

DEVELOPMENT OF AN APPROXIMATE CONSTRUCTION DURATION PREDICTION MODEL DURING THE PROJECT PLANNING PHASE FOR GENERAL OFFICE BUILDINGS

Dong-Jun YEOM¹, Hae-Mi SEO¹, Yoo-Jun KIM¹, Chung-Suk CHO², Youngsuk KIM^{1*}

¹*Department of Architectural Engineering, Inha University, Incheon, Korea*

²*Department of Civil Infrastructure and Environmental Engineering, Khalifa University of Science and Technology, Abu Dhabi, UAE*

Received 09 August 2017; accepted 20 March 2018

Abstract. Accurate prediction of the construction duration is imperative to the reliable cash flow analysis during the project planning phase when feasibility analysis is carried out. However, lack of information and frequent changes that occur as a result of a negotiation process between the owner and the designer in defining the project scope make it difficult to compute real-time construction duration. Domestic and foreign models for calculating the construction durations cannot be readily applied to computation of construction duration for general office buildings in Korea specifically during the project planning phase as there is a limit in its applicability due to numerous restrictions. Moreover, there are no preceding studies suggesting different computational approaches to predict the entire construction duration for office buildings with the approximate construction duration concept during planning phase. Therefore, based on the collected performance data, this study proposes a multiple linear regression model that facilitates reliable prediction of approximate construction duration for office buildings in the project planning phase. The model will allow the owner and other stakeholders to predict the real-time construction duration using the basic information on office buildings and to assess the construction durations incorporating frequent changes during the project planning phase.

Keywords: general office building, approximate duration, prediction model, multiple linear regression analysis, project planning phase, construction schedule.

Introduction

Since 2000, with the vitalization of the second new city development project, which strengthens self-sufficiency through the formation of a multifunctional administrative city, innovation city, enterprise city, etc., there is the trend of a continuous increase in the demand for general office buildings in Korea.

Because these general office buildings are sale-in-lots or rent businesses, and cash flow for the owner acts as a major component determining the success and failure of the project, it is important to manage the cash flow smoothly by predicting the real-time construction duration from the project planning phase when feasibility analysis is carried out (Hwang 2002; Ko, Han 2017). However, there is a limit on the information that can be obtained in the project planning phase by the owner. This can occur because changes frequently happen during that phase. Any change is most likely to take place as a result of a negotia-

tion process between the owner and the designer, directly affecting the construction duration and the project cost. Therefore, in computing the real-time construction duration it is necessary to include all anticipated changes because they impact on both the owner and other stakeholders in the project.

The need for an accurate estimation of construction duration from an early stage is apparent (Kaka, Price 1991; Love *et al.* 2005; Hoffman *et al.* 2007; Jin *et al.* 2016). Inaccurate estimation of project duration, whether under or over, results in poor project performance as well as failure of meeting project objectives (Khosrowshahi, Kaka 1996; Lin *et al.* 2011; Jin *et al.* 2016). Therefore, predicting the real-time construction duration by incorporating the characteristics of a pertinent project and facilities can have persuasive power towards various stakeholders of a project, including the users of the facility while providing the

*Corresponding author. E-mail: youngsuk@inha.ac.kr

contractor with a real-time construction duration that is necessary for carrying out the construction. This can help prevent many problems, such as compromising the quality of construction due to fast-phased construction work, possibility of disputes due to the generation of additional costs, increase in the project budget, etc. (Thomas *et al.* 2001; Owolabi *et al.* 2014; Kim 2014; Mačková, Bašková 2014).

Domestic and foreign models for calculating the construction durations are difficult to apply in practice when computing a construction duration for office buildings in project planning phase, because they only adopt a part of the items related to the construction duration of particular facilities, such as an apartment building ordered by a public organization, military facilities, etc. as variables. Moreover, there is a limit in its applicability due to the limitations caused by obsolescence of old standards and size of constructions. According to a literature review, studies related to the computation of construction duration for office buildings have been limited to analyzing the trends of the construction durations in accordance with influential factors or suggesting a limited approach for computing construction duration for particular work type, such as the construction duration for steel-frame work. In addition, the analysis showed that there are no preceding studies suggesting different approaches to predict the entire construction duration for office buildings with the concept of approximate construction duration in the project planning phase.

Therefore, the purpose of this study was to suggest a multiple linear regression model that can help predict approximate construction duration of general office buildings in the project planning phase for the owner and contractor. The model is based on the performance data collected from a number of Korean general contracting companies regarding the construction duration of general office buildings. The results of this study are expected to allow the owner and other stakeholders of the project to predict the real-time construction duration using the basic information on office buildings and to assess the construction durations incorporating frequent changes during the project planning phase. Therefore, it is expected that owners and project stakeholders will be able to improve the reliability as well as the appropriate timing of their decision making based on reliable construction duration prediction.

1. Scope and method of the study

The scope of this study is limited to general office buildings in Korea, where specialized models for computing construction duration during the project planning phase do not exist. In accordance with the Enforcement Decree of the Building Act, office buildings are defined as financial business offices, service business offices, such as matchmaking services, publishers, newspaper company, other similar buildings (hereinafter referred to as the “office”), and efficiency apartments. In addition, approximate construction duration in this research refers to the total construction duration from the beginning to the end of

the construction during the construction phase, which is predicted in the project planning phase. The research methods used in this study are as follows:

First, based on an analysis of domestic and foreign models for computing the construction duration and a review of preceding studies, this study analyzed the status and problems of calculating approximate construction duration for general office buildings in Korea during the project planning phase.

Second, by establishing independent variables for an analysis of the influential factors and development of prediction models in relation to the computation of the construction duration of office buildings, which are the subjects of this study, the performance was collected from general contracting companies. In addition, the current status of the collected performance data was analyzed, and an approximate construction duration prediction model was developed for domestic office buildings through statistical analysis methods, such as correlation analysis, multiple linear regression analysis, and analysis of the variance. Third, based on the case studies applied to local office buildings, the accuracy and validity of the developed model was verified by comparison analysis between the actual construction duration and the predicted construction duration calculated from an approximate construction duration prediction model. Figure 1 presents a diagram summarizing the method and procedures of this research.

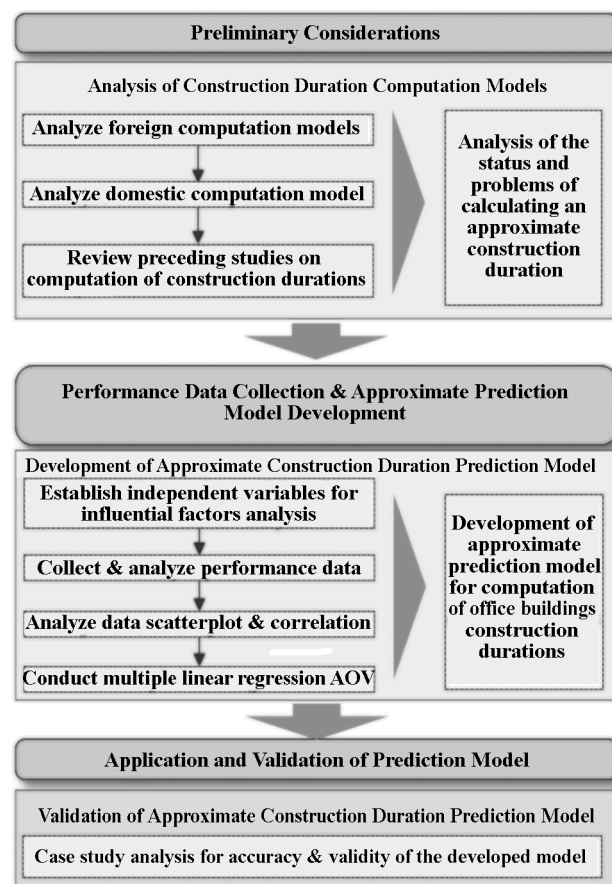


Figure 1. Research method and procedures

2. Preliminary study

2.1. Analysis of the status and problems of calculating an approximate construction duration for general office buildings during the project planning phase

Generally, the construction duration for office buildings are determined in two ways: the owner defines the project duration directly (Jin *et al.* 2016; Bayram 2017; Peško *et al.* 2017) or the contractor suggests construction duration to the owner. On the other hand, the construction duration computation model utilized by contractors during the project planning stage (Table 1) is not a specialized standard for general office buildings (Hwang *et al.* 2002; Koo *et al.* 2010; Jin *et al.* 2016; Kim *et al.* 2016; Thomas, N., Thomas, A. V. 2016). In addition, variables applied to this model are also limited thereby increasing the chance of error.

Table 1. Construction duration computation model by contractors during the project planning phase

$T = (\text{Total floor}D)\alpha\beta\gamma$	
T : Total project duration	α : Correction for gross floor area
D : Duration per each floor (= Total construction duration/Total floor)	β : Correction for construction method
	γ : Correction for project delivery method

Note: Total floor = Number of basement floor + Number of above ground floor.

The investigation (Hwang *et al.* 2002) showed that Ito's and Asakura's computation formulae were utilized universally for domestic computation models for calculating the construction durations. Both formulae have an on-site construction speed coefficient for offices and a calibrating constant while applying variables that can be held during the project planning phase, making them applicable when calculating construction duration for office buildings in the project planning phase. Nevertheless, updating the data is necessary, as the performance data has become old (Hwang *et al.* 2002). In particular, in the case of the Asakura computation formula, there is a limit in its applicability, because it can be applied only to newly constructed buildings that are no taller than 30 floors with a total floor area no less than 500 m².

In addition, domestic computation models for computing construction durations have been prescribed for particular facilities, such as apartment buildings, military facilities, etc., so there are limitations in their applicability to computing construction duration for general office buildings. Koo *et al.* (2010) developed a case-based reasoning hybrid model for predicting construction cost and duration and based on project characteristics in multi-housing projects. Jin *et al.* (2016) developed a case-based reasoning model for multi-housing projects estimating the construction duration in the project planning phase and validated its applicability and effectiveness through application of 83 multi-housing projects. Czarnigowska and So-

botka (2014) presented a model of the road construction duration based on relationships between the project qualities. These relationships were assumed to be determined while analysing project qualities likely to be known or estimated in the project planning phase. Kim *et al.* (2016) suggested a prediction model for dam construction duration based on activities that have a main effect on the overall dam construction duration. Because specific items (amount of rocks, pile depth, etc.) are established as influential factors, there is a limit in utilizing them during the project planning phase (Lee *et al.* 2010). General construction companies in Korea utilize their own construction duration computation models that are specialized for office buildings. On the other hand, they are also difficult to utilize during the project planning phase because they apply specific variables, such as the amount of soil excavated, rock composition ratio, number of tower cranes needed, etc. (Lee *et al.* 2010).

Despite aforementioned problems and limitations of these computation models, preceding studies on the construction duration of office buildings have been limited to influential factors when computing a construction duration, trend analysis of the track record construction durations based on construction companies, or different approaches for computing the construction durations of particular work types (Table 2). Ji (1990) suggested a specific regression equation for computing the construction duration for steel-frame work through regression analysis based on the performance data of high-rise office buildings. Such studies on computing the construction durations of specific work types have limitations, as they can only be utilized during the design phase during when specific information regarding the corresponding work type can be obtained and they are unsuitable for predicting the entire construction duration.

Martin *et al.* (2006) suggested a regression equation for computing the construction duration for various facilities including office buildings with the purpose of predicting the construction duration during the project planning phase. On the other hand, because it establishes only the cost of construction as the influential factor, there is a limitation such that a prediction of the construction duration is not possible without an accurate prediction of the cost in advance, and updating the performance data in accordance with changes in the construction environment is also necessary (Lee *et al.* 2010).

2.2. Theoretical study of the statistical analysis method

Correlation analysis is a method for determining the relationship between two variables objectively, where the correlation coefficient generally means the Pearson correlation coefficient. The correlation coefficient, r , between two variables X and Y , when the values of each case are $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, was calculated using Eqn (1):

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (1)$$

Table 2. Preceding studies on computing the construction duration for general office buildings

Classifications	Research(s)	Contents of research
Research on influential factors and analysis of construction duration computation models by construction companies	Hwang (2002)	Conducted comparative analysis of actual durations based on the influential factors for office building construction and suggested the need for construction duration reduction and target completion duration through a comparison of national and international actual construction durations of building constructions.
	Ryu <i>et al.</i> (2006)	Analyzed the competitiveness of the construction durations of Korean construction companies by comparing the actual construction durations of multipurpose buildings and efficiency apartment with virtual cases. Suggested measures to secure the construction duration competitiveness during each construction project phase.
Research on assessment of specific construction as well as total construction duration	Ji (1990)	Suggested a specific regression equation ($R^2 = 0.9504$) for computing the construction duration for steel-framed, high-rise office buildings through regression analysis using independent variables, such as the number of ground and base floors, steel-beam size, and quantity.
	Martin <i>et al.</i> (2006)	Suggested a regression equation for computing construction duration for various facilities including office buildings using the performance data from British construction projects with the purpose of predicting construction duration during the project planning phase. Used the log value of the construction cost as an independent variable.

The correlation coefficient, r , is between -1.0 and $+1.0$, and a negative and positive correlation exists if $-1.0 < r < 0$ and $0 < r < 1.0$, respectively. This study examined the correlation between the independent and dependent variables of a model for an approximate construction duration prediction for general office buildings using the Pearson correlation coefficient.

This paper proposes a prediction model that allows the owners and other stakeholders to predict an approximate construction duration by multiple linear regression analysis, which facilitates an understanding of the relationship between the dependent and independent variables while predicting the dependent variable under various situations. Multiple linear regression analysis is a linear regression model that considers several independent variables simultaneously when there are at least two independent variables explaining the dependent variable. From the dependent variable denoted by Y and k independent variables denoted by X_1, \dots, X_k , the multiple linear regression model can be written as Eqn (2):

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2). \quad (2)$$

Here, ε is the error term that was assumed to be normally distributed with a mean of 0 and variance of σ^2 . In this regression model, regression coefficients, $\beta_0, \beta_1, \dots, \beta_k$, are estimated from data using the least square method. β_i indicates the amount of change in the dependent variable, Y , when the independent variable, X_i , increases by one unit. For example, when the total floor area, which is an independent variable, increases by 1 m^2 , the construction duration, which is a dependent variable, changes by as much as the regression coefficient of the total floor area variable. This study derived an approximate construction duration prediction model for office buildings that is statistically significant and has high explanatory power by incorporating several independent variables through a multiple linear regression model using SPSS 18.0.

Variance analysis was carried out to verify the suitability of a regression model. Because a regression model is derived under the assumption that the dependent variable is affected linearly by at least one independent variable, in order for the derived regression model to be meaningful, a null hypothesis, such as “None of the independent variable affects the dependent variable” has to be set, and the hypothesis test should return a statistically significant result. To achieve this objective, this study verified the suitability of a prediction model (significance level of 0.05 or lower) based on the significance probability (P-value) for the F value in the ANOVA table derived from SPSS 18.0.

3. Development of approximate construction duration prediction model for office buildings in project planning phase

3.1. Analysis of the influential factors in relation to the computation of construction duration for domestic office buildings and establishment of independent variables for prediction model

In this study, prior to collecting the performance data, the influential factors related to the computation of a construction duration for office buildings in Korea were collected through research on previous studies on the existing influential factors for computation models to establish independent variables to be reflected in the prediction model.

As a result, while there are influential factors, such as the purpose, region, structural form, land area, total floor area, number of basement floors, and above ground floors that can be obtained in the project planning phase, specific factors, such as the characteristics of the foundation underground, drilling area, volume of soil excavated, cut-off area, etc., which require the user’s expertise and can be known only upon reaching the design phase, and influential factors, such as the structural form, which require a

Table 3. Analysis of the influential factors for office buildings construction duration computation

Classification		Construction Duration Influential Factors
Previous studies	Hwang (2002)	Structural form, foundation underground characteristics, climate and construction conditions, total floor area, number of basement and above ground floors, number of finishing construction, owner sponsors
	Ryu <i>et al.</i> (2006)	Region, lot area, total floor area, number of basement and above ground floors, site condition, retaining wall methods, foundation form, structural form, finishing requirement, building use, climatic conditions, heating system
Computation models	Ito Computation Model	Total floor area, structural form, facility use, field construction speed, contractors' possible mobilization of daily workforce (economic construction speed)
	Asakura Computation Model	Building area, total floor area, number of basement and above ground floors, penthouse floors, pile type, soil type, structural form, facility use
	Engineering & construction firm	Excavation area, daily excavation volume, volume of soil excavated, cut-off area, number of basement and above ground floors, steel-frame structure, total floor area, building form (simple/complex)

subjective judgment, were applied. Table 3 summarizes the influential factors related to the computation of construction duration for general office buildings in Korea derived from this study.

Based on this, the independent variables of an approximate construction duration prediction model for domestic office buildings have been established, as shown in Table 4. The variables includes the type of a facility, region, lot area, building area, total floor area, number of basement and above ground floors, and the structural form that can be obtained in the project planning phase and whose performance data are easy to collect. In addition, the months of the start of construction, which affect the climate and construction process and number of buildings that affects the

construction size, are also considered. By forming specific subcategories for each category, a total of 19 variables were derived.

The land area, construction area, total floor area, number of basement and above ground floors, and number of buildings were established as the ratio scales, which are the scale variables, because they were qualitative factors. The office, efficiency apartment, metropolitan area, Chungcheong region, Jeonla region, Gyeongsang region, Gangwon region, reinforced concrete structure, steel-framed reinforced concrete structure, steel-frame structure, winter season, summer season, and others were established as the nominal scales that are the categorical variables, as they are quantitative factors.

Table 4. Establishment and characteristics of the independent variables for the prediction model

Category	Subcategory	Scale
Facility type	Office	Nominal scale*
	Efficiency apartment	
Project region	Metropolitan area Chungcheong region Jeonla region Gyeongsang region Gangwon region	Nominal scale*
Lot area	() m ²	Ratio scale
Building area	() m ²	Ratio scale
Total floor area	() m ²	Ratio scale
Number of floors	Above ground	Ratio scale
	Basement	
Structural form	Reinforced concrete structure	Nominal scale*
	Steel-framed reinforced concrete structure	
	Steel-frame structure	
Construction commencement month	Winter season (Nov – Feb)	Nominal scale*
	Summer season (Jun – Aug)	
	Other seasons	
	(Mar – May / Sep – Oct)	
Number of buildings	Building number ()	Ratio scale

* Nominal variables are transformed into dummy variables so as to include them in the multiple regression model. Therefore, they are binary variables: "0" if the attribute is absent and "1" if the attribute is present.

3.2. Collection and status of the performance data for general office buildings in Korea

This study collected the performance data on the construction of office buildings from four Korean general construction companies based on the independent variables established previously. The collected data comprised of 47 cases in total that are made up of 22 cases of offices and 25 cases of efficiency apartments, and except for one case (began construction in 1995), all were constructions that began and were completed after 2000 (between 2002–2013).

Figure 2 presents the distribution status of the performance data. The distribution of the performance data according to the region of construction showed that 90% of the entire performance data were records of constructions carried out in the metropolitan area, as shown in Figure 2a. Accordingly, statistical analyses were carried out after removing the variable for the regions.

The distribution status by structural forms (Figure 2b) also concentrated on reinforced concrete (RC) and steel-framed reinforced concrete (SRC) structures due to the limit in collecting the performance data. Therefore, it would have been reasonable to remove the corresponding variable, but the statistical analyses were carried out without removing that variable because it was determined to be a variable that affects the framework construction duration, which comprises a large proportion of the entire construction duration. Supplementation in relation to the accumulation of additional performance data of steel-frame structures will be necessary in the future. In addition, an analysis of the histogram of the dependent variable (construction duration) showed a distribution that is similar to the normal distribution, as shown in Figure 3 below.

In this study, 47 cases of collected performance data were divided randomly into a group of 37 cases of analysis data to be used for statistical analysis for the development of an approximate construction duration prediction model

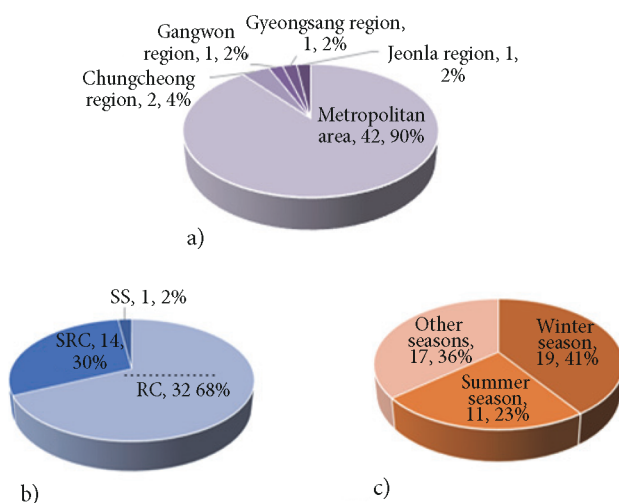


Figure 2. Distribution status of the performance data: a – distribution status per construction region; b – distribution status per structural forms; c – distribution status per construction commencement month

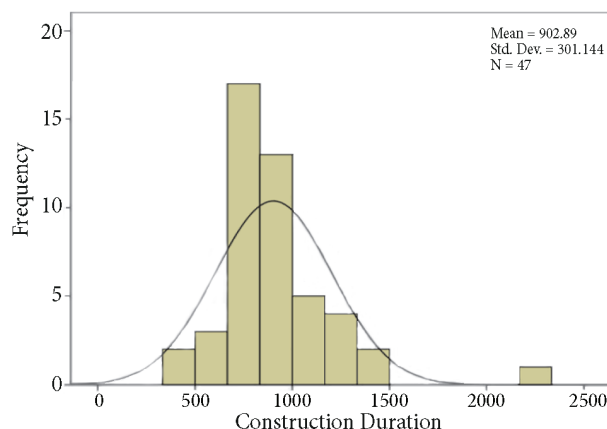


Figure 3. Distribution status of the performance data per construction duration

and a group of 10 cases of validation data to be used for validation of the prediction model.

3.3. Scatter plot by variable and correlation analysis

This study reviewed the linear relationship between the dependent variable, construction duration, and the independent variables based on 37 cases of analysis data. In the case of the categorical variables of which a linear relationship is difficult to review by the nature of their characteristics (type of facilities, structural forms, month of the construction commencement, etc.), the trend of the average construction duration was analyzed through a line chart depending on the categories. For the scale variables (land area, construction area, total floor area, number of floors above the ground, number of basement floors, number of buildings), a linear relationship with the dependent variable, construction duration, was reviewed through a scatter plot. After confirming the linear relationship, numerical correlations between the construction duration and the independent variables and between the independent variables were understood through correlation analysis.

Figure 4 shows the trend of the average construction duration of the analysis data based on the type of facilities, structural forms, and months of the construction commencement. Based on the type of facilities, the average construction duration of the offices tended to be higher than the average construction duration of the efficiency apartments (Figure 4a). This is because the offices have more diverse floor plans than efficiency apartments and because of the need for additional procedures, such as construction of curtain walls resulting in a slight increase in construction duration. Based on the structural forms, the average construction duration increased in the order of RC, SRC, and SS (Figure 4b), and based on the months of the start of construction, it increased in the order of winter, summer, and other months (Figure 4c).

Figure 5 presents the matrix of the scatter plots between the dependent variable, construction duration, and the scale variables among independent variables – lot area, building area, total floor area, number of floors above the

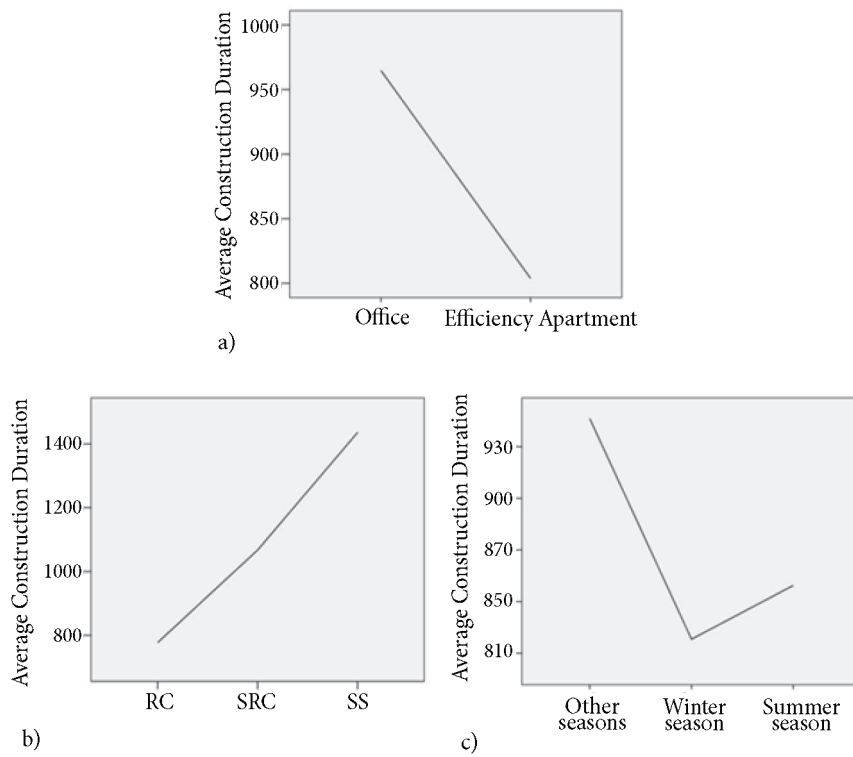


Figure 4. Trend of average construction duration based on the analysis data: a – average construction duration trend based on facility type; b – average construction duration trend based on structural form; c – average construction duration trend based on construction commencement month

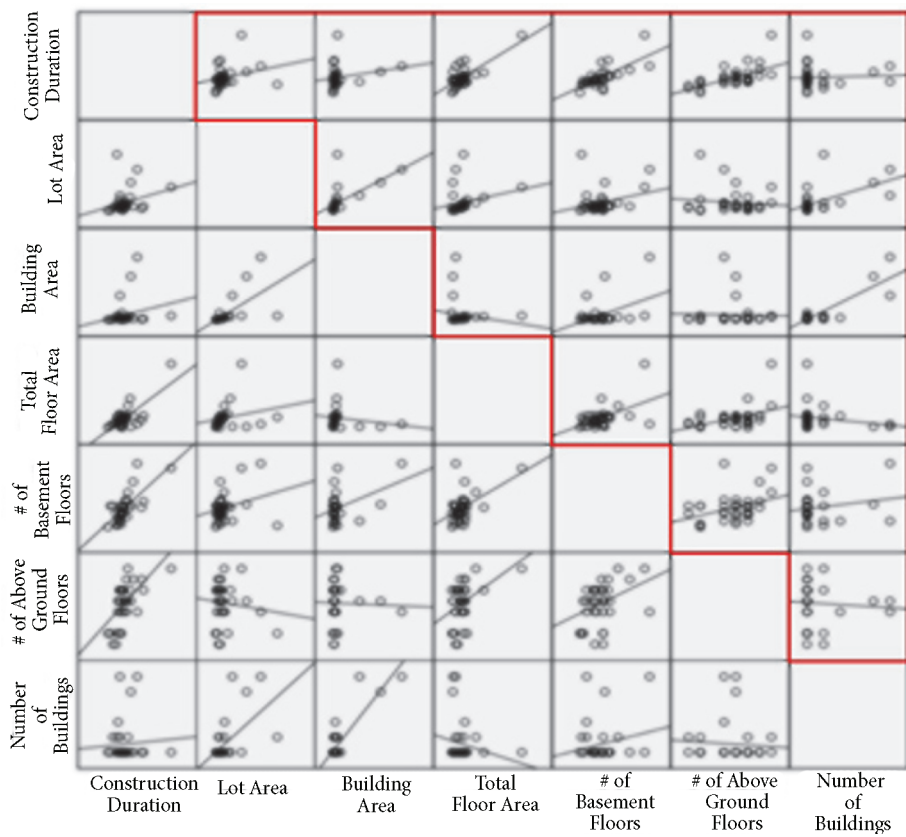


Figure 5. Matrix of scatter plots between the construction duration and scale variables

Table 5. Results of correlation analysis – Pearson correlation coefficient

Classification	Construction duration	Office	Efficiency apartment	Lot area	Building area	Total floor area	# of above ground floors	# of basement floors	RC	SRC	SS	Winter season	Summer season	Other seasons	Number of buildings
Construction duration	1														
Office	.250	1													
Efficiency apartment	-.250	-1.000**	1												
Lot area	.269	.057	-.057	1											
Building area	.216	-.213	.213	.624**	1										
Total floor area	.740**	.271	-.271	.227	-.146	1									
# of above ground floors	.714**	-.085	.085	.276*	.423**	.510**	1								
# of basement floors	.602**	.211	-.211	-.113	-.030	.435**	.370*	1							
RC	-.464**	-.506**	.506**	-.163	.140	-.322*	-.086	-.198	1						
SRC	.370*	.452**	-.452**	.188	-.124	.318*	.060	.164	-.936**	1					
SS	.294*	.191	-.191	-.058	-.054	.036	.078	.109	-.256	-.101	1				
Winter season	-.143	.057	-.057	.035	-.219	.097	-.019	-.035	.055	-.131	.202	1			
Summer season	-.040	.072	-.072	-.068	.091	-.207	-.179	-.018	-.089	.124	-.088	-.434**	1		
Other seasons	.178	-.119	.119	.023	.145	.077	.172	.051	.020	.027	-.130	-.644**	-.410**	1	
Number of buildings	.055	-.319*	.319*	.573**	.874**	-.194	.193	-.071	.224	-.205	-.069	-.215	-.016	.232	1

Note: ** – correlation coefficient of 0.01 at significance level, * – Correlation coefficient of 0.05 at significance level.

ground, number of basement floors, and number of buildings. Looking at the relationship between the construction duration and independent variables (first column in Figure 5), most are related linearly. In particular, a strong correlation exists with the total floor area, number of floors above the ground, and the number of basement floors. In addition, although the analysis shows that most of the relationships between the independent variables are not linearly related and have a low correlation, it is believed that there is a slight correlation between the lot area and building area, the lot area and number of buildings, the building area and number of buildings, and the total floor area and number of floors above ground.

A linear relationship was confirmed using the scatter plot derived previously. Because there was no particular curvilinear relationship, it was determined to be reasonable to understand the numerical relationships between the independent and dependent variables and between the independent variables, based on 37 cases of analysis data. The result is as shown in Table 5.

The independent variables with the strongest correlation with the dependent variable, i.e., construction duration, were the total floor area with a correlation coefficient 0.740 and the number of floors above the ground with a correlation coefficient 0.714, showing a very strong relationship. The correlation coefficient between the construction duration and number of basement floors was 0.602, and -0.464 with RC, showing a significant relationship. Correlation coefficients with other variables showed 0.370 for SRC and 0.294 for SS, showing a slight relationship.

An examination of the correlation coefficients between the independent variables revealed a correlation between the following: the lot area and the building area, the total floor area and the number of floors above the ground, the lot area and the number of buildings, and the building area and the number of buildings. Because the building area or

the number of buildings is likely to increase with increasing lot area, and because the total floor area increases with increasing number of floors above the ground, they have an influence on each other. Collinearity can be understood through relationships among the independent variables, and in general, the problem of collinearity is considered to exist if there is a strong correlation ($r \geq 0.90$). Therefore, to inspect such collinearity more rigorously, collinearity statistics derived through regression analysis need to be verified.

3.4. Analysis of results using multiple linear regression analysis method

In this study, multiple linear regression analysis was carried out to derive an approximate construction duration prediction model for office buildings in the project planning phase. Dependent and independent variables were established, as listed in Table 6, and a dummy variable was applied to carry out the analysis when the independent variable was a nominal scale.

To determine the status of collinearity, which was a matter of concern in previous correlation analysis, this study performed regression analysis using an input method that incorporates all the independent variables into the regression model at the same time as that for the method of selecting the variables. Tables 7, 8 and 9 present the results of a multiple linear regression analysis using the input method.

According to the analysis, the adjusted R^2 of the regression model was 0.868, showing high explanatory power, and variance analysis showed a significance probability of 0.000, which is lower than 0.05, so the model was found to be statistically significant. Table 9 lists the results of an analysis of the regression coefficient for each independent variable, but only the coefficients for the constant, total floor area, number of basement floors, SRC, SS, and other

Table 6. Establishing dependent and independent variables for multiple linear regression analysis

Dependent variable	Variable		Unit		Scale	
	Construction Duration		Day		Ratio scale	
Independent variable	Variable	Unit	Scale	Variable	Unit	Scale
	Office	-	Nominal	RC	-	Nominal
	Efficiency Office	-	Nominal	SRC	-	Nominal
	Lot Area	m ²	Ratio	SS	-	Nominal
	Building Area	m ²	Ratio	Winter Season	-	Nominal
	Total Floor Area	m ²	Ratio	Summer Season	-	Nominal
	# of Above Ground Floors	Floor	Ratio	Other Seasons	-	Nominal
	# of Basement Floors	Floor	Ratio	Number of Building	Bldg.	Ratio

Table 7. Multiple linear regression analysis using the input method – model summary

R	R ²	Adjusted R ²	Standard Error of Estimated Value	Durbin-Watson
.953 ^a	.908	.868	117.684	1.893

Notes: a – Predicted value: (constant), number of building, winter season, number of basement floors, SS, SRC, number of above ground floors, other seasons, efficiency apartment, total floor area, lot area, building area; b – Dependent variable: construction duration.

Table 8. Multiple linear regression analysis using the input method – variance analysis

Model	Sum of Square	Degree of freedom	Mean Square	F	Significance Probability
Regression model	3418700.406	11	310790.946	22.441	.000 ^a
Residual	346236.297	25	13849.452		
Sum	3764936.703	36			

Notes: a – Predicted value: (constant), number of building, winter season, number of basement floors, SS, SRC, number of above ground floors, other seasons, efficiency apartment, total floor area, lot area, building area.

Table 9. Multiple linear regression analysis using the input method – coefficient

Independent variable	Unstandardized coefficient	Standardized coefficient	t	Significance probability	Collinearity statistic	
	B	Beta			Tolerance	VIF
(Constant)	256.930	–	2.571	0.016	–	–
Efficiency apartment	14.182	0.022	0.275	0.786	0.573	1.745
Lot area	–0.001	–0.021	–0.198	0.845	0.332	3.014
Building area	0.006	0.321	1.618	0.118	0.093	10.696
Total floor area	0.005	0.501	4.815	0.000	0.340	2.939
# of above ground floors	8.141	0.229	1.992	0.057	0.278	3.592
# of basement floors	40.912	0.233	3.059	0.005	0.636	1.572
SRC	139.438	0.194	2.544	0.018	0.632	1.583
SS	586.248	0.298	4.559	0.000	0.861	1.162
Summer season	113.686	0.147	1.930	0.065	0.637	1.571
Other seasons	107.161	0.163	2.216	0.036	0.681	1.469
Number of building	–30.816	–0.126	–0.769	0.449	0.136	7.329

Table 10. Multiple linear regression analysis using the stepwise selection method – model summary

Step	R	R ²	Adjusted R ²	Standard error of estimated value	Durbin-Watson
1	.740 ^a	.548	.535	220.584	–
2	.837 ^b	.701	.684	181.881	
3	.882 ^c	.778	.758	159.134	
4	.908 ^d	.825	.803	143.430	
5	.922 ^e	.850	.826	134.905	
6	.941 ^f	.886	.863	119.572	
7	.951 ^g	.905	.882	111.131	

Notes: a – Predicted value: (constant), total floor area; b – Predicted value: (constant), total floor area, # of above ground floors; c – Predicted value: (constant), total floor area, # of above ground floors, RC; d – Predicted value: (constant), total floor area, # of above ground floors, RC, # of basement floors; e – Predicted value: (constant), total floor area, # of above ground floors, RC, # of basement floors, SRC; f – Predicted value: (constant), total floor area, # of above ground floors, RC, # of basement floors, SRC, winter season; g – Predicted value: (constant), total floor area, # of above ground floors, RC, # of basement floors, SRC, winter season, building area; h – Dependent variable: construction duration.

seasons whose significance probabilities are less than 0.05 can be considered to be statistically significant. Collinearity exists if the tolerance is less than 0.1 or if VIF > 10. In this model, collinearity exists because VIF of the construction area variable was greater than or equal to 10. Therefore, the final regression analysis was performed using a stepwise selection method to increase the accuracy of the model by increasing the number of statistically significant variables and minimizing the effects of collinearity.

The dependent and independent variables of multiple linear regression analysis utilizing the stepwise selection method were established to be the same as those established by the input method (Table 6). According to multi-

ple linear regression analysis carried out using a stepwise selection method, seven types of linear regression models were suggested, as listed in Tables 10 and 11, and the total floor area variable was entered first. Because the adjusted R² was generally higher for a regression model at each step (Table 10), the independent variables of the regression model are considered to explain the dependent variable sufficiently well. The variable for number of floors above the ground was entered at step 2, RC at step 3, number of basement floors at step 4, SRC at step 5, winter season at step 6, and building area at step 7.

The variable(s) of low importance can be removed if a stepwise selection method is used, but in the current

Table 11. Multiple linear regression analysis using the stepwise selection method – variance analysis

Step	Model	Sum of square	Degree of freedom	Mean square	F	Significance probability
1	Regression model	2061925.875	1	2061925.875	42.376	.000 ^a
	Residual	1703010.828	35	48657.452		
	Sum	3764936.703	36			
2	Regression model	2640198.774	2	1320099.387	39.906	.000 ^b
	Residual	1124737.929	34	33080.527		
	Sum	3764936.703	36			
3	Regression model	2929252.464	3	976417.488	38.557	.000 ^c
	Residual	835684.238	33	25323.765		
	Sum	3764936.703	36			
4	Regression model	3106630.287	4	776657.572	37.753	.000 ^d
	Residual	658306.416	32	20572.075		
	Sum	3764936.703	36			
5	Regression model	3200753.656	5	640150.731	35.174	.000 ^e
	Residual	564183.046	31	18199.453		
	Sum	3764936.703	36			
6	Regression model	3336009.534	6	556001.589	38.888	.000 ^f
	Residual	428927.168	30	14297.572		
	Sum	3764936.703	36			
7	Regression model	3406785.904	7	486683.701	39.407	.000 ^g
	Residual	358150.799	29	12350.028		
	Sum	3764936.703	36			

Notes: a – Predicted value: (constant), total floor area; b – Predicted value: (constant), total floor area, # of above ground floors; c – Predicted value: (constant), total floor area, # of above ground floors, RC; d – Predicted value: (constant), total floor area, # of above ground floors, RC, # of basement floors; e – Predicted value: (constant), total floor area, # of above ground floors, RC, # of basement floors, SRC; f – Predicted value: (constant), total floor area, # of above ground floors, RC, # of basement floors, SRC, winter season; g – Predicted value: (constant), total floor area, # of above ground floors, RC, # of basement floors, SRC, winter season, building area; h – Dependent variable: construction duration.

Table 12. Multiple linear regression analysis using the stepwise selection method – coefficient

Independent variable	Unstandardized coefficient	Standardized coefficient	t	Significance probability	Collinearity statistic	
	B	Beta			Tolerance	VIF
(constant)	897.272		6.456	.000		
Total floor area	.004	.467	5.765	.000	.500	2.001
# of above ground floors	9.896	.278	3.283	.003	.456	2.192
RC	-578.234	-.829	-4.949	.000	.117	8.544
# of basement floors	40.646	.231	3.513	.001	.758	1.319
SRC	-437.319	-.609	-3.599	.001	.115	8.724
Winter season	-110.373	-.170	-2.775	.010	.875	1.143
Building area	.003	.176	2.394	.023	.607	1.649

regression mode, the variables that were entered were not removed. Durbin-Watson statistics, which determine the autocorrelation status, was 1.926, which is very close to 2, so autocorrelation does not appear to exist. According to the ANOVA result (Table 11), the goodness of fit for all 7 regression models showed a significance probability of 0.000; thus, all 7 models appear to be appropriate models for explaining the dependent variable. Hence, any model can be selected from models 1–7. In this study, model 7 was selected because it incorporates various independent variables and has the highest modified R^2 value, a value that represents the explanatory power of construction duration.

Table 12 lists the results of coefficients analysis of the final model. The estimated regression equation can be

written using unstandardized regression coefficients in Table 12. Because the significance probabilities of t statistics are lower than 0.05 for all regression coefficients, each of the independent variables is considered to have a significant effect on the dependent variable. Because the measurement units of the independent variables are all different, the relative influence of the independent variables on the dependent variable can be assessed through beta coefficients that have been analyzed by standardizing the analysis data. In model 7, the independent variables were shown to be influential in the order of RC, SRC, total floor area, number of floors above the ground, number of basement floors, building area, and winter season. Looking at the collinearity statistics to examine the issues of collin-

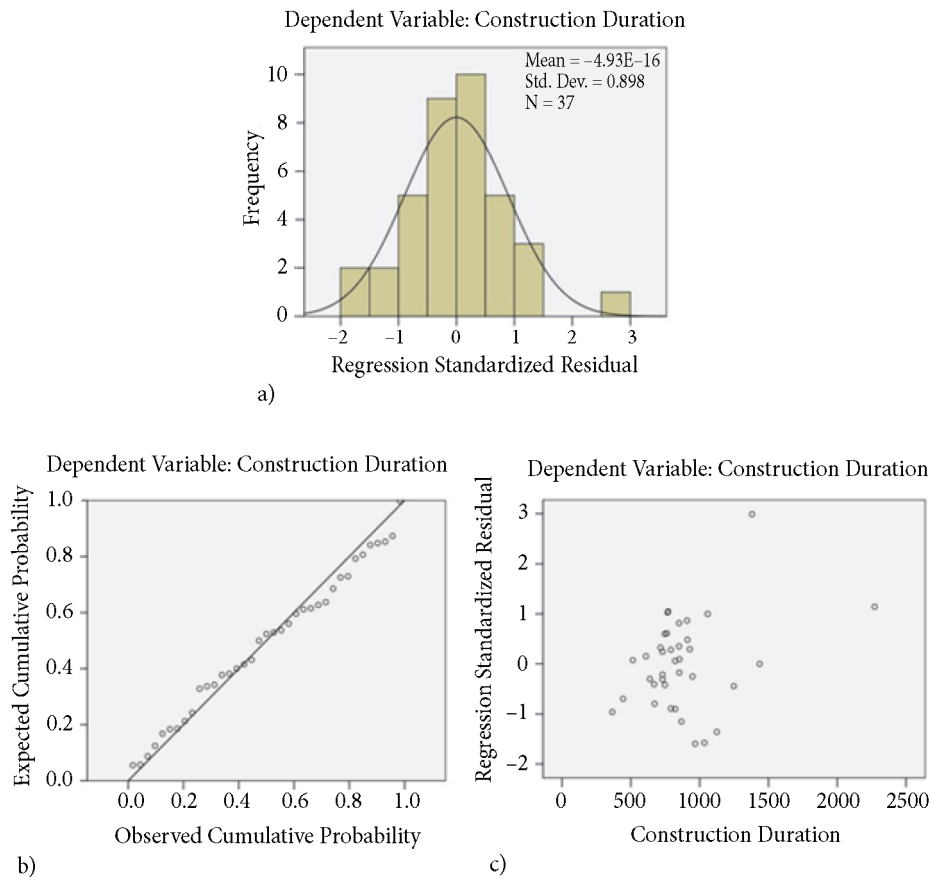


Figure 6. Result of residual analysis: a – regression standardized residual histogram; b – normal probability plot of standardized residual; c – scatter plot for dependent variable of standardized residual

earity, the tolerance limit was bigger than 0.1 and the VIF value was less than 10, confirming that there is no collinearity problems.

Residual analyses, such as the residual histogram, normal probability plot, etc. were carried out to determine if the residual is distributed normally, as shown in Figure 6. According to the result, the residual histogram showed a distribution that is close to a normal distribution, and the normal probability plot also showed points that are lined up linearly, suggesting that the residual is distributed normally. In addition, if there is a clear trend shown by the points of a scatter plot for the dependent variables of a standardized residual, it is interpreted as lacking randomness in the data or having not considered an important independent variable(s). As there was no clear trend in this study, this model is considered to be appropriate.

3.5. Suggestion of an approximate construction duration prediction mode for general office buildings in Korea during project planning phase

This study carried out correlation analysis, multiple linear regression, etc. by selecting the independent variables that have an effect on the construction duration of domestic office buildings. Through these analyses, a highly statistically significant regression model was derived. Conse-

quently, based on the derived regression model, this study proposed an approximate construction duration prediction model for general office buildings during the project planning phase, as expressed in Eqn (3):

$$\begin{aligned}
 T = & 897.272 + 0.004 \times \text{Total Floor Area} + 0.003 \times \\
 & \text{Building Area} + 9.896 \times \text{Number of Above Ground} \\
 & \text{Floors} + 40.646 \times \text{Number of Basement Floors} - \\
 & 437.319 \times \text{SRC} - 578.234 \times \text{RC} - 110.37 \times \text{Winter} \\
 & \text{Season Construction Commencement.} \quad (3)
 \end{aligned}$$

Here, T represents the approximate construction duration, and the variables reflected in the prediction model comprise a total of seven types – total floor area, building area, number of floors above the ground, number of basement floors, SRC, RC, and winter season construction commencement. The units of the total floor area and construction area are m^2 and the units of the number of floors above the ground and basement are floors. For the other categorical variables, such as SRC, RC, and winter season construction commencement, 1 is substituted if they are applicable and 0 otherwise to calculate the construction duration.

This prediction model will allow the end user to make simple and quick predictions of an approximate construc-

Table 13. Outline of the case study

Case	Case Summary				
	Duration	Facility type	Structural form	Commencement month	Number of buildings
	Lot area	Building area	Total floor area	# of above ground floors	# of basement floors
1	763	Office	RC	Other seasons	1
	2648.9	1318.6	26395.1	15	4
2	822	Office	RC	Summer season	1
	4466.6	2661.4	51264.0	14	5
3	850	Office	SRC	Winter season	1
	2129.5	1194.5	20473.9	12	5
4	945	Office	SRC	Other seasons	1
	3794.0	2384.5	36521.8	16	6
5	973	Efficiency apartment	RC	Winter season	2
	4592.0	2695.0	60909.0	26	6
6	1033	Office	SRC	Winter season	1
	7886.0	4680.5	85603.4	24	4
7	1124	Efficiency apartment	RC	Winter season	2
	10133.0	2061.0	75702.0	37	6
8	1186	Efficiency apartment	RC	Other seasons	3
	9490.0	4897.0	99463.7	32	3
9	1188	Efficiency apartment	RC	Summer season	2
	6325.0	3013.0	78309.0	37	4
10	1237	Office	SRC	Summer season	1
	5914.4	2184.7	91829.7	33	6

tion duration based only on the basic information in the project planning phase when there is insufficient information on the project. In addition, as a tool for evaluating the suitability of construction duration, it is expected that this model will ensure reliability and proper timing of decision making when establishing construction duration.

4. Application and verification of the approximate construction duration prediction model for office buildings using case studies

4.1. Outline of the case study

This study validated the prediction model based on 10 cases of validation data that were separated from the performance data on office buildings collected from general construction companies in Korea; Table 13 outlines the case study. The case studies consisted of six cases of offices and four cases of efficiency apartments, and the number of floors above the ground ranges from 12 to 33. As for the structural form, there are six cases of RC and four cases of SRC. The months of the construction commencement consist of four cases of winter, three cases of summer, and three cases for other seasons.

4.2. Computation of the construction duration through case studies and verification of the prediction model validity through analysis of results

Based on the above outline of the case studies, this study calculated the predicted construction duration by substituting the total floor area, building area, number of floors above the ground, number of basement floors, SRC and

RC, and information on the winter season into the prediction model, and all the predicted construction durations were rounded to the first decimal place. In addition, the prediction performance of the developed model was analyzed by calculating the error rate and prediction accuracy through comparative analysis between the predicted construction duration, i.e., calculated and the actual construction duration of the performance data. The results are shown in Table 14. Eqn (4) was used to calculate the error rate (APE; Absolute Percentage Error); and the prediction accuracy (PA) was computed by subtracting the error rate from 100%L:

$$f_{APE(x)} = \frac{|T_{Actual} - T_{Predicted}|}{T_{Actual}} \times 100, \quad (4)$$

here: $f_{APE(x)}$ – absolute percentage error (error rate) computation formula; T_{Actual} – actual construction duration; $T_{Predicted}$ – predicted construction duration.

Figure 7 presents the results of graphing the actual construction duration, predicted construction duration, and the 95% confidence interval for the actual construction duration; the actual construction duration lies inside the 95% confidence interval. In particular, the predicted construction duration derived from a prediction model showed a larger error for the construction duration of offices compared to the efficiency apartments. The error occurrence rate is higher because offices have more variable floor plans than efficiency apartments and because of the need for additional procedures, such as curtain wall construction.

Table 15 lists the results of descriptive statistics for the error rate and prediction accuracy. The average error rate

Table 14. Calculation of the predicted construction duration, error rate, and prediction accuracy

Case	Predicted duration (A)	Actual duration (B)	Difference (A-B)	Error rate	Prediction accuracy
1	740	763	-23	3.01%	96.99%
2	874	822	52	6.33%	93.67%
3	757	850	-93	10.94%	89.06%
4	1016	945	71	7.51%	92.49%
5	962	973	-11	1.13%	98.87%
6	1107	1033	74	7.16%	92.84%
7	1128	1124	4	0.36%	99.64%
8	1171	1186	-15	1.26%	98.74%
9	1170	1188	-18	1.52%	98.48%
10	1405	1237	168	13.58%	86.42%

Table 15. Descriptive statistics of the error rate and prediction accuracy of the prediction model

Classification	N	Average	Standard deviation	Median	Minimum	Maximum
Error rate	10	5.28%	4.56%	4.67%	0.36%	13.58%
Prediction accuracy	10	94.72%	4.56%	95.33%	86.42%	99.64%

of the prediction model and the average prediction accuracy was 5.28% and 94.72%, respectively. Considering that the prediction model is applied in the early stage of business, i.e., in the project planning phase, this is considered to be a prediction result at a considerably reliable level.

In this study, a prediction model was estimated based on the performance data on 37 cases of office buildings. On the other hand, it is expected that the error would be reduced further if a prediction model can be developed by accumulating the additional performance data and applying new variables. At present, various case studies could not be performed for validation because of the limitations in collecting performance data, but in the future it will be necessary to carry out additional validity analyses of a prediction model through the verification of case studies on various office buildings.

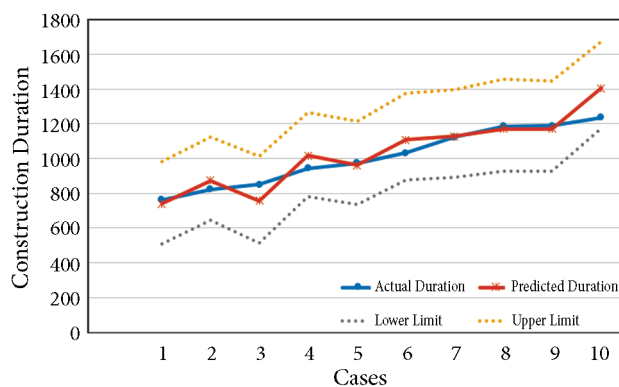


Figure 7. Actual construction duration vs. predicted construction duration

Conclusions

To develop an approximate construction duration prediction model for general office buildings, this study analyzed and identified influential factors and independent variables related to the computation of a construction duration with which 47 cases of performance data (divided into a group of 37 cases of analysis data and a group of 10 cases of validation data) on office buildings in Korea were collected and analyzed. Statistical studies pursued resulting in the development of an approximate duration prediction model. The study further verified the developed model using 10 cases of validation data. The results obtained through the prediction model developed in this study are as follows:

1. The need for estimating approximate construction duration for general office buildings in Korea, especially during the project planning phase, is paramount but existing computation models are found to be limited in their applications.
2. Through the analysis of influential factors related to the computation of construction duration for office buildings; type of facilities, region of construction, lot area, building area, total floor area, number of floors (number of floors above the ground and number of basement floors), months of the construction commencement, and number of buildings were established as independent variables.
3. Number of statistical analysis performed using independent variables and the construction duration as a dependent variable revealed a linear relationship between the construction duration, and most of the independent variables. According to correlation analy-

sis, independent variables that are strongly correlated with the construction duration were derived to be total floor area, number of floors above the ground, and number of basement floors with which there is a strong linear relationship. Using multiple linear regression and ANOVA analyses, an approximate construction duration prediction model was derived with high explanatory power ($R^2 = 0.905$, adjusted $R^2 = 0.882$) and statistical significance. The developed model was a multiple linear regression model with excellent applicability to the end user's prediction of an approximate construction duration. Independent variables, such as the total floor area, building area, number of floors above the ground, number of basement floors, RC, SRC, and winter season were incorporated.

4. The verification results showed that the developed model had an average error of 5.28% (i.e. an average prediction accuracy of 94.72%), hence, deriving a reliable result.

The approximate construction duration prediction model presented in this paper enhanced the accuracy of the construction duration forecast since the model not only can reflect the historical data available during the project planning phase but also can easily incorporate frequent changes of owner's requirement. In addition, the presented model fully addressed the latest construction technologies by incorporating performance data collected from latest building constructions.

For the future related study, it is recommended to collect additional performance data, such as the structural form, and thereby supplementing the prediction mode. In doing so, it is expected that the accuracy and reliability of the prediction model can be improved further if there is a good circulation among continuously accumulating the performance data on general office buildings, a corresponding update of the approximate construction duration prediction model, and carrying out comparative analyses with construction durations of the performance data.

Acknowledgements

This work was supported by a research grant from Inha University.

Author Contributions

Dong-Jun Yeom conceived the idea for this study and wrote the paper. Hae-Mi Seo designed the early methodology. Yoo-Jun Kim partially contributed to the development of the model and to the editing of the paper. Youngsuk Kim and Chung-Suk Cho conceived the idea, supervised the research, and revised the paper. All authors have read and approved the final paper.

Disclosure Statement

The authors declare that we do not have any financial, professional, personal interests or benefit arising from the

direct applications of this research nor competitions from other parties.

References

- Bayram, S. 2017. Duration prediction models for construction projects: In terms of cost or Physical characteristics?, *KSCE Journal of Civil Engineering* 21(6): 2049–2060. <https://doi.org/10.1007/s12205-016-0691-2>
- Czarnigowska, A.; Sobotka, A. 2014. Estimating construction duration for public roads during the preplanning phase, *Journal of Engineering, Project, and Production Management* 4(1): 26–35.
- Hoffman, G. J.; Thal Jr., A. E.; Webb, T. S.; Weir, J. D. 2007. Estimating performance time for construction projects, *Journal of Management in Engineering* 23(4). [https://doi.org/10.1061/\(ASCE\)0742-597X\(2007\)23:4\(193\)](https://doi.org/10.1061/(ASCE)0742-597X(2007)23:4(193))
- Hwang, H. S. 2002. *Analysis of actual duration by effecting elements to duration estimate – focused on standard duration of the office building construction*: Master's thesis. Hangyang University.
- Hwang, H. S.; Kim, K. R.; Suh, S. W.; Kim, C. D.; Shin, D. W. 2002. Analysis of actual duration by effecting elements to duration estimate, *Korean Journal of Construction Engineering and Management* 3(3).
- Ji, W. S. 1990. *A study on computation of construction period for high-rise office buildings*: Master's thesis. Chung-Ang University.
- Jin, R. Z.; Han, S. W.; Hyun, C. T.; Cha, Y. W. 2016. Application of case-based reasoning for estimating preliminary duration of building project, *Journal of Management in Engineering* 142(2). [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001072](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001072)
- Kaka, A.; Price, A. D. F. 1991. Relationship between value and duration of construction projects, *Construction Management and Economics* 9(4): 383–400. <https://doi.org/10.1080/01446199100000030>
- Khosrowshahi, F.; Kaka, A. P. 1996. Estimation of project total cost and duration for housing projects in the U.K., *Building and Environment* 31(4): 375–383. [https://doi.org/10.1016/0360-1323\(96\)00003-0](https://doi.org/10.1016/0360-1323(96)00003-0)
- Kim, J.; Yun, W.; Kim, I. 2016. Estimating approximate construction duration of CFRD in the planning stage, *KSCE Journal of Civil Engineering* 20(7): 2604–2613. <https://doi.org/10.1007/s12205-016-0810-0>
- Kim, S. H. 2014. *Estimation model for no-construction activity duration reflecting characteristics of each construction work type*: Master's thesis. Inha University.
- Ko, Y.; Han, S. 2017. A duration prediction using a material-based progress management methodology for construction operation plans, *Sustainability* 9: 635. <https://doi.org/10.3390/su9040635>
- Koo, C.; Hong, T.; Chang, T.; Koo, K. 2010. A CBR-based hybrid model for predicting a construction duration and cost based on project characteristics in multi-family housing projects, *Canadian Journal of Civil Engineering* 37(5): 739–752. <https://doi.org/10.1139/L10-007>
- Lee, H. D.; Lee, S. H.; Seo, Y. C.; Lee, S. B. 2010. *Development of a CBR-Based construction duration estimation model for construction projects*: Dissertation. Architectural Institute of Korea.
- Lin, M. C.; Tseng, H. P.; Ho, S. P.; Young, D. L. 2011. Developing a construction-duration model based on a historical dataset

- for building project, *Journal of Civil Engineering and Management* 17(4): 529–539.
<https://doi.org/10.3846/13923730.2011.625641>
- Love, P. E. D.; Tse, R. Y. C.; Edwards, D. J. 2005. Time-cost relationships in Australian building construction projects, *Journal of Construction Engineering and Management* 131(2): 187–194.
[https://doi.org/10.1061/\(ASCE\)0733-9364\(2005\)131:2\(187\)](https://doi.org/10.1061/(ASCE)0733-9364(2005)131:2(187))
- Mačková, D.; Bašková, R. 2014. Applicability of Bromilow's time-cost model for residential projects in Slovakia, *Selected Scientific Papers – Journal of Civil Engineering* 9(2): 5–12.
<https://doi.org/10.2478/sspjce-2014-0011>
- Martin, J.; Burrows, T. K.; Pegg, I. 2006. Predicting construction duration of building projects”, in *XXIII FIG Congress “Shaping the Change”*, 8–13 October 2006, Munich, Germany.
- Owolabi, J. D.; Amusan, L. M.; Oloke, C. O.; Olusanya, O.; TunjiOlayeni, P.; Owolabi, D.; Peter, J.; Omuh, I. 2014. Causes and effect of delay on project construction delivery time, *International Journal of Education and Research* 2(4): 197–208.
- Peško, I.; Mučlenki, V.; Šešljija, M.; Radović, N.; Vujkov, A.; Bibić, D.; Krklješ, M. 2017. Estimation of costs and durations of construction of urban roads using ANN and SVM, *Complexity*, Article ID 2450370.
<https://doi.org/10.1155/2017/2450370>
- Ryu, H. G.; Kim, S. G.; Lee, H. S. 2006. A Competitive advantage analysis of construction duration through the comparison of actual data of domestic construction firms, *Korean Journal of Construction Engineering and Management* 7(1).
- Thomas, N.; Mak, M. Y.; Skitmore, M.; Lam, K. C.; Varnam, M. 2001. The predictive ability of Bromilow's time–cost model, *Construction Management and Economics* 19(2): 165–173.
<https://doi.org/10.1080/01446190150505090>
- Thomas, N.; Thomas, A. V. 2016. Regression modelling for prediction of construction cost and duration, *Applied Mechanics and Materials* 857: 195–199.
<https://doi.org/10.4028/www.scientific.net/AMM.857.195>