iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IX **Month of publication:** September 2023

DOI: https://doi.org/10.22214/ijraset.2023.55786

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

Multicollinearity in Multiple Linear Regression: Detection, Consequences, and Remedies

Siddamsetty Upendra[1], Dr. R. Abbaiah[2], Dr. P. Balasiddamuni[2]
[1, 2, 3]*Dept.of Statistics, S V University, Tirupathi*

Abstract: ***Multiple linear regression is a widely used statistical tool for modeling relationships between a dependent variable and multiple explanatory variables. However, it assumes that these explanatory variables are independent, which is not always the case in practical scenarios, leading to a phenomenon known as multicollinearity.***
Multicollinearity occurs when explanatory variables in a regression model are strongly correlated with each other, causing several issues in regression analysis. This paper discusses the detection and remedies for multicollinearity in detail.
Detection methods include examining the determinant of the correlation matrix, inspecting correlation coefficients, using partial regression coefficients, calculating Variance Inflation Factors (VIFs), and assessing the condition number and condition index. These techniques help researchers identify the presence and severity of multicollinearity in their dataset.
To address multicollinearity, several remedies are proposed, including obtaining more data, dropping collinear variables, using relevant prior information, employing generalized inverses, and employing principal component regression. Ridge regression, which introduces bias to reduce variance, is also discussed as an effective technique to combat multicollinearity.
Understanding multicollinearity and employing appropriate detection and remediation strategies is crucial for obtaining reliable and meaningful results from multiple linear regression models.
Keywords: Multicollinearity, Detection Methods, Remedies, Correlation Matrix, Variance Inflation Factors (VIFs), Condition Number, and Ridge Regression. etc.

I. MULTICOLLINEARITY

A basic assumption of the multiple linear regression models is that the rank of the matrix of observations on the explanatory variables is equal to the number of explanatory variables. In other words, such a matrix is of full column rank. This indicates that all the explanatory variables are independent, i.e. there is no linear relationship between the explanatory variables. The explanatory variables are called orthogonal.
In many practical situations, explanatory variables may not be independent for various reasons. A situation where the explanatory variables are strongly correlated is called multicollinearity.

Consider a multiple regression model

$$Y_{n\times 1} = X_{n\times k}\beta_{k\times 1} + \varepsilon_{n\times 1},\ \ \varepsilon \sim \mathrm{N}(0,\ \sigma^2\ \mathrm{I})$$

With k explanatory variables $\mathrm{X}_1, \mathrm{X}_2, ---, \mathrm{X}_\mathrm{k}$ with usual assumptions including $Rank(X) = k$.

Assume observations on all the $X_i's$ and the $Y_i's$ are entered and scaled to unit length. So

* 'X becomes a k X k matrix of correlation coefficients between explanatory variables and
* Y will be the k X1 vector of correlation coefficients between explanatory variables and study variables.

Let $X = [X_1, X_2, ---, X_K]$ where X_j a column of X is represents the n observations on X_j. The Zero column vectors, such as $X_1, X_2, ---, X_K$ are linearly dependent if there is a set of constants $C_1, C_2, ---, C_k$ not all zero, such that

$$\sum_{j=1}^{k} C_j X_j = 0$$

If this is true for exactly one subset $X_1, X_2, ---, X_K$, then rank (X'X) = k.

Therefore $(X'X)^{-1}$ does not exist. If the condition $\sum_{j=1}^{k} C_j X_j = 0$ is approximately true for a subset of $X_1, X_2, ---, X_K$ and then there has a quasi-linear dependence on (X'X). In such a case, there is a problem of multicollinearity. That too (X'X) will become ill-conditioned.

II. SOURCE OF MULTICOLLINEARITY

1) *Method of Data Collection:* Data are expected to be collected on a complete sample of variables. Data may be collected in a subspace of explanatory variables where the variables are linearly dependent. For example, sampling is conducted on only a limited range of explanatory variables in the population
2) *Model and Population Constraints:* There may be some restrictions on the sample or the population from which the sample is drawn. A sample can be drawn from a portion of the population that has linear combinations.
3) *Existence of Identities or Definitional Relationships:* Relationships between variables may be due to the definition of the variables or any identity relationship between them. For example, if data is collected on variables such as income, savings, and expenditures, then income = savings + expenditures. Such a relationship does not change even when the sample size is increased.
4) *Imprecise Formulation of Model:* Model formulation is unnecessarily complicated. For example, quadratic (or polynomial) terms or cross-product terms appear as explanatory variables. For example 3 variables X_1, X_2 and X_3, therefore k = 3. Suppose their cross product terms X_1X_2, X_2X_3 and X_1X_3 are equal added. Then k increases to 6.

A. *An over-determined Model*

Sometimes, in an overzealous manner, a large number of variables are included to further refine the model realistic. Therefore, the number of observations (n) becomes smaller than the number of interpretations variables (k) . Such a situation may arise in clinical research where the number of patients is small, but information is collected on a large number of variables. In another example, if there is a 50-year time series of data on a consumption pattern, the consumption pattern is expected to remain unchanged.50 years is the same. So the best option is to select a small number of variables, hence the results n<k.

III. CONSEQUENCES OF MULTICOLLINEARITY

To illustrate the effects of the presence of multicollinearity, they considered a model

$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, $E[\varepsilon] = 0$ and $Var(\varepsilon) = \sigma^2 I$

Where x_1, x_2 and y are scaled to length unity**.**

The normal equation (X ' *X) b = X ' y* in this model becomes

$$\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} r_{1y} \\ r_{2y} \end{pmatrix}$$

Where r is the correlation coefficient between X_1 and X_2 ; r_{jy} is the correlation coefficient between x_j and $y(j=1,2)$ and $b = (b_1, b_2)'$ is the OLSE of β

$$(X'X)^{-1} = \frac{1}{1-r^2}\begin{bmatrix} 1 & -r \\ -r & 1 \end{bmatrix}$$

$$\Rightarrow b_1 = \frac{r_{1y} - rr_{2y}}{1-r^2}$$

$$b_2 = \frac{r_{2y} - rr_{1y}}{1-r^2}$$

So the variance matrix is $\text{Var(b)} = \sigma^2(\text{X}'\text{X})^{-1}$

$$\Rightarrow Var(b_1) = Var(b_2) = \frac{\sigma^2}{1-r^2}$$

$$Cov(b_1, b_2) = -\frac{r\sigma^2}{1-r^2}$$

If X_1 and X_2 are uncorrelated, then r=0 and $\text{X}'\text{X} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$$Rank(\text{X}'\text{X}) = 2$$

If X_1 and X_2 are perfectly correlated, then $\text{r} = \pm 1$ and $\text{Rank}(\text{X}'\text{X}) = 1.$

If $\text{r} \to \pm 1$, then $\text{Var}(\text{b}_1) = Var(b_2) \to \infty$

So, if the variables are perfectly parallel, the variance of OLSEs becomes larger. This implies very unreliable estimates and is an unacceptable situation.

If near or high multicollinearity occurs, the following potential consequences are encountered.

1. OLSE is an unbiased estimator of β, but its sample variance becomes very large. Thus, the OLSE becomes inaccurate and the blue property no longer exists.
2. Due to large standard deviations, the regression coefficients do not appear to be significant. Therefore, essential variables can be omitted

For example, to test $H_0 : \beta_1 = 0$, we use t- test statistic as

$$t_0 = \frac{b_1}{\sqrt{\hat{V}ar(b_1)}}$$

Since $\hat{V}ar(b_1)$ is large, so t_0 is small and consequently H_0 is more often accepted.

Thus, deleterious multicollinearity tends to eliminate important variables.

3. Due to large standard deviations, a large confidence zone may appear. For example, confidence interval is provided as $b_1 \pm t_{\frac{\alpha}{2}, n-1}\sqrt{\hat{V}ar(b_1)}$. When $\hat{V}ar(b_1)$ it becomes large, then the confidence interval becomes wider.

4. OLSE can be sensitive to small changes in the values of explanatory variables. If some observations are added or removed, the OLSE can change dramatically in amplitude and sign. Ideally, OLSE should not change by adding or removing variables. Thus OLSE loses consistency and robustness.

If the number of explanatory variables is greater than two, say k as $X_1, X_2, ---, X_k$ then the j^{th} diagonal element of $C = (\text{X}'\text{X})^{-1}$ is $C_{jj} = \frac{1}{1-R_j^2}$

Where R_j^2 are the multiple correlation coefficients or the coefficient of determination of the regression of X_j among other (k-1) explanatory variables.

If X_j is closely related to a subset (k -1) explanatory variables then R_j^2 is large and close to 1. Therefore, the variation of j^{th} OLSE is $Var(b_j) = C_{jj}\sigma^2 = \frac{\sigma^2}{1-R_j^2}$ becomes very large.

Covariance between b_i and b_j If so, it would be high, X_i and X_j are involved in a linear relation leading to multicollinearity.

Least squares estimates b_j becomes much larger in absolute value in the presence of multicollinearity. For example, consider the squared distance between b and β as

$$L^2 = (b-\beta)'(b-\beta)$$

$$E(L^2) = \sum_{j=1}^{k} E(b_j - \beta_j)^2 \quad = \sum_{j=1}^{k} Var(b_j) \quad = \sigma^2 trace\left(X' X\right)^{-1}$$

The trace of a matrix is equal to the sum of its Eigen values. If $\lambda_1, \lambda_2, ---, \lambda_k$ are Eigen values of $(X' X)$, then $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, ---, \frac{1}{\lambda_k}$ are the Eigen values of $\left(X' X\right)^{-1}$ and hence

$$E(L^2) = \sum_{j=1}^{k} \frac{1}{\lambda_j},\ \lambda_j > 0$$

If (X ' X) is ill-conditioned due to multicollinearity, at least one of the eigenvalues is small. Thus, the distance between b and β may also be significant. Thus

$$E(L^2) = E(b-\beta)'(b-\beta)$$

$$\sigma^2 trace\left(X' X\right)^{-1} = E\left(b'b - 2b'\beta + \beta'\beta\right)$$

$$\Rightarrow E(b'b) = \sigma^2 trace\left(X' X\right)^{-1} + \beta'\beta$$

$$\Rightarrow \text{b is generally longer than } \beta$$

$$\Rightarrow \text{OLSE is too large in absolute value}$$

Least squares provide erroneous parameter estimates in the presence of multicollinearity. This does not mean that the fitted model will also provide incorrect predictions. If the estimates are restricted to the space x with harmless multicollinearity, then the estimates are satisfactory.

IV. MULTICOLLINEARITY DIAGNOSTICS

An important question arises as to how to detect the presence of multicollinearity in the data given the sample information. There are many diagnostic measures, each of which is based on a specific approach. It is difficult to say which of the diagnoses is the best or the most definitive. Some popular and important diagnostics are described in more detail. The detection of multicollinearity involves 3 aspects:

1) Determining its presence
2) Determining its severity
3) Determining its form or location

A. *Determinant of* $X' X$ $\left(\left|X' X\right|\right)$:

This measure is based on the fact that the matrix X'X degenerates in the presence of multicollinearity. The value of the multicollinearity determinant of X'X increases.

If Rank(X ' X) <k then $\left|X' X\right|$ will be singular and so $\left|X' X\right|$**=0.** Then as $X' X \to 0$, the degree of multicollinearity increases and it becomes precise or perfect at $\left|X' X\right|$=0. Thereby $\left|X' X\right|$ acts as an action multicollinearity and $\left|X' X\right|$=0 indicates perfect multicollinearity.

Limitations:

This procedure has the following limitations

(i) It is unlimited because $0 < \left|X' X\right| < \infty$.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

(ii) It is affected by the distribution of the explanatory variables. For example, if k<2, then

$$\left|X' X\right| = \begin{vmatrix} \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} \\ \sum_{i=1}^{n} x_{2i}x_{1i} & \sum_{i=1}^{n} x_{2i}^2 \end{vmatrix} = \left(\sum_{i=1}^{n} x_{1i}^2\right)\left(\sum_{i=1}^{n} x_{2i}^2\right)\left(1-r_{12}^2\right)$$

Where r_{12} is the correlation coefficient between X_1 and X_2. So $\left|X' X\right|$ depends on X Correlation coefficient and variance of the explanatory variable. If there are explanatory variables very small variances, then $\left|X' X\right|$ can be zero, indicating the presence Diversity and this is not the case.

(iii) It gives no idea of the relative effects on the individual coefficients. If multicollinearity currently it does not indicate which variable in $\left|X' X\right|$ causes the multicollinearity It's hard to determine.

B. Inspection of Correlation Matrix

Inspection of off-diagonal elements r_{ij} **i**n X ' X gives an idea about the presence of multicollinearity. If the X_i and X_j are almost linearly dependent, and then $\left|r_{ij}\right|$ is close to 1. Note the observations in X Each observation is subtracted from the mean of that variable and divided by the square root of the corrected squares of that variable.

When more than two explanatory variables are considered and if they are involved Dependence, it does not require any r_{ij} will be large. Usually, pairwise check Correlation coefficients are not sufficient to detect multicollinearity in the data.

C. Determinant of Correlation Matrixrity

Let D be the determinant of the correlation matrix, then $0 \le D \le 1$.
If D = 0, it indicates the existence of perfect linear dependence between the explanatory variables.
If D = 1, then the columns of the matrix X are orthonormal.
Thus, a value closer to 0 is indicative of a higher degree of multicollinearity. Any value of D between 0 and 1 gives an idea about the degree of multicollinearity.

Limitations
It does not provide information on the number of linear dependencies between explanatory variables.

Advantages over $\left|X' X\right|$:

(i) It is a finite measure, $0 \le D \le 1$.
(ii) It is not affected by the dispersion of explanatory variables. For example, when k = 2,

$$\begin{vmatrix} \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} \\ \sum_{i=1}^{n} x_{2i}x_{1i} & \sum_{i=1}^{n} x_{2i}^2 \end{vmatrix} = \left(1-r_{12}^2\right)$$

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

D. Measure Based on Partial Regression

A measure of multicollinearity can be obtained based on the coefficients of determination Partial regression. R^2 is the coefficient of determination in the full model, i.e. dependent on all explanatory variables and R^2 is the coefficient of determination in the model when the i[th] explanatory variable is deleted, i=1, 2, ---, k and $\mathrm{R}_{\mathrm{L}}^2 = \max\left(R_1^2, R_2^2, ---, R_k^2\right)$

Procedure:

(i) Remove one of the explanatory variables out of the k variables, say X_1.

(ii) Fit the regression of y on the remainder of (k -1) variables $X_2, X_3 -, --, X_k$

(iii) Calculate R^2.

(iv) Similarly, calculate $R_2^2, R_3^2, ---, R_k^2$.

(v) Find $\mathrm{R}_{\mathrm{L}}^2 = \max\left(R_1^2, R_2^2, ---, R_k^2\right)$

(vi) Determine $\mathrm{R}^2 - \mathrm{R}_{\mathrm{L}}^2$.

Value of ($\mathrm{R}^2 - \mathrm{R}_{\mathrm{L}}^2$) provides a measure of multicollinearity. If multicollinearity exists, $\mathrm{R}_{\mathrm{L}}^2$ will be high. The higher the degree of multicollinearity, the higher the value $\mathrm{R}_{\mathrm{L}}^2$. So in the presence of Multicollinearity,($\mathrm{R}^2 - \mathrm{R}_{\mathrm{L}}^2$) be small. So if ($\mathrm{R}^2 - \mathrm{R}_{\mathrm{L}}^2$) Closer to 0, it indicates a higher degree of multicollinearity.

Limitations:

(i) It does not give any information about the underlying relationships with respect to explanatory variables, i.e. how many relationships exist or how many explanatory variables are responsible for multicollinearity.

(ii) A small value of ($\mathrm{R}^2 - \mathrm{R}_{\mathrm{L}}^2$), it can also happen due to wrong model specification. In such a situation multicollinearity can be expected.

E. Variance Inflation Factors (VIF)

In the presence of multicollinearity in the data the matrix (X ' X) becomes ill. So the diagonal elements of $C = \left(\mathrm{X}' \mathrm{X}\right)^{-1}$ helps detect multicollinearity. Whether there is or not specifies a multiplier of R_i^2 denotes the coefficient of determination taken when X_j is regressed on the remaining (k -1).Except for variables X_j. Then the j[th] diagonal element of C is

$$C_{jj} = \frac{1}{1 - R_j^2}$$

If X_j is approximately orthogonal to the rest of the explanatory variables, then R_j^2 is small and therefore C_{jj} close to 1.

If X_j depends almost linearly on the subset of the remaining explanatory variables, then R_j^2 close to 1 and C_{jj} is very wide. Since by the variance of j[th] OLS estimator of β_j exists as $Var(b_j) = \sigma^2 C_{jj}$.

So C_{jj} factor of variation of b_j increases when the explanatory variables are non-linear addictive. Based on this concept, the variance inflation factor for j[th] explanatory variable is defined as

$$VIF_j = \frac{1}{1 - R_j^2}$$

This factor increases the sample variance. The combined effect of dependencies between explanatory variables on the variance of a term is measured by the VIF of that term in the model.
One or more large VIFs indicate the presence of multicollinearity in the data.
In practice, usually a IVF >5 or 10 indicates that the associated regression coefficients are incorrect estimated due to multicollinearity. If the regression coefficients are estimated by OLSE and their variance is $\sigma^2(\mathrm{X}'\mathrm{X})^{-1}$. Then VIF indicates that part of this variance is contributed by VIF_j.

Limitations:
(i) It does not shed light on the number of dependencies between explanatory variables.
(ii) VIF > 5 or 10 rules may vary from situation to situation.

Another interpretation of VIF_j
The *VIF*s can also be viewed as follows.
The confidence intervals of j[th] OLS estimator of β_j is obtained by

$$\left(b \pm \sqrt{\hat{\sigma}^2 C_{jj}}\, t_{\frac{\alpha}{2},n-k-1}\right)$$

The length of the confidence interval is $L_j = 2\sqrt{\hat{\sigma}^2 C_{jj}}\, t_{\frac{\alpha}{2},n-k-1}$

Now consider the situation where X is an orthogonal matrix such that $(\mathrm{X}'\mathrm{X})^{-1} = I$ so 1, $C_{jj} = 1$, sample size is same as mean square root as $\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)^2\right)$, then the confidence interval is

$$\mathrm{L}^* = 2\hat{\sigma}\, \mathrm{t}_{\frac{\alpha}{2},\mathrm{n}-\mathrm{k}-1}$$

Consider the ratio $\frac{L_j}{L^*} = \sqrt{C_{jj}}$

Thus the $\sqrt{VIF_j}$ represents the increase in the length of the confidence interval of the j[th] regression coefficient because of multicollinearity.

F. Condition Number and Condition Index

Let $\lambda_1, \lambda_2, ---, \lambda_k$ are the Eigen values (or characteristic roots) of $\mathrm{X}'\mathrm{X}$. Let

$$\lambda_{\max} = \max(\lambda_1, \lambda_2, ---, \lambda_k)$$
$$\lambda_{\min} = \min(\lambda_1, \lambda_2, ---, \lambda_k)$$

The condition number (CN) is given by

$$CN = \frac{\lambda_{\max}}{\lambda_{\min}},\ 0 < CN < \infty$$

Small values of the characteristic roots indicate a quasi-linear dependence in the data. The CN provides a measure of the spectral dispersion of characteristic sources of X′ X.

The condition number provides a measure of multicollinearity.

- If CN <100, it is considered harmless multicollinearity.
- If 100 < CN <1000, then this indicates that multicollinearity is moderate to strong. This threshold is called the risk level.
- Whether or not there is CN >1000, this indicates severe (or strong) multicollinearity.

The number of conditions depends on only two Eigen values: minimum maximum λ_{min} and λ_{max}. Another measurement situation index that use information about other eigenvalues.

The condition indices of X′ X are defined as

$$C_j = \frac{\lambda_{max}}{\lambda_j}, \quad j=1, 2, ---. K$$

Suppose, if the largest C_j=CN

More than 1000 conditional codes represent many semi-linear dependencies on X 'X.

The restriction of CN and j C is that they are unbounded functions such that $0 < CN < \infty$, and $0 < C_j < \infty$.

G. Measure Based on Characteristic Roots and Proportion of Variances

Let $\lambda_1, \lambda_2, ---, \lambda_k$ are the Eigen values of X 'X, $\Lambda = diag(\lambda_1, \lambda_2, ---, \lambda_k)$ is kxk and V is a kxk matrix is constructed by the eigenvectors of X ' X . Obviously, V is an orthogonal matrix. So X ' X can be rewrite as $X'X = V\Lambda V'$ .Let $V_1, V_2, ---, V_k$ be a column of V. If semi-linear dependency on the data, then λ_j is close to zero and is explained by linear dependence by the components of the corresponding eigenvector V_j .

The covariance matrix of the OLS estimator is given by

$$V(b) = \sigma^2 (X'X)^{-1} = \sigma^2 (V\Lambda V')^{-1} = \sigma^2 V\Lambda^{-1}V'$$

Thus $\mathrm{Var}(b_i) = \sigma^2 \left(\frac{v_{i1}^2}{\lambda_1}, \frac{v_{i2}^2}{\lambda_2}, ---, \frac{v_{ik}^2}{\lambda_k} \right)$

Where $v_{i1}, v_{i2}, ---, v_{ik}$ are the elements of V.

The condition number indices are $C_j = \frac{\lambda_{max}}{\lambda_j}, \quad j = 1,2, ---, k$

Procedure:

(i) Find the condition index C_1, C_2, -----, C_k.

(ii) (a) Identify λ_i **'s** for which risk level C_j is greater than 1000.

(b) It gives the number of linear dependencies.

(c) Ignore C_j's below the hazard level.

(iii) For such λ_i **'s** condition above the hazard level, select such an eigenvalue, say λ_j.

(iii) Find the value of the coefficient of variation for λ_j in $\mathrm{Var}(b_1), \mathrm{Var}(b_2), ---, \mathrm{Var}(b_k)$ as

$$p_{ij} = \frac{\left(\frac{v_{ij}^2}{\lambda_j} \right)}{VIF_j} = \frac{\frac{v_{ij}^2}{\lambda_j}}{\sum_{j=1}^{k} \left(\frac{v_{ij}^2}{\lambda_j} \right)}$$

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

Note that $\left(\frac{v_{ij}^2}{\lambda_j}\right)$ is obtained from $\mathrm{Var}(b_i)=\sigma^2\left(\frac{v_{i1}^2}{\lambda_1},\frac{v_{i2}^2}{\lambda_2},---,\frac{v_{ik}^2}{\lambda_k}\right)$ it is corresponding to j[th] factor.

The ratio of variance p_{ij} provides a measure of multicollinearity.

If $p_{ij}>0.5$, this indicates that b_i is negatively affected by multicollinearity, i.e., the estimate of β_i is suffers from the presence of multicollinearity.

It is a good diagnostic tool represented by the number of linear dependencies responsible for multicollinearity. This diagnosis better than other diagnoses.

Condition indices are defined by the singular value decomposition of the X matrix as follows:

$$X = UDV'$$

Where U is an nxk matrix, V is a kxk matrix, U ' U = I, V ' V = I, D is a kxk matrix, $D=diag(\mu_1,\mu_2,---,\mu_k)$ and $\mu_1,\mu_2,---,\mu_k$ are singular values of X, V is a column matrix is a matrix whose columns are the eigenvectors of the eigenvalues of X ' X and U are the eigenvectors corresponding to the k nonzero eigenvalues of X ' X.

The condition indices of matrix X are defined as

$$\eta_i=\frac{\mu_{\max}}{\mu_j},\ j=1,2,---,k$$

Where $\mu_{\max}=\max(\mu_1,\mu_2,---,\mu_k)$

If $\lambda_1,\lambda_2,---,\lambda_k$ are the eigenvalues of X ' X then

$$X'X=(UDV)'UDV'=VD^2V'=V\Lambda V'$$

So, $\mu_j^2=\lambda_j$ for $j=1,2,---,k$.

Note that with $\mu_j^2=\lambda_j$,

$$Var(b_j)=\sigma^2\sum_{i=1}^{k}\frac{v_{ji}^2}{\mu_i^2}$$

$$VIF_j=\sum_{i=1}^{k}\frac{v_{ji}^2}{\mu_i^2}$$

$$p_{ij}=\frac{\left(\frac{v_{ji}^2}{\mu_i^2}\right)}{VIF_j}$$

The worst case in X is reflected in the range of singular values. There is a small singular value for each linear dependence. The degree of ill-conditioning is described by the quantity μ_j compared to $\mu_{\max}$.

Explanatory variables are recommended to be measured by unit length but not centered, when calculating p_{ij}. This helps confirm the role of the intercept term non-linear dependence. There is no guidance in the literature on centering explanatory variables. The centering makes the intercept orthogonal to the explanatory variables. So it can eliminate bad conditioning because of the intercept period in the model.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

V. REMEDIES FOR MULTICOLLINEARITY

Many methods have been proposed to solve the problems arising from existence Multicollinearity in data.

A. Obtain More Data

Pernicious multicollinearity arises when X'X has rank less than k and $|X' X|$ is close to zero. Additional data help reduce the sample variance of the estimates. Data required Collected in a way that helps break down multicollinearity in the data.
It is not always possible to collect additional data for various reasons as indicated below:

- The run and process terminates and is no longer available.
- Financial constraints do not allow additional data collection.
- Additional data may be inconsistent with previously collected data and may also be abnormal.
- If the data is in a time series, a longer time series will force you to discard more backward data in past.
- If multicollinearity is caused by some exact identity or relationship, increase the sample size doesn't help.
- Sometimes data is available but it is not advisable to use it. For example, if the data from Usage policy is available for the years 1950-2010, then one may not want to use it. The consumption pattern usually does not stay the same for a long time.

B. Drop Some Variables that are Collinear

If possible, identify variables that cause multicollinearity. These variables can be collinear dropped to match the drop rank state of the X- matrix. The process of leaving variables form organized based on some kind of sequence of explanatory variables, for example, those variables are eliminated first which are smaller value of t-ratio. In another example, suppose the experimenter is not interested in all parameters. In such cases, estimates of the parameters of interest can be obtained OLS estimator has smaller squared errors than the variance when removing some variables. If some variables are omitted, this reduces the predictive power of the model. Sometimes there is there is no guarantee that the model will exhibit low multicollinearity.

C. Use Some Relevant Prior Information

You can search for some relevant prior information on regression coefficients. This may lead to explanation of some coefficient estimates. The most common situation involves specification some exact linear constraints and stochastic linear constraints. Procedures such as constrained regression and mixed regression can be used for this purpose. Relevance and accuracy of information play an important role in such analysis, but difficult to ascertain in practice. For example, expectations derivation in UK may not be valid in India.

D. Employ Generalized Inverse

If rank (X ' X) <k, the generalized inverse can be used to find the inverse of X' X. So β can be estimated as $\hat{\beta} = (X' X)^{-1} X' y$.
In such a case, the estimates are not unique except to use the Moore-Penrose inversion of (X ' X). Different methods for finding the generalized inverse may give different results. So, we will get different results. Also, it is not known what method to find the generalized inverse favorable.

E. Use of Principal Component Regression

Principal components regression is based on principal component analysis technique. The k-explanatory variables are transformed into a new set of orthogonal variables called principal components. Generally, this technique is used to reduce the size of data by retaining some levels variance of explanatory variables expressed by variance in the study variable. The principal components represent the determination of a set of linear combinations of explanatory variables they preserve the total diversity of the system and these linear combinations are mutually exclusive are independent of each other. The principal components obtained are classified in their order Significance. Significance is determined in terms of the variance explained by the principal component regarding the overall diversity of the system. The process involves removing some principal components help explain the relatively small variation. After removing a most significant principal components, multiple regression setup is used instead explanatory variables with principal components. The study variable Regressed against the principal components by selected using ordinary least squares method. From all the main

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

The components are orthogonal; they are independent of each other, so OLS can be used without any problem. After obtaining the estimates of the regression coefficients for the reduced orthogonal variables (principal components), they are mathematically transformed into new estimated regression coefficients for the original set of correlated variables. These new expectations principal component estimates of coefficients regression coefficients.

Suppose there are k explanatory variables $X_1, X_2, ---, X_k$. Consider a linear function of $X_1, X_2, ---, X_k$ such as

$$Z_1 = \sum_{i=1}^{k} a_i X_i$$

$$Z_2 = \sum_{i=1}^{k} b_i X_i \text{ etc.}$$

The constants $a_1, a_2, ---, a_k$ are determined such that the variance of Z_1 is maximized normalization status $\sum_{i=1}^{k} a_i^2 = 0$. The constants $b_1, b_2, ---, b_k$ are determined such that the variance of Z_2 is maximized under normal condition $\sum_{i=1}^{k} b_i^2 = 1$ and is independent of the first principal component.

We continue such a process and obtain k such linear combinations which are orthogonal and their previous linear combinations and satisfy the normal condition. We get their differences. Let these linear combinations be $Z_1, Z_2, ---, Z_k$ and for them, $Var(Z_1) > Var(Z_2) > --- > Var(Z_k)$. The linear combination with the largest variance is the first principal component. Linear combination having the second largest variance is the second largest principal component and so on. These are the main ones components have that property

$$\sum_{i=1}^{k} Var(Zi) = \sum_{i==}^{k} Var(X_i).$$

Also, $X_1, X_2, ---, X_k$ are correlated but $Z_1, Z_2, ---, Z_k$ are orthogonal or uncorrelated. Hence there is zero multicollinearity between $Z_1, Z_2, ---, Z_k$.

The problem of multicollinearity arises because $X_1, X_2, ---, X_k$ are not independent. From the main $X_1, X_2, ---, X_k$ dependent components are independent of each other, so they can be used descriptively variables, and such regression struggles with multicollinearity.

Let $\lambda_1, \lambda_2, ---, \lambda_k$ are the eigenvalues of $X'X$, $\Lambda = diag(\lambda_1, \lambda_2, ---, \lambda_k)$ is a kxk diagonal matrix, V is an kxk orthogonal matrix containing the eigenvectors associated with $\lambda_1, \lambda_2, ---, \lambda_k$. Enter the account a canonical form of a linear model

$$y = X\beta + \varepsilon$$
$$y = XVV'\beta + \varepsilon$$
$$y = Z\alpha + \varepsilon$$

Here Z=XV, $\alpha = V'\beta$, $V'X'XV = Z'Z = \Lambda$.

The columns of $(Z_1, Z_2, ---, Z_k)$ define new explanatory variables, called principal components.

The OLS estimator of α is

$$\hat{\alpha} = (Z'Z)^{-1} Z'y$$
$$\hat{\alpha} = \Lambda^{-1} Z'y$$

and the covariance matrix is

$$Var(\hat{\alpha}) = \sigma^2 (Z'Z)^{-1}$$

$$Var(\hat{\alpha}) = \sigma^2 \Lambda^{-1}$$

$$Var(\hat{\alpha}) = \sigma^2 diag\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, ---. \frac{1}{\lambda_k}\right)$$

Note that λ_j is the variance of the j[th] principal component and $Z'Z = \sum_{i=1}^{k}\sum_{j=1}^{k} Z_i Z_j = \Lambda$. A small eigenvalue of X ' X means that there is a linear relationship between the original explanatory variable and the variance of the corresponding orthogonal regression coefficient is large, indicating there is multicollinearity. If one or more λ_j is small, then it indicates that multicollinearity exists.

1) Retainment of Principal Components

The new sets of variables, i.e. the principal components, are orthogonal and of the same magnitude deviation from the original set. If multicollinearity is severe, then there is at least one small value of eigenvalue. Elimination of one or more principal components associated with the smallest eigenvalues will reduce the total variance of the model. Moreover, the main component responsible for creating Multicollinearity is removed and the resulting model is significantly improved.

The principal component matrix $Z = [Z_1, Z_2, ---, Z_k]$ with $Z_1, Z_2, ---, Z_k$ has exactly the same information than the original data in X, the total variance in X and Z is the same. The difference between them is that the original data is handled as new variables are uncorrelated and can be classified according to the magnitude of their eigenvalues. The j[th] vertical vector Z_j corresponds to the largest λ_j **r**epresents the highest proportion of variation in original data. Thus, Z_j is indexed so that $\lambda_1 > \lambda_2 > --- > \lambda_k > 0$ and λ_j is the variance of Z_j.

One strategy for removing core components is to start by removing the associated component is the smallest eigenvalue. The idea behind this is to have a main body with a smaller one, the eigenvalue contributes the least to the variance and is therefore the least informative.

F. Ridge Regression

OLS estimator is the best linear unbiased estimator of the regression coefficient that is the minimum variance in class of linear and unbiased estimates. However, if there is an unbiased situation relaxed, and then it is possible to find a biased estimator of the regression coefficient, say $\hat{\beta}$, which is small and they are unbiased OLS estimator of b. The Mean Square Error (MSE) of $\hat{\beta}$ is given by

$$MSE(\hat{\beta}) = E\left[\hat{\beta} - \beta\right]^2$$

$$MSE(\hat{\beta}) = E\left[\left\{\hat{\beta} - E[\hat{\beta}]\right\} + \left\{E[\hat{\beta}] - \beta\right\}\right]^2$$

$$MSE(\hat{\beta}) = Var(\hat{\beta}) + E\left[\left\{E[\hat{\beta}] - \beta\right\}\right]^2$$

$$MSE(\hat{\beta}) = Var(\hat{\beta}) + [Bias]^2$$

Therefore, MSE($\hat{\beta}$ **)** can be made smaller than Var($\hat{\beta}$ **)** by introducing a small bias $\hat{\beta}$. One of the procedures doing so is ridge regression. The peak regression estimator is obtained by solving the general equations of least squares estimation. The general equations are modified as

$$(X'X + \delta I)\hat{\beta}_{Ridge} = X'y$$

$$\Rightarrow \hat{\beta}_{Ridge} = (X'X + \delta I)^{-1} X'y$$

Is a ridge regression estimator of β and $\delta \geq 0$ is any characterizing scalar called the bias parameter.

$$\text{As}\,\delta \to 0\,,\ \hat{\beta} \to b(\text{OLS Estimator}) \ \text{and as}\,\delta \to \infty\,,\ \hat{\beta} \to 0$$

So the higher the value of δ, the greater the shrinkage towards zero. Note that OLSE is not appropriate for use in when there is multicollinearity in the data it accounts for much more variance. On the other hand, a lot A small value of $\hat{\beta}$ may accept the null hypothesis $H_0 : \beta = 0$, which Variables are irrelevant. The value of the bias parameter controls the amount of shrinkage expectations.

1) Bias of Ridge Regression Estimator

The bias of $\hat{\beta}_{Ridge}$ is

$$Bias\left(\hat{\beta}_{Ridge}\right) = E\left[\hat{\beta}_{Ridge}\right] - \beta$$

$$Bias\left(\hat{\beta}_{Ridge}\right) = \left(X'X + \delta I\right)^{-1} X'E[y] - \beta$$

$$Bias\left(\hat{\beta}_{Ridge}\right) = \left(X'X + \delta I\right)^{-1} X'X\beta - \beta$$

$$Bias\left(\hat{\beta}_{Ridge}\right) = \left[\left(X'X + \delta I\right)^{-1} X'X - I\right]\beta$$

$$Bias\left(\hat{\beta}_{Ridge}\right) = \left(X'X + \delta I\right)^{-1}\left[X'X - X'X - \delta I\right]\beta$$

$$Bias\left(\hat{\beta}_{Ridge}\right) = -\delta\left(X'X + \delta I\right)^{-1}\beta$$

Thus, the ridge regression estimator is a biased estimate of β.

2) Covariance Matrix

The covariance matrix of $\hat{\beta}_{Ridge}$ is defined as

$$Var\left(\hat{\beta}_{Ridge}\right) = E\left[\left\{\hat{\beta}_{Ridge} - E\left[\hat{\beta}_{Ridge}\right]\right\}\left\{\hat{\beta}_{Ridge} - E\left[\hat{\beta}_{Ridge}\right]\right\}'\right]$$

Since

$$\hat{\beta}_{Ridge} - E\left[\hat{\beta}_{Ridge}\right] = \left(X'X + \delta I\right)^{-1} X'y - \left(X'X + \delta I\right)^{-1} X'X\beta$$

$$\hat{\beta}_{Ridge} - E\left[\hat{\beta}_{Ridge}\right] = \left(X'X + \delta I\right)^{-1} X'(y - X\beta)$$

$$\hat{\beta}_{Ridge} - E\left[\hat{\beta}_{Ridge}\right] = \left(X'X + \delta I\right)^{-1} X'\varepsilon$$

So $$Var\left(\hat{\beta}_{Ridge}\right) = \left(X'X + \delta I\right)^{-1} X'V(\varepsilon)X\left(X'X + \delta I\right)^{-1}$$

$$Var\left(\hat{\beta}_{Ridge}\right) = \sigma^2\left(X'X + \delta I\right)^{-1} X'X\left(X'X + \delta I\right)^{-1}$$

3) Mean Squared Error

The mean squared error of $\hat{\beta}_{Ridge}$ is defined as

$$MSE\left(\hat{\beta}_{Ridge}\right) = Var\left(\hat{\beta}_{Ridge}\right) + \left[Bias\left(\hat{\beta}_{Ridge}\right)\right]^2$$

$$MSE\left(\hat{\beta}_{Ridge}\right) = trace\left[Var\left(\hat{\beta}_{Ridge}\right)\right] + \left[Bias\left(\hat{\beta}_{Ridge}\right)\right]^2$$

$$MSE\left(\hat{\beta}_{Ridge}\right) = \sigma^2 trace\left[\left(X'X + \delta I\right)^{-1} X'X\left(X'X + \delta I\right)^{-1}\right] + \delta^2\beta'\left(X'X + \delta I\right)^{-2}\beta$$

$$MSE\left(\hat{\beta}_{Ridge}\right)=\sigma^2\sum_{j=1}^{k}\frac{\lambda_j}{\left(\lambda_j+\delta^2\right)}+\delta^2\beta'\left(X'X+\delta I\right)^{-2}\beta$$

Here $\lambda_1,\lambda_2,---,\lambda_k$ are the eigenvalues of $X'X$

Thus, as δ increases, so does the bias of $\hat{\beta}_{Ridge}$ increases but its variance decreases. Thus, the trade-off between bias and the difference depends on the value of δ . It can be shown that δ has such a value

$$MSE\left(\hat{\beta}_{Ridge}\right)<Var(b)$$

given that $\beta'\beta$ is bounded.

4) Idea Behind Ridge Regression Estimator

The problem of multicollinearity arises because some roots of the eigenvalues of X'X are close to zero (or are zero). So $\lambda_1,\lambda_2,---,\lambda_p$ are characteristic roots if , and if

$$X'X=\Lambda=diag(\lambda_1,\lambda_2,---,\lambda_k)$$

Then

$$\hat{\beta}_{Ridge}=\left(I+\delta\Lambda^{-1}\right)^{-1}b$$

Here b is OLS estimator of β is given by

$$b=\left(X'X\right)^{-1}X'y$$

$$b=\Lambda^{-1}X'y$$

So a particular element is of the form

$$\frac{1}{1+\frac{\delta}{\lambda_i}}b_i=\frac{\lambda_i}{\lambda_i+\delta}b_i$$

So a small quantity δ is added to λ_i so if $\lambda_i=0$, then $\frac{\lambda_i}{\lambda_i+\delta}$ is meaningful.

5) Another interpretation of ridge regression estimator:

In the $y=X\beta+\varepsilon$ model, obtain the least squares estimate of β when $\sum_{i=1}^{k}\beta_i^2=C$, where C is a constant. So minimize

$$S(\beta)=\left(y-X\beta\right)'\left(y-X\beta\right)+\delta\left(\beta'\beta-C\right)$$

Where δ is a Lagrangian coefficient.

Differentiating $S(\beta)$ with respect to β, simple normal equations are

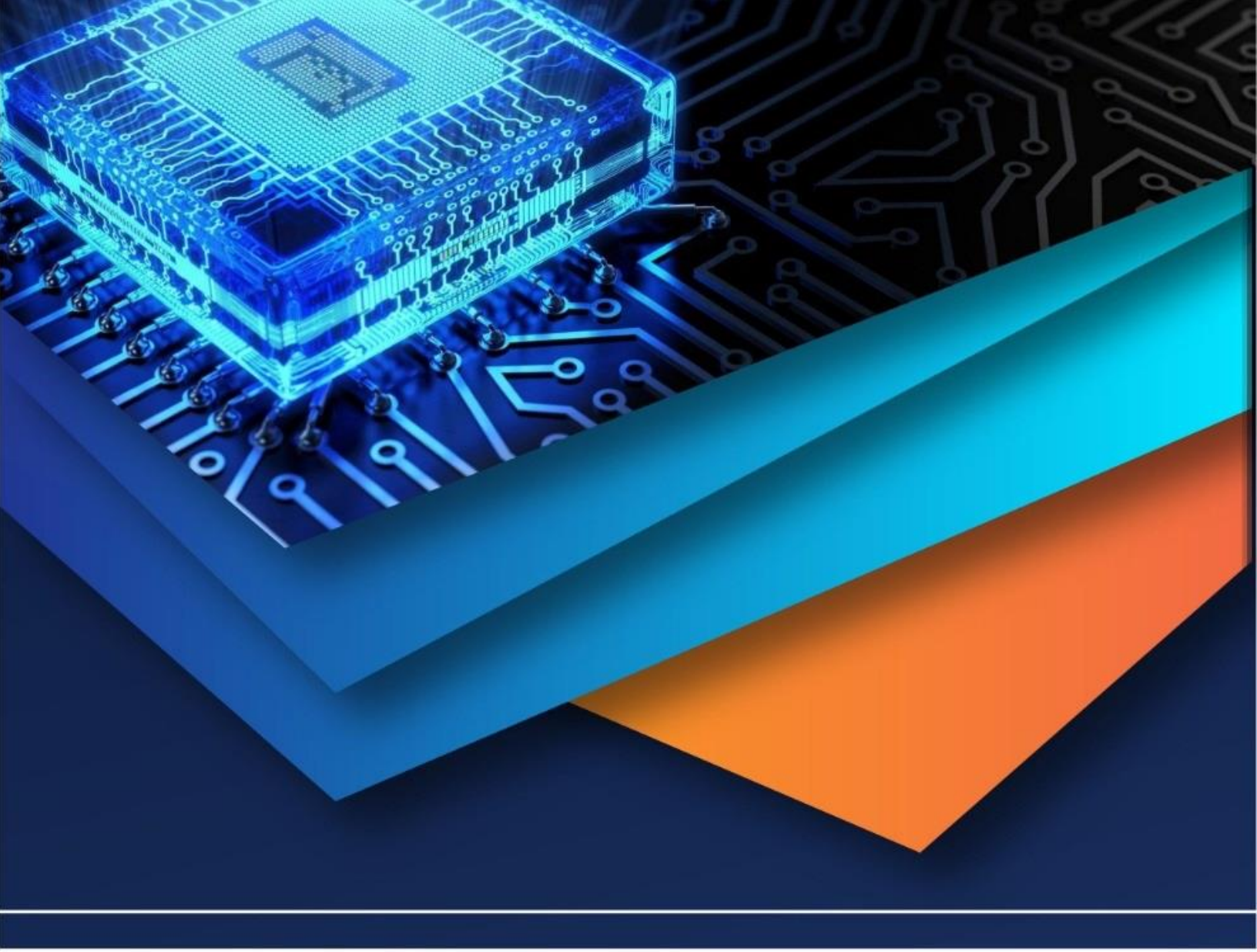
$$\frac{\partial S(\beta)}{\partial\beta}=0\Rightarrow-2X'y+2X'X\beta+2\delta\beta=0$$

$$\Rightarrow\hat{\beta}_{Ridge}=\left(X'X+\delta\,I\right)^{-1}X'y$$

If C is very small, this may indicate that most of the regression coefficients are close to zero, and if C is large, which may indicate that the regression coefficients are far from zero. So C keeps a kind penalty on the regression coefficients to allow its estimation.

REFERENCES

[1] Vatcheva, K.P., Lee, M., McCormick, J.B., and Rahbar, M.H., "Multicollinearity in regression analysis conducted in epidemiologic studies," Epidemiology (Sunnyvale, Calif.), 6 (2). 227. 2016.
[2] "Applied Multivariate Statistical Analysis" by Richard A. Johnson and Dean W. Wichern.
[3] Gunst, R.F. and Webster, J.T., "Regression analysis and problems of multicollinearity," Communications in Statistics, 4 (3). 277-292. 1975.
[4] Kleinbaum and David G 2008 Applied regression analysis and other multivariable methods (Australia; Belmont, CA: Brooks/Cole) 906
[5] "Regression Analysis and Its Application: A Data-Oriented Approach" by Richard F. Gunst and Robert L. Mason.
[6] Debbie J Dupuis and Maria-Pia Victoria-Feser 2013 Robust VIF regression with application to variable Selection in large data sets The Annals of Applied Statistics 7 319-341
[7] Jensen D.R and Ramirez D.E. 2012 Variance Inflation in Regression, Advances in Decision Sciences 1-15 2013
[8] "Linear Regression Analysis" by George A. F. Seber and Alan J. Lee.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY