Available: https://doi.org/10.54117/iimj.v3i1.5 *Research article*

IPS Intelligentsia Multidisciplinary Journal, 3(2): 1-8 (2023)
Received: 18 Dec 2022, Accepted: 22 Jan 2023, Published: 08 Feb 2023

Comparative Analysis of Ridge, Bridge and Lasso Regression Models In the Presence of Multicollinearity

Kelachi Enwere[1*], Ethelbert Nduka[2], Uchenna Ogoke[3]

[1]*Department of Mathematics and Statistics, University of Port Harcourt, Nigeria.*
[2]*Department of Mathematics and Statistics, University of Port Harcourt, Nigeria.*
[3]*Department of Mathematics and Statistics, University of Port Harcourt, Nigeria.*

*Corresponding author e-mail: uchenna.ogoke@uniport.edu.ng

Abstract

Aims: This research work investigated the best regression technique in handling multicollinearity using the Ridge, Least Absolute Shrinkage and Selection Operator (LASSO) and Bridge regression models in comparison to Analysis and Prediction.
Study Design: Two sets of secondary data on Body Size and Heart Rate gotten from the Lulu Briggs Health Center, University of Port Harcourt were used for comparison for model fit and in handling multicollinearity between the regression techniques. Tables were used to present Comparisons made using MSE, RMSE, VIF, AIC and BIC for efficiency. Scatter plots were employed to show fitted regression models. R Software was used to perform data analysis.
Methodology: The data were tested for the presence of multicollinearity using VIF respectively, before proceeding to apply Ridge, LASSO and Bridge regression techniques to solve the problem of multicollinearity. Then comparison was made in analysis and prediction between the regression techniques.
Results: The results from the study show that, for analysis on body size, we found that none of the Regression Techniques handled the problem of multicollinearity, even though the degree of multicollinearity present in the data set reduces, with VIF values of 11.36762 for Ridge, 10.8042 for LASSO, and Bridge which are 10.95578, 11.24945, 12.22628 and 12.14645 respectively. For Heart Rate analysis, we see that all the regularized regression techniques handled the problem of multicollinearity. The results show that the Bridge regression technique performed better with a VIF of 1.744461 when $\gamma = 1$. Second to it was the Ridge regression with VIF of 1.914978, and lastly the LASSO regression with VIF of 2.184537 respectively. In comparison for best model fit, Bridge regression performed better for both datasets. For body size analysis, with MSE of 13.79458 when $\gamma = 1.5$, AIC of 274.4276 and BIC of 290.0586 respectively. Also for heart rate analysis, with MSE, AIC and BIC of 8.063168, 220.7307 and 236.3617 when $\gamma = 0.5$ respectively.
Conclusion: It was found from this study that Bridge, LASSO and Ridge regression techniques can be used to solve the problem of multicollinearity and address overfitting in model building. Though, the choice of technique to be used depends on the type of data under consideration.
Keywords: *Regression, Ridge, Bridge, LASSO, Regularize, Multicollinearity.*

Cite as: Enwere, K., Nduka, E., & Ogoke, U. Comparative Analysis of Ridge, Bridge and Lasso Regression Models In the Presence of Multicollinearity. *IPS Intelligentsia Multidisciplinary Journal*, *3*(1), 1–8. https://doi.org/10.54117/iimj.v3i1.5.

1. Introduction

In many field of study, Regression is a statistical method that is used to determine the strength and the character of the relationship that exist between a dependent variable and a set of independent variables. Such relationships allows researchers to examine the effects variables have on each other, while simultaneously controlling the effects that other variables in the model may also have.

In recent studies, regression has become a powerful machine learning technique used to make predictions, and hence, modelling in regression analysis has been used to find out to what extent a response (dependent variable) can be predicted by the independent variables (explanatory variables). Again, due to the increasing availability of data in recent years, regression techniques have been employed in various applications to produce flexible solutions to problems in various settings (Weeraratne, 2016).

However, in estimating parameters of linear regression using the least square method, one of the assumptions made, is that the independent variables, say W_i's are not linearly correlated. When this assumption is violated and there is correlation among the independent variables we say that their exist collinearity between the independent variables. Thus, collinearity in regression analysis refers to the event of two (or multiple) covariates being strongly linearly related. That is, one or more than one explanatory variable is determined by other variable. Consequently, multicollinearity becomes a problem that arises in multiple regression analysis when there is a strong correlation or relationship between two or more explanatory variables which results in inaccurate estimates of the regression coefficients, inflation of the standard errors of the regression coefficients, deflating the partial t-tests for the regression coefficients, giving false, nonsignificant, p-values, and degrading the predictability of the model (Herawati *et al*. 2018).

Multicollinearity can cause values of least squares estimators to be unstable i.e. subject to change with slight variation in the data. It also makes it difficult to establish the effects of each explanatory variable on the response variable. (Fox, 2015). Though, it is not possible to eliminate multicollinearity completely but the degree of multicollinearity can be reduced by adopting regularized regression techniques such as ridge regression, LASSO regression, etc. (Ranjit, 2006) [4]. Moreover, this study will explore Ridge regression, Bridge regression and Least Absolute Shrinkage and Selection Operator (LASSO) regression which performs best as a method for handling multicollinearity problem in multiple regression analysis.

Consequently, in order to estimate parameters in a regression analysis, consider the case of a **Simple linear regression** which allows us to look at the linear relationship between a dependent variable and an independent variable.

The model is given as,

$$v = \delta_0 + \delta_1 w_i + \varepsilon_i \qquad i = 1, 2, 3, \dots, n \tag{1}$$

Where v is the dependent variable, w is the independent variable, δ_0 is the y-intercept and δ_1 is the regression coefficient representing the change in v with respect to the change in w, also called the slope.

And ε is the residual error which is the difference between the regression line (which represents our regression prediction) and the actual observation. The predictions made by the "best" regression line are indicated by $\hat{v}$. Thus, $\varepsilon = v - \hat{v}$

Nevertheless, under the assumption that v_i ~ normal, independent and identical distribution ($\delta_0 + \delta_1 w_i$), the Ordinary Least Square Estimation of δ_0, δ_1 and σ^2 are given as follows:

$$\hat{\delta}_0 = \bar{v} - \hat{\delta}_1 \bar{w} \quad (2)$$

$$\hat{\delta}_1 = \frac{\sum_{i=1}^{n}(v_i - \bar{v})(w_i - \bar{w})}{\sum_{i=1}^{n}(w_i - \bar{w})^2} \quad \text{for } \sum_{i=1}^{n}(w_i - \bar{w})^2 \neq 0$$

$$\equiv \frac{n\sum_{i=1}^{n} w_i v_i - (\sum_{i=1}^{n} w_i)(\sum_{i=1}^{n} v_i)}{n\sum w_i^2 - (\sum w_i)^2} \quad (3)$$

$$\sigma^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(v_i - \hat{\delta}_0 - \hat{\delta}_1 w_i\right)^2 \quad (4)$$

Where σ^2 is the variance and the residual error $\varepsilon^2 = \left(v_i - \hat{\delta}_0 - \hat{\delta}_1 w_i\right)^2$

Nevertheless, when there is more than one independent variable, the regression becomes a **Multiple Linear Regression**. Thus, a multiple linear regression model and its estimation using Ordinary Least Square method allows one to estimate the relation between a dependent variable and a set of explanatory variables.

Thus, multiple linear regression is given as,

$$v = \delta_0 + \delta_1 w_1 + \delta_2 w_2 + \ldots + \delta_n w_n + \varepsilon \quad (5)$$

Where w_1, w_2, w_n are the independent variables, δ_0 is the y-intercept, all other δ_n's could represent other parameters.
OLS estimation for two predicator variables δ_1 and δ_2 from a given dataset, are given by:

$$\delta_0 = \hat{v} - \delta_1 \bar{w}_1 - \delta_2 \bar{w}_2 \quad (6)$$

$$\delta_1 = \frac{(\sum w_2^2)(\sum w_1 v) - (\sum w_1 w_2)(\sum w_2 v)}{(\sum w_1^2)(\sum w_2^2) - (\sum w_1 w_2)^2} \quad (7)$$

$$\delta_2 = \frac{(\sum w_1^2)(\sum w_2 v) - (\sum w_1 w_2)(\sum w_1 v)}{(\sum w_1^2)(\sum w_2^2) - (\sum w_1 w_2)^2} \quad (8)$$

For multiple regression, we minimize sum of square error (SSE) to obtain the estimates of δ's.

Thus SSE is given by,

$$\varepsilon^2 = \sum_{i=1}^{n}\left(v_i - \hat{\delta}_0 - \sum_{j=1}^{p}\hat{\delta}_j w_{ij}\right)^2 \quad (9)$$

When SSE is small it implies low variance and when large – high variance of estimates. SSE measures the discrepancies between the data and the estimation model.

Therefore, the above estimates of regression parameters are used in regression analysis for model fitting and prediction of future events. Thus, the understanding of the behavior of explanatory variables from best fitted regression models can help manage and improve the predictability and forecasting of the response variables, since better fitted models can help increase the accuracy of predictions. However, this study focuses on predicting body size and heart rate from known determining factors, using the Ridge, Bridge and LASSO regression techniques.

The understanding of LASSO, Bridge and Ridge regression techniques informs how best to solve the problem of multicollinearity and address overfitting in model building. Also, the use of better fitted models can help interpret analysis on data sets and improve performance of estimates of regression coefficients. It is interesting to know that, these three types of regression models to be considered makes use of a regularization technique to solve the problem of multicollinearity.

1.1 Regularization

Regularization is a technique of adding a penalty to certain models which shrinks the coefficient estimates of the model towards zero in order to minimize error. The regularization techniques often allow reducing the estimated variance at the cost of introducing a small bias, as a result, the prediction accuracy increases. (Melkumovaa and Shatskikhb, 2017).

1.2 Fitted Models

Fitted regression models can be described as: Under-fitting, Overfitting or Best fitting. Nevertheless, Training data points as well as Test data points are used in predicting the best fit line. *Under fitting* occurs as a result of high bias and high variance from the fitted line, also both the training data and test data have low accuracy. *Overfitting* occurs as a result of low bias and high variance from the fitted line, but the training data have high accuracy and test data have low accuracy. However, *Best fitting* occurs as a result of low bias and low variance from the fitted line, which is the case for a generalized model. Note that *Bias* is the error of the training data and *Variance* is the error of the test data. However, in analytical work, the comparison of data, or sets of data, is important to quantify accuracy (bias) and precision.

1.3 Literature Review

Modern Regression Methods such as the Ridge, LASSO and Bridge regressions makes use of regularization techniques in order to deal with Severe Multicollinearity (Herawati *et al*. 2018), that is present in a data set. These modern regression models thus, extend the OLS regression. In overall, the advantages of such regression techniques is that they can reduce the variance by paying the price of an increasing bias. This can improve the prediction accuracy of a model (Frank and Matthias 2019). On like the Ridge regression that only solve the problem of collinearity in the data set (Dorugade 2014), the LASSO simultaneously estimates and selects the coefficients of a given model (Ehsanes *et al*. 2019). On the other hand, the bridge estimators can be used to distinguish between covariates whose coefficients are zero and covariates whose coefficients are nonzero, which is a useful alternative to the existing methods for variable selection and parameter estimation (Jian *et al*. 2006).

Now consider some previous research works carried out on different regression models, in model comparison, prediction and handling multicollinearity.

Abhishek (2021) stated in his work that Supervised Learning is a prominent task of machine learning which maps inputs to corresponding outputs. And that Regression is one such supervised learning technique that models a relationship between independent and dependent variables. His study concluded that, the choice of regression algorithm to be used should be based on the type of data being considered, the distribution of the data and the parameters under considerations.

Serkan and Mehmet (2016) presented a study, comparing the regression models explaining the profitability base on financial data. They have made use of and evaluated multiple linear regression and logistic regression. From their study, they concluded that the multiple linear regression model returned an R^2 value of 0.912 and Logistic regression returned an R^2 value of 0.47 and multiple linear regression gave a better performance. They concluded that the optimal model must be selected based on the purpose of the analysis.

Acharya *et al*. (2019) provided a comparative study of regression models to predict graduate admissions. They have compared different regression algorithms such as Linear Regression, Support Vector Regression (SVR), Decision Trees Regression, and Random Forest Regression, for the given profile of the student. To select the best model, they have computed the error functions for these models and compared their performance. They concluded that linear regression performed best on their dataset with a low Mean Squared Error value and a high R^2 value.

Cheolwoo and Young (2011 carried out a research work on Bridge regression. Their study shows that bridge regression adaptively selects the penalty order from data and produces flexible solutions in various settings. The numerical study shows that the proposed bridge estimators are a robust choice in various circumstances compared to other penalized regression methods such as the ridge, lasso, and elastic net and it shows superior performances in comparisons with other existing methods.

Frank and Matthias (2019) in there research paper, surveyed modern regression models that extend OLS regression. They discussed the regularization terms responsible for inducing coefficient shrinkage and variable selection leading to improved performance metrics of modern regression models. A common feature of all extensions of OLS regression and ridge regression is that these models perform variable selection (coefficient shrinkage to zero). This allows to obtain interpretable models because the smaller the number of variables in a model, the easier it is to find plausible explanations. Considering this, the adaptive LASSO has the most satisfying properties because it possesses the oracle property, making it capable to identify only the coefficients that are non-zero in the true model.

Jian *et al*. (2006) studied the asymptotic properties of bridge estimators in sparse, high- dimensional, linear regression models when the number of covariates increases to infinity with the sample size. Their work was particularly interested in the use of bridge estimators to distinguish between covariates whose coefficients are zero and covariates whose coefficients are nonzero. The results show that under appropriate conditions, bridge estimators correctly select covariates with nonzero coefficients with probability converging to one and that the estimators of nonzero coefficients have the same asymptotic distribution that they would have if the zero coefficients were known in advance. Thus, the bridge estimators have an oracle property.

Noora (2020) carried out a research on detecting multicollinearity in regression analysis. She discusses on the three primary techniques for detecting the multicollinearity using the questionnaire survey data on customer satisfaction, which were: the correlation coefficients and the variance inflation factor, and the eigenvalue method. It was observed that the product attractiveness is more rational cause for the customer satisfaction than other predictors. The study also concluded that advanced regression procedures such as principal components regression, weighted regression, and ridge regression method can be used to determine the presence of multicollinearity.

Dorugade (2014) introduced the ridge regression estimator as an alternative to the ordinary least squares estimator in the presence of multicollinearity. In his article, he introduce alternative ordinary and generalized ridge estimators and study their performance by means of simulation techniques where he compared their evaluated estimators. The results indicates that under certain conditions the performance of proposed ridge parameters is better than OLS used in the simulation study. Hence the study concluded that the performance of the proposed estimators is satisfactory over the other estimators in the presence of multicollinearity.

Jamal (2017) research work focused on multicollinearity, reasons and consequences on the reliability of regression models. He concluded that Multicollinearity is a serious problem that should be resolved before starting the process of data modeling. It is highly recommended that all regression analysis assumption should be met as they contribute to accurate conclusions and helps to make inferences on the population.

Nevertheless, it is shown above that regression techniques have gained more popularity in modern times. Studies in regression analysis had been applied in various field of study. For instance, Ridge regression have been applied in Genetic studies (Arashi *et al*. 2021). Due to ''the advancement of technology, analysis of large-scale data of gene expression is feasible and has become very popular in the era of machine learning''. Also in agriculture, Ridge regression have been proposed to predict wheat production (Nasir and Rind 2007) with the use of fertilizers and manure to improve wheat yield. The LASSO have also been used to investigate changes in in weather conditions (Bappa *et al*. 2018) that impact crop yield and food security. Also, in machine learning, combination of regression models (Thomas *et al*. 2015) could be used together to yield a lower total error of prediction, depending on the requirements of the user.

2. Methodology

This study compared the Bridge Regression model, Ridge Regression model and LASSO Regression model in order to determine the best fit regression technique to produce the better performance. The regression analysis was based on prediction on Body Size and Heart Rate. Thus, from the study, the relationship among the response variable Body size (or Heart rate) and its determining factors using regression techniques can help explain the predictability of future occurrences accurately.

For the purpose of this study, two sets of secondary data gotten from the Lulu Briggs Health Center, University of Port Harcourt was used for comparison and data analysis. First, data on Body Size, Weight, Body Fat, Height and Age were used for the study. Secondly, data on Heart Rate, Body Temperature, Blood Pressure, Blood Volume and Cholesterol Level were also used for the study.

The data used for the study where tested for the presence of multicollinearity using VIF respectively, before proceeding to apply Ridge, LASSO and Bridge regression techniques to solve the problem of multicollinearity.

R software was used to perform data analysis on Ridge, LASSO and Bridge regressions. Comparison for best model fit between Ridge, Bridge and LASSO regression models was made using MSE, RMSE, R^2, AIC and BIC from the regression analysis. Scatter plots were employed to show fitted regression models on Actual values against Predicted values.

The R Software package used in analyzing Ridge and LASSO regression is the 'glmnet' package. Whereas, the R Software package used in analyzing the Bridge regression is the 'rbridge' package. The data sets were fitted for Ridge, Bridge and LASSO regression models to determine the best regression technique in handling multicollinearity.

2.1 Ridge Regression

In solving system of simple linear equation using ordinary least square (OLS) which leads to non-linear normal equations, it depends upon a reduction of the residuals to linear form by first order Taylor approximations (Lavenberg, 1944). If the least square procedure performed with these linear approximations, yields new values for the parameters which are not sufficiently close to the initial values, then the neglect of second and higher order terms may invalidate the process and may actually give rise to larger sum of squares of the residuals, which does not corresponds to the initial solution. This failure of the OLS to improve the initial solution has frequently been encountered in various field of study (Lavenberg, 1944). Moreover, when multicollinearity exists between the independent variables, the OLS method becomes ineffective for computing the parameter estimates. Hence, Hoerl A. E. proposes the Ridge Regression estimation in 1962.

The Ridge Regression (RR) becomes a technique for analyzing multiple regression data that suffer from multicollinearity. By adding a degree of bias to the regression estimates, RR reduces the standard errors and obtains more accurate regression coefficients estimation than the OLS.

The ridge regression parameter is estimated by minimizing the sum square of errors which added a constraints on squares that shrink the coefficient close to zero. Ridge regression penalizes those features that have higher slopes that is we add a penalty to the square of the magnitude of coefficients, to reduce the cost function (Abhishek, 2021).

For the multiple regression model,

$$v = \delta_0 + \delta_1 w_1 + \delta_2 w_2 + \delta_3 w_3 + \delta_4 w_4 + \varepsilon \quad (10)$$

Ridge regression minimizes

$$\varepsilon^2 + \lambda({\delta_1}^2 + {\delta_2}^2 + {\delta_3}^2 + {\delta_4}^2) \qquad \lambda > 0 \quad (11)$$

Where λ is the ridge regression penalty that controls the penalization. λ is determined by cross validation and the parameters are scaled by their measurements.

Hence, Ridge regression model can be stated as adding a regularization term to the multiple regression model

$$v = \delta_0 + \delta_1 w_1 + \delta_2 w_2 + \delta_3 w_3 + \delta_4 w_4 + \varepsilon + \lambda\sum({\delta_1}^2 + {\delta_2}^2 + {\delta_3}^2 + {\delta_4}^2) \quad \lambda > 0, \quad \sum_{i=1}^{p} {\delta_i}^2 \le t \quad (12)$$

where $\lambda\sum({\delta_1}^2 + {\delta_2}^2 + {\delta_3}^2 + {\delta_4}^2)$ is the regularization term. This regularization term adds a penalty to the multiple regression model which minimizes the SSE. Thus SSE is expressed as,

$$\varepsilon^2 = (v - \hat{v})^2 + \lambda\sum({\delta_1}^2 + {\delta_2}^2 + {\delta_3}^2 + {\delta_4}^2) \quad \lambda > 0 \quad (13)$$

where λ is the penalty, $\sum_{i=1}^{p} {\delta_i}^2$ is the slopes or magnitude of the coefficients, t controls the amount of shrinkage. w_1, w_2, w_3 and w_4 are independent variables representing weight, body fat, height and age (or body temperature, blood pressure, blood volume and cholesterol level) used to predict the response variable v represented by body size (or heart rate).

Hence, Ridge regression helps reduce variance by making the predictions less sensitive to training data. This is done by adding the ridge regression penalty to the parameter that must be minimized. This penalty term penalizes the size of coefficient – it reduces the coefficients toward zero, but never all the way to zero. Ridge regression penalty would have all the parameters squared, except for the y-intercept. Moreover, Ridge regression is better when there are many features that have importance.

2.2 LASSO Regression
Robert Tibshirani in 1996 proposes a new method for estimation in linear models. Since the Ridge regression is a continuous process that only shrinks coefficients but it does not set any coefficients to 0 and hence does not give an easily interpretable model. Thus the new technique - Least Absolute Shrinkage and Selection Operator (LASSO), shrinks some coefficients and sets others to 0, and hence tries to retain the good features of both subset selection and ridge regression. The LASSO solves the l_1 - penalized regression problem of finding $\delta = \{\delta_j\}$ to minimize $\sum_{i=1}^{n}\left(v_i - \sum_j w_{ij}\delta_j\right)^2 + \lambda\sum_{j=1}^{p}|\delta_j|$

This is equivalent to minimizing the sum of squares with a constraint of the form $\sum|\delta_j| \le t$. It is similar to ridge regression, which has constraint $\sum_j \delta^2{}_j \le t$. If we consider a more general penalty of the form $\left(\sum_{j=1}^{p}\delta^{\gamma}{}_j\right)^{\frac{1}{\gamma}}$, then the lasso uses $\gamma = 1$ and ridge regression has $\gamma = 2$ (Tibshirani, 2011).

Techniques, such as Least Absolute Shrinkage and Selection Operator (LASSO) is also very common to overcome the problem of multicollinearity. LASSO regression, unlike Ridge regression, results in a model where some coefficient estimates are exactly equal to zero when λ (regularization parameter) is large. That is, LASSO regression will keep the correlated feature with higher coefficients but keep out lower coefficients that are nearly zero. In other words, the LASSO regularization additionally performs variable selection which makes the model easier to interpret (Melkumovaa and Shatskikhb, 2017).

Nevertheless, the aim of using Lasso regression is somewhat similar to that of Ridge regression that is to reduce overfitting. Additionally, Lasso regression also serves the purpose of feature selection. Slope values that tend to zero will be removed meaning those feature that are removed are not important for predicting the line of best fit, since they are insignificant.
For the multiple regression model,

$$v = \delta_0 + \delta_1 w_1 + \delta_2 w_2 + \delta_3 w_3 + \delta_3 w_4 + \varepsilon \quad (14)$$

LASSO regression minimizes

$$\varepsilon^2 + \lambda\sum|\delta_1| + |\delta_2| + |\delta_3| + |\delta_4| \qquad \lambda > 0 \quad (15)$$

LASSO regression model can be stated as adding a regularization term to the multiple regression model

$$v = \delta_0 + \delta_1 w_1 + \delta_2 w_2 + \delta_3 w_3 + \delta_4 w_4 + \varepsilon + \lambda\sum|\delta_1| + |\delta_2| + |\delta_3| + |\delta_4| \qquad \lambda > 0, \quad \sum_{k=1}^{q}|\delta_k| \le t \quad (16)$$

where $\lambda\sum|\delta_1| + |\delta_2| + |\delta_3| + |\delta_4|$ is the regularization term. This regularization term adds a penalty to the multiple regression model which minimizes the SSE. Thus SSE is expressed as,

$$\varepsilon^2 = (v - \hat{v})^2 + \lambda\sum|\delta_1| + |\delta_2| + |\delta_3| + |\delta_4| \qquad \lambda > 0 \quad (17)$$

Where t is the quality that controls the amount of shrinkage in the estimation of coefficients of LASSO with $t \ge 0$

LASSO regression will be better when there are a few important features in the model because more information is lost, since unimportant features are completely removed. LASSO regression can be applied in financial analysis which is used for variable selection, evaluation and interpretation of financial data with other information in investment and financing decision-making process (Serkan and Mehmet, 2016).

2.3 Bridge Regression
Frank and Friedman (1993) introduced the Bridge Regression which minimizes the residual sum of squares, subject to a constraint $\sum|\beta_j|^{\gamma} \le t$ with $\gamma \ge 0$. Overall, bridge regression achieves small MSE and performs well in estimation and predictions compared to the LASSO and the Ridge for linear regression models in general.

Bridge regression estimator generalizes both ridge regression and LASSO regression estimators, because it minimizes the SSE with a λ penalty. Thus, the bridge regression method provides a way of combining parameter estimation and variable selection in a single minimization problem. The bridge estimator can effectively identify large and moderate nonzero covariate effects and zero covariate effects. However, it penalizes small coefficient values excessively.

Moreover, when it comes to high-dimensional data, the bridge estimator with $0 < \gamma \le 1$ becomes a useful alternative to the existing methods for variable selection and parameter estimation. For the case $0 < \gamma \le 1$, the resulting estimator will be nearly unbiased for large values of the unknown parameter δ. Consequently, when $\gamma > 1$, the bridge regression method shrinks the regression coefficients, but does not provide variable selection. For when $0 < \gamma < 1$ the objective function is non-convex and the singularity at $\beta = 0$ makes it difficult to minimize the penalized objective function. Thus for larger values of γ, the shrinkage increases with the magnitude of the regression parameters being estimated (Olcay, 2015).

Hence, for the multiple regression model,

$$v = \delta_0 + \delta_1 w_1 + \delta_2 w_2 + \delta_3 w_3 + \delta_3 w_4 + \varepsilon \quad (18)$$

Bridge regression minimizes

$$\varepsilon^2 + \lambda(\delta_1{}^{\gamma} + \delta_2{}^{\gamma} + \delta_3{}^{\gamma} + \delta_4{}^{\gamma}) \qquad \lambda > 0, \qquad \gamma > 0 \quad (19)$$

Bridge regression model can be stated as adding a regularization term to the multiple regression model

$$v = \delta_0 + \delta_1 w_1 + \delta_2 w_2 + \delta_3 w_3 + \delta_4 w_4 + \varepsilon + \lambda(\delta_1{}^{\gamma} + \delta_2{}^{\gamma} + \delta_3{}^{\gamma} + \delta_4{}^{\gamma}) \qquad \lambda > 0, \quad 0 < \gamma \le 2, \quad \sum_{j=1}^{r}|\delta_j|^{\gamma} \le q \quad (20)$$

where $\lambda(\delta_1{}^{\gamma} + \delta_2{}^{\gamma} + \delta_3{}^{\gamma} + \delta_4{}^{\gamma})$ is the regularization term. This regularization term adds a penalty to the multiple regression model which minimizes the SSE. Thus SSE is expressed as,

$$\varepsilon^2 = (v - \hat{v})^2 + \lambda(\delta_1{}^{\gamma} + \delta_2{}^{\gamma} + \delta_3{}^{\gamma} + \delta_4{}^{\gamma}) \qquad \lambda > 0, \qquad 0 < \gamma < 2 \quad (21)$$

Where q is a positive parameter representing the tuning constant that controls the amount of shrinkage. γ is the shrinkage parameter.

Also, the optimal γ selection will increase the efficiency of the model. Moreover, by setting $\gamma = 0$, $\gamma = 1$ and $\gamma = 2$ in optimization problem we obtain the least-squares regression model, the lasso regression model, and the ridge regression model respectively (Delara *et al*. 2020).

The Bridge estimator correctly identifies zero coefficients with higher probability than the LASSO and Ridge estimators. It performs well in terms of predictive mean square errors. Bridge regression is known to possess many desirable statistical properties such as oracle, sparsity, and unbiasedness. However, a disadvantage of bridge is that it lacks a systematic approach to inference, reducing its flexibility in practical applications.

3. Results and Discussion

3.1 Multiple Regression Analysis Results

From the multiple regression analysis results below for body size, we see that 'weight, body fat and height' have high VIF values of 37.13, 60.20 and 68.38 respectively, indicating the presence of high multicollinearity among these independent variables. Thus, the strong correlation between those independent variables means that they can be predicted by other independent variables in the data set. Similarly, the multiple regression analysis results for heart rate shows that 'blood pressure, blood volume and cholesterol level' have high VIF values of 27.44, 36.36 and 46.28 respectively, indicating the presence of high multicollinearity among those explanatory variables.

Regression Equation for Body Size

Body Size = 68.6 + 0.234 Weight + 1.221 Body Fat + 0.018 Height - 1.015 Age

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	68.6	47.0	1.46	0.148	
Weight	0.234	0.145	1.61	0.110	37.13
Body Fat	1.221	0.337	3.62	0.000	60.20
Height	0.018	0.849	0.02	0.984	68.38
Age	-1.015	0.302	-3.36	0.001	4.18

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.78317	91.94%	91.60%	90.74%

Regression Equation for Heart Rate

Heart Rate = 293.0 + 0.268 Body Temperature - 1.056 Blood Pressure - 1.442 Blood Volume - 0.2821 Cholesterol Level

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	293.0	48.7	6.02	0.000	
Body Temperature	0.268	0.276	0.97	0.333	1.28
Blood Pressure	-1.056	0.394	-2.68	0.009	27.44
Blood Volume	-1.442	0.342	-4.21	0.000	36.36
Cholesterol Level	-0.2821	0.0939	-3.00	0.003	46.28

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.91333	54.44%	52.52%	46.29%

From the multiple analysis results above, it is known that there is presence of multicollinearity among the explanatory variables. Therefore, we then proceed to use Ridge, LASSO and Bridge regression analysis techniques to solve the problem of multicollinearity among these data sets.

3.2 Data Analysis and Results for Body Size

The Table 1 below shows the comparative analysis results of Ridge regression, LASSO regression and Bridge regression for Body Size, in testing for the best model fit and in handling the problem of multicollinearity.

3.2.1 Scatter Plot for Fitted Regression models on Body Size

Figure 1 shows that the data points are linear and close to the line of best fit indicating less variation, except for few data points which are far away from the line of best fit which may be outliers. The data points which are closer to the best fit line indicates a strong positive correlation or relationship between the actual values and predicted values. Figure 2 also indicates a strong positive correlation between the actual values and the predicted values. The data points are close to the best fit line and linear.

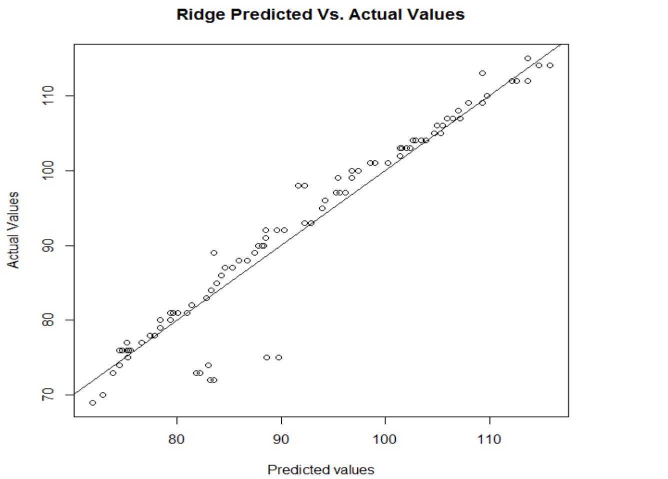

Figure 1: Ridge regression plot for Actual values against Predicted values

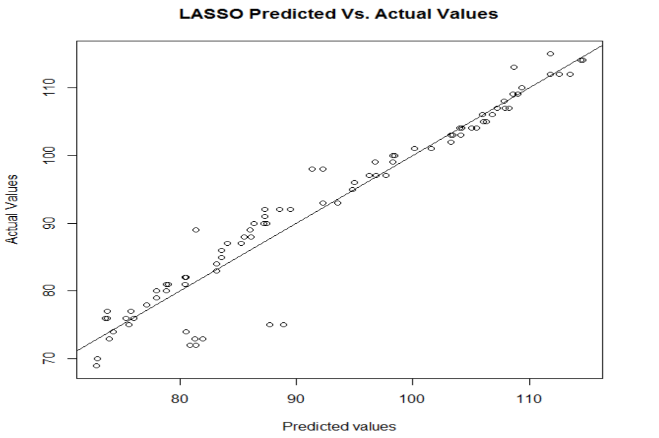

Figure 2: LASSO regression plot for Actual values against Predicted values

Figure 3 for Bridge regression when $\gamma = 1$, shows that the data points are close to the best fit line, but with slight variation in some data points which are outliers that stands far from the line of best fit. This plot indicates a strong positive correlation between the actual data points and the predicted data points. This similar to figure 4 when $\gamma = 2$.

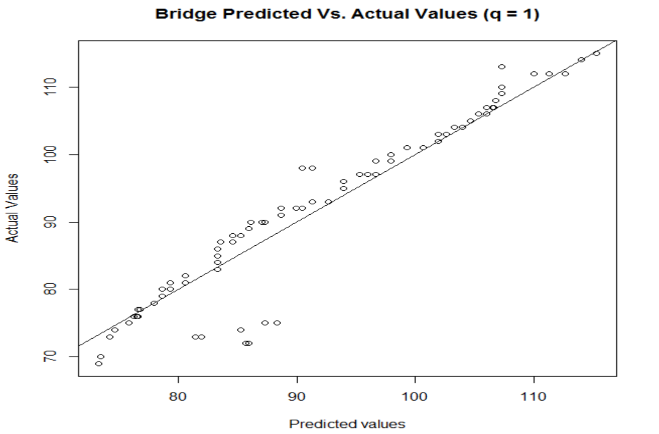

Figure 3: Bridge regression plot for Actual values against Predicted values when $\gamma = 1$

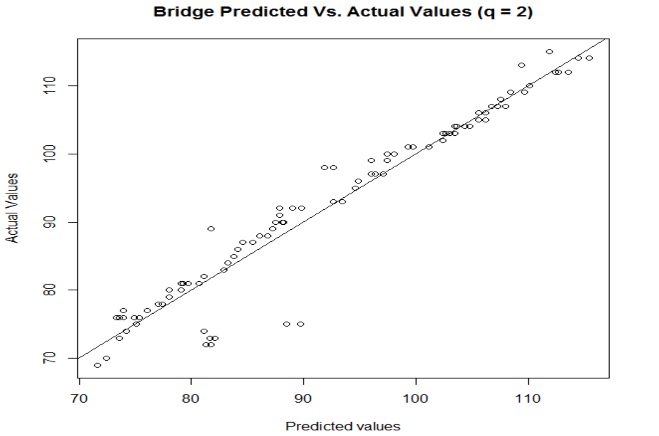

Figure 4: Bridge regression plot for Actual values against Predicted values when $\gamma = 2$.

3.3 Data Analysis and Results for Heart Rate

The Table 2 below shows the comparative analysis results of Ridge regression, LASSO regression and Bridge regression for Heart Rate, in testing for the best model fit and in handling the problem of multicollinearity.

Table 1: Comparative Results for Body Size, Weight, Body Fat, Height and Age.

Analysis Criterion	Regression Techniques					
			Bridge			
	Ridge	LASSO	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$
Best λ value	1.238038	0.4770814	151.1523	70.58116	6.693683	2.215357
MSE	14.83657	15.61026	15.39428	14.99242	13.79458	13.88525
RMSE	3.851827	3.950983	3.923555	3.872004	3.714106	3.726291
R^2	0.9120308	0.9074434	0.908724	0.9111067	0.918209	0.9176714
R^2_{adj}	0.9083268	0.9035463	0.9048808	0.9073639	0.9147651	0.9142049
AIC	281.7095	286.7929	285.3996	282.7545	274.4276	275.0827
BIC	297.3405	302.4239	301.0306	298.3855	290.0586	290.7137
VIF	11.36762	10.8042	10.95578	11.24945	12.22628	12.14645

Table 2: Comparative Results for Heart Rate, Body Temperature, Blood Pressure, Blood Volume and Cholesterol Level.

Analysis Criterion	Regression Techniques					
			Bridge			
	Ridge	LASSO	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$
Best λ value	0.2849412	0.01556428	0.5	151.1523	0.05	0.2413889
MSE	9.241047	8.100754	8.063168	10.14433	8.063249	8.089838
RMSE	3.039909	2.846182	2.839572	3.185017	2.839586	2.844264
R^2	0.4778007	0.5422372	0.5443611	0.4267572	0.5443566	0.542854
R^2_{adj}	0.4558134	0.522963	0.5251763	0.4026206	0.5251716	0.5236058
AIC	234.3655	221.1957	220.7307	243.6915	220.7317	221.0609
BIC	249.9965	236.8267	236.3617	259.3226	236.3627	236.6919
VIF	1.914978	2.184537	2.19472	1.744461	2.194699	2.187485

3.3.1 Scatter Plot for Fitted Regression models on Heart Rate

Figure 5 shows a weak, nonlinear, positive correlation between the actual values and the predicted values. More data points stands far from the fitted line which may be outliers, thus indicating some variation in the model. In Figure 6, we see a moderate, positive, correlation with two clusters of actual data points against the predicted data points. We also see the presence of outliers in the plot.

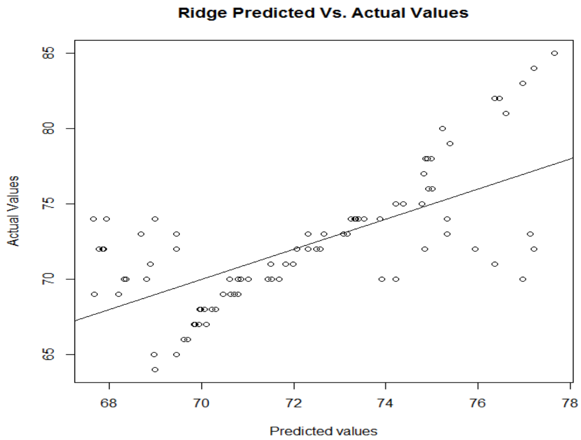

Figure 5: Ridge regression plot for Actual values against Predicted values

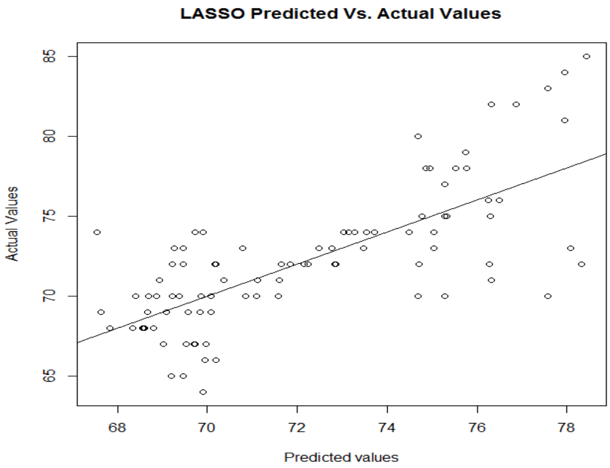

Figure 6: LASSO regression plot for Actual values against Predicted values

Figure 7 is similar to the Bridge regression plot in Figure 5 when $\gamma = 1$. The plot indicates a weak, nonlinear, positive correlation between the actual data points and the predicted data values. Again, Figure 8 is similar to the LASSO plot in Figure 6, when the Bridge shrinkage parameter $\gamma = 2$. We see from the plot two clusters of the data points and a moderate, positive correlation between the actual data points and the predicted data points.

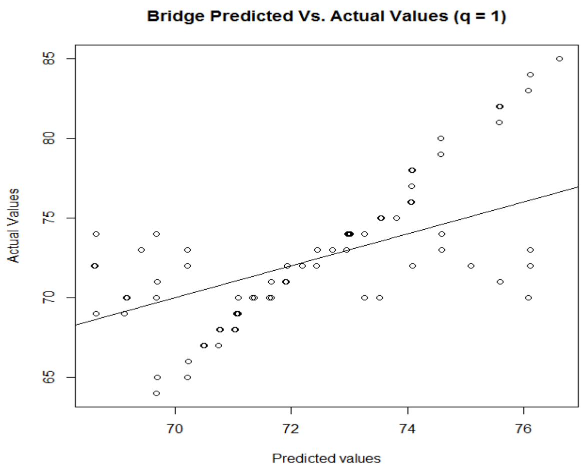

Figure 7: Bridge regression plot for Actual values against Predicted values when $\gamma = 1$

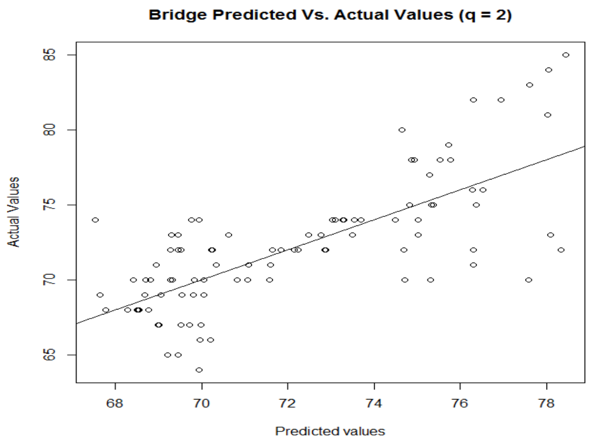

Figure 8: Bridge regression plot for Actual values against Predicted values when $\gamma = 2$

Discussion of Findings

Based on our numerical results, from Table 1, when comparing the best regression technique to solve the problem of multicollinearity between Ridge, LASSO, and Bridge regressions, we see that none of the Regression Techniques handled the problem of multicollinearity, even though the degree of multicollinearity in the data set reduces (Ranjit, 2006; Jong, 2019) with VIF values of 11.36762 for Ridge, 10.8042 for LASSO, and Bridge which are 10.95578, 11.24945, 12.22628 and 12.14645 respectively. This is in contrast with some studies done on regularized techniques, (Herawati *et al*. 2018; Cheolwoo and Young, 2011) which result to further studies or modifications on existing regularized regression techniques which was the case of (Frank and Matthias, 2019; Olcay, 2015; Himel and Nengjun, 2018).

Also in our comparison for best model fit, Bridge regression performed better with MSE of 13.79458 when $\gamma = 1.5$, AIC of 274.4276 and BIC of 290.0586 respectively. This is in line with Jian *et al*. (2006) that the Bridge estimator performs well in terms of predictive mean square errors. Moreover, about 92% of the variation is explained in the model. Thus producing more accurate results for predicting Body Size using Weight, Body Fat, Height and Age. Ridge regression came second with MSE of 14.83657, AIC of 281.7095 and BIC of 297.3405 respectively. Also, about 91% of variation is explained in our ridge model. From the R^2 for LASSO regression, we also have about 92% of our variation explained in the model with MSE of 15.61026.

Consequently, from Table 2, in comparing Ridge, LASSO, and Bridge regression techniques in handling the problem of multicollinearity, we see that all the regularized regression techniques handled the problem of multicollinearity, which is why studies in regularized regressions have become popular in recent years (Dorugade, 2014; Jian *et al*., 2006; Sarojamma and AnilKumar, 2018; Cheolwoo and Young, 2011). We see that the Bridge regression technique performed better with VIF of 1.744461 when $\gamma = 1$ respectively. Second to it was the Ridge regression with VIF of 1.914978, and lastly the LASSO regression with VIF of 2.184537 respectively. The findings of all three techniques used to solve multicollinearity where quite similar to Gursev (2020).

In comparing best model fit, Bridge regression performed better with MSE, AIC and BIC of 8.063168, 220.7307 and 236.3617 when $\gamma = 0.5$ respectively. Next to it is the Ridge regression with MSE, AIC and BIC of 9.241047, 234.3655 and 249.9965 respectively. R^2 for Bridge regression, shows that about 54% of our variation was explained in the model. Thus Bridge produced more accurate results for predicting Heart Rate using Body Temperature, Blood Pressure, Blood Volume and Cholesterol Level.

Nevertheless, based on the two sets of data used in this study, we notice that from Table 1, we see that none of regularized regression techniques solved the problem of multicollinearity which is in contrast to Table 2 where all three regression techniques successfully handled the problem of multicollinearity. Meaning that the choice of regression technique to be used also depends on the type of data to be considered and the parameters under consideration which is also similar to the findings of Abhishek, (2021). However, in both data sets that where analyzed, the numerical results show that Bridge regression wins our comparison, thus showing superior performance to Ridge regression and LASSO regression which was a similar conclusion from the study done by Cheolwoo and Young (2011).

Conclusion

From the study, we have been able to show the best regression technique that can be used in handling multicollinearity that exist in multiple regression analysis. Though the choice of technique to be used should be based on the type of data being considered. Consequently, Bridge regression wins our comparison for being a better technique in handling the problem of multicollinearity. When comparing which regression model produced the best model fit for better prediction and accurate results, Bridge regression performed better for both sets of data used.

Granted, access to the medical records for the two sets of data used for the study was granted after approval and ethical clearance from the relevant authority of the Health Center. The relationship among Body size or Heart rate and their determining factors using regression techniques can help explain the predictability of future occurrences accurately.

The understanding of Ridge regression, LASSO regression and Bridge regression techniques can inform how best they can be used by Statisticians and other Researchers to solve the problem of multicollinearity and in addressing over fitting in model building. Nevertheless, the use of Bridge regression, Ridge regression and LASSO regression can help interpret analysis on data sets and improve performance of estimates of regression coefficients and predict possible response behaviors.

References

Abhishek, T. (2021). Comparative Assessment of Regression Models Based On Model Evaluation Metrics. *International Research Journal of Engineering and Technology*, *9*(8), 853-860.

Acharya, M., Armaan, A., & Antony, A. (2019, February 1). *A Comparison of Regression Models for Prediction of Graduate Admissions*. [Paper presentation]. Second International Conference on Computational Intelligence in Data Science, Mysuru, India. DOI: 10.1109/ICCIDS.2019.8862140

Arashi, M., Roozbeh, M., Hamzah, A., & Gasparini, M. (2021). Ridge regression and its Applications in Genetic studies. *PLoS One, 16*(4).

Bappa, D., Bhakti, N., Viswanatha, K., & Paramesh, V. (2018). Evaluation of Multiple Linear, Neural Network and Penalized Regression Models for Prediction of Rice Yield Based on Weather Parameters for West Coast of India. *International Journal of Biometeorology*, *62*(10), 1809–1822.

Cheolwoo, P., & Young, Y. (2011). *Bridge regression: adaptivity and group selection.* University of Georgia. DOI: 10.1016/J.JSPI.2011.05.004

Delara K., Mohammad, R., & Alireza, N. (2020). *A Generalized Bridge Regression in Fuzzy Environment and Its Numerical Solution by a Capable Recurrent Neural Network.* Shah Rood University of Technology. https://doi.org/10.1155/2020/8838040

Dorugade, V. (2014). New Ridge Parameters for Ridge Regression. *Journal of the Association of Arab Universities for Basic and Applied Sciences*, *15*(1), 94–99. DOI: 10.1016/ j.jaubas.2013.03.005

Ehsanes, S., Golam, K., & Florence, G. (2019). Comparative Study of LASSO, Ridge Regression, Preliminary Test and Stein-type Estimators for the

Sparse Gaussian Regression Model. *Statistics, Optimization and Information Computing, 7,* 626–641. DOI: 10.19139/soic-2310-5070-713

Fox, J. (2015). *Applied regression analysis and generalized linear models*. Sage publications.

Frank, E., & Friedman, J. (1993). A Statistical View of Some Chemometrics Regression Tools,*Technometrics*, *35,* 109–148.

Frank, E., & Matthias, D. (2019). High-Dimensional LASSO-Based Computational Regression Models: Regularization, Shrinkage, and Selection. *Machine Learning & Knowledge Extraction, 1*(1), 359-383. DOI: *10.3390/make1010021*

Gursev, P. (2020). *Comparison of Regression Analysis Algorithms*. https://gursev- pirge.medium.com/comparison-of-regression-analysis-algorithms-db710b6d7528.

Herawati, N., Nisa, K., Setiawan, E., & Nusyirwan, T. (2018). Regularized Multiple Regression Methods to Deal with Severe Multicollinearity. *International Journal of Statistics and Applications, 8*(4), 167-172. DOI: 10.5923/j.statistics.20180804.02

Himel, M., & Nengjun, Y. (2018). Bayesian Bridge Regression. *Journal of Applied Statistics, 45*(6), 988-1008.

Hoerl, A. (1962). Application of Ridge Analysis to Regression Problems. *Chemical Engineering Progress, 58*, 54-59.

Jamal, D. (2017, August 8-9). Journal of Physics: Conference Series [Conference Session]. 4th International Conference on Mathematical Applications in Engineering, Kuala Lumpur, Malaysia. DOI: 10.1088/1742-6596/949/1/012009

Jian, H., Joel, H., & Shuangge, M. (2006). *Asymptotic Properties of Bridge Estimators in Sparse High-Dimensional Regression Models.* University of Iowa. DOI: 10.1214/009053607000000875

Jong, H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology, 72*(6), 558-569.

Levenberg, K. (1944). A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quart of Applied Mathematics*, *2*, 164-168.

Melkumovaa, L., & Shatskikhb, S. (2017). *Comparing Ridge and LASSO estimators for Data Analysis*. Elsevier Inc. DOI: 10.1016/j.proeng.2017.09.615

Noora, S. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics, 8*(2), 39-42. DOI: 10.12691/ajams-8-2-1

Olcay, A. (2015). Penalized MM Regression Estimation with L_γ Penalty: A Robust Version of Bridge Regression. *A Journal of Theoretical and Applied Statistics*. *50*(6), 1236-1260. DOI: 10.1080/02331888.2016.1159682

Ranjit, P. (2006). Multicollinearity: Causes, Effects and Remedies. *Indian Agricultural Statistics Research Institute*, *1*(1), 58-65.

Robert, T. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, *58*(1), *267*-288.

Sarojamma, B., & AnilKumar, K. (2018). A Study on Comparison among Ridge, Lasso, and Elastic Net Regressions. *International Journal of Creative Research Thoughts, 6*(1), 600-604.

Serkan, E. & Mehmet, I. (2016). A Research on Comparison of Regression Models Explaining the Profitability Base on Financial Data. *International Journal of Business & Management*, *4*(10), 470 475.

Thomas, F., Marie, V., Marco, L., Carmosino, M., Darby, D., Sridhar, M., Elly, A., Katherine, H., Samuel, M. (2015). A Study of Machine Learning Regression Methods for Major Elemental Analysis of Rocks Using Laser-Induced Breakdown Spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 107, 1–10.

Tibshirani, R. (2011). Regression Shrinkage and Selection via the Lasso: A Retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*(3), 273-282.

Weeraratne, C. (2016). The Effects of Multicollinearity in Ordinary Least Squares (OLS) Estimation. *International Journal of Advance Research and Innovative Ideas in Education, 2*(4), 1002-1008.

Thank you for publishing with us.

References to Equations

Equation 1, Linear regression model.
Equation 2, Formula for estimation of the parameter δ_0 for one predicator variable.
Equation 3, Formula for estimation of the parameter δ_1 for one predicator variable.
Equation 4, Variance for Linear regression model.
Equation 5, Multiple regression model.
Equation 6, Formula for estimation of the parameter δ_0 for two predicator variables.
Equation 7, Formula for estimation of the parameter δ_1 for two predicator variables.
Equation 8, Formula for estimation of the parameter δ_2 for two predicator variables.
Equation 9, Sum of Square Error.
Equation 10, Multiple regression model for four predicator variables.
Equation 11, Ridge regression regularized error term for four predicator variables.
Equation 12, Ridge regression model.
Equation 13, Sum of Squares for Ridge regression.
Equation 14, See equation 10 above.
Equation 15, LASSO regression regularized error term for four predicator variables.
Equation 16, LASSO regression model.
Equation 17, Sum of Squares for LASSO regression.
Equation 18, See equation 10 above.
Equation 19, Bridge regression regularized error term for four predicator variables.
Equation 20, Bridge regression model.
Equation 21, Sum of Squares for Bridge regression.