Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 4 (2015), pp. 2123-2126

http://www.ripublication.com

Multiple Linear Regression for Estimation of Monthly Average Benzene in the Air at Rayong Province

J. Mekparyup[1], K. Saithanu[2*] and P. Thongprasert [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]prow_nooch@hotmail.com

Abstract

The purpose of this study is to estimate monthly average benzene in the air by multiple regression analysis. For building linear regression equation, the data using for analysis was Benzene in the air, Nitrogen monoxide (NO), Carbon monoxide (CO), Ozone (O_3), Nitrogen dioxide (NO_2), Sulfur dioxide (SO_2), Oxides of nitrogen (NO_X), Particulate matter smaller than 10 microns (PM_{10}) and Hydrocarbon. The results of this study found that regression equation for estimation monthly average benzene in the air is $\widehat{BENZENE}=0.742-0.546CO$ with standard error of estimation 0.1478 and adjusted coefficient of determination 0.317.

Mathematics Subject Classification: 62J05

Keywords: multicollinearity, variance inflation factor, best subset method

INTRODUCTION

Thailand's economy is developing rapidly. The growth of both manufacture and agroindustry is expanding enormously. As of such a development, it leads to many problem of air pollution especially volatile organic compounds (VOCs). United States Environmental Protection Agency (EPA) has given priority to organize the approximate 114 types of VOCs such as benzene, toluene and dichloroethane, etc. International Agency for Research on Cancer (IARC) and World Health Organization (WHO) then seriously defined some VOCs which cause to cancer.

The problem of air pollution in Rayong province, Thailand, currently holds the urgent priorities that need to be resolved particularly the problem of VOCs into the air released from the petrochemical industry. As this mentioned, the present study intended to fit the model for estimation the monthly average benzene in the air. The regression equations derived from the study is to estimate the amount of benzene in the air to prevent benzene content not exceed the standard amount.

MATERIALS AND METHODS

The monthly average data [1][2][3][4], benzene in the air, carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), nitrogen monoxide (NO), sulfur dioxide (SO_2), oxides of nitrogen (NO_X), particulate matter smaller than 10 microns (PM_{10}) and hydrocarbon (TOTALHC), was collected from Air Quality and Noise Management Bureau, Pollution Control Department, Thailand. For building multiple linear regression (MLR) equation to estimate monthly average benzene in the air, MLR model is used by Equation 1

$$BENZENE = \beta_0 + \beta_1 CO + \beta_2 NO + \beta_3 O_3 + \beta_4 NO_2 + \beta_5 SO_2 + \beta_6 NO_X$$

$$+\beta_7 PM_{10} + \beta_8 TOTALHC + \varepsilon \qquad (1)$$

Simple correlation coefficients (R) were calculated to consider relationship among these pollutants. All possible MLR equations were determined by best subset method considering Mallows' C_p [5], adjusted coefficients of determination (R^2_{adj}) and standard error of estimation (S). After obtained the best fitted MLR equation, then checking assumptions for MLR analysis was proceeded. There are four assumptions to be tested;

(i) normality of the error distribution [6],
(ii) independence of the errors [7],
(iii) homoscedasticity (constant variance) of the errors [8],
(iv) multicollinearity among independent variables.

RESULTS

The relationship between one dependent variable and eight independent variables using R showed that most values were positive and highly significant (p-value<0.01) and the highest R was determined between BENZENE and CO (R=0.687, p-value=0.000). The results was in the same previous studies [1][2][3][4]. All possible MLR equations were employed in the selection of a best fitted MLR equation and the analysis displayed the MLR equation for estimation BENZENE consisted two independent variables, CO and SO_2, with Mallows' C_p=0.7, R^2_{adj}=49.7 and S=1.0343 and the MRL equation was $BENZENE = -1.12 + 5.31CO + 0.333SO_2$. Then testing the assumptions of MLR analysis was determined. However, the study result was not

satisfied. The Box-Cox method [9] was then used to transform data. The MLR equation was rebuilt and the equation was $BENZENE' = 0.742 - 0.546CO$ with $R^2_{adj} = 31.7$ and S=1.478 where $BENZENE' = 1/BENZENE$. The MLR assumptions were retested as following;

(i) **The test of normality:** The method to test normality of the error distribution used Anderson-Darling statistic values [6], AD=0.559 (p-value=0.752) so the distribution of error was normal.

(ii) **The test of independence:** Durbin-Watson statistic was calculated to test the independence [7]. The results illustrated that DW statistic values 1.454 (d_L=1.341) so the errors were independent.

(iii) **The test of homoscedasticity:** Breush-Pagan statistic was determined to test homoscedasticity [8]. The BP values (χ^2_{BP} =0.316) was less than critical values (3.842) so the variance of error was constant**.**

(iv) **Test of multicollinearity:** Variance Inflation Factor (VIF) was used to detect multicollinearity represented that all VIF values of both independent variables (VIF=4.1) were less than 5 then there was no relationship among independent variables [10]. After checking all assumptions of MLR analysis, the monthly average benzene in the air was estimated by using the MLR equation then comparison between the real and the estimated values was illustrated by graph shown as Figure 1.

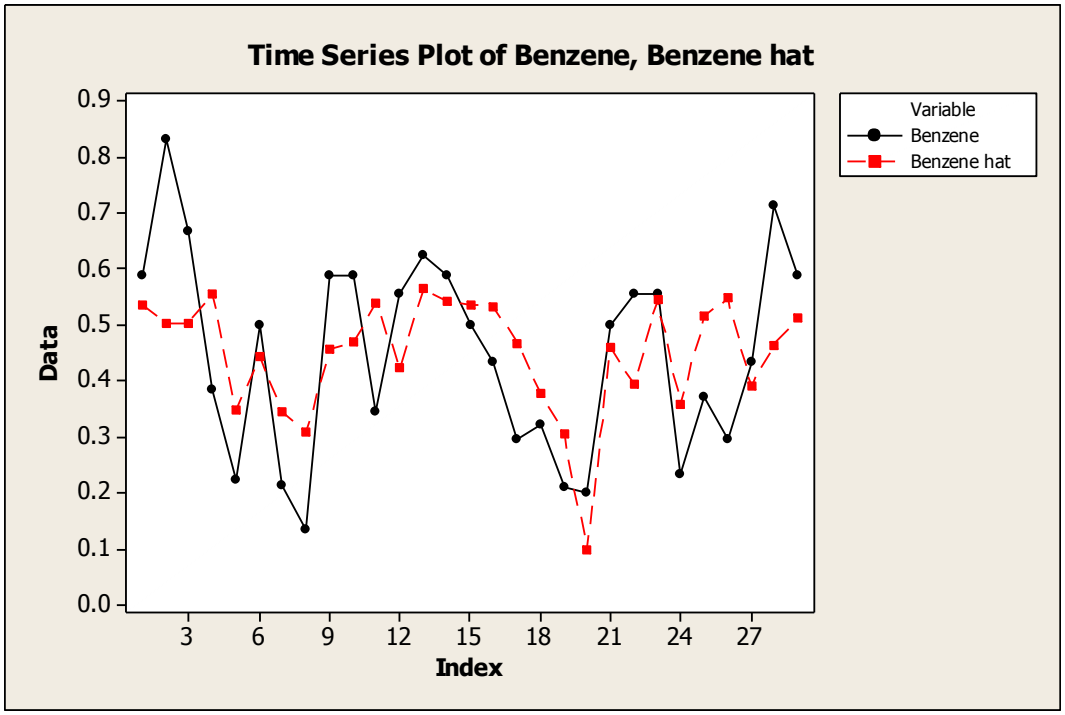

Figure 1. Monthly average benzene in the air plot between the real and the estimated values

DISCUSSION

The regression coefficient of carbon monoxide has shown that such an estimator could be used to estimate the monthly average benzene in the air. The comparison result of present study using time series plot as of Figure 1 showed that estimation of monthly average benzene in the air using carbon monoxide would be useful to estimate monthly average benzene in the air.

ACKNOWLEDGEMENT

The authors wish to thank to Air Quality and Noise Management Bureau, Pollution Control Department, Thailand for kind support in data collecting.

REFERENCES

[1] Kajihara, H., Ishizuka, S., Fushimi, A., & Masuda, A., 1999, "Exposure assessment of benzene from vehicles in Japan, " Environmental Science and Technology, 240, 62-70.

[2] Leong, S. T., Muttamara, S., & Laortanakul, P., 2001, "Influence of benzene emission from motorcycles on Bangkok air quality, " Atmospheric Environment, 36, 651-661.

[3] Oftedal, B., Nafstad, P., Magnus, P., Bjørkly, S., & Skrondal, A., 2002, "Traffic related air pollution and acute hospital admission for respiratory diseases in Drammen, Norway 1995–2000, " European Journal of Epidemiology, 18, 671-675.

[4] Moussiopoulos, N., Vlachokostas, Ch., Tsilingiridis, G., Douros, I., Hourdakis E., Naneris, C., & Sidiropoulos, C., 2008, "Air quality status in Greater Thessaloniki Area and the emission reductions needed for attaining the EU air quality legislation, " Science of the total environment, 407, 1268-1285.

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

[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 heteroscedasticity and Random Coefficient Variation, " Econometrica, 47(5), 1287-1294.

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

[10] Kutner, M.H., Christopher, J.N., Neter, J., 1996, "Applied liner regression models. 4th ed, " McGraw-Hill/Irwin; (USA).