CAUCHY –Jurnal Matematika Murni dan Aplikasi
Volume 5(3) (2018), Pages 140-149
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Geographically Weighted Regression to Predict the Prevalence of Hypertension Based on the Risk Factors in South Kalimantan

Suroto[1,3], Bambang Widjanarko Otok[2], Suharto[3], Arief Wibowo[3]

[1] Health Polytechnic Banjarmasin, Ministry of Health & Ph.D Student Faculty of Public Health, Airlanga University, Surabaya
[2] Laboratory of Environmental and Health Statistic, 'Sepuluh Nopember' Institute of Technology (ITS), Surabaya
[3]Department of Biostatistics and Demography, Faculty of Public Health, Airlanga University, Surabaya, Indonesia

Email: [1]surotojahrani@yahoo.com, [2]dr.otok.bw@gmail.com

ABSTRACT

Hypertension is one of the disease is not contagious diseases which is a public health problem. Uncontrolled Hypertension can trigger a degenerative disease such as congestive heart failure, renal failure and vascular disease. Hypertension is called the silent killer because his nature the condition is asymptomatic and can cause a fatal stroke. With the increasing prevalence of cases of degenerative diseases, one only hypertension, then the researchers want to predict the variables very big role as one of the risk factors of Genesis hypertension. With clearly know the risk factors that play against genesis hypertension is expected to be used as a reference for the prevention and control so that they can reduce the prevalence of hypertension and prevent deaths from degenerative diseases, especially hypertension. The results of the study showed that the results of the modeling the prevalence of hypertension in South Kalimantan Province using linier regression there is no factor that affect the genesis of hypertension. The prevalence of hypertension spread spatially because there are heterogenitas between the location of the observation that means that observations of a location depends on the observations in another location that the distance is near so do spatial regression modeling with adaptive gaussian kernel function, the result 5 groups. Group I consists of the districts *Tanah Laut* and *Tanah Bumbu;* group II, *Kota Baru*; Group III consists of *Banjar, Kota Banjar Baru, Kota Banjarmasin*; Group IV on the *Barito Kuala* Regency and the Group V consists of *Tapin, H S Selatan, H S Tengah, H S Utara, Tabalong, Balangan*.

Keywords: GWR, Kernel function, adaptive Gaussian, prevelance hypertension

INTRODUCTION

A research influenced by aspects of territorial characteristic (spatial) then need to be considered spatial data on the model. Spatial data is data that contains the location information. On spatial data, often observations in a location dependent on observation in other locations near (*neighboring*).

The first law of geography advanced by Tobler in 1979, stated that all things are related to each one with the other but something close more had the effect of something far [1]. The law is the basis of the examination of the problems based on the effect of the location or spatial method. In the modeling language, when the classic regression model used as a tool of analysis on spatial data, then can cause the conclusion that less accurate

Submitted: 03 July 2018 Reviewed: 15 August 2018 Accepted: 30 November 2018
DOI: http://dx.doi.org/10.18860/ca.v5i3.5879

because of the assumption of the error free to each other and the assumption homogenitas not met [2].

Hypertension is one of the disease is not contagious diseases which is a public health problem. Uncontrolled Hypertension can trigger a degenerative diseases such as congestive heart failure, renal failure and vascular disease. Hypertension is called the *silent killer* because his nature the condition is asymptomatic and can cause a fatal stroke [3]. Although not treated, prevention and managements can decrease the occurrence of hypertensi and the accompanying disease. Hypertension is the cause of the death of number 3 after stroke and tuberculosis, namely 6.7 % from the population of death in all age in Indonesia. The problem of hypertension that occurs in South Kalimantan would not escape from the factors causing the hypertension essential where genetic factors, environment, behavior, health services also contribute to cause a high case of hypertension in South Kalimantan Province [4]. To identify the multitude of causative factors that affect it may need to be an analysis or the development of the model is spatial [5].

Spatial effects testing done with test heterogenitas and spatial dependencies. If there is a settlement of securities heterogenity is by using point approach. Spatial regression points between the other Geographically Weighted Regression (GWR) with the scale of the measurement of the response variable is the interval and ratio,Geographically Weighted Poisson Regression (GWPR) with the variable data is response count, Geographically Weighted Logistic Regression (GWLR) with the scale of the measurement of the response variable is the nominal [6].

Based on the [7], the prevalence of hypertension in Indonesia that acquired through the measurement at the age $\geq$ 18 years of 25.8%, the highest in Bangka Belitung securities have totaled 30.9%, followed South Kalimantan 30,8 % , East Kalimantan 29.6% and West Java 29.4%. The prevalence of hypertension in Indonesia that is obtained through the questionnaire diagnosed health workers of 9.4%, diagnosed health workers or is drinking drugs of 9.5%. So there are 0.1% that medication itself. Respondents who have a normal blood but is drinking Hypertension medications 0.7%. So the prevalence of hypertension in Indonesia by 26.5%.

The problem of hypertension that occurs in South Kalimantan would not escape from the factors causing the hypertension essential where genetic factors, environment, behavior, health services also contribute to cause a high case of hypertension in South Kalimantan Province. To identify the multitude of causative factors that affect it may need to be an analysis or development models. With the increasing prevalence of hypertension in a region, then the researchers want to predict the variables very big role as one of the risk factors of Genesis hypertension. With clearly know the risk factors that play against genesis hypertension with GWR approach is expected to be used as a reference for the prevention and control so that they can reduce the prevalence of hypertension.

METHODS

The data used is the Basic Health Research data [7]. The data will be analyzed in this research is data genesis hypertension in regency/city of South Kalimantan Province. Based on the results of the analysis of the previous library, hypertension served on the following conceptual framework [8].

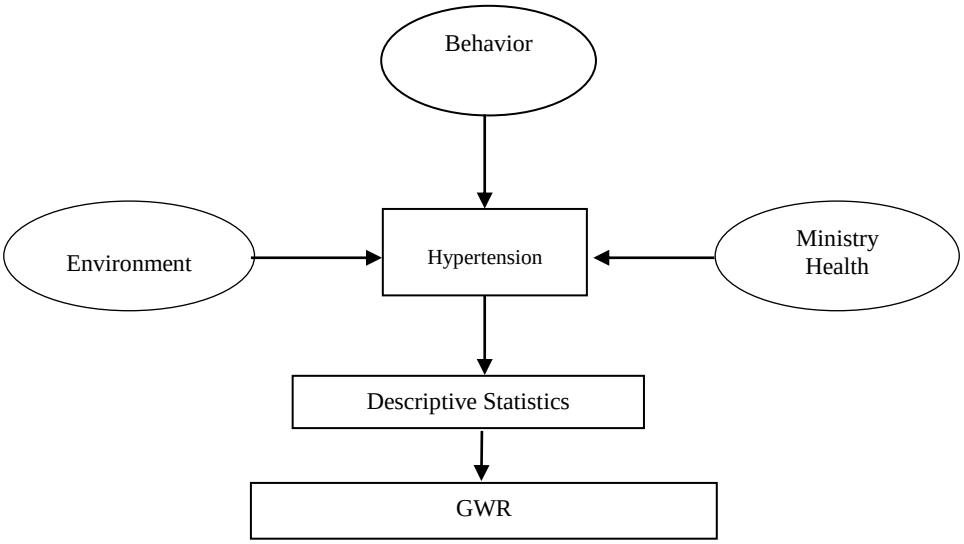

Figure 1. GWR Modeling flow in the case of hypertension

The response variable Y that is used in this research is the percentage of patients with hypertension are diagnosed per sub-districts in South Kalimantan Province. While the variables predictors Xi used there in Table 1.

Table 1. The variables Predictors [9]

Variables	Description
X_1	The percentage of the inhabitants of gender male
X_2	The percentage of population with education Completed SD/MI
X_3	The percentage of population with smoking habit every day
X_4	The percentage of the population of physical activity
X_5	Percentage of the population who consume the fruits of 7 times in 1 weeks
X_6	Percentage of the population who consume vegetables 7 times in 1 weeks
X_7	Percentage of the population who consume salty food more than 1 times per day
X_8	The percentage of the population consuming fatty food consumption/ order/ fried more than 1 times per day
X_9	The percentage of the population with ownership of health insurance

The variables X1 and X2 used to describe the spread of environmental aspects according to their geographic region. Aspects of the behavior described by the variables X3, X4, X5, X6, X7 and X8. While aspects of the Ministry of Health is portrayed through the variable X9.

RESULT AND DISCUSSION

Description of the Prevalence of Hypertension

Description of this research includes the mean and standard deviation from each of the research variables. Now in detail is presented in the following table.

Table 2. Descriptive Data the prevalence of hypertension in South Kalimantan Province [9]

Variables [7]	Minimum	Maximum	Mean	StDev
Persentase patients with hypertension is diagnosed (Y)	1.471	29.508	14.978	5.764
The percentage of the inhabitants of gender male (X1)	32.787	55.556	47.522	3.904
The percentage of population with education Completed SD/MI (X2)	7.320	69.230	33.660	12.340
The percentage of population with smoking habit every day (X3)	9.375	66.667	23.750	7.507
The percentage of the population physical activity (X4)	54.348	100,000	86.855	9.988
Percentage of the population who consume the fruits of 7 times in 1 weeks (X5)	0.000	41.584	20.289	7.070
Percentage of the population who consume vegetables 7 times in 1 weeks (X6)	13.390	93.100	55.470	15.200
Percentage of the population who consume salty food more than 1 times per day (X7)	22.220	86.360	45.080	13.170
The percentage of the population consuming fatty food consumption/ order/ fried more than 1 times per day (X8)	0.000	79.550	31.670	15.620
The percentage of the population with ownership of health insurance (X9)	0.000	54.023	15.002	11.946

Table 2. shows that the percentage of the highest Hypertension Prevalence on the sub-district *Pulau Laut Tanjung Selayar* of 29.508 percent and lowest on sub-district *Tatah Makmur* of 1.471 percent while the average percentage of the prevalence of hypertension by 14.978 percent with standard deviation of 5.764 percent. The percentage of the inhabitants of gender male (X1) highest on the sub-district *Danau Panggang* (55.556 percent), and the lowest on the sub-district *Pulau Laut Tanjung Selayar* (32.787). The percentage of population with education Completed SD/MI (X2) highest on the sub-district *Pamukan Barat* of 69.230 percent and lowest on sub-district *Tatah Makmur* of 7.320 per cent while the average percentage of the prevalence of hypertension by 33.660 percent with standard deviation of 12.340 percent. The percentage of population with smoking habit every day (X3) has an average of 23.750 percent with standard deviation 7.507 percent. The percentage of the population physical activity (X4) has an average of 86.855 percent with standard deviation 9.988 percent. Percentage of the population who consume the fruits of 7 times in 1 weeks (X5) has an average of 20.289 percent with standard deviation 7.070 percent. Percentage of the population who consume vegetables 7 times in 1 weeks (X6) has an average of 55.470 percent with standard deviation 15.2 percent. Percentage of the population who consume salty food more than 1 times per day (X7) has an average of 45.08 percent with standard deviation 13.170 per cent. The percentage of the population consuming fatty food consumption/ order/ fried more than 1 times per day (X8) has an average of 31.670 percent with standard deviation 15.62 percent and the percentage of the population with ownership of health insurance (X9) has an average of 15.002 percent with standard deviation 11.946 percent.

Regression modeling the prevalence of hypertension

Before performing the analysis using spatial regression method, done regression modeling double linier first. Regression modeling linear to the prevalence of hypertension and factors suspected to influence using parameter assessment method Ordinary Least Square (OLS) [10] which aims to know the variables significant on the prevalence of hypertension globally. The first step is to detect multicolinearity to know whether or not the relationship between the free variable (predictors), then continued with double linier regression modeling (global) covers the review of the significance of the parameters simultaneously or partially, and residual assumptions IIDN test [11].

Multicolinearity Detection

One of the conditions in the multiple regression analysis with some variables predictors is no cases multikolinieritas or not there is a variable predictors that have a correlation with other predictors variable. Tracing muticolinearity done based on the value of Variance Inflation Factor (VIF). The following is the value of VIF in each of the variables predictors.

Table 3. The value of their respective VIF Predictors Variables

Variables	X1	X2	X3	X4	X5	X6	X7	X8	X9
VIF	1.287	1.279	1.313	1.093	1.153	1.347	1.194	1.208	1.422

Based on the Table 3, obtained the information that all variables predictors have a VIF value less 10. This detects that there are not cases multikolinieritas or not there is a variable predictors that have a correlation with other predictors variable.

The Significance of Parameters Linear Regression tests the prevalence of hypertension

The following is the significance test good linier regression parameters simultaneously or partial to know the influence of predictors variables used. The hypothesis to test the significance of the parameters simultaneously on the linier regression is as follows [6].

H0 : $\beta 1 = \beta 2 = ... = \beta 9 = 0$ (the parameters do not affect the significant impact on model)

H1 : at least one $\beta k \neq 0$; $k = 1.2, ... ,9$ (at least one parameters that affect the significant impact on model)

Table 4. ANOVA table the prevalence of hypertension in South Kalimantan Province

Source	Sum of Squares	df	Mean of Square	F	p
Regression	598.32	9	66.48	2.14	0.030
Residual	4418.32	142	31.11		
Total	5016.64	151			

Table 4., produce F-Statistic value of 2.14 and p-value of 0.030. Based on the level of the significance (α) of 5 percent and $F_{(0.05;9;142)}$ of 1,.46, obtained the decision Reject H0 because the value of F-Statistic > $F_{(0.05;9;142)}$ or p-value < 0.05. This can be interpreted that there is at least one parameters that affect the significant impact on the prevalence of hypertension.

Next to know the variables predictors anywhere that provides significantly influence, then the test is done the significance of parameters partially presented in Table 5. The following is the hypothesis test the significance of the parameters spatially against linier regression model (global) [6].

H0 : $\beta k = 0$,

H1 : $\beta k \neq 0$, $k = 1,2,3,...,9$

Table 5. The results of the Test Regression Model Parameters Linear Partial

Parameters	coefficient	SE Coefficient	T-Statistic	*Sig.*	The Decision
β_0	28.9260	6.53000	4.43	0.000	
β_1	-0.1728	0.08935	-1.93	0.055	Fail to reject H0
β_2	0.0181	0.04155	0.44	0.663	Fail to reject H0
β_3	0.0031	0.05605	0.05	0.956	Fail to reject H0
β_4	-0.0447	0.04784	-0.94	0.351	Fail to reject H0
β_5	-0.1015	0.06937	-1.46	0.146	Fail to reject H0
β_6	0.0284	0.03464	0.82	0.414	Fail to reject H0
β_7	-0.0160	0.03780	-0.42	0.673	Fail to reject H0
β_8	-0.0555	0.03205	-1.73	0.086	Fail to reject H0
β_9	0.0252	0.04528	0.56	0.579	Fail to reject H0

Based on the results of the test in table 5, with significant level (α) of 5 percent and $t_{(\alpha/2;n-p-1)} = t_{(0.025,142)} = 1.977$, Obtained the information that all values T count smaller than t table. This shows that there is no predictors variables that affect the prevalence of hypertension.

Testing the Assumptions of a Residual IIDN

After testing the significance of the parameters simultaneously and partially, then the next step is to test the assumptions residual of identical, independent, and normal distribution (IIDN).

- **Test the assumption of identical Residual**

One of the assumption of the test in the OLS regression is a residual must be homoscedasticitys variance (is identical) or in cases of heteroscedasticity. How to identify the existence of the case of heteroskedastisitas is to create a regression model between a residual and predictor variables. When there are variables predictors that affect the model significantly, it can be said that the residual is not identical or in cases of heteroscedasticity. Testing the assumption of identical residual provides information that there are not cases heteroscedasticity or identic residual with significant (α) of 0.05 and $\alpha\, F_{(\alpha;n-p-1)} = F_{(0,05;9,142)} = 1.946$. This is due to the value of the p-*Value* of 0.002 smaller

than α and F-Statistic greater 3.06 than $F_{(0,05;9,142)} = 1.946$ then happened heteroscedasticity.

- **Test the assumption of Independent Residual [11]**

Test the assumption of independent residual used to know whether or not the relationship between the residual periods. The test statistics used is Durbin-Watson. Based on the attachment 4 obtained the value $d = 2.07875$ With the value $d_L = 1.0201$ and $d_U = 1.9198$. So that the decision can be taken is to fail to reject H0 because $d_U = 1.9198 < d < (4 - d_U) = 2.0802$. It shows that there is no relationship between the residual, so that the assumption of independent residual have been fulfilled.

- **Test the assumption Normal distribution**

The assumption of the normal distribution test is done with the following Kolmogorov-Smirnov test.

H0: Data normal distribution

H1: Data not normal distribution

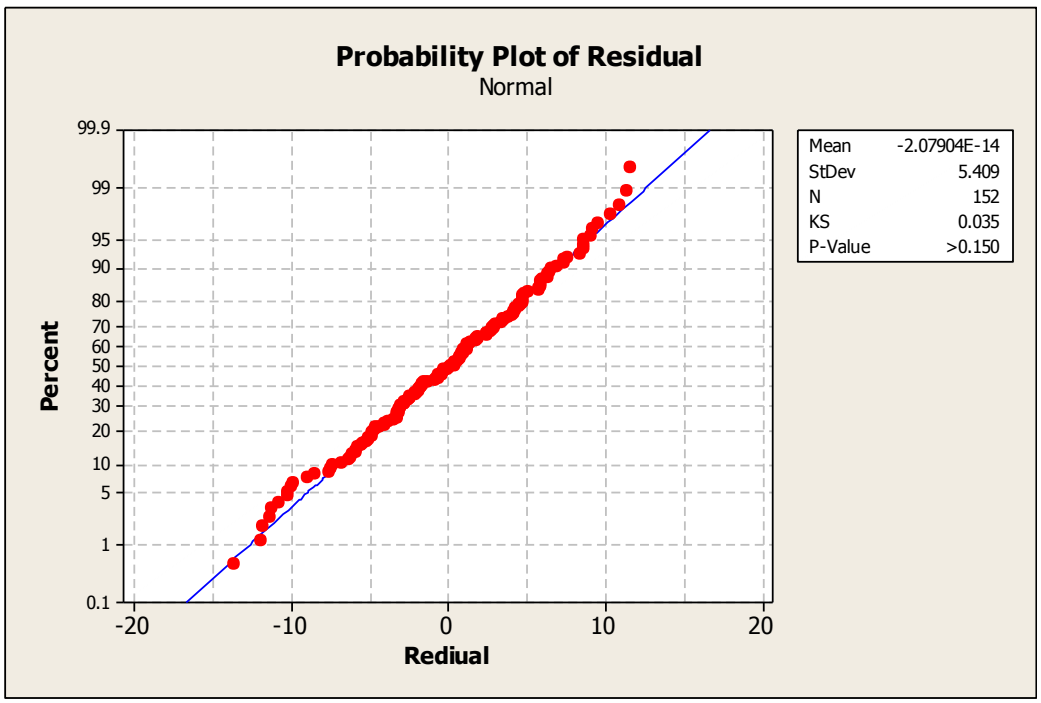

Figure 2. Probability plots of Residual of the Hypertension prevalence

Based on the Figure 2 obtained the information that the points of red spread near linear line (normal) which means that the data has been normal distribution. In addition, also can be seen from the value of *P-value* that is more > 0.15. So that the decision can be taken is to fail to reject H0 on equal significant (α) of 5%, because the value of the *P-value* is greater than α. This means that the data has been meet the assumption normal distribution. Based on the results of the test of assumption, it can be concluded that a residual on the linear regression model (global) meets the independent assumption, and data has been normal distribution, but the assumption is identical not fulfilled.

Spatial Regression Modeling the Prevalence of Hypertension

The analysis using the GWR method aims to know the variables that affect the prevalence of Genesis hypertension on each observation location regency/city in South Kalimantan province. The first step is done to get the GWR model is to determine the coordinates of the point latitude and longitude on each location to count the distance euclidean and determine the optimum bandwidth values based on the criteria of Cross validation (CV). The next step is to determine the matrix pembobot with kernel functions: Fixed Gaussian, fixed bi-square, Adaptive Gaussian, Adaptive Bi-Square and assess the GWR model parameters. The matrix pembobot obtained for each location and then used to form a model, so that obtained the model vary in each location of observation [12].

GWR model hypothesis testing consists of two test, test the suitability of the GWR model and test the significance of the parameters GWR model. The following is the results of the hypothesis testing GWR model [6].

H0 : $\beta_k(u_i, v_i) = \beta_k$;

(There is no significant difference between the linear regression model (global) and GWR model)

H1 : at least one $\beta_k(u_i, v_i) \neq \beta_k$ K = 1, 2, ...,9

(No difference between significant linear regression model (global) and GWR model)

Table 6. The estimation of GWR on the weight of the Kernel Function

Statistic	Weight			
	Fixed Gaussian	**Fixed Bi-Square**	**Adaptive Gaussian***	**Adaptive Bisquare**
Bandhwith	0.879990	1.802952	148.831349	152.000000
MSE	28.141	27.681	26.794	28.429
R2	0.265928	0.290850	0.302022	0.252414
AICc	956.170600	956.662511	**956.153605**	957.642999

Note: *) Best GWR Model

Table 6 shows the comparison of the model with pembobot GWR estimates that vary. Testing the suitability of the GWR model is done by using the difference in the number of residual square GWR model and global regression model. GWR Model will vary significantly with global regression model if can reduce the amount of residual square significantly. Table 6., shows that the value of the smallest AICc is the GWR model with adaptive Gaussian kernel pembobot amounting 956.153. So using the significance level (α) 5 percent so it can be concluded that the GWR model differ significantly with global regression model. This means that the kernel pembobot GWR model with adaptive Gaussian more worthy to illustrate the percentage of the prevalence of hypertension in South Kalimantan Province.

Next is a test of the significance of the parameters GWR model with adaptive Gaussian kernel pembobot partially to know any parameters that affect the prevalence of hypertension in each location of observation. Sub Division that have common

variables which affect the significant impact on the prevalence of hypertension is presented in Table 7.

Table 7. Grouping of Districts based on a significant Variable

Group	Significant variables	Variable Description
Tanah Laut	X3,X4,X6	Persentase patients with hypertension is diagnosed (Y) The percentage of the inhabitants of gender male (X1) The percentage of population with education Completed SD/MI (X2) The percentage of population with smoking habit every day (X3) The percentage of the population physical activity (X4) Percentage of the population who consume the fruits of 7 times in 1 weeks (X5) Percentage of the population who consume vegetables 7 times in 1 weeks (X6) Percentage of the population who consume salty food more than 1 times per day (X7) The percentage of the population consuming fatty food consumption/ order/ fried more than 1 times per day (X8) The percentage of the population with ownership of health insurance (X9)
Tanah Bumbu	X3,X4,X6	
Kota Baru	X4,X8	
Banjar	X4,X5,X6,X7	
Kota Banjarmasin	X4,X5,X6,X7	
Banjar Baru	X4,X5,X6,X7	
Barito Kuala	X3,X4,X9	
Tapin	X1,X5	
Hulu Sungai Selatan	X1,X5	
Hulu Sungai Tengah	X1,X5	
Hulu Sungai Utara	X1,X5	
Tabalong	X1,X5	
Balangan	X1,X5	

Clustering based on a significant variable is divided into 5 groups. There are districts that have common variables which significantly influence with the surrounding district, but there is a district that has its own uniqueness because the variables that influence significant is not the same.

CONCLUSION

The results of the modeling the prevalence of hypertension in South Kalimantan Province based on using sub-linier regression there is no factor that affect the genesis of hypertension. The prevalence of hypertension spread spatially because there are heterogenitas between the location of the observation that means that observations of a location depends on the observations in another location that the distance is near so do spatial regression modeling with Adaptive Gaussian kernel function, to result 5 groups. Group I consists of the districts of *Tanah Laut* and *Tanah Bumbu* with the characteristics of the percentage of the population with the smoking habit every day (X3), the percentage of the population physical activity (X4), the percentage of the population who consume vegetables 7 times in 1 weeks (X6). The group II, *Kota Baru* with the characteristics of the percentage of physical activity (X4), the percentage of the population who consume fatty food consumption/ order/ fried more than 1 times per day (X8). Group III consists of *Banjar*, *Kota Banjar Baru*, *Kota Banjarmasin*, with characteristic of the percentage of the population physical activity (X4), the percentage of the population who consume the fruits of 7 times in 1 weeks (X5), the percentage of the population who consume vegetables 7 times in 1 weeks (X6), the percentage of the population who consume salty food more than 1 times per day (X7). The group IV on the *Barito Kuala* Regency with the characteristics of a percentage of the population with the smoking habit every day (X3), the percentage of the population physical activity (X4),

the percentage of the population with the ownership of health insurance (X9), and the Group V consists of *Tapin, H S Selatan, H S Tengah, H S Utara, Tabalong, Balangan* with characteristics of the percentage of the population of the sexes men (X1), the percentage of the population who consume the fruits of 7 times in 1 weeks (X5).

REFERENCES

[1] L. Anselin, *Spatial Econometrics: Methods and Models.* Kluwer Academic, Dordrecht, 1988

[2] H. J. Miller, 'Tobler's First Law and Spatial Analysis'. *Annals of the Association of America Geographers,* 94(2), hal.284-289, 2004.

[3] H. J. Tudor, F. Tom, *Tanya Jawab Seputar Tekanan Darah Tinggi*, Edisi 2, Arcan, Jakarta, 2010.

[4] L. Marliani, *100 Question & Answers Hipertensi.* PT Alex Media Komputindo Gramedia. Jakarta, 2007.

[5] C. Chasco, I. Garcia, and J. Vicens, ''Modeling Spatial Variations in Household Disposible Income with Geographically Weighted Regression*", Munich Personal RePEc Arkhive (MPRA) Working Papper* No. 1682, 2007.

[6] A. S. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression*, Jhon Wiley & Sons, Chichester, UK, 2002.

[7] Kementrian Kesehatan RI, *Riset Kesehatan Dasar (Riskesdas 2013),* Badan Penelitian dan Pengembangan Kesehatan, Jakarta, 2014.

[8] A. Mansjoer, *Kapita Selekta Kedokteran,* Edisi 3 , Media Aesculapius, FK UI, Jakarta, 2005.

[9] Dinas Kesehatan Propinsi Kalimantan Selatan, *Profil Kesehatan Propinsi Kalimantan Selatan*, Dinkes Prop. Kalsel, Banjarmasin, 2014.

[10] J. P. LeSage, *A Family of Geographically Weighted Regression*, Departement of Economics University of Toledo, 2001.

[11] Y. Leung, C. L. May, and W. X. Zhang, "Testing for spatial autocorrelation among the residuals of the geographically weighted regression", *Environment and Planning A,* 32, 871-890, 2000.

[12] Y. Leung, C. L. May, and W. X. Zhang, "Statistic Tests for Spatial Non-Stationarity Based on the Geographically Weighted Regression Model", *Environment and Planning* A, 32 9-32, 2000.