UAD Universitas Ahmad Dahlan
Bulletin of Applied Mathematics and Mathematics Education
Volume 5, Number 1, April 2025 pp. 23-34
https://dx.doi.org/10.12928/bamme.v5i1.12759
e-ISSN 2776-1029, p-ISSN 2776-1002

Multicollinearity problem-solving with Jackknife Ridge Regression: A case study on slum conditions in Bone Bolango

Dwi Putri Juniar Adam, Salmun K. Nasib, Amanda Adityaningrum*

Statistics Department, Universitas Negeri Gorontalo, Jl. Prof. Dr. Ing. B. J. Habibie, Tilongkabila, Kabupaten Bone Bolango, Gorontalo 96119, Indonesia

*Corresponding e-mail: amanda@ung.ac.id

ARTICLE INFO

Article History
Received 10 February 2025
Revised 20 May 2025
Accepted 21 May 2025

Keywords
Jackknife Ridge Regression
Multicollinearity
Slum conditions

How to cite this article:
Adam, D. P. J., Nasib, S. K., & Adityaningrum, A. (2025). Multicollinearity problem-solving with Jackknife Ridge Regression: A case study on slum conditions in Bone Bolango. *Bulletin of Applied Mathematics and Mathematics Education*, *5*(1), 23-34.

ABSTRACT

Slum conditions in Indonesia, particularly in Gorontalo Province's Bone Bolango District, are a significant challenge to sustainable development. This research aims to identify the key factors contributing to slum conditions in the strategic economic areas of Kabila, Suwawa, and Tilongkabila using Jackknife Ridge Regression (JRR) analysis to address multicollinearity and overfitting issues. Data from the Regional Development Planning Board (BAPPEDA) Bone Bolango District's 2023 document was used, with a sample of 40 urban villages and villages. The result showed that there is a high collinearity between two independent variables, necessitating the use of JRR. The JRR model identified seven independent variables significantly related to slum value. The regression model explained 83% of the variability in slum conditions. This study provides methodological depth through the JRR framework, which enables accurate slum analysis where traditional models (like OLS) tend to fall short. It emphasizes the need for Bone Bolango to prioritize its policy initiatives by focusing on the seven independent variables. Additionally, the framework demonstrates scalability, making it adaptable to other Indonesian provinces that face similar challenges with slum data.

Introduction

In Indonesia, approximately 55,000 hectares of slums remain unaddressed. Specifically, in the Bone Bolango District of Gorontalo Province, a strategic economic zone, there are 459,81 hectares of slums, hindering sustainable development. Kabila, Suwawa, and Tilongkabila are three subdistricts included in the strategic economic area (Arifin et al., 2023; Kabupaten Bone Bolango, 2021). Previous studies have identified several key determinants of slum conditions, including inadequate waste infrastructure and facilities that fail to meet technical requirements, inadequate waste management systems, and lack of fire protection facilities, the lack of safe and quality drinking water, the length of roads with damaged surfaces, the length of damaged drainage channels, inadequate wastewater management infrastructure and facilities that fail to meet technical standards, and poorly constructed buildings (Afrina et al., 2021; Gobel & Zees, 2022; Prajamandana et al., 2021). Additionally, the Regional Development Planning Board (BAPPEDA) Bone Bolango

District's document identifies several contributing factors, such as buildings that do not meet technical requirements, ideal road lengths, existing road lengths, ideal drainage channel lengths, and existing drainage channel lengths (Dinas PUPR, 2023). This results in a total of 15 variables that require thorough analysis.

Regression analysis is the traditional methodological framework for investigating the relationship between slum conditions and the factors significantly associated with them. In some cases, however, multicollinearity (the presence of highly correlated independent variables) may occur in the model, leading to instability in the parameter estimates of Ordinary Least Squares (OLS) regression (Montgomery et al., 2012; Shrestha, 2020; Yitshak-Sade et al., 2020). While removing some variables may help to overcome this issue, it can also result in a loss of valuable information and skewed or misleading results. Ultimately, this may result in a wider range of confidence intervals and a lower reliability of probability values for predictors, making the model's conclusions less accurate (Kutner et al., 2005; Montgomery et al., 2012; Shrestha, 2020). In addition, in cases of overfitting, where the number of predictors approaches the number of observations in the analyzed population, the variance can be inflated, resulting in insufficient statistical power to detect small effects. Further, overfitting can significantly impact low-dimensional data, particularly if the relationship between the outcome and the predictor variables is weak (Madrid-García et al., 2022; Yitshak-Sade et al., 2020).

General Ridge Regression (GRR) is one of the analyses that can be applied to eliminate the effects of multicollinearity and overfitting for small sample sizes. However, GRR often exhibits significant bias, which has been largely overlooked in the literature (Khurana et al., 2014). A development in GRR called Jackknife Ridge Regression (JRR) involves removing and repeating one data sample as many times as the number of data samples (Tinungki, 2019). JRR integrates ridge penalties with Quenouille's resampling method, effectively reducing bias and Mean Squared Error (MSE), particularly in smaller samples (Arrasyid et al., 2021; Khurana et al., 2014; Tinungki, 2019). Lasso and Principal Component Regression (PCR) are two reasonable alternatives; however, Lasso's variable elimination approach may omit relevant predictors (Tibshirani, 1996), and PCR's component transformation may compromise interpretability (Xie & Kalivas, 1997). In contrast, JRR maintains all variables, stabilizes estimates in high multicollinearity, and surpasses GRR in terms of bias reduction (Khurana et al., 2014; Arrasyid et al., 2021). Therefore, JRR can be used to address multicollinearity, especially in small datasets.

Prior research conducted using JRR to address multicollinearity issues includes research by Fatihah (2022), which examined the factors significantly influencing the poverty rate in Grobogan District, then Arrasyid et al. (2021) and Winda (2021), both of which focused on the factors significantly influencing the human development index in Central Java. Collectively, these research results show the effectiveness of JRR in overcoming problems related to multicollinearity.

Until now, no research has been conducted on slums in Bone Bolango Regency, Gorontalo Province. Furthermore, there has been no research into factors significantly related to slums considering multicollinearity cases. JRR is particularly suitable for studies in slum areas, where comprehensive variables are retained and accuracy is critical for informing policy decisions. Thus, this research aims to identify key factors that significantly influence the slum conditions using JRR, and to evaluate the performance of JRR as a solution to multicollinearity and overfitting. It is hypothesized that at least one variable will significantly influence slum conditions. Furthermore, it is expected that JRR will outperform traditional methodology, such as OLS.

 10.12928/bamme.v5i1.12759

Method

Standardized Regression Model

The general linear regression model, characterized by normal error terms, is defined in terms of X variables, as expressed in Equation (1):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \tag{1}$$

where $\beta_0, \beta_1, \ldots, \beta_{p-1}$ represents parameters; $X_{i1}, \ldots, X_{i,p-1}$ represents known constants; ε_i are independent $N\left(0, \sigma^2\right)$, $i = 1, \ldots, n$. By setting $X_{i0} \equiv 1$, the regression model in Equation (1) can be rewritten as shown in Equation (2).

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \,; \quad Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \tag{2}$$

In matrix notation, the model from the Equation (1) is expressed as in Equation (3).

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}} \underset{n\times p}{\boldsymbol{\beta}} + \underset{n\times 1}{\varepsilon} \tag{3}$$

Meanwhile, the OLS estimators are represented in Equation (4).

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Y} \tag{4}$$

Due to significant differences in the units of the variables, the first step in the regression process is to standardize the variables (Zulyanti & Noeryanti, 2022). The correlation transformation is used for this purpose, as cited in Kutner et al. (2005), it controls round-off errors and expresses regression coefficients in consistent units, facilitating comparison. In this transformation, the standardization process is modified, followed by an explanation of the resulting standardized regression model (Kutner et al., 2005).

Citing Kutner et al. (2005), standardizing a variable involves the variable's centering and scaling. Observations are centered by taking the difference between each observation and the mean of all observations for the variable, whereas observations are scaled by using the standard deviation of the observation for the variable. Accordingly, the correlation transformation of the dependent variable Y and the independent variables $X_1, \ldots, X_{p-1}$ are expressed in Equation (5) and (6) (Kutner et al., 2005).

$$Y_i^* = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i - \bar{Y}}{S_Y}\right) \qquad ; S_Y \left(\sqrt{\frac{\sum_i \left(Y_i - \bar{Y}\right)^2}{n-1}}\right) \tag{5}$$

$$X_{ik}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ik} - \bar{X}_k}{S_k}\right) \quad \left(k = 1, \ldots, p-1\right) \;\; ; S_k \left(\sqrt{\frac{\sum_i \left(X_{ik} - \bar{X}_k\right)^2}{n-1}}\right) \tag{6}$$

The model of standardized regression refers to the model of regression that uses the transformed variables Y^* and X^*, which are described by the correlation transformation in Equation (5) and (6). The model is shown in Equation (7) (Kutner et al., 2005).

$$Y_i^* = \beta_1^* X_1^* + \beta_2^* X_2^* + \ldots + \beta_n^* X_n^* + \varepsilon \tag{7}$$

The Equation (7) does not include an intercept parameter since OLS or maximum likelihood estimations would yield a zero intercept term if an intercept parameter were included. In the context of transformed variables, the expression $\mathbf{X}'\mathbf{Y}$ in Equation (4) is modified as follows:

$$\underset{(p-1)\times 1}{\mathbf{X}'\mathbf{Y}} = \mathbf{r}_{YX} \tag{8}$$

where $\mathbf{r}_{YX}$ represents the vector of simple correlation coefficients between *Y* and each *X* variable. The Equation (9), therefore, shows the estimators for the coefficients of regression for the standardized regression model in Equation (7) (Kutner et al., 2005).

$$\mathbf{b} = \mathbf{r}_{XX}^{-1}\mathbf{r}_{YX} \quad ; \underset{(p-1)\times 1}{\mathbf{b}} = \begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_{p-1}^* \end{bmatrix} \tag{9}$$

The regression coefficients $b_1^*,\ldots,b_{p-1}^*$ are commonly referred to as standardized regression coefficients. The estimated regression coefficients for the regression model in Equation (1) can be expressed in terms of the original variables by utilizing the following relationships in Equation (10) and (11) (Kutner et al., 2005).

$$b_k = \left(\frac{S_Y}{S_k}\right) b_k^* \qquad \left(k = 1,\ldots,p-1\right) \tag{10}$$

$$b_0 = \bar{Y} - b_1\bar{X}_1 - \ldots - b_{p-1}\bar{X}_{p-1} \tag{11}$$

Multicollinearity test

The term multicollinearity refers to a situation in which two or more independent variables in a dataset are correlated with each other, and if multicollinearity exists, it leads to instability in the parameter estimates of OLS regression (Montgomery et al., 2012; Shrestha, 2020; Yitshak-Sade et al., 2020). The multicollinearity test is, therefore, necessary. A common method for detecting multicollinearity is variance inflation factors, which show the extent to which the variances of regression coefficients increase when the predictors have a linear relationship (Kutner et al., 2005).

Referring to Kutner et al. (2005), the Equation (12) illustrates how multicollinearity is identified using a VIF (Variance Inflation Factor) value:

$$\left(VIF\right)_k = \left(1 - R_k^2\right)^{-1} \tag{12}$$

where R_k^2 is the determination coefficient. The largest VIF value among all independent variables is often used to assess the severity of multicollinearity. A maximum VIF value exceeding 10 is commonly interpreted as a sign that multicollinearity may significantly affect the OLS estimates (Kutner et al., 2005).

Jackknife Ridge Regression

Jackknife Ridge Regression (JRR) is a development from GRR. As noted in Singh et al. (1986) and Khurana et al. (2014), consider a matrix *G*, where the columns represent the normalized eigenvectors of the $(p \times p)$ matrix $X'X$. So, the Equation (3) can be written as:

$$\mathbf{Y} = Z\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{13}$$

with $Z = XG$ and $\boldsymbol{\gamma} = G'\boldsymbol{\beta}$. Also, $Z'Z = G'X'XG = \Lambda = diag(\lambda_1,\ldots,\lambda_p)$, with λ_i being the *i*-th eigenvalue of $X'X$. By adding a positive constant k_i to the *i*-th diagonal element of matrix Λ, the GRR estimator of $\boldsymbol{\gamma}$ can be expressed as:

10.12928/bamme.v5i1.12759

$$\hat{\boldsymbol{\gamma}}_{GRR} = (\Lambda + K)^{-1} Z'\mathbf{Y} = A^{-1}Z'\mathbf{Y} = A^{-1}\Lambda\mathbf{g} = (I - A^{-1}K)\mathbf{g} \tag{14}$$

with $\mathbf{g}$ is the OLS estimator of $\boldsymbol{\gamma}(\hat{\boldsymbol{\gamma}}_{OLS})$, which is expressed as $\mathbf{g} = \Lambda^{-1}Z'\mathbf{Y}$. Moreover, $A = \Lambda + K$ and $\mathbf{K}$ is the diagonal matrix consists of non-negative constants $\left(K = diag(k_1, ..., k_p); \hat{k}_i = \hat{\sigma}^2 / \hat{\gamma}_i^2; \hat{\sigma}^2 = \text{MSE}\right)$ (Hoerl & Kennard, 1970; Khurana et al., 2014; Singh et al., 1986). Then, since $\boldsymbol{\gamma} = G'\boldsymbol{\beta}$ and $GG' = I$, so the GRR estimator of β can be expressed as in Equation (15) (Khurana et al., 2014; Singh et al., 1986).

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{GRR} &= G\hat{\boldsymbol{\gamma}}_{GRR} \\ &= A_*^{-1}\mathbf{XY} \\ &= (\mathbf{X}'\mathbf{X} + K_*)\mathbf{XY} \\ &= (\mathbf{X}'\mathbf{X} + GKG')\mathbf{XY} \end{aligned} \tag{15}$$

From Singh et al. (1986) and Khurana et al. (2014), the JRR estimator of $\boldsymbol{\gamma}$ can be expressed as in Equation (16).

$$\hat{\boldsymbol{\gamma}}_{JRR} = (I + A^{-1}K)\hat{\boldsymbol{\gamma}} = \left[I - (A^{-1}K)^2 \mathbf{g}\right] \tag{16}$$

Meanwhile, the JRR estimator of β can be expressed as in Equation (17).

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{JRR} &= G\hat{\boldsymbol{\gamma}}_{JRR} \\ &= G\left[I + (A^{-1}K)\right]\left[I - (A^{-1}K)\right]\mathbf{g} \\ &= G\left[I + (A^{-1}K)\right](A^{-1}\Lambda)\Lambda^{-1}Z'\mathbf{Y} \end{aligned} \tag{17}$$

A JRR consists of the following steps (Hoerl & Kennard, 1970; Khurana et al., 2014; Montgomery et al., 2012; Singh et al., 1986; Tinungki, 2019):

a. Calculate the initial value of $\hat{k}$ using the values of $\hat{\sigma}^2$ and $\hat{\boldsymbol{\gamma}}$ from OLS estimator;
b. Calculate the initial GRR estimator $(\hat{\boldsymbol{\gamma}}_{GRR})$ in Equation (14);
c. Calculate the next value of $\hat{k}$ using the values of $\hat{\sigma}^2$ and $\hat{\boldsymbol{\gamma}}$ from GRR estimator;
d. Calculate the next GRR estimator;
e. Repeat step (c), until convergence or $\left|(\hat{\boldsymbol{\gamma}}'_{GRR}\hat{\boldsymbol{\gamma}}_{GRR})^i - (\hat{\boldsymbol{\gamma}}'_{GRR}\hat{\boldsymbol{\gamma}}_{GRR})^{i-1}\right| \le 0.0001$;
f. Calculate the JRR estimator of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ using Equation (16) and (17).

Data

This research was conducted from September 2024 to December 2024. Data is obtained from the document of the 2023 BAPPEDA Bone Bolango District (Dinas PUPR, 2023). Purposive sampling is employed in this study, in which samples are selected according to their relevance to the conditions of slums in strategic economic areas, namely Kabila, Suwawa, and Tilongkabila. All subdistricts classified as slums in three subdistricts were sampled, resulting in a total sample of 40 urban villages and villages. The total number of urban villages and villages for each subdistrict as follows. We selected 22 urban villages and villages in Kabila, 6 urban villages and villages in Suwawa, and 12 urban villages and villages in Tilonkabila.

The extent of slum conditions in an area can be measured by a slum value, which indicates the degree of unfitness of the living environment. This slum status is typically derived from assessing various indicators, as it's referred to in the 2023 document from the Office for Regional Planning in

Bone Bolango (Dinas PUPR, 2023).

The dependent variable used in the research is the slum value (Y). Meanwhile, the independent variables include poorly constructed buildings (X_1), buildings that fail to meet technical requirements (X_2), length of Ideal road (X_3), length of existing road (X_4), length of roads with damaged surfaces (X_5), lack of safe and quality drinking water (X_6), length of ideal drainage channel (X_7), length of existing drainage channel (X_8), length of damaged drainage channels (X_9), inadequate wastewater management systems that fail to meet with technical standards (X_{10}), inadequate wastewater management infrastructure and facilities that fail to meet technical requirements (X_{11}), inadequate waste infrastructure and facilities that fail to meet technical requirements (X_{12}), inadequate waste management systems that fail to meet technical requirements (X_{13}), lack of fire protection infrastructure (X_{14}), and lack of fire protection facilities (X_{15}).

Step analyses

Step analyses were applied to Bone Bolango's dataset as follows:

1. Multicollinearity check: VIF values exceeded 10 for 15 independent variables;
2. Correlation transformation: Variables standardized;
3. JRR estimates: Iteration stops when $\left|\left(\hat{\gamma}'_{GRR}\hat{\gamma}_{GRR}\right)^{i} - \left(\hat{\gamma}'_{GRR}\hat{\gamma}_{GRR}\right)^{i-1}\right| \leq 0.0001$; the flowchart is displayed in Figure 1;
4. Hypothesis testing: F test and t test for 15 independent variables.

In all steps, R software is used.

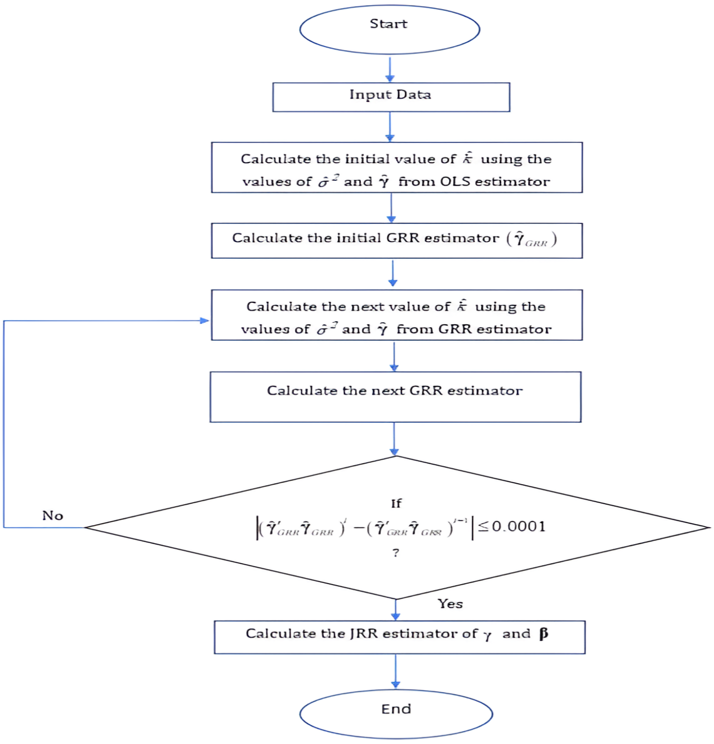

Figure 1. Flowchart of JRR Estimates

10.12928/bamme.v5i1.12759

Results and discussion

Multicollinearity test

The first step is to determine whether the data contain multicollinearity. This step was performed using the Equation (12). The result of the multicollinearity test is shown in Table 1.

Table 1. Result of the multicollinearity test

	X_1	X_2	X_3	X_4	$\cdots$	X_{12}	X_{13}	X_{14}	X_{15}
X_1	1	1.004	1.007	1.004	$\cdots$	1.175	1.217	1.009	1.075
X_2	1.004	1	1.000	1.002	$\cdots$	1.036	1.023	1.041	1.074
X_3	1.007	1.000	1	15.780*	$\cdots$	1.001	1.001	1.024	1.061
X_4	1.004	1.002	15.780*	1	$\cdots$	1.002	1.000	1.021	1.048
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_{12}	1.175	1.036	1.000	1.002	$\cdots$	1	1.337	1.002	1.117
X_{13}	1.217	1.023	1.001	1.000	$\cdots$	1.337	1	1.029	1.001
X_{14}	1.009	1.040	1.024	1.021	$\cdots$	1.002	1.029	1	1.064
X_{15}	1.075	1.074	1.061	1.048	$\cdots$	1.117	1.001	1.064	1

Table 1 shows that the length of ideal road (X_3) and length of existing road (X_4) have VIF values greater than 10 (15.780). Hence, the result indicates a highly collinear relationship between these two independent variables. Refer to the document of the 2023 BAPPEDA Bone Bolango District, to calculate the length of ideal road is by summing the length of existing road and the projected new road needs (Dinas PUPR, 2023). So, by default, the nearly identical calculation method leads to a near-perfect correlation between the two variables. Despite the potential for bias, the decision to retain both variables is made to preserve policy-relevant benchmarks for infrastructure planning, as removing one of these variables might lead to loss of valuable information and skewed or misleading results (Kutner et al., 2005; Montgomery et al., 2012; Shrestha, 2020).

Standardized variables

Standardizing variables is the next step in the analysis. The correlation transformation is used for this step, which is written in Equation (5) and (6).

Table 2. Result of the standardized variables

Y	X_1	X_2	X_3	$\cdots$	X_{12}	X_{13}	X_{14}	X_{15}
0.552	0.113	0.017	0.195	$\cdots$	-0.132	0.050	-0.158	0.648
0.762	-0.229	-0.918	0.710	$\cdots$	0.364	1.100	-0.158	1.182
0.341	-0.314	-0.517	-1.120	$\cdots$	-1.207	-1.246	-0.158	-0.057
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.551	-0.229	0.951	2.716	$\cdots$	-0.049	0.653	-0.158	0.840
0.341	-0.101	-0.517	1.889	$\cdots$	-1.166	-0.554	-0.158	-0.078
2.232	-0.272	3.621	-1.060	$\cdots$	-0.938	-0.308	-0.158	0.178

The results of the standardized variables test are shown in Table 2.

JRR estimation

The first step is to conduct an OLS estimator. The result is used to calculate the initial value of $\hat{k}$ using the values of $\hat{\sigma}^2$ and $\hat{\gamma}$ from OLS estimator. The result is shown in Table 3.

Table 3. *Result* of initial values of $\hat{\sigma}^2$, $\hat{\gamma}$, and $\hat{k}$

Variables	$\hat{\sigma}^2$	$\hat{\gamma}$	$\hat{k}$
X_1		0.045	133.652
X_2		0.011	2015.646
X_3		0.183	7.925
⋮	0.265	⋮	⋮
X_{13}		-0.361	2.034
X_{14}		-0.160	10.419
X_{15}		-0.516	0.997

The next step is to conduct a GRR estimator. The result is used to calculate the next value of $\hat{k}$ using the values of $\hat{\sigma}^2$ and $\hat{\gamma}$ from GRR estimator using Equation (14). The result is shown in Table 4.

Table 4. Result of next values of $\hat{\sigma}^2$, $\hat{\gamma}$, and $\hat{k}$ using GRR estimator

Variables	$\hat{\sigma}^2_{GRR}$	$\hat{\gamma}_{GRR}$	$\hat{k}$
X_1		0.022	793.465
X_2		0.001	1062088
X_3		0.167	13.165
⋮	0.367	⋮	⋮
X_{13}		-0.266	5.206
X_{14}		-0.047	168.549
X_{15}		-0.197	9.445

The process of iteration continues until $\left|\left(\hat{\boldsymbol{\gamma}}'_{GRR}\hat{\boldsymbol{\gamma}}_{GRR}\right)^i-\left(\hat{\boldsymbol{\gamma}}'_{GRR}\hat{\boldsymbol{\gamma}}_{GRR}\right)^{i-1}\right|\leq 0.0001$ (Hoerl & Kennard, 1970; Khurana et al., 2014; Montgomery et al., 2012; Singh et al., 1986; Tinungki, 2019). The result showed that the value of $\left|\left(\hat{\boldsymbol{\gamma}}'_{GRR}\hat{\boldsymbol{\gamma}}_{GRR}\right)^i-\left(\hat{\boldsymbol{\gamma}}'_{GRR}\hat{\boldsymbol{\gamma}}_{GRR}\right)^{i-1}\right|$ is equal to 0.0000 in the fifth iteration; thus, the iteration process stopped at fifth (as it is already convergence, less than 0.0001). The last step is to calculate the JRR estimator of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ using Equation (16) and (17). The result is shown in Table 5.

Table 5. Result of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}$ using JRR estimator

Variables	$\hat{\gamma}_{JRR}$	$\hat{\beta}_{JRR}$
X_1	0.045	0.131
X_2	0.011	0.035
X_3	0.183	-0.539
⋮	⋮	⋮
X_{13}	-0.361	0.268
X_{14}	0.000	0.081
X_{15}	-0.516	0.262

According to the result shown in Table 5, the regression model for the estimator of JRR is as follows in Equation (18).

$$\begin{aligned}\hat{Y}=&-1.475\times 10^{-17}+0.131X_1+0.035X_2-0.539X_3+0.178X_4\\&+0.459X_5+0.232X_6+0.047X_7-0.373X_8+0.291X_9+0.182X_{10}\\&+0.110X_{11}-0.636X_{12}+0.268X_{13}+0.081X_{14}+0.262X_{15}\end{aligned} \quad (18)$$

 10.12928/bamme.v5i1.12759

Hypothesis tests

The first hypothesis test conducted is the simultaneous regression test, commonly known as the F test. This test evaluates whether the independent variables significantly affect the dependent variables simultaneously. The null hypothesis suggests that all regression coefficients for the independent variables are equal to zero $\left(\beta_1 = \beta_2 = \ldots = \beta_k = 0\right)$, indicating no significant effect on the dependent variable (Priyono, 2021). The F test statistic is calculated using Equation (19) (Montgomery et al., 2012).

$$F = \frac{SSR / k}{SSE / (n-k-1)} = \frac{MSR}{MSE} \tag{19}$$

The null hypothesis is rejected if the calculated *F* value exceeds $F_{(k;(n-(k+1));\alpha)}$ (Montgomery et al., 2012). The result of the calculated *F* value is 8.04, so it rejected the null hypothesis, as $F_{(15;24;0.05)}$ is equal to 2.11. These results indicate that at least one independent variable significantly affects the slum value (*Y*) in the Bone Bolango District.

The next hypothesis test is the partial regression test, commonly known as the t test. This test evaluates the significance of the effect of each independent variable on the dependent variable individually. The null hypothesis asserts that the parameter $\beta_i \left(i = 1, 2, \ldots k\right)$ equals zero, indicating that the independent variable has no significant effect on the dependent variable (Priyono, 2021). The t test statistic is calculated using Equation (20) (Montgomery et al., 2012).

$$t = \frac{\hat{\beta}_i}{se\left(\hat{\beta}_i\right)} \tag{20}$$

The null hypothesis is rejected if the calculated $|t|$ value exceeds $t_{\left(n-(k+1);\alpha/2\right)}$ (Montgomery et al., 2012). The result of the calculated $|t|$ is shown in Table 6.

Table 6. Result of the $|t|$ test for JRR and OLS

JRR				OLS			
Variables	$\|t\|$	Variables	$\|t\|$	Variables	$\|t\|$	Variables	$\|t\|$
X_1	1.251	X_9	2.514*	X_1	1.328	X_9	2.413*
X_2	0.279	X_{10}	1.508	X_2	0.596	X_{10}	1.103
X_3	1.194	X_{11}	0.647	X_3	-1.162	X_{11}	1.116
X_4	0.399	X_{12}	5.236*	X_4	0.497	X_{12}	-5.467*
X_5	2.674*	X_{13}	2.066*	X_5	2.351*	X_{13}	2.264*
X_6	2.131*	X_{14}	0.516	X_6	1.913	X_{14}	0.120
X_7	0.262	X_{15}	2.368*	X_7	0.003	X_{15}	2.510*
X_8	2.180*			X_8	-1.993		

Based on the results shown in Table 6, seven variables, namely X_5, X_6, X_8, X_9, X_{12}, X_{13}, and X_{15}, rejected the null hypothesis, as $t_{(24;0.025)}$ is equal to 2.064. These results indicate that when using JRR, the length of roads with damaged surfaces (X_5), lack of safe and quality drinking water (X_6), length of existing drainage channel (X_8), length of damaged drainage channels (X_9), inadequate waste infrastructure and facilities that fail to meet technical requirements (X_{12}), inadequate waste management systems that fail to meet technical requirements (X_{13}), and lack of fire protection facilities (X_{15}) significantly affect the slum value (*Y*) in Bone Bolango District.

In contrast, in Table 6, only five variables rejected the null hypothesis when using OLS, excluding lack of safe and quality drinking water (X_6) and length of existing drainage channel (X_8). Based on the results, OLS missed some important variables.

These results indicated that JRR estimation is preferable when datasets contained multicollinearity and small datasets, such as the slums condition in the Bone Bolango District, compared to OLS. Accordingly, it is recommended that JRR be used in future research that analyzes similar datasets.

Furthermore, according to the results in Table 6, another JRR is conducted to obtain a new model by eliminating significant independent variables that do not affect the dependent variable. Following this, a transformation of JRR into original variables is performed using Equation (10) and (11). The result is shown in Equation (21).

$$\begin{aligned}\hat{Y} = & -3.597\times 10^{-17} + 0.178X_5 + 0.046X_6 - 0.047X_8 \\ & -0.373X_9 + 0.110X_{12} - 0.636X_{13} + 0.081X_{15}\end{aligned} \tag{21}$$

According to Equation (21), X_5, which represents the length of roads with damaged surfaces, has a coefficient of 0.178. Accordingly, for every unit increase in length of roads with damaged surfaces, slum values increase by 0.178. Research conducted by Qonita and Rahmawati (2021) also indicates that greater road damage is associated with a greater incidence of slum conditions. The cause of this relationship may be attributed to factors such as diminished accessibility, an increased risk of accidents, or a decline in the environmental quality of surrounding neighborhoods.

Coefficient of determination

The coefficient of determination (R^2), which ranges from 0 to 1, measures the effectiveness of the independent variable in explaining the dependent variable, with a higher value indicating a more accurate model (Adityaningrum et al., 2023; Najib et al., 2024). The coefficient of determination is calculated using the Equation (22) (Montgomery et al., 2012).

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{22}$$

The result of the calculated is displayed in Table 7.

Table 7. Calculated of R^2 for JRR and OLS

	JRR	OLS
R^2	0.830	0.735

As indicated in Table 7, the JRR model shows a higher R^2 value of 0.830, which is higher than the OLS model R^2 value of 0.735. This shows that the JRR model improves both accuracy and reliability by reducing the effect of multicollinearity in the dataset. The model explains 83% of the variability in strength. While the remaining 17% is influenced by factors that are not included in this research. For example, previous research has identified several factors related to slums, including spatial planning, building density, and population density (Irawan et al., 2020). Additionally, Afrina et al. (2021) highlighted land legality. Furthermore, Annas et al. (2018) elaborated on the government's role, access to green open spaces, social conditions, economic conditions, and the role of the community.

The results show that the Bone Bolango government should prioritize all seven independent variables significantly related to slum conditions. Making changes or fixing these factors can greatly impact daily life and health.

10.12928/bamme.v5i1.12759

Conclusion

The analysis results indicate a highly collinear relationship between these two independent variables. Therefore, JRR is performed. The iteration process stopped at the fifth. The result also shows key factors contributing to slum conditions using JRR. Significant variables identified include the length of roads with damaged surfaces (X_5), lack of safe and quality drinking water (X_6), length of existing drainage channel (X_8), length of damaged drainage channels (X_9), inadequate waste infrastructure and facilities that fail to meet technical requirements (X_{12}), inadequate waste management systems that fail to meet technical requirements (X_{13}), and lack of fire protection facilities (X_{15}). The model explains 83% of the variability in slum conditions, indicating the importance of these factors. However, 17% of the variance is unaccounted for, suggesting other factors may be involved.

By applying JRR to the slum dataset, this research showed how the modeling method can effectively tackle multicollinearity without eliminating the variable. This research demonstrates its effectiveness in slum datasets that are small in sample size and high in dimensionality, which has never been done before. This method offers a replicable framework for other resource-limited areas where traditional regression methods may not perform adequately.

References

Adityaningrum, A., Ladjali, S. I., Djakaria, I., Yahya, L., Payu, M. R. F., Nashar, L. O., & Jusuf, H. (2023). Comparing Gaussian kernel and quadratic spline of nonparametric regression in modeling infectious diseases. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, *17*(4), 2135–2146. https://doi.org/10.30598/barekengvol17iss4pp2135-2146

Afrina, S., Fuady, Z., & Yusuf, M. A. (2021). Identifikasi faktor penyebab utama kekumuhan permukiman di Dusun Tengku Muda, Gampong Jawa, Kecamatan Kuta Raja, Kota Banda Aceh. *Jurnal Ilmiah Mahasiswa Arsitektur Dan Perencanaan*, *5*(2), 58–64.

Annas, A., Soemitro, R. A. A., & Santoso, E. B. (2018). Analisis penentuan infrastruktur prioritas pada kawasan kumuh lingkungan kerantil Kota Blitar. *Journal of Civil Engineering*, *33*(2), 56–67.

Arifin, C., Muhibuddin, A., & Bahri, S. (2022). Evaluasi program kotaku dalam penanganan kumuh Kota Parepare: Studi kasus Kelurahan Ujung Lare Kecamatan Soreang. *Urban and Regional Studies Journal*, *5*(1), 28–43. https://doi.org/10.35965/ursj.v5i1.1965

Arrasyid, A. H., Ispriyanti, D., & Hoyyi, A. (2021). Metode modified Jackknife Ridge Regression dalam penanganan multikolinieritas (Studi Kasus Indeks Pembangunan Manusia Di Jawa Tengah). *Jurnal Gaussian*, *10*(1), 104–113.

Fatihah, A. I. (2022). Penerapan Metode Jackknife Ridge Regression dalam Penanganan Multikolinearitas (Studi Kasus Tingkat Kemiskinan di Kabupaten Grobogan 2011-2021) [Universitas Islam Negeri Walisongo] https://eprints.walisongo.ac.id/id/eprint/20479/

Gobel, F. F., & Zees, E. A. T. (2022). Faktor pengaruh dalam pembentukan permukiman kumuh di Desa Pohuwato Timur. *ARSITEKTURA*, *20*(1), 99. https://doi.org/10.20961/arst.v20i1.55383

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55. https://doi.org/10.2307/1267351

Irawan, R. C., Nugraha, A. L., & Firdaus, H. S. (2020). Analisis tingkat kekumuhan pada permukiman menggunakan model Geographically Weighted Regression (GWR)(Studi Kasus: Kota Surakarta, Jawa Tengah). *Jurnal Geodesi Undip*, *9*(2), 145–154.

Kabupaten Bone Bolango. (2021). *Peraturan Daerah Nomor 5 Tahun 2021 tentang Rencana Tata Ruang Wilayah Kabupaten Bone Bolango Tahun 2021–2041*. Pemerintah Kabupaten Bone Bolango: Bone Bolango.

Khurana, M., Chaubey, Y. P., & Chandra, S. (2014). Jackknifing the Ridge Regression Estimator: A Revisit. *Communications in Statistics - Theory and Methods*, *43*(24), 5249–5262. https://doi.org/10.1080/03610926.2012.729640

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models*. McGrawhill.

Madrid-García, A., Merino-Barbancho, B., Rodríguez-González, A., Fernández-Gutiérrez, B., Rodríguez-Rodríguez, L., & Menasalvas-Ruiz, E. (2022). *Understanding the role and adoption of artificial intelligence techniques in rheumatology research: an in-depth review of the literature*. https://doi.org/10.1101/2022.11.04.22281930

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

Najib, M. K., Fauzan, M. D., Nurdiati, S., Khoerunnisa, N., Maulia, S. D., Triwulandari, R. R. C., & Aziz, M. F. (2024). Prediksi Angka Harapan Hidup Menggunakan Regresi Linear Berganda, Lasso, Ridge, Elastic Net, dan Kuantil Lasso. *Jurnal Sains Matematika Dan Statistika*, *10*(2). https://doi.org/10.24014/jsms.v10i2.27916

Prajamandana, A. P., Sukmono, A., & Firdaus, H. S. (2021). Analisis tingkat permukiman kumuh menggunakan geographically weighted regression semiparametric (GWRS)(Studi Kasus: Kota Bogor, Jawa Barat). *Jurnal Geodesi Undip*, *10*(3), 41–51.

Priyono, M. S. (2021). *Analisis Regresi dan Korelasi untuk Penelitian Survei (Panduan Praktis Olah Data dan Interpretasi: D*. GUEPEDIA.

PUPR. (2023). *Dokumen Rencana Pencegahan dan Peningkatan Kawasan Perumahan Kumuh dan Kawasan Permukiman Kumuh (RP2KPKPK) Kabupaten Bone Bolango*. Kementrian Pekerjaan Umum dan Perumahan Rakyat

Qonita, C. D., & Rahmawati, D. (2021). Strategi Penanganan Pemukiman Kumuh di Area Pendukung Industri Kelurahan Krian, Kabupaten Sidoarjo. *Jurnal Teknik ITS*, *9*(2). https://doi.org/10.12962/j23373539.v9i2.56302

Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39–42. https://doi.org/10.12691/ajams-8-2-1

Singh, B., Chaubey, Y. P., & Dwivedi, T. D. (1986). An almost unbiased ridge estimator. *Sankhyā: The Indian Journal of Statistics, Series B*, 342–346.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tinungki, G. M. (2019). Orthogonal iteration process of determining K value on estimator of Jackknife ridge regression parameter. *Journal of Physics: Conference Series*, *1341*(9), 092001. https://doi.org/10.1088/1742-6596/1341/9/092001

Winda, B. M. (2021). *Penerapan metode jackknife ridge regression dalam kasus multikolinieritas pada indeks pembangunan manusia di kabupaten/kota provinsi jawa tengah* [Universitas Andalas]. http://scholar.unand.ac.id/72586/

Xie, Y.-L., & Kalivas, J. H. (1997). Local prediction models by principal component regression. *Analytica Chimica Acta*, *348*(1–3), 29–38. https://doi.org/10.1016/S0003-2670(97)00036-6

Yitshak-Sade, M., Fabian, M. P., Lane, K. J., Hart, J. E., Schwartz, J. D., Laden, F., James, P., Fong, K. C., Kloog, I., & Zanobetti, A. (2020). Estimating the Combined Effects of Natural and Built Environmental Exposures on Birthweight among Urban Residents in Massachusetts. *International Journal of Environmental Research and Public Health*, *17*(23), 8805. https://doi.org/10.3390/ijerph17238805

Zulyanti, T., & Noeryanti, N. (2022). Perbandingan pengelompokan usaha mikro kecil dan menengah di Kabupaten Klaten Tahun 2019 dengan Metode k-Means dan Clustering Large Application. *Jurnal Statistika Industri Dan Komputasi*, *7*(01), 46–59.

10.12928/bamme.v5i1.12759