Development of a Maintenance and Repair Cost Estimation Model for Educational Buildings Using Regression Analysis

Ji-Myong Kim¹, Taehui Kim², Yeong-Jin Yu³ and Kiyoung Son*⁴

¹ Ph.D., Construction Science Department, Texas A&M University, USA
² Professor, Department of Architectural Engineering, Mokpo National University, Republic of Korea
³ M.S. Student, School of Architectural Engineering, University of Ulsan, Republic of Korea
⁴ Assistant Professor, School of Architectural Engineering, University of Ulsan, Republic of Korea

Abstract
The aim of this study is to identify key performance indicators as well as the correlations among these indicators to develop a maintenance and repair cost estimation model for educational buildings based on actual payment records. The importance of financial estimation for facility management systems has increased. Especially, the estimation of maintenance and repair costs is essential to facility management for educational buildings considering the long-life cycle of a building. However, data regarding facility management is still limited. This study adopts payment records related to maintenance and repair work from educational institutions to develop a quantitative approach. Statistical analyses and a multiple regression analysis are conducted to examine the record and generate a cost estimation model. The findings and results of this study provide a guide for maintenance and repair cost estimation of integrated facility management and could be used as a guideline for budgeting for school maintenance.

Keywords: Florida; educational buildings; maintenance and repair cost; regression analysis

1. Introduction
Educational facilities have differences in the quality of education according to the quality level of the facility. Securing and maintaining facilities at a suitable level is necessary to support the quality of education. However, facility maintenance expenses are significant and the existing post-maintenance system cannot maintain the original functions of school buildings. Accordingly, it is necessary to be prepared in advance for possible damages and to set aside reserves for failures and accidents. Therefore, setting up strategies to manage facilities budgets for schools is essential. In order to address this requirement, the industry has focused on establishing an integrated facility management system (FMS) to support the operation of educational properties. FMS attempt to maintain the functions of a building, prevent errors in design and construction caused by various factors after building completion, and restrict factors that accelerate the deterioration of the building associated with misuse and poor management.

Cost estimates for an FMS are limited (Amaratunga 2000). The range of expenditures for maintenance and repair work are broader than for any other FMS, since the work includes damages caused by human activity, natural hazards, and the deterioration of facilities and equipment. In addition, educational properties have a higher uncertainty than general properties owing to the comprehensive and long life cycle of the building, which also makes it difficult to predict maintenance costs. Moreover, educational buildings have relatively less interest than other buildings. Although, well operated and maintained educational properties can significantly improve the satisfaction and accomplishment of occupants, i.e., students, staff, and teachers (Lavy and Bilbo 2009). However, educational properties are on average older than other building types and 25% of such buildings are not properly maintained. For this reason, there are many challenges concerning maintenance and repair for educational buildings (Li et al. 2005).

Therefore, this study proposes a model that can determine maintenance and repair costs using statistical analysis of the actual cost records. This study investigates expenses including maintenance and repair at educational facilities to determine the Key Performance Indicators (KPIs) and to develop a cost estimation model for integrated facility management. The outcomes and results of this study include cost estimates for maintenance and repair that are expected to improve budgetary control.
2. Research Methods

The purpose of this study is to analyze maintenance and repair costs for the FM of educational facilities. To reach this goal, this research includes five phases to define KPIs and to determine interactions between payments regarding maintenance and repair work and indicators. First, the key performance indicators and estimation models of FM are explored based on former studies. Second, the payment records are gathered as a dependent variable. Third, payment records are explored in the category of season and loss causes, and are used to create a frequency and severity matrix based on the analysis of causes. Fourth, various categories of KPIs are utilized, i.e., building information, environmental vulnerability, and natural hazards, to examine the wide range of expenditures via comprehensive research. Lastly, the KPIs and record are analyzed utilizing a multiple regression analysis method.

3. Literature Reviews

The FM’s role in maximizing investment and improving the effectiveness of education is widely recognized. Lavy and Bilbo (2009) found that many US students study in schools that are poorly maintained (Lavy and Bilbo 2009). He also maintains that accurate FM performance evaluation for US educational facilities is required to overcome this problem in order to improve both FM and education performance.

Therefore, studies on the FM performance evaluation of educational facilities have been carried out. For instance, a balanced scorecard was proposed by Kaplan and Norton (2005). It was intended to link the long-term vision and outcomes of short-term operational strategies based on four perspectives of finance, management, users, learning, and growth (Kaplan and Norton 2005). Baldry et al. (2000) presented a study of facility performance assessment approaches related to the characteristics of tertiary education, and developed a framework founded on balanced scorecards (Amaratunga and Baldry 2000). Kok et al. (2011) presented a literature review of services and effects on educational achievement, as well as the function of facilities due to FM activities in the educational environment (Kok et al. 2011). Lavy et al. (2014) identified and classified KPIs based on previous studies for FM evaluation, classifying them as financial, functional, physical, and user-based, based on the literature (Lavy et al. 2014). Tamosaitiene et al. (2013) developed a program based on game theory to select the optimal service using the FM factor. Each characteristic was classified into general management, security, cleaning, and building characteristics (Tamosaitienë et al. 2013).

However, even if research on FM performance evaluation and indicators occurred, it is still necessary to define indicators to reflect particular circumstances (Woodhouse 2011) because it is challenging to perform FM continuously for asset management for non-commercial buildings. Furthermore, a quantitative evaluation and a focused classification system are required for highly reliable evaluations and facility management on site. To address these problems, this study analyzes damage based on the characteristics of each educational facility using historical data on quantified maintenance and repair costs.

In addition, there are many studies on the maintenance and repair of educational facilities. However, there are few studies of maintenance and repair frequency and cost estimation methods. Kim et al. (2010) explained that there is a vast difference in building reliability and repair costs. Sometimes, early replacement is advantageous (Kim et al. 2010). For these reasons, estimating maintenance and repair costs is problematic. Maintenance and repair uncertainty makes it hard to decide when to build new facilities. Consequently, statistical analysis and numerical estimation are needed to define the causes of damage and establish a cost estimation model for maintenance and repair work.

4. Data Collection

This research adopted records from Florida Independent Colleges and the Universities Risk Management Association (FICURMA) to reveal the real pecuniary loss on maintenance and repair costs for educational buildings. FICURMA is a risk management pool founded in 2003, which covers financial risk for educational institutions for private schools, colleges, and universities in Florida. The association insures the properties, machines, vehicles, vessels, etc. Nonetheless, this study only sorted expenses that were related to maintenance and repair costs from claims since 2004. Total incurred costs, which exclude insurance conditions such as deductible, liability of limit, and so on, were used to assess the

| Category         | % of total | Mean    | SD     | Max     | Min     |
|------------------|------------|---------|--------|---------|---------|
| Leakage          | 24.47%     | 37,192  | 55,383 | 378,461 | 238     |
| Weather Event    | 18.43%     | 102,085 | 135,413| 500,000 | 6       |
| Overflow         | 10.88%     | 94,444  | 183,389| 988,888 | 995     |
| Vehicle          | 9.97%      | 7,966   | 11,566 | 42,141  | 18      |
| Lightning        | 7.55%      | 27,354  | 26,585 | 94,285  | 71      |
| Mold             | 6.95%      | 85,179  | 149,771| 656,015 | 836     |
| Crime            | 4.83%      | 17,258  | 20,094 | 60,142  | 85      |
| Others           | 4.83%      | 54,536  | 143,516| 603,591 | 475     |
| Fire             | 4.23%      | 96,598  | 101,032| 264,862 | 722     |
| Sprinkler        | 2.11%      | 214,407 | 284,220| 801,664 | 10,791  |
| Mechanical       | 1.81%      | 60,085  | 42,115 | 121,019 | 3,181   |
| Failure          | 1.81%      | 52,070  | 107,633| 295,832 | 1,891   |
| Vandalism        | 0.60%      | 1,958   | 1,358  | 3,315   | 600     |

Table 1. Descriptive Statistics for Cause
policies that are referred to as natural hazards. Table 3. denotes the indicators of each category. First, building information is measured by the type of building. Second, environmental vulnerability is assessed by the type of building. Third, location is categorized as either a rural area, urban area, or metropolitan area. Finally, the campus is divided into four groups by season.

5. Data Analysis

Payment records in this section are identified by season and cause. Each instance is identified in terms of frequency and severity. Frequency is the number of losses that occur, and severity is the mean value of the loss.

5.1 Causes of Payment

The horizontal line represents the cause of payment and the vertical line represents the mean value of the payment for each group. The frequency matrix can be segmented into four zones. Zone A indicates that both frequency and severity are small. Zone B indicates that the frequency is small, but the severity is large. Zone C indicates that the frequency is large, but the severity is small. Zone D indicates that both the frequency and severity are large. For example, zone B includes a sprinkler failure with extraordinarily large severity and another with small severity. This suggests that properties that have suitable regular maintenance and repair work may avoid significant damage. On the other hand, leakage in zone C has a remarkably large frequency and small severity. This suggests that the loss could happen anywhere and anytime. Accordingly, the FM manager and worker should focus on leakage prevention.

5.2 Seasonal Variation

The record is grouped into four classes by season and analyzed as shown in Table 2. The most repeatedly affected season is summer, with 34.3% of instances. Fall has a mean payment value of 85K USD. The summer, winter, and spring have 74K, 62K, and 50K USD individually in that order. Furthermore, each season has a different pattern of payment causes. For instance, leakage, overflow, and weather events are the major causes of payment in the spring, accounting for 29.0%, 13.0%, and 10.1%, respectively. However, even if the pattern of payment causes are various, leakage and weather events are the most common causes of payment throughout the seasons. Hence, the analysis shows that FM administrators and employees should be aware of the expenses that can occur due to various causes.

5.3 Statistical Examination of Data

This research employs the multiple regression analysis method to generate a model for maintenance and repair work. After collection, the payments are converted to ratios, as shown in Equation (1).

\[ \text{Ratio} = \frac{\text{Maintenance and Repair Cost (USD)}}{\text{Total Area of Buildings (Ac)}} \]  

(1)

5.3.1 Independent Variables

This study used three categories: building information, environmental vulnerability, and natural hazards. Table 3. denotes the indicators of each category. First, building information is measured by two indicators, total building area and year the building was built. Second, environmental vulnerability is grouped into an indicator as the location of the campus. The location of the campus is categorized as either a rural area, urban area, or metropolitan area. Third, three indicators given as natural hazards, including the rating of tropical cyclone risk at the campus, the rating of lightning risk at the campus, and the rating of a FEMA flood zone, are all used to represent the overall natural hazard rating. This research adopts the risk rating of a natural hazard risk map, the Natural hazards. The statistics are shown in Table 1. Water leakage, weather event, overflow, vehicle, and lightning were the most commonly arising payment events, at 24.47%, 18.43%, 10.88%, 9.97%, and 7.55% respectively. Sprinkler, weather event, fire, overflow, and mold were the most critical payment events in view of average payment amounts of 214, 102, 96, 94, and 85 K USD.
Hazards Assessment Network (NATHAN), from the Munich Reinsurance Company. The online risk map is settled to estimate the natural hazards worldwide. The map can explain the various risks of natural hazards by rating them depending on the location information. The tropical cyclone risk and lighting at the campus are adopted to understand the nature of the natural hazards at a campus. Furthermore, this study uses the ratings of FEMA Flood Map, FEMA flood zone X and FEMA flood zone AE to describe the risk of flood. Zone X denotes that the zone is outside the 500-year flood field. Zone AE denotes that a flood will likely occur within 100 years.

Table 3. Explanation of Independent Variables

| Category            | Indicator | Explanation                        | Unit |
|---------------------|-----------|------------------------------------|------|
| Building information| Building area | Total area of building (acres)   | Number |
|                     | Building age | Year the building was built | Number |
| Environmental vulnerability | Campus location | Location of the campus | 1. Rural area 2. Urban area 3. Metropolitan area |
| Natural Hazards     | Tropical cyclone rating | Total area of building | Year the building was built |
|                     | Frequency of lightning (yearly per km²) | Frequency of lightning | 0: 0-2.1 1: 2.1-4 2: 4.1-10 3: 10.1-20 4: 20.1-40 5: 40.1-80 |
|                     | Lighting rate | Rating of lightning risk at the campus | Rating of lightning risk | 0: 0-2.1 1: 2.1-4 2: 4.1-10 3: 10.1-20 4: 20.1-40 5: 40.1-80 |
|                     | FEMA Flood Zone rating | Rating of FEMA Flood Zone | 0. FEMA flood zone X 1. FEMA flood zone AE |

5.3.2 Multiple Regression Analysis

The descriptive statistics describe the dependent and independent variables as seen in Table 4.

Table 4. Descriptive Statistics

| Category       | N  | Min. | Max. | Mean | SD  |
|----------------|----|------|------|------|-----|
| Ratio          | 331| 0.00 | 7716.33 | 461.00 | 977.21 |
| Building area  | 331| 49.00 | 314.00 | 192.40 | 98.55 |
| Building age   | 331| 36.00 | 0     | 128.00 | 64.41 |
| Campus location| 331| 1.00 | 3.00  | 1.32  | 0.48 |
| Tropical cyclone| 331| 3.00 | 4.00  | 3.87  | 0.32 |
| Lightning      | 331| 4.00 | 5.00  | 4.35  | 0.48 |
| FEMA Flood Zone| 331| 0.00 | 1.00  | 0.48  | 0.50 |

The model summary is shown in Table 5. In the dependent variable, the ratio, is transformed by the natural log. The P-value of 0.005 is smaller than 0.05, which implies that the regression model is statistically significant. The adjusted R-square value of 0.355 identifies the relationship between the dependent and independent variable that defines 35.5% of the variance.

Table 5. Models Summary

| Model             | Sum of Squares | Mean Square | F   | Sig. | R² | Adj. R² |
|-------------------|----------------|-------------|-----|------|----|---------|
| Regression        | 66.537         | 22.179      | 33.098 | 0.000 | 0.366 | 0.355   |
| Residual          | 115.257        | 0.67        | 0.000 | 1.000 |
| Total             | 181.794        |             |      |      |    |         |

Table 6. shows the coefficients of the model. The three indicators, total area of building, location of the campus, and rating of FEMA Flood Zone, have statistically significant relations to the ratio. Additionally, there is no serious multicollinearity among these three indicators, since the range of the variance inflation factor (VIF) is from 1.392 to 3.998.

Table 6. Coefficients of the Model

| Indicators      | B     | Std. Error | Beta | Sig. | VIF |
|-----------------|-------|------------|------|------|-----|
| Constant        | 3.610 | 0.324      | 0.000|      |     |
| Construction information and ability |       |            |      |      |     |
| Area            | -0.003| 0.001      | -0.291| 0.002*| 2.272|
| Age             | 0.000 | 0.005      | 0.002| 0.981| 2.426|
| Environmental vulnerability |       |            |      |      |     |
| Location        | 1.037 | 0.170      | 0.439| 0.000*| 1.394|
| Natural disaster |       |            |      |      |     |
| Tropical cyclone| 0.260 | 0.278      | 0.074| 0.937| 1.666|
| Lightning       | -0.121| 0.274      | -0.054| -0.441| 3.998|
| FEMA Flood Zone | 0.773 | 0.165      | 0.379| 0.000*| 1.782|

*Significance at the 0.05 level (2-tailed)

Equation 2 presents the multiple regression model, which uses the three above variables to estimate the natural log transformed ratio. The model accounted for 35.5% of the variance of the dependent variable.

\[
\ln (\text{Ratio}) = 3.610 + (-0.003) \cdot \text{Total area of building} + 1.037 \cdot \text{Location of the campus} + (0.773) \cdot \text{Rating of FEMA Flood Zone}
\]  

(2)

5.3.3 Examination of the Model

As shown in Table 7., the Kolmogorov-Smirnov test was chosen to consider the normality of the residual. The P-value of 0.059 is greater than 0.05, which means that the residuals are customarily discrete. Fig.2 explores the homoscedasticity of the residuals. The unsystematic extent forms of the residual verify that the residuals are unsystematically spread. This confirms that the variance of the residual is constant. Furthermore, the residual histogram (a) and the Q-Q plot (b) demonstrate that the residuals are ordinarily distributed, as seen in Fig.3.

Table 7. Model Normality Test

| Kolmogorov-Smirnov Statistic | Sig. |
|-----------------------------|------|
| In (ratio)                  | 0.059| 0.200|
5.3.4 Model Validation

As shown in Fig. 4., the scatter plot compares the results of the authentic natural log transformed ratio and anticipated natural log transformed ratio. The defined indicators are able to explain 35.5% of the variability in the dependent variable following the adjusted R-square value of the model, 0.355. The remainder of the variability of the dependent variable can be identified utilizing unrevealed indicators. The scatter plot verifies that the values are consistent.

This study employs a cross-validation methodology that compares the coefficient of the sum of the squared error (SSE) and the predicted error sum of square statistic (PRESS). Table 8. shows that the model likely can be fitted with other data sets, since the value of PRESS, 120.497, is similar to the value of SSE, 115.257. Subsequently, the validation results suggest that this model will be able to reliably predict the dependent variable ratio.

6. Conclusion

Facility management affects not only the operation, maintenance, and monitoring of buildings, but also the occupants. Educational buildings significantly affect students and researchers as well. Therefore, it is essential to budget appropriately for building maintenance. A cost estimation model is a useful approach. This study examines payment records related to maintenance and repair work at educational facilities quantitatively. A frequency and severity matrix was created and seasonal variation is explored to define the severity, frequency, and cause of expense by season based on the records.

In addition, this study identifies key performance indicators and uses multiple regression analysis to establish a cost estimation model for integrated facility management. The findings and results of this research may be employed as a guideline for educational buildings. For instance, building maintenance managers in education facilities would not only reduce the losses but also make a mitigation strategy to prevent losses based on the analysis results. Besides, it is possible for real estate and property managers to estimate the annual operation and maintenance cost and to set up financial plans and long-range investment strategies using the cost estimation model. Furthermore, the framework and indicator of this study can be strongly applied for other types of building such as hospital, industrial, and commercial buildings, which need to develop maintenance and repair cost estimation models.

Acknowledgement

This research was supported by a grant (NRF-2016 R1A2B4009909) from the National Research Foundation of Korea by the Ministry of Science, ICT and Future Planning.

References

1) Amaratunga, D. (2000). Assessment of facilities management performance, Property Management, 18(4), pp.258-266.
2) Lavy, S., and Bilbo, D. L. (2009). Facilities maintenance management practices in large public schools, Texas, Facilties, 27(1/2), pp.5-20.
3) Li, P. P., Locke, J., Nair, P., and Bunting, A. (2005). "Creating 21st century learning environments." PEB Exchange, Programme on Educational Building, 2005(10), OECD Publishing, Paris.
4) Kaplan, R. S., and Norton, D. P. (2005). The balanced scorecard: measures that drive performance, Harvard business review, 83(7), p.172.
5) Amaratunga, D., and Baldry, D. (2000). Assessment of facilities management performance in higher education properties, Facilities, 18(7/8), pp.293-301.
6) Kok, H. B., Mobach, M. P., and Omta, O. S. (2011). The added value of facility management in the educational environment, Journal of Facilities Management, 9(4), pp.249-265.
7) Lavy, S., A. Garcia, J., and K. Dixit, M. (2014). KPIs for facility's performance assessment, Part I: identification and categorization of core indicators, Facilities, 32(5/6), pp.256-274.
8) Tamošaitienė, J., Peldschus, F., and Al Ghanem, Y. (2013). Assessment of Facility Management Candidates by Applying Game Theory, Procedia Engineering, 57, pp.1145-1150.
9) Woodhouse, J. Optimal timing for replacing aging or obsolete assets, Proc., Asset Management Conference 2011, IET and IAM, IET, pp.1-10.
10) Kim, J.-R., Jung, Y.-H., and Son, J.-H. (2010). A study on reliability analysis model of the repair and replacement cycle of a building which utilizes Monte Carlo Simulation, Journal of the Korea Institute of Building Construction, 10(2), pp.41-50.