GEOGRAPHICALLY WEIGHTED POISSON REGRESSION (GWPR) FOR ANALYZING THE MALNUTRITION DATA (Case study: Java Island in 2008)

DIDIN SAEPUDIN

DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
BOGOR AGRICULTURAL UNIVERSITY
BOGOR
2012
ABSTRACT

DIDIN SAEPUDIN. Geographically Weighted Poisson Regression (GWPR) for Analyzing The Malnutrition Data (Case Study: Java Island in 2008). Advised by ASEF SAEFUDDIN and DIAN KUSUMANINGRUM.

Poisson regression, namely global model is a statistical method used to analyze the relationship between the dependent variable and the explanatory variables, where the dependent variable is a counted data and has a Poisson distribution. The result of its parameter estimation is homogeneous for all of the observations. However, especially in spatial data, its estimation will produce biased estimation. The parameter estimates in each location will vary among regions as it is influenced by territorial or geographical factors, which is known as spatial variability or spatial non-stationarity. Therefore the appropriate analysis for this data is Geographically Weighted Poisson Regression (GWPR) model. GWPR parameter estimation used a weighting matrix which depends on the proximity between the locations. Fisher scoring iteration is used for solving the iteratively parameter estimation. In this research, GWPR will be used in malnutrition data because malnutrition is counted data which is assumed to have a Poisson distribution and the indirect factors of differences in the number of malnourished patients in every region is possible due to spatial factors. The results showed that GWPR model has better performance than global model based on AICc difference. Poverty aspect was the most influencing factor to the number of malnourished patients in a region compared to health, education, and food aspect. The spatial variability map is created for eight variables used in selected global model where every map showed the variability of local parameter estimates. There were five groups of the local parameter estimates in each map based on percentiles grouping which showed the low until high relation of parameter estimates groups to the number of malnourished patients. This research also created a significant variables map which detects the variables that were significant in each region.

Keywords: spatial non-stationarity, Geographically Weighted Poisson Regression (GWPR), weighting matrix, spatial variability map
GEOGRAPHICALLY WEIGHTED POISSON REGRESSION (GWPR) FOR ANALYZING THE MALNUTRITION DATA
(Case study: Java Island in 2008)

DIDIN SAEPUDIN

Scientific Paper to complete the requirement for graduation of Bachelor Degree in Statistics at Department of Statistics Faculty of Mathematics and Natural Sciences Bogor Agricultural University

DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
BOGOR AGRICULTURAL UNIVERSITY
BOGOR
2012
Title : GEOGRAPHICALLY WEIGHTED POISSON REGRESSION (GWPR) FOR ANALYZING THE MALNUTRITION DATA (Case study: Java Island in 2008)
Name : Didin Saepudin
NRP : G14080012

Approved by,

Advisor I
Advisor II

Dr. Ir. Asep Saefuddin, M.Sc
NIP. 19570316 198103 1 004

Dian Kusumaningrum, M.Si

Acknowledged by,

Head of Department of Statistics
Faculty of Mathematics and Natural Sciences
Bogor Agricultural University

Dr. Ir. Hari Wijayanto, M.Si
NIP. 19650421 1999002 1 001

Graduation date:
ACKNOWLEDGEMENTS

Alhamdulillahi rabbil ‘alamin, praise be to Allah SWT, God of the universe. The author felt deeply grateful to Allah SWT for all the blessing and guidance that made this paper entitled “Geographically Weighted Poisson Regression (GWPR) for Analyzing The Malnutrition Data (Case Study: Java Island in 2008)”.

The author realizes that the completion of this research would not be possible without the support and help from many people. The author would like to express his sincere gratitude to the advisors, Mr. Asep Saefuddin and Mrs. Dian Kusumaningrum for his/her guidance and suggestion for this research and also to Mr. Tomoki Nakaya, Professor of Geography Department in Ritsumeikan University, for the enlightening discussion and giving the GWPR references by e-mail. The author is also grateful to his beloved family for their never ending prayer, love, and support, especially to his mother, alm. Warti for her forever lasting love. Gratitude is shown to all his friends in Statistika 45, Sylvasari Dormitory, and Sylvapinus Dormitory for their patience, spirit, and togetherness in finding knowledge. Gratitude is also to Rama, Liara, and Dania as Mr. Asep Saefuddin’s advised friends for giving the spirit and support in his research. Hopefully, this paper would be useful for all.

Bogor, December 2012

Didin Saepudin
BIOGRAPHY

The author was born in Cilacap, October 13th, 1989 as the son of Nono Darsono and alm. Warti. He is the first son with two brothers. The author graduated from MIN 1 Dayeuhluhur in 2002 and from SMPN 1 Dayeuhluhur in 2005. By the year 2008, he graduated from SMAN 1 Dayeuhluhur and at the same year, he enrolled in Bogor Agricultural University through USMI (Undangan Seleksi Masuk IPB) as a student of Department of Statistics, Faculty of Mathematics and Natural Sciences and took Actuarial and Finance Mathematics as a minor programme.

During his study, the author was active in college organization. In 2009, he was a staff of survey and research department in Statistics Student Association, Gamma Sigma Beta (GSB). In the next year, he was mandated as chairman of Gamma Sigma Beta. He was also active in dormitory organization (Sylvasari Dormitory) as a staff of HRD in 2010. From February 13th until April 6th, 2012 he had fulfilled his internship in PT. Asuransi Cigna, Jakarta on subdivision pricing, division of actuary. The author also had experience as a lecturer assistant for Statistical Method subject for two semesters and Categorical Data Analysis for one semester. The author also received a 2nd place winner on National Statistics Essay Competition in 2011.
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURE</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF TABLE</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF APPENDIX</td>
<td>viii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Background</td>
<td>1</td>
</tr>
<tr>
<td>Objectives</td>
<td>1</td>
</tr>
<tr>
<td>LITERATURE REVIEW</td>
<td>1</td>
</tr>
<tr>
<td>Definition of Malnutrition</td>
<td>1</td>
</tr>
<tr>
<td>Poisson Regression Model</td>
<td>2</td>
</tr>
<tr>
<td>Breusch-Pagan (BP) test</td>
<td>2</td>
</tr>
<tr>
<td>Geographically Weighted Regression (GWR)</td>
<td>2</td>
</tr>
<tr>
<td>Parameters Non-stationarity</td>
<td>3</td>
</tr>
<tr>
<td>Geographically Weighted Poisson Regression (GWPR)</td>
<td>3</td>
</tr>
<tr>
<td>Standard Error of Parameters in GWPR</td>
<td>4</td>
</tr>
<tr>
<td>Parameters Significance Test in GWPR</td>
<td>4</td>
</tr>
<tr>
<td>Kernel Weighting Function</td>
<td>4</td>
</tr>
<tr>
<td>Akaike Information Criterion</td>
<td>4</td>
</tr>
<tr>
<td>METHODOLOGY</td>
<td>5</td>
</tr>
<tr>
<td>Data Sources</td>
<td>5</td>
</tr>
<tr>
<td>Methods</td>
<td>5</td>
</tr>
<tr>
<td>RESULT AND DISCUSSION</td>
<td>5</td>
</tr>
<tr>
<td>Data Exploration</td>
<td>5</td>
</tr>
<tr>
<td>Variables Selection</td>
<td>6</td>
</tr>
<tr>
<td>Global Model</td>
<td>6</td>
</tr>
<tr>
<td>Spatial Variability Test</td>
<td>6</td>
</tr>
<tr>
<td>Parameters Estimate Interpretation of Global Model</td>
<td>7</td>
</tr>
<tr>
<td>Bandwidth Selection</td>
<td>8</td>
</tr>
<tr>
<td>GWPR Model</td>
<td>8</td>
</tr>
<tr>
<td>Stationary Parameter and Fitted Model of GWPR</td>
<td>9</td>
</tr>
<tr>
<td>Spatial Variability Map</td>
<td>9</td>
</tr>
<tr>
<td>Significant Explanatory Variables Map</td>
<td>12</td>
</tr>
<tr>
<td>CONCLUSION AND RECOMMENDATION</td>
<td>13</td>
</tr>
<tr>
<td>Conclusion</td>
<td>13</td>
</tr>
<tr>
<td>Recommendation</td>
<td>13</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>13</td>
</tr>
</tbody>
</table>
LIST OF FIGURE

Figure 1 Stationary versus non-stationary based on concept from Fotheringham et al. (2002) .................................................................................. 3
Figure 2 Effective number of GWPR parameters (K) and the AICc plotted against adaptive kernel bandwidth .......................................................................................................................... 8
Figure 3 Absolute residual of GWPR and global model plotted against observations .......................................................................................................................... 8
Figure 4 Map of eight parameter estimates of GWPR model ................................................................................................................................. 9-10
Figure 5 Significant Explanatory Variables Map ................................................................................................................................. 12

LIST OF TABLE

Table 1 Parameter estimate and \( p \)-value of \( t \)-test of global model with seven explanatory variables .................................................................................................................. 6
Table 2 Comparison of AICc value between two global models ................................................................................................................................. 6
Table 3 Breusch-Pagan test statistics ................................................................................................................................................................................ 7
Table 4 Comparison of AICc value between global model and GWPR model ................................................................................................................. 8
Table 5 Simple test for non-stationary parameter estimate ................................................................................................................................. 9

LIST OF APPENDIX

Appendix 1 Description of dependent variable and explanatory variables .................................................. 15
Appendix 2 Algorithm of research method ........................................................................................................ 16
Appendix 3 Map of malnourished patients distribution for each regency and city in Java island (2008) ................................................................................................................................. 17
Appendix 4 Boxplot of dependent variable and explanatory variables ............................................................................................................... 18
Appendix 5 Pearson’s correlation between variables (dependent variable and explanatory variables) .............................................................................................................................................. 19
Appendix 6 Summary statistics for varying (local) coefficients ................................................................................................................................. 20
INTRODUCTION

Background

Poisson regression is a statistical method used to analyze the relationship between the dependent variable and the explanatory variables, where the dependent variable is in the form of counted data that has a Poisson distribution. For example, the case of relation between cervical cancer disease incidence cases in England and socio-economic deprivation (Cheng et al. 2011). In the parameter estimation, it is assumed that the fitted value for all of the observations or regression points are the same. This assumption is referred to as homogeneity (Harlton & Fotheringham 2009).

However, there is a problem if the observations are recorded in a regional or territorial data, which is known as spatial data. If the spatial data was analyzed by Poisson regression, it will ignore the variation across study region. The decision to ignore the local spatial variation in parameters can lead to biased results (Cheng et al. 2011). Whereas, the local spatial variation can be important and such information may have important implications for policy makers.

In the case of spatial processes which is referred to spatial non-stationarity (Theringham et al. 2002). Geographically Weighted Poisson Regression (GWPR) will be the most appropriate analysis if the data has a non-stationary condition and the dependent variable was assumed Poissonly distributed. GWPR model uses weighting matrix which depends on the proximity between the location of the observation. In this research, a modified local Fisher scoring procedure is used to estimate the local parameters of GWPR iteratively and using the bi-square adaptive Kernel function to find the weighting matrix is expected. The Kernel function determines the weighting matrix based on window width (bandwidth) where the value is accordance with the conditions of the data and adaptive because it has a wide window (bandwidth) that varies according to conditions of observation points.

Malnutrition data in Java island was chosen in this research because the number of malnourished patients was counted data which was assumed to have a Poisson distribution. Based on the data taken from National Institute of Health Research and Development (2008), there are differences of malnutrition prevalence in every province in a island. The malnutrition prevalence was based on indicators of Weight/Age (BB/U), East Java (4.8%) had the highest malnutrition prevalence, followed by Banten (4.4%), Central Java (4.0%), West Java (3.7%), DKI Jakarta (2.9%), and Jogjakarta (2.4%). These difference can be related to the phenomenon of spatial variabiliy. The variability of malnutrition prevalence in each province might vary in each city or regency. Directorate of Public Health Nutrition (2008) described that the indirect factors that influence malnutrition are food supply, sanitation, health services, family purchasing power and the level of education. Nevertheless geographical aspect could also influence malnutrition. Handling the malnutrition case is very important because malnutrition will directly or indirectly reduce the level of children intelligence, impaired growth and children development and decreased the productivity.

The research that analyzed the affected factors that affect malnutrition which is also associated with spatial effect has been carried out by several researchers. Two researchers who have analyzed it were Rohimah (2011) with the title of thesis "Spatial Autoregressive Poisson model for Detecting the Influential Factors of the number of Malnourished patients in East Java" and A’yunin (2011) who analyzed "Modeling of Malnutrition Infants Under Five (Toddlers) in Surabaya with Spatial Autoregressive Model (SAR)". However, there was only one of the research, which is related with Geographically Weighted Poisson Regression (GWPR), was analyzed by Aulele (2010) in his thesis entitled “Geographically Weighted Poisson Regression Model (Case Study: The Number of Infant Mortality in East Java and Central Java 2007)”. The difference of this research with the previous research is the researcher has created a local model for each regency or city in Java island. The researcher also created a spatial variability map for making it more easy to interpret the local parameter estimates.

Objectives

The objectives of this research are:

1. To compare the better model between global Poisson model and GWPR model for the malnutrition data in Java
2. To create the spatial variability map
3. To analyze the factors influencing the number of malnourished patients for each regency and city in Java.
LITERATURE REVIEW

Definition of Malnutrition
Based on Statistics Indonesia (2008), definition of malnutrition is a condition of less severe level of nutrients caused by low consumption of energy and protein in a long time characterized by the incompatibility of body weight with age. Directorate of Public Health Nutrition (2008) described that malnutrition is directly caused by the lack of food intake and infectious diseases and it is indirectly caused by the availability of food, sanitation, health care, parenting, family purchasing power, education and knowledge.

Poisson Regression Model
The standard model for count data is the Poisson regression model, which is a nonlinear regression model. This regression model is derived from the Poisson distribution by allowing the intensity parameter $\mu$ to depend on explanatory variables (Cameron & Trivedi 1998).

Probability to count the numbers of events $y_i$ (dependent variable for $i$th observation) given variable $x$ can be defined as:
$$
Pr(y_i|x_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}
$$
where, $y_i = 0, 1, 2, ..., i = 1, 2, ..., n$,
$\mu_i = E(y_i|x_i) = \exp(x_i' \beta)$
$x_i' = [1, x_{i1}, ..., x_{ik}]$
$k$ = the number of explanatory variables

Fleiss et al. (2003) has defined the Poisson regression model as,
$$
ln \mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_{p-1} x_{i(p-1)} = x_i' \beta
$$
where $p$ is the number of parameters regression, $\mu_i$ is the expected value of Poisson distribution for observation $ith$ and
$n = [\beta_0, \beta_1, ..., \beta_{p-1}]$

The factor change in the expected count for a change of $\delta$ in $x_k$ equals,
$$
\frac{\Delta E(y|x, x_k + \delta)}{\Delta E(y|x, x_k)} = \exp(\beta_k \delta)
$$
therefore, the parameter can be interpreted as for a change of $\delta$ in $x_k$, expected count increase by a factor of $\exp(\beta_k \delta)$, while all other variables are considered constant (Long 1997).

Breusch-Pagan Test (BP-test)
Breusch and Pagan (1979) in Arbia (2006) proposed a generic form of homoscedasticity expressed by the following equation,
$$
E(u_i^2|x) = \alpha_1 x_{1i} + \alpha_2 x_{2i} + ... + \alpha_k x_{ki}
$$
where, $\alpha' = (\alpha_1, \alpha_2, ..., \alpha_k)'$ is a set of constants. $x_1$ is the constant term of the regression and $x_2, ..., x_k$ are the constant terms for the regressors. The null and alternative hypothesis of BP-test are:
$H_0: \alpha_2 = \alpha_3 = ... = \alpha_k = 0$
$H_1: \exists i$, where $\alpha_i \neq 0$ : ($i=2, 3, ..., k$)
Under $H_0$, we assume that $\sigma_i^2 = \sigma_i^2 = \text{constant}$.

Spatial variability is simply structural instability in the form of non-constant error variances (heteroscedasticity) or model coefficients (variable coefficients, spatial regimes) and can be tackled by means of the standard econometric toolbox. One of the reasons why it is important to consider spatial variability, explicitly the “structure” of data behind the instability is spatial (or geographic) in the sense that the location of the observations is crucial in determining the form of the instability. For example, groupwise heteroscedasticity could be modeled as different error variances for different compact geographic subsets of the data (Anselin 1999).

Anselin (1988) in Arbia (2006) has described that Breusch-Pagan test statistics can be derived using the general expression of the Lagrange multiplier test, which could be expressed as:
$$
BP = \frac{1}{2} \left( \sum_{i=1}^{n} x_i f_i \right) \left( \sum_{i=1}^{n} x_i f_i \right) \left( \sum_{i=1}^{n} x_i f_i \right)
$$
where,
$f_i = \hat{u}_i - 1$
$\hat{u}_i = (y_i - \hat{\beta} x_i)$
$\sigma^2 = \sum_{i=1}^{n} \hat{u}_i^2$

The test-statistic in equation (1) has a $\chi^2$ distribution with $k-1$ degrees of freedom ($k$ is the number of explanatory variables).

Geographically Weighted Regression
Geographically Weighted Regression (GWR) is an alternative method for the local analysis of relationships in multivariate data sets. The underlying model for GWR is,
where \( \{ \beta_0(u_i, v_i), ..., \beta_{k+1}(u_i, v_i) \} \) are \( k+1 \) continuous functions of the location \((u_i, v_i)\) in the geographical study area. The \( \varepsilon \)'s are random error terms (Fotheringham et al. 2002).

**Parameters Non-stationarity**

Fotheringham et al. (2002) has described that the data is in a stationary condition if the data is homogenous across the study region, otherwise it is known to be non-stationary. In a non-stationary condition, it investigates that the parameter estimates might not be constant over space. Clearly, any relationship that is non-stationary over space will not be well-represented by a global statistic and this global value may be misleading locally. It is therefore useful to conduct an indepth study why relationships might vary over space. If there is no reasonable suspect that the parameters vary, there is no need to develop local statistical methods.

![Figure 1: Stationary versus non-stationary based on concept from Fotheringham et al. (2002)](image)

**Geographically Weighted Poisson Regression**

Geographically Weighted Poisson Regression (GWPR) model is a developed model of GWR where it brings the framework of a simple regression model into a weighted regression. Coefficients of GWPR regression can be estimated by calibrating a Poisson regression model where the likelihood is geographically weighted, with the weights being a Kernel function centred on \( u_i \) (\( u_i \) is a vector of location coordinates \((u_i, v_i)\)) (Nakaya et al. 2005).

The steps used to estimate the coefficients of GWPR are:

1. Make a likelihood function of Poisson distribution from the \( n \)-numbers of response variable \( y_i \)-Poisson \( \mu(y_i) = \exp(x_i' \beta) \):

\[
\mu(y_i) = \exp(x_i' \beta) \quad (2)
\]

2. Maximize the log-likelihood function in equation (2):

\[
\ln L(\beta) = - \sum_{i=1}^{n} \exp(x_i' \beta) + \sum_{i=1}^{n} y_i(x_i' \beta) - \sum_{i=1}^{n} \ln y_i!
\]

3. Fotheringham et al. (2002) has described that an observation in GWPR is weighted in accordance with its proximity to location \( i^0 \). So, the equation in the 2nd step can be expressed as:

\[
\ln^* L(\beta(u_i, v_i)) = - \sum_{i=1}^{n} \exp(x_i' \beta(u_i, v_i)) + \sum_{i=1}^{n} y_i(x_i' \beta(u_i, v_i)) - \sum_{i=1}^{n} \ln y_i! \quad w_i(u_i, v_i)
\]

4. Make a partial derivatives of equation above with respect to the parameters in \( \beta'(u_i, v_i) \) and its result must be equal to zero:

\[
\frac{\ln^* L(\beta(u_i, v_i))}{\beta'(u_i, v_i)} = - \sum_{i=1}^{n} x_i \exp(x_i' \beta(u_i, v_i)) + \sum_{i=1}^{n} x_i y_i w_i(u_i, v_i) = 0
\]

However, this equation is iterative.

5. Based on Nakaya et al. (2005), The maximization problem in the equation (3) can be solved by a modified local Fisher scoring procedure, a form of Iteratively Reweighted Least Squares (IRLS). The iterative procedure is necessary except for the special case of Kernel mapping. In this local scoring procedure, the following matrix computation of weighted least squares should be repeated to update parameter estimates until they converge:

\[
\beta^{(l+1)}(u_i, v_i) = (XW(u_i, v_i)A(u_i, v_i)^{0.5}X)^{-1}XW(u_i, v_i)A(u_i, v_i)^{0.5}z(u_i, v_i)^{0.5}
\]

where \( \beta^{(l+1)} \) is a vector of local parameter estimates specific to location \( i \) and superscript \((l+1)\) indicates the number of iterations. The vector of local parameter estimates specific to location \( i \) when \( l^\text{th} \) iteration is defined as:

\[
\beta^{(l)}(u_i, v_i) = (\beta_0^{(l)}(u_i, v_i), \beta_1^{(l)}(u_i, v_i), ..., \beta_k^{(l)}(u_i, v_i))
\]
The local pseudo t-statistic for the local version of the $k^{th}$ parameter estimate is then computed by:

$$t_k(u_i,v_i) = \frac{\hat{\beta}_k(u_i,v_i)}{Se(\hat{\beta}_k(u_i,v_i))}$$

This can be used for local inspection of parameter significance. The usual threshold of p-values for a significance test is effectively $|t| > 1.96$ for tests at the five percent level with large samples (Nakaya et al. 2005).

### Kernel Weighting Function

The parameter estimates at any regression point are dependent not only on the data but also on the Kernel chosen and the bandwidth for that Kernel. Two types of Kernel that can be selected are a fixed Kernel and an adaptive Kernel. The adaptive Kernel permits using a variable bandwidth. Where the regression points are widely spaced, the bandwidth is greater when the regression points are more closely spaced (Fotheringham et al. 2002).

The particular function in adaptive Kernel’s method is a bi-square function. A bi-square function where the weight of the $j^{th}$ data point at regression point $i$ is given by:

$$w_j = \begin{cases} 
(1-(d_{ij}/b_{ik})^2) & \text{when } d_{ij} \leq b_{ik}, \\
0 & \text{when } d_{ij} > b_{ik},
\end{cases}$$

where $w_j$ is weight value of observation at the location $j$ for estimating coefficient at the location $i$. $d_{ij}$ is the Euclidean distance between the regression point $i$ and the data point $j$ and $b_{ik}$ is an adaptive bandwidth size defined as $k^{th}$ nearest neighbour distance (Fotheringham et al. 2002).

### Akaike Information Criterion

The Akaike Information Criterion (AIC) of the model with bandwidth $b$ is defined as:

$$AIC(b) = D(b) + 2K(b)$$

where $D$ and $K$ denote the deviance and the effective number of parameters in the model with bandwidth $b$ respectively (Fotheringham et al. 2002). The deviance of Poisson regression model is defined as:

$$D = 2 \sum_{i=1}^{n} \left[ y_i \ln \left( \frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right]$$

where $y_i$ is a dependent variable for $i^{th}$ observation and $\hat{\mu}_i$ is the expected value of Poisson distribution for $i^{th}$ observation (Hardin & Hilbe 2007). The model with the smallest AIC should be selected as an optimal model.
called MAICE (minimum AIC estimator) (Fotheringham et al. 2002).

Sugiura (1978) in Burnham & Anderson (2002) derived a second-order variant of AIC that he called AICc (corrected Akaike Information Criterion). Formula of AICc is:

$$AICc(b) = AIC(b) + \frac{2(K(b)(K(b)+1))}{n-K-1}$$

If sample sizes is large with respect to the number of estimated parameters or when the ratio $n/K$ is small (say < 40), using AICc is recommended (Burnham & Anderson 2002). If the ratio $n/K$ is sufficiently large, then AIC and AICc are similar and will strongly tend to select the same model. A common rule-of-thumb in the use of AICc is that if the difference in AICc values between two models is less than or equal to 2, then two models are good as model. Otherwise, the difference in AICc values between two models is greater than 2, there is significant difference between the two models and the minimum AICc as a good model (Nakaya 2005).

Fotheringham et al. (2005) have described that AIC and AICc also can be used to select the optimal bandwidth with the minimum AIC or AICc.

**METHODOLOGY**

**Data Sources**

The data used in this study are secondary data from Potensi Desa (Podes) data in 2008. The objects of this research are 112 regencies and cities in Java island. The dependent variable used in this research is the number of malnourished patients in each regency and city in Java island. The explanatory variables is selected from four aspects, which include poverty, health, education, and food aspect. There were 14 explanatory variables used which is described in detail in Appendix 1.

**Methods**

The methods used in this research are:

- Explore the data
- Check the multicolinearity among the explanatory variables by using the Pearson’s correlation
- Make a best model for global Poisson regression model, namely global model with the following steps:
  a) Modeling with all of selected explanatory variables from previous step
  b) Select the explanatory variables which were significant to the dependent variable (with $p$-value of parameter estimates are equal to or lower than five percent and 15 percent)
  c) Make a global model from combinations of selected explanatory variables
  d) Compare the AICc value among the models created
  e) Select the best model (with criteria of AICc difference is less than 2)
- Test the spatial variability of the data by Breusch-Pagan test
- Interpretation of parameter estimate in global model
- Analyze a local model of GWPR, by using the following steps:
  a) Determine the value of longitude and latitude $(u_i,v_i)$ for every regency and city in Java island
  b) Determine the optimum bandwidth by the minimum AICc
  c) Calculate the weighted matrix by the bi-square adaptive Kernel function for each regency or city
  d) Estimate the parameter of GWPR with a modified local Fisher scoring procedure for solving the parameter estimate iteratively
  e) Test the spatial non-stationary
  f) Create a fitted model of GWPR
- Test the goodness of fit between global model and GWPR model by the minimum AICc estimation (with criterion of AICc difference is less than 2)
- Interpretation of the GWPR parameters
- Interpretation of the spatial variability map

The algorithm of research method can be seen in Appendix 2. The Softwares that was applied for this research are GWR 4.0, R 2.15.0, and ArcGIS 3.2.

**RESULT AND DISCUSSION**

**Data Exploration**

The distribution of malnourished patients in Java island can be seen in Appendix 3. From the map, we can see the variability of malnourished patients for each regency and city in Java island. For an example, the number of malnourished patients in Purworejo (2,002), Karawang (2,125), and Majalengka (2,137) are very high. But, the number of
malnourished patients in the regions of DKI Jakarta province are low which is lower than 500 patients. The regions with high numbers of malnourished patients must be handled seriously than the others. One of the solution is by determining the influencing factors to the malnourished patients for each regions. If the factors have been known, it can be used by policy maker for solving this problem.

A boxplot was also used for exploring the data. The figure of boxplot for each variable was shown in Appendix 4. From the boxplot, it can be seen that there were five explanatory variables that does not have outliers, which are FamLabour, Illiteracy, ISP, ActISP, and HlthWrk variable. The other explanatory variables have upper extreme values as outliers. The outliers were found in the regions with high values of FamLabour, Illiteracy, ISP, ActISP, and HlthWrk. As for an example, Bogor regency have upper extreme value in FamSlums, JHSch, and ElmSch variable. It means that Bogor regency have high number of families residing in the slums, and many of their people have junior high schools or equivalent, or elementary schools or equivalent education. Therefore the outliers in each of the variables are still used in further analysis because every region will be created a local model of GWPR.

**Variables Selection**

Variable selection in developing the GWPR model is done by exploring the correlation between two explanatory variables using Pearson’s correlation, which is showed in Appendix 5. The explanatory variable with higher correlation to dependent variable (if there are two variables that are highly correlated) is chosen for the global model. From 14 explanatory variables, there are only seven explanatory variables that are not highly correlated to each other. They were FamLabour, FamRiver, FamSlums, ISP, Doctor, DiffPlace, and JHSch variable. Therefore, these seven variables were used in the further analysis. All of the aspects are represented in the selected explanatory variables.

**Global Model**

A traditional global Poisson regression model, namely global model, was fitted to the malnourished patients in one regency or city based on the selected explanatory variables. As explained above, seven explanatory variables were selected and used to create global model. Their parameter estimates and \( p \)-values are displayed in Table 1.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimate</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.039338</td>
<td>(&lt; 2 \times 10^{-16})</td>
</tr>
<tr>
<td>FamLabour</td>
<td>0.000002</td>
<td>(&lt; 2 \times 10^{-16})</td>
</tr>
<tr>
<td>FamRiver</td>
<td>0.000053</td>
<td>(&lt; 2 \times 10^{-16})</td>
</tr>
<tr>
<td>FamSlums</td>
<td>-0.000008</td>
<td>(&lt; 1.46 \times 10^{-12})</td>
</tr>
<tr>
<td>ISP</td>
<td>0.00074</td>
<td>(&lt; 2 \times 10^{-16})</td>
</tr>
<tr>
<td>Doctor</td>
<td>-0.00029</td>
<td>0.124</td>
</tr>
<tr>
<td>DiffPlace</td>
<td>0.001539</td>
<td>(&lt; 7.74 \times 10^{-9})</td>
</tr>
<tr>
<td>JHSch</td>
<td>-0.001113</td>
<td>(&lt; 2 \times 10^{-16})</td>
</tr>
</tbody>
</table>

*significant at \( \alpha = 5\% \)

A Global model from seven explanatory variables, namely global model 1, showed that only one variable was not significant at five percent level of significance which was Doctor variable. Therefore another global model candidate will be created without using Doctor variable, namely global model 2. Global model 2 resulted that all of variables were significant. But, we can not select directly that global model 2 was the best model. To compare it with global model 1, the difference of AICc values was used as a criterion to select the best model.

<table>
<thead>
<tr>
<th>Model</th>
<th>AICc</th>
<th>AICc Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global model 1</td>
<td>11657.58</td>
<td>0</td>
</tr>
<tr>
<td>Global model 2</td>
<td>11657.67</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Based on the criterion of selecting the best model with the AICc difference between two models is less than or equal to 2 (Nakaya et al. 2005). Both models are considered as a good model because the AICc difference is lower than 2. But, global model 1 will be selected as best model to create GWPR model because its AICc difference is very small and this model accommodated all factors. Therefore, it will explain the influencing factors of the number of malnourished patients better.

**Spatial Variability Test**

The test used for detecting spatial variability was Breusch-Pagan test (BP test). The null hypothesis for this test is error variance for all of the observations is constant. While, the alternative hypothesis is heteroscedasticity in error variance (there are variability among observations).
From Table 3, the \( p \)-value of BP test was significant at a five percent significance level. It means that there are variability among observations. Therefore, there are special characteristics for each region. Hence, the data must be analyzed by local modeling in order to capture the variation of data and explain the characteristic of each region better.

### Parameters Estimate Interpretation of Global Model

In the previous discussion, the global model used for developing the next modeling is global model 1. The selected explanatory variables were FamLabour, FamRiver, FamSlums, ISP, Doctor, DiffPlace, and JHSch variable. The final fitted global model is:

\[
\text{HalPat} = 5.039338 + 0.000002 \text{FamLabour} + 0.000053 \text{FamRiver} - 0.000008 \text{FamSlums} + 0.000740 \text{ISP} - 0.000029 \text{Doctor} + 0.001539 \text{DiffPlace} - 0.001113 \text{JHSch}
\]

From equation (4), all of the aspects are represented in the final fitted global model. The aspect of poverty is represented by FamLabour, FamRiver, and FamSlums variable. The health aspect are represented by ISP and Doctor variable. Education and food aspect is represented only by JHSch variable as an education aspect and DiffPlace variable as a food aspect.

In the poverty aspect, FamLabour and FamRiver variables are positively related, but FamSlums variable is negatively related to the number of malnourished patients in a region. Actually, poverty affects health, so the poor people are vulnerable to various kinds of diseases, one of them is malnutrition. In this case, FamLabour and FamRiver variable was enough to explain the positive relation of poverty aspect with the number of malnourished patients. It means that the regions with high counts of poor people would have high counts of malnourished patients. In parameter interpretation, the parameter estimate of FamLabour was used for an example. With the increasing of 1,000 family members that are farm labour, the expected number of malnourished patients will increase around 1,000 persons in each regency or city (assuming that all of other parameter estimates are constant).

Next, the health aspect represented by ISP and Doctor variables. Relation of ISP and Doctor variables to the number of malnourished patients are different. The ISP variable is positively related, but the Doctor variable is negatively related. In expectation, the better organizing of health aspect in a region, will cause the number of malnourished patients to decrease. One of the reason that the ISP variable has a positive parameter estimate is that when there are many ISP (Posyandu), there will be many malnourished patients being recorded. For this aspect, there will not be any parameter estimates that will be interpreted because the ISP variable has positive parameter estimate and the Doctor variable is not significant (\( p \)-value = 0.124). Therefore, it might produce a bias or wrong interpretation.

The education aspect is only represented by JHSch variable. The JHSch variable has a negative parameter estimate. It means that indirectly the education aspect tends to decrease the malnourished patients. The interpretation of JHSch parameter estimate is that the decrease of 1,000 building units of junior high schools or equivalent, the expected number of malnourished patients increase up to 999 persons in each regency or city (assuming that all of other parameter estimates are constant). Junior high school or equivalent is the last level for nine-years compulsory education programme from the government. In a regency or city that has many buildings of junior high school or equivalent, the regions might have many people studying in these school. Therefore it might increase the knowledge of the people, especially in malnutrition knowledge. Hence, if more people knew about malnutrition, the number of malnourished patients in this regions is expected to decrease.

The last aspect is food aspect. Similar with the education aspect, the food aspect is only represented by one variable, which is the DiffPlace variable. In expectation, the need of food in a region is sufficient, so the malnourished patients will decrease. In this case, the food aspect is positively related to the number of malnourished patients. It is shown by the positive parameter estimate of DiffPlace variable. The interpretation of this variable is when there is an increase of 10,000 families with poor access to village office or sub-district capital, the expected number of malnourished patients will increases up to
10,002 persons in each regency or city (assuming that all of other parameter estimates are constant). This variable represents the food aspect because usually more markets, shopping centers, and food stalls are found near the village office or sub-district capital. The farther people are from the centre of village and sub-district, it will be more difficult to buy food needs. The regency or city that has many villages like this also might have high malnourished patients.

**Bandwidth Selection**

The method used for selecting an optimal bandwidth is bi-square adaptive Kernel. The optimum bandwidth selection is a feature of the golden section search method in the GWR software (see Fotheringham et al. 2005).

![Figure 2: Effective number of GWPR parameters (K) and the AICc plotted against adaptive kernel bandwidth](image)

From Figure 2, it can be seen that the selected optimum bandwidth from this method is 46 neighboring regencies or cities nearest to the \( i \)th regency or city. In selecting the optimum bandwidth, GWPR model with bi-square adaptive Kernel has an AICc value of 6705.39. The effective number of GWPR parameters is 32.23. Afterwards the selected optimum bandwidth will be used to calculate the weighting matrix in a regency or city.

**GWPR Model**

The next analysis is modeling with GWPR. From this model with an optimum bandwidth, GWPR model also has a minimum AICc value. Before further interpretation of GWPR model, GWPR model must be compared with a global model to select the best model between them.

Table 4 showed that the AICc difference with the global model is 4952.14. It indicates that GWPR model is better than the global model. This difference of AICc of global model is very high. Based on The AICc difference criterion (Nakaya et al. 2005), GWPR model was better than global model.

<table>
<thead>
<tr>
<th>Model</th>
<th>AICc</th>
<th>AICc Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global model</td>
<td>11657.58</td>
<td>4952.14</td>
</tr>
<tr>
<td>GWPR model</td>
<td>6705.39</td>
<td>0</td>
</tr>
</tbody>
</table>

The performance of GWPR model compared to the global model also can be seen from the residual values. According to Figure 3, residual of GWPR model is relatively less than global model for each region. The total absolute residual value is 14,467.94 for GWPR model and 22,020.07 for the global model. Similar with the AICc difference, the difference of the total absolute residual value for two models was very high. Therefore GWPR model will be considered as the best model.

![Figure 3: Absolute residual of GWPR and global model plotted against observations](image)

**Stationary Parameter and Fitted Model of GWPR**

The method used to test the stationarity in GWPR parameters is done by comparing the inter-quartile range of GWPR model and the standard error of the global model. If the local inter-quartile range (IQR) is twice the global standard error or more, then the variable requires a non-stationary model to represent it adequately (Cheng et al. 2011).

Simple test in Table 5 showed that all of the variables have an inter-quartile range more than twice the global standard error. This indicates that all of variables are better fitted by a non-stationary model in GWPR than using a global model. A non-stationary model allows greater prediction power (Cheng et al. 2011), and as a result, leads to greater understanding of the relation between the number of malnourished patients and seven explanatory variables selected which come...
from four aspects and how it varies over space in relation to deprivation conditions regionally.

Table 5 Simple test for non-stationary parameter estimate

<table>
<thead>
<tr>
<th>Variable</th>
<th>Standard Error</th>
<th>IQR (GWPR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.010759</td>
<td>0.237853</td>
</tr>
<tr>
<td>FamLabour</td>
<td>0.000000*</td>
<td>0.000003</td>
</tr>
<tr>
<td>FamRiver</td>
<td>0.000002</td>
<td>0.000031</td>
</tr>
<tr>
<td>FamSlums</td>
<td>0.000001</td>
<td>0.000094</td>
</tr>
<tr>
<td>Doctor</td>
<td>0.000008</td>
<td>0.000377</td>
</tr>
<tr>
<td>DiffPlace</td>
<td>0.000266</td>
<td>0.010788</td>
</tr>
<tr>
<td>JHSch</td>
<td>0.000059</td>
<td>0.002455</td>
</tr>
</tbody>
</table>

* value in 10 digit decimals is 0.0000000089

The final fitted model of GWPR is given

\[
\hat{\ln y}_i = \hat{\beta}_0(u_i, v_i) + \hat{\beta}_1(u_i, v_i) \text{FamLabour} + \hat{\beta}_2(u_i, v_i) \text{FamRiver} - \hat{\beta}_3(u_i, v_i) \text{FamSlums} + \hat{\beta}_4(u_i, v_i) \text{ISP} - \hat{\beta}_5(u_i, v_i) \text{Doctor} + \hat{\beta}_6(u_i, v_i) \text{DiffPlace} - \hat{\beta}_7(u_i, v_i) \text{JHSch}
\]

The symbol of \((u_i, v_i)\) besides the parameter estimates in equation (5) defined that the parameter estimates is in a non-stationary condition. Summary of statistics for varying (local) coefficients are shown in Appendix 6.

Spatial Variability Map

Concerning the individual significance of the parameters, though they are all statistically significant in the global regression, there are several non-significant parameters in the GWPR models. Therefore, spatial variability map is needed. GWPR parameter estimates map has some advantages. First, the map provides specific information for each area including parameter estimates which are calculated in the case of global models. Second, the map also can lead to some regions which have highly relation to dependent variable for each explanatory variable.

Figure 4 below shows the maps of local parameter estimates of GWPR for intercept, FamLabour, FamRiver, FamSlums, ISP, Doctor, DiffPlace, and JHSch variable. The maps indicated that the eight parameter estimates are not equal for all locations. The darker color of the map indicates a higher value of a local parameter estimate.
The local parameter estimates of intercept were very high in the regencies and cities in border area of West Java and Central Java province, and also in all of regencies and cities in Banten province. This indicates the higher malnourished patients in this regions when the other variables are constant. However, they were pretty low in most of the regencies and cities in border area of Central Java and East Java province and also in the west area of DKI Jakarta province, but pretty low in area of Banten province.
the number of malnourished patients. Therefore, the regions with high number of families residing along the river would have a higher increase of malnourished patients. Therefore, the regions with high parameter estimates must be given a solution for this problem by the government, especially local government. For example, the central government in these regions might build more flat house units like in DKI Jakarta to relocate the families residing along the river. This recommendation is expected to decrease the number of malnourished patients.

In FamSlums variable, the parameter estimates of all of the regencies and cities in the central and western area of East Java and Tuban regency were very high. However, the regencies and cities in West Java, DKI Jakarta, Banten province, some regencies and cities in the west region of Central Java province, Ponorogo, and Tenggalek regency have a pretty low negative local parameter estimates. Their negative value indicated that increasing the number of families residing in slums would decrease the number of malnourished patients. It was clear that poverty aspect in general can increase the number of malnourished patients. Therefore, the regions with high number of families residing along the river would have a pretty low negative local parameter estimates. Their negative value indicated that these regions might decrease the number of malnourished patients in each region.

Next, the local parameter estimates of FamRiver were pretty low in Banyumas, maggane, Pekalongan, Tegal, Purbalingga regency, Tegal, Pekalongan city, and all of the regencies and cities in the central and east area of East Java province. However, the parameter estimates in the regions of Yogyakarta, Blora, Tuban, Bojonegoro, and west of the regencies and cities in West Java province were very high. The higher value of parameter estimates indicated that these regions would have a higher increase of malnourished patients. Therefore, the regions with high parameter estimates must be given a solution for this problem by the government, especially local government. For example, the central government in these regions might build more flat house units like in DKI Jakarta to relocate the families residing along the river. This recommendation is expected to decrease the number of malnourished patients.

In FamSlums variable, the parameter estimates of all of the regencies and cities in the central and western area of East Java and Tuban regency were very high. However, the regencies and cities in West Java, DKI Jakarta, Banten province, some regencies and cities in the west region of Central Java province, Ponorogo, and Tenggalek regency have a pretty low negative local parameter estimates. Their negative value indicated that increasing the number of families residing in slums would decrease the number of malnourished patients. It was clear that poverty aspect in general can increase the number of malnourished patients. Therefore, the regions with high number of families residing along the river would have a pretty low negative local parameter estimates. Their negative value indicated that these regions might decrease the number of malnourished patients.

The FamRiver variable was the highest influencing factors to the number of malnourished patients compared to the other variables which represent the poverty aspect, which is based on the comparisson of the mean of local parameter estimates from FamLabour, FamRiver, and FamSlums variable according to Appendix 6. This indicated that the regions with high number of families residing along the river must be given the serious attention by local government.

All of the local parameter estimates in ISP variable were positive. This indicated that the number of integrated services post (Posyandu) in each region are positively related to the expected number of malnourished patients. The possible reason is that increasing of integrated services post (Posyandu) units, which have been built by the government, through the Ministry of Health in each region was done without considering that a region is an endemic area or not to malnutrition case. The other reason is that increasing the number of Posyandu in a region may cause the detection of many malnourished patients. These patients are also allowed to have a consultation about malnutrition to the health workers in Posyandu.

Then, the local parameter estimates of Doctor were low and negative in Ciamis regency, Banjar city, and all of the regencies and cities in Central Java, East Java, and DI Yogyakarta province. In these regions, increasing the number of doctors would decrease the expected number of malnourished patients. This condition may be used as a recommendation to the government that these regions need additional health workers, especially doctors in order to decrease the malnutrition case. However, the local parameter estimates of Doctor variable were positive in the other regions. If we look into the map of parameter estimate in Doctor variable, the regions with positive local parameter estimates includes all of regencies or cities in DKI Jakarta, Banten, and most of the regions in West Java province. In these regions it was possible that the need of the number of doctors was high, especially in big cities like in DKI Jakarta province.

In the education aspect, the local parameter estimates of JHSch variable were negative in most regencies and cities in Java island, except for Wonogiri, Ngawi, Magetan, and Pacitan regency. Their values indicated that increasing the number of junior high school or equivalent buildings would decrease the expected number of malnourished patients. In the four other regions, it was possible that the increase of junior high school or equivalent buildings as an education aspect.
A representation did not have a direct influence to the number of malnourished patients.

The last variable is DiffPlace. All of the regencies and cities in Central Java and DI Yogyakarta have positive local parameter estimates. The model indicated that increasing the total village having insufficient access towards the village office and the capital of sub-district would increase the expected number of malnourished patients. Meanwhile, in other regions, DiffPlace variable was negatively related to the number of malnourished patients. It might be caused by more sufficient access towards the village office and the capital of sub-district, so there are many malnourished patients being recorded by the local government officials, especially by the health workers.

The local parameter estimates that were inversely related with some variables might produce bad interpretation. For better results, the GWPR model, the study of regions for the next research should be done at a lower, such as sub-district or village level. Using lower levels might enhance the precision of parameter estimates in spatial analysis.

**Significant Explanatory Variables Map**

GWPR model usually displays the map of non-stationary parameter estimates. However, every point of the regression, in this case each regency or city, it also has a pseudo t-value which is formulated as by the parameter estimate divided with the standard error in each point of the regression. As a result, the GWPR local parameters of explanatory variables (we exclude the intercept) have t-maps (Mennis 2006). Each t-map has an advantage that is it can detect the significancy of an explanatory variables for each region. The other advantage is it can detect the combination of significant explanatory variables for each region. But, the disadvantage of this mapping approach is that potentially interesting patterns may not be observed regarding the magnitude of the relationship between the explanatory and dependent variable as contained in the actual parameter estimate values, as well as in the magnitude of the significance (Mennis 2006).

In this research, every t-map of GWPR local parameter of explanatory variable would be joined as a map which is named as map of significant explanatory variable. Figure 5 shows the significant explanatory variable map. The explanatory variables used in the map are the explanatory variables used in the selected global model.

The map in Figure 5 displays six groups of color. The colors indicate that there were different groups without implying a rank like the map in Figure 4. The dominant color was dark brown, it indicates that in these regions, all of the explanatory variables were significant. A region where FamSlums and DiffPlace variable were not significant is Ponorogo regency. Tulungagung, Pacitan, and Majalengka regency only had DiffPlace variable as a significant variable. Then, FamRiver variable was significant in eastern area of East Java. FamSlums variable was not significant in Pemalang, Trenggalek, and Purbalingga regency. The last group is the regions which JHSch was not a significant variable. The regions include Blora, Wonogiri, and Ngawi regency. The other advantage for creating this map is it can give recommendations to policy maker or the local government and central government especially Ministry of Health about the factors influencing the number of malnourished patients in each regency or city in Java island.
CONCLUSION AND RECOMMENDATION

Conclusion
GWPR model was proven to be a better model than global model. GWPR model had highly difference of AICc compared to the global model. A condition of non-stationarity in parameters was fulfilled, and its residual relatively less than global model for each region. Map of parameter estimates showed more meaningful results. The FamRiver variable was the highest influencing factors to the number of malnourished patients compared to the other variables which represent the poverty aspect. There were six groups of the factors influencing the number malnourished patients for each regency and city in Java based on the map of significant explanatory variable.

Recommendation
For better results of GWPR model, the study of regions for the next research should lower level than regency or city level, such as the sub-district or village level. Conducting GWPR at this level might give a better local parameter estimates.

REFERENCES
### Appendix 1 Description of dependent variable and explanatory variables

#### Table 1 Description of dependent variable

<table>
<thead>
<tr>
<th>Name of dependent variable</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>MalPat</td>
<td>Total number of malnourished patients during last three years in each regency or city</td>
<td>Persons</td>
</tr>
</tbody>
</table>

#### Table 2 Description of explanatory variable

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Name of explanatory variable</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poverty</td>
<td>FamLabour</td>
<td>Total number of families whose members are farm labour in each regency or city</td>
<td>Families</td>
</tr>
<tr>
<td></td>
<td>FamRiver</td>
<td>Total number of families residing along the river in each regency or city</td>
<td>Families</td>
</tr>
<tr>
<td></td>
<td>FamSlums</td>
<td>Total number of families residing in the slums in each regency or city</td>
<td>Families</td>
</tr>
<tr>
<td></td>
<td>FamPovIns</td>
<td>Total number of families which received poor family health insurance (Asuransi kesehatan keluarga miskin/Askeskin) in each regency or city</td>
<td>Families</td>
</tr>
<tr>
<td>Health</td>
<td>Midwife</td>
<td>Total number of midwifes in each regency or city</td>
<td>Persons</td>
</tr>
<tr>
<td></td>
<td>Doctor</td>
<td>Total number of general doctors in each regency or city</td>
<td>Persons</td>
</tr>
<tr>
<td></td>
<td>ISP</td>
<td>Total number of integrated services post (Posyandu) in each regency or city</td>
<td>Units</td>
</tr>
<tr>
<td></td>
<td>ActISP</td>
<td>Total number of villages that all of integrated services post (Posyandu) in a village is actively operated in each regency or city</td>
<td>Villages</td>
</tr>
<tr>
<td>Education</td>
<td>HlthWrk</td>
<td>Total number of mantri kesehatan and dukun bayi in each regency or city</td>
<td>Persons</td>
</tr>
<tr>
<td></td>
<td>ElmSch</td>
<td>Total number of elementary schools or equivalent in each regency or city</td>
<td>Units</td>
</tr>
<tr>
<td></td>
<td>JHSch</td>
<td>Total number of junior high schools or equivalent in each regency or city</td>
<td>Units</td>
</tr>
<tr>
<td></td>
<td>Illiteracy</td>
<td>Total number of villages that there are eradication programme of illiteracy during one last year</td>
<td>Villages</td>
</tr>
<tr>
<td>Food</td>
<td>RiceField-Land</td>
<td>Total area of agricultural land for rice field in each regency or city</td>
<td>100 Ha</td>
</tr>
<tr>
<td></td>
<td>DiffPlace</td>
<td>Total number of villages having insufficient or poor access towards the village office and the capital of sub-district in each regency or city</td>
<td>Villages</td>
</tr>
</tbody>
</table>

Notes: 1). Regencies or cities in this research are in Java island (Regencies in Madura island and Seribu Archipelago are not included)  
2). Variable of FamLabour, FamRiver, FamSlums, FamPovIns, and ISP were similar with Rohimah’s research (2011)
Appendix 2  Algorithm of research method

Data exploration (using dependent variable distribution map and boxplot)

Explanatory variables selection (using Pearson’s correlation (r))

Higher Pearson’s correlation (r) with dependent variable

No

Explanatory variables reduction

Yes

GWPR modeling

Goodness of fit test between Global model and GWPR model

AICc difference ≤ 2

Yes

Global model

No

GWPR model

Parameter stationarity test (using simple test based on Chang et al., 2011)

Parameter estimate with IQR of GWPR model ≥ 2 times std. error of Global model

Non-stationary parameter estimate

Yes

Stationary parameter estimate

Interpretation of local parameter estimates

Non-stationary parameter estimate

Spatial variability map

Interpretation of Spatial variability map

The other spatial analysis (example: spatial dependency test)
Appendix 3  Map of malnourished patients distribution for each regency and city in Java island (2008)

Note: unit of malnourished patients in persons
Appendix 4 Boxplot of dependent variable and explanatory variables

Notes:
1. Poverty aspect: FamLabour, FamSlums, FamRiver, and FamPovIns variable
2. Health aspect: ISP, ActISP, Doctor, Midwife, and HlthWrk variable
3. Education aspect: ElmSch, JHSch, and Illiteracy variable
4. Food aspect: RiceFieldLand and DiffPlace variable
### Pearson’s correlation between variables (dependent variable and explanatory variables)

<table>
<thead>
<tr>
<th>Variable</th>
<th>FamLabour</th>
<th>FamRiver</th>
<th>FamSlums</th>
<th>FamPovIns</th>
<th>Illiteracy</th>
<th>ISP</th>
<th>ActISP</th>
<th>RiceField-Land</th>
<th>Doctor</th>
<th>Midwife</th>
<th>HlthWrk</th>
<th>DiffPlace</th>
<th>ElmSch</th>
<th>JHSch</th>
</tr>
</thead>
<tbody>
<tr>
<td>MalPat</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>FamLabour</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>FamRiver</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>FamSlums</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>FamPovIns</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>Illiteracy</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>ISP</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>ActISP</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>RiceField-Land</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>Doctor</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>Midwife</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>HlthWrk</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>DiffPlace</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
<tr>
<td>ElmSch</td>
<td>0.356</td>
<td>0.323</td>
<td>0.102</td>
<td>0.140</td>
<td>0.124</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.214</td>
<td>0.097</td>
<td>0.036</td>
<td>0.377</td>
<td>0.257</td>
<td>0.296</td>
<td>0.421</td>
</tr>
</tbody>
</table>

*Underlined variable is dependent variable. Underlined variable is dependent variable.*

*|r|≥0.7 is strong Correlation (source: C:\data\StatPrimer\correlation.wpd).*
Appendix 6  Summary statistics for varying (local) coefficients

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.817929</td>
<td>0.164236</td>
<td>4.601174</td>
<td>4.716957</td>
<td>4.871907</td>
<td>4.954810</td>
</tr>
<tr>
<td>FamLabour</td>
<td>0.000005</td>
<td>0.000002</td>
<td>0.000001</td>
<td>0.000003</td>
<td>0.000005</td>
<td>0.000006</td>
</tr>
<tr>
<td>FamRiver</td>
<td>0.000052</td>
<td>0.000024</td>
<td>-0.000010</td>
<td>0.000038</td>
<td>0.000053</td>
<td>0.000069</td>
</tr>
<tr>
<td>FamSlums</td>
<td>0.000033</td>
<td>0.000055</td>
<td>-3.20x10^-5</td>
<td>-0.000025</td>
<td>0.000036</td>
<td>0.000069</td>
</tr>
<tr>
<td>ISP</td>
<td>0.000858</td>
<td>0.000212</td>
<td>0.000406</td>
<td>0.000674</td>
<td>0.000940</td>
<td>0.001051</td>
</tr>
<tr>
<td>Doctor</td>
<td>-0.000710</td>
<td>0.000924</td>
<td>-0.001850</td>
<td>-0.001490</td>
<td>-0.001271</td>
<td>0.000502</td>
</tr>
<tr>
<td>DistPlace</td>
<td>-0.001510</td>
<td>0.005036</td>
<td>-0.010250</td>
<td>-0.006885</td>
<td>-0.003167</td>
<td>0.000390</td>
</tr>
<tr>
<td>Sch</td>
<td>-0.002040</td>
<td>0.001421</td>
<td>-0.005300</td>
<td>-0.002985</td>
<td>-0.002292</td>
<td>-0.000529</td>
</tr>
</tbody>
</table>

IQR = Inter-quartile range