Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 3 (2015), pp. 1393-1397

http://www.ripublication.com

Analysis of Effect of Meteorological Factors on The Number of Dengue Fever Patients With Multiple Linear Regression

J. Mekparyup[1], K. Saithanu[2*] and N. Treewong[3]
[1,2,3]Department of Mathematics, Faculty of Science, Burapha University
169 Muang, Chonburi, Thailand
[1]jatupat@buu.ac.th, [2]corresponding author: ksaithan@buu.ac.th,*
[3]be_bear424@hotmail.com

Abstract

The purpose of this research was to forecast the number of monthly dengue fever patients (y) in Pantong district, Chonburi, with meteorological factors (temperature, humidity, atmospheric pressure, pressure, wind speed and rain full) by multiple linear regression (MLR). The results of study found that the estimated regression equation of monthly number of dengue fever patients was $\hat{y}' = 28.84 - 1.2885x_7 + 0.00000033x'_{10} + 0.13098x_{12}$ with 0.68594 for standard error of estimation and 0.276 of adjusted coefficient of determination.

Mathematics Subject Classification: 62J05

Keywords: multicollinearity, variance inflation factor, best subset method

Introduction

An outbreak of dengue fever is a major problem throughout the world. Due to 30 years ago, the disease has spread to 100 countries and people have been suffering from dengue fever reaching to 2,500 million all over the world. The first outbreak was discovered in Philippines in 1954, the name was first called Philippine Hemorrhagic Fever (PHF), and later distributed to Thailand in 1958 with 2,158 cases. The Bureau of Vector-Borne Diseases in Thailand announced the rate of dengue hemorrhagic fever was 234.86 per hundred thousand populations in 2012 which increased dramatically in double from the year 2012 [1]. Obviously, an outbreak of dengue fever raised in every year so this study was intended to forecast the number of dengue fever patients for control planning of dengue outbreaks in the future.

Materials and Methods

Department of Health in Chonburi, Thialand provided data since 2007 to 2012. The three steps of this research were as follows.

Building The Multiple Linear Regression Equation

The multiple linear regression (MLR) equation to estimate monthly number of dengue fever patients was generated by regression model [2][3][4][5] as of Equation 1.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} + \beta_{12} x_{12} + \beta_{13} x_{13} + \beta_{14} x_{14} + \varepsilon \quad (1)$$

The model was composed of one dependent variable; y=number of monthly dengue fever patients in Pantong district, Chonburi, and fourteen independent variables; x_1=monthly average temperature, x_2=monthly average relative humidity, x_3=monthly average atmospheric pressure, x_4=monthly average pressure, x_5= monthly average wind speed, x_6=monthly average rainfall, x_7=days of monthly average rainfall ≥ 20 ml, x_8=days of monthly average rainfall ≥ 25 ml, x_9=days of monthly average rainfall ≥ 30 ml, x_{10}=days of monthly average relative humidity ≥ 40%, x_{11}=days of monthly average relative humidity ≥ 50%,, x_{12}=days of monthly average relative humidity ≥ 60%,, x_{13}=days of monthly average relative humidity ≥ 70%,, x_{14}=days of monthly average relative humidity ≥ 80%,, and ε=error of regression model.

Checking Assumptions of MLR

After obtained the best fitted MLR equation, the assumptions of MLR was considered. There are four assumptions to be checked; (I) normality of the error distribution using Anderson-Darling statistic by $AD = -n - \sum_{i=1}^{n} \frac{2i-1}{n} \left[\ln F(Y_i) + \ln(1 - F(Y_{n+1-i})) \right]$ [6]; (II) independence of the errors using Durbin-Watson statistic by $DW = \sum_{i=1}^{n} (\varepsilon_i - \varepsilon_{i-1})^2 \Big/ \sum_{i=1}^{n} \varepsilon_i^2$ [7]; (III) homoscedasticity (constant variance) of the errors using Breusch-Pagan statistic by $BP = \frac{SSR^*}{2} \div \left(\frac{SSE}{n} \right)^2$ [8] where SSR^* = Sum of squares in regression between e_j^2 = jth residual and x_{ij}, SSE = Sum of squares in regression error between y_j and x_{ij}; (IV) multicollinearity among predictor variables using Variance Inflation Factor (VIF) by $VIF_j = \frac{1}{1 - R^2_{j|others}}$ where $R^2_{j|others}$ = Multiple coefficient of determination between x_{ij} and all x_i. After testing all assumptions, the real values (RV) and the estimated values (EV) of the number of monthly dengue fever patients from the obtained MLR equation were plotted for making comparison.

Validating Between The RV And The PV

The comparison between RV and PV of the number of monthly dengue fever patients was lastly considered by the obtained MLR equation. Then, time series plot and scatter plot were generated to check how well this estimated regression equation forecast the number of monthly dengue fever patients.

Results and Discussion

Many positive correlations between y and x_i were found such as y and x_{13} with R=0.266 (p-value=0.000), y and x_3 with R=0.221 (p-value=0.000) which was the same previous studies [2][3][4][5]. According to the best subsets method, x_{13} was selected to build the possible MLR equation $\hat{y} = 3.701 + 0.3269x_{13}$ with the Mallow C-p= –1.8, standard error of estimation (S=10.366) and the adjusted coefficient of determination (R^2_{adj} =0.055). Once receiving the equation, (I) the test of normality was determined and found that hypothesis testing of normality was not satisfied with AD=3.483 (p-value=0.005). Box-Cox method was used to transform y and the MLR was rebuilt. The fitted MLR equation was $y' = 28.84 - 1.2885x_7 + 0.00000033x'_{10} + 0.13098x_{12}$ where $y' = \ln y$ and $x'_{10} = (x_{10})^5$. All assumptions were then retested as follows; (I) the test of normality was satisfied with *AD*=0.387 (p-value=0.374). (II) The test of independence of errors was tested by Durbin-Watson statistic value (DW=1.342, DL=1.338, DU=1.659) so the errors were independent. (III) The test of homoscedasticity of error variation was calculated by Breush-Pagan statistic (BP=7.629, p-value=0.054) so the error variances were constant. (IV) Test of multicollinearity: the VIF values of x_7, x_{10} and x_{12} were manifested and all values were less than 5 then there was no relationship among independent variables in multiple regression equation [9]. After all assumptions were validated, plotting between RV and PV was compared by graph of time series and scatter plot as Figure (1a) and Figure (1b) respectively. It showed that Figure 1(a) was closely plotted with the correlation coefficient 0.562 illustrated in Figure 1(b).

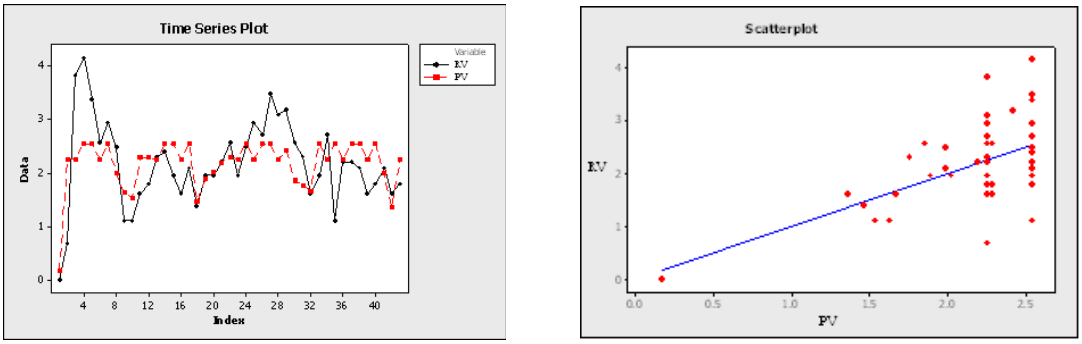

Figure 1: Comparison between RV and PV of the number of monthly dengue fever patients; (a) Time series plot, (b) Scatter plot

Discussion

The meteorological factors used to estimate the number of monthly dengue fever patients in Pantong district, Chonburi, were days of monthly average rainfall ≥ 20 ml (x_7), days of monthly average relative humidity ≥ 40% (x_{10}) and days of monthly average relative humidity ≥ 60% (x_{12}), with R^2_{adj}=0.276 and the S=0.68594. The accuracy of estimation was displayed rather good prediction with 0.562 of the coefficient of correlation in Figure 1(a) and 1(b).

Acknowledgement

We are grateful to Department of Health in Chonburi, Thialand for kindly providing all data.

References

[1] Bureau of Vector-Borne Diseases, 2013, "Dengue monitoring," Retrieved December 20, 2013, from Web site: http://www.thaivbd.org/n/dengues?module=%E0%B8%AA%E0%B8%96%E0%B8%B2%E0%B8%99%E0%B8%81%E0%B8%B2%E0%B8%A3%E0%B8%93%E0%B9%8C%E0%B9%84%E0%B8%82%E0%B9%89%E0%B9%80%E0%B8%A5%E0%B8%B7%E0%B8%AD%E0%B8%94%E0%B8%AD%E0%B8%AD%E0%B8%81&type=week&year=2556

[2] Goto, K., Kumarendran, B., Mettananda, S., Gunasekara, D., Fujii, Y., & Kaneko, S., 2013, "Analysis of Effects of Meteorological Factors on Dengue Incidence in Sri Lanka Using Time Series Data," PloS one, 8(5), e63717.

[3] Shang, C. S., Fang, C. T., Liu, C. M., Wen, T. H., Tsai, K. H., & King, C. C., 2010, "The role of imported cases and favorable meteorological conditions in the onset of dengue epidemics," PLoS neglected tropical diseases, 4(8), e775.

[4] Azil, A., Long, S. A., Ritchie, S. A., & Williams, C. R., 2010, "The development of predictive tools for pre‐emptive dengue vector control: a study of Aedes aegypti abundance and meteorological variables in North Queensland, Australia," Tropical Medicine & International Health, 15(10), 1190-1197.

[5] Thai, K. T., Cazelles, B., Van Nguyen, N., Vo, L. T., Boni, M. F., Farrar, J., ... & de Vries, P. J., 2010, "Dengue dynamics in Binh Thuan province, southern Vietnam: periodicity, synchronicity and climate variability," PLoS neglected tropical diseases, 4(7), e747.

[6] Lewis, P.A.W., 1961, "Distribution of the Anderson-Darling Statistic," The Annals of Mathematical Statistics, 32(4), 1118-1124.
[7] Durbin, J., & Watson, G.S., 1951, "Testing for Serial Correlation in Least Squares Regression II," Biometrika, 38(2), 159-177.
[8] Breusch T.S., & Pagan, A.R., "A Simple Test for heteroscedasti city and Random Coefficient Variation," Econometrica, 47(5), 1287-1294.
[9] Kutner, M.H., Christopher, J.N., & Neter, J., 1996, "Applied liner regression models, 4th ed," McGraw-Hill/Irwin , USA.