e-ISSN: 2714-5506
p-ISSN: 2714-609X

SQUARE : Journal of Mathematics and Mathematics Education
Volume 7, No. 1, 2025, pp. 1-11
http://dx.doi.org/10.21580/square.2025.7.1.28037

Handling Multicollinearity using Least Absolute Shrinkage and Selection Operator Regression on Per Capita Expenditure Data

Della Setya Rahma[1], Eva Khoirun Nisa[2,*]
[1,2] UIN Walisongo Semarang
[*] evakn@walisongo.ac.id

ABSTRACT

Multicollinearity is a problem that must be addressed when using regression. Multicollinearity often occurs in socioeconomic data, such as Per Capita Expenditure. Several relevant studies have shown that Least Absolute Shrinkage and Selection Operator (LASSO) regression is a good method for handling multicollinearity. Additionally, it produces the simple model. Meanwhile, the Least Angle Regression (LAR) algorithm works effectively in model optimization, especially when multicollinearity occurs in multiple variables. Therefore, this study aims to handle multicollinearity with LAR LASSO regression in the specific case of per capita expenditure data in Wonosobo with many variables experiencing multicollinearity. The result study is LAR LASSO regression successfully eliminated two of the four predictor variables that exhibited multicollinearity by reducing the regression coefficients on the two predictor variables to zero. The best regression model obtained produces two significant coefficients so that Per Capita Expenditure in Wonosobo was influenced by the Human Development Index and Average Years of Schooling.
Keywords*: LASSO regression, LAR, multicollinearity, per capita expenditure.*

1. INTRODUCTION

Multicollinearity is a problem that often arises in regression analysis. Multicollinearity occurs when the independent variables in the regression model have a strong linear relationship, indicated by a high correlation coefficient or even a value of one (Gujarati & Porter, 2009)). A good regression model should not have correlation in the independent variables because it can affect the results of the model estimation (Ghozali, 2011).

One of the methods that can overcome multicollinearity is the Least Absolute Shrinkage and Selection Operator (LASSO) regression. Rahmawati and Suratman (2022) compared LASSO regression with Ridge and Elastic Net. Their research found LASSO regression to be the most effective method with the simplest model among the others. Meanwhile, Nasution & Pane (2024) compared it with Principal Component Regression (PCR). The results showed that LASSO regression performed better than PCR. Therefore, LASSO regression is used in this study.

LASSO regression can shrink the regression coefficient to exactly zero or close to zero, in addition to simplifying the model through the variable selection process (Tibshirani, 2011). In its implementation, LASSO regression can be solved using the Least Angle Regression (LAR) algorithm which makes it easy to understand the order in which variables enter the model and easy to calculate manually (Hastie et al., 2009)

LASSO regression is widely applied to determine the factors that influence the dependent variable but experience multicollinearity. One example of its application is to determine the factors that are suspected of influencing population expenditure (Rahayu & Husein, 2023). Although many studies have used LASSO regression in a national or global context, studies that focus on specific areas such as Wonosobo Regency are still very limited. It shows that the application of the model in a local context can provide new insights into the factors suspected of influencing per capita expenditure in Wonosobo Regency.

The per capita expenditure condition in Wonosobo Regency fluctuated in 2015–2024. After increasing until 2018, there was a decline in 2019 and 2020 due to rising commodity prices, inflation, and the impact of the pandemic. Although it recovered in 2021–2023, the figure declined again in 2024. This decline was influenced by an increase in the number of workers that was not accompanied by the availability of jobs, as well as the low quality of education and skills, which caused a mismatch between graduates and the needs of the labor market (Badan Pusat Statistik, 2025)

Based on these conditions, an analysis of the factors causing per capita expenditure in Wonosobo using LASSO regression. The factors suspected of influencing per capita expenditure are taken from the three main dimensions of the Human Development Index, namely longevity and healthy living, knowledge, and a decent standard of living (BPS, 2022). Longevity and healthy living can be measured by life expectancy, knowledge can be represented by schooling expectancy and average years of schooling, and a decent standard of living can be seen from per capita expenditure (Ginting & Lubis, 2023). All variables are theoretically correlated, so multicollinearity management is necessary. One good statistical method for handling multicollinearity is LASSO regression.

Several previous studies have applied LASSO to handle multicollinearity. (Jamco et al., 2023) addressed multicollinearity in the gross regional domestic product percentage in Maluku. Rahmawati et al. (2022) compared the performance of LASSO regression with Ridge regression in overcoming multicollinearity. The results showed that LASSO regression was better, especially in selecting predictor variables, resulting in a simpler model and no multicollinearity. Nasution & Pane (2024) compared the performance of LASSO with Principal Component Regression (PCR). LASSO regression was found to be more efficient than PCR because it produced MSE value was smaller than PCR. Based on these advantages, LASSO regression will be used in this study to handle multicollinearity in per capita expenditure in Wonosobo.

2. METHOD

The data is secondary data obtained from Central Statistics Agency (BPS) Wonosobo Regency in 2024. This research uses one response variable (Y) and four predictor variables (X). A detailed explanation of the variables is provided in Table 1.

Table 1. Research Variables

Variables	Symbol	Unit
Per Capita Expenditure	Y	Rupiah
Human Development Index	X_1	Percent
Average Years of Schooling	X_2	Year
School Life Expectancy	X_3	Year
Life Expectancy	X_4	Year

Determination of predictor variables based on the results of several studies which state that Human Development Index (Febriani Sagala et al., 2024), Average Years of Schooling (Sianturi et al., 2024), School Life Expectancy (Manurung & Hutabarat, 2021), and Life Expectancy (Nizar & Arif, 2023) are suspected of having an influence on per capita expenditure.

This research began with a multicollinearity detection on the per capita expenditure data in multiple linear regression. Multicollinearity detection to show there is a linear relationship between one or more independent variables in the regression model (Kadir, 2008). The multicollinearity test is carried out by Variance Inflation Factor (VIF). The following is the VIF formula (Mubarak, 2021)):

$$VIF_k = \frac{1}{1 - R_k^2} \tag{1}$$

with R_k^2 being the value of the coefficient of determination and k is number of predictor variables. If VIF > 10 then there is multicollinearity (Nisa & Maulina, 2024).

Once multicollinearity was identified, the process was continued with LASSO regression using the LAR algorithm. The first step in handling multicollinearity with LASSO regression is to determine the LASSO regression estimator. With penalty condition $\sum_{k=1}^{p} |\beta_k| \leq t$ (Hastie et al., 2009) formula LASSO regression estimator is (Tibshirani, 2011)

$$\hat{\beta}^{lasso} = \arg min \sum_{i=1}^{n} (Y_i - \beta_0 - \sum_{k=1}^{p} \beta_k X_{ik})^2 \quad (2)$$

where

Y_i : Response variable for the $i - th$ observation,
β_0 : Constant in the regression model,
β_k : Regression coefficient of the $k - th$ predictor variable,
X_{ik} : Value of the $k - th$ independent variable for the $i - th$ observation,
i : $1, 2, \ldots, n$; n is the number of observations,
k : $1, 2, \ldots, p$; p is the number of predictor variables.

The t parameter is a tuning parameter used to control the amount of shrinkage in the LASSO regression coefficient.

$$t = \sum_{k=1}^{p} |\hat{\beta}_k| \quad (3)$$

It acts as a limit on the total shrinkage of the coefficient by providing constraints so that the sum of the absolute values of all regression coefficients does not exceed t, provided that: $t \geq 0$. The following are the terms of analysis:

1. If the value $t < t_0$ with $t_0 = \sum_{k=1}^{p} |\hat{\beta}_k^0|$, the multiple linear regression coefficients will shrink towards zero or exactly at zero. This is a desirable property of LASSO regression.
2. If the value $t \geq t_0$, the LASSO regression coefficient gives the same results as the multiple linear regression coefficient.

The second step is apply Least Angle Regression (LAR) algorithm with the following algorithm (Efron et al., 2004):

1. Find a vector that is proportional to the correlation vector between the independent variables and the residuals of each independent variable, namely:
$$\hat{\boldsymbol{\mu}} = \mathbf{X}'(\mathbf{Y} - \hat{\boldsymbol{\mu}}) \quad (4)$$
2. Determine the largest correlation value of the correlation vector, through:
$$\mathrm{C} = \max\{|\hat{c}_k|\} \quad (5)$$
So it is obtained $s_j = \text{sign}\ \{|\hat{c}_k|\}$ for $k \in A$.
3. Determine $\boldsymbol{X_A}$, where **A** is a collection of indices of active variables. Active variables are independent variables that are currently being used in the model because they have the largest correlation value. These variables are collected into a submatrix called $\boldsymbol{X_A}$, which is defined as:
$$\boldsymbol{X_A} = (\ldots s_k \boldsymbol{X_k} \ldots)_{k \in A} \quad (6)$$
with a sign $s_k \pm 1$, so $\boldsymbol{G_A}$ and A_A can be written as:
$$\boldsymbol{G_A} = \boldsymbol{X_A}'\boldsymbol{X_A} \text{ dan } A_A = (1'_A \boldsymbol{G_A}^{-1} \boldsymbol{1_A})^{-\frac{1}{2}} \quad (7)$$
4. Calculate the equiangular vector, which is a unit vector that forms the same angle to each active variable in the model. This vector determines the direction of change of the regression coefficient when a variable is added, by ensuring that the columns in $\boldsymbol{X_A}$ have the same angle (less than 90°) to the vector. The value of the equiangular vector is found using the following formula:

$$\boldsymbol{u_A} = \boldsymbol{X_A W_A} \text{ dengan } \boldsymbol{W_A} = A_A \boldsymbol{G_A}^{-1} \boldsymbol{1_A} \tag{8}$$

5. Calculating the inner product vector

$$\boldsymbol{a} \equiv \boldsymbol{X' u_A} \tag{9}$$

6. Model Prediction Updates and New Variable Selection

$$\widehat{\boldsymbol{\mu}}_{A+} = \widehat{\boldsymbol{\mu}}_A + \widehat{\boldsymbol{\gamma}} \boldsymbol{u}_A \text{ dan } \boldsymbol{A}_+ = A - \{k\} \tag{10}$$

7. Optimal steps $\hat{\gamma}$ can be obtained from the following equation:

$$\hat{\gamma} = min^{+}_{k \epsilon A^c} \{ \frac{\widehat{\boldsymbol{C}} - \widehat{\boldsymbol{c}}_{\boldsymbol{k}}}{\boldsymbol{A_A} - \boldsymbol{a_k}}, \frac{\widehat{\boldsymbol{C}} + \widehat{\boldsymbol{c}}_{\boldsymbol{k}}}{\boldsymbol{A_A} + \boldsymbol{a_k}} \} \tag{11}$$

min^{+} shows the smallest positive selection γ_k, counted for all $k \epsilon A^c$. The goal is to determine which new variables are most eligible to be added to the model.

8. In the LAR algorithm for LASSO, the sign of the coefficient $\hat{\beta}_k$ must be the same as the correlation sign $\hat{c}_k$, yaitu: $sign\ (\hat{\beta}_k) = sign\ (\hat{c}_k) = s_k$. This means, that if $\hat{c}_k > 0$, so $\hat{\beta}_k$ must also be positive, and vice versa. This provision ensures that the sign of the active variable coefficient remains consistent with the direction of its correlation to the residual, so that the stability of the model is maintained throughout the process.

9. This step is carried out when there is a potential change in sign in the active variable coefficient, namely when $\tilde{\gamma}$, where $\beta_k(\gamma)$ to zero, smaller than the optimal step $\hat{\gamma}$. Mark $\tilde{\gamma}$ calculated as:

$$\tilde{\gamma} = min_{\gamma_k > 0} \{ \frac{-\hat{\beta}_k}{d_k} \} \tag{12}$$

with $d_k = s_k w A_k$. If $\tilde{\gamma} < \hat{\gamma}$, then the coefficient will be zero before the addition of the new variable, so the process must be stopped to prevent a sign change, according to the LASSO rule. The variable is then removed from the active set, but can still be re-entered if it meets the requirements in the next step.

10. Calculate the candidate value of the LASSO coefficient, calculated using the formula

$$\beta_k^{(t+1)} = \beta_k^{(t)} + \hat{\gamma} \times d_k \tag{13}$$

11. This process is repeated until all relevant variables are included in the model or there are no more variables to select, signaling the completion of the variable selection stage.

The third step is to apply cross validation. One method for estimating the performance and robustness of models is cross-validation (CV) (Berrar, 2019). Cross validation divides training data and testing data. Training data is used to build models $\hat{\beta}$ and testing data is used to test the goodness of prediction $X\hat{\beta}$ (Soleh & Aunuddin, 2013). The following is a cross-validation estimate (Hastie et al., 2009):

$$CV = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i) \tag{14}$$

One of the cross-validation methods is *k*-folds. *K*-fold cross validation has been used in numerous case studies and is a simple method for determining the relative effectiveness of different models (Aprihartha & Idham, 2024). Using k-folds will produce an estimate of c of the test error $\mathrm{MSE}_1, \mathrm{MSE}_2, \ldots, \mathrm{MSE}_k$. The selection of the number of folds (c) commonly used in K-fold cross-validation is usually $c = 5$ or $c = 10$ because it produces a small data distribution towards the average value (Wijiyanto et al., 2024).

The MSE value can be calculated using the formula (Kutner et al., 2005):

$$MSE = \frac{\sum\left(Y_i - \hat{Y}_i{}^2\right)}{n-1} = \frac{\sum e_i^2}{n-1} \tag{15}$$

In LASSO regression, the best model is selected based on the minimum MSE value and optimal parameters, which are determined through cross-validation to improve model accuracy (Jamco et al., 2023).

The last step is to re-detect multicollinearity with Variance Inflation Factor (VIF). After obtaining the best-performing regression model and applying it to the data, multicollinearity detection is continued. Multicollinearity detection is carried out in the same way as the initial stage, using VIF. It is hoped that the multicollinearity that occurs can be effectively addressed.

3. RESULT AND DISCUSSION

The first thing to do before checking for multicollinearity is to form a multiple linear regression model on per capita expenditure data. The following is the estimated multiple linear regression model:

$$\hat{Y} = 30{,}425 + 1{,}589X_1 + 1{,}037X_2 + 2{,}395X_3 - 2{,}299X_4. \tag{16}$$

After the model estimator is obtained, the next step is to detect multicollinearity. The following are the VIF values for each predictor variable:

Table 2. VIF on each predictor variable

Variable	VIF
X_1	31, 246585
X_2	9, 073431
X_3	4, 734037
X_4	26, 075121

Based on Table 2, the VIF value for X_1 and X_2 has a VIF value of more than 10. It is a multicollinearity problem. After multicollinearity is detected, data standardization is carried out. By defining the standard content of data measurement points, which covered data collection, monitoring and alarm, data analysis and optimization, etc., the data was standardized (Zhang et al., 2024). Table 3 displays the results of standardization of per capita expenditure data in Wonosobo (Y) with four predictor variables.

Table 3. Data Standardization Results

Observation	Y	X_1	X_2	X_3	X_4
1	0.7764	1.4687	0.8365	0.8984	1.5465
2	1.441	1.0997	0.8036	0.8077	1.0441
3	0.8252	0.7061	0.7707	0.6262	0.782
4	0.4049	0.3289	0.5731	0.4445	0.5418
5	−0.1892	0.1567	0.5401	0.354	0.2796
6	0.198	0.1977	0.0791	0.2632	−0.201
7	0.4502	−0.1796	0.3425	−0.1906	−0.5068
8	−1.0409	−0.9341	−0.4479	−0.2814	−0.8563
9	−1.3631	−1.5081	−1.7325	−0.3721	−1.1621
10	−1.5024	−1.3359	−1.7652	−2.55	−1.4679

Furthermore, estimate the LASSO regression parameters using the LAR algorithm. The steps are as follows:

1. Determine a vector that is comparable to the correlation vector between the predictor variables and the residuals in the first selection. The value obtained is:

$$\hat{\boldsymbol{\mu}} = \mathbf{X}'(\mathbf{Y} - \hat{\boldsymbol{\mu}}) = \begin{bmatrix} 8{,}4040 \\ 8{,}2222 \\ 7{,}0884 \\ 7{,}9294 \end{bmatrix}$$

From the vector, the largest correlation value of the correlation vector is $\mathrm{C} = \max\{|\hat{\boldsymbol{c}}_k|\} = 8{,}4040$. This result explaines that the first active variable is X_1, which contains data from variables X_1 which has been standardized. For the value $\boldsymbol{G_A} = 9{,}00$ and $A_A = 3{,}00$.

2. Determine the equiangular vector. The procedure begins by determining the weight value of the active variable $\boldsymbol{W_A}$. Vector size $\boldsymbol{W_A} = 0{,}3333$. will change the active variables increase. The equiangular vector ($\boldsymbol{u_A}$) for the first active variable entered as follows

$$\boldsymbol{u_A} = \begin{bmatrix} 0,4896 \\ 0,3666 \\ 0,2354 \\ 0,1097 \\ 0,0523 \\ 0,0659 \\ -0,0598 \\ -0,3114 \\ -0,5026 \\ -0,4453 \end{bmatrix}$$

3. Find out the inner product vector for the first active variable $\boldsymbol{a}$.

$$\boldsymbol{a} \equiv \boldsymbol{X}'\boldsymbol{u_A} = \begin{bmatrix} 0,7191 \\ 0,4031 \\ 0,1662 \\ 0,0361 \\ 0,0082 \\ 0,0130 \\ 0,0108 \\ 0,2908 \\ 0,7581 \\ 0,5949 \end{bmatrix}$$

4. Calculate the prediction vectors $\hat{\boldsymbol{\mu}}_{A+} = \hat{\boldsymbol{\mu}}_A + \hat{\boldsymbol{\gamma}}\boldsymbol{u}_A$. This equation need $\hat{\boldsymbol{\gamma}}$ from variables selection. The results are as follows:

$$\hat{\boldsymbol{\gamma}} = min^{+}_{k \epsilon A^c} \left\{ \frac{\widehat{\boldsymbol{C}} - \hat{\boldsymbol{c}}_{\boldsymbol{k}}}{\boldsymbol{A}_{\boldsymbol{A}} - \boldsymbol{a}_{\boldsymbol{k}}}, \frac{\widehat{\boldsymbol{C}} + \hat{\boldsymbol{c}}_{\boldsymbol{k}}}{\boldsymbol{A}_{\boldsymbol{A}} + \boldsymbol{a}_{\boldsymbol{k}}} \right\} = 0{,}75$$

The following prediction vector value obtained

$$\hat{\boldsymbol{\mu}}_{A+} = \hat{\boldsymbol{\mu}}_A + \hat{\boldsymbol{\gamma}}\boldsymbol{u}_A = \begin{bmatrix} 0,3672 \\ 0,2749 \\ 0,1765 \\ 0,0822 \\ 0,0392 \\ 0,0494 \\ -0,0448 \\ -0,2335 \\ -0,3769 \\ -0,3339 \end{bmatrix}$$

5. The candidate value of the regression coefficient for the first active variable

$$\beta_k^{(t)} + \hat{\gamma} \times d_k = 0{,}2499$$

6. Checking the coefficient sign is correct $sign\ (\hat{\beta}_{\boldsymbol{k}}) = sign\ (\hat{c}_{\boldsymbol{k}}) = s_{\boldsymbol{k}}$. In the first variable selection obtained $sign\ (\hat{\beta}_{\mathbf{1}}) = sign\ (\hat{c}_{\mathbf{1}}) = s_{\mathbf{1}} = 1$. Because $\hat{\beta}_1$ and $\hat{c}_1$ has the same sign as s_1 then you can continue with the selection of the second variable.
7. The same steps are repeated in each selection process until all predictor variables are selected.

The result of parameter LASSO estimation at each step are given in Table 4.

Table 4. LASSO Parameter at Each Step

Step	$\boldsymbol{\beta}_1\ (\boldsymbol{X}_1)$	$\boldsymbol{\beta}_2\ (\boldsymbol{X}_2)$	$\boldsymbol{\beta}_3\ (\boldsymbol{X}_3)$	$\boldsymbol{\beta}_4\ (\boldsymbol{X}_4)$
0	0,0000000	0,0000000	0,00000000	0,0000000
1	0,2498419	0,0000000	0,00000000	0,0000000
2	0,5865269	0,3366850	0,00000000	0,0000000
3	0,5973010	0,3408444	0,01239943	0,0000000
4	1,2804893	0,2081409	0,17446217	−0,6956909

Based on the parameter value (coefficient) displayed Table 4, the order of entry of variables in LASSO model can be seen in Table 5. From Table 5 information is obtained that the Human Development Index (X_1) predictor variable is first followed by Average Length of Schooling (X_2), Old School Expectations (X_3), and finally is Life Expectancy (X_4).

Table 5. Order of Entry Variables in LASSO Model

Steps	**Incoming Variables**
1	X_1
2	X_2
3	X_3
4	X_4

A *k*-fold cross validation process was implemented using the steps mode. In this study, the *k*-fold cross validation method was used with the number of parts $c = 10$ to evaluate model performance and select optimal s parameters. Here is a table of cross-validation (CV) values:

Table 6. CV values with mode steps

Step	CV
0	1,1111111
1	0,8134334
2	0,2397145
3	0,2425746
4	0,4819177

Table 6 shows that the minimum CV is 0,2397145. It is mean the minimum mean squared error (MSE) value from the cross-validation process. Therefore, the model in step 2 is selected as the best model. In addition, from the results of the cross-validation, the value obtained $s = 0{,}5$ this value reflects the penalty.

A good LASSO regression model is a model that shrinks the regression coefficient to zero. The more coefficient that shrinks, the more multicollinearity decreases. In other words the performance of the Lasso regression model is good. Following are the results of the shrinkage of the resulting lasso regression coefficient:

Table 7**.** Coefficients of LASSO Regression

Variable	Coefficient
X_1	0,5866
X_2	0,3367
X_3	0
X_4	0

Table 7 shows that almost the coefficient values of LASSO tends to shrink towards zero. Additionally two coefficients on X_3 and X_4 are zero so the School Life Expectancy and Life Expectancy are not significant and the LASSO regression on expenditue per capita data is the best model. Further, multicollinearity checks were performed to verify the presence of multicollinearity. Finally the best LASSO regression model is

$$\hat{Y} = -2{,}4277 \times 10^{-15} + 0{,}5866X_1 + 0{,}3367X_2. \quad (17)$$

Equation (17) means that the variables that influence per capita expenditure in Wonosobo is Human Development Index and Average Length of Schooling.

Table 8 shows that the VIF values for all predictor variables are less than 10, indicating that there is no multicollinearity. It can be concluded that the LASSO regression model obtained is effective in overcoming multicollinearity in per capita expenditure data in Wonosobo.

Table 8. VIF of LASSO Regression

Variable	VIF
X_1	6,4428
X_2	6,4428
X_3	0
X_4	0

4. CONCLUSION

LASSO regression with LAR algorithm can select School Life Expectancy (X_3) and Life Expectancy (X_4) as insignificant predictor variables influencing Average per Capita Expenditure by reducing the regression coefficients closer to zero for both predictor variables. This results in the following LASSO regression model.

$$\hat{Y} = -2{,}4277 \times 10^{-15} + 0{,}5866X_1 + 0{,}3367X_2$$

The model interprets that Per Capita Expenditure in Wonosobo is influenced by the Human Development Index (X_1) and Average Years of Schooling (X_2). This result aligns with the theory in measuring the Human Development Index (HDI) that Per Capita Expenditure is an indicator of the primary dimension of a decent standard of living and Average Length of Schooling is an indicator of the HDI's primary dimension of knowledge.

5. REFERENCES

Aprihartha, Moch. A., & Idham. (2024). Optimization of Classification Algorithms Performance with k-Fold Cross Validation. *Eigen Mathematics Journal*, *7*(2).

Badan Pusat Statistik. (2025, April 10). *Adjusted Per Capita Expenditure (PPP) of Male (Thousand Rupiah/Person/Year)*. BPS-Statistics Indonesia Wonosobo Regency.

Berrar, D. (2019). Cross-Validation. In *Encyclopedia of Bioinformatics and Computational Biology* (pp. 542–545). Elsevier.

BPS. (2022). *https://www.bps.go.id/subject/26/indeks-pembangunan-manusia.html#subjekViewTab1*.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*(2).

Febriani Sagala, I., Romadhoni, A. F., Mardiana, A., Widyasari, A., Simamora, D. R., Nurfadiah, D., Ananda, E., Sembiring, B., Harahap, I. A., Siregar, E., Alif, M. F., Adella, N., Augustian, R., 12, S., & Sihotang, R. F. (2024). Analisis Pengaruh Indeks Pembangunan Manusia dan Pengangguran Terhadap Tingkat Kemiskinan di Sumatera Utara Menggunakan Metode Regresi Linear Berganda. *Jurnal Matematika dan Ilmu Pengelatuan Alam*, *2*(2), 309–324.

Ghozali, I. (2011). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 19*. Badan Penerbit Universitas Diponegoro.

Ginting, D. I., & Lubis, I. (2023). Pengaruh Angka Harapan Hidup dan Harapan Lama Sekolah terhadap Indeks Pembangunan Manusia. *Jurnal Bisnis Net*, *6*.

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5th ed.). McGraw-Hill.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

Jamco, J. C., Kondolembang, F., & Noya Van Delsen, M. S. (2023). Penanganan Multikolinearitas pada Regresi Linier Berganda Menggunakan Regresi Lasso (Studi Kasus: Distribusi Presentase Produk Domestik Regional Bruto di Provinsi Maluku Tahun 1999-2021). *PARAMETER: Jurnal Matematika, Statistika dan Terapannya*, *2*(02), 145-154.

Kadir. (2008). Kemampuan Komunikasi Matematik dan Keterampilan Sosial Siswa dalam Pembelajaran Matematika. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika*.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models.* McGraw-Hill.

Manurung, E. N., & Hutabarat, F. (2021). Pengaruh Angka Harapan Lama Sekolah, Rata-Rata Lama Sekolah, Pengeluaran per Kapita Terhadap Indeks Pembangunan Manusia. *Jurnal Ilmiah Akuntansi Manajemen*, *4*(2), 121-129.

Mubarak, R. (2021). *Pengantar ekonometrika* (1st ed.). Duta Media Publishing.

Nasution, E. F., & Pane, R. (2024). Perbandingan Regresi Lasso dan Principal Component Regression dalam Mengatasi Masalah Multikolinearitas. *In Journal of Mathematics Education and Science, 10(1)*.

Nisa, E. K., & Maulina, R. (2024). The Comparison of Inverse Gaussian and Gamma Regression: Application on Stunting Data in Jepara. *Jurnal Matematika, Statistika Dan Komputasi*, *21*(1), 334–344.

Nizar, F., & Arif, M. (2023). Pengaruh Rata Lama Sekolah, Pengeluaran Perkapita, Pendapatan Asli Daerah, Investasi, Tingkat Pengangguran Terbuka Terhadap Tingkat Kemiskinan di Nusa Tenggara Barat Tahun 2012-2021. *KOMITMEN: Jurnal Ilmiah Manajemen, 4*(1).

Rahayu, A., & Husein, I. (2023). Comparison of Lasso and Adaptive Lasso Methods in Identifying Variables Affecting Population Expenditure. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, *8*(3), 1435-1445.

Rahmawati, F., Yoga Suratman, R., Magister Matematika, A., & Gadjah Mada, U. (2022). Performa Regresi Ridge dan Regresi Lasso pada Data dengan Multikolinearitas. *Leibniz : Jurnal Matematika*.

Sianturi, A., Tampubolon, A., & Nasution, M. (2024). Pengaruh Pengeluaran Perkapita dan Angka Harapan Hidup terhadap Pertumbuhan Ekonomi Kota Medan 2010-2021. *Jurnal Sosial Humaniora Sigli*, *7*(2).

Soleh, A. M., & Aunuddin. (2013). Lasso : Solusi Alternatif Seleksi Peubah dan Penyusutan Koefisien Model Regresi Linier. *FSK : Indonesian Journal of Statistics*, *18*(1), 21–27.

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

Wijiyanto, W., Pradana, A. I., Sopingi, S., & Atina, V. (2024). Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa. *Jurnal Algoritma*, *21*(1).

Zhang, Y., Jin, W., Li, Z., & Pan, W. (2024). A brief discussion on the role of data standardization in regional centralized control construction on the refined management of wind power production. *E3S Web of Conferences*, *561*, 02006.