Study on Influencing Factors and Prediction of Construction Project Duration under Multiple Linear Regression Method

Jie Ju

Henan Vocational College of Applied Technology, Zhengzhou, Henan 45000, China

Keywords: Multiple Linear Regression, Project Duration, Influencing Factors, Forecast

Abstract: Multiple linear regression analysis prediction method refers to the method of establishing prediction model for prediction by correlation analysis of two or more independent variables and one dependent variable. When there is a linear relationship between the independent variable and the dependent variable, it is called multiple linear regression analysis. A complete construction project can be divided into the early decision-making stage, design stage, construction stage and maintenance stage, etc. The construction stage is concentrated in the construction activities of the engineering entity, with many participants, many influencing factors, high cost and long time, etc. Features are a key stage of the entire project. Accurately predicting the construction period is an important prerequisite for ensuring the smooth implementation of the entire project. It is the basis for formulating project plans, controlling investment budgets, managing and controlling projects, and resolving disputes.

1. Introduction

The construction project has the characteristics of large engineering volume, complicated structure, many construction links, long cycle, strict technical quality requirements, many related parties and many influencing factors {Chang, 2014 #16}. Duration and quality, cost, and safety have become the four most important control objectives in project management. The four control objectives are interrelated, and the construction period management is particularly critical under the premise of ensuring project quality and construction safety. It is of great greatly theoretical and practical significance to compare and analyze existing achievements, integrate innovation, and solve important problems in construction period management. In order to explore the impact of different factors on the construction period of construction projects, different scholars have made a lot of efforts from different angles {Eberly, 2007 #22}. The commonly used independent variables in the construction forecasting forecast model models include the construction period, the logarithm of the construction period, the construction speed and the logarithm of the construction speed, etc. The common commonly dependent variables include including total cost, total number of layers, type of bidding, construction use and management methods, etc. The research methods all adopt multiple linear regression method {Hong, 2011 #23}.

Regression analysis is one of the many methods used in data analysis. Regression analysis is a quantitative method that gives the variation between among variables. It not only provides the regression equation between among variables, but judges the validity of the established regression equation. Under the premise of the validity of the equation, the equation can be used for prediction and control, and the accuracy of prediction and control can be understood. The multiple regression analysis prediction method refers to the method of establishing a prediction model for prediction by correlation analysis of two or more independent variables and one dependent variable. When there is a linear relationship between the independent variable and the dependent variable, it is called multiple linear regression analysis.

  DOI: 10.25236/icepms.2018.228

2. Factors affecting the construction period of construction projects

2.1 Factors affecting the construction period and data sources

The construction period of the construction project is the time from the start of the project to the completion of the project and the handover to the user after the completion of the acceptance. Before the project starts, the construction unit needs to determine a reasonable construction period as the construction period is the basis for the implementation of construction schedule control. The traditional methods of determining the construction period can be divided into three categories according to the period quota, historical experience, network planning method, and expert system, as shown in Figure 1.

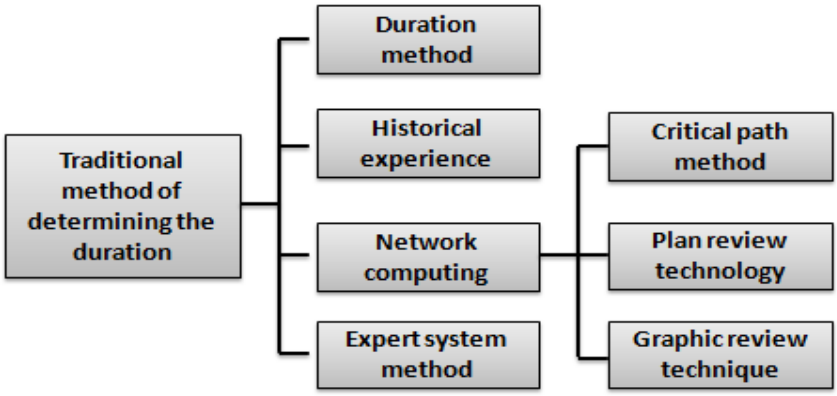

Figure.1 Traditional method of determining the duration

Most of the literature believes that project cost is the main factor affecting the construction period, but other factors affecting the project duration are divided. The domestic literature mainly studies the time of micro-process activities of the project, but has less research on the total project duration {Ming-ChiaoLin, 2011 #20}. Therefore, based on the Shanghai construction market data, this paper will summarize the influencing factors and prediction models of the construction period of domestic construction projects. The data comes from the completion of construction projects in Shanghai from 2008 to 2015. The road works, bridge works and reconstruction projects were removed. After the consolidation, 212 valid samples were obtained. To ensure traceability of data sources, each sample is numbered according to its date of completion. The sample types are divided into six categories according to the purpose of construction: office buildings, public buildings, teaching buildings, commercial buildings, hospital buildings, residential buildings, and basic information of various buildings, which are shown in Table 1.

Table.1 sample data overview of building construction projects

Sample type	Sample type	Sample cost (ten thousand yuan)	Overall floorage(㎡)	Total.number of floors
Office building	59	273~395059	1413~47208	3~59
Public building	8	6431~29871	12500~55825	2~15
Teaching building	33	771~13960	2685~44393	3~13
Commercial Building	25	376~105423	1511~108561	1~20
Hospital building	7	1512~47743	5091~83334	5~22
Residential building	80	233~52182	2017~131349	4~38

According to the results of the literature review, the factors affecting the construction period are

initially determined as total cost, total floor area, total number of floors, height, construction use and structure type, etc., while introducing some derivative factors, such as standard floor construction area and unit area cost. Variable types are classified into continuous, discrete, and subtypes sub-types {Moy, 2016 #17}. When analyzing the sample data, these influencing factors are used as independent variables of multiple linear regression analysis to explore the relationship between these independent variables and the dependent variables such as project duration and construction speed.

2.2 Establishment of the construction period prediction model

According to the sample data, the multivariate regression method is used to explore the influence of different factors on the project construction period and construction speed, and the variables with high correlation are selected to construct the project duration prediction model. According to the research objectives and literature review results, the candidate dependent variables of the construction forecasting model are formulated as the construction period T, the construction speed V, and the logarithmic construction speed lnV. According to Gauss-Markov Theorem, in the linear regression model, when the dependent variable obeys the normal distribution and the dependent variable satisfies the mutual irrelevance (independence), the mean meaning of the error term is 0 (zero mean), the variance of the error term is a fixed value (equal variance) and other assumptions, then the best unbiased linear estimate of the regression coefficient (BLUE) is the minimum variance estimate. The frequency frequent distribution of the dependent variable T, the construction speed V, and the logarithmic construction speed lnV are plotted as shown in Fig. 2. It can be seen that the sample frequency of the construction period T and the logarithmic construction speed lnV is approximately approximate in accordance with the normal distribution. The construction period T and the logarithmic construction speed lnV can be selected as the dependent variables of the construction period prediction regression model.

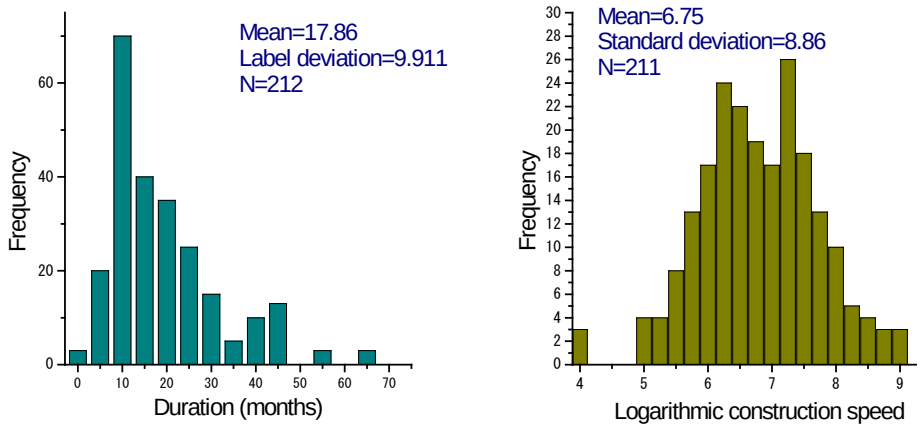

Figure.2 Frequency distribution of candidate dependent variables

3. Stability analysis of the project forecasting model

4. Table.2 Range of independent variables in stability analysis

Independent variable	Minimum value	Maximum	Average value
Total floor area A(㎡)	1412.60	472087	26386.30
Total number of layers N (layer)	1	59	11.8
Unit cost area Ca(Yuan/㎡)	982.34	13198.03	3236.65

After selecting the duration prediction model, it is necessary to evaluate its stability and effectiveness, and optimize its effect {Nathans, 2012 #18}. By changing the different values of the forecasting model variables, the stability of the model is evaluated. The residual analysis is used to diagnose the validity of the forecasting model. The residual scatter plot is used to eliminate the sample outlier optimization period prediction model, as shown in table.2.

In the model stability analysis, when one of the independent variables (such as the total building area A, the total number of floors N, the unit area cost Ca) changes and the other independent variables are fixed (recommended to take the average), respectively. The value of the construction speed V can be predicted when the variable changes. When the independent building area A, the total number of floors N, the unit area cost Ca and other independent variables change, the construction period prediction model is stable. The specific performance is as follows: (1) As the total construction area A increases, the construction speed V will increase simultaneously, but the increase will be smaller and smaller. (2) As the total number of layers N and the unit area cost Ca increase, the construction speed V will continue to decrease, and the reduction will be slower and slower. In a multiple linear regression equation, if an independent variable has a greater influence on the result of the regression equation, that is, the dependent variable, the regression coefficient is larger. The expression of different independent variables will have different effects on the regression equation. Of course, when the variables of the construction forecasting model are in a logarithmic form, the predicted value of the construction period speed exhibits an exponential change with the independent variable, rather than a linear change, which is fully expressed in the stability analysis curve.

5. Conclusion

Through multiple linear regression analysis of construction engineering data, this paper finds that the construction speed in a logarithmic form is the most suitable dependent variable for multivariate linear regression analysis of project duration. Through stepwise regression analysis and extensive comparison, the construction forecasting model finally considers three key continuous variables: total floor area, total floor area, and unit area cost, with and two important categorical variables: six different building uses and Five five different types of structures, etc., and give an intuitive explanation of the relationship between the duration and these factors. The multivariate linear regression analysis shows that the total construction area and total cost are important factors in the construction period of the corresponding construction project. The total cost can better reflect the change law of the construction period through the unit cost, and the construction speed can replace the project construction period more accurately. Reflecting the impact of complex construction performance on the construction period, the construction speed index also reflects the impact of the total construction area. Therefore, the total construction area and total cost are the fundamental factors affecting the construction project duration. The larger the total construction area, the more effective the use of various resources, the higher the management capacity and the faster construction speed. The research results also show that different construction uses and different structural types have a certain impact on the project construction period, but the impact is limited.

References

[1] Chang, J.G. and H.W. Li, The Study on Key Factors Influencing Time Delay of Public Construction Project. Advanced Materials Research, 2014. 912-914: p. 1838-1841.

[2] Eberly, L.E., Multiple linear regression. Methods in Molecular Biology, 2007. 404(2): p. 165.

[3] Hong, T., P. Wang, and H.L. Willis. A Naïve multiple linear regression benchmark for short term load forecasting. in IEEE Power and Energy Society General Meeting. 2011.

[4] Ming-ChiaoLin, et al., Developing a Construction-Duration Model Based on a Historical Dataset for Building Project. Statyba, 2011. 17(4): p. 529-539.

[5] Moy, R.L., L.S. Chen, and L.J. Kao, Multiple Linear Regression. World Scientific Book Chapters, 2016. 193(12): p. 364-368.

[6] Nathans, L.L., F.L. Oswald, and K. Nimon, Interpreting Multiple Linear Regression: A Guidebook of Variable Importance. Practical Assessment Research & Evaluation, 2012. 17: p. 19.