Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 3 (2015), pp. 1387-1391

http://www.ripublication.com

Modeling For Predicting The Monthly Average Number of Influenza Cases With Weather Factors In Chonburi

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

Abstract

The objective of present study was to build an appropriate model to predict the monthly average number of Influenza cases (y) in Chonburi, Thailand, by using a multiple linear regression equation with 27 weather factors; 13 temperature factors, 7 relative humidity factors, 3 atmospheric pressure factors, 3 wind speed factors and 1 rainfall factor. The study results revealed that the multiple linear regression equation for prediction the monthly average number of Influenza cases was $\hat{y}' = 0.4 + 0.007x_2 - 0.026x_6 - 0.002x_{16} + 0.020x_{21} + 0.021x_{23} - 0.025x_{27}$ with 0.031 for standard error of estimation and 0.342 for adjusted coefficient of determination.

Mathematics Subject Classification: 62J05

Keywords: multicollinearity, variance inflation factor, best subset method

Introduction

According to the Bureau of Epidemiology in 2013, it was found that Chonburi province had the highest number of cases of influenza in the eastern area especially in winter. Chonburi Public Health Office began a campaign to decrease the spread of influenza outbreak, but the epidemic was still progressing. As this such problems mentioned, the present study intended to fit the model for forecasting the monthly average number of influenza cases due to the numerous weather factors. The results of study will be used for making prevention and treatment plan of epidemic and pandemic of influenza that will occur in the future.

Materials and Methods

The monthly average number of influenza cases (y) and the 27 weather factors [1][2][3][4][5] were collected from Chonburi Public Health Office since 2007 to 2012 for building multiple linear regression (MLR) equation as follows: average temperature (x_1), difference between maximum and minimum temperature of the day (x_2), average relative humidity (x_3), average station atmospheric pressure (x_4), average vapor pressure (x_5), average wind speed (x_6), average rainfall (x_7), average atmospheric pressure at MSL (x_8), days of average temperature ≥ 20°C (x_9), days of average temperature ≥ 25°C (x_{10}), days of average temperature ≥ 30°C (x_{11}), days of relative humidity ≥ 40% (x_{12}), days of relative humidity ≥ 50% (x_{13}), days of relative humidity ≥ 60% (x_{14}), days of relative humidity ≥ 70% (x_{15}), days of relative humidity ≥ 80% (x_{16}), days of relative humidity ≥ 90% (x_{17}), days of maximum wind speed ≥ 5 m/s (x_{18}), days of maximum wind speed ≥ 10 m/s (x_{19}), days of maximum temperature ≥ 25°C (x_{20}), days of maximum temperature ≥ 30°C (x_{21}), days of maximum temperature ≥ 35°C (x_{22}), days of maximum temperature < 30°C (x_{23}), days of minimum temperature ≥ 20°C (x_{24}), days of minimum temperature ≥ 25°C (x_{25}), days of minimum temperature ≥ 30°C (x_{26}), days of minimum temperature < 30°C (x_{27}), Firstly, simple correlation coefficients (R) were calculated to identify relationship among these factors. Then the MLR equation for prediction the monthly average number of Influenza cases was generated following model (1)

$$\begin{aligned} y = {} & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} \\ & + \beta_{11} x_{11} + \beta_{12} x_{12} + \beta_{13} x_{13} + \beta_{14} x_{14} + \beta_{15} x_{15} + \beta_{16} x_{16} + \beta_{17} x_{17} + \beta_{18} x_{18} \\ & + \beta_{19} x_{19} + \beta_{20} x_{20} + \beta_{21} x_{21} + \beta_{22} x_{22} + \beta_{23} x_{23} + \beta_{24} x_{24} + \beta_{25} x_{25} + \beta_{26} x_{26} \\ & + \beta_{27} x_{27} + \varepsilon \qquad (1) \end{aligned}$$

where $\beta_i =$ the regression coefficient $(i = 0,1,2,...,27)$ and ε = error of the regression model. The best subset method was used to select the appropriate MLR equation by considering Mallows' C_p [6], standard error of estimation (S) and adjusted coefficient of determination (R^2_{adj}). The selected equation was then tested by F statistic of analysis of variance (ANOVA). After receiving the appropriate MLR equation, 4 assumptions of MLR analysis were testified in accordance with (I) Normal distribution of the error was tested by Anderson-Darling statistic [7], (II) Independence of the errors was examined by Durbin-Watson statistic [8], (III) Homoscedasticity of the errors was monitored by Breusch-Pagan statistic [9] and (IV) Multicollinearity among independent variables was verified by Variance Inflation Factor (VIF), $VIF_j = \dfrac{1}{1 - R^2_{j|others}}$, where $R^2_{j|others}$ is the coefficient of multiple determination with independent variable x_j on the $p-2$ other independent variables x in the MLR model (p is the number of independent variables). Box-Cox transformation [10] was used when any of these 4 assumptions was violated. Finally, validation of the fitted MLR equation was represented by comparing time series graphs between the observed data (OBS) and the predicted data (PRE).

Results and Discussion

The two highest positive and negative correlations were indicated between y and x_{17} (R=0.308, *p-value*=0.000) and y and x_8 (R= –0.259, *p-value*=0.000) which were the same previous studies [1][2][3][4][5]. The results of the best subsets method showed that x_8 and x_{17} were selected to generate the MLR equation $\hat{y} = 12848 - 12.6x_8 + 71.1x_{17}$ with the Mallow C_p = –2.1, S=87.147 and R^2_{adj}=0.215. The assumptions were tested following; (I) the test of normality was determined with AD=4.032 (*p-value*=0.005) so this assumption was not satisfied. Box-Cox transformation method was used to regenerate y and the MLR was reconsidered by the best subset method. The fitted MLR equation was $\hat{y}' = 0.400 + 0.00748x_2 - 0.0255x_6 - 0.00181x_{16} + 0.0202x_{21} + 0.0214x_{23} - 0.0254x_{27}$ where $\hat{y}' = 1/\sqrt{y}$. Then 4 assumptions were retested following; (I) the test of normality was satisfied with AD=0.189 (*p-value*=0.752), (II) The test of independence of the errors was tested by Durbin-Watson statistic value (DW=1.28, DL=1.25 and DU=1.64) so the errors were independent, (III) The test of homoscedasticity of error variation was tested by Breush-Pagan statistic (BP=4.3695, *p-value*=0.6268) so the error variances were constant. (IV) Test of multicollinearity: the *VIF* values of x_2, x_6, x_{16}, x_{21}, x_{23} and x_{27} were calculated and all values were less than 5 then there was no relationship among independent variables in multiple regression equation [11]. After all assumptions were validated, plotting between predicted y (PRE) and the observed values (OBS) was compared by graph of time series and scatter plot as Figure (1a) and Figure (1b) respectively. It was shown that the both graphs were closely plotted with the correlation coefficient 0.635.

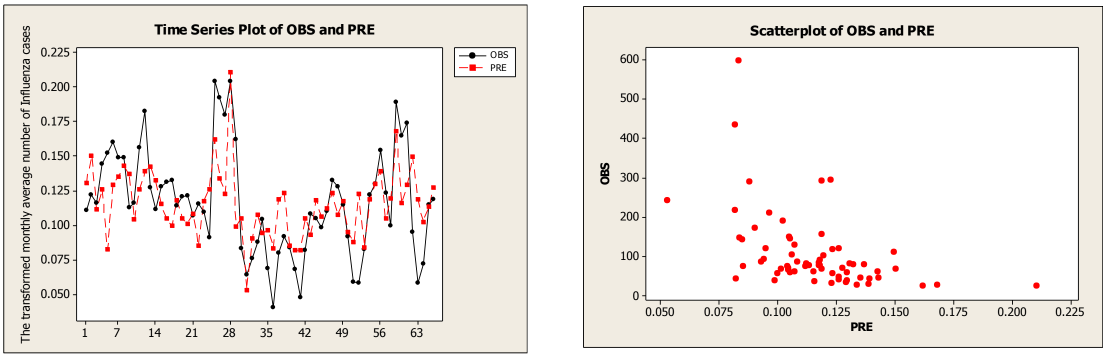

Figure 1: Comparison between OBS and PRE (a) Time series plot, (b) Scatter plot

Discussion

The factors used to predict the monthly average number of influenza cases in Chonburi were difference between maximum and minimum temperature of the day (x_2), average wind speed (x_6), days of relative humidity ≥ 80% (x_{16}), days of maximum temperature ≥ 30°C (x_{21}), days of maximum temperature < 30°C (x_{23}), days of minimum temperature < 30°C (x_{27}) with adjusted coefficient of determination (R^2_{adj}) 0.342 and the standard error of the estimation (S) 0.031. The accuracy of estimation was shown by comparing the graph between the observe values and the predicted values from the MLR equation.

Acknowledgement

We are grateful to Chonburi Public Health Office, Thailand for kindly providing all data.

References

[1] Shaman, J., & Kohn, M., 2009, "Absolute humidity modulates influenza survival, transmission, and seasonality," Proceedings of the National Academy of Sciences, 106(9), 3243-3248.

[2] Barreca, A.I., & Shimshack, J.P., 2012, Absolute Humidity, Temperature, and Influenza Mortality: 30 Years of County-Level Evidence from the United States," American Journal of Epidemiology, 176(7), 114-122.

[3] Baumgartner, E.A., Dao, C.N., Nasreen, S., Bhuiyan, M.U., Muneer, S.M., Mamun, A.A., Sharker, M.A., Zaman, R.U., Cheng, P.Y., Klimov, A.I., Widdowson, M.A., Uyeki, T.M., Ludy, T.M., Mounts, A., & Bresee, J., 2012, "Seasonality, Timing, and Climate Drivers of Influenza Activity Worldwide," The Journal of Infectious Diseases, 206, 838-846.

[4] Fang, L.Q., Wang, L.P., Vlas, S.J., Liang, S., Tong, S.L., Yan, L.L., Ya, P.L., Qian, Q., Yang, H., Zhou, M.G., Wang, X.F., Richardus, J.H., Ma, L.Q., & Cao, W.C., 2012, "Distribution and Risk Factors of 2009 Pandemic Influenza A (H1N1) in Mainland China," American Journal of Epidemiology, 175(9), 890-897.

[5] Yaari, R., Katriel, G., Huppert, A., Axelsen, J.B., & Stone, L., 2013, "Modelling seasonal influenza: the role of weather and punctuated antigenic drift," Journal of the Royal Society Interface, 10, 1-12.

[6] Hocking, R.R., & Leslie, R.N., 1967, "Selection of the best subset in regression analysis," Technometrics, 9(4), 531-540.

[7] Lewis, P.A.W., 1961, "Distribution of the Anderson-Darling Statistic," The Annals of Mathematical Statistics, 32(4), 1118-1124.

[8] Durbin, J., & Watson, G.S., 1951, "Testing for Serial Correlation in Least Squares Regression II," Biometrika, 38(2), 159-177.

[9] Breusch T.S., & Pagan, A.R., "A Simple Test for heteroscedasticity and Random Coefficient Variation," Econometrica, 47(5), 1287-1294.

[10] Sakia, R.M., 1992, "The Box-Cox transformation technique: a review," The statistician, 169-178.

[11] University Social Resposibility, 2015, "Malaria," Retrieved February 4, 2015, from Mahidol University Web site: http://www.sc.mahidol.ac.th/usr/?p=51