

Modelling Infant and Maternal Mortality Rate based on slum household uses Bivariate Spline Regression Method

Alan Prahutama¹, Budi Warsito², and Moch.Abdul Mukid³

^{1,2,3} Department of Statistics, Faculty of Science and Mathematics, Diponegoro University Semarang-Indonesia

Abstract: Maternal mortality and infant mortality rate is an interrelated issue. Therefore, maternal and infant mortality modeling can be done bivariate. One method used to model the relationship between response variables and predictor variables is regression. The regression approach that does not use the assumption is spline regression. Spline regression is a regression method based on spline function. Spline function is a polynomial piece that has high flexibility. In this study the response variable used is bivariate, the maternal mortality rate and infant mortality rate, while the predictor variable used is the percentage of slum households. The weighting used is based on the value of the covariance variant. Determination of point knots using Mean Square Error (MSE). The results obtained modeling maternal and infant mortality rates based on the percentage of slum households resulted in MAPE 55.55%. Number of knots obtained as much as 5 point knots with linear order.

Keywords: Infant and Maternal Mortality rate; bivarite dependent; spline regression.

1. Introduction

Multivariate statistics is statistical analysis methods that relate to more than one variable. One of the methods in multivariate statistics is multivariate regression. Multivariate regression is a method to model the relationship between two or more dependent variables with some independent variables [1]. Multivariate regression modeling gives the advantage one of which is a model that obtained more than one model in one analysis. From the model can be known variable independent of the two or more variable take effect to dependent variables. Regression modeling can be done by parametric and nonparametric approaches. The parametric approach is very strict with the model assumptions, whereas the nonparametric approach does not require model assumptions. The parametric approach is made when the shape of the curve is known, while the nonparametric approach is made when the curve shape is unknown [2]. One method of nonparametric regression approach is spline truncated regression. The truncated spline is segmented of segmentation polynomials. Spline truncated regression provides high flexibility due to segmented of segmentation. Spline regression can be established multivariate regression. The development of multivariate spline modeling requires a weighted value. Also in the truncated spline modeling requires a knot point [3]. Optimal knot selection and order in spline modeling are very important. One determination of optimal knot point is to use Mean Square Error (MSE) method. The steps in spline truncated modeling include the determination of knots and order with MSE followed by spline regression parameter estimation [4]. Reducing the number of maternal and infant mortality is one of the goals of Sustainable Development Goals (SDG's) for developing countries. The maternal and infant mortality rate

in Indonesia is decreasing every year. The World Health Organization (WHO) defines the number of maternal mortality as the number of maternal deaths during pregnancy or 42 days after the end of pregnancy which is not caused by an accident. While the number of Infant Mortality is the number of infant deaths under one year, per 1000 live births in a given year. One of the factors causing increased maternal and infant mortality rates is environmental factors. In this study we will examine the model of maternal and infant mortality rate using spline bivariate regression. Environmental factors that made the independent variable is the percentage of slum households.

2. Spline Regression

The general form of the regression model is as follows [5]:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i$$

for $i = 1, 2, \dots, n$ and $p = 1, 2, \dots, P$ with n is number of observations and m is the number of independent variables. For estimate β in parametric regression approach's uses Ordinary Least Square (OLS).

From the equation can be made in the form of:

$$y_i = \hat{f}(x_i) + \varepsilon_i$$

If $\hat{f}(x_i)$ is unknown curve, then $\hat{f}(x_i)$ can be estimated by nonparametric regression approach. One of nonparametric regression approach is spline truncated. Spline truncated function can be follow as:

$$\hat{f}(x_i) = \sum_{j=0}^{m-1} \hat{\beta}_j x_i^j + \sum_{j=1}^r \hat{\beta}_{j+m-1} (x_i - k_j)_+^{m-1}$$

with truncatednya function as follow as:

$$(x_i - k_j)_+^{m-1} = \begin{cases} (x_i - k_j)^{m-1}; & x_i - k_j \geq 0 \\ 0; & x_i - k_j < 0 \end{cases}$$

$\hat{\beta}_j$ is constant with real values and $j = 0, 1, \dots, m + r - 1$ and k is knot function. So that spline regression model can be written as [6]:

$$y_i = \sum_{j=0}^{m-1} \hat{\beta}_j x_i^j + \sum_{j=1}^r \hat{\beta}_{j+m-1} (x_i - k_j)_+^{m-1} + \varepsilon_i$$

According to Eubank (1999), there are several rules in placing the knot position, among others [1]:

1. For the linear spline ($m = 2$), place the knot at the point where the data slope changes.
2. For the quadratic spline ($m = 3$), place the knot near the maximum, minimum or point of change in the data.
3. For the cubic spline ($m = 4$), arrange the knots to close to the point of change in the data and no more than one extreme point (maximum or minimum) and one change point occurs between two knots.

3. Spline Bivariate regression

Bivariate regression analysis is an analysis involving two dependent variables and among the dependent variables there is a correlation. In general, the model for nonparametric bivariate regression can be written as follows [6]:

$y_{1i} = \hat{f}_1(x_i) + \varepsilon_{1i}$ and $y_{2i} = \hat{f}_2(x_i) + \varepsilon_{2i}$ with y_{1i} first dependent variable and y_{2i} is second dependent variable.

The curves of $\hat{f}_1(x_i)$ and $\hat{f}_2(x_i)$ are unknown shape, if its would like approaches by spline function then follow as [7]:

$$\hat{f}_1(x_i) = \sum_{j=0}^{m_1-1} \hat{\beta}_{1j} x_i^j + \sum_{j=1}^r \beta_{1(j+m_1-1)} (x_i - k_j)_+^{m_1-1}$$

$$\hat{f}_2(x_i) = \sum_{j=0}^{m_2-1} \hat{\beta}_{2j} x_i^j + \sum_{j=1}^r \beta_{2(j+m_2-1)} (x_i - k_j)_+^{m_2-1}$$

4. Results and Discussions

The following is given Scatter plot of maternal and infant mortality on the number of slum percentage. The data were taken as many as 35 districts / cities in Central Java in 2016.

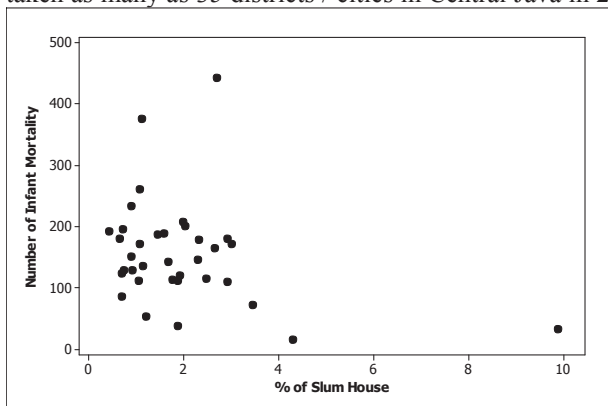


Figure 1. Scatter plot between The number of Infant Mortality and % of Slum House

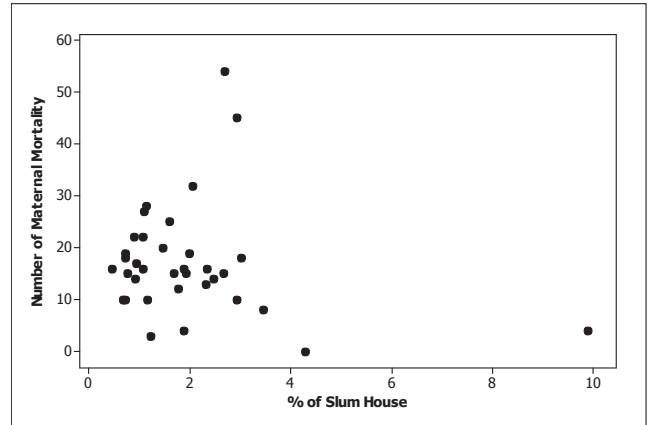


Figure 2. Scatter plot between The number of Maternal Mortality and % of Slum House

Based on Figure 1 and Figure 2 it appears that the data is spreading randomly. Both images have almost the same distribution. The following statistics are presented for each variable.

Table 1. Statistics Descriptive of Data

Variable	Min	Max	Mean	Variance
The Number of Infant Mortality (Y_1)	16	444	156.7	7221.2
The Number of Maternal mortality (Y_2)	0	54	17.2	114.99
% of Slum House (X)	0.44	9.89	1.958	2.746

Table 1 shows that the highest number of infant mortality reached 444 people in Brebes district. While the lowest is 16 people in the city of Magelang with an average infant mortality rate in Central Java reached 156 inhabitants. While the lowest number of maternal deaths reached 0 people in the city of Magelang, the highest reached 54 people in the district Brebes. The average number of maternal deaths reached 17 people. For percentage of Slum House variables, the average ranges from 1.958%.

Based on the analysis of correlation value between the number of maternal and infant mortality is 0.822 (significant with alpha 5%), meaning that can be said that both variables have high linkage. So that the variables of y_1 and y_2 can be modeled bivariate. To model the bivariate spline, the weight used is the variance value of covariance. The value of the covariant is 748,765. Determination of optimum knot using MSE method. Can be seen in Appendix 1.

Based on Appendix 1, the optimum knots are at 5 point knots with the order for y_1 and y_2 are two (linear). The value of optimum knot points is at point 1.14; 2.66; 2.69; 2.92; 1.12 with MSE is 3192.491. So the equation of the spline bivariate regression model for the number of maternal and infant mortality in Central Java with 5 knots is as follows:

For the first independent variable is as follows:

$$\hat{f}_1(x_i) = \sum_{j=0}^{m_1-1} \hat{\beta}_{1j} x_i^j + \sum_{j=1}^r \beta_{1(j+m_1-1)} (x_i - k_j)_+^{m_1-1}$$

$$\hat{f}_1(x_i) = \sum_{j=0}^{2-1} \hat{\beta}_{1j} x_i^j + \sum_{j=1}^5 \beta_{1(j+m_1-1)} (x_i - k_j)_+^{2-1}$$

For the second independent variable is as follows:

$$\hat{f}_2(x_i) = \sum_{j=0}^{m_1-1} \hat{\beta}_{2j} x_i^j + \sum_{j=1}^r \beta_{2(j+m_1-1)} (x_i - k_j)_+^{m_1-1}$$

$$\hat{f}_2(x_i) = \sum_{j=0}^{2-1} \hat{\beta}_{2j} x_i^j + \sum_{j=1}^5 \beta_{2(j+m_1-1)} (x_i - k_j)_+^{2-1}$$

with truncated function as follow as:

$$(x_i - k_j)_+ = \begin{cases} (x_i - k_j); & x_i - k_j \geq 0 \\ 0; & x_i - k_j < 0 \end{cases}$$

The result of parameter estimation is as follows:

Table 2. estimate spline bivariate with variancovarian weighted

Estimate for \hat{y}_1	Estimate for \hat{y}_2
$\beta_{10} = 231.1423; \beta_{11} = -86.4778$	$\beta_{10} = 29.31716; \beta_{11} = -16.3303$
$\beta_{12} = 2229.823; \beta_{13} = -2150.6$	$\beta_{12} = 251.8139; \beta_{13} = -230.136$
$\beta_{14} = 9269.625; \beta_{15} = -10885.7$	$\beta_{14} = 999.1448; \beta_{15} = -1197.6$
$\beta_{16} = 1622.223$	$\beta_{16} = 192.3811$

The model formed for the first dependent variable is as follows:

$$\hat{y}_1 = 231.1423 - 86.4778x + 2229.823(x - 1.12)_+ - 2150.6(x - 1.14)_+ + 926.626(x - 2.66)_+ - 10885.7(x - 2.69)_+ + 1622.223(x - 2.92)_+$$

With truncated function as follow as

$$\hat{y}_1 = \begin{cases} 231.1423 - 86.4778x; & x < 1.12 \\ -3405.797 + 2143.345x; & 1.12 \leq x < 1.14 \\ 4912.648 - 7.25382x; & 1.14 \leq x < 2.66 \\ -19744.6 + 9262.372x; & 2.66 \leq x < 2.69 \\ 9537.949 - 1623.32x; & 2.69 \leq x < 2.92 \\ 4801.058 - 1.09482x; & x \geq 2.92 \end{cases}$$

The model formed for the second dependent variable is as follows:

$$\hat{y}_2 = 29.31716 - 16.3303x + 251.8139(x - 1.12)_+ - 230.136(x - 1.14)_+ + 999.1448(x - 2.66)_+ - 1197.6(x - 2.69)_+ + 192.3811(x - 2.92)_+$$

With truncated function as follow as

$$\hat{y}_2 = \begin{cases} 29.31716 - 16.3303x; & x < 1.12 \\ 281.1311 + 235,4836x; & 1.12 \leq x < 1.14 \\ 543.4862 + 5.34751x; & 1.14 \leq x < 2.66 \\ -2114.24 + 1004.492x; & 2.66 \leq x < 2.69 \\ 1107.302 - 193.107x; & 2.69 \leq x < 2.92 \\ 545.5495 - 0.72559x; & x \geq 2.92 \end{cases}$$

The resulting MAPE value is 55.55%. The value of MAPE produced quite a bit, it is necessary to examine the variables that affect the maternal and infant mortality

5. Conclusion

There is a significant correlation between the number of maternal deaths and the number of infant deaths. In modeling maternal and infant mortality rates against percentage of slum households using spline bivariate resulted in MAPE 55.55%. In the spline bivariate model the number of knots used is 5 knots with the order $m = 2$ (linear). In the next research should be developed multivariable spline bivariate model, meaning the number of independent variables can be reproduced. Bivariate Spline can be used to model ambient air. The ambient air has impact to Global Warming. The higher global warming cause damage ecosystem on earth. Low carbon city's concept is very usefull to reduce global warming. One of method as a model, to create low carbon city we ca use bivariate spline model. For example modelling CO and NO based on density population.

References

1. R.L. Eubank, C. Huang, Y.M. Maldonado, N. Wang, S. Wang, R.J. Buchanan, *J. R. Statist. Soc.*, **66**, Part 3, pp. 653-667, (2004).
2. K.M.O., Bryson, M. Ko, *Information and Management*, **42**: 1-13, (2004).
3. H. Dette, V.B. Melas, *Journal of Statistical Planning and Inference*, **140**: 2037-2045, (2010).
4. B. Lestari, I.N. Budiantara, S. Sunaryo, S., M. Mashuri, *Journal of Mathematics and Statistics*, Vol. **6**, No. 3, hal. 327-332, (2010).
5. L.L. Schumaker, *Spline Functions: Basic Teory*, Cambridge University Press, (2007).
6. R.H. Wang, *Multivariate Spline Functions and Their Applications*, Springer Sciences, (2001)
7. Wu, H. dan Zhang, J.T., *Nonparametric Regression Methods for Longitudinal Data Analysis*. A John-Wiley and Sons Inc. Publication, New Jersey, (2006).

Appendix 1. The Knots of Spline Bivariate

1 knot			2 knot			3 knot		
Orde	MSE	Knot	Orde	MSE	Knots	Orde	MSE	Knots
$y_1 : 2$ $y_2 : 2$	6351.042	2.69	$y_1 : 2$ $y_2 : 2$	5713.734	4.29; 2.69	$y_1 : 2$ $y_2 : 2$	3810.772	2.69; 2.92; 2.66
$y_1 : 2$ $y_2 : 3$	6376.895	2.69	$y_1 : 2$ $y_2 : 3$	5721.527	4.29; 2.69	$y_1 : 2$ $y_2 : 3$	5096.163	1.12; 1.14;1.06
$y_1 : 2$ $y_2 : 4$	6348.382	2.69	$y_1 : 2$ $y_2 : 4$	5722.356	4.29; 2.69	$y_1 : 2$ $y_2 : 4$	4947.307	2.69; 3.45; 2.47
$y_1 : 3$ $y_2 : 2$	6437.108	2.66	$y_1 : 3$ $y_2 : 2$	5963.247	4.29; 2.92	$y_1 : 3$ $y_2 : 2$	5425.142	2.69; 3.45; 2.66
$y_1 : 3$ $y_2 : 3$	6150.084	4.29	$y_1 : 3$ $y_2 : 3$	5392.073	2.69; 2.66	$y_1 : 3$ $y_2 : 3$	4900.964	2.66; 2.69; 2.47
$y_1 : 3$ $y_2 : 4$	6146.769	4.29	$y_1 : 3$ $y_2 : 4$	5437.061	2.69; 2.66	$y_1 : 3$ $y_2 : 4$	5089.421	2.66; 2.69; 1.59
$y_1 : 4$ $y_2 : 2$	6290.021	2.47	$y_1 : 4$ $y_2 : 2$	5896.554	4.29; 3.02	$y_1 : 4$ $y_2 : 2$	5305.076	2.69; 3.45; 2.32
$y_1 : 4$ $y_2 : 3$	6155.771	4.29	$y_1 : 4$ $y_2 : 3$	5871.355	3.02; 2.93	$y_1 : 4$ $y_2 : 3$	5193.655	2.66; 2.69; 2.47
$y_1 : 4$ $y_2 : 4$	5873.856	3.02	$y_1 : 4$ $y_2 : 4$	5582.254	2.3; 1.99	$y_1 : 4$ $y_2 : 4$	4930.292	2.69; 2.92; 2.66
4 knot			5 knot					
Orde	MSE	Knots	Orde	MSE	Knots			
$y_1 : 2$ $y_2 : 2$	3475.96	2.69; 2.92; 4.29; 2.66	$y_1 : 2$ $y_2 : 2$	3192.491	1.14; 2.66; 2.69; 2.92; 1.12			
$y_1 : 2$ $y_2 : 3$	4501.621	2.69; 2.92; 2.93; 2.66	$y_1 : 2$ $y_2 : 3$	3951.782	1.12; 1.14; 2.69; 4.29; 1.06			
$y_1 : 2$ $y_2 : 4$	4718.737	2.47; 2.69; 3.45; 2.32	$y_1 : 2$ $y_2 : 4$	4001.933	2.66; 2.69; 2.92; 4.29; 2.47			
$y_1 : 3$ $y_2 : 2$	4464.858	2.69; 2.92; 2.93; 2.66	$y_1 : 3$ $y_2 : 2$	3957.308	2.69; 2.92; 2.93; 3.45; 2.66			
$y_1 : 3$ $y_2 : 3$	3923.419	2.69; 2.92; 2.93; 2.66	$y_1 : 3$ $y_2 : 3$	3782.793	2.66; 2.69; 2.92; 2.93; 1.59			
$y_1 : 3$ $y_2 : 4$	4603.589	2.69; 2.92; 2.93; 2.66	$y_1 : 3$ $y_2 : 4$	3981.814	2.66; 2.69; 2.92; 2.93; 2.47			
$y_1 : 4$ $y_2 : 2$	5002.222	2.69; 2.93; 3.02; 2.66	$y_1 : 4$ $y_2 : 2$	4280.55	2.66; 2.69; 2.93; 3.02; 2.47			
$y_1 : 4$ $y_2 : 3$	4854.535	2.47; 2.66; 2.69; 2.32	$y_1 : 4$ $y_2 : 3$	4124.681	2.66; 2.69; 2.92; 2.93; 2.47			
$y_1 : 4$ $y_2 : 4$	4702.049	2.69; 2.92; 2.93; 2.66	$y_1 : 4$ $y_2 : 4$	4047.077	2.66; 2.69; 2.92; 2.93; 2.47			