International Journal of Innovation Scientific Research and Review
Vol. 05, Issue, 07, pp.4908-4911, July 2023
Available online at http://www.journalijisr.com
SJIF Impact Factor 2023: 6.599

IJISRR

ISSN: 2582-6131

Research Article

FORECASTING ELECTRICITY CONSUMPTION USING REGRESSION ANALYSIS

*** Kidakan Saithanu and Jatupat Mekparyup**

Burapha University, Department of Mathematics, Faculty of Science,169 Muang, Chonburi, Thailand.

Received 24th May 2023; Accepted 25th June 2023; Published online 30th July 2023

ABSTRACT

This research focused on forecasting monthly average electricity consumption in Rayong province, Thailand, using regression analysis method with 3 factors, monthly average number of electricity customers, monthly average rainfall and monthly average temperature. Data of 80 samples was divided into two groups; the first group was to build a regression model and the other was to validate the performance of forecasting electricity consumption. After fitting the model, assumptions of regression analysis were detected by Anderson-Darling statistic, Durbin-Watson statistic, Breusch-Pagan statistic and Variance inflation factor. Finally, the performance of forecasting electricity consumption values was calculated by mean absolute percentage error with 0.0416, mean absolute error with 407,061.084 and root mean squared error with 596,706.8834. The results illustrated that the regression standard error was 231,204 with the adjusted coefficients of determination of 0.931.

Keywords: electricity consumption, regression analysis.

INTRODUCTION

Rayong is a province in the eastern region of Thailand and one of the famous attractions in Thailand. Rayoung has 8 districts; Muaeng, Klaeng, Ban Chang, Ban Khai, Khao Chamao, NikhomPhatthana, Pluak Daeng and Wang Chan and 10 industrial estates; Map Ta Phut industrial estate, WHA Eastern industrial estate, PhaDaeng industrial estate, Eastern Seaboard industrial estate, Amata City industrial estate, Hemaraj Eastern Seaboard industrial estate, Asia industrial estate, RIL industrial estate, Rayong industrial estate and LakchaiMuaeng Yang industrial estate. Regarding to the report of Provincial Electricity Authority Rayong Province since 2013 to 2019, the highest amount of electricity is in Muaeng district while the least amount of electricity is in Khao Chamaodistrict as seeing in Figure 1. It's obviously found the electricity consumption tended to increase steadily every year.Therefore, studying the forecasting electricity consumption is absolutely necessary to plan for the production and use of electricity in Rayong province. Many research articles focused on studying forecasting electricity consumption(Mohamed and Bodger, 2005; Bianco *et al.,* 2009; Kaytez *et al.,* 2015; Amber *et al.,* 2018; Fan *et al.,* 2020). For this research, multiple linear regression (MLR) is used tobuildan MLR model for forecasting electricity consumption in Rayong province.

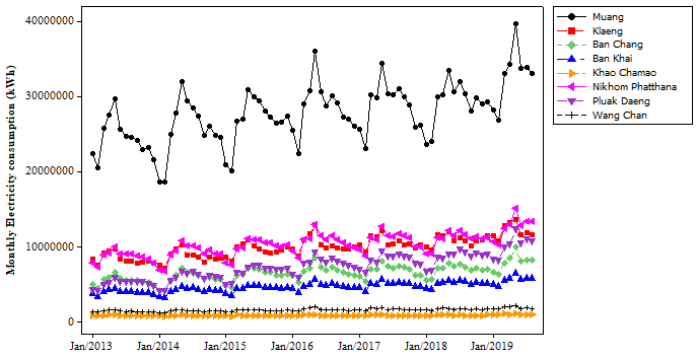


Figure 1: Monthly electricity consumption of all districts in Rayong province since January 2013 – August 2019

MATERIALS AND METHODS

The data, the monthly average electricity consumption, the monthly average number of electricity customers, the monthly average rainfall and the monthly average temperature, was collected from District 2 Central Electricity Authority in Chonburisince January 2013 to August 2019 showed in Table 1.

Table 1: **Data set**

Data set	Size	%
Training (January 2013 – December 2017)	60	75
Validation (January 2018 – August 2019)	20	25
Total	80	100

Monitoring linear relationship

Simple correlation coefficients (R) were firstly calculated to identify relationship among these data, the monthly average electricity consumption, the monthly average number of electricity customers, the monthly average rainfall and the monthly average temperature.

Building the multiple linear regression equation

The multiple linear regression equation to estimate the monthly average electricity consumption in Rayong province was generated by regression model as Equation (1).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \qquad (1)$$

The model is composed of one response variable; y = the monthly average electricity consumption (kWh), and 3 predictor variables; x_1 = the monthly average number of electricity customers, x_2 = the monthly average rainfall (millimeters),x_3 is the monthly average temperature (Celsius) andx_3 = the monthly average temperature (Celsius), where β_i=the regression coefficient (i = 0, 1, 2, 3) and ε = the error of regression model.

Checking assumptions for multiple linear regression analysis

After obtained the best fitted multiple linear regression equation, the assumptions checking for multiple regression analysis was proceeded. There are four assumptions to be tested; (I) normality of the error distribution using Anderson-Darling statistic by Equation (2) (Anderson and Darling, 1952); (II) independence of the errors using Durbin-Watson statistic by Equation (3) (Durbin and Watson, 1950); (III) homoscedasticity (constant variance) of the errors using Breusch-Pagan statistic by Equation (4) (Breusch and Pagan, 1979); (IV) multicollinearity among predictor variables using Variance Inflation Factor (VIF) Equation (5). After tested all assumptions, the comparison between the real values and the estimated values of the electricity consumption from the obtained multiple regression equation was plotted.

$$\mathrm{AD} = -n - \sum_{i=1}^{n}\left(\frac{2i-1}{n}\right)\{\ln F(y_i) + \ln[1 - F(y_{n+1-i})]\} \quad (2)$$

$$\mathrm{DW} = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \quad (3)$$

$$\mathrm{BP} = \frac{SSR^*/2}{(SSE/n)^2} \sim \chi^2_{(k)} \quad (4)$$

where SSR^* = the sum of squares in regression between e_j^2 = j^{th} residual and x_{ij}, SSE = the sum of squares in regression error between y_j and x_{ij}.

$$\mathrm{VIF}_j = C_{jj} = \frac{1}{1 - R^2_{j|\text{others}}};\ j = 1,\ 2,\ \ldots,\ k \quad (5)$$

where $R^2_{j|\text{others}}$ = the multiple coefficient of determination between x_{ij} and all x_i.

Comparison between the real values and the forecasted values from the multiple linear regression equation

The comparison between the real values and the estimated values calculated from January 2018 – August 2019 by the obtained multiple regression equation was plotted when checking all assumptions of multiple regression analysis was determined.

Validation of the multiple linear regression equation

The performance of forecasting accuracy was calculated by 3 criteria; mean absolute percentage error (MAPE), mean absolute error (MAE) and root mean squared error (RMSE) as of Equation (6), (7) and (8), respectively.

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i} \quad (6)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \quad (7)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (8)$$

RESULTS

Descriptive statistics was calculated by average (MEAN), standard deviation (SD), minimum value (MIN) and maximum value (MAX) split by types of variable and data set presented in Table 2.

Table 2: **Data summary**

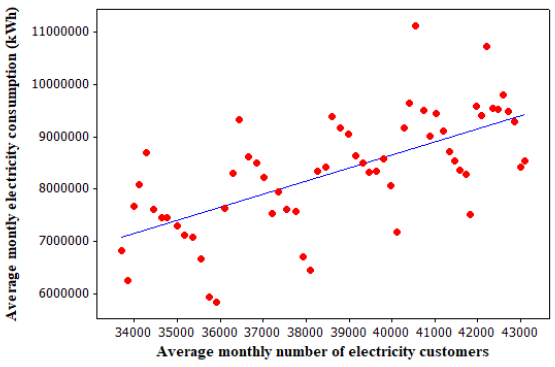
VARIABLE	DATA SET	MEAN	SD	MIN	MAX
y	All	8,705,681	1,293,648	5,839,223	12,620,105
	Training	8,320,538	1,108,015	5,839,223	11,122,843
	Validation	9,861,109	1,127,971	7,776,912	12,620,105
x_1	All	40,170	3,614	33,693	45,999
	Training	38,683	2,874	33,693	43,112
	Validation	44,633	838	43,312	45,999
x_2	All	10.716	6.008	0.000	24.631
	Training	10.957	5.938	0.000	24.631
	Validation	9.990	6.310	1.240	24.400
x_3	All	28.289	1.328	24.176	30.794
	Training	28.213	1.395	24.176	30.794
	Validation	28.514	1.105	26.574	30.683

The values of Pearson correlation coefficients (*R*) for the 4 variables, the monthly average electricity consumption (*y*), the monthly average number of electricity customers (x_1), the monthly average rainfall (x_2) and the monthly average temperature (x_3) were tabulated in Table 3 and the scatter plots were also displayed in Figure 2.

Table 3: **Pearson correlation coefficients among four variables**

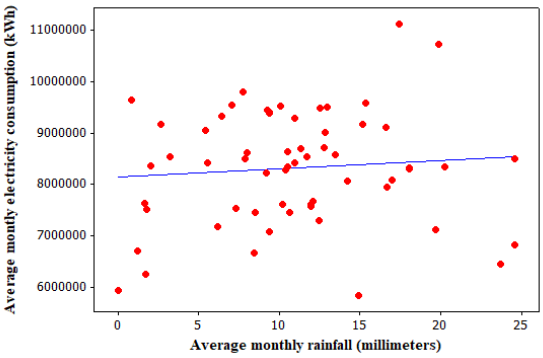
Variables	Variables		
	y	x_1	x_2
x_1	0.650 (0.000)		
x_2	0.085 (0.517)	-0.080 (0.545)	
x_3	0.696 (0.000)	0.073 (0.579)	0.076 (0.564)

p-values in bracket

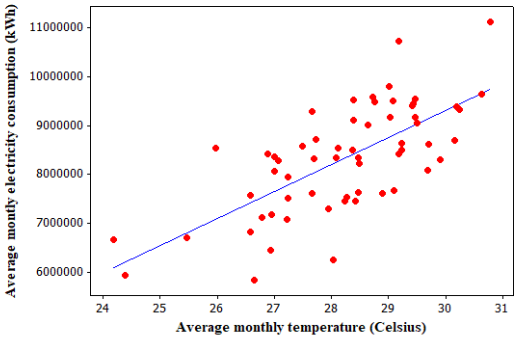


(a)

***Corresponding Author: Kidakan Saithanu,**
Burapha University, Department of Mathematics, Faculty of Science,169 Muang, Chonburi, Thailand.



(b)



(c)

Figure 2: Scatter plot between the monthly average electricity consumption (y) and (a) the monthly average number of electricity customers (x_1), (b) the monthly average rainfall (x_2), (c) the monthly average temperature (x_3)

Therefore, the multiple linear regression equation was generated by multiple linear regression analysis with stepwise method. The estimators of regression coefficients $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ and the equation of forecasting monthly average electricity consumption in Rayong province was exhibited as of Equation (9).

$$\hat{y} = -13,206,280 + 218.68x_1 + 464,424x_3 \qquad (9)$$

The multiple linear regression model was tested by analysis of variance (ANOVA), illustrated in Table 4.

Table 4: ANOVA

Source of variation	Degree of freedom	Sum of square	Mean of square	*F*	p-value
Regression	2	3.7085×10^{13}	1.8542×10^{13}	346.88	0.000
Residual	49	2.6193×10^{12}	53,455,169,770		
Total	51	3.9704×10^{13}			

It was showed that the model was appropriated with *F*-statistic values 346.88 and was found that there was significant. The standard error of regression (S) and the adjusted coefficients of determination (r^2_{adj}) of this model also displayed as 231,204 and 0.931, respectively. Then the regression coefficients of the model were tested consequently as in Table 5. It was displayed that all coefficient regressions were significant.

Table 5: Regression coefficients

Predictor	Coefficient values	*t*-value	*p*-value
$\hat{\beta}_0$	-13,206,280	-15.55	0.000
$\hat{\beta}_1$	218.68	19.34	0.000
$\hat{\beta}_3$	464,424	18.37	0.000

p-values in brackets

Table 6: **Assumption testing**

Assumptions	Test statistic	Critical value	p-value
Normality	AD = 0.817		0.223
Auto-correlation	DW = 2.5537	D_U = 1.6334	
Homoscedasticity	BP = 0.965	5.991	0.383
Multicollinearity	VIF (x_1) = 1.00		
	VIF (x_3) = 1.00		

After fitting the multiple linear regression equation to forecast monthly average electricity consumption in Rayong province, (I) the test of normality was determined and found that hypothesis testing of normality was significant with AD=0.817and p-value =0.223. (II) The test of independence: Durbin-Watson statistic was calculated by Equation (3) and the results showed that the test statistic values (DW= 2.5537) was more than upper critical values (D_U = 1.6334) so the errors were independent. (III) The test of homoscedasticity: Breush-Pagan statistic was determined by Equation (4) and the results illustrated that the test statistic values (BP= 0.965) was less than the critical values (5.991) with p-value =0.383 so the variance of error was constant. (IV) Test of multicollinearity: the VIF values were calculated by Equation (5) and all VIF values were less than 5 then there was no relationship among predictor variables in multiple linear regression mode l(Kutner *et al.,* 1996).

Finally, the accuracy of prediction was numerically presented by performance of forecasting monthly average electricity consumption was subsequently found with MAPE = 0.0416, MAE = 407,061.084and RMSE = 596,706.8834. Moreover, time series were plotted to compare between the real and forecasted data displayed in Figure 3.

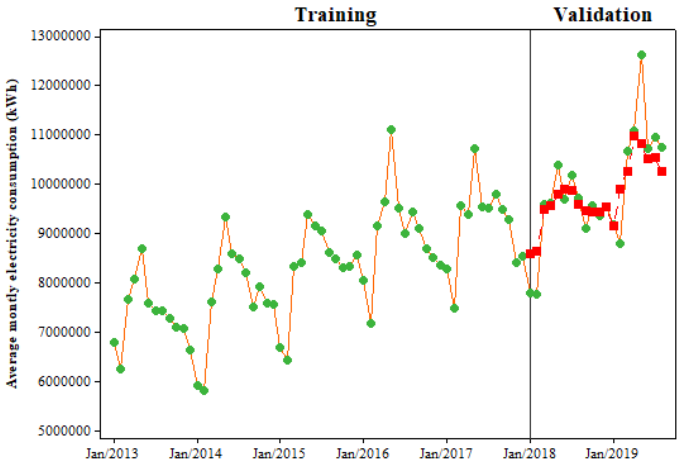


Figure 3: Time series plot between real and predicted data of monthly average electricity consumption in Rayong province

CONCLUSION AND DISCUSSION

Multiple linear regression model was generated to forecast monthly average electricity consumption in Rayong province found in Equation (9) with r^2_{adj} =0.931and $S = 231{,}204$. The forecasting efficiency was consequently calculated by MAPE = 0.0416, MAE = 407,061.084 and RMSE = 596,706.8834.

ACKNOWLEDGEMENTS

Specially, the authors were appreciated to provide the secondary data from District 2 Central Electricity Authority in Chonburi.

REFERENCES

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. The annals of mathematical statistics, 193-212.

Amber, K. P., Ahmad, R., Aslam, M. W., Kousar, A., Usman, M., & Khan, M. S. (2018). Intelligent techniques for forecasting electricity consumption of buildings. Energy, 157, 886-893.

Bianco, V., Manca, O., & Nardini, S. (2009). Electricity consumption forecasting in Italy using linear regression models. Energy, 34(9), 1413-1421.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica: Journal of the econometric society, 1287-1294. Durbin, Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. Biometrika, 37(3/4), 409-428.

Energy Regulatory Commission. (2016). Overview of electricity consumption by economic sector. Available: https://www.erc.or.th/ERCWeb2/default.aspx

Fan, G. F., Wei, X., Li, Y. T., & Hong, W. C. (2020). Forecasting electricity consumption using a novel hybrid model. Sustainable Cities and Society, 61, 102320.

Hussain, A., Rahman, M., & Memon, J. A. (2016). Forecasting electricity consumption in Pakistan: The way forward. Energy policy, 90, 73-80.

Kaytez, F., Taplamacioglu, M. C., Cam, E., & Hardalac, F. (2015). Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. International Journal of Electrical Power & Energy Systems, 67, 431-438.

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

Mohamed, Z., & Bodger, P. (2005). Forecasting electricity consumption in New Zealand using economic and demographic variables. Energy, 30(10), 1833-1843.

****