Development of linear regression model for brick waste generation in Malaysian construction industry

S H Hassan1*, H A Aziz2, Z Zakariah3, S M Noor4, M R A Majid5 and N A Karim6
1Faculty of Civil Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang, 13500, Permatang Pauh, Pulau Pinang, Malaysia.
2Faculty of Civil Engineering, Universiti Sains Malaysia, 14300 Nibong Tebal, Pulau Pinang, Malaysia.

* sitihaizaz741@uitm.edu.my

Abstract. Construction waste generation is one of the challenging tasks that related to the materials used at site. Brick waste (BW) is also part of the waste generated in construction sites. This paper determines the brick waste generated in construction stages and conducts linear regression analysis for the amount of brick waste generated. The brick waste management practices are also revealed in this study. The method used in this study was Linear Regression Model. The regression model established based on the sample data reported an R2 value of 0.78; therefore, the model can predict approximately 78% of the factor (area) involved in brick waste generation. Specifically, the prediction model is focusing on the relationship between the area of work and the amount of brick waste generated. Moreover, the linear regressions can be applied as tools to predict the brick waste generated at construction sites and as a tool for the contractor to track the brick waste sources in the future.

1. Introduction

Construction waste is an economic burden in the construction industry. Brick used in construction sites is one of the waste generators in this industry. Quantifying the brick waste (BW) generated at sites is the most crucial step in recording and predicting the waste data [1]. Previous study stated the unreliable quantification method often leads to increase wastage up to between 15 and 20 times compared to the original estimation during the construction process [2]. However, most of the data related to construction waste is based mostly on assumptions. The collection of true data is important in order to be able to establish treatment solutions for construction waste that will be based on reliable data on the quantities of construction waste [3]. Hence; an accurate quantification strategy for BW is needed to address these issues. Development of waste quantification model such as linear regression model is found to be the future strategy for accurate brick waste estimation. The linear regression predicts the amount of brick waste generated per unit/built area. Regression analysis is a statistical technique to investigate and modelling the relationship between variables [4]. The study of the regression analysis techniques will also provide certain insights on how to plan the data collected [5]. The approach can be utilised by decision makers in a pro-active manner with the objective to predict the amount of construction waste and establish a benchmark to reduce waste generation, the key and can be considered as the most effective step in waste management.
2. Method

Multiple Linear Regression (MLR) requires several assumptions to be satisfied in order to obtain the best model with efficient, unbiased and consistent estimation (Stage 2). The four main assumption for residual (error term) are:

a. Mean of residual is zero
b. Variance of residual is constant
c. Residuals follow normal distribution
d. Residuals are uncorrelated with the independent variables

Further study was reported that to establish an MLR model, a few stages are required. Figure 1 shows in Stage 1, 80% of the data for each variable was randomly selected with MATLAB R2014a software [6]. In Stage 2, the classification as depicted in Table 1 was used to interpret the obtained correlation coefficient values. The most important variables are derived with IBM SPSS Statistic 23. In stage 3, variance inflation Factor (VIF) was used to measure the effect of multicollinearity on the variance of estimated regression coefficients. When the VIF for one or more independent variables is large, then it has a multicollinearity problem. Field (2005) suggests VIF of 10 as a critical threshold.

Another test for MLR is Durbin Watson (D-W) statistical test. This test is for lag-one autocorrelation of residual. Stage 4 is for model validation based on PI value using 20% complete data sampling. OLS assumptions were checked in Stage 4 (Table 2); then, the models were validated by a performance indicator (NAE, PA, IA, and R2) in accordance with a 20% sampling data. The MLR model was obtained in Stage 5.

![Figure 1. Stages of Linear Regression Models Development.](image-url)
Table 1. Interpretation of the strength of correlation results [7].

| Correlation coefficient range | Strength of correlation |
|-------------------------------|-------------------------|
| 0.00-0.30                    | Weak                    |
| 0.31-0.50                    | Moderate                |
| 0.31-0.50                    | Strong                  |
| 0.81-1.00                    | Very strong             |

Table 2. Ordinary least square assumption [8].

| Assumption                              | Checked by                           |
|-----------------------------------------|--------------------------------------|
| Residuals follow a normal distribution  | Normal P-P plot                      |
| Residual has constant variance          | Scatter plot (the spread of point)   |
| Residuals are uncorrelated with the Independent variables | Durbin Watson test statistics (no \autocorrelation) |

To evaluate the performance models for predictions, performance indicators are used [9]. The performance models (Table 3) consist of accuracy measures (PA, IA and R\(^2\)) and error measures (NAE).

Table 3. Performance indicators [10].

| Performance Indicator (PI)            | Notes                               |
|---------------------------------------|-------------------------------------|
| Normalized Absolute Error (NAE)       | Close to 0, model is appropriate    |
| Index of Agreement (IA)               | Close to 1, model is appropriate    |
| Prediction Accuracy (PA)              | Close to 1, model is appropriate    |
| Coefficient of determination (R\(^2\))| Close to 1, model is appropriate    |

3. Result and Discussion

Linear regressions have been developed from sites sampling data (Site Taman Ilmu, Seri Akasia and Seri Putera) using IBM SPSS Statistics 23. The Durbin-Watson test statistic tests for auto-correlations between errors. It tests whether adjacent residual is correlated because one of the assumptions of residual is that it should be independent. From Table 4, the values of Durbin-Watson are closer to two indicating that the assumption is satisfied. The models do not have any auto correlation problem.

Table 4 Result for Durbin-Watson Test.

| Sites       | Linear Regression Models | Durbin-Watson |
|-------------|--------------------------|---------------|
| Taman Ilmu  | BW = 1.605A - 2.359      | 1.613         |
| Seri Akasia | BW = 4.050 + 1.675A      | 1.944         |
| Seri Putera | BW = 6.916 + 1.235A      | 1.921         |
|             | BW=Brick Waste, A=Area   |               |

The plot of the residuals shown in Figure 2 fits the expected pattern well enough to support the conclusion that these residuals are normally distributed with zero means.
Figure 2. Histogram of brick waste residual.
This residual plot exhibits a random scatter; as such, no obvious pattern can be observed in Figure 2. The assumption that the residual displays a constant variance is satisfied when the scatter plot exhibits an equal spread and approach to the regression line (homoscedacity). Moreover, the assumption that the residuals are uncorrelated with the independent variable is satisfied because the Durbin–Watson value (2.2) is close to 2. In summary, the analysis described above verifies that the developed model can be used to predict the brick waste generated by housing construction projects.

Figure 3. Scatter plot of residual versus fitted values.
The plot of the residuals shown in Figure 3 fits the expected pattern well enough to support the conclusion that these residuals are normally distributed with zero means. This residual plot exhibits a random scatter; as such, no obvious pattern can be observed in Figure 3. The assumption that the residual displays a constant variance is satisfied when the scatter plot exhibits an equal spread and approach to the regression line (homoscedacity). Moreover, the assumption that the residuals are uncorrelated with the independent variable is satisfied because the Durbin–Watson value (2.2) is close to 2. In summary, the analysis described above verifies that the developed model can be used to predict the brick waste generated by housing construction projects.

Performance indicators were used to measure the accuracy such as Prediction Accuracy (PA), Coefficient of Determination ($R^2$) and Index of Agreement (IA) and error Normalised Absolute Error (NAE) for linear regression models. Table 5 shows the performance indicators for brick waste. The results show that Site Taman Ilmu and Seri Putera prediction is more accurate for predicting the brick waste generated. Accuracy measure for Site Seri Putera and Taman Ilmu model (PA, IA and $R^2$) is greater than 0.7 indicating that predicted values are a highly representative model.

| Sites       | Linear Regression Models | NAE   | IA    | PA    | $R^2$ |
|-------------|--------------------------|-------|-------|-------|-------|
| Taman Ilmu  | $BW = 1.605A - 2.359$    | 0.2663| 0.9328| 0.8785| 0.7718|
| Seri Akasia | $BW = 4.050 +1.675A$     | 0.2794| 0.9217| 0.8635| 0.7456|
| Seri Putera | $BW = 6.916 + 1.235A$    | 0.2691| 0.8400| 0.7409| 0.5489|

$BW =$ Brick waste, $A =$ Area, Prediction Accuracy =$PA$, Coefficient of Determination =$R^2$, Index of Agreement =$IA$, Normalised Absolute Error =$NAE$
Previous research was conducted a study using multiple linear regression and found the set of these variables was able to predict approximately up to 69% of the factors involved in waste generation in high-rise residential buildings of the sample, since the statistical model presented coefficient of determination ($R^2$) = 0.784 and the adjusted coefficient of determination (adjusted $R^2$) 0.694 [11]. This sampling involved 18 buildings with a different company.

A study on estimation of construction and demolition waste generation in Lisbon Metropolitan Area, Portugal found that population density, building density and percentage of urban [12]. These explained the waste generation as the factors contributes to high correlation coefficients with 0.82, 0.68, and 0.79, respectively. Apart from that, in China a model was developed to integrates the mass balance principle, break-down the work structure, take-off of material quantity, conversion ratios between different waste measurement units and the wastage levels of various materials used in different work packages [13]. The proposed model can predict the quantities of various kinds of construction waste from a building project, to track the origin of construction waste (i.e., from which work package is a particular kind of waste generated and how much) and to help contractors investigate the potential improvements for waste management.

A quantification model of waste generated in high-rise building construction in Brazil using statistical multiple produced dependent (i.e. the amount of waste generated) and independent variables associated with the design and the production system used [14]. The best regression model obtained from the sample data resulted in an adjusted $R^2$ value of 0.694, which means that it predicts approximately 69% of the factors involved in the generation of waste in similar constructions. Most independent variables show a low determination coefficient when assessed in isolation, which emphasises the importance of assessing their joint influence on the response (dependent) variable.

3.1 Validation and Verification of Models

Validation and verification are independent procedures that are used together for checking that a models meets requirements and specifications and that it fulfills its intended purpose. Table 6 shows the comparison of performance indicator between validation and verification at the three sites for the brick waste.

| Brick Waste | NAE | RMSE | A |
|-------------|-----|------|---|
| Site        | Validate | Verify | Validate | Verify | Validate | Verify |
| Taman Ilmu  | 0.2653 | 0.5094 | 20.377 | 18.513 | 0.9103 | 0.6024 |
| Seri Akasia | 0.3269 | 0.7742 | 19.309 | 26.244 | 0.8779 | 0.5606 |
| Seri Putera | 0.2490 | 0.4421 | 12.552 | 15.790 | 0.8882 | 0.5871 |

4. Conclusion

This study recommends the use of linear regressions to investigate the amount of brick waste generated from brick installation activities in housing construction sites. The analyses were only conducted from the construction stages. The projected amount may be utilized for future improvement in brick activities. The most common brick waste producers are the unskilled workers who had been assigned to do the skill work. Because educational work (to become the skill one) is a process, waste may come from many areas within this process. The site management is the person who responsible in controlling the amount of waste produces at sites.

Acknowledgments

The authors would like to acknowledge Universiti Teknologi MARA Cawangan Pulau Pinang and Faculty of Civil Engineering for supporting this research and publication.

References

[1] Li J Ding Z Mi X and Wang J 2013 A model for estimating construction waste generation index for building project in china, *J Resour Conserv Recy.* 74, p.20-26.
[2] Osmani M 2011 Construction Waste. Waste: A Handbook for Management p.207-218.
[3] Katz A and Baum H 2011. A novel methodology to estimate the evolution of construction waste in construction sites. Waste manag, 31(2), p.353-358.
[4] Montgomery D C Peck E A and Vining G G 2012 Introduction to linear regression analysis 821, John Wiley & Sons.
[5] Draper N R and Smith H 2014 Applied regression analysis. Wiley series in probability and statistics New York, USA: John Wiley & Sons, Inc., 4.
[6] Muhamad M Ul-Saufie A Z and Deni S M 2015 Three days ahead prediction of daily 12 hours ozone (O₃) concentrations for urban area in Malaysia J Environ Sc and Tech, 8(3) p.102-112.
[7] Nangolo C and Musingwini C 2011 Empirical correlation of mineral commodity prices with exchange-traded mining stock prices J South Afr Inst Min Metall, 111, p.459-468.
[8] Chatterjee S and Hadi A S 2006 Regression analysis by example John Wiley, New Jersey, 4, p.416.
[9] Lu W Chen W X Peng Y and Shen L 2015 Benchmarking construction waste management performance using big data, J Resour Conserv Recy. 105, p.49-58.
[10] Gervasi O 2008 Computational science and its applications Springer, New York, USA.
[11] Andrea P K Dias M F Kulakowski M P and Gomes L P 2015 Waste generated in high-rise buildings construction, a quantification model based on statistical multiple regression J Waste Manag, 39, p.35–44.
[12] Bernardo M Gomes M C, and Brito J D 2016 Demolition waste generation for development of a regional management chain model J Waste Manag, 49, p.156–169.
[13] Lia Y Zhang X Ding G And Zhouquan F 2016 Developing a quantitative construction waste estimation model for building construction projects J Resour Conserv Recy., 106, p 9–20.
[14] Kern A P Dias M F Kulakowski M P and Gomes L P 2015 Waste generated in high-rise buildings construction: a quantification model based on statistical multiple regression J Waste Manag, 39, p. 35–44.