

Correlation Analysis and Multilinear Regression Model for Prediction on Solid Waste Generation in Malaysia

Faridah Zulkipli^a, Zulkifli Mohd Nopiah^{b*}, Noor Ezlin Ahmad Basri^c & Cheng Jack Kie^d

^aFaculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia

^bDepartment of Engineering Education, Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

^cDepartment of Civil Engineering, Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

^dFaculty of Industrial Management, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 23600 Gambang, Kuantan, Pahang, Malaysia

*Corresponding author: zmn@ukm.edu.my

Received 19 May 2021, Received in revised form 22 May 2021
Accepted 27 May 2021, Available online 30 August 2021

ABSTRACT

Increased volumes of solid waste generation continue unabated due mainly to rapid population increase, urban migration, economic enhancements, and modern lifestyles. Two significant factors contributing directly to the higher volumes of solid waste generation are population and gross domestic product (GDP). In Phase 1 of this study, a correlation analysis with a Pearson coefficient of more than 93% ($r > 0.93$) shows a strong positive linear relationship between the amount of solid waste generation and population and GDP. Phase 2 provides a multilinear regression analysis and the development of a regression model. They indicate that a 1-unit increment in the amount of solid waste generation is caused by 0.06 of the population and 0.119 of GDP while β_0 remains constant at 124.449. Therefore, the multilinear regression model in this study can be applied by solid waste management authorities in Malaysia to accurately forecast future solid waste generation volumes. However, further investigation on other significance factors are suggested for future work in order to develop a holistic model for solid waste management in Malaysia.

Keywords: Solid waste generation; correlation analysis; multilinear regression model; population; gross domestic product

INTRODUCTION

Solid waste generation and the protection of the environment is of much interest to governments, policy makers, and the public. The generation of solid waste is directly related to rapid population growth, technological advancement, changing lifestyles, multiracial citizens, and economic development (Zulkipli et al. 2020; Abd Manaf et al. 2009; Tchobanoglous & Kreith. 2002). The tremendous increase in the overall rate of solid waste generation has serious implications to the environment. In developed countries, various policies have been implemented aimed at achieving environment sustainability. This has spread to developing countries such as Malaysia and Indonesia as they strive to strengthen enforcement towards attaining a sustainable

environment for the present and future generations. In addition, the most important variables for solid waste generation in Vietnam were the urban population, average monthly consumption expenditure, and overall retail sales (Nguyen et al. 2021).

Over the last decade, Malaysia has experienced rapid economic growth and urban transformation. Apart from water and air pollution, municipal solid waste (MSW) is one of Malaysia's three major environmental issues (Rodzi 2019). The 2-3% annual population growth in Malaysia increased the amount of solid waste generation to more than 33 000 tons in 2012, 38 200 tons in 2016, and the figure is forecast to reach about 45 900 tons by 2020 (Solid Waste Association of North America 2017). Malaysia is a multiracial country with Malays forming the majority of

the population followed by the Chinese, Indians, and others. Due to this, celebrations and festivals are held in Malaysia almost every month and this also contributes to an increase in the annual generation of solid waste.

There is a similarity in terms of how each country handle the solid waste including Malaysia. General processes involved in solid waste management begins with solid waste generation up until waste disposal in landfills. Figure 1 represent the process flow of solid waste management.

As shown in Table 1, the solid waste management authorities have set a goal to make Malaysia a clean country by 2020 (Agumuthu et al. 2009). Malaysia Vision 2020 indicates that the planned solid waste disposal scenarios

of the total solid waste generated (Younes et al 2015). The amount of waste dumped in landfills will decrease to more than 50% compared to the current practice where all waste are collected and sent to landfills.

Figure 2 illustrates the various states that fall under the supervision of the Solid Waste Corporation (SWCorp) which are Perlis, Kedah, Pahang, Wilayah Persekutuan Kuala Lumpur, Putrajaya, Johor, Negeri Sembilan and Melaka. In developing countries, it is difficult to obtain reliable historical data for solid waste characteristics (Rimaityte et al. 2012; Wu et al. 2020), so modelling is particularly important (Beigl et al. 2008; Abbasi & El Hanandeh 2016; Kannangara et al. 2018) to provide a theoretical foundation for the future scenario.



FIGURE 1. Solid Waste Management Process

TABLE 1. Percentage of Solid Waste Disposed

Treatment	Percentage of solid waste disposed		
	2002	2004	2020*
Recycling	5.0	5.5	22.0
Composting	0.0	1.0	8.0
Incineration	0.0	0.0	16.8
Inert landfill	0.0	3.2	9.1
Sanitary landfill	5.0	30.9	44.1
Others	90.0	59.4	0.0

The accuracy of data for predicting future amounts of waste generation has been highlighted as an urgent research area as it will contribute to better future outcomes. The planning and design departments of solid waste management agencies can address this issue more efficiently through systematic approach. Previous research work has produced some forecasting techniques for future waste generation such as multiple regression analysis, system dynamics simulation, ARIMA model, and others (Zulkipli et al. 2017; Younes et al. 2015; Dyson & Chang 2005). In this study, multiple regression analysis has been adopted from Zulkipli (2017) due to the similarity of scenarios and the complexity of the problem. Further explanations on the method are described in methodology section.



FIGURE 2. States Under SWCorp Supervision

the study while third section briefly describes the results analysis and discussion regarding the results. Finally, the conclusion section is discussed of the study and recommendations for future work.

METHODOLOGY

PROPOSED FRAMEWORK

In this study, secondary data was obtained from the Department of Statistics Malaysia (DOSM) and the World Bank. The data consists of population (in millions), gross domestic product (GDP US\$), and waste generation quantities (kg/capita/year) from 1981 to 2011. The software IBM SPSS version 25 is implemented in this study in order to perform correlation and multilinear regression

The aim of this study is to predict waste generation from the year 2012 to 2030 using multilinear regression model. In order to achieve this aim, several objectives need to be investigated, which are:

1. To examine the relationship between solid waste generation towards population and GDP.
2. To develop a regression model for solid waste generation based on population and GDP.
3. To predict the quantity of solid waste generation from 2012 until 2030 using the regression model in (2).

The organization of this paper begins with the introduction to the study and highlights of issues in the first section. Second section presents the methodology of



FIGURE 3. Dependent Variable & Independent Variable

analyses. Figure 3 depicts the dependent and independent variables employed in this study.

ANALYSIS PROCEDURE

A correlation analysis is a statistical method used to assess the power of the connection between variables. The hypothesis set and tested in this study is as follows:

- H_0 : There is no relationship between variables
 H_1 : There is a relationship between variables

In addition, multilinear regression analysis is a statistical method that can be used to achieve the equation for dependent and independent variables. It also enables more than one independent variable to be investigated with the dependent variable [8]. The independent variables

measured in this study are population and GDP while the dependent variable is solid waste generation. The hypothesis set and tested in this study is as follows:

- H₀: Population and gross domestic product are not the contributing factors towards the amount of solid waste generation.
- H₁: Population and gross domestic product are the contributing factors towards the amount of solid waste generation.

H0 will be rejected if the p-value ≤ significance value (α) = 0.05. It indicates that the variable(s) significantly contributes towards solid waste generation. The multilinear regression model is as in (1).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots + \beta_n X_n + \varepsilon \quad (1)$$

where Y is the dependent variable, X₁, X₂, ..., X_n are the independent variables, ε is the residual term while the β's are the regression coefficients with β₀ the constant term.

There are four assumptions of the regression analysis that must be met in order to make the analysis reliable and valid. The assumptions are: (1) the values of the residuals are normally distributed; (2) the values of the residuals are independent; (3) multicollinearity does not exist; and (4) no outlier exists in the dependent variable.

RESULTS AND DISCUSSION

CORRELATION ANALYSIS

Correlation analysis was conducted for examine the strength of relationship between two variables. In this study, the authors are interested to examine whether there is a relationship between the variables. Based on the Pearson correlation coefficient values in Table 2, both independent variables (population and GDP) are highly correlated with solid waste generation. Therefore, population and GDP can be used in order to predict waste generation by applying multiple linear regression.

Variable	Solid Waste Generation
Population	0.988
Gross Domestic Product	0.941

ASSUMPTIONS OF MULTILINEAR REGRESSION

There are four assumptions of multilinear regression must be fulfilled. First, the normality test for the residuals. The normal P-P plot for the residuals of the model was designed

to test for normality as shown in Figure 4. The closer the points are to the diagonal line, the nearer the residuals are to standard. In this case, the assumption of normality for residues is not violated because all the points are closer to the diagonal line.

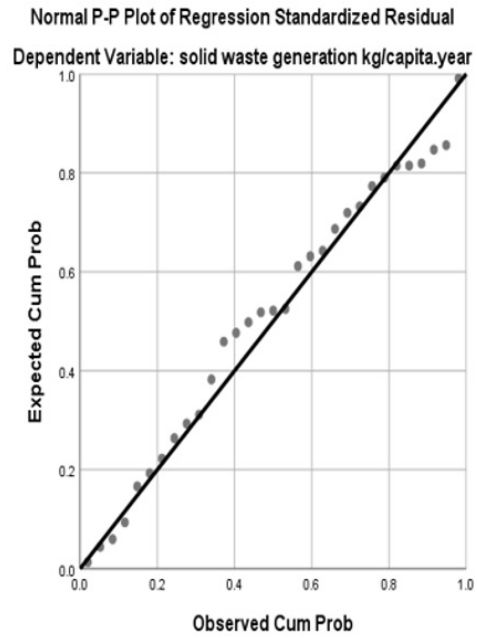


FIGURE 4. The normal P-P plot for the residuals of the model

Second, the independent for the residuals using Durbin-Watson. Durbin-Watson statistic was used to test the assumption that the residuals were independent or not correlated. As a result, the Durbin-Watson statistic came up as 1.247 in this study, indicating that the residuals were independent, since the value was between 1.0 and 3.0. Figure 5 shows the variance of residual scatter plot where all the data positions is between +3 and -3.

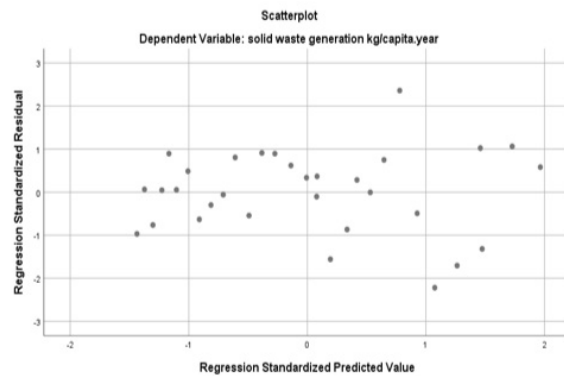


FIGURE 5. Variance of Residual Scatterplot

Third, the multicollinearity test. The collinearity statistics shown in Table 3 are used to test the existence of multicollinearity. Analysis of statistics on collinearity show that this hypothesis was met as VIF scores were below 10

and tolerance scores above 0.1. Thus, no multicollinearity exists in the data.

TABLE 3. Collinearity Statistics

Variables	Collinearity statistics	
	Tolerance	VIF
Population	0.176	5.668
Gross Domestic Product	0.176	5.668

Fourth, the existence of the outliers. For each individual, the Cook's Distance is calculated and is the difference between the predicted regression values with and without an individual observation. The distance of a large Cook shows an influential observation. Based on Table 4, Cook's distance values were all below 1, indicating that individual instances did not influence the model incorrectly. This indicates no significance outliers exist which may influence the model.

TABLE 4. Residuals Statistics

Cook's Distance	Statistics
Minimum	0.001
Maximum	0.216
Mean	0.041
Std. Deviation	0.058
N	30

MULTILINEAR REGRESSION MODEL

The coefficient of determination (r^2) = 0.986 shows that 98.6% of the total variation in solid waste generation can be explain by population and GDP. Therefore, both

independent variables are significantly used in order to describe the amount of solid waste generation in this study.

TABLE 5. Model Summary

Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std.Error of the estimated
1	.933a	.987	.986	4.131492938307917

a. Predictors: (Constant), GDP in Billion (US\$ 2017), pop (in 100)

b. Dependent Variable: Solid waste generation kg/capita. year

The multilinear regression is accurate with a p-value of 0.001 which is less than 0.05. we can conclude that the multilinear regression model is accurate in order to be applied for predicting the amount of solid waste generation. The F-Statistics obtained was 1046.996 with a p-value less than 0.05 indicating that the estimated regression is valid and is statistically significant at the significance level of 0.05. Based on Table 5, the regression results show that population and GDP are significant at 0.05 implying that both significantly affects solid waste generation. Hence, population and GDP are positively correlated with waste generation. The estimated coefficients are 0.06 and 0.119 respectively. Therefore, the null hypothesis is rejected and states that population and GDP have an impact on the amount of solid waste generated. The equation is as in (2).

$$\hat{Y} = 124.499 + 0.06 X_1 + 0.119 X_2 \quad (2)$$

The 0.06 associated with the amount of waste generation indicates that for each addition to the population, the waste generation level will increase 0.06, whereas other independent variables are held constant. Then for each

TABLE 6. Multilinear Regression Results

Model	B	Std.Error	Std Coef .Beta	t	p-value
β_0	124.499	6.342		19.632	0.000
X_1	0.06	0.000	0.761	14.716	0.000
X_2	0.119	0.025	0.250	4.846	0.000

addition to GDP, the solid waste generation level will increase 0.119, whereas the other independent variables are held constant. In terms of the best predictor variable for the amount of solid waste generation is population compared to GDP. It is because the value of standard coefficient beta for population is 0.761 which is higher than the 0.250 GDP. Therefore, population is more influence factor towards predicting the amount of solid waste generation.

Figure 6 presents the trend of waste generation based on actual data (1981-2011) and predicted data (2012-2030). The graph shows that there is an increasing trend in the generation of waste with a parallel increase in population and gross domestic product. The trend is expected to continue to rise annually.

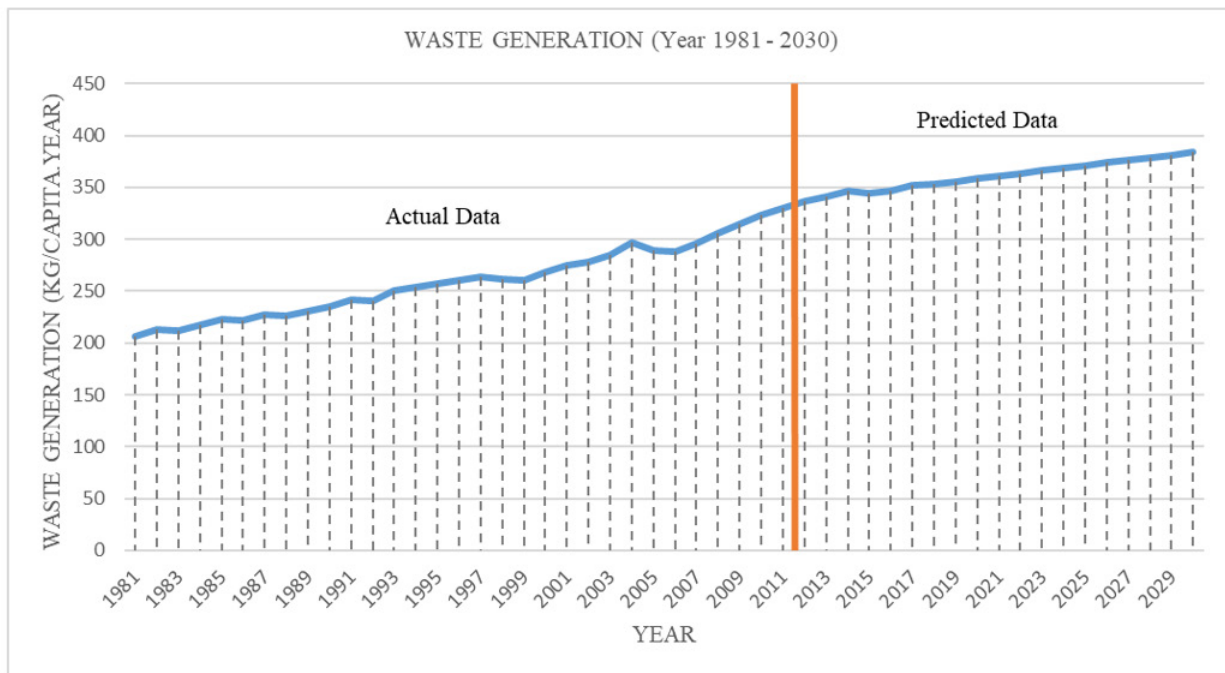


FIGURE 6. Solid Waste Generation: Actual vs Prediction

CONCLUSION

This study sought to investigate issues related to solid waste generation in Malaysia. Two main analyses were conducted in the investigation process. First, correlation analysis was adopted in this study in order to evaluate the relationship between the amount of solid waste generation, population and GDP. The two independent variables were highly significant towards the dependent variable with the coefficient of Pearson score at more than 0.93 and the coefficient of determination also more than 0.88. Second, multilinear regression analysis was implemented which concluded that population and GDP significantly contributed to the amount of solid waste generated. Finally, future research should investigate other factors that may contribute directly to increased volumes of waste generation.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Department of Engineering Education and Centre of Engineering Education Research, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia for supporting this study.

DECLARATION OF COMPETING INTEREST

None

REFERENCES

- Abbasi, M. & El Hanandeh, A. 2016. Forecasting municipal solid waste generation using artificial intelligence modelling approaches. *Waste Management* 56: 13-22.
- Abd Manaf, L., Samah, M.A.A. & Zukki, N.I.M. 2009. Municipal solid waste management in Malaysia: Practices and challenges. *Waste management* 29(11): 2902-2906.
- Agumuthu, P., Hamid, F.S. & Khidzir, K. 2009. Evolution of solid waste management in Malaysia: impacts and implications of the solid waste bill, 2007. *Journal of Material Cycles and Waste Management* 11(2): 96-103.
- Beigl, P., Lebersorger, S. & Salhofer, S. 2008. Modelling municipal solid waste generation: A review. *Waste Management* 28(1): 200-214.
- Department of Statistics Malaysia, Available: https://www.dosm.gov.my/v1/index.php?r=column/ctwoByCat&parent_id=115&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09
- Dyson, B. & Chang, N.B. 2005. Forecasting municipal solid waste generation in a fast-growing urban region with system dynamics modeling. *Waste Management* 25(7): 669-679.
- Global Recycling, The Magazine for Business Opportunities and International Markets, Global Report, Available: <https://global-recycling.infoarchives/1451>
- Kannangara, M., Dua, R., Ahmadi, L. & Bensebaa, F. 2018. Modeling and prediction of regional municipal solid waste generation and diversion in Canada using

- machine learning approaches. *Waste Management* 74: 3-15.
- Nguyen, X.C., Nguyen, T.T.H., La, D.D., Kumar, G., Rene, E.R., Nguyen, D.D. & Nguyen, V.K. 2021. Development of machine learning-based models to forecast solid waste generation in residential areas: A case study from Vietnam. *Resources, Conservation and Recycling* 167: 105381.
- Rimaitytė, I., Ruzgas, T., Denafas, G., Račys, V. & Martuzevicius, D. 2012. Application and evaluation of forecasting methods for municipal solid waste generation in an eastern-European city. *Waste Management & Research* 30(1): 89-98.
- Rodzi, R.M. 2019. Analysis of solid waste generation and composition in Malaysia TVET campus. *International Journal of Integrated Engineering* 11(2).
- Solid Waste Association of North America, Available: <https://swana.org/Events/WASTECON/ConferenceProgram/2017WasteManagementinMalaysia.aspx>
- Tchobanoglous, G. & Kreith, F. 2002. Handbook of Solid Waste Management. McGraw-Hill Education.
- Wu, F., Niu, D., Dai, S. & Wu, B. 2020. New insights into regional differences of the predictions of municipal solid waste generation rates using artificial neural networks. *Waste Management* 107: 182-190.
- Younes, M.K., Basri, N.E.A., Nopiah, Z.M., Basri, H., Abushammala, M.F. & Maulud, K.N.A. 2015. Integrating approach to size and site at a sanitary landfill in Selangor state, Malaysia. *Environmental Engineering Research* 20(3): 268-276.
- Zulkipli, F., Jamian, N.H. & Zulkifli, I.Z. 2020. Forecasting Model for Organic Waste Generation at Administration Cafe in UiTM Tapah Campus.
- Zulkipli, F., Nopiah, Z.M., Basri, N.E.A., Kie, C.J. & Januri, S.S. 2017. Multilinear regression analysis on solid waste generation quantity in Malaysia towards sustainable development. *International Journal of Advanced and Applied Sciences* 4(9): 46-52.