Application of the Block Ridge Logistic Regression Model on Libyan Imports Data

By the student : Magde M Abdalah

Supervisor : DR.RAMI SALAH M.GEBRIL

spring 2012 - 2013

Contents

Chapter 3

Methodology of Block Ridge Regression

Chapter 4

Application and Results

Acknowledgement

Praise be to God first and foremost Praise be to God, who provided us health and wellness for the completion of this work, which would not have been without the will of the Almighty, and ask him luck and success.

Thanks to my supervisor, Dr. Rami S. Gebril for the suggestion of this field of study, and to give unlimited in providing tips and suggestions scientific task, that actually contributed to the success of this research and I wish him success and progress in the practical and scientific life.

Finally I would like to thank my family and my close friends, for their support to me throughout the study period, and do not forget to thank the faculty members in the department of statistics on the moral *support and to give scientific advice thanks to all.*

ABSTRACT

In logistic regression analysis, as in ordinary linear regression when dealing with large number of variables, may have a high probability that multicollinearity to be present in the model, which effect in the parameters of model. This problem has actually bad impact on model parameters and results, such as large variances and unreliable predictions.

In this thesis, the block ridge logistic regression is applied as a suggest solution to deal with the presence of moderate and severe cases of multicollinearity.

The Block ridge solution attempts to give each collinear variable its appropriated "weight" so no "extra" bias is gained in the logistic model.

This approach will be applied on a local data concerned with the Libyan imported an commodities from different countries around the world. The logistic model will classify these countries into major or minor exporters to Libya. Using the proposed block ridge solution, will lead to extra improvement in the application of the model.

List of Tables

List of Figures

CHPTER 1

General introduction

1.1 Introduction

Multivariate approach is the most important part of the statistical analysis, in which more than one variable are considered simultaneously, give more information of phenomena under study rather than use only one variable. Although in some studies may isolate each variable and study it individually, to discover how much this variable help of understanding, but often fail to understanding the full structure of the data. Advantage of multivariate analysis, variables need to be examined simultaneously in order to uncover the manner and key features in the data.

Multivariate analysis is split to descriptive and inferential part. The first, particularly in new areas of research, involves *data exploring* in an attempt to recognize any structure requiring explanation. At this stage, finding the question is often of more interest than seeking the subsequent answer. Instead, methods are sought that allow possibly unanticipated patterns in the data to be detected, opening up a wide range of competing explanations.

1.2 Overview Of Multivariate Methods

William R, Dillon and Matthew Goldstein [11] may an interest exists about the association between two sets of variables, where one set is the realization of a dependent or criterion measure then the appropriate class of techniques would be those designated as *dependence methods*. may an interested with the mutual association across all variables with no distinction made among variable types, one uses *interdependence methods*. The dependence methods seek to explain or predict one or more criterion measures based upon the set of predictor variables. Interdependence methods, on the other hand, are less predictive in nature and attempt to provide seeing into the underlying structure of the data by simplifying the complexities, primarily through data reduction.

1.2.1 Dependence Methods

Depending on the nature and the number of variables the researcher wishes to study, there are several multivariate techniques that can analyze dependence structures among these methods;

1. *Multiple Regression*: This is perhaps the most commonly known and used multivariate method . Multiple regression concerned with the study of the dependence of one variable, the dependent variable, on a set of variables, the predictor variables, with a view toward estimating or predicting the mean value of the dependent variable on the basis of the known values of the predictor variables.
2. *Discriminant Analysis*: Discriminant analysis, and specifically the two group models, is one of the more popular used techniques in the analysis of multiple measurements. Given a vector of p observed scores, denoted by X, known to belong to one of two groups, the basic problem is to find some function of the p scores which can accurately assign observations with the reading X into one of the two groups.

3. *Logit Analysis*: Logit analysis is appropriate when the single criterion measure is discrete and all the predictor variables are also categorical in nature.
4. *Multivariate Analysis Of Variance* (MANOVA): When multiple criterion measures are available and the goal is to assess the impact of various levels of one or more (experimental) variables on the criterion measures, multivariate analysis-of-variance is the appropriate data analysis technique.
5. *Canonical Correlation Analysis*: Canonical correlation analysis usually seek to determine the linear association between a set of predictor variables and a set of response variables.

1.2.2 Interdependence Methods

The choice of a multivariate technique for the analysis of interdependence structure among a set of variables depends on the nature of the data input. If the variables have at least interval scale properties then the following multivariate techniques form the most popular interdependence methods are;

1. *Principal Components Analysis*: Principal components analysis is a data reduction technique where the primary goal is to construct linear combinations of the original variables that account for as much of the total variation as possible.
2. *Factor Analysis*: Factor analysis, or more precisely the *common factor analysis model,* is also a data reduction technique
3. *Cluster Analysis*: Cluster analysis can be considered as another technique for data reduction. The goal in most studies that have used

cluster analysis technique is to identify a smaller number of groups such that elements residing in a particular group are, in some sense, more similar to each that other than to elements belonging to other group .

4. *Non-metric Multidimensional Scaling*: The goal of non-metric multidimensional scaling is to transform the perceived (dis)similarities between a set of objects into distances by placing those objects in multidimensional space of some dimensionality. In this sense , non-metric multidimensional scaling is identical to its metric counterpart.

1.3 Regression Analysis

In variety of statistical methodologies which are different from discipline to another, one of the most popular methodologies is the regression analysis, which is a statistical tool uses to study the relationship between one response variable and one or more of independent variables, to find the best model to fit the data. This technique is widely used in business, the social and behavioural sciences, the biological sciences, and many other disciplines.

1.3.1 Linear Regression Mode

Michael H, Christopher J& William Li [8] the model in linear regression is a linear combination of the parameter in the model (but need not be linear in the independent variables). In *simple linear regression*, will have only two variables suppose them as x and y and suppose that they are related by an expression of the form $y = \beta_0 + \beta_1 x + \varepsilon$. will leave aside for a moment the nature of the variables and focus on the x , y relationship. $y = b_0 + b_1 x$ is the equation of a straight line; b_0 is the *intercept* (or *constant*) and b_1 is the x *coefficient*, which represents the slope of the straight line the equation describes.

- The main assumptions of regression model

1. y is related to x by the simple linear regression model.
   $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, (i = 1,...,n).
2. The errors $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are independent of each other.
3. The errors $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ have a common variance σ^2.
4. The errors are normally distributed with a mean 0 and variance σ^2, that is, $\varepsilon \backslash X \sim N(0, \sigma^2)$.

The *multiple linear regression* has the same manner as in the simple linear regression model, but with more than one independent variable and the outcome effect by those dependent variables to give more information and nature relation between them and the outcome rather than one variable , in this case the model is called multiple linear regression model.

In general, can state a linear regression model in the form:

$$y_i = f(X_i , \beta) + \varepsilon_i \qquad ;i = 1,2,3, \dots, n$$

where X_i is the victor of the observation on the prediction variables for the ith case:

$$X_i = \begin{bmatrix} 1 \\ X_{i1} \\ . \\ . \\ . \\ X_{i,p+1} \end{bmatrix}$$

Where p is number of independent variables, β is the vector of regression coefficients , and $f(X_i, \beta)$ represents the expected value $E(y_i)$ where,

$$f(X_i, \beta) = X'_i\beta$$

1.3.2 Nonlinear Regression Model

Michael H, Christopher J& William Li [8] nonlinear regression models are of the same basic form as for linear regression models :

$$y_i = f(X_i, \gamma) + \varepsilon_i \qquad ; i = 1,2,3, \dots, n$$

An observation y_i is still the sum of a mean response $f(X_i, \gamma)$ given by the nonlinear response function $f(X_i, \gamma)$ and the error ε_i. The error usually assumed to have expectation zero, constant variance and to be uncorrelated , just as for linear regression model. Often, a normal error model is used which assumes that the error terms are independent normal random variables with constant variance. The parameter vector in the response function $f(X_i, \gamma)$ is now denoted by γ rather than β as before that the response function here is nonlinear in the parameters. The nonlinear model is similar in general form to linear regression model, each y_i observation is postulated to be the sum of a mean response $f(X_i, \gamma)$ based on the given nonlinear response function and random error term ε_i.

1.4 Goal of The Study

In his study we applied logistic regression on import data, to modeling the data from classify the response variable (minor export, major export), some of important points of study are;

- To deal with the problem of multicollinearity in logistic regression model, using the block ridge estimators, which includes modified ridge parameters that give weights to each collinear block of variables.
- To introduce a statistical strategy that deals with datasets having relatively large number of predictors in the nominator of exp $(\beta' X)$ in the multiple logistic model.

CAPTER 2

Data Description and exploration

2.1 Introduction

Every statistical data analysis task starts by gathering, characterizing, and organizing a new, unfamiliar data set. After this process, the data can be analyzed and the results delivered. In our experience, the first step is far more difficult and time consuming than the second. To start with, data gathering is a challenging task complicated by problems both sociological (such as turf sensitivity) and technological (different software and hardware platforms make transferring and sharing data very difficult). Once the data are in place, acquiring the metadata (data descriptions, business rules) is another challenge. Very often the metadata are poorly documented. When we finally are ready to analyze the data, its quality is suspect. Furthermore, the data set is usually too large and complex for manual inspection.
Sometimes, improved data quality is itself the goal of the analysis, usually to improve processes in a production database. Although the goal seems different than that of making an analysis, the methods and procedures are quite similar in both cases we need to understand the data, then take steps to improve data quality.

2.2 Dataset of Study

In this study an interest of economic dataset which have large number of variables comparing to small number of cases, the depended or response variable is a categorical variable describes the nature of the exporting country.

The source of the data is the " National Authority of Information and Documentation " in Libya . And they represent the imports of individual commodities by countries of origin during 2003.

The data features are quantities of imported commodities (measured by weight in kg and number of units) from a set of countries . The data matrix contains 53 columns represent $p = 53$ imported commodity and 39 rows represent $n = 39$ countries, that are classified into 21 minor exporters and 18 major exporters, for Libya .

Table (2.1) gives a brief description of the variables of the data set.

Table (2.1): Description of the variables

Variable Number	**Variable Name**
Y	Nature of the exporting country (minor exporter, major exporter)
1	Different dairy products and cheese
2	Fish and Aquatic Products
3	Wheat, barley and rice (milled and non-milled)
4	Other food industries
5	Dry beans
6	Fruits and vegetables (fresh or preserved)
7	Juices (natural or concentrated)
8	Sugar confectionery

Variable Number	Variable Name
9	Food containing chocolate
10	Drinks
11	Seeds (vegetables and fruits)
12	Cooking oil
13	Coffee (roasted and non-roasted)
14	Raw sugar and refined
15	Products and petroleum oils
16	Industrial chemical compounds
17	Supplies Plastics and rubber industry
18	Chemical fertilizers and natural
19	Coating material
20	Medical equipment and furniture (human and veterinary)
21	Medicines and medical supplies (human and veterinary)
22	Cosmetics
23	Pesticides (Insecticides and rodents)
24	Building Materials
25	Tires
26	Paper Crafts
27	Articles made of glass and ceramic
28	Supplies iron industry
29	Manufactures aluminium
30	Ware and Bathroom
31	Drilling machines and their parts
32	Water treatment and gas

Variable Number	Variable Name
33	Machines and various industrial pumps
34	Air-conditioning and Refrigeration
35	Parts and computer peripherals
36	Projectors visual and audible and phone
37	Parts monitors visual and audible and phone
38	Electricity and lighting kits
39	Electrical household appliances (MMS)
40	Cars and mobile machines
41	Auto Parts and Machinery
42	Aircraft parts
43	Furniture
44	Different clothes and accessories
45	Leather Shoes and plastic
46	Stationery and School tools
47	Carpeting
48	Tools and Home Furniture (non-electric)
49	Different fabrics
50	Books and Publications
51	Toys and entertainment
52	Parts and accessories of photography
53	Installations and tanks (iron and aluminium) and components

Some statistical procedures in a previous are applied on crude data (clustering and transformation), so the data will be proper for study. The use of transformation is to overcome the presence of extreme values, and the use of clustering is performed to classify the original dataset into homogeneous groups or blocks.

Table(2.2) shows the crude data after applied the procedures, where classified to five homogeneous blocks.

Table (2.2): Membership of variables to the blocks

Block Number	Variable name (X_j) j
1	16 , 25 , 28 , 30 , 31 , 34 , 35 , 40 , 41
2	8 , 14 , 20 , 23 , 37 , 38 , 39 , 47 , 48 , 49 , 53
3	2 , 7, 13 , 22 , 27 , 32 , 33 , 36 , 42 , 43 , 44 , 45 , 46
4	3 , 9 , 18 , 21 , 24 , 26 , 50 , 51 , 52
5	1 , 4 , 5 , 6 , 10 , 11 , 12 , 15 , 17 , 19 , 29

The following part shows some diagnostic plots for predictors, which can be helpful in exploring the behavior of the data in general. First scatter plot of predictor variable against each of other predictor variables in the blocks are done to study the nature relationship among the predictor variables and for finding gabs and detecting hidden outliers .

Figure(2.1) of the scatter plot of block one shows no clear manner between the intersection variables in the plot. Also there is no clear direction in the data generally, because there are many extreme values in the variables.

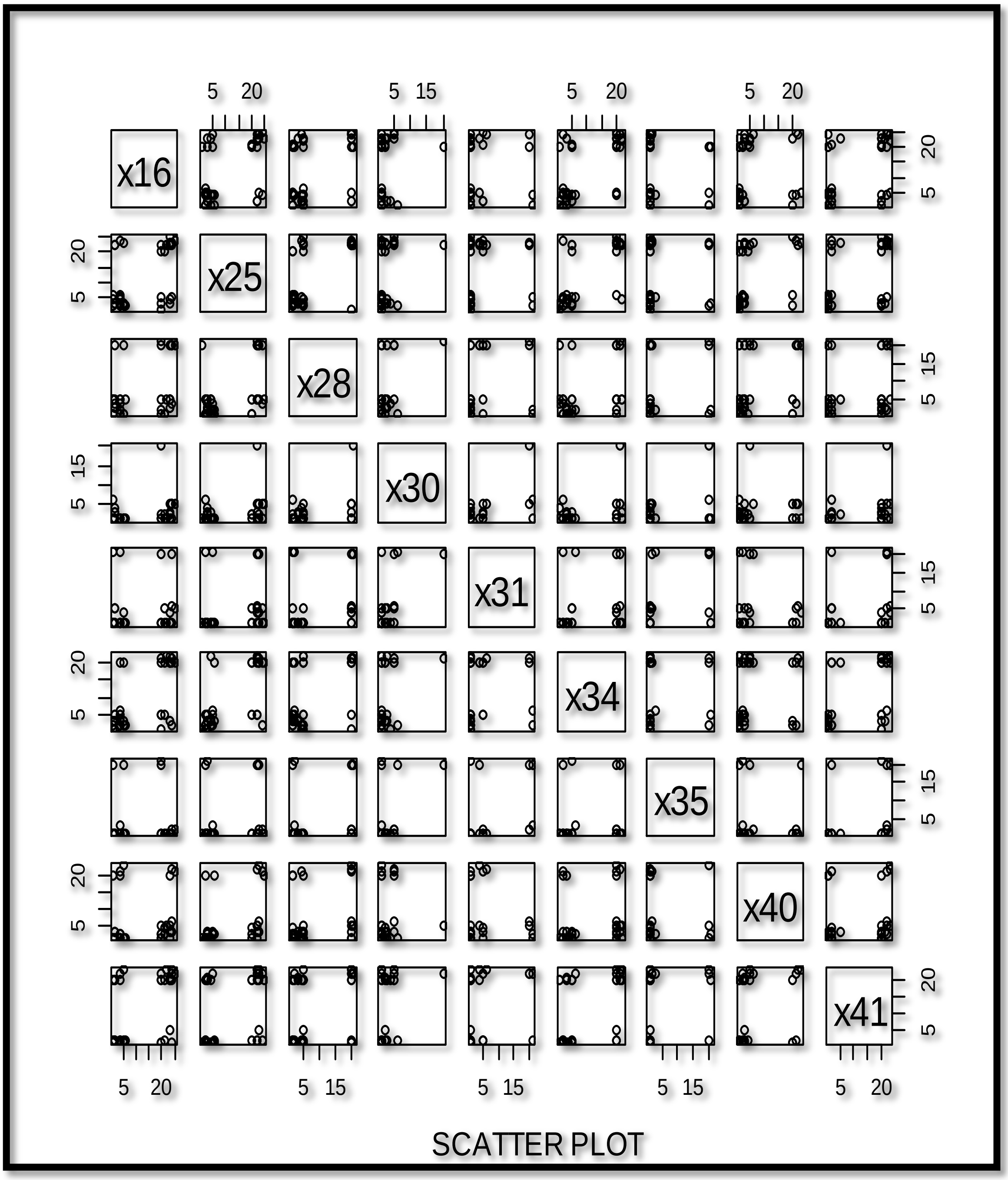


Figure (2.1) Scatter plot of block1

Other type of plot either can give us important and useful information of the data is boxplot, which is gives distribution of variables, dispersion, variation and extreme values in each variable.

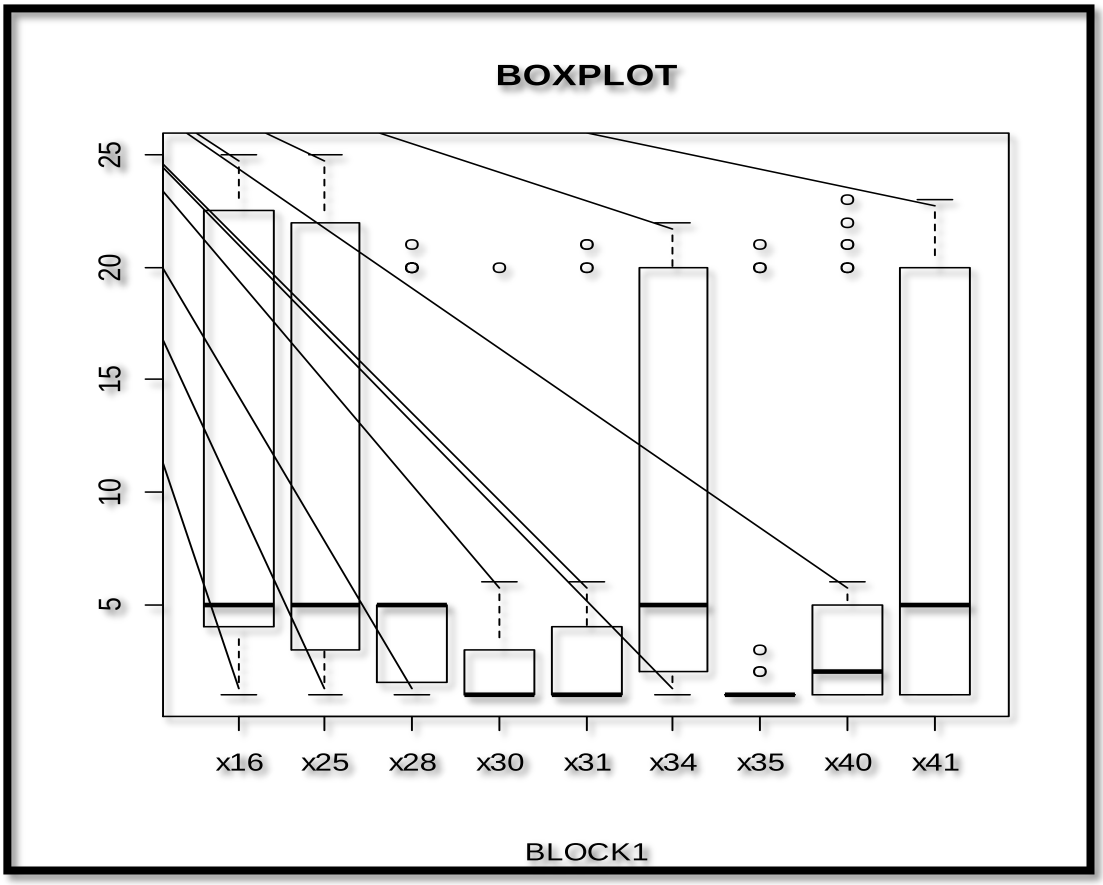


Figure (2.2) Box plot of block1

Figure (2.2) illustrates the description of the variables in block one, which a reflect an obvious difference between variable distributions, especially between($X_{16}, X_{25}, X_{34}, X_{41}$) and the other variables, which have less variation.

Figure(2.3) shows the boxplot of blocks(2,3,4,5), where in some variables (for example $X_2, X_8, X_{20}, X_{22}, X_{15}$) the extreme values a clearly present. Also it can be differences in variations between variables.

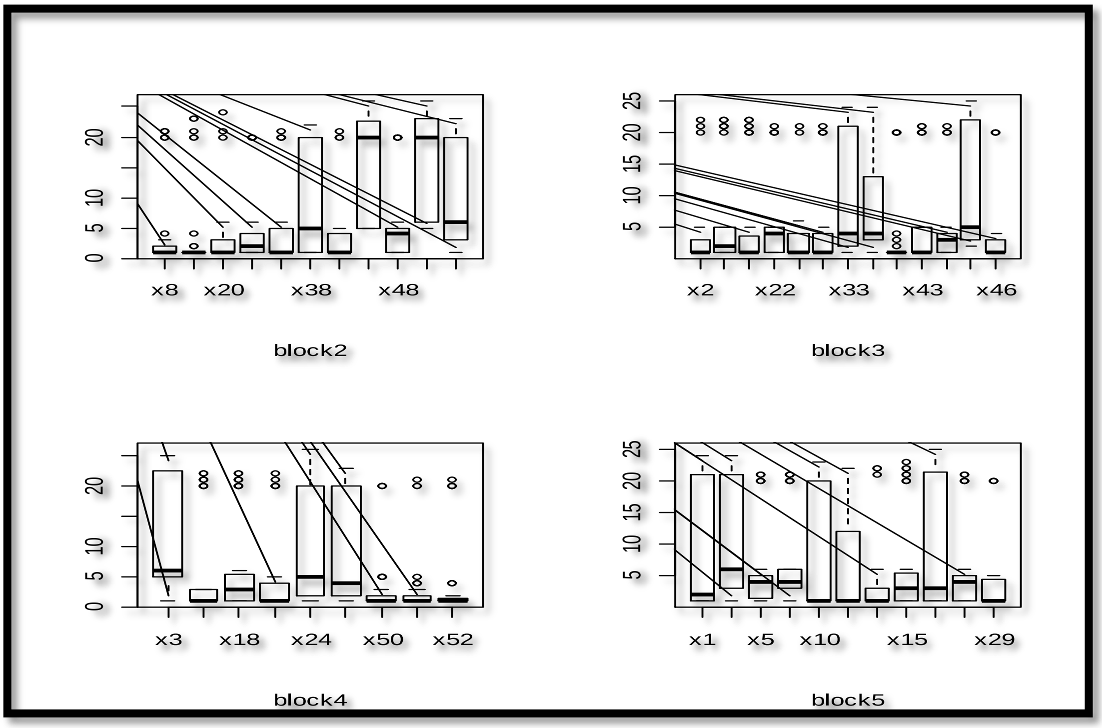


Figure (2.3) Boxplot of other blocks

CHAPTER 3

Methodology of Block Ridge Logistic Regression

3.1 Logistic Regression

This is chapter start with a brief explanation of the general case of logistic regression, Simon J. Sheather. et. al [9] where the response variable(y) has two outcomes or binomial random variable based on a single variable X. before considering the logistic regression we briefly review a few facts about the binomial distribution on which the logistic model based on.

A binomial process has the following properties :

1 . There are m identical trials.

2 . Each trial results in one of two outcomes , either a "success" S or a "failure" F.

3 . π , the probability of "success" is the same for all trials.

4 . Trials are independent.

The trials of the binomial process are called *Bernoulli trials*.

Let y be the number of successes in m trials of a binomial process ($y \leq m$) which means that y is a binomial distribution with parameters m and π, the short-hand notation for this is as follows :

$$y \sim Bin(m, \pi)$$

The probability that y takes an integer value j (j =0,1,2,...,m) is given by

$$(y= j) = \binom{m}{j} \pi^{\,j} (1-\pi)^{m-j} \qquad j=0,\ldots,m$$

the mean and variance of y are given by

$$E(y) = m\pi \ , V(y) = m\pi(1-\pi)$$

In logistic regression setting, to model π and hence y on the basis of predictors $x_1, x_2, \ldots, x_p$, where p is the number of predictors.

Begin with considering the case of a single predictor variable x. In this case

$$(y|\, x_i) \sim \text{Bin}(m_i\, , \pi(x_i)) \qquad i = 1, \ldots, n$$

The sample proportion of "successes " at each i is given by y_i/m_i. Notice that the variance of the response y_i/m_i depends on $\pi(x_i)$ and as such it is not constant. In addition, this variance is also therefore unknown. Thus, least square regression is an inappropriate technique for analyzing Binomial response.

1. Explanation of deviance

In logistic regression the concept of the residual sum of squares is presented by a measurement known as the *deviance*. In logistic regression the deviance is defined to be

$$G^2 = 2\sum_{i=1}^{n} \left[y_i log\left(\frac{y_i}{\hat{y}_i}\right) + (m_i - y_i) log\left(\frac{m_i - y_i}{m_i - \hat{y}_i}\right) \right]$$

where,

$$\hat{y}_i = m_i \pi(x_i).$$

2. Residuals for logistic regression

There are at least three types of residuals for logistic regression, namely,

• Response residuals.

• Pearson residuals and standardized Pearson residuals.

• Deviance residuals and standardized deviance residuals.

Response residuals is defined as the response minus the fitted values, that is,

$$r_{res} = \frac{y_i}{m_i} - \pi\,(x_i) \qquad i = 1, \dots, n$$

where $\pi\,(x_i)$ is the ith fitted value from the logistic regression model. However, since the variance of y_i/m_i is not constant, response residual can be difficult to interpret in practice.

$$r_{p_i} = \frac{\left(\frac{y_i}{m_i} - \pi\,(x_i)\right)}{\sqrt{\hat{V}ar(y_{i/}m_i)}} \qquad i = 1, \dots, n$$

where

$$\sum_i^n r_{p_i}^2 = \sum_{i=1}^n \frac{\left(\frac{y_i}{m_i} - \pi\,(x_i)\right)^2}{\pi\,(x_i)\left(1 - \pi\,(x_i)\right)/\mathrm{m_i}} = X^2$$

this is commonly cited as the reason for the name Pearson residuals. Pearson residuals do not account for the variance of $\pi\,(x_i)$. This issue is overcome by *standardized Pearson residuals* , which are defined to be

$$sr_{p_i} = \frac{\left(\frac{y_i}{m_i} - \pi\,(x_i)\right)}{\sqrt{\hat{V}ar(y_i/m_i - \pi\,(x_i))}} \qquad i = 1, \dots, n$$

$$sr_{p_i} = \frac{\left(\frac{y_i}{m_i} - \pi\,(x_i)\right)}{\sqrt{(1 - h_{ii})\pi\,(x_i)(1 - \pi\,(x_i))/\mathrm{m_i}}} = \frac{r_{p_i}}{\sqrt{(1 - h_{ii})}} \qquad i = 1, \dots, n$$

where h_{ii} is the $i$the diagonal element of the matrix obtained from the weighted least squares approximation to the Maximum Likelihood Estimate.

Deviance residuals are defined in an analogous manner to Pearson residuals with the Pearson goodness-of-fit statistic replaced by the deviance G^2, that is

$$\sum_i^n r_{Dev_i}^2 = G^2$$

thus, *deviance residuals* are defined by

$$r_{Dev_i} = \text{sign}(\frac{y_i}{m_i} - \pi\,(x_i))g_i \qquad i = 1, \dots, n$$

where $G^2 = \sum_i^n g_i^2$. Furthermore, *standardized deviance residuals* are defined to be

$$sr_{Dev_i} = r_{Dev_i} / \sqrt{1 - h_{ii}} \qquad i = 1, \dots, n$$

3.2 Binary Logistic Regression

In regression analysis when the dependent variable of interest has only two possible outcomes dichotomous, in this case called logistic regression. Therefore can be represented by a binary indicator variable taking two values, namely, 0 or 1.

Michael H, Christopher J and William Li [8] consider first the meaning of the response function when the outcome variable is binary, and then takeing up some special problems that arise with this type of response variable.

Consider the simple linear regression model:

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad y_i = 0,1; \quad i = 1,2,\dots,n$$

where the outcome y_i is a binary variable takes only the value of either 0 or 1. The expected response $E\ (y_i)$ has a special meaning in this case. Since $E\ (\varepsilon_i)$ =0 which can state:

$$E\ (y_i) = \beta_0 + \beta_1 X_i \qquad i = 1,\dots,n$$

consider y_i to be a Bernoulli random variable for which can state the probability distribution as follows :

Table (3.1) Probability of Bernoulli Distribution

$\boldsymbol{y_i}$	**probability**
1	$\mathbf{p}(\boldsymbol{y_i} = \mathbf{1}) = \boldsymbol{\pi_i}$
0	$\mathbf{p}(\boldsymbol{y_i} = \mathbf{0}) = \mathbf{1} - \boldsymbol{\pi_i}$

Thus, $\boldsymbol{\pi}_i$ is the probability that y_i = 1, whereas 1- $\boldsymbol{\pi}_i$ is the probability that y_i= 0. By the definition of expected value of a random variable obtain:

$$E\ (y_i) = 1(\boldsymbol{\pi}_i) + 0(1 - \boldsymbol{\pi}_i) = \boldsymbol{\pi}_i = p(y_i = 1) \qquad i = 1,\dots,n$$

then:

$$E\ (y_i) = \beta_0 + \beta_1 X_i = \boldsymbol{\pi}_i \qquad i = 1,\dots,n$$

The mean response E(y_i)= β_0 + $\beta_1 X_i$ as given by the response function is therefore simply the probability that y_i = 1 when the level of the predictor variable is X_i . This interpretation of the mean response applies whether the response function is a simple linear one , as here , or a complex multiple regression one, the mean response , when the outcome variable is a 0,1 indicator variable, always represents the probability that y = 1 for the given levels of the predictor variables.

Special problems arise, unfortunately, when the response variable is an indicator variable, consider three now, using a simple linear regression model as an illustration.

1 . *Non-normal Error Terms*

For a binary 0,1 response variable, each error term ε_i = y_i- (β_0 + $\beta_1 X_i$)

can take on only two values:

for y_i = 1 : $\varepsilon_i = 1 - \beta_0 - \beta_1 X_i \qquad i = 1, \dots, n$

and for y_i = 0 : $\varepsilon_i = -\beta_0 - \beta_1 X_i \qquad i = 1, \dots, n$

Clearly, normal error for linear regression model, which assumes that the ε_i are normally distributed, is not appropriate.

2 . *Non-constant Error Variance*

Another problem with the error terms ε_i is that they do not have equal variances when the response variable is an indicator variable. To see this, shall obtain $var(y_i)$ for the simple linear regression model

$$var(y_i) = E\,\{(y_i - E\{y_i\})^2\} = (1 - \pi_i)^2\pi_i + (0 - \pi_i)^2(1 - \pi_i)$$

or

$$var(y_i) = \pi_i(1 - \pi_i) = (E\{y_i\})(1 - E\{y_i\}) \quad i = 1, \dots, n$$

the variance of ε_i is the same as that of y_i because $\varepsilon_i = y_i - \pi_i$ and π_i is a constant:

$$var(y_i) = \pi_i(1 - \pi_i) = (E\{y_i\})(1 - E\{y_i\}) \quad i = 1, \dots, n$$

or

$$var(y_i) = (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) \qquad i = 1, \dots, n$$

Not from the last equation that $var(y_i)$ depends on X_i. Hence, the error variances will differ at different levels of X, and ordinary least squares will no longer be optimal.

3 . *Constraints on Response Function*

since the response function represents probabilities when the outcome variable is a 0,1 indicator variable, the mean responses should be constrained as follows :

$$0 \le (\{y\} = \pi) \le 1$$

Many response functions, do not automatically possess this constraint. A linear response function, for instance, may fall outside the constraint within the range of the predictor variable in the scope of model.

The difficulties created by the need for the constrain on the response function are the most serious. One could use weighted least squares to handle the problem of unequal error variances. In addition, with large sample sizes the method of least squares provides estimators that are asymptotically normal. However, the constraint on the mean responses to fall between 0 and 1 frequently will rule out a linear response function.

3.2.1 Simple Logistic Regression Model

The whole text of this section is totally taken from Michael H, Christopher J and William Li [8].

In simple logistic regression model use maximum likelihood to estimate the parameters of the logistic regression model. This method is well suited to deal with the problems associated with the response y_i being binary. First need to develop the joint probability function of the sample observations. Instead of using the normal distribution for Y observations as that of the ordinary linear regression, using the Bernoulli distribution for a binary random variable.

First, a formal statement of the simple logistic regression model is required. Recall that when the response variable is binary, taking on the values 1 or 0 with probabilities π and $1-\pi$ respectively, y is a Bernoulli random variable with parameter $E(y) = \pi$. The simple logistic regression model in the usual form :

$$y_i = E(y_i) + \varepsilon_i \qquad\qquad i = 1, \dots, n$$

Since the distribution of the error term ε_i depends on the Bernoulli distribution of the response y_i, it is preferable to state the simple logistic regression model in the following fashion:

y_i are independent Bernoulli random variables with expected values $E(y_i) = \pi_i$, where:

$$E(y_i) = \pi_i = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)} \qquad i = 1, \dots, n$$

The X observations are assumed to be known constants. Alternatively, if the X observations are random, $E(y_i)$ is viewed as a conditional mean, given the value of X_i.

- *Likelihood function*

Since each y_i observation is an ordinary Bernoulli random variable , where:

$$p(y_i = 1) = \pi_i \qquad i = 1, \dots, n$$

$$p(y_i = 0) = 1 - \pi_i \qquad i = 1, \dots, n$$

The probability distribution can be represented as follows:

$$f_i(y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i} \qquad y_i = 0,1 \;; \quad i = 1, \dots, n$$

Note that $f_i(1) = \pi_i$ and $f_i(0) = 1 - \pi_i$. Hence, $f_i(y_i)$ simply represents the probability that $y_i = 1 \; or \; 0$.

Since the y_i observations are independent, their joint probability function is:

$$g(y_{1,\dots,}y_n) = \prod_{i=1}^{n} f_i(y_i) = \prod_{i=1}^{n} \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$$

Again, it will be easier to find the maximum likelihood estimates by working with the logarithm of the joint probability function:

$$log_e g(y_{1,\ldots,}y_n) = log_e \prod_{i=1}^{n} \pi_i{}^{y_i} (1 - \pi_i)^{1-y_i}$$

$$= \sum_{i=1}^{n} [y_i \, log_e \, \pi_i + (1 - y_i) log_e \, (1 - \pi_i)]$$

$$= \sum_{i=1}^{n} \left[y_i log_e \left(\frac{\pi_i}{1 - \pi_i} \right) \right] + \sum_{i=1}^{n} log_e (1 - \pi_i)$$

Since $E(y_i) = \pi_i$ for a binary variable , then,

$$1 - \pi_i = [1 + \exp{(\beta_0 + \beta_1 X_i)}]^{-1} \quad i = 1, \ldots, n$$

and

$$log_e \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 X_i \qquad i = 1, \ldots, n$$

Therefore,

$$log_e L(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i(\beta_0 + \beta_1 X_i) - \sum_{i=1}^{n} log_e [1 + \exp{(\beta_0 + \beta_1 X_i)}]$$

where $L(\beta_0, \beta_1)$ replaces $g(y_{1,\ldots,}y_n)$ to show explicitly that now view this function as a likelihood function of the parameters to be estimated, given the sample observations.

- *The Maximum Likelihood Estimation*

The maximum likelihood estimates of β_0 and β_1 in the simple logistic regression model are those values of β_0 and β_1 that maximize the log-likelihood function. No closed-form solution exists for the values of β_0 and β_1 for the last equation that maximize the log-likelihood function.

There are several widely used numerical search procedures, one of these employs *iteratively reweighted least squares*, will depend on standard statistical software programs specifically designed for logistic regression to obtain the maximum likelihood estimates.

- *Interpretation of* β_1

The interpretation of the estimated regression coefficient β_1 in the fitted of simple logistic regression response function is not the straightforward interpretation of the slope in a linear regression model. the reason is that effect of a unit increase in X varies for the logistic regression model according to the starting point on the X scale. An interpretation of β_1 is found in the property of the fitted logistic function that the estimated odds $\pi/(1-\pi)$ are multiplied by exp (β_1) for any unit increase in X.

3.2.2 Multiple Logistic Regression Model

The simple logistic regression model is easily extended more than one predictor variable. In fact, several variables are usually required with logistic regression to obtain adequate description and useful predictions. In extending the simple logistic regression model, by replace $\beta_0 + \beta_1 X$ by $\beta_0 + \beta_1 X_1 + \cdots + \beta_{p+1} X_p$. To simplify the formulas, the matrix notation and the following two vectors are used :

$$\beta_{(p+1)\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_{p+1} \end{bmatrix} \qquad X_{(p+1)\times 1} = \begin{bmatrix} 1 \\ X1 \\ X2 \\ . \\ . \\ . \\ X_p \end{bmatrix}$$

This yields the following two equations:

$$X'\beta = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p+1} X_p$$

$$X_i'\beta = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p+1} X_{i,p}$$

with this notation, the simple logistic response function extends to multiple logistic response function as follows :

$$E(y) = \frac{\exp(X'\beta)}{1 + \exp(X'\beta)}$$

and the equivalent simple logistic regression response is extends to:

$$E(y) = [1 + \exp(-X'\beta)]^{-1}$$

the multiple logistic regression model can therefore be stated as follows

$$E(y_i) = \pi_i = \frac{\exp(X_i'\beta)}{1 + \exp(X_i'\beta)} \qquad i = 1, \dots, n$$

Again, the X observations are considered to be known constants. Alternatively, if the X variables are random, $E(y_i)$ is viewed as a conditional mean, given the values of $X_{i1}, \dots, X_{i,p-1}$.

Like the simple logistic response function, the multiple logistic response function is monotonic and sigmoidal in shape with respect to $X'\beta$ and is almost linear when π is between 0.2 and 0.8. The X variables may be different predictor variables, or some may represent curvature and/or interaction effects. Also, the predictor variables may be quantitative, or they may be qualitative and represented by indicator variables. This flexibility makes the multiple logistic regression model very attractive.

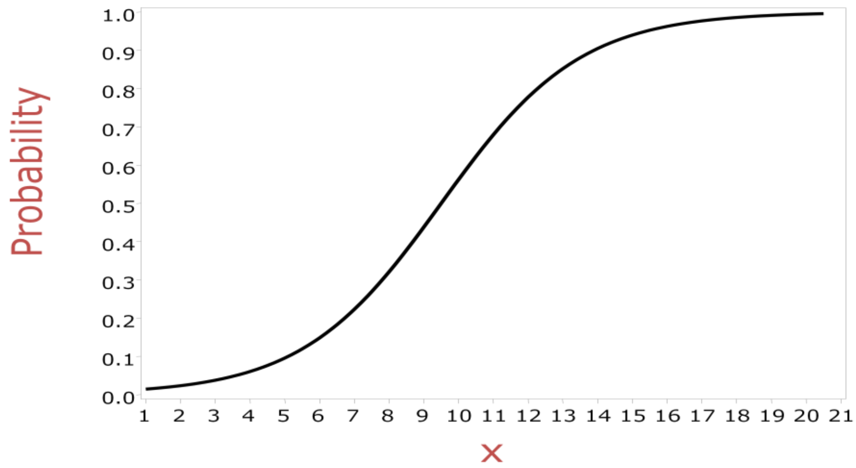


Figure (3.1) Logistic Regression Curve

- *Fitting of logistic model*

Again, use the method of maximum likelihood to estimate the parameters of the multiple logistic regression. The log-likelihood function for simple logistic regression of simple logistic model extends directly for multiple logistic regression:

$$log_e L(\beta\ ) = \sum_{i=1}^{n} y_i(X_i{}'\beta) - \sum_{i=1}^{n} log_e\left[1 + \exp\left(X_i'\beta\right)\right]$$

Numerical search procedures are used to find the values of $\beta_0 , \beta_1 , \dots , \beta_{p+1}$ that maximize $log_e L(\beta\ )$. These maximum likelihood estimates will be denoted by $b_0 , b_1 , \dots , b_{p+1}$. Let b denote the vector of the maximum likelihood estimates:

$$b_{(p+1)\times 1} = \begin{bmatrix} b_0 \\ b_1 \\ . \\ . \\ . \\ b_{p+1} \end{bmatrix}$$

The fitted logistic response function and fitted values can then be expressed as follows:

$$\pi = \frac{\exp(X'b)}{1 + \exp(X'b)} = [1 + \exp(-X'b)]^{-1}$$

$$\pi_i = \frac{\exp(X_i'b)}{1 + \exp(X_i'b)} = [1 + \exp(-X_i'b)]^{-1} \qquad i = 1, \dots, n$$

where,

$$X'b = b_0 + b_1X_1 + \dots + b_{p+1}X_p$$

$$X_i'b = b_0 + b_1X_{i1} + \dots + b_{p+1}X_{i,p}$$

3.3 Inferences of Logistic Regression Model

The same type of inferences are of interest in logistic regression as for linear regression models inferences about the regression coefficients, estimation of mean responses , and predictions of new observations.

The inference procedures will rely on large sample size. For large samples, under generally applicable conditions, maximum likelihood estimators for logistic regression are approximately normally distributed, with little or no bias. Also with approximate variance and matrix that are function of the second-order partial derivatives of the logarithm of the likelihood function.

The estimated approximate variance and covariance matrix are routinely provided by most logistic regression computer packages.

Inferences about the regression coefficients for the simple logistic regression model or the multiple logistic regression model are based on the following approximate result when the sample size is large :

$$\frac{b_k - \beta_k}{s\{b_k\}} \sim z \qquad k = 0,1,\dots,p+1$$

where z is a standard normal random variable and $s\{b_k\}$ is the estimated approximate standard deviation of b_k.

- *Test concerning a single* β_k :(*Wald test*)

A large sample test of a single regression parameter can be constructed based on the last condition as,

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

an appropriate test statistic is : $z^* = \dfrac{b_k}{s\{b_k\}}$

and the decision rule is :

If $|z^*| \leq z_{(\propto/2)}$, accept H_0

If $|z^*| > z_{(\propto/2)}$, reject H_0

One-sided alternatives will involve a one-sided decision rule. The testing procedure for coefficients is commonly referred to as Wald test. On occasion , the square of z^* is used instead , and the test is then based on a chi-square distribution with 1 degree of freedom. This also referred to as the Wald test.

Interval estimation of single β_k

From the last condition, the approximate $(1-\propto)$ confidence limits for β_k is obtained directly as follows:

$$b_k \pm z_{(\propto/2)} s\{b_k\}$$

where β_k is the parameter under test and $z_{(\propto/2)}$ is the $(z_{(\propto/2)})100$ percentile of the standard normal distribution.

The corresponding limits for the odds ratio exp (b_k) are:

$$\exp\,[b_k \pm z_{(\propto/2)}]$$

- *Likelihood Ratio Test* : *Test whether several* $\beta_k = 0$

Frequently there is interest in determining whether a subset of the X variables in a multiple logistic regression model can be dropped, that is, in testing whether the associated regression coefficients β_k equal zero.

The test procedure employ here is a general one for use with maximum likelihood estimation, and is analogous to the general linear test procedure for linear models. The test is called the *likelihood ratio test*, and, like general linear test, is based on a comparison of full and reduced models. The test valid for large sample size.

will begin with the full logistic model with response function:

$$\pi = [1 + \exp\,(-X'\beta_F)]^{-1}$$

where

$$X'\beta_F = \beta_0 + \beta_1 X_1 + \dots + \beta_{p+1} X_p$$

Where F represent of model with all parameters (all variables inter to fit model). Now finding the maximum likelihood estimates for the full model, denoted by b_F and evaluate the likelihood function $L(\beta)$ when $\beta_F = b_F$. The value of the likelihood function for the full model is denoted by $L(F)$.

The hypothesis aimed to be tested is:

$$H_0 : \beta_q = \beta_{q+1} = \dots = \beta_{p+1} = 0$$

Against

$$H_1\text{: Not all of the } H_0 \text{ equal zero}$$

Where, for convenience, arrange the model so that the last $p - q$ coefficients are those tested. The reduced logistic model therefore has the response function:

$$\pi = [1 + \exp(-X'\beta_R)]^{-1}$$

where,

$$X'\beta_R = \beta_0 + \beta_1 X_1 + \dots + \beta_{q-1} X_{q-1}$$

Now obtain the maximum likelihood estimates b_R for the reduced model and evaluate the likelihood function for the reduced model containing q parameters when $\beta_R = b_R$.

Will denote this value of the likelihood function for the reduced model by $L(R)$. It can be shown that $L(R)$ cannot exceed $L(F)$ since one cannot obtain a large maximum for the likelihood function using a subset of the parameters.

The actual test statistic for the likelihood ratio test, denoted by G^2, is:

$$G^2 = -2log_e\left[\frac{L(R)}{L(F)}\right] = -2[log_e L(R) - log_e L(F)]$$

Note that if the ratio $L(R)/L(F)$ is small , indicating H_1 is the appropriate conclusion , then G^2 is large . Which leads to conclusion of H_1.

Large-sample theory states that when n is large, G^2 is distributed approximately as χ^2 with degrees of freedom correspond to $df_R - df_F = (n-q)-(n-p) = p-q$. The appropriate decision rule therefore is:

$$If\ G^2 \leq \chi^2(1-\alpha\,; p-q), \text{accept } H_0$$

$$If\ G^2 > \chi^2(1-\alpha\,; p-q), \text{reject } H_0$$

Note that if the large-sample conditions for inferences are not met, the bootstrap procedure can be employed to obtain confidence limits for the regression coefficients.

3.4 Model Selection Criterion

In the context of multiple regression models, some procedures are proposed to chose the variables of data. For logistic regression modelling, the AIC_p and SBC_p criteria are easily adapted and are generally available in commercial software. The focus will be on the use of AIC_p, $LRT(G_p^2)$ criteria. The modifications is as follows :

$$AIC_p = -2Log_e L(b) + 2p$$

The $Log_e L(b)$ is the log-likelihood function for logistic regression model. Promising models will relatively have small values for this criteria .

1. *Best Subsets Procedures*

These procedures identify a group of subset models from all combination of independent variables(2^p model), those give the best values of a specified criterion. As long as the number of parameters is not large, these procedures can be useful.

2. *Stepwise Model Selection*

When the number of predictors is large the use of all-possible-regression procedures for model selection may not be feasible. Stepwise selection procedures are generally employed. Since obtain an analogous procedure by basing the decision on the Wald statistic z^* for kth estimated regression parameter, such as the forward stepwise, forward selection, and backward elimination algorithms is straightforward.

3.5 Multicollinearity problem

In multiple regression analysis, when there are highly correlated among the predictor variables (often referred to as multicollinearity), particularly, in the situation where regression coefficients can have the wrong sign and/or many of the predictor variables are not statistically significant where as the overall F-test is highly significant.

- *Effects of multicollinearity*

1. The fact that some or all predictor variables are correlated among themselves does not, in general, inhibit our ability to obtain a good fit nor does it tend to affect inferences about mean response or predictions of now observations, provided these inferences are made within the region of observations.
2. The counterpart in real life to many different regression function providing equally good fits to the data and the estimated regression coefficients tend to have large sampling variability when the predictor variables are highly correlated. Thus, the estimated regression coefficients tend to vary widely from one sample to another when the predictor variables are highly correlated.
3. The common interpretation of a regression coefficient as measuring in the expected value of the response variable when the given predictor variable is increased by one unit while all other predictor variables are held constant is not fully applicable when multicollinearity exists.

- *Informal diagnostics*

1. Large change in the estimated regression coefficients when a predictor variable is added or deleted.

2. Non significant results in individual tests on the regression coefficients for important predictor variables .

3. Estimated regression coefficients with an algebraic sign that is the opposite of what expected from theoretical consideration or prior experience.

4. Large coefficients of simple correlation between pairs of predictor variables in the correlation matrix .

5. Wide confidence intervals for the regression coefficients representing important predictor variables.

- *Variance inflation factor*

A formal method of detecting the presence of multicollinearity that is widely accepted is the use of variance inflation factor. These factors measure how much the variances of the estimated regression coefficients are inflated as compared to those where the predictor variables are not linearly related. This gives us quantitative measurements of the impact of multicollinearity.

The diagonal element $(VIF)_k$ is called the *variance inflation factor* (VIF) for b_k. It can be shown that this variance inflation factor is equal to :

$$(VIF)_k = (1 - R_k^2)^{-1} ; \qquad k = 1,2, \dots , p - 1.$$

Where R_k^2 is the coefficient of determination when X_k is regressed on the $p - 2$ other X variables in the model .

The largest VIF value among all X variables is often used as an indicator of the severity of multicollinearity. A maximum VIF value in excess of 10 is frequently taken as an indication that multicollinearity may be unduly influencing the least squares estimates.

- *Some remedial methods*

1. Eliminate some predictors from the model.
2. Design an experiment in which the pattern of correlation is broken.
3. Principal Components Regression (PCR).
4. Ridge Regression.

3.6 Ridge Regression

The ridge regression estimator(b^R), is defined by Hoerl and Kennard, for some $k \geq 0$, by

$$(b^R) = (X'X + kI)^{-1}X'Y$$

where k is ridge parameter, If $k = 0$, $(b^R) = (b)$, the least squares estimator, while large k will move (b^R) away from least squares, and increase the bias in the estimate.

Ridge regression is one of several methods that have been proposed to remedy multicollinearity problems by modifying the method of least squares to allow biased estimators of the regression coefficients. When an estimator has only a small bias and is substantially more precise than an unbiased estimator, it may well be the preferred estimator since it will have a large probability of being close to the true parameter value.

Figure (3.2) illustrates this situation. Estimator (b) is unbiased but imprecise, whereas estimator (b^R) is much more precise but has a small bias. The probability that (b^R) falls near the true value β is much greater than that for the unbiased estimator b.

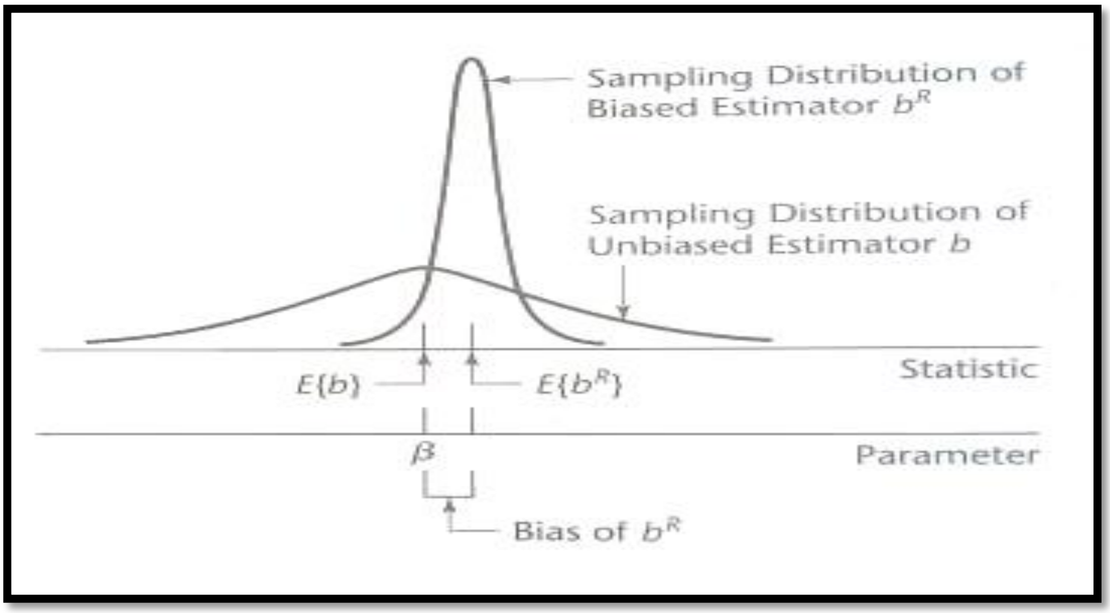


Figure (3.2) Ridge regression estimator

source: Michael H, Christopher J and William Li [8]

Generally, k is an unknown parameter which can be set by the analyst. A commonly used method of determining the ridge parameter k is based on the ridge trace. The ridge trace is a simultaneous plot of the values of p estimated ridge regression coefficients for different values of k, usually between 0 and 1, as illustrated in figure(3.3) .

Extensive experience has indicated that the estimated regression coefficients b_k^R may fluctuate widely as k is changed slightly from 0, and some may even change signs. Large value of k correspond to increased bias but lower variance, so a value of k must be chosen to balance bias against variance.

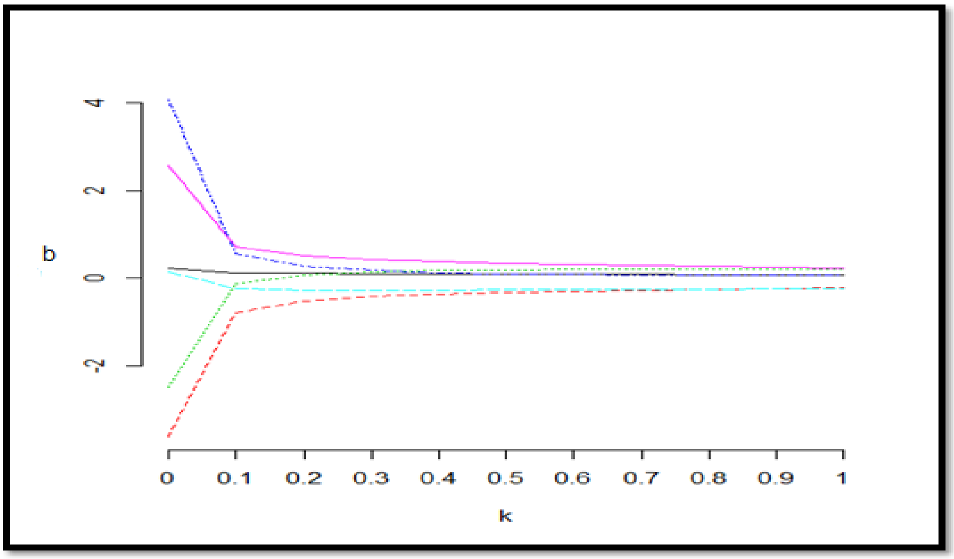


Figure (3.3) Ridge Trice(program with R software)

3.6.1 Generalized Ridge Regression

One generalization of the ridge regression rule is replace the ridge parameter k by a victor of parameters $(k_{1\,,}\, k_2\,, \dots, k_p)$. In terms of the original coordinates , defined G to be the $p \times p$ matrix such that, if k is the diagonal matrix with $k_{1\,,}\, k_2\,, \dots, k_p$ on the diagonal, $G = UKU'$, with U as an orthogonal transformation . Then the generalized ridge estimator is

$$b^{GR} = (X'X + G)^{-1}X'Y$$

Note in particular if $k_1 = k_2 = \cdots = k_p = k$, then $K = kI$, and

$G = U(kI)U' = kI$, so ridge regression is a special case of generalized ridge regression. (Hoerl and Kennard; Bingham and Larntz.1970).

3.6.2 Block Ridge Logistic Regression

Since the ridge regression is procedure to overcome of multicollinearity problem, this procedure is looking for one value (usually a small value between 0 and 1) to solve this problem, but the correlation between the independent variables difference from pair to pair or group to another in severity of multicollinearity, the ordinary ridge logistic regression introduced by Schaefer et. al. (1984), then by Duffy and Santner. (1989) the ridge logistic regression estimator is introduced as follows:

$$(b^R) = (X'VX + kI)^{-1}X'VXb$$

Where k is the ridge parameter, $V = diag(v_i) = diag(\pi_i(1 - \pi_i))$ is a diagonal matrix that contains the variance of y_i .

Block ridge logistic regression procedure work by partitioning the data to (m) blocks, with respect to severity of correlation between the predictors, which each block have variables that to be close in severity of multicollinearity, next step determining the ridge logistic regression parameter of each block, which the severity of multicollinearity difference from block to anther then will get different ridge parameter at each block, which it is work as weight of problem at each block, where replaced k in ordinary ridge logistic regression with diagonal matrix that contain ridge parameters of each block $(k_1, k_2, \dots, k_m)$, then the block ridge logistic regression can be written as:

$$b^{BR} = (X'VX + K)^{-1}X'VXb$$

Where K is a diagonal matrix contained $(k_1, k_2, \dots, k_m)$, if $K = k$, return to ordinary ridge logistic regression, and $K = k_1 = k_2 = \cdots = k_m = 0$, return to ordinary logistic regression.

On the other hand using the ordinary ridge logistic regression, will have one ridge parameter that work as weight for all variables to solve the problem, maybe some variables need less weight because the collinearity is weak, and some other high weight, then estimators don't be optimal in present of multicollinearity.

Figure(3.4), illustrates the idea of block ridge logistic regression as solution of multicollinearity, it can be seen that each block will has an appropriate ridge parameter (weight) of collinear in the block.

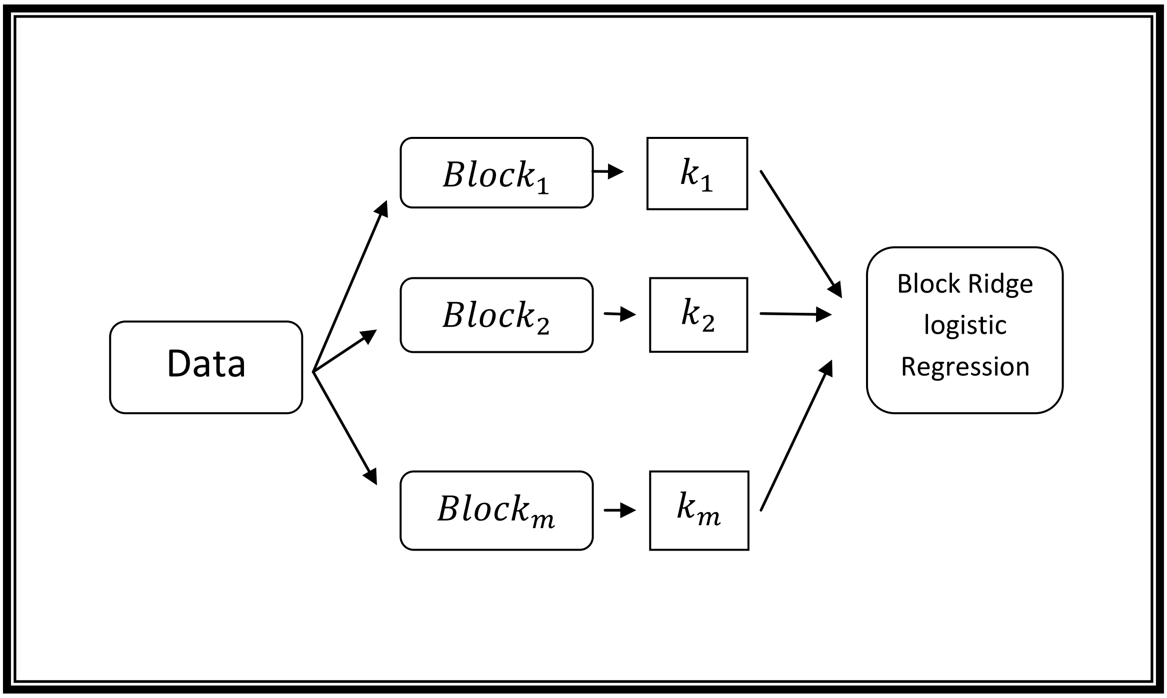


Figure (3.4) Block ridge logistic regression

CAPRER 4

Application And Results

4.1 Introduction

In this chapter, the results of performing the strategy of block ridge logistic regression on the Libyan import data in the following order ;

First, the procedure of "best subset variable selection" is applied on each block in order to choose those variables which possess most causal relation multicollinearity within each block of variables. Secondly, focusing on the effects of the presence of multicollinearity in each block. Third, the generalized matrix of ridge parameters will obtain use the original logistic regression parameters, and a justification is presented on the statistical importance of using this solution.

4.2 Results of The Best Subset Procedure

The method that adopted in the application to choose the best variables, the best subset procedure that give the best logistic model, which mean that the best variables from all, the best subset procedure to choose that variables at each block will be used, which will make combinations from all variables, from one to number of variables in block minus one (all possible subset available), the criteria to choose the best subset of variables is check the significant of the likelihood ratio test of the model, this feature of the study rather than get the best subset from all data, will get the best subset variables from blocks.

Table(4.1) gives summary of the best subset method, all subset of models those contain one predictor variable, which have $C_1^9 = 9$ available logistic model, with p-value of likelihood ratio test, which test of whether model of logistic regression with intercept difference of model with full parameters.

Table (4.1) Summary of best subset procedure with one variable model of block 4

Variable in the model	p-value
X_3	0.000
X_9	0.737
X_{18}	0.608
X_{21}	0.603
X_{24}	0.000
X_{26}	0.000
X_{50}	0.745
X_{51}	0.206
X_{52}	0.728

Table(4.1) shows that the variables those included in the models. There are three models those contain (X_3, X_{24}, X_{26}) the alternative (H_1) were accepted, which means that, each variable when included in model gives information about the response variable from the model, it is found that (X_3, X_{24}, X_{26}) are the best variables of block four in case of one predictor variable included in the model.

Table(4.2) illustrates models those contain two predictor variables (all subset available of models contains two independent variables). Which yields $C_2^9 = 36$ models, where can see the effect in models those contain two variable in the logistic regression, with p-value of likelihood ratio test, which distributed with χ^2distribution with degrees of freedom $(p - q)$.

Table (4.2) Summary of best subset procedure with two variables model of block 4

	p-value							
Variable	X_{52}	X_{51}	X_{50}	X_{26}	X_{24}	X_{21}	X_{18}	X_9
$\boldsymbol{X_3}$	0.000	0.00	0.000	0.000	0.000	0.000	0.000	0.000
$\boldsymbol{X_9}$	0.919	0.419	0.933	0.000	0.000	0.872	0.877	
$\boldsymbol{X_{18}}$	0.866	0.444	0.868	0.000	0.000	0.830		
$\boldsymbol{X_{21}}$	0.860	0.449	0.871	0.000	0.000			
$\boldsymbol{X_{24}}$	0.000	0.000	0.000	0.000				
$\boldsymbol{X_{26}}$	0.000	0.000	0.000					
$\boldsymbol{X_{50}}$	0.925	0.449						
$\boldsymbol{X_{51}}$	0.383							

Table(4.2) shows the variables included in the models, there are three variables those effect in the models. The alternative (H_1) was accepted when one or two of these variables (X_3 , X_{24} , X_{26}) included in the model, which the effects is clear, if one or two of these variables included in the models. In this step we note that (X_3 , X_{24} , X_{26}), are the best variables in this step too for block four.

Applied this procedure on this block for four ,five ,…,eight predictor variables, conclude that the same variables (X_3 , X_{24} , X_{26}) are the best in the block four, where the effect of these variables is clear when included in all models.

The same procedure is applied in order on the other blocks, the summary of best subset procedure for variables selection is shown in table (4.3). Then from the summary of the best subset procedure, there is no variables were chosen in the block one and block two, one variable were chosen from block three, two variables were chosen from block five.

Table (4.3) Summary of best subset procedure of all blocks

Block number	1	2	3	4	5
Variable name	-	-	X_{33}	X_3 X_{24} X_{26}	X_4 X_{17}

4.3 Check of Multicollinearity

One of most important procedure in regression analysis, is detecting the multicollinearity in the regression model. There are many ways to detect this problem, one of them is Variance Inflation Factor (VIF), which is special detecting technique, since it gives the numerical value of the severity of the multicollinearity in the model. Adopting this criterion leads to detect the severity of the multicollinearity in this study, which is an indication of the presence of multicollinearity if $(VIF) \geq 10$. From the best subset procedure, there is no variable were chosen in the blocks one and two, have one variable in block three then there is no problem in this block, there are three variables in block four, two variables in block five. Using (VIF) is to detect the presence of multicollinearity problem in the last two blocks.

Table(4.4) Multicollinearity diagnostic

Block	Response	R^2	(ANOVA) p-value	Multicollinearity VIF
1	-	-	-	-
2	-	-	-	-
3	X_{33}	-	-	-
4	X_3 X_{24} X_{26}	0.983 0.988 0.993	0.000 0.000 0.000	58.82 83.33 142.85
5	X_4 X_{17}	0.980	0.000	50.00

Table(4.4) shows the result of applying linear regression for the independent variables, in blocks four and five with p-value (F-test) of each model. Coefficient of determination (R^2), variance infliction factor (VIF). To determine severity of multicollinearity, in the block three ($VIF > 10$) high multicollinearity is presented in each of the three models, which means that presence of the problem in this block, again in the block five ($VIF > 10$) which is very high among the variables, which means that multicollinearity is presented in this block too between the two variables where is very high.

From the last summary it is clear that the variables in some blocks are high correlated, applying the logistic regression model, at time of expectation that, the coefficients have wrong sign and/or many of the predictor variables are not statistically significant, because the multicollinearity is present in the model.

4.4The Effects of Multicollinearity

To see the effect of multicollinearity, fit logistic regression model to the best variables. In the previous section it is seen that, the variables chosen are correlated in the blocks four and five. The parameters of the model will have large variance and MSE, because the multicollinearity is presented in the model.

Table (4.5) Summary of fitted logistic regression model

Variable	Coefficient	Std. error	MSE	AIC
X_{33}	0.214	0.651		
X_3	-3.612	3.744		
X_{24}	-2.497	3.173	53.711	19.424
X_{26}	4.062	4.704		
X_4	0.136	0.825		
X_{17}	2.564	2.525		

Table(4.5) shows the summary of fitting the logistic regression model, note that the mean square error MSE is large, also the standard errors of the estimated parameters have large values. The AIC is unrealistic, because it has a relatively small value with respecting to the number of variables, due to a high multicollinearity problem in the model.

A Big change will be seen in the results of the fitting (coefficients, standard error, MSE and AIC) after applying the block ridge logistic regression approach, as a solution of the multicollinearity.

4.5 Determining the Ridge Parameters

The beginning would be by applying the ordinary ridge logistic regression to the best variables. And the R software is used to compute MSE, and AIC against set values usually between (0,1). First looking for the optimal value of the ridge parameter, in order to achieve more precise estimates. Some methods are proposed, such as *ridge trace*, in which the change in model parameters against the set of ridge parameter, are monitoring as follows.

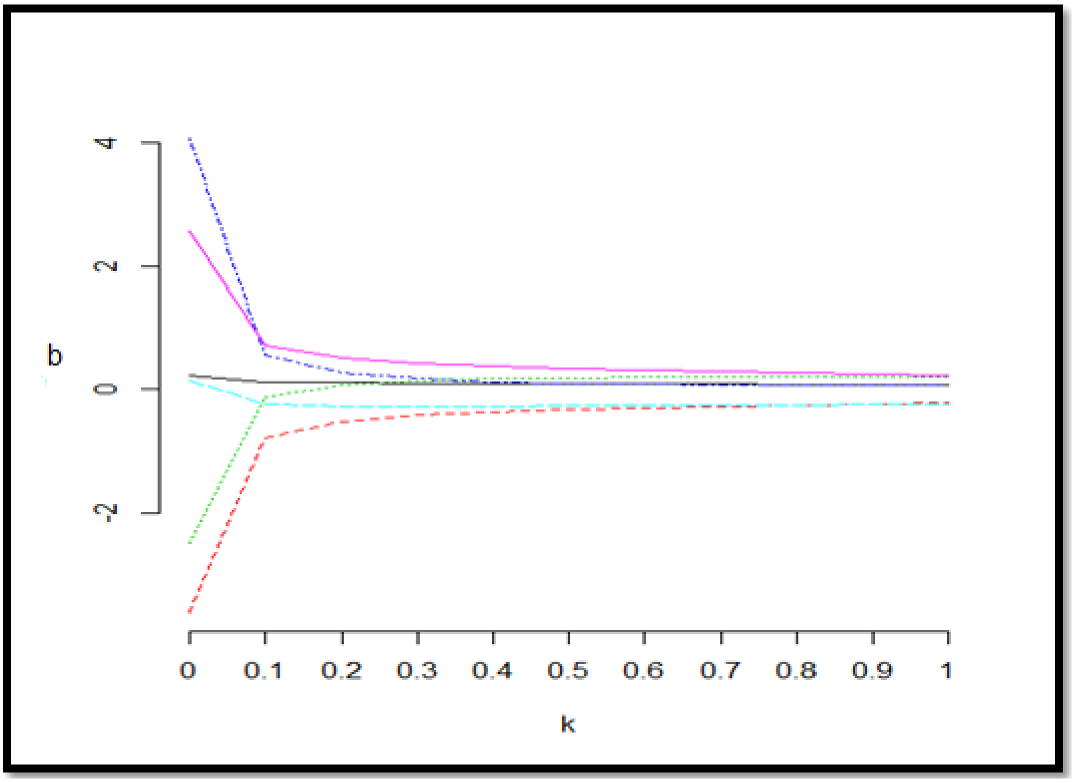
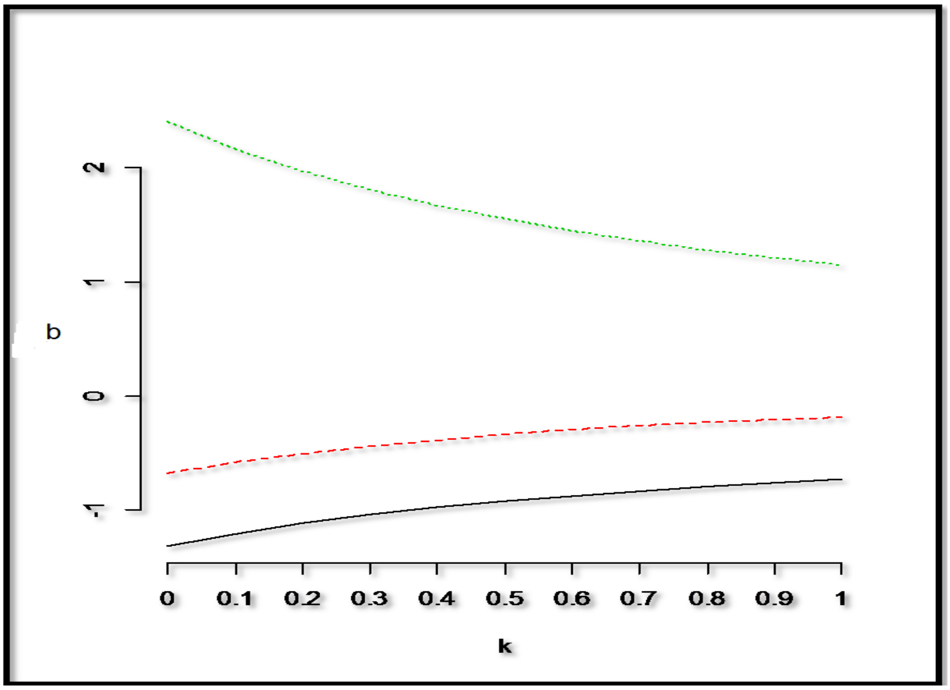


Figure (4.1) Ridge parameter of full logistic model(program with R software)

From the above graph, it is clear that all coefficients change rabidly after the ridge parameter change from zero, the coefficients at zero point is ordinary logistic regression coefficients. Note that the stationary of the coefficients start at 0.2, choose the ridge parameter for the model at this point, fitting the ridge logistic regression at this point. Note that the fitted gives MSE $= 2.288$ and $AIC = 29.719$, summary of ordinary ridge logistic regression, which note that MSE value of ordinary ridge is less than MSE of logistic regression in Table(4.5), and AIC of ordinary ridge is greater than AIC in Table(4.5).

Note that the ridge logistic regression approach is proposed to solve the problem of multicollinearity, but it gives the same weight (ridge parameter) for all variables in the model. From table(4.4) the problem of multicollinearity is not present in block3, because there is one variable in this block, where the ordinary ridge solution gave X_{33} weight (0.2), this is overweight or much more bias, the severity of multicollinearity in block4 different from block5, where the ordinary ridge gave the same weight. In this study using block ridge solution, which gives weight with respect to severity of problem.

To deal with this problem Gebril, R[5] proposed "Block Ridge Logistic Solution", in which each block can take weight according to the severity of problem in the block. In this study Table(4.4) gives three blocks those have the best variables, which gives the optimum results for the logistic model(minimum deviance). Now to find the optimum ridge regression parameters of each block, fit ridge logistic regression model to overcome the problem in each block. After applying the best subset method in each block, it was found that there is no variables selected from the block1 and block2, one variable is selected from block3 with weight zero(ridge parameter). Using ridge trace to find ridge parameter for block4 and block5, which multicollinearity is present in them. The ridge trace for block4 is presented as follows.



Figure(4.2) Ridge trace of block4(program with R software)

From Figure(4.2),and since the coefficients almost stationary at value 0.65, choose ridge parameter to be 0.65 for block4, the ridge value is relatively large, because there is high problem in this block. Similarly as in block4 select the ridge value of block5, and accordingly the ridge parameter found to be 0.05.

After determining the ridge parameters for all blocks, determine matrix K, where K is a diagonal matrix contains the values of ridge parameters of all blocks, the feature of this matrix giving a proper weight for each variable to solve the problem of multicollinearity.

Now applying the block ridge logistic regression, as a solution of the problem of multicollinearity in blocks, where the estimator of block ridge logistic regression as following;

$$b^{BR} = (X'VX + K)^{-1}X'VXb$$

where

$$K = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.65 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.65 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.65 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.05 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.05 \end{bmatrix}$$

In the above matrix K the first element in the diagonal matrix is 0 because the first variable X_{33} from block3 where is no problem in this block, the second, third and fourth are 0.65 for the variables of block4, which is proper value (ridge parameter) for these variables, the fifth and sixth are 0.05 for block5, which is proper weight of variables in this block.

Now fitting the block ridge logistic regression, which was expected to give a result better than that when applying ordinary logistic regression or ordinary ridge logistic regression.

Table(4.6) Fitting block ridge logistic regression model

Variable	Coefficients	Std. error	MSE	AIC
X_{33}	0.107	0.600	1.880	33.340
X_3	-0.339	0.491		
X_{24}	0.139	0.406		
X_{26}	0.093	0.393		
X_4	-0.353	0.583		
X_{17}	0.476	0.786		

Table(4.6) gives the summary of fitting block ridge logistic regression model. Compare this with the summary of Table (4.5), note that the values of the standard error of this model are smaller than those of logistic regression, indicating that those estimators are obtained with much more accuracy, the MSE of block ridge logistic model is less than both MSE of ordinary logistic model and MSE of ordinary ridge logistic model.

4.6 Summary & Conclusion

From the statistical analyses carried on imported commodities data can conclude that:

- The block ridge logistic regression model gives a less MSE compared to logistic regression or ordinary ridge logistic regression.
- The coefficients of block ridge logistic regression model are much more precise.
- Cannot depend on AIC as a criterion for variables selection, because it showed unrealistic quantity with high severity of multicollinearity.

Appendix A1

Results of check of multicollinearity problem for variables.

- H_0: the regression model is insignificant.
- H_1: the regression model is significant.
- p-value>0.05 accept H_0,   p-value $<$ 0.05 reject H_0.

lm(formula = x3 ~ x24 + x26 - 1)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
x24	-0.0270	0.2117	-0.128	0.899
x26	1.1413	0.2231	5.116	9.87e-06 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.122 on 37 degrees of freedom

Multiple R-squared: 0.983,   Adjusted R-squared: 0.9821

F-statistic: 1068 on 2 and 37 DF, p-value: < 2.2e-16

lm(formula = x24 ~ x3 + x26 - 1)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
x3	-0.01627	0.12758	-0.128	0.899	
x26	1.06562	0.14323	7.440	7.41e-09	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.647 on 37 degrees of freedom

Multiple R-squared: 0.9884,   Adjusted R-squared: 0.9877

F-statistic: 1570 on 2 and 37 DF, p-value: < 2.2e-16

lm(formula = x26 ~ x3 + x24 - 1)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
x3	0.36297	0.07095	5.116	9.87e-06	***
x24	0.56246	0.07560	7.440	7.41e-09	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.197 on 37 degrees of freedom

Multiple R-squared: 0.9932,   Adjusted R-squared: 0.9928

F-statistic: 2692 on 2 and 37 DF, p-value: < 2.2e-16 lm(formula = x4 ~ x17 - 1)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
x17	0.97631	0.02209	44.19	<2e-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.065 on 38 degrees of freedom

Multiple R-squared: 0.9809,   Adjusted R-squared: 0.9804

F-statistic: 1953 on 1 and 38 DF, p-value: < 2.2e-16

Appendix A2

Results of the best subset procedure of block one, with one predictor in the model, where we have $C_1^9 = 9$ models.

Logistic Model	**p-value**
$\boldsymbol{y \sim X_{16}}$	0.236
$\boldsymbol{y \sim X_{25}}$	0.097
$\boldsymbol{y \sim X_{28}}$	0.074
$\boldsymbol{y \sim X_{30}}$	1

$y \sim X_{31}$	0.577
$y \sim X_{34}$	0.110
$y \sim X_{35}$	0.296
$y \sim X_{40}$	0.613
$y \sim X_{41}$	0.123

Results of the best subset procedure of block one, with two predictors in the model, where we have $C_2^9 = 36$ models.

Logistic Model	p-value
$y \sim X_{16} + X_{25}$	0.247
$y \sim X_{16} + X_{28}$	0.202
$y \sim X_{16} + X_{30}$	0.324
$y \sim X_{16} + X_{31}$	0.237
$y \sim X_{16} + X_{34}$	0.265
$y \sim X_{16} + X_{35}$	0.136
Logistic Model	p-value
$y \sim X_{16} + X_{40}$	0.491
$y \sim X_{16} + X_{41}$	0.305
$y \sim X_{25} + X_{28}$	0.187
$y \sim X_{25} + X_{30}$	0.094
$y \sim X_{25} + X_{31}$	0.057
$y \sim X_{25} + X_{34}$	0.241
$y \sim X_{25} + X_{35}$	0.041
$y \sim X_{25} + X_{40}$	0.166
$y \sim X_{25} + X_{41}$	0.225

$y \sim X_{28} + X_{30}$	0.040
$y \sim X_{28} + X_{31}$	0.026
$y \sim X_{28} + X_{34}$	0.189
$y \sim X_{28} + X_{35}$	0.016
$y \sim X_{28} + X_{40}$	0.162
$y \sim X_{28} + X_{41}$	0.188
$y \sim X_{30} + X_{31}$	0.724
$y \sim X_{30} + X_{34}$	0.099
$y \sim X_{30} + X_{35}$	0.396
$y \sim X_{30} + X_{40}$	0.854
$y \sim X_{30} + X_{41}$	0.130
$y \sim X_{31} + X_{34}$	0.070
$y \sim X_{31} + X_{35}$	0.576
$y \sim X_{31} + X_{40}$	0.658
$y \sim X_{31} + X_{41}$	0.075
$y \sim X_{34} + X_{35}$	0.041
Logistic Model	p-value
$y \sim X_{34} + X_{40}$	0.235
$y \sim X_{34} + X_{41}$	0.250
$y \sim X_{35} + X_{40}$	0.366
$y \sim X_{35} + X_{41}$	0.032
$y \sim X_{40} + X_{41}$	0.279

Results of the best subset procedure of block two, with one predictor in the model, where we have $C_1^{11} = 11$ models.

Logistic Model	p-value
$y \sim X_8$	0.285
$y \sim X_{14}$	0.710
$y \sim X_{20}$	0.841
$y \sim X_{23}$	0.452
$y \sim X_{37}$	0.899
$y \sim X_{38}$	0.145
$y \sim X_{39}$	0.222
$y \sim X_{47}$	0.816
$y \sim X_{48}$	0.714
$y \sim X_{49}$	0.855
$y \sim X_{53}$	0.167

Results of the best subset procedure of block two, with two predictors in the model, where we have $C_2^{11} = 55$ models.

Logistic Model	p-value
$y \sim X_8 + X_{14}$	0.265
$y \sim X_8 + X_{20}$	0.523
$y \sim X_8 + X_{23}$	0.522
$y \sim X_8 + X_{37}$	0.539
$y \sim X_8 + X_{38}$	0.539
$y \sim X_8 + X_{39}$	0.394
$y \sim X_8 + X_{47}$	0.393
$y \sim X_8 + X_{48}$	0.565
$y \sim X_8 + X_{49}$	0.388

$y \sim X_8 + X_{53}$	0.059
$y \sim X_{14} + X_{20}$	0.931
$y \sim X_{14} + X_{23}$	0.560
$y \sim X_{14} + X_{37}$	0.931
$y \sim X_{14} + X_{38}$	0.319
$y \sim X_{14} + X_{39}$	0.296
$y \sim X_{14} + X_{47}$	0.933
$y \sim X_{14} + X_{48}$	0.806
$y \sim X_{14} + X_{49}$	0.933
$y \sim X_{14} + X_{53}$	0.359
$y \sim X_{20} + X_{23}$	0.626
$y \sim X_{20} + X_{37}$	0.979
$y \sim X_{20} + X_{38}$	0.272
$y \sim X_{20} + X_{39}$	0.433
$y \sim X_{20} + X_{47}$	0.968
$y \sim X_{20} + X_{48}$	0.864
Logistic Model	p-value
$y \sim X_{20} + X_{49}$	0.975
$y \sim X_{20} + X_{53}$	0.348
$y \sim X_{23} + X_{37}$	0.708
$y \sim X_{23} + X_{38}$	0.025
$y \sim X_{23} + X_{39}$	0.461
$y \sim X_{23} + X_{47}$	0.529
$y \sim X_{23} + X_{48}$	0.749
$y \sim X_{23} + X_{49}$	0.576
$y \sim X_{23} + X_{53}$	0.080

$y \sim X_{37} + X_{38}$	0.276
$y \sim X_{37} + X_{39}$	0.357
$y \sim X_{37} + X_{47}$	0.973
$y \sim X_{37} + X_{48}$	0.848
$y \sim X_{37} + X_{49}$	0.983
$y \sim X_{37} + X_{53}$	0.322
$y \sim X_{38} + X_{39}$	0.012
$y \sim X_{38} + X_{47}$	0.193
$y \sim X_{38} + X_{48}$	0.081
$y \sim X_{38} + X_{49}$	0.292
$y \sim X_{38} + X_{53}$	0.325
$y \sim X_{39} + X_{47}$	0.282
$y \sim X_{39} + X_{48}$	0.442
$y \sim X_{39} + X_{49}$	0.292
$y \sim X_{39} + X_{53}$	0.037
$y \sim X_{47} + X_{48}$	0.756
Logistic Model	p-value
$y \sim X_{47} + X_{49}$	0.970
$y \sim X_{47} + X_{53}$	0.231
$y \sim X_{48} + X_{49}$	0.803
$y \sim X_{48} + X_{53}$	0.109
$y \sim X_{49} + X_{53}$	0.170

Results of the best subset procedure of block three, with one predictor in the model, where we have $C_1^{13} = 13$ models.

Logistic Model	p-value
$y \sim X_2$	0.607
$y \sim X_7$	0.504
$y \sim X_{13}$	0.510
$y \sim X_{22}$	0.840
$y \sim X_{27}$	0.789
$y \sim X_{32}$	0.658
$y \sim X_{33}$	8.7660e-05
$y \sim X_{36}$	0.485
$y \sim X_{42}$	0.887
$y \sim X_{43}$	0.685
$y \sim X_{44}$	0.685
$y \sim X_{45}$	0.051
$y \sim X_{46}$	0.942

Results of the best subset procedure of block four, with one predictor in the model, where we have $C_1^9 = 9$ models.

Logistic Model	p-value
$y \sim X_3$	0.000
$y \sim X_9$	0.737
$y \sim X_{18}$	0.608
$y \sim X_{21}$	0.603
$y \sim X_{24}$	9.2724e-05
$y \sim X_{26}$	5.1542e-05
$y \sim X_{50}$	0.745

$y \sim X_{51}$	0.206
$y \sim X_{52}$	0.728

Results of the best subset procedure of block four, with two predictors in the model, where we have $C_2^9 = 36$ models.

Logistic Model	**p-value**
$y \sim X_3 + X_9$	0.000
$y \sim X_3 + X_{18}$	4.0825e-05
$y \sim X_3 + X_{21}$	0.000
$y \sim X_3 + X_{24}$	2.6389e-05
$y \sim X_3 + X_{26}$	1.0656e-08
$y \sim X_3 + X_{50}$	0.000
$y \sim X_3 + X_{51}$	2.5538e-07
$y \sim X_3 + X_{52}$	0.000
Logistic Model	p-value
$y \sim X_9 + X_{18}$	0.876
$y \sim X_9 + X_{21}$	0.871
$y \sim X_9 + X_{24}$	1.6944e-05
$y \sim X_9 + X_{26}$	9.6584e-06
$y \sim X_9 + X_{50}$	0.932
$y \sim X_9 + X_{51}$	0.418
$y \sim X_9 + X_{52}$	0.919
$y \sim X_{18} + X_{21}$	0.830
$y \sim X_{18} + X_{24}$	1.7949e-06
$y \sim X_{18} + X_{26}$	1.0140e-06

$y \sim X_{18} + X_{50}$	0.867
$y \sim X_{18} + X_{51}$	0.444
$y \sim X_{18} + X_{52}$	0.865
$y \sim X_{21} + X_{24}$	9.1165e-06
$y \sim X_{21} + X_{26}$	4.0496e-06
$y \sim X_{21} + X_{50}$	0.870
$y \sim X_{21} + X_{51}$	0.448
$y \sim X_{21} + X_{52}$	0.859
$y \sim X_{24} + X_{26}$	0.000
$y \sim X_{24} + X_{50}$	1.6520e-05
$y \sim X_{24} + X_{51}$	6.2422e-09
$y \sim X_{24} + X_{52}$	3.0761e-05
$y \sim X_{26} + X_{50}$	1.1128e-05
$y \sim X_{26} + X_{51}$	1.2208e-11
$y \sim X_{26} + X_{52}$	1.3150e-05
Logistic Model	p-value
$y \sim X_{50} + X_{51}$	0.448
$y \sim X_{50} + X_{52}$	0.924
$y \sim X_{51} + X_{52}$	0.382

Results of the best subset procedure of block five, with one predictor in the model, where we have $C_1^{11} = 11$ models.

Logistic Model	**p-value**

$y \sim X_1$	0.348
$y \sim X_4$	0.000
$y \sim X_5$	0.547
$y \sim X_6$	0.890
$y \sim X_{10}$	0.920
$y \sim X_{11}$	0.490
$y \sim X_{12}$	0.822
$y \sim X_{15}$	0.809
$y \sim X_{17}$	1.8662e-05
$y \sim X_{19}$	0.667
$y \sim X_{29}$	0.496

Results of the best subset procedure of block five, with two predictors in the model, where we have $C_2^{11} = 55$ models.

Logistic Model	p-value
$y \sim X_1 + X_4$	4.1634e-05
$y \sim X_1 + X_5$	0.644
$y \sim X_1 + X_{10}$	0.436
$y \sim X_1 + X_{11}$	0.211
$y \sim X_1 + X_{12}$	0.392
$y \sim X_1 + X_{15}$	0.606
$y \sim X_1 + X_{17}$	2.3802e-07
$y \sim X_1 + X_{19}$	0.643

$y \sim X_1 + X_{29}$	0.209
$y \sim X_4 + X_5$	2.9032e-05
$y \sim X_4 + X_6$	2.6471e-05
$y \sim X_4 + X_{10}$	2.6040e-05
$y \sim X_4 + X_{11}$	1.9086e-05
$y \sim X_4 + X_{12}$	7.1575e-05
$y \sim X_4 + X_{15}$	5.2793e-05
$y \sim X_4 + X_{17}$	8.3886e-08
$y \sim X_4 + X_{19}$	3.6041e-05
$y \sim X_4 + X_{29}$	3.3007e-07
$y \sim X_5 + X_6$	0.525
$y \sim X_5 + X_{10}$	0.707
$y \sim X_5 + X_{11}$	0.433
$y \sim X_5 + X_{12}$	0.740
$y \sim X_5 + X_{15}$	0.834
$y \sim X_5 + X_{17}$	1.3606e-09
$y \sim X_5 + X_{19}$	0.818
Logistic Model	p-value
$y \sim X_5 + X_{29}$	0.487
$y \sim X_6 + X_{10}$	0.990
$y \sim X_6 + X_{11}$	0.732
$y \sim X_6 + X_{12}$	0.975
$y \sim X_6 + X_{15}$	0.934
$y \sim X_6 + X_{17}$	4.1743e-09
$y \sim X_6 + X_{19}$	0.832
$y \sim X_6 + X_{29}$	0.786

$y \sim X_{10} + X_{11}$	0.778
$y \sim X_{10} + X_{12}$	0.975
$y \sim X_{10} + X_{15}$	0.949
$y \sim X_{10} + X_{17}$	6.8407e-07
$y \sim X_{10} + X_{19}$	0.874
$y \sim X_{10} + X_{29}$	0.683
$y \sim X_{11} + X_{12}$	0.779
$y \sim X_{11} + X_{15}$	0.653
$y \sim X_{11} + X_{17}$	3.9221e-07
$y \sim X_{11} + X_{19}$	0.403
$y \sim X_{11} + X_{29}$	0.714
$y \sim X_{12} + X_{15}$	0.827
$y \sim X_{12} + X_{17}$	2.5789e-06
$y \sim X_{12} + X_{19}$	0.772
$y \sim X_{12} + X_{29}$	0.776
$y \sim X_{15} + X_{17}$	1.6773e-06
$y \sim X_{15} + X_{19}$	0.910
Logistic Model	p-value
$y \sim X_{15} + X_{29}$	0.665
$y \sim X_{17} + X_{19}$	6.0081e-07
$y \sim X_{17} + X_{29}$	5.1459e-12
$y \sim X_{19} + X_{29}$	0.550

References

1. ALIAN AGRESTI(1996). *An Introduction to Categorical Data analysis*. John Wiley & Sons.
2. ANNA. O'CONNELL. (2006). *Logistic Regression Model for Ordinal Response Variables*. Sage Publications.
3. Barbara G. et . al . (2007). *Using Multivariate Statistics*. Pearson Education.
4. David W. Hosmer & Stanley Lemeshow. (2000). *Applied Logistic Regression*. John Wiley & Sons.
5. Gebril, R.(2005). *Combination of statistical techniques for data modeling in large database*.
6. Hoerl, A. E . et . al. (1975). *ridge regression; biased regression estimators*. Marcel Dekker.
7. Hoerl, A. E. & Kennard, R. W. (1970). *Ridge regression. Biased estimation of non-orthogonal problems*.
8. Michael H, Christopher J & William Li. (2005). *Applied Linear Statistical Models*. Mc Graw Hill.
9. Simon J. Sheather. et. al. (2009). *A Modern Approach to Regression with R*. Springer Science & Business Media.
10. S. Le. Cessle & J. C. Van Houwelingen. (1990). Ridge Estimators in Logistic Regression. Rolay statistical Society.
11. Vipin Kumar. (2011). *Data Mining with R* (Learning with Case Studies).Taylor &Francis Group .
12. William R, Dillon & Matthew Goldstein.(1984). *Multivariate Analysis* (*Methods and Application*). John Wiley & Sons.