that $\beta_i=\beta_{io}$ for i=0,1,2,...,k

i.e each β_i is equal to some specified value β_{io}

3 : H : $l'\underline{\beta} = \underline{m}$ that some linear combination of $\underline{\beta}$ equals a specified constant.

4 : H $=\underline{\beta}_q = 0$ that some of the β_i's, q of them (q <k)are zero.

1.6 Need for selection of variables:

The purpose of this dissertation is to eliminate the irrelavent variables from the model. At the time of elimination or selection of the variables from the model, naturally the question arises is "which variables are to be deleted from the model or which variables are to be included in the model ?". The answer to this question is given in the dissertation. Now, we discuss need of subset selection.

A regression analysis may be carried out with one or more following objectives.

(a) To predict a variable which may depend on the regressors.

(b) To asses quantitatively the nature of dependence of the dependent variables on the regressors of the subset in the presence of other regressors.

(c) To build a working model to explain the association between the dependent variable and regressors.

(d) To examine statistically certain empirical belifes regarding the model.

In a study of regression analysis, sometimes the form of the model is known to the Statistician. In such a case, the problem of selection of variables is not serious problem. But in many practical situations model is not known clearly, only there is rough idea regarding the variable involved.

The object of analysis in above case is to arrive at a working model which may be adequate for the intended goal. For instance, consider an example: In official publication data, several variables are likely to be involved in collinear relationship with dependent variables. Building a model for such type of data involving a large set of variables is so difficult. In such a situation the problem of elimination of variables which exhibit insignificant effect on Y is more important.

The major advantage of selection of subset of variables is that a regression equation with fewer variables have the appeal of simplicity as well as economic advantages in terms of obtaining the necessary information to use the equation. In addition, other advantage in eliminating irrelavent variables is that the variance decreases for both estimation and prediction, if variables are deleted from the model.

Once the decision of reducing the number of variables is

taken, the question naturally arises as to which subset to select. The statistician may have some ideas regarding the regressors which are "absolutely" important. There are many known variable selection methods most of which falls in one of the two main categories.

1) Exhaustive search method: This method is based on examining all possible subset of predictor variables with respect to some criterion.These are discussed in detail in the chapter II.

2) Sequential (Systematic) selection algorithm such as the forward selection, backward elimination and stepwise method. These methods are given in the chapter III.

Other methods are also developed for subset selection in regression analysis. Kudo and Tarumi (1974) develop a algorithm for generating all subsets of p variables out of k. Beale etl.(1967) and Hocking and Leslile (1967) develop branch and bound technique for selection of the p variables subset out of k variables. Furnival and Wilson (1974) have described algorithm for branch and bound techniques.