Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 03, April-May 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.33.1.11

Comparison of Some Estimator Methods of Regression Mixed Model for the Multilinearity Problem and High – Dimensional Data

Thaer Hashim Abdul Muttaleb*

**College of Medicine/ Wasit University/ Family and Community Medicine Branch, Iraq*

*Corresponding Email: *t.hashim@uowasit.edu.iq*

Received: 02 December 2022 **Accepted:** 28 February 2023 **Published:** 03 April 2023

Abstract: In order to obtain a mixed model with high significance and accurate alertness, it is necessary to search for the method that performs the task of selecting the most important variables to be included in the model, especially when the data under study suffers from the problem of multicollinearity as well as the problem of high dimensions. The research aims to compare some methods of choosing the explanatory variables and the estimation of the parameters of the regression model, which are Bayesian Ridge Regression (unbiased) and the adaptive Lasso regression model, using simulation. MSE was used to compare the methods.

Keywords: Mixed Model, Lasso Regression, Bayesian Ridge.

1. INTRODUCTION

Problem of the research

Many applied studies, the number of explanatory variables is large, greater than the size of the sample affecting the dependent variable and overlapping with each other, so the problem of linear multiplicity arises between variables as well as a problem of high dimensions, and many methods were used to avoid the problem of linear multiplicity between explanatory variables as well as choosing and estimating the best significant variables, Therefore, it is necessary to search for methods whose task is to solve the problem of multicollinearity and to choose the number of independent variables that have an intestinal effect and to obtain a model with significant and high efficiency.

Aim of the research

The aim of this research is to compare the methods of selecting explanatory variables and estimating the parameters of the mixed regression model, which are Bayesian regression and the adaptive lasso regression model, in the presence of the problems of multicollinearity and

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 03, April-May 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.33.1.11

the curse of dimensionality, using some known statistical criteria to reach the best mixed model.

The mixed model

Linear Mixed Models (which are also called Linear Mixed Effect Models) proposed by Hartly & Rao (1967) have become one of the means of analyzing Repeated Measures Data that appears in many fields such as: the agricultural field, Biological domain, medical domain, economic domain, geographic domain, the increasing generality of these models is explained by the flexibility they offer in representing correlations within items usually provided in refined measurements data by treating both balanced and unbalanced data, as well as by facilitating efficient and reliable programs them to fit.

Especially in research related to medical experiments, an attempt is made to determine if the rate of improvement as a result of treatment (A) is faster than treatment (B), and for those measurements of the item are taken repeatedly during time and the process of change can be represented within the items, and these repeated measurements on the item are Usually Correlated.

This type of research is included in the so-called (longitudinal studies), which represents an observational study, as the longitudinal study is a study of research on related items such as medical and social research that includes repeated observations on the same item (Units). Over a long period of time, such as months or years.

The importance of these longitudinal data focuses on the general effects of within and between individual covariates on the response, as well as on the behavior of the specific individual, and the linear mixed model (LMM) is usually used to analyze this type of data for which correlation is the defining characteristic.

The Linear Mixed Model (LMM) assumes that random effects and errors within the items have a normal distribution (N.D.), as the normality (symmetry) of the random effects and errors within the items represent the usual assumptions of the Linear Mixed Model.

Mixed models are powerful tools for analyzing cluster data and many extensions of the classic linear mixed model in which the response variable follows a normal distribution as with all parametric models.

In general, regression models are used to describe the relationship between a response variable (Y) and a group of explanatory variables (Xs). For example, the linear relationship between a response variable and a single explanatory variable can be written in the following form:

$$Yi = \beta \circ + \beta_1 X_i + \varepsilon i$$

As they $\beta\circ, \beta_i$ are unknown regression landmarks. The classical regression model assumes that these parameters in the model are constant effects. In other words, the parameters are constants and the data is used to obtain estimates of these constants. However, there are many

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 03, April-May 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.33.1.11

cases where the features in the regression model can be random, in which case the features are known as random effects or random features.
Mixed models are models that contain at least one fixed effect and at least two random effects with an error limit. When mixed models appear in practice, data are often grouped together by a common feature. These groups are known as clusters. Cluster data includes repeated measurements on the items (as the item represents the cluster) as in the design of split plots in the agricultural experiments, the whole sector represents the cluster.

Lasso Regression
This method was presented for the first time in the geophysical literature in the year (1982), then it was rediscovered independently in the year (1996) by the researcher Robert Tibshirani, where he formulated this method and provided many ideas about its performance.
Least Absolute Shrinkage and Selection Operator is an abbreviation for the English words (Least Absolute Shrinkage and Selection Operator), and it is a penalty function for the linear regression model. It is a method for estimating the parameters of the regression model, as well as for selecting and organizing the variables included in the model to increase the explanatory accuracy of the regression models used in analyzing the phenomenon under study. Through the processes of adapting the model to select a subset of the covariates in the final model instead of using them all, in the Lasso method the sum of squares of random errors is minimized to the maximum of the sum of the absolute values of the coefficients of the regression model LASSO was originally designed for Least squares models, where LASSO reveals a large amount of behavior of the estimator via soft thresholding, including the relationship of the LASSO estimator with a ridge regression estimator and a best subset of variables selection estimator. subset selection, which is analogous to the stepwise selection method, and also reveals (as in linear regression) that the Lasso coefficient estimates do not have to be unique if the independent variables suffer from the problem of multicollinearity. And the Lasso method has the ability to choose a partial group that depends on the constraint formula, and although the Lasso method has been defined for least squares, the Lasso method can easily be used in a wide range of many statistical models, including generalized linear models, generalized estimation coefficients, relative risk models, and M estimators. , and Lasso can be used in many areas such as geometry, Bayesian statistics and convex analysis.

Before the lasso regression method, the most used method for selecting the independent variables that are included in the model was the Stepwise Selection method, which improves the accuracy of the model in certain cases, especially when some of the independent variables have a strong relationship with the response variable, which makes the prediction inaccurate, as well as the method Ridge regression is the most popular one that is used to improve the prediction accuracy of the regression model. It improves prediction error by reducing large regression coefficients in order to reduce recurrence, but it does not perform co-selection and therefore does not help make the model more interpretable. While Lasso can achieve both of these goals by making the set of absolute values of the regression coefficients have less than a fixed value, which forces some coefficients to be equal to zero, while choosing a simpler model that does not include these coefficients.

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 03, April-May 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.33.1.11

LASSO Regression Principle

The principle of the Lasso regression method is to reduce the sum of the squares of the residuals according to a constraint that represents the absolute sum of the coefficients that are smaller than a certain constant. In order to do this, Lasso applies the shrinking (regulating) process, as it successively regresses the regression coefficients and shrinks some of them to zero. During the selection process, the variables that contain a non-zero symbol will be determined after the shrinking process and will be part of the model. The goal of this process is to reduce the prediction error.

In the Lasso method, there is a parameter (adjustment) that controls the force of punishment (penalty) of the regression coefficients, and it occupies great importance in that. When the adjustment parameter is large enough, the coefficients are forced to be equal to zero, and this is restricted in reducing the variables in the model, that is In other words, the larger the value of the adjustment parameter, the greater the number of coefficients equal to zero. And if the adjustment parameter is equal to zero, we will get the OLS Regression.

LASSO Regression Advantages

There are many advantages in using the lasso method as follows:

1. Lasso can provide very good predictive accuracy because shrinking and removing variables can lower variance without greatly increasing bias, and this is especially useful when we have a small number of observations and a large number of variables.
2. Lasso helps to increase the possibility of interpreting the model by eliminating irrelevant variables that are not related to the response variable.
   Thus, the Lasso method is a method for selecting and organizing the variables involved in the regression model.

LASSO Regression Formula

The parameters of the Lasso regression were estimated according to the principle of least squares from the basic formula as follows:

Let us have a sample consisting of N cases, each case composed of P independent variables and one intentional variable yi, and let x represent the vector of the independent variables for the case j^th, so the goal of lasso regression is to solve the following equation:

$\min\{\frac{1}{N}\sum_{i=1}^{N}(yi - \beta_0 - x_i{}^T\beta)^2\}$ Subject to $\sum_{j=1}^{p}|\beta_j| \leq t$

t: represents a predefined free parameter that specifies the amount of flattening.

X: matrix of independent variables.

and $X_{ij} = (x_i)_j$

$x_i{}^T$ is the j^{th} row of matrix X

The lasso formula can be written in the following form:

$\min_{\beta_0,\beta}\{\frac{1}{N}||y - \beta_0 I_N - X\beta||_2^2\}$ Subject to $\|\beta\| \leq t$

$||\beta||_p = (\sum_{i=1}^{N}|\beta_i|^p)^{1/p}$ is the standard length ℓ^p and that I_N is a vector of units (Nx1).

$\bar{x}$ stands for the standard mean of data points x_i and $\bar{y}$ stands for the mean of the dependent variable (response variable y_i) and the estimate $\beta_0 = \bar{y} - \bar{x}_i{}^T\beta$ If that:

$yi - \beta_0 - x_i{}^T\beta = yi - (\bar{y} - x_i{}^T\beta) - x_i{}^T\beta$

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 03, April-May 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.33.1.11

$$= (yi - \bar{y}) - (x_i - \bar{x})^T\beta$$

Therefore, it is natural to work with variables that have been centralized (making their average equal to zero), in addition to the independent variables that are typically standardized $\sum_{i=1}^{N} {x_i}^2 = 1$
The formula for minimizing the sum of squares of errors can be rewritten as follows:

$$\min_{\beta_0,\beta}\{\frac{1}{N}||y - X\beta||_2^2\} \quad \text{Subject to } ||\beta||_1 \leq t$$

The Lagrange multiplier is in the following form:

$$\min_{\beta \in R^P}\{\frac{1}{N}||y - X\beta||_2^2 + \lambda||\beta||_1\}$$

λ It is the parameter that controls the strength of the penalty (punishment) on the regression parameters.

The Lasso Estimator properties
There are some characteristics of the lasso estimator that we can list as follows:

1- Orthonormal covariates
Suppose that the covariates are naturally orthogonal so that $(x_i|x_j) = \delta_{ij}$, where:
$(.|.)$ Inner product and δ_{ij} Kroncher delta expresses :

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

and by using the iterative Subgradient method, which is one of the numerical analysis methods for solving least convexity problems, we get:

$$\hat{\beta}_j = S_{N\lambda}\left(\hat{\beta}_j^{OLS}\right) = \hat{\beta}_j^{OLS} Max\left(0,1 - \frac{N\lambda}{\left|\hat{\beta}_j^{OLS}\right|}\right)$$

$$\hat{\beta}_j^{OLS} = (X^TX)^{-1}X^TY$$

and that $S_{N\lambda}$ refers to the Smooth Threshold.
Since it converts the values to zero (make them exactly zero if they are small enough) instead of setting smaller values to zero and leaving the larger values untouched as a hard threshold factor denoted by $H_{N\lambda}$, this can be compared with the slope of the letter, so the goal is to minimize the amount as follows:

$$\min_{\beta \in R^P}\{\frac{1}{N}||y - X\beta||_2^2 + \lambda||\beta||_2^2\}$$

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 03, April-May 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.33.1.11

$$\hat{\beta}_j = (1 + N\lambda)^{-1}\hat{\beta}_j^{OLS}$$

As long as the slope of the character is reduced, all coefficients are reduced by the variable factor of $(1 + N\lambda)^{-1}$ and none of the coefficients is set to zero.

$$\min_{\beta \in R^P}\{\frac{1}{N}||y - X\beta||_2^2 + \lambda||\beta||_0\}$$

where $\| \quad \|_0$. It is the lengthℓ^0, which is defined as $\left||z|\right| = m$. If m are exactly components of z, it is non-zero, and it can be written as follows:

$$\hat{\beta}_j = H_{\sqrt{N\lambda}}\left(\hat{\beta}_j^{OLS}\right) = \left(\left|\hat{\beta}_j^{OLS}\right| > \sqrt{N\lambda}\right)$$

Where $H_{\sqrt{N\lambda}}$ is the solid threshold term and I is the indicator function.
We note from this that Lasso estimates combine the characteristics of the letter regression and the best partial regression, as it converts all coefficients to zero with a fixed value and adjusts them to zero if they are reached.

Correlated covariates
Returning to the general form of the lasso in which the different covariates may not be independent, in the case where two independent variables i,j are identical for each case so that: $x_i = x_j$ the parameter values B_i and B_j Which minimizes the lasso's objective function is not uniquely specific, but there is $\widehat{B}_i, \widehat{B}_j \geq 0$ It is if $s \in [0,1]$ by replacing $\widehat{B}_i$ of $s(\widehat{B}_i + \widehat{B}_j)$ and $\widehat{B}_j$ of $(1 - s)(\widehat{B}_i + \widehat{B}_j)$ With the rest retained $\widehat{B}_i$ And we get a new solution and the lasso function continues to reduce the coefficients.

Bayesian Ridge Regression (unbiased)
The biased letter regression method has been used, which has been explained in many studies, as this method has proven effective in addressing the problem of collinearity, which leads to a large size of the estimators' variance and causes confusion in the causal relationship between the independent variables and the dependent variable in the regression model. However, in the unbiased letter regression method, prior information will be entered into the letter regression model to get rid of the bias problem in estimating parameters.

When applying the OLS least squares method to non-orthogonal data, we will get unstable estimators that have a large sum of squares error (MSE). To deal with the problem of collinearity, the usual (biased) character regression method is used, which was developed by (Horel and Kennard (1970)). This method is summarized by adding the constant K to the matrix (X'X) before taking its inverse, as follows:

$$\widehat{B}(KI, J) = \hat{\beta}_R = (X'X + KI_P)^{-1}(X'Y)$$

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 03, April-May 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.33.1.11

Swindel (1976) suggested entering Prior Information into the letter declension formula, so the Bayesian letter declension formula would be as follows:

$$\hat{\beta}_R = (X'X + KI_P)^{-1}(X'Y + KJ)$$

Where we notice from the above formula that the constant K has been introduced with the vector J into the vector X'Y, which is represented by the prior information.
We note from the above formula that in the case of (K = 0) we will get the OLS estimate.
And through the concept of Bayes' theorem represented in the following:

Posterior dist. $\propto$ Prior Prob.× Likelihood
We note from the above formula that to obtain the final distribution, the initial probability and the weighting function must be provided, i.e. providing preliminary information about the estimator in order to obtain the final probability.

If we have the vector J, which represents the initial information vector, the value of K gives estimates that have MSE less than the MSE of least squares. And if J is any random vector (not informational), K also gives estimators that have an MSE less than the MSE of least squares.
As is known from the assumptions of the linear model, that error u_i It has a normal distribution with a mean of 0 and variance σ^2_{In} i.e $u_i \sim N(0, \sigma^2_{In})$ The least squares estimates are normally distributed with mean B and variance $\sigma^2(X'X)^{-1}$

Let us suppose that the vector J containing the previous information has a normal distribution with a mean of B, and the covariance-variance matrix is V, i.e. J$\sim$ $N(B, V)$.
Assuming that V is a full rank matrix and represents the covariance matrix - the covariance, thus the convex estimator:
$B(C, J) = C\widehat{B}_{Ols} + (I - C)J$

$B(C, J)$ It is the convex estimation, which is a convex real-value function and one of its properties is that if there are any two points in the domain of this function such as x and y, then if the straight line that connects any two points on its graph is located above the graph of the function, then if it is:

$B(C, J) = cf(x) + (I - c)f(y)$
I It is a unary matrix of rank P×P
C is a dimensioned matrix P×P
We note from the above formula that the matrix C is an unknown matrix. To find this matrix, we will find the mean squares sum of the error of the convex estimation, which is in the following formula:
$\widehat{B}_{Bays} = (X'X + \sigma^2 V_0^{-1})^{-1} (X'Y + \sigma^2 V_0^{-1} B_0)$

Simulation
There are many definitions of the concept of simulation, all of which lead to one goal, which is to re-enact complex systems in which a large number of variables overlap in a mathematical

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 03, April-May 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.33.1.11

manner and with the help of an electronic computer. Simulation may be defined as a mathematical method that includes a set of mathematical equations and logical relationships that describe the general behavior of different experiments with the help of an electronic computer.

Simulation can also be defined as an attempt to find an exact copy of any system without taking that system. This is done by taking mathematical and statistical models and with the help of an electronic computer. The most important advantage of the simulation method is the shortening of the time needed to analyze the system. This method also helps in discovering obstacles and problems when adding new variables to the system. It also provides the ability to control the conditions and restrictions of the system and provides the possibility of verifying the validity of the analytical results of the system before its actual application. However, the simulation method is not without flaws. Given that this method works on repeating the experiment many times, obtaining the same sample size in each iteration may lead to an increase in cost and time. It may also be very difficult to obtain the same practical case for each iteration or implementation of the experiment, and simulation does not always lead to obtaining to the optimal solution. In order to avoid these defects, the method of constraint simulation was used, as this method in the simulation relies on conducting the data generation process according to the assumed model by adopting the estimated or estimated model parameters in the light of the real data of the issue under study.

Simulation experiments were carried out using four sample sizes as follows (n=20,50,100,150) with replicates equal to (1000) for each experiment. The variables have been generated as follows:

Linear explanatory variables

It is the $X_i's$ of a normal distribution, i.e:

$X \sim N_P(\underline{0}, \Sigma)$ Where $P = 50$

So, the correlation between X_I and X_j is

$\rho^{|i-j|}$ and $\rho = 0.4, 0.7, 0.9$

random errors

which is e_i has a normal distribution:

$e_i \sim N(0, \sigma^2)$, $i = 1,2,\ldots,n$

$\sigma = 1, 3, 5$

Noting that e and X are independent.

The dependent variable

The response or dependent variable Yi is generated directly through the additive partial linear model:

$Y = X^T\beta + + U$

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 03, April-May 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.33.1.11

Table 1: represents to the results of Mean Squared errors (MSE) for the aforementioned simulated cases with correlation coefficient($\rho = 0.4$) and sample sizes (n= 20, 50, 100, 150).

N	Methods	σ		
		1	3	5
20	Lasso	7.3367	8.5122	13.3531
	Bayesian Ridge Regression (unbiased)	9.0995	11.2172	14.0014
50	Lasso	6.0333	6.9999	10.9808
	Bayesian Ridge Regression (unbiased)	7.4829	9.2244	11.5139
100	Lasso	4.7924	5.5602	8.7224
	Bayesian Ridge Regression (unbiased)	5.9439	7.3272	9.1458
150	Lasso	3.7303	4.3280	6.7894
	Bayesian Ridge Regression (unbiased)	4.6266	5.7034	7.1190

Table 2: represents to the results of Mean Average Squared errors (MASE) for the aforementioned simulated cases with correlation coefficient($\rho = 0.7$) and sample sizes (n= 20, 50, 100, 150).

n	Methods	σ		
		1	2	6
20	Lasso	7.9190	9.1878	14.4129
	Bayesian Ridge Regression (unbiased)	9.8217	12.1075	15.1126
50	Lasso	6.5121	7.5555	11.8523
	Bayesian Ridge Regression (unbiased)	8.0768	9.9564	12.4277
100	Lasso	5.1727	6.0015	9.4146
	Bayesian Ridge Regression (unbiased)	6.4156	7.9087	9.8717
150	Lasso	4.0264	4.6715	7.3282
	Bayesian Ridge Regression (unbiased)	4.9938	6.1560	7.6840

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 03, April-May 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.33.1.11

Table 3: represents to the results of Mean Average Squared errors (MSE) for the aforementioned simulated cases with correlation coefficient($\rho = 0.9$) and sample sizes (n= 20, 50, 100, 150).

n	Methods	σ		
		1	**2**	**6**
20	Lasso	8.2683	9.5931	15.0487
	Bayesian Ridge Regression (unbiased)	10.2550	12.6416	15.7794
50	Lasso	6.7994	7.8888	12.3752
	Bayesian Ridge Regression (unbiased)	8.4331	10.3957	12.9760
100	Lasso	5.4009	6.2663	9.8300
	Bayesian Ridge Regression (unbiased)	6.6986	8.2576	10.3072
150	Lasso	4.2040	4.8776	7.6515
	Bayesian Ridge Regression (unbiased)	5.2141	6.4276	8.0230

2. CONCLUSION

Through the results shown in Tables (1),(2) and (3), the Bayesian Ridge Regression (unbiased) is more efficient than the Lasso method estimation method, for all sample sizes and with different correlation coefficients and standard deviation values, so it is preferable to use it to estimate the components of the linear model.

3. REFERENCES

1. ALKHAMISI M. A. , SHUKUR G., (2007) , " A Monte Carlo Study of Recent Ridge Parameters " , Communications in Statistics—Simulation and Computation, Taylor & Francis, 36: 535–547.
2. Al-Sadoun , Muhannad Faiz , (2005) . " Empirical Bayes and Bayes for Ridge regression" , Quarterly Specialized Refereed Journal, vol. (7) no. (1).
3. Buhlmann, Peter ; van de Geer, Sara,(2011), " Statistics for High-Dimensional Data , Methods, Theory and Applications, Springer , Heidelberg Dordrecht London New York.
4. Duzan , Hanan; Shariff, Nurul Sima Mohammed, (2015) , " Ridge Regression for solving the Multicollinearity Problem: review of Methods and Models" , Journal of Applied Statistics, 15(3) : 392-404.
5. F. Benee, Swindle, (2013) , " Good ridge estimators based on prior information", Moscow State Univ Bibliotic, Taylor & Francis.
6. Hans ,Christ , (2009) , " Bayesian lasso regression" , Biometrika ,96, 4, Biometrika Trustpp. 835–845 .
7. Li Fan; Yang Yiming ; P. Xing ,Eric, (2006) , " From Lasso regression to Feature vector machine ",Pittsburgh, PA USA 15213, fhustlf,yiming,epxingg@cs.cmu.edu.

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 03, April-May 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.33.1.11

8. Lindley, D.V.; Smith, A.F.M, (1972), " Bayes Estimation for the linear Model" , Journal of Royal Statistical Society , Series B (Methodological) , Volume 34, Issue 1 , 1-41.
9. Peihua Qiu,Changliang Zou and Zhaojun Wang (2010) "Nonparametric profile monitoring By mixed Effects Modeling" School of statistics, University of Minnesota Department of statistics ,Nankai university ,China
10. Ranstam, J.; Cook, J. A., (2018), " Lasso Regression " , BJS Statistical Editors .
11. Robert H. Crouse & Chun Jin.(1995)."Unbiased RIDGE Estimation With prior Information And Ridge Trace" Commun. Statist.-Theory Meth.27(9),2341-2354.
12. Saleh, A.K.Md. Ehsanes; Kibria ,B.M. Golam, (1993), " Performance of some new Preliminary tests Ridge Regression estimators and other properites" , COMMUN. STATIST.-THEORY METH., 22(10), 2747-2764 .
13. Samira, M.S., Dhafir, H.R., Nawzad,M.A.,(2011) Comparison among some estimation method in Generalized linear mixed models by simulation and practical data" Sulimania university, Iraq
14. Tibshirani , Robert, (1996) ," Regression Shrinkage and Selection via the Lasso" , J. R. Statist. Soc. B 58, No. 1, pp. 267-288.
15. Tibshirani , Robert, (1997) ," The LASSO method for variable selection in Cox Model " , J. R. Statist. Soc. B 58, No. 1, pp. 267-288.
16. Tibshirani , Ryan , (2013) , " Modern regression " , Optional reading: ISL 6.2.1, ESL 3.4.1, Data Mining: 36-462/36-662.
17. Zellner, A. (1971).An introduction to Bayesian Inference in Econometrics, Wiley, New York.