# Ridge Regression for the Prediction of Compressive Strength of Concrete

Palika Chopra
*Lecturer and Ph. D Scholar,School of Mathematics and Computer Applications*
*Thapar University, Patiala-147004, INDIA*

Dr. R. K. Sharma
*Professor,School of Mathematics and Computer Applications*
*Thapar University,Patiala-147004, INDIA*

Dr. Maneek Kumar
*Professor and Head,Department of Civil Engineering*
*Thapar University, Patiala-147004, INDIA*

**Abstract- In this paper, regression analysis method i.e., traditional method and ridge regression is applied on the dataset collected for the prediction of compressive strength of concrete. When regression analysis is applied and data set is increased or decreased, the regression coefficients have drastically changed their values like the negative coefficients changed to positive or vice versa. So, traditional method is not reliable method for the prediction of compressive strength of concrete. And when we applied ridge regression, then there is no effect on the coefficients either there is increase or decrease in the data set or very minimal effect that is negligible.**

**Keywords – ridge, regression, prediction, concrete, compressive strength.**

## I. INTRODUCTION

Multicollinearity refers to a situation in which one or more predictor variables in a multiple regression model are highly correlated if multicollinearity is perfect , the regression coefficients are indeterminate and their standard errors are infinite, if it is less than perfect. The regression coefficients although determinate but posses large standard errors, which means that the coefficients cannot be estimated with great accuracy (Gujarati, 1995).Multicollinearity has several effects, these are described as follow :
1.) High variance of coefficients may reduce the precision of estimation. 2.)Multicollinearity can result in coefficients appearing to have the wrong sign.3.)Estimates of coefficients may be sensitive to particular sets of sample data.4.)Some variables may be dropped from the model although, they are important in the population.5.)The coefficients are sensitive of to the presence of small number inaccurate data values (more details in Judge 1988, Gujarat; 1995).Because multicollinearity is a serious problem when we are working for predictive models. So it is very important for us to find a better method to deal with multicollinearity. Therefore, in this paper, we are comparing different models using the ordinary least square method with the ridge regression to find the accurate results. The aim of the present study is to predict compressive strength of concrete for a given sample, as accurately as possible. For this purpose, multiple regression analysis is used for predicting the compressive strength of concrete using four variables, namely, water-cementations ratio, fine aggregate-cementitious ratio, coarse aggregate-cementitious ratio and cementitious content. Regression models have been developed for concrete with medium and high workability at different curing ages(28,56 and 91 days). For models developed for 56 and 91 days curing ages, the compressive strength at lower age has also been considered as a parameter.

## II. EXPERIMENTAL DATASET

The most popular regression equation, which has been used by many researchers, in the prediction of compressive strength is the linear regression equation:

$$f_c = a_0 + a_1(^{w}/_{cm}) \qquad (1)$$

Where, $f_c$= compressive strength of concrete; $^{w}/_{cm}$= water to cementitious material ration; and $a_0$ and $a_1$ are regression coefficients.
The origin of this equation is Abram's Law which states that the concrete materials, for a mixture of workable consistency, the strength of concrete is determined by the ration of water to cement (*Popovices & Ujhelyi,* 2008). So according to this law, increasing the *w/cm* ration will definitely lead to decrease in concrete strength. The original formula for Abram is as given by Eq. (2) below:

$$f_c = \frac{A}{B^{W}/_{cm}} \qquad (2)$$

Where, $f_c$ = compressive strength of concrete; $^{w}/_{cm}$ = water to cementitious material ration; and A and B are empirical constants.
Lyse (*Jee et al., 2004*) made a formula similar to Abram but he related compressive strength to cement/water ratio and not water/cement ratio. According to Lyse (*Zain et al,2009*), strength of concrete increases linearly with an increase in the *cm/w* ratio, and a general form of this popular model was:

$$f_c = A + B\,(^{cm}/_{w}) \qquad (3)$$

where, $f_c$ = compressive strength of concrete; $^{cm}/_{w}$= cementitious material to water ration; and *A* and *B* are empirical constants.
The model as proposed by Abram and Lyse did not account for the quantities of fine aggregates and coarse aggregates for the prediction of concrete strength. So, for various concrete mixes where their $^{w}/_{cm}$ratio is constant, the strength will be the same is not true. Thus, it made it imperative to accommodate all the constituent materials into the predicting equation to have more reliable and accurate results for the prediction of concrete strength.
For the reason, Abram's Law has been extended by various researchers, to include other variable in the form of multiple linear regression equation and used widely to predict the compressive strength of various types of concrete.

$$f_c = b_0 + b_1\,(^{w}/_{cm}) + b_2\,(^{FA}/_{cm}) + b_3\,(^{CA}/_{cm}) \qquad (4)$$

Where, $f_c$= compressive strength of concrete; $^{w}/_{cm}$= water to cementitious material ; $^{FA}/_{cm}$= Fine aggregate to cementitous material ration; $^{CA}/_{cm}$= Coarse aggregate to cementitious material ratio; and $b_0$, $b_1$ and $b_2$ are regression coefficients.
According to Eq4, all the variables related to the compressive strength in a linear fashion, but this is not always true because the variable involved in a concrete mix and affecting its compressive strength are interrelated with each other and the additive action is not always true. Here, it appears that there is a need to develop another type of mathematical model that can reliably predict the compressive strength of concrete with acceptable high accuracy. So, if we take the general form of the multiple linear regression as below:

$$Y = a_0 + a_1x_1 + a_2x_2 + \ldots\ldots\ldots\ldots\ldots\ldots + a_mx_m$$

For situation there the multiple dependencies are non-linear, the logarithmic transformation can also be applied to this type of regression as:

$$\log(Y) = \log(a_0) + a_1\log(x_1) + a_2\log(x_2) + \ldots\ldots\ldots\ldots + a_m\log(x_m) \qquad (5)$$

The Eq. 5 could be transformed back to a form that predicts the dependent variable (*Y*) by taking antilogarithm to yield and equation of type:

$$Y = a_0 x_1^{a_1} x_2^{a_2} x_3^{a_3} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\; x_{m1}^{a_m} \qquad (6)$$

This Equation is known as multivariable power equation and, in engineering as variable are often dependent on several independent variable, this functional dependency is best characterized by the equation mentioned earlier and is said to give results that are more realistic too The equation has been successfully used to predict the compressive strength for Ordinary Portland Cement Concrete also (*Kheder et. al, 2003*) and for compressive strength prediction of high performance concrete (*Zain et al, 2009*).
In this study, the multivariable power equation has been used for prediction of compressive strength of concretefor varying workability specifically manufactured in the laboratory under control conditions. Factors affecting this strength were the elements of the concrete mix itself. for developing the model, in this study, the mean value of compressive strength of concrete at 28,56 and 91 days are predicted and the output of one is used as one of the input variable for the prediction of another like the predicted compressive strength of 28 days(output of 28 days) is entered as one of theinput variable to predict the compressive strength of 56 days and the output of 28 days and 56 days are accommodated asinput variables to predict the compressive strength of 91 days. Other than this, 91 days

compressive strength is also predicted using only strength of 28 days and also using only strength of 56 days. The experimental data generated and as provided in Tables 3 to 6 (*Kumar, 2003*) is used for regression analysis. The final form of the regression equations for different cases as per Model is provided as below:

$$f_{c,28} = exp\ [B_0 + B_1\left(\frac{w}{cm}\right) + B_2\,(cm) + B_3\left(\frac{cm}{FA+CA1}\right) + B_4\left(\frac{cm}{FA+CA2}\right)] \quad (7)$$

$$f_{c,56} = exp\ [B_0 + B_1\left(\frac{w}{cm}\right) + B_2\,(cm) + B_3\left(\frac{cm}{FA+CA1}\right) + B_4\left(\frac{cm}{FA+CA2}\right) + B_5\left(\frac{cm}{f_{c,28}}\right)] \quad (8)$$

$$f_{c,91} = exp\ [B_0 + B_1\left(\frac{w}{cm}\right) + B_2\,(cm) + B_3\left(\frac{cm}{FA+CA1}\right) + B_4\left(\frac{cm}{FA+CA2}\right) + B_5\left(\frac{cm}{f_{c,28}}\right) + B_6\left(\frac{cm}{f_{c,56}}\right)] \quad (9)$$

$$f_{c,91,28} = exp\ [B_0 + B_1\left(\frac{w}{cm}\right) + B_2\,(cm) + B_3\left(\frac{cm}{FA+CA1}\right) + B_4\left(\frac{cm}{FA+CA2}\right) + B_5\left(\frac{cm}{f_{c,28}}\right)] \quad (10)$$

$$f_{c,91,56} = exp\ [B_0 + B_1\left(\frac{w}{cm}\right) + B_2\,(cm) + B_3\left(\frac{cm}{FA+CA1}\right) + B_4\left(\frac{cm}{FA+CA2}\right) + B_6\left(\frac{cm}{f_{c,56}}\right)] \quad (11)$$

IV.RIDGE REGRESSION

Coefficient estimates for the models described in linear regression rely on the independence of the model terms. When terms are correlated and the columns of the design matrix *X* have an approximate linear dependence, the matrix $(XTX)^{-1}$ becomes close to singular. As a result, the least-squares estimate

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad \ldots\ldots \quad (12)$$

becomes highly sensitive to random errors in the observed response *y*, producing a large variance. This situation of *multicollinearity* can arise, for example, when data are collected without an experimental design. *Ridge regression* addresses the problem by estimating regression coefficients using

$$\hat{\beta} = (X^T X + kI)^{-1} X^T y \quad \ldots\ldots. \quad (13)$$

where *k* is the *ridge parameter* and *I* is the identity matrix. While biased, the reduced variance of ridge estimates often result in a smaller mean square error when compared to least-squares estimates.

V. RESULTS AND DISCUSSIONS

Resgression Analysis and ridge regression is applied on the data set of 14 and data set of 8 also.Table 1 to Table 4 are the results regression analysis and Table 5 to Table 8 are the results of ridge regression.It has been seen that during regression analysis when we reduced the data set of 14 to dataset of 8, the value of the coefficients have been changed. It is not the small or negligible change, the value sometimes changed from the negative value to positive or sometimes positive value to negative. It means as the increase or decrease in the number of data set would effect on the prediction of strength. Therefore , for the accurate prediction , ridge regression is applied in which small values of k, ridge parameter improves the conditioning problem and reduce the variance of the estimates as shown in the tables below.

| DATA SET =14 | | | | | |
|---|---|---|---|---|---|
| | M28DAYS | M56DAYS | M91DAYS | M91WITH28 | M91WITH56 |
| a0 | 6.9549 | 4.5112 | 4.3499 | 4.5507 | 3.8336 |
| a1 | -5.3138 | -1.1579 | -1.454 | -1.9211 | -0.67283 |
| a2 | -0.00045 | 0.001612 | 0.001303 | 0.000879 | 0.0016408 |
| a3 | -0.32213 | 0.027984 | 0.16602 | 0.21862 | 0.25983 |

| | | | | | |
|---|---|---|---|---|---|
| a4 | -0.77362 | -0.23342 | -0.06629 | -0.1884 | 0.018516 |
| a5 | | -0.06724 | 0.055108 | -0.00135 | -0.03724 |
| a6 | | | -0.09157 | | |

Table 1: Regression coefficients of multiple regression models predicting compressive strength of concrete with medium workability as per Model ,when data set is of 14.

| DATA SET =08 | | | | | |
|---|---|---|---|---|---|
| | M28DAYS | M56DAYS | M91DAYS | M91WITH28 | M91WITH56 |
| a0 | 0.000484 | 0 | 0 | 0 | 0 |
| a1 | 2.7178 | 4.6605 | 5.0901 | 4.1002 | 4.7504 |
| a2 | 0.002793 | 0.003575 | 0.003311 | 0.002789 | 0.0031205 |
| a3 | 0.51339 | 0.41178 | 0.39998 | 0.56934 | 0.46552 |
| a4 | 2.6187 | 2.2912 | 2.3583 | 2.4017 | 2.3618 |
| a5 | | -0.10391 | 0.046315 | -0.04423 | -0.09782 |
| a6 | | | -0.1763 | | |

Table 2: Regression coefficients of multiple regression models predicting compressive strength of concrete with medium workability as per Model , when data set is of 8.

In Table1 & Table 2, the regression coefficients of multiple regression model with medium workability having data set of 14 and data set of 08 respectively. As the data set is reduced to 08, the coefficients predicted for data set of 14 is drastically changed. As the coefficient a0 reached to 0 from 4.5112 for M56days , 0 from 4.3499 for M91days,0 from 4.5507 for M91with28 and 0 from 3.8336 for m91with56. Some coefficients have been changed from negative to positive like for M28days, the coefficient a1 is changed to 2.7178 from -5.3138.

| DATA SET =14 | | | | | |
|---|---|---|---|---|---|
| | H28DAYS | H56DAYS | H91DAYS | H91WITH28 | H91WITH56 |
| a0 | 0.003984 | 0.002273 | 0.002245 | 0.002275 | 4.5541 |
| a1 | 2.7745 | 6.1736 | 5.9661 | 5.9441 | 0.0022949 |
| a2 | 0.004182 | 0.006259 | 0.005736 | 0.005748 | 5.4525 |
| a3 | 1.5709 | 0.25336 | 0.37106 | 0.3808 | -0.1973 |
| a4 | 0.31243 | 0.3211 | 0.38516 | 0.38781 | -0.10159 |
| a5 | | -0.20873 | -0.18476 | -0.17617 | |
| a6 | | | 0.00949 | | |

Table 3: Regression coefficients of multiple regression models predicting compressive strength of concrete with high workability as per Model, when dataset is of 14.

| DATA SET =08 | | | | | |
|---|---|---|---|---|---|
| | H28DAYS | H56DAYS | H91DAYS | H91WITH28 | H91WITH56 |
| a0 | 0 | 0 | 0 | 0 | 0 |
| a1 | 3.0099 | 5.6218 | 4.6651 | 5.1006 | 4.5541 |
| a2 | 0.000752 | 0.005441 | 0.002536 | 0.003905 | 0.0022949 |
| a3 | 6.8585 | 1.3285 | 5.1375 | 3.0826 | 5.4525 |
| a4 | -0.35974 | 0.11246 | -0.17079 | 0.000487 | -0.1973 |
| a5 | | -0.16853 | -0.01545 | -0.12089 | -0.10159 |
| a6 | | | -0.09293 | | |

Table 4: Regression coefficients of multiple regression models predicting compressive strength of concrete with high workability as per Model, when dataset is of 8.

In Table3 and Table 4 , there are regression coefficients of multiple regression model for predicting compressive strength of concrete with high workability for data set of 14 and data set of 08 respectively. In these tables, when data set is reduced to 08, the regression coefficients have changed their values, same as for medium workability in Table 1 and Table 2.

| DATA SET =14 | | | | | |
|---|---|---|---|---|---|
| | M28DAYS | M56DAYS | M91DAYS | M91WITH28 | M91WITH56 |
| a0 | 3.717 | 3.8277 | 3.8853 | 3.8853 | 3.8853 |
| a1 | -0.13254 | -0.0405 | -0.03936 | -0.04583 | -0.03189 |
| a2 | 0.032653 | 0.056586 | 0.051427 | 0.045574 | 0.052365 |
| a3 | 0.002634 | 0.005696 | 0.001501 | 0.018571 | 0.016262 |
| a4 | -0.02432 | -0.01168 | 0.002095 | -0.00197 | 0.0023376 |
| a5 | | -0.04871 | 0.020147 | -0.00895 | -0.018083 |
| a6 | | | -0.03282 | | |

Table 5 : Ridge coefficients of multiple regression models predicting compressive strength of concrete with medium workability as per Model ,when data set is of 14.

| DATA SET =08 | | | | | |
|---|---|---|---|---|---|
| | M28DAYS | M56DAYS | M91DAYS | M91WITH28 | M91WITH56 |
| a0 | 3.6262 | 3.7354 | 3.8006 | 3.8006 | 3.8006 |
| a1 | -0.11943 | -0.03451 | -0.02898 | -0.03551 | -0.025849 |
| a2 | 0.022534 | 0.048913 | 0.039778 | 0.037564 | 0.040624 |
| a3 | 0.001481 | 0.006686 | 0.013977 | 0.017045 | 0.01531 |
| a4 | 0.002255 | 0.017026 | 0.022154 | 0.024607 | 0.024007 |
| a5 | | -0.04584 | 0.014979 | -0.00372 | -0.013573 |
| a6 | | | -0.0278 | | |

Table 6: Ridge coefficients of multiple regression models predicting compressive strength of concrete with medium workability as per Model ,when data set is of 8.

| DATA SET =14 | | | | | |
|---|---|---|---|---|---|
| | H28DAYS | H56DAYS | H91DAYS | H91WITH28 | H91WITH56 |
| a0 | 3.7056 | 3.8217 | 3.8746 | 3.8746 | 3.8746 |
| a1 | -0.06717 | 0.00554 | -0.002 | 0.004257 | -0.012078 |
| a2 | 0.024296 | 0.06078 | 0.050362 | 0.05029 | 0.04541 |
| a3 | 0.01838 | 0.000873 | 0.004372 | 0.004226 | 0.006656 |
| a4 | -0.00663 | -0.00256 | 0.006852 | 0.006279 | 0.006275 |
| a5 | | -0.04384 | -0.01302 | -0.03204 | -0.023921 |
| a6 | | | -0.01759 | | |

Table 7: Ridge coefficients of multiple regression models predicting compressive strength of concrete with high workability as per Model ,when data set is of 14.

| DATA SET =08 | | | | | |
|---|---|---|---|---|---|
| | H28DAYS | H56DAYS | H91DAYS | H91WITH28 | H91WITH56 |
| a0 | 3.658 | 3.7669 | 3.8203 | 3.8203 | 3.8203 |
| a1 | -0.06122 | -0.01942 | -0.01883 | -0.01958 | -0.023402 |
| a2 | 0.01635 | 0.060404 | 0.042513 | 0.046058 | 0.037657 |
| a3 | 0.041484 | -0.01674 | 0.016566 | 0.00382 | 0.023005 |
| a4 | -0.01338 | 0.002151 | -0.00152 | 0.001902 | -0.0032944 |
| a5 | | -0.04028 | -0.01012 | -0.03032 | -0.025112 |
| a6 | | | -0.02006 | | |

Table 8: Ridge coefficients of multiple regression models predicting compressive strength of concrete with high workability as per Model , when data set is of 8.

In Tables 5, 6, 7 and 8, the ridge regression is applied for the data set of 14 and data set of 08 with medium and high workability. It is shown in the tables, it has no effect or very minimal effect on the ridge coefficients, when the data set is reduced. Therefore, ridge regression is more reliable method for the prediction purposes.

REFERENCES

[1] Gujarati, Damodar,1998. “Basic Econometrics, Singapore: McGraw-Hill Book Company”.

[2] J.Namyong, Y.Sangchun, and C.Hong Bum, 2004.”Prediction of compressive strength of concrete based on mixture proportions”. Asian Architect Build. Engineering. 3: 9-16.

[3] M.F.M.Zain, H.B.,Mahmud, A.Ilham, andM. Faizal, 2002.” Prediction of splitting tensile strength of high-performance concrete”. Cement Concrete Research, 32: 1251-1258.

[4] M.Kumar, 2003.” Reliability Based Design of Structural Elements”. Thesis submitted for degree of Doctor of Philosophy, T.I.E.T Patiala

[5] S.Popovics, and J . Ujhelyi, 2008. “Contribution to the concrete strength versus water-cement ratio relationship. J.Mater.Civil Eng.” 20:459-463.

[6] S.Popovics, 1990. “Analysis of concrete strength versus water-cement ratio relationship. ACI Mater”. J., 87: 517-529.

[7] M. El-Dereny and N. I. Rashwan.” Solving Multicollinearity Problem Using Ridge Regression Models”. Int. J. Contemp. Math. Sciences, Vol. 6, 2011, no. 12, 585 - 600

[8] A. A. Reddy and B. N. Chatterji, "A new wavelet based logo-watermarking scheme", Pattern Recognition Letters, vol. 26, pp. 1019-1027, 2005.

[9] S. Angshuman, “Application of Ridge Regression for Improved Estimation of Parameters in compartmental Models”, Ph.D. Dissertation, Washington University, 1998.

[10] D.K. Guilkey, and J.L. Murphy, Directed Ridge Regression Techniques in cases of Multicollinearity, Journal of American Statistical Association,70(1975) , 767-775.