International Journal of Mathematical Sciences and Optimization: Theory and Applications
11(4), 2025, Pages 46 - 56
https://doi.org/10.5281/zenodo.18986070

Comparing Ordinary Least Squares, Ridge, and Lasso Regression for Multicollinearity Mitigation in Linear Models

F. A. Okolie [1,2], B. O. Fagbemigun [1*], O. E. Taiwo [1], E. O. Adekola [1]

1. Department of Mathematical Sciences, Ajayi Crowther University, Oyo State, Nigeria.
2. Department of Statistics, Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria.
* Corresponding author: bo.fagbemigun@acu.edu.ng[2], fa.okolie@acu.edu.ng, oe.taiwo@acu.edu.ng, adekolaeunice100@gmail.com

Article Info
Received: 18 February 2025 Revised: 23 June 2025
Accepted: 28 June 2025 Available online: 10 December 2025

Abstract

Ordinary Least Squares (OLS) regression provides unbiased estimates but performs poorly when predictor variables are highly correlated, due to increased variance and model instability. This study compares the effectiveness of OLS, Ridge regression, and the Least Absolute Shrinkage and Selection Operator (LASSO) in mitigating multicollinearity and improving predictive accuracy in linear models. Using academic data of University of Ilorin undergraduate students, Nigeria, we evaluated model performance using Root Mean Squared Error (RMSE), Variance Inflation Factors (VIF), and cross-validation. Ridge regression applies an L_2 penalty to shrink coefficients, while LASSO uses an L_1 penalty that also enables variable selection by setting some coefficients to zero. The results show that Ridge regression achieved the best generalization performance with the lowest test RMSE (0.2358), while LASSO provided a more interpretable model through coefficient sparsity. OLS exhibited overfitting and the poorest generalization due to high multicollinearity. The findings highlight the importance of regularization techniques in regression modeling, especially in high-dimensional data environments. This study offers practical guidance on model selection when predictive accuracy and feature interpretability are essential.

Keywords: Linear Modeling, Regression Analysis, Multicollinearity, Ridge Regression, LASSO Regression.
MSC2010: 00A71.

1 Introduction

Regression analysis is a foundational statistical technique for modeling the relationship between a dependent variable and one or more independent variables [1]. Among the various regression approaches, Ordinary Least Squares (OLS) regression is widely used due to its computational simplicity, interpretability, and optimality under classical assumptions. However, its performance significantly deteriorates in the presence of multicollinearity—a condition where two or more predictors are highly correlated [2]. This results in inflated standard errors, unstable coefficient estimates,

https://doi.org/10.5281/zenodo.18986070

and poor predictive performance.

To address these limitations, regularization techniques have been introduced. These methods modify the OLS objective function by adding penalty terms to shrink regression coefficients, thereby reducing model complexity and improving generalization. Two widely used regularization techniques are Ridge Regression, proposed by Hoerl and Kennard [3], which applies an L_2 penalty to the sum of squared coefficients, and the Least Absolute Shrinkage and Selection Operator (LASSO), introduced by Tibshirani [4], which employs an L_1 penalty that not only shrinks coefficients but can also set some to exactly zero, effectively performing variable selection.

These techniques have proven effective in high-dimensional settings and are particularly useful when predictors exhibit strong linear dependence [5]. Ridge regression stabilizes coefficient estimates under multicollinearity, while LASSO enhances interpretability by excluding irrelevant features. Furthermore, hybrid approaches such as the Elastic Net [6] combine both penalties to balance the strengths of Ridge and LASSO.

Despite the growing adoption of regularization methods across many fields, empirical evaluations using real-world educational data remain limited, especially within Nigerian higher institutions. This study addresses this gap by assessing the predictive performance of OLS, Ridge and LASSO regression models using undergraduate academic records from the University of Ilorin. It further evaluates the ability of each method in managing multicollinearity and the extent to which this shapes model performance. This approach also contributes to widening the empirical application of penalized regression within African educational research. Though such techniques are common in areas such as genomics, economics and engineering, they have seen very limited use with real academic datasets from Nigerian universities.

2 Literature Review

Multicollinearity poses a major challenge in multiple linear regression, as it inflates the variances of coefficient estimates and undermines the reliability of statistical inference. In such scenarios, Ordinary Least Squares (OLS) regression becomes unstable and often fails to generalize well on unseen data. To address this issue, penalized regression techniques have gained significant traction, especially in high-dimensional and collinear data environments.

Ridge regression, introduced by Hoerl and Kennard [3], addresses multicollinearity by adding an L_2 penalty term to the loss function. This penalty shrinks the coefficients toward zero but does not force any of them to zero, thus retaining all variables while reducing model variance. Ridge is particularly effective when many predictors contribute small effects to the response variable.

In contrast, the Least Absolute Shrinkage and Selection Operator (LASSO), proposed by Tibshirani [4], incorporates an L_1 penalty on the regression coefficients. Unlike Ridge, LASSO performs both coefficient shrinkage and variable selection by shrinking some coefficients exactly to zero, resulting in a sparser, more interpretable model. This property makes LASSO attractive when there is a need to identify key predictors among many.

To bridge the limitations of Ridge and LASSO, Zou and Hastie [6] developed the Elastic Net, which combines both L_1 and L_2 penalties. The Elastic Net retains the sparsity of LASSO while preserving the grouping effect of Ridge, making it especially useful when predictors are highly correlated or when $p > n$.

Further advancements in regularization techniques include the Adaptive LASSO, introduced by Zou [7], which modifies the LASSO by applying data-dependent weights to each coefficient in the penalty term. This approach improves the consistency of variable selection, especially under conditions of multicollinearity. Another notable method is the Smoothly Clipped Absolute Deviation (SCAD) estimator proposed by Fan and Li [8], which uses a non-convex penalty function to reduce bias in large coefficients and satisfies the oracle property under certain regularity conditions.

From a theoretical standpoint, Bühlmann and Van De Geer [5] provided a comprehensive framework for understanding penalized regression methods in high-dimensional statistics. They demonstrated the asymptotic properties of various estimators and established conditions for consistency and sparsity.

Empirical studies have further confirmed the utility of penalized regression in handling multicollinearity. For example, Hastie, Tibshirani, and Wainwright [9] provide a comprehensive theoretical and practical overview of penalized regression methods, including LASSO, Ridge, Elastic Net, and their extensions, with applications in high-dimensional and multicollinear settings.

Although regression techniques are widely applied, there is limited empirical evidence regarding their effectiveness in modeling educational data within African contexts. To address this shortcoming, the present study evaluates the predictive performance of OLS, Ridge, and LASSO regression models, using academic course scores from undergraduate students at the University of Ilorin, Nigeria. The analysis focuses on how each model addresses multicollinearity and improves generalization performance in real-world educational settings.

3 Methods

3.1 Multiple Linear Regression Model

Multiple linear regression model describes the relationship between a continuous response variable and a set of predictors. It is expressed as:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{3.1}$$

where y_i is the response variable for the i^{th} observation, x_{ij} denotes the value of the j^{th} predictor for observation i, β_0 is the intercept, β_j are the regression coefficients, and ε_i is the random error term assumed to be independently and identically distributed with mean zero and constant variance σ^2.

The various parameter estimation approaches for model (3.1) considered in this study are:

The OLS estimator which minimizes the residual sum of squares (RSS):

$$\hat{\beta}^{\text{OLS}} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2. \tag{3.2}$$

The estimator (3.2) performs well under classical assumptions, but becomes unstable when predictors are highly correlated, leading to inflated variances and overfitting , hence the need for a better Regression estimation approach.

The Ridge regression addresses multicollinearity by introducing an L_2 penalty on the regression coefficients. The ridge estimator minimizes the following objective function:

$$\hat{\beta}^{\text{Ridge}} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}, \tag{3.3}$$

where $\lambda \geq 0$ is a tuning parameter controlling the amount of shrinkage. Ridge shrinks coefficients towards zero but does not set any of them exactly to zero, thus retaining all predictors in the model.

LASSO regression, on the other hand, imposes an L_1 penalty on (3.2), allowing it to perform variable selection by shrinking some coefficients exactly to zero as:

$$\hat{\beta}^{\text{LASSO}} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}. \tag{3.4}$$

This makes LASSO particularly useful when only a subset of predictors are relevant, improving both model interpretability and generalization.

3.2 Model Diagnostics and Evaluation

To detect multicollinearity in the data, the Variance Inflation Factor (VIF) is computed for each predictor. VIF values exceeding 10 suggest severe multicollinearity, which may distort OLS estimates.

Model predictive accuracy is evaluated using the *Root Mean Squared Error (RMSE)*, defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \tag{3.5}$$

where y_i is the observed value, $\hat{y}_i$ is the predicted value from the model, and n is the number of observations. Lower RMSE values indicate better predictive performance.

To select the optimal tuning parameter λ for Ridge and LASSO models, k-fold cross-validation is performed. This procedure ensures that the chosen λ generalizes well to unseen data and avoids overfitting.

4 Results

The dataset used in this study comprises academic records of undergraduate students from the University of Ilorin, Nigeria, spanning four years. A total of 117 students were observed, with each student having scores recorded in 52 core and elective courses taken across four academic years. The target variable is the students' Cumulative Grade Point Average (CGPA), which is modeled as a function of their course performances.

This dataset is well-suited for multicollinearity studies due to the natural correlation that exists among certain course grades (e.g., foundational and advanced versions of similar statistical-STAs, mathematical-MATs or related-GNSs, GSEs, CHMs, CSCs courses). These relationships provide the opportunity to explore how Ridge and LASSO manage redundant information in high-dimensional linear models.

4.1 Multicollinearity Assessment

To evaluate multicollinearity, the heatmap plot was used to visualize the strength and direction of linear relationships between predictors Figure 1 and the Variance Inflation Factor (VIF) was computed for all predictors. A subset of results is presented in Table 1.

IJMSO
International Journal of Mathematical Sciences and Optimization: Theory and Applications
11(4), 2025, Pages 46 - 56
https://doi.org/10.5281/zenodo.18986070

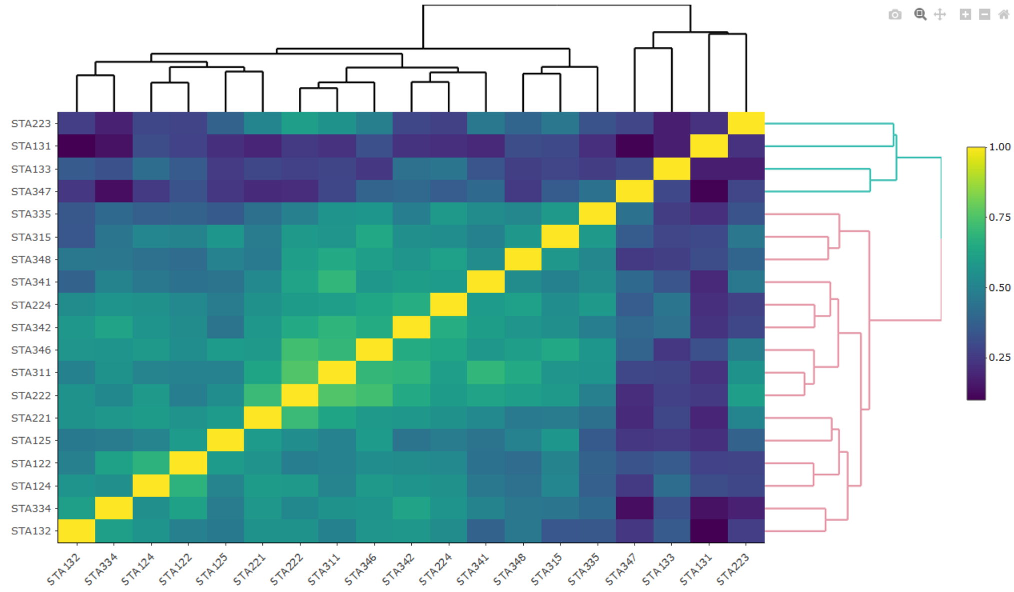


Figure 1: Heatmap showing the relationship between some predictors

The heatmap in Figure 1 reveals strong correlations among certain independent variables, indicating potential multicollinearity. High correlation coefficients suggest that some predictors share significant linear relationships, which can distort regression estimates and affect model interpretability. This further reinforces the need for techniques such as regularization or variable selection to mitigate multicollinearity and improve model reliability.

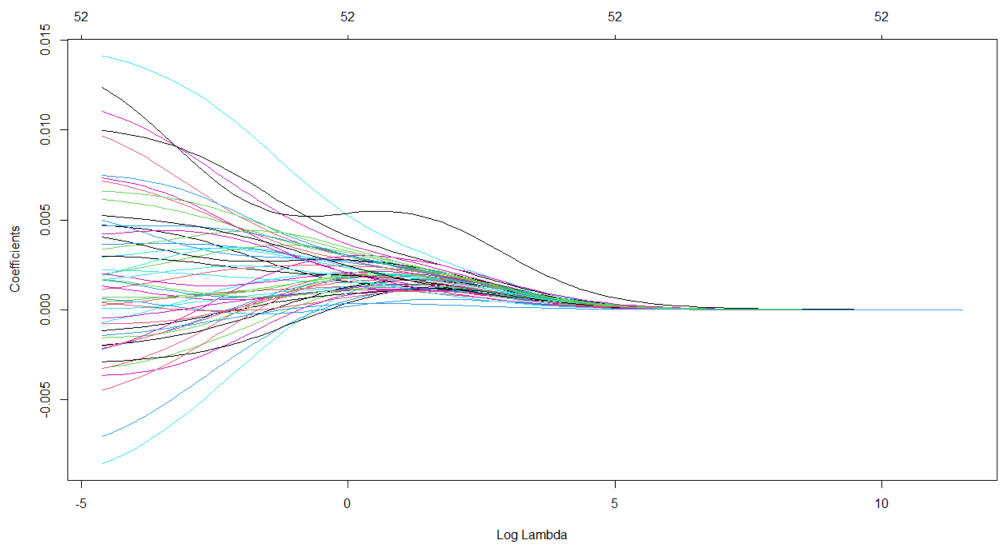
Table 1: Variance inflation factors (VIF) for the data

STA122	**STA124**	**STA125**	**STA131**	**STA132**
6.22	8.15	3.55	3.30	5.46
STA133	**STA221**	**STA222**	**STA223**	**STA224**
5.11	7.63	13.34	6.15	13.10
STA311	**STA315**	**STA334**	**STA335**	**STA341**
8.69	6.60	9.23	7.13	5.45
STA342	**STA346**	**STA347**	**STA348**	**STA354**
11.51	8.01	3.78	5.54	5.41
STA422	**STA424**	**STA435**	**STA436**	**STA437**
18.79	10.12	16.15	4.32	7.08
STA438	**STA441**	**STA443**	**STA444**	**STA498**
8.21	10.63	3.91	3.76	4.80
STA499	**MAT101**	**MAT102**	**MAT103**	**MAT104**
4.43	4.29	7.91	11.57	5.47
MAT201	**MAT202**	**MAT203**	**MAT204**	**MAT312**
5.15	2.89	5.57	8.34	4.18
GNS111	**GNS112**	**GNS211**	**GNS212**	**GNS311**
4.81	3.82	5.61	3.86	4.05
CSC101	**CSC102**	**CSC201**	**CSC203**	**CHM101**
3.37	4.18	9.36	3.44	3.27
CHM102	**GSE301**			
2.54	2.50			

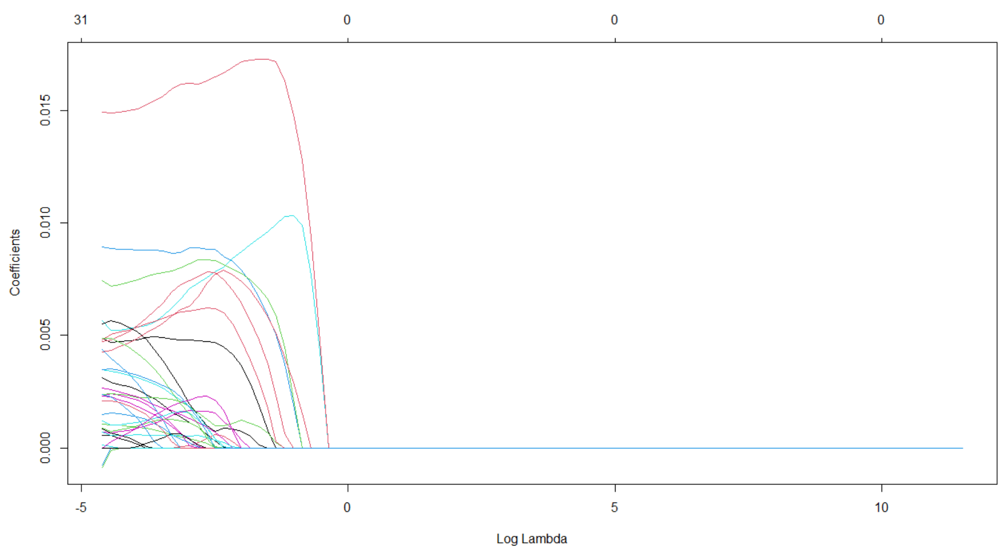
Several predictors exhibited VIF values above 10, particularly STA222, STA224, STA342, STA424, STA435, and MTH103, confirming the presence of significant multicollinearity in the dataset. This supports the need for regularization methods.

4.2 Cross-Validation and Model Selection

To determine optimal regularization parameters, 10-fold cross-validation was applied to both Ridge and LASSO regressions. The mean squared prediction error (MSE) was plotted against log(λ), and the best λ was selected based on either the minimum MSE or the one-standard-error rule. For Ridge, the optimal λ was 0.7035. For LASSO, two values were considered: $\lambda_{\text{min}} = 0.0166$ and $\lambda_{\text{1se}} = 0.0508$.



Trace plot for ridge regression



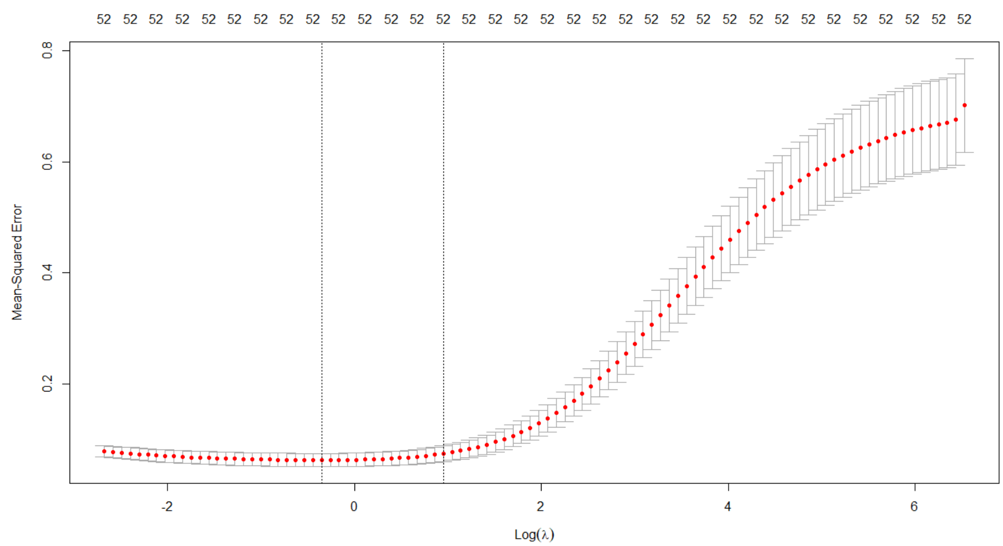
Trace plot for lasso regression

Figure 2: Coefficient estimates $\hat{\beta}_i$ for Ridge regression (up) and the LASSO (down) for the red wine data plotted versus $\log \lambda$. The upper part of the plot shows the number of non-zero coefficients $\hat{\beta}_i$ in the regression model for a given $\log \lambda$.

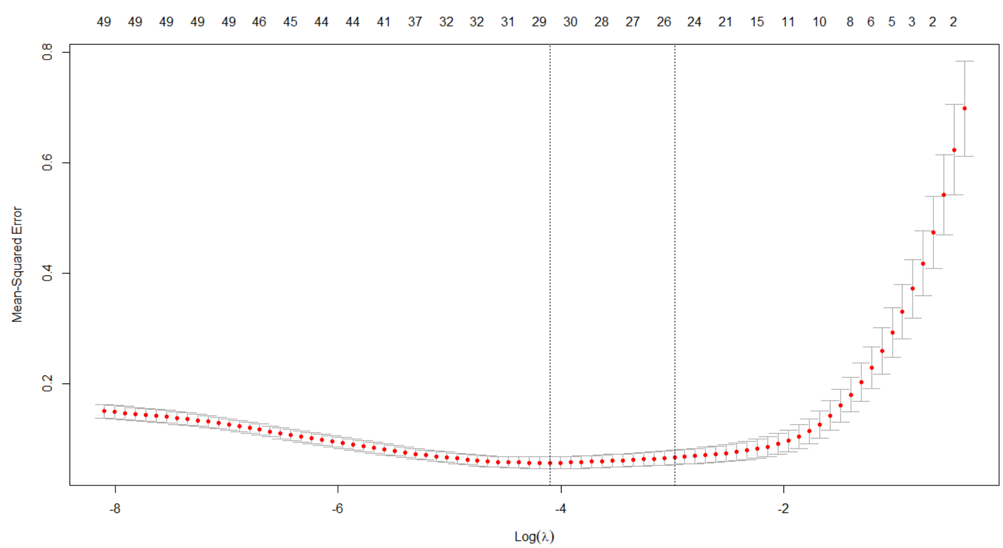
Figure 2 describes the trace plot for the ridge and lasso regression models. As λ increases, the coefficient estimates are progressively "shrunk" toward zero, reducing the norm of the estimate

International Journal of Mathematical Sciences and Optimization: Theory and Applications
11(4), 2025, Pages 46 - 56
https://doi.org/10.5281/zenodo.18986070

vector. However, individual coefficients may increase within specific intervals. The upper part of the plot illustrates the number of nonzero coefficient estimates for a given λ value. Notably, in Ridge regression, this number remains constant across all λ values, equaling the total number of predictors in the dataset. While Ridge regression effectively shrinks coefficient estimates toward zero, even large λ values do not result in exact zero estimates.



The cross-validation plot for ridge regression



The cross-validation plot for lasso regression

Figure 3: Cross-validated estimate of the mean squared prediction error for Ridge (up) and LASSO (down), as a function of $\log\lambda$. The upper part of the plot shows the number of non-zero coefficients $\hat{\beta}_i$ in the regression model for a given $\log\lambda$. The dashed lines show the location of the function minimum and the "one-standard-error" location

The plots illustrate the mean squared prediction error (MSE_λ) against $\log\lambda$. The gray bars at each point represent $MSE_\lambda\mp$ one standard error. One vertical dashed line marks the location of the minimum MSE, while the second dashed line indicates the value selected by the "one-standard-error" rule. Using cross-validation results, we determine the optimal λ values. The best λ for Ridge regression is $\lambda_{best} = 0.7035$, and for LASSO, it is $\lambda_{best} = 0.0166$ and $\lambda_{1se} = 0.0508$. The corresponding coefficient estimates for these λ values are presented in Table 2.

4.3 Coefficient Estimates and Variable Selection

Table 2 shows the estimated coefficients from OLS, Ridge, and LASSO models.

Table 2: Coefficient Estimates Across Different Regression Techniques

Coefficients	**OLS**	$Ridge_{\lambda=0.7035}$	$Ridge_{\lambda=2.5877}$	$LASSO_{\lambda=0.0166}$	$LASSO_{\lambda=0.0508}$
(Intercept)	-3.2064	-3.27	-2.6960	-2.7357	-1.9520
STA122	0.0049	0.00172	0.00131	0.00272	0.00112
STA124	-0.00013	0.00124	0.00136	.	0.00013
STA125	-0.00365	0.00081	0.00165	.	.
STA131	0.00092	0.00003	0.00050	.	.
STA132	0.00215	0.00153	0.00141	0.00094	0.00044
STA133	-0.00367	0.00056	0.00110	.	.
STA221	0.01014	0.00452	0.00314	0.00879	0.00886
STA222	0.01128	0.00317	0.00240	0.00543	0.00702
STA223	0.00153	0.00335	0.00253	0.00215	0.00013
STA224	0.00763	0.00326	0.00253	0.00522	0.00192
STA311	0.01514	0.00594	0.00373	0.01504	0.01618
STA315	0.01219	0.00400	0.00284	0.00732	0.00819
STA334	-0.00128	0.00078	0.00105	.	.
STA335	-0.00388	0.00180	0.00196	.	.
STA341	-0.00006	0.00173	0.00194	.	.
STA342	-0.00306	0.00163	0.00190	.	.
STA346	-0.01017	0.00043	0.00165	.	.
STA347	0.00189	0.00099	0.00115	0.00111	0.00152
STA348	0.00479	0.00279	0.00251	0.00078	0.00163
STA354	0.00796	0.00297	0.00218	0.00473	0.00478
STA422	0.00611	0.00320	0.00242	0.00455	0.00634
STA424	0.00372	0.00279	0.00224	0.00240	0.00184
STA435	0.00039	0.00122	0.00136	.	.
STA436	0.00778	0.00234	0.00152	0.00334	.
STA437	-0.00226	0.00104	0.00163	.	.
STA438	-0.00553	0.00160	0.00182	.	.
STA441	-0.00163	0.00168	0.00199	.	.
STA443	0.00470	0.00328	0.00235	0.00324	0.00184
STA444	0.00254	0.00253	0.00188	0.00239	0.00131
STA498	-0.00320	0.00291	0.00286	.	.
STA499	0.01498	0.00532	0.00542	.	.
MAT101	-0.00254	0.00145	0.00164	0.00038	.
MAT102	0.00661	0.00354	0.00262	0.00516	0.00749
MAT103	0.00632	0.00282	0.00241	0.00089	0.00114
MAT104	-0.00154	0.00206	0.00195	0.00143	0.00023

MAT201	0.00266	0.00198	0.00207	0.00053	0.00051
MAT202	0.00296	0.00195	0.00170	0.00201	0.00039
MAT203	0.00077	0.00248	0.00224	0.00050	.
MAT204	0.00276	0.00372	0.00262	0.00519	0.00612
MAT312	-0.00865	0.00018	0.00108	.	.
GNS111	0.00083	0.00264	0.00176	0.00169	.
GNS112	-0.00083	0.00107	0.00142	.	.
GNS211	-0.00317	0.00003	0.00103	.	.
GNS212	-0.00068	0.00105	0.00106	.	.
GNS311	0.00071	0.00184	0.00220	0.00065	0.00217
GSE301	0.00219	0.00027	0.00034	.	.
CSC101	0.00111	0.0016	0.00121	.	0.00041
CSC102	0.00383	0.00194	0.00123	0.00179	.
CSC201	0.00552	0.00235	0.00141	0.00450	0.00153
CSC203	0.00065	0.00273	0.00175	0.00335	0.00182
CHM101	0.00214	0.00052	0.00100	.	.
CHM102	-0.00185	0.00097	0.00134	.	.

Table 2 presents the coefficient estimates for various regression techniques, including Ordinary Least Squares (OLS), Ridge regression with different λ values, and LASSO with different λ values. The coefficients vary across models, highlighting the impact of regularization on parameter estimation.

For instance, in the OLS model, the intercept coefficient is -3.206. Ridge regression with $\lambda = 2.5877$ reduces it to -2.696, while LASSO with $\lambda = 0.0508$ further shrinks it to -1.952. Similarly, coefficients for individual predictors like STA122 and STA132 change across models, illustrating how Ridge and LASSO adjust parameter estimates based on the regularization strength (λ).

Notably, LASSO sets some coefficients, such as those for STA124, STA125, and STA335, to zero, indicating that these predictors contribute minimally to the model's predictive performance when penalization is applied. In contrast, Ridge regression retains all predictors but shrinks their magnitudes, reducing overfitting while preserving model complexity.

These results emphasize the role of regularization techniques in mitigating multicollinearity and enhancing model generalization.

4.4 Predictive Performance

Root Mean Squared Error (RMSE) was computed for both training and test datasets to evaluate generalization performance. The results are summarized in Table 1.

Table 1: Root Mean Squared Error (RMSE) for Training and Test Sets Across Different Regression Techniques

RMSE	OLS	$Ridge_{\lambda=0.7035}$	$Ridge_{\lambda=2.5877}$	$LASSO_{\lambda=0.0166}$	$LASSO_{\lambda=0.0508}$
RMSE (Train)	0.1180	0.1836	0.2364	0.1500	0.1828
RMSE (Test)	0.3398	0.2358	0.2853	0.2616	0.2803

For the training set, the OLS model has the lowest RMSE (0.1180), indicating the best fit to the training data. However, as λ increases in Ridge and LASSO regression, RMSE also increases, with the highest RMSE recorded for Ridge regression at $\lambda = 2.5877$ (0.2364).

For the test set, OLS performs the worst, with the highest RMSE (0.3398), suggesting overfitting and poor generalization to unseen data. In contrast, Ridge regression with $\lambda = 0.7035$ achieves the lowest test RMSE (0.2358), demonstrating the best generalization capability. The LASSO models also show competitive performance, with test RMSE values ranging from 0.2616 to 0.2803, slightly higher than Ridge at $\lambda = 0.7035$.

5 Discussion

The results of this study provide empirical support for the theoretical claims made in the literature regarding the strengths and limitations of different regression approaches in the presence of multicollinearity. As highlighted in the literature review, multicollinearity can severely distort OLS estimates by inflating the variance of regression coefficients, thereby reducing the model's predictive reliability [5,9]. This theoretical weakness was confirmed in our analysis: although OLS performed well on training data, it exhibited significant overfitting, resulting in the highest RMSE on the test set.

In contrast, Ridge regression, consistent with the findings of Hoerl and Kennard [3], stabilized coefficient estimates by introducing an L_2 penalty that shrank large coefficients while retaining all predictors. This method achieved the lowest test RMSE (0.2358), demonstrating superior generalization performance under conditions of high predictor correlation. This outcome aligns with the conclusions of Hastie et al. [9], who noted Ridge's effectiveness in balancing bias and variance trade-offs in multicollinear contexts.

LASSO regression, further improved model interpretability by reducing several coefficients to exactly zero, effectively performing variable selection. This behavior is particularly useful in educational modeling, where identifying the most influential courses can inform curriculum design and targeted intervention. While LASSO did not outperform Ridge in terms of test RMSE, its sparsity-producing nature yielded a simpler model with fewer predictors, consistent with the variable selection benefits discussed by Zou [7].

The comparative performance of Ridge and LASSO also confirms theoretical insights about their distinct behaviors: Ridge is preferable when all predictors contribute to the response with small effects, while LASSO is beneficial when a few variables carry most of the predictive signal. These findings are further supported by Bühlmann and Van De Geer [5], who emphasized the role of penalization in managing overfitting and enhancing generalizability in high-dimensional settings.

By applying Ridge and LASSO methods to academic performance data from the University of Ilorin, Nigeria, this study demonstrates their practical value in educational analytics for predicting outcomes and identifying key academic drivers. The findings validate the theoretical foundations of penalized regression and confirm their relevance in real-world applications. This indicates that regularization methods such as Ridge and LASSO not only reduce multicollinearity but also improve the reliability and interpretability of regression models in educational research and related fields.

6 Conclusion

This study conducted a comparative analysis of Ordinary Least Squares (OLS), Ridge, and LASSO regression methods to address the problem of multicollinearity in linear modeling. Using real-world academic data comprising 52 course scores and CGPA records from undergraduate students, the performances of these models based on their predictive accuracy and ability to handle collinearity among predictors were evaluated.

The findings confirm that while OLS is easy to implement and performs well on training data, it is highly sensitive to multicollinearity, resulting in poor generalization to test data. Ridge regression effectively mitigated this issue by applying L_2 regularization, which stabilized coefficient estimates and yielded the lowest test RMSE of 0.2358 at $\lambda = 0.7035$. LASSO, on the other hand, introduced sparsity through L_1 regularization, shrinking some coefficients to zero and enhancing model interpretability without severely sacrificing predictive performance.

These results align with existing literature on regularization methods and highlight the practical value of Ridge and LASSO in predictive modeling, especially in high-dimensional or correlated feature spaces. For education data analysts and data scientists working with student performance data, these models provide robust alternatives to traditional regression when multicollinearity is a concern.

Future studies could extend this comparison by incorporating Elastic Net, SCAD, or Bayesian shrinkage methods, and by testing these techniques in larger and more diverse datasets across different educational institutions.

References

[1] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis. *John Wiley & Sons*.

[2] Chan, J. Y., Leow, S. M., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics*, 10(8), 1283.

[3] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.

[4] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267-288.

[5] Bühlmann, P., & Van De Geer, S. (2011). Statistics for high-dimensional data: methods, theory and applications. *Springer Science & Business Media*.

[6] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), 301-320.

[7] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418–1429.

[8] Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348–1360.

[9] Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press.