African Journal of Advanced Pure and Applied Sciences (AJAPAS)
Online ISSN: 2957-644X
Volume 5, Issue 1, 2026
Page No: 459-466
Website: https://aaasjournals.com/index.php/ajapas/index

ISI 2025: 1.126 SJIFactor 2024: 6.752 معامل التأثير العربي: 1.62

Comparing Penalized Regression Analysis of Binary Logistic Regression Model with Multicollinearity Using Log Loss

Radi A. Othman *
Department of Statistics, Faculty of Science - Al Maraj, University of Benghazi, Libya

مقارنة أداء تحليل الانحدار المقيد لنموذج الانحدار اللوجستي الثنائي مع التعدد الخطي باستخدام الخسارة اللوغاريتمية

راضي عبد الرحيم عثمان *
قسم الإحصاء، كلية العلوم – المرج، جامعة بنغازي، ليبيا

*Corresponding author: radi.aljafil@uob.edu.ly

Received: December 22, 2025	Accepted: March 05, 2026	Published: March 17, 2026

Abstract
In this study compares the performance of the ridge, lasso, and elastic net approaches for controlling multicollinearity issues among independent variables in binary logistic regression analysis. A binary logistic model with eight independent variables (P=8) and a high degree of multicollinearity (ρ=0.99) at various sample sizes (20,50,100,200, and 300) is applied in data simulation. The best approach is found using the minimum logarithmic loss (Log Loss) and Akaike's information criterion (AIC) values. According to the study results, the lasso method consistently offers the lowest Log Loss and AIC, making it the best model. For every sample size tested, ridge method seems to be the least successful of the three models, whereas elastic net is a powerful substitute for lasso method.

Keywords: Binary logistic model, Multicollinearity, Log Loss, Ridge, Lasso, Elastic Net.

الملخص
في هذه الدراسة تمت مقارنة أداء أساليب (Ridge ،Lasso ،Elastic Net) للتحكم في مشكلة الارتباط الخطي المتعدد بين المتغيرات المستقلة في تحليل الانحدار اللوجستي الثنائي. اعتمدت الدراسة على بيانات محاكاة، حيث تم تطبيق نموذج لوجستي ثنائي بثمانية متغيرات مستقلة (P = 8) مع درجة عالية من التعددية الخطية (ρ = 0.99) بأحجام عينات مختلفة (20، 50، 100، 200، 300). الطريقة الأفضل هي التي تقدم أقل قيمة للخسارة اللوغاريتمية (Log Loss) ومعيار معلومات أكايكي (AIC). وفقًا لنتائج الدراسة، توفر طريقة (Lasso) باستمرار أقل قيم للخسارة اللوغاريتمية ولمعيار معلومات أكايكي أيضا، مما يجعلها النموذج الأفضل لأغلب احجام العينات التي تم اختبارها، اما طريقة (Ridge) تبدو هي الأقل نجاحًا من بين النماذج الثلاثة، بينما تُعد طريقة (Elastic Net) بديلاً قويًا لطريقة (Lasso).
.

الكلمات المفتاحية: نموذج الانحدار اللوجستي ثنائي، ارتباط خطي متعدد، خسارة لوغاريتمية، تحليل الانحدار المقيد.

1 Introduction

Logistic regression is an essential technique in statistical modeling, especially for binary classification scenarios, such as (pass, fail), (present, absent) or (diseased, not diseased), where the outcome variable is dichotomous. The extensive application across diverse domains including as, epidemiology, social sciences, medicine, economics, engineering, and machine learning is attributable to its capacity to model the probability of an event occurring based on one or more predictor variables[1]. One assumption of logistic regression is that predictor variables must be uncorrelated. The logistic model becomes unstable when there is a correlation among the independent variables, this may lead to inaccuracies in interpreting the relationship between the dependent variable and each independent variable about odds ratios, as the confidence intervals around the odds ratios expand, complicating the ability to make exact inferences about the effect sizes, this complicates using the model's findings in practical

applications[2]. When logistic regression encounters multicollinearity, the model experiences inflated errors in the estimated parameters and unstable parameter estimations, this causes parameters to seem important when they are not, or conversely, the interpretation and accuracy of logistic regression models are further complicated as a result[1,2,3]. The variance inflation factor (VIF) is one of the most popular techniques to detect multicollinearity in data, multicollinearity is stated to be present when the (VIF) is 10 or more, A (VIF) of 1.0 indicates no multicollinearity[5,6]. Researchers frequently employ regularization techniques like ridge, lasso, and elastic net regression to reduce multicollinearity in logistic regression[7,8], These techniques are meant to reduce predictor variable overlap, improve coefficient estimate consistency, and increase the model's overall reliability and clarity. Since ridge regression was first used by Hoerl and Kennard (1970) to handle multicollinearity in engineering data[9], a number of studies have been conducted to improve and refine the regularized regression model methods, for instance, McDonald & Schwing,1973; McDonald & Galarneau,1975; J. F. & P,1976; Dempster et al.,1977; Gibbons,1981; Schaeffer et al.,1984; Kibria,2003; Muniz & Kibria,2009; Mansson et al.,2010; Kibria et al.,2012; Aslam,2014; A.V. Dorugade,2014; Arashi & Valizadeh,2015; Melkumova & Shatskikh,2017; Lukman et al.,2018, 2019; Herawati et al.,2018; Yuzba， sı et al.,2020; Golam Kibria & Lukman,2020; Kibria,2023; Hoque & Kibria,2023; Mermi et al.,2024; Nayem et al.,2024; Hoque & Kibria,2024; Yasmin & Kibria, 2025[10-34]. Despite widely research on ridge regression, its application in logistic regression has received very little attention [35]. mean squared error (MSE) is the main measurement instrument used in a large portion of current research to evaluate the classification performance of logistic regression models. this measure cannot be trusted to evaluate the performance of binary logistic regression models, because the estimated values $\hat{y}_i$ is a nonlinear function. Consequently, the MSE equation in logistic regression will produce a non-convex function, we will have difficulties when classifying, since we have target values like (0, 1), so $(\hat{y} - y)^2$ will always fall within (0-1), It could make error tracking and storing high precision floating integers difficult.
This gap encourages more analytical simulation-based research, the most essential measure that outperforms MSE is the minimum logarithmic loss (Log Loss), which offers a continuous and differentiable assessment of the binary logistic regression model's performance, where Log loss is the most important probability-based classification measure[36]. The aim of the current study is to add to the body of literature regarding binary logistic regression when multicollinearity is present, where a logistic model with binary responses and a set of continuous predictor variables was used to compare the ridge, lasso, and elastic net approaches, simulation data with high multicollinearity and varying sample sizes were used to compare each approach, the best technique was determined based on the smallest value of Log Los and the best model is characterized by AIC value.

2 Methodology

2.1 Binary logistic regression model

Assume the response variable of the regression application of interest has two possible results or that Y_i is a Bernoulli random variable with a probability distribution,

$P(y_i = 0) = 1 - \pi_i$ and $P(y_i = 1) = \pi_i$

The probability function for each observation is[4,5,6],

$$f(y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}, \; i = 1,2,\dots,n \tag{1}$$

The multiple logistic regression model of the response variable is expressed as follows[37],

$$\pi_i(x_i) = P[y_i = 1|x_i] = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})} \tag{2}$$

Where, $x_i = (x_{i1}, x_{i2} \dots, x_{ik})$ is the input vector of k independent variables and $\beta_{k\times 1}$ is the vector of estimated parameters.

when the logit transformation is applied, the model turns into a linear model as follows[37],

$$\text{logit}[\pi_i(x)] = \ln\left(\frac{\pi_i(x_i)}{1 - \pi_i(x_i)}\right) \tag{3}$$

or can be expressed in linear form as,

$$\text{logit}[\pi_i(x)] = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} \tag{4}$$

By maximizing the likelihood function, the parameters were determined as [38],

$$L(\beta) = \prod_{i=1}^{n} \pi_i(x_i)^{y_i}(1 - \pi_i(x_i))^{1-y_i} \tag{5}$$

Consequently, the log likelihood function is,

$$\ell(\beta) = \sum_{i=1}^{n}\left[y_i \ln(\pi_i(x_i)) + (1 - y_i)\ln(1 - \pi_i(x_i))\right] \tag{6}$$

When the log-likelihood is differentiated with regard to β equal to zero in matrix notation, we obtain[4],

$$\hat{\beta}_{ML} = (X'\widehat{W}X)^{-1}X'\widehat{W}Z \tag{7}$$

Where, $Z_{n\times 1}$ is a column vector with elements,

$$Z_i = \text{logit}(\hat{\pi}_i) + \frac{y_i \hat{\pi}_i}{\hat{\pi}_i(1-\hat{\pi}_i)} \text{ and } \hat{W} = \text{diag}[\hat{\pi}_i(1-\hat{\pi}_i)] \quad (8)$$

The matrix $X'WX$ is (almost) singular when there is multicollinearity among the independent variables in the logistic model, since we are unable to obtain the matrix's inversion, the maximum likelihood method is not appropriate for estimating the model's parameters. Because of this, the maximum likelihood technique for estimating the logistic model's parameters is unstable and cannot be uniquely estimated. Variation inflation factor (VIF) is one method for detecting multicollinearity. The independent variables' multicollinearity is determined by,

$$\text{VIF}_j = \frac{1}{\left(1-R_j^2\right)} \quad (9)$$

The coefficient of determination for the regression of that explanatory variable on the other independent variables is denoted by R_j^2, there is multicollinearity between the independent variables if the VIF value is more than 10.

In this case, regularized binary logistic regression models can be used to stabilize the coefficient estimates by applying a penalty to the diagonal matrix of $X'WX$. All penalized regression techniques provide notable advantages over ordinary maximum likelihood estimation, especially when multicollinearity is high[7].

2.2 Regularized binary logistic regression models

In recent years, regularized regression models have become more and more popular among machine learning algorithms. Ridge, lasso, and elastic net logistic regression models are frequently used as regularized models in various fields to control multicollinearity problems among independent variables, in regularized logistic regression models to estimate the unknown regression, a penalty term is added to the log-likelihood functions of these models[21].

2.2.1 Ridge logistic regression model

Ridge regression is particularly useful when predictors are highly correlated, since it stabilizes coefficient estimates by reducing them towards zero, its penalty parameter is usually selected using cross-validation[39]. Despite the fact that it will result in bias in the model's coefficient estimations, this approach offers a lower variance of the coefficient estimates than the unpenalized model. Ridge likelihood estimator of the logistic model is determined by maximize the ridge penalized loglikelihood[15,19,35,40], the ridge logistic regression parameter estimates are expressed as follows,

$$\ell^{\text{Ridge}}(\beta) = \ell(\beta) \mid + \lambda \sum_{j=1}^{k} {\beta_j}^2 \quad (10)$$

This model uses the $L2$-norm penalty ($\sum_{j=1}^{k} {\beta_j}^2$) to solve the log-likelihood function of the binary logistic regression model, the tuning parameter (λ>0) regulates the amount of shrinkage, when the value of (λ=0) the estimator (ML) can be seen as a particular case of equation (3).

2.2.2 Lasso logistic regression model

Multicollinearity issues can be resolved by using the least absolute shrinkage and selection operator (lasso) approach [41]. In contrast to ridge regression, which depends on the squared $L2$-norm, lasso uses the $L1$-norm to minimize the sum of differences while limitation to a constraint on the total absolute magnitude of the coefficients β, so that it correlates to zero or nearly zero, this limitation motivates the model to be sparse, finding a subset of actually relevant predictors and successfully reducing multicollinearity [42,43]. The Lasso logistic regression model is created by adding a penalty term to the negative log-likelihood function, known as $L1$-norm ($\sum_{j=1}^{k}|\beta_j|$), with tuning parameter λ. Thus, using lasso, we obtain an estimate of the logistic regression parameter as follows,

$$\ell^{\text{Lasso}}(\beta) = -\ell(\beta) + \lambda \sum_{j=1}^{k} |\beta_j| \quad (11)$$

The tuning parameter (λ>0), which controls the penalty severity in the lasso approach, can be acquired by generalized cross validation [41].

2.2.3 Elastic net logistic regression model

Another regularization and variable selection method called the elastic net, which was first presented by Zou and Hastie in 2005, where the lasso and ridge models are combined to create the elastic net model. To put it another way, it transmits the characteristics of both lasso and ridge by using a combination of $L1$-norm and $L2$-norm penalties. The Elastic net approach offers a balance between lowering the parameters to zero and reducing their size by using an additional parameter to tune is (α) [44]. The elastic net logistic regression parameter estimates are given as follows,

$$\ell^{\text{Elastic net}}(\beta) = \ell(\beta) + \lambda[\tfrac{1}{2}(1-\alpha)\sum_{j=1}^{k} {\beta_j}^2 + \alpha \sum_{j=1}^{k}|\beta_j|] \quad (12)$$

The elastic net logistic regression model is he identical as ridge logistic regression when (α=0) and lasso logistic regression when (α=1). The value of α is usually set at 0.5 for an equal combination of ridge and lasso models.

2.3 Simulation study
Using the R programming language, a Monte Carlo simulation is applied for comparing the performance of ridge regression, lasso, and elastic net for a binary logistic regression model with high multicollinearity. Eight independent variables (k=8) are used in the simulated logistic regression model, to obtain multicollinearity in the data collection, employed the following formula[16],

$$X_{ij} = \sqrt{(1-\rho^2)}\ Z_{ij} + \rho Z_{i(k+1)}, \qquad i = 1,2,\ldots\ldots,n\ ;\ j = 1,2,\ldots.,k \qquad (13)$$

Z_{ij} represents the independent pseudo-random variable $Z_{ij} \sim N(0,1)$, where ρ is the correlation between any two independent variables, with β_0=0 and $\beta_1 = \beta_2 \ldots \beta_k$=1, the dependent variable Y is produced by the binary logistic regression probability given by equation (2). The experiment was conducted 1000 times using the model with different sample sizes (n = 20,50,100,200, and 300) with high multicollinearity amongst independent variables (ρ=0.99).

2.3.1 Choice of tuning parameter
The values of estimated parameters are determined by the amounts of λ, therefore choosing the right value of λ is essential to avoiding overfitting and underfitting problems for the model. Cross-validation is frequently used to choose the appropriate λ values that provide a good balance between variance and bias and lower the misclassification error. Because of this, the optimal value of λ in this study is found using cv. glmnet() from the glmnet package [45], [46]. 10-fold cross-validation was used to determine the optimal λ tuning parameter for Ridge, Lasso, and Elastic net logistic regression model.

2.3.2 Measurement of performance
The models studied in multicollinearity handling will be evaluated for performance using,
a- logarithmic loss or cross-entropy loss (Log Los)
The Log Loss is an important measure for binary classification models, it evaluates a model's performance by measuring the difference between predicted probabilities and actual values [36], we can apply this formula,

$$\text{Log Los} = -\frac{1}{n}\sum_{i=1}^{n} y_i.\log[p(y_i)] + (1-y_i).\log[1-p(y_i)] \qquad (14)$$

The actual class is represented by y_i, and is the probability of that class is $\log[p(y_i)]$, where $p(y_i)$ is the probability of 1and $[1-p(y_i)]$ is the probability of 0. Better model performance is shown by a lower Log Loss.
b- Akaike's information criterion (AIC)
A measure of fit used to evaluate the various models is the AIC. The number of estimated parameters in the model is added as a penalty term to the log-likelihood value [37]. The calculation of the AIC value is as follows,

$$\text{AIC} = -2(\ell(\beta)) + 2d \qquad (15)$$

Where the number of estimated parameters is denoted by d. Better model fit is indicated by smaller AIC values.

2.3.3 Modelling steps
The following steps were used to run the ridge, lasso, and elastic net logistic regression models,
a. Split the data into training and testing sets
The data was split into two sets: a training set of 66% and a test set of 44% in order to build, validate and compare each approach. The quality has been evaluated by applying the estimated models to the test set.
b. Binary logistic regression model
- Conducted binary logistic regression model with high multicollinearity (ρ=0.99).
- Detecting multicollinearity amongst the independent variables using (VIF).

c. Ridge logistic regression model
- Fitting ridge model with training data and (α=0).
- Determined the optimal λ with 10-fold cross-validation.
- Fitting final ridge model with optimal lambda.
- Calculate the Log Loss for the final Ridge logistic regression model.
- Calculate the AIC for the final Ridge logistic regression model.

d. Lasso logistic regression model
- Fitting lasso model with training data and (α=1).
- Determined the optimal λ with 10-fold cross-validation.
- Fitting final lasso model with optimal lambda.
- Calculate the Log Loss for the final lasso logistic regression model.
- Calculate the AIC for the final lasso logistic regression model.

e. Elastic net logistic regression model
- Fitting elastic net model with training data and (0< α <1).
- Determined the optimal α and λ with 10-fold cross-validation.
- Fitting final Elastic net model with optimal lambda and alpha.
- Calculate the Log Loss for the final Elastic net logistic regression model.
- Calculate the AIC for the final Elastic net logistic regression model.

3. Results and discussion

The simulated data's initial VIF values were created to exhibit a strong correlation ($\rho=0.99$) among all independent variables, this requirement applies to all sample sizes. Consequently, if the VIF value of the related variable is greater than 10, a multicollinearity problem exists. The experiment was repeated 1000 times to acquire reliable estimation results, display the outcomes of the simulation in Table 1,

Table 1. Multicollinearity in independent variables for all sample sizes studied**.**

independent variables	VIF for all sample sizes studied				
	$n_1 = 20$	$n_2 = 50$	$n_3 = 100$	$n_4 = 200$	$n_5 = 300$
X_1	238.03	180.48	34.10	58.15	11.54
X_2	2451	210.89	46.98	69.9	33.75
X_3	3413.88	265.99	58.94	31.96	32.43
X_4	2560.38	338.52	42.14	9.22	71.55
X_5	1529.2	603.1	35.84	16.43	11.55
X_6	840.47	323.68	20.15	107.32	7.40
X_7	2324.9	1071.34	105.67	62.57	58.98
X_8	2054.42	552.89	37.66	56.43	19.95

From Table 1, Based on the simulation results, the independent variables demonstrate extreme multicollinearity. Although VIFs typically decrease with larger sample sizes (e.g., X_3 from 3413.88 to 32.43), they still remain problematic because the VIF values for almost all variables across all sample sizes significantly exceed the threshold of 10, this indicates that the intrinsic correlation structure causes problems even with more data, this shows that the estimations of the individual regression coefficients may be unstable or incorrect. The value of Log Los and AIC will be examined to determine which method best controls the multicollinearity in this study, if the Log Los and AIC value is close to zero then, the technique will be the best. The Log Los and AIC values of ridge, lasso, and elastic net binary logistic regression under multicollinearity in the model on various sample sizes are displayed in Table 2.

Table 2. Log Loss and AIC of Ridge, Lasso and Elastic net for all sample sizes studied**.**

Model	Sample size	Log Loss	AIC
Ridge	20	0.2060	18.5387
	50	0.1118	21.9486
	100	0.1159	30.1822
	200	0.1102	47.8914
	300	0.1398	63.4968
Lasso	20	0.0213	4.0167
	50	0.0442	4.6455
	100	0.0053	4.5810
	200	0.0474	6.3639
	300	0.0483	13.4020
Elastic net	20	0.0278	16.0304
	50	0.0418	18.6965
	100	0.0172	16.7611
	200	0.0296	15.9906
	300	0.0513	21.3568

From Table 2, ridge logistic regression model consistently demonstrates the highest Log Loss and AIC values compared to both lasso and elastic net across all sample sizes, ridge logistic regression model has the greatest Log Loss values, ranging from 0.1102 to 0.206. The AIC values for ridge logistic regression model increase dramatically as the sample size expands (from 18.5387 at n=20 to 63.4968 at n=300), showing that its relative model complexity or poor fit worsens with larger datasets in this context, this indicates that the ridge logistic regression generally provides a lower fit, a higher error rate, and a more complex model. In terms of both measures, the lasso and elastic net logistic regression models consistently produce far lower Log Loss and AIC values than the ridge logistic regression model. Across all sample sizes, lasso typically exhibits very low Log Loss; the lowest value of all models was found with a sample size of 100 (0.0053), the AIC values for lasso are consistently the lowest among the three models for most sample sizes, showing it often delivers the most parsimonious (least complex) and best-fitting model. Elastic net's performance is competitive with lasso, consistently demonstrating low Log Loss and AIC values. Its measurements are nevertheless significantly lower

than those of the ridge logistic regression model, while often slightly higher than lasso best performance, its measurements remain much lower than ridge logistic regression model.
To offer clearer findings about the ridge, lasso and elastic net techniques in overcoming high multicollinearity for all sample sizes, we evaluated the Log Loss and AIC values of the three methods as shown in Figures 1and 2.

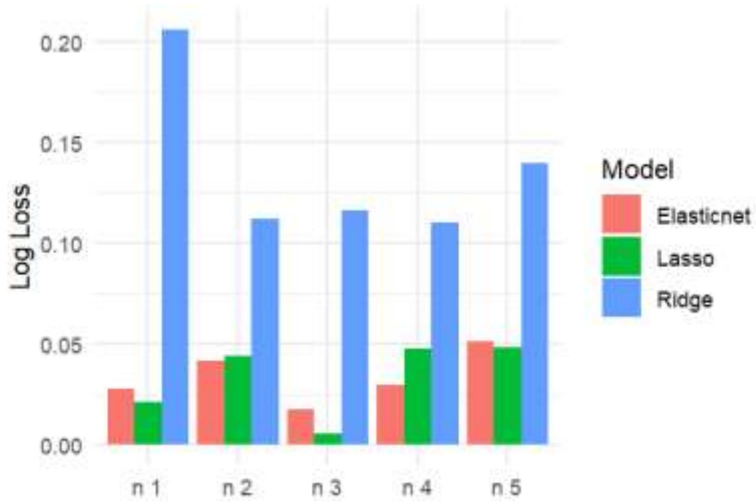

Figure 1. Log Loss of Ridge, Lasso and Elastic net for all sample sizes studied.

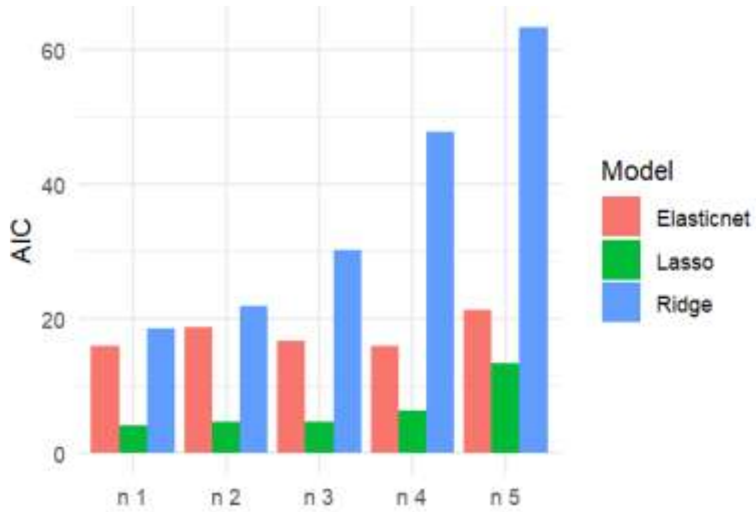

Figure 2. AIC of Ridge, Lasso and Elastic net for all sample sizes studied.

The lasso logistic regression model was the most successful under high multicollinearity for all sample sizes, based on the values of Log Los from figure 1 and AIC from figure 2. This supports the conclusions drawn from the above table.

4. Conclusions

Based on the simulation results for eight independent variables (P=8) and the number of data (n=20, 50, 100, 200 and 300) containing severe multicollinearity between independent variables, multicollinearity problems can be resolved with the Lasso logistic regression model. Overall, lasso emerges as the superior model in this specific scenario based on the provided metrics, consistently offering the lowest Log Loss and AIC. Elastic net is a strong alternative, while the ridge logistic regression model appears to be the least effective model among the three for this study.
Since the lasso and elastic net models performed similarly in this study, the researcher generally suggests using the lasso model when the study calls for extreme simplification of the model, and the elastic net model when the study requires maintaining a balance of related variables.

Compliance with ethical standards

Disclosure of conflict of interest

The author(s) declare that they have no conflict of interest.

References

[1] D. W. Hosmer and S. Lemeshow, “Applied Logistic Regression,” in *Applied Logistic Regression*, 2000. doi: 10.1002/0471722146.

[2] A. Agresti, *Foundations of Linear and Generalized Linear Models*. 2015.

[3] L. Bruno, *LOGISTIC REGRESSION USING THE SAS SYSTEM Theory and Application*, vol. 53, no. 9. 2019.

[4] T. P. Ryan, *Modern regression methods, second edition*. 2008. doi: 10.1002/9780470382806.

[5] H. Midi, S. K. Sarkar, and S. Rana, “Collinearity diagnostics of binary logistic regression model,” *Journal of Interdisciplinary Mathematics*, vol. 13, no. 3, 2010, doi: 10.1080/09720502.2010.10700699.

[6] M. Kutner, C. Nachtsheim, J. Neter, and W. Li, “Applied Statistical Linear Models,” in *McGraw Hill*, 2005, ch. 9, pp. 406–410.

[7] T. Hastie, R. Tibshirani, and J. Friedman, “The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics) (9780387848570): Trevor Hastie, Robert Tibshirani, Jerome Friedman: Books,” in *The elements of statistical learning: dta mining, inference, and prediction*, 2011.

[8] G. James, D. Witten, T. Hastie, and R. Tibishirani, *An Introduction to Statistical Learning with Applications in R (older version)*. 2013.

[9] A. E. Hoerl and R. W. Kennard, “Ridge Regression: Biased Estimation for Nonorthogonal Problems,” *Technometrics*, vol. 12, no. 1, 1970, doi: 10.1080/00401706.1970.10488634.

[10] G. C. McDonald and R. C. Schwing, “Instabilities of regression estimates relating air pollution to mortality,” *Technometrics*, vol. 15, no. 3, 1973, doi: 10.1080/00401706.1973.10489073.

[11] G. C. McDonald and D. I. Galarneau, “A monte carlo evaluation of some ridge-type estimators,” *J. Am. Stat. Assoc.*, vol. 70, no. 350, 1975, doi: 10.1080/01621459.1975.10479882.

[12] J. F. Lawless and P. Wang, “A Simulation Study Of Ridge And Other Regression Estimators,” *Commun. Stat. Theory Methods*, vol. 5, no. 4, 1976, doi: 10.1080/03610927608827353.

[13] A. P. Dempster, M. Schatzoff, and N. Wermuth, “A simulation study of alternatives to ordinary least squares,” *J. Am. Stat. Assoc.*, vol. 72, no. 357, 1977, doi: 10.1080/01621459.1977.10479910.

[14] D. G. Gibbons, “A simulation study of some ridge estimators,” *J. Am. Stat. Assoc.*, vol. 76, no. 373, 1981, doi: 10.1080/01621459.1981.10477619.

[15] R. L. Schaefer, L. D. Roi, and R. A. Wolfe, “A ridge logistic estimator,” *Commun. Stat. Theory Methods*, vol. 13, no. 1, 1984, doi: 10.1080/03610928408828664.

[16] B. M. Golam Kibria, “Performance of some New Ridge regression estimators,” *Communications in Statistics Part B: Simulation and Computation*, vol. 32, no. 2, 2003, doi: 10.1081/SAC-120017499.

[17] G. Muniz and B. M. G. Kibria, “On some ridge regression estimators: An empirical comparisons,” *Commun. Stat. Simul. Comput.*, vol. 38, no. 3, 2009, doi: 10.1080/03610910802592838.

[18] K. Månsson, G. Shukur, and B. M. Golam Kibria, “A simulation study of some ridge regression estimators under different distributional assumptions,” *Commun. Stat. Simul. Comput.*, vol. 39, no. 8, 2010, doi: 10.1080/03610918.2010.508862.

[19] B. M. G. Kibria, K. Månsson, and G. Shukur, “Performance of Some Logistic Ridge Regression Estimators,” *Comput. Econ.*, vol. 40, no. 4, 2012, doi: 10.1007/s10614-011-9275-x.

[20] M. Aslam, “Performance of Kibria’s method for the heteroscedastic ridge regression model: Some Monte Carlo evidence,” *Commun. Stat. Simul. Comput.*, vol. 43, no. 4, 2014, doi: 10.1080/03610918.2012.712185.

[21] A. V. Dorugade, “On Comparison of Some Ridge Parameters in Ridge Regression,” *Sri Lankan J. Appl. Stat.*, vol. 15, no. 1, 2014, doi: 10.4038/sljastats. v15i1.6792.

[22] M. Arashi and T. Valizadeh, “Performance of Kibria’s methods in partial linear ridge regression model,” *Statistical Papers*, vol. 56, no. 1, 2015, doi: 10.1007/s00362-014-0578-6.

[23] L. E. Melkumova and S. Y. Shatskikh, “Comparing Ridge and LASSO estimators for data analysis,” in *Procedia Engineering*, 2017. doi: 10.1016/j.proeng.2017.09.615.

[24] N. Herawati, K. Nisa, and E. Setiawan, “Regularized Multiple Regression Methods to Deal with Severe Multicollinearity,” *Int. J. Stat. Appl.*, vol. 8, no. 4, 2018.

[25] A. F. Lukman, O. A. Oluyemi, O. B. Akanbi, and O. A. Clement, “Classification-Based Ridge Estimation Techniques of Alkhamisi Methods,” *Journal of Probability and Statistical Science*, vol. 16, no. 2, 2018.

[26] A. F. Lukman, K. Ayinde, S. Binuomote, and O. A. Clement, “Modified ridge-type estimator to combat multicollinearity: Application to chemical data,” *J. Chemom.*, vol. 33, no. 5, 2019, doi: 10.1002/cem.3125.

[27] B. M. Golam Kibria and A. F. Lukman, “A new ridge-type estimator for the linear regression model: Simulations and applications,” *Scientifica (Cairo).*, vol. 2020, 2020, doi: 10.1155/2020/9758378.
[28] B. Yüzbaşı, M. Arashi, and S. Ejaz Ahmed, “Shrinkage Estimation Strategies in Generalised Ridge Regression Models: Low/High-Dimension Regime,” *International Statistical Review*, vol. 88, no. 1, 2020, doi: 10.1111/insr.12351.
[29] B. M. Golam Kibria, “More than hundred (100) estimators for estimating the shrinkage parameter in a linear and generalized linear ridge regression models,” *Journal of Econometrics and Statistics*, vol. 2, no. 2, 2022.
[30] M. A. Hoque and B. M. Golam Kibria, “SOME ONE AND TWO PARAMETER ESTIMATORS FOR THE MULTICOLLINEAR GAUSSIAN LINEAR REGRESSION MODEL: SIMULATIONS AND APPLICATIONS,” *Surveys in Mathematics and its Applications*, vol. 18, 2023.
[31] S. Mermi, Ö. Akkuş, A. Göktaş, and N. Gündüz, “A new robust ridge parameter estimator having no outlier and ensuring normality for linear regression model,” *J. Radiat. Res. Appl. Sci.*, vol. 17, no. 1, 2024, doi: 10.1016/j.jrras.2023.100788.
[32] H. M. Nayem, S. Aziz, and B. M. G. Kibria, “Comparison among Ordinary Least Squares, Ridge, Lasso, and Elastic Net Estimators in the Presence of Outliers: Simulation and Application,” *International Journal of Statistical Sciences*, 2024, doi: 10.3329/ijss.v24i20.78212.
[33] M. A. Hoque and B. M. G. Kibria, “Performance of some estimators for the multicollinear logistic regression model: theory, simulation, and applications,” *Research in Statistics*, vol. 2, no. 1, 2024, doi: 10.1080/27684520.2024.2364747.
[34] H. M. Nayem, S. Aziz, and B. M. G. Kibria, “Evaluating Estimator Performance Under Multicollinearity: A Trade-Off Between MSE and Accuracy in Logistic, Lasso, Elastic Net, and Ridge Regression with Varying Penalty Parameters,” *Stats (Basel).*, vol. 8, no. 2, 2025, doi: 10.3390/stats8020045.
[35] J. Wu and Y. Asar, “On almost unbiased ridge logistic estimator for the logistic regression model,” *Hacettepe Journal of Mathematics and Statistics*, vol. 45, no. 3, 2016, doi: 10.15672/HJMS.20156911030.
[36] M. Setia, “Log Loss vs. Mean Squared Error: Choosing the Right Metric for Classification,” Analytics Vidhya.
[37] A. Agresti, *Book - An Introduction to Categorical Data Analysis*. 2019.
[38] S. A. Czepiel, “Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation,” *Class Notes*, 2012.
[39] A. Araveeporn and C. Kuharatanachai, “Comparing penalized regression analysis of logistic regression model with multicollinearity,” in *ACM International Conference Proceeding Series*, 2019. doi: 10.1145/3343485.3343487.
[40] D. E. Duffy and T. J. Santner, “On the small sample properties of norm-restricted maximum likelihood estimators for logistic regression models,” *Commun. Stat. Theory Methods*, vol. 18, no. 3, 1989, doi: 10.1080/03610928908829944.
[41] R. Tibshirani, “Regression Shrinkage and Selection Via the Lasso,” *Journal of the Royal Statistical Society. Series B: Methodological*, vol. 58, no. 1, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.
[42] M. Yuan and Y. Lin, “On the non-negative garrotte estimator,” *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 69, no. 2, 2007, doi: 10.1111/j.1467-9868.2007.00581.x.
[43] K. Lounici, “Sup-norm convergence rate and sign concentration property of lasso and dantzig estimators,” *Electron. J. Stat.*, vol. 2, 2008, doi: 10.1214/08-EJS177.
[44] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 67, no. 2, 2005, doi: 10.1111/j.1467-9868.2005.00503.x.
[45] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity: The lasso and generalizations*. 2015. doi: 10.1201/b18401.
[46] Ş. TOPTANCI, N. ERGİNEL, and I. P. ACAR, “Predicting the severity of occupational accidents in the construction industry using standard and regularized logistic regression models,” *Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 2023, doi: 10.28948/ngumuh.1212385.

