**International Journal of Engineering and Information Systems (IJEAIS)**
**ISSN: 2000-000X**
**Vol. 3 Issue 4, April – 2019, Pages: 14-20**

# The Development of Regression Models for Preliminary Prediction of Road Construction Duration

**Ibrahim Mahamid**

Associate professor, University of Prince Mugrin, Civil Engineering Department, Madinah,.
Email: imahamid@ymail.com

***Abstract:*** *This study aims at developing mathematical models for preliminary prediction of road construction duration using multiple regression techniques, based on 112 sets of data collected in the West Bank in Palestine. 5 regression models are developed to estimate the road construction duration in working days; 3 of them include bid quantities as input variables and 2 include road length and road width. The findings indicate that the coefficient of multiple determination (R square) of the developed models is ranging from 0.88 to 0.93 which confirms a good correlation between dependent and independent variables. The value of the mean absolute percentage error (MAPE) of the developed regression models is ranging from 19.1% to 31.4%. As the developed models are early phases ones, considerations were given to the fact that the input data could be easily extracted from sketches or scope definition of the projects.*

**Keywords:** regression, road construction, prediction, duration, preliminary models.

## 1. INTRODUCTION

"Construction duration is the number of days needed to perform the work required in the contract. This amount of time considers only the number of days actually worked and does not consider non-working days such as weekends, holidays, or weather impacted days" (Williams, 2008). Accurate and realistic duration estimates are important to every aspect of the construction project. Unreasonably short and long contract durations affect the project performance in a way or another. For example, unreasonably short contract might raise the bid price, increase disputes between construction parties, restrict qualified bidders from submitting bids, and lead to delay and cost overrun. Conversely, unreasonably long contract durations lead to inconvenience to the traveling public and encourage contractors with less experience and qualifications to submit a bid (FHWA 2002). Therefore, accurate estimate of construction duration in the early phases of construction projects is highly important and crucial for proper project planning. However, many difficulties arise when conducting duration estimate during the early phases which affect the accuracy and reliability of estimates. These difficulties could be: lack of design information, lack of productivity database, wrong estimation method, lack of estimator experience and many environmental, political, social and external uncertainties.

Preliminary duration estimates present the lowest expected accuracy; however they are often used by key people involved in the construction process. This study addresses the development of easy-to-use and reliable duration estimates during the early stages of road construction projects. It describes the development of duration prediction models using regression analysis. The estimating models are developed based on collected data for 112 awarded road construction projects in the West Bank in Palestine. The developed models rely on the conceptual design of the project and use only basic design technologies to provide fast and reliable results that can be very useful in the early stages of construction projects.

## 2. OBJECTIVES

The objectives of this study are:

- To develop preliminary duration estimates for road construction projects as a function of bid quantities.
- To develop preliminary duration estimates for road construction projects as a function of project size (i.e. road length and road width).

## 3. PREVIOUS STUDIES

Many studies have taken place in a number of construction industry sectors to develop prediction models of construction duration (Bromilow, 1980; Nkado, 1992; Chan and Kumaraswamy, 1995; Chan, 1999; Skitmore and Ng, 2001, Burrows et al., 2005; Williams, 2008; Mahamid, 2013). Work has been performed in the residential and building construction areas. However, the domestic highway construction industry has not seen the same level of attention (Williams, 2008).

Mahamid (2011) conducted a study to develop early cost estimating models for road construction projects using multiple regression techniques. The study based on 131 sets of data collected in the West Bank in Palestine. He developed 11 regression

estimates for cost of road construction project. Mahamid (2011) established his models based on bid quantities and project size. Williams (2008) conducted a study to develop mathematical models for preliminary prediction of highway construction duration. He identified the project factors which influence highway construction duration in the early stages. He also quantified the relationship between the duration-influential factors and highway construction duration. The quantity, magnitude, and sign of the factor coefficient yields evidence regarding the importance of the project factor to highway construction duration. In summary, Williams (2008) incorporates the duration-influential project factors and their relationship with highway construction duration into mathematical models which assist in the prediction of construction duration. Full and condensed models are presented for Full-Depth Section and Highway Improvement project types. He used statistical regression analysis to identify, quantify, and model these early-known, duration-influential project factors.

Abu Hammad et al. (2008) conducted a study to develop a prediction model for construction cost and duration in Jordan. Statistical regression models and sample tests are developed using real data of 140 projects. The prediction models were developed based on historic data of similar projects. Al-Momani (2000) developed a quantitative regression model for estimating the actual time using the data of 130 public building projects constructed during 1990-1997 in Jordan. The study also identified the most critical factors affecting projects delays. He concluded that the main causes of delay in construction projects are caused by: designers, owner changes, weather, differing site conditions, delays in material deliveries, economic conditions and increase in quantities.). Mahamid (2013) conducted a study aims at developing of regression models to predict the total cost of road construction project during the project early phases. 52 sets of data collected in Saudi Arabia were used to develop the models. Mahamid (2013) concluded 5 regression models; 2 of them include bid quantities of the major construction activities as independent variables and the other 3 include road length and road width as independent variables. Chan and Chan (2004) conducted a study in Hong Kong. They developed a benchmark model for project time performance in Hong Kong. In their study, they identified the critical factors influencing construction durations of high-rise public housing buildings in Hong Kong. Through their analysis, they determined five statistically significant regression variables: (1) total construction cost, (2) type of housing scheme, (3) use of pre-cast facades, (4) building volume, and (5) ground floor area per floor.

## 4. RESEARCH METHOD

The method approach in this research is as follow:

1) Define the objectives: The first step in the study is to define the objectives; the objective of the study is to develop simple mathematical models to predict the total duration of road construction projects as a function of bid quantities and project size.
2) Collect data: As estimating techniques require an extensive historical data, 112 sets of data for road construction projects were collected in the West Bank in Palestine. The data were collected from contracts awarded by some governmental agencies (i.e ministries and municipalities) who are the client of road projects. The data collected comprised of projects awarded over the years 2013-2018. Considerations taken to have approximately equal number of awarded projects over the years 2013-2018.
3) Select prediction method: Previous studies show different methods which are used to predict the project duration. As the objective of this study is to establish simple prediction models, regression analysis is used to develop the models. In general, regression models are effective due to a well-defined mathematical approach, as well as because of being able to explain the significance of each variable and relationship between independent variables (Sodikov, 2005). Regression models are intended to find the linear combination of independent variables which best correlates with dependent variables.
4) Propose model's variables: As the objective of this study is to develop estimating models that could be handled easily using calculator or simple computer programs in the early stages of project and due to the lack of information at the early stages of projects, the models were developed using variables that could be easily extracted from project scope and early sketches. The following variables are used: project duration in days (dependent variable), road length in meter (independent variable), road width in meter, earthwork quantity in m3 (independent variable), basecoarse quantity in m3 (independent variable), asphalt quantity in m2 (independent variable).
5) Develop the models: Once the variables to be included in the prediction model have been identified, a series of mathematical models were developed using multiple regression analysis techniques. Excel statistical tools were employed to perform regression analyses and to test the significance of the model. The regression equation can be expressed as:

$$Y = C + b_1X_1 + b_2X_2 + \ldots + b_nX_n \qquad \textit{Equation (1)}$$

Where,

C: regression constant.
$b_1$, $b_2$, …, $b_n$: regression estimates .
$X_1$, $X_2$, …, $X_n$: independent variables.
Y: dependent variable.

## 5. Results and Discussion

### 5.1 Duration as a function of project size

As the aim of this paper is to develop mathematical equations that easily can be used to estimate project duration at early stages of the project, an equation uses road length as the only input variable is formulated as shown in Table 1. The regression statistics analysis of variance test confirmed the statistical significance of the model at a significance level of 0.05 ($P < 0.05$). The results show good correlation between dependent and independent variable (The coefficient of multiple determination (R square) = 0.88). The formulated model is useful in estimating project duration at early stages of the project since the input data, road length, for the required regression equation could be easily extracted from scope definition of the project, and so the estimation could be achieved within minutes. The formulated model is:

$$Project\ duration\ (days) = -14.37 + 0.05x_1 \qquad model\ (1)$$

Where,
$x_1$ = Road length (m)

**Table 1: Regression model relates road construction duration with road length**

| *Regression Statistics* | |
|---|---|
| R Square | 0.88 |
| Adjusted R Square | 0.88 |
| Observations | 112 |
| F | 840.68 |

| *Variable* | *Coefficients* | *t Stat* | *P-value* |
|---|---|---|---|
| Intercept | -14.37 | -2.65 | 0.01 |
| Road length (m) | 0.05 | 28.99 | 0.00 |

Another mathematical equation that describes project duration as a function of road length and road width is developed. The coefficient of multiple determinations (R square) of the developed model is 0.90. It indicates a better correlation between dependent and independent variables than model 1. The regression statistics results for the developed equation are shown in Table 2. It shows that the analysis of variance test confirmed the statistical significance of the equation at a significance level of 0.05 ($P < 0.05$ for all variables). The developed model is:

$$Project\ duration\ (days) = -73.74 + 0.05x_1 + 10.27x_2 \qquad model\ (2)$$

Where,
$x_1$ = Road length (m)
$x_2$ = Road width (m)

**Table 2: Regression model relates road construction duration with road length and width**

| *Regression Statistics* | |
|---|---|
| R Square | 0.90 |
| Adjusted R Square | 0.89 |
| Observations | 112 |
| F | 472.24 |

| *Variable* | *Coefficients* | *t Stat* | *P-value* |
|---|---|---|---|
| Intercept | -73.74 | -4.26 | 0.00 |
| Road length (m) | 0.05 | 20.11 | 0.00 |
| Road width (m) | 10.27 | 3.59 | 0.00 |

It should be noticed that model 2 will give unrealistic results for small projects (i.e. for a project with road length = 1000 m and road width = 3m, the project duration equals 0). Therefore the model may give better results when road length is higher than 1000m and road width is higher than 5m.

A mathematical model with better correlation between dependent and independent variables is developed. The developed model uses variables interaction; road area which is a multiplication of road length and road width. The coefficient of multiple determination (R square) of the developed model = 0.92. The variance test confirmed the statistical significance of the model at a significance level of 0.05 as shown in Table 3. The model is as follow:

$$Project\ duration\ (days) = 14.48 + 0.005x_1x_2 \qquad model\ (3)$$

Where,
$x_1$ = Road length (m)
$x_2$ = Road width (m)

**Table 3: Regression model relates road construction duration with road area**

| *Regression Statistics* | |
|---|---|
| R Square | 0.92 |
| Adjusted R Square | 0.92 |
| Observations | 112 |
| F | 1291.95 |

| *Variable* | *Coefficients* | *t Stat* | *P-value* |
|---|---|---|---|
| Intercept | 14.48 | 3.77 | 0.00 |
| Road area ($m^2$) | 0.005 | 35.94369 | 0.00 |

### 5.2 Project duration as a function of bid quantities

The project duration as a function of bid quantities is developed. The developed model includes the quantities of major construction activities as independent variables, namely: earthwork ($m^3$), basecoarse ($m^3$), and asphalt ($m^2$). The coefficient of multiple determination (R square) of the developed model = 0.93, indicating a good correlation between input and output variables. The regression statistics shown in Table 4 confirmed the statistical significance of the model at a significance level of 0.05 (P-value less than 0.05). The developed mathematical model is:

$$Project\ duration\ (days) = 11.26 + 0.0047x_3 + 0.0038x_4 + 0.005x_5 \qquad model\ (4)$$

Where,
$x_3$ = Earthwork quantity ($m^3$)
$x_4$ = Basecoarse quantity ($m^3$)
$x_5$ = Asphalt quantity ($m^2$)

**Table 4: Regression model relates road construction duration with road construction activities**

| *Regression Statistics* | |
|---|---|
| R Square | 0.93 |
| Adjusted R Square | 0.92 |
| Observations | 112 |
| F | 451.36 |

| *Variable* | *Coefficients* | *t Stat* | *P-value* |
|---|---|---|---|
| Intercept | 11.26 | 2.73 | 0.01 |
| Earthwork ($m^3$) | 0.0047 | 4.78 | 0.00 |
| Base coarse ($m^3$) | 0.0038 | 2.77 | 0.01 |
| Asphalt ($m^2$) | 0.005 | 15.65 | 0.00 |

The correlation between independent variables included in model 4 is tested. The results show a good correlation between asphalt quantity and basecoarse quantity (R square = 0.76), and average correlation between asphalt quantity and earthwork quantity (R square = 0.52) as shown in Table 5. Therefore, a mathematical model describes the project duration as a function of asphalt quantity is developed. This model is helpful in the early stages of the project since the information needed can be easily extracted

from project scope and sketches. Table 6 indicates the statistical significance of the model at a significance level of 0.05 and shows a good correlation between project duration and asphalt quantity with "R square" value of 0.89. The developed model is:

$$Project\ duration\ (days) = 16.00 + 0.007x_5 \qquad model\ (5)$$

Where,
$x_5$ = Asphalt quantity ($m^2$)

**Table 5: Correlation between road construction activities**

| Correlation between: | R square |
|---|---|
| Asphaltt - Basecoarse | 0.76 |
| Asphalt – Earthwork | 0.52 |

**Table 6: Regression model relates road construction duration with Asphalt quantity**

| *Regression Statistics* | |
|---|---|
| R Square | 0.89 |
| Adjusted R Square | 0.89 |
| Observations | 112 |
| F | 868.17 |

| *Variable* | *Coefficients* | *t Stat* | *P-value* |
|---|---|---|---|
| Intercept | 16.00 | 3.48 | 0.0007 |
| Asphalt ($m^2$) | 0.007 | 29.46 | 0.00 |

### 5.3 Testing accuracy of the developed models

The accuracy for the developed mathematical models is tested using the mean absolute percentage error (MAPE). The following formula is used to compute the MAPE (Lowe et al., 2006):

$$MAPE = (1/n) * \sum_{i=1}^{n} \left| (A_i - F_i)/A_i \right| \qquad (5)$$

Where,
$A_i$ is the actual value
$F_i$ is the forecast value
n is number of fitted points

Table 7 shows the developed mathematical models for forecasting of duration of road construction projects. 5 regression models are developed in this study: 3 models are developed to estimate the duration of road construction project in working days as a function of project size (model 1, model 2, and model 3). The other 2 models predict the project duration as a function of project bid quantities, namely: asphalt quantity, basecoarse quantity, and earthwork quantity (model 4 and model 5).

It should be noticed that models 1, 2 and 3 are fit to be used at the early stages of the project since they are using only road length and road width as input variables which could be easily extracted from project scope. Later, when the BOQ is available, the models based on BOQ (model 4 and model 5) may be used.

**Table 7: Summary of the developed equations**

| Model No. | Mathematical model | Variables |
|---|---|---|
| 1 | *Project duration (days) = -14.37 + 0.05$x_1$* | $x_1$ = Road length (m) |
| 2 | *Project duration (days) = -73.74 + 0.05$x_1$ + 10.27$x_2$* | $x_1$ = Road length (m).<br>$x_2$ = Road width (m). |
| 3 | *Project duration (days) = 14.48 + 0.005$x_1x_2$* | $x_1$ = Road length (m).<br>$x_2$ = Road width (m). |
| 4 | *Project duration (days) = 11.26 +0 .0047$x_3$ + 0.0038$x_4$ + 0.005$x_5$* | $x_3$ = Earthwork quantity ($m^3$).<br>$x_4$ = Basecoarse quantity ($m^3$).<br>$x_5$ = Asphalt quantity ($m^2$). |
| 5 | *Project duration (days) = 16.00 + 0.007$x_5$* | $x_5$ = Asphalt quantity ($m^2$) |

Table 8 shows the MAPE resulting from using the mathematical models to estimate the duration of 112 data sets of road construction projects. The results show the following:

- The MAPE of the developed models is ranging from 19% to 31% which compare favorably with past researches which have shown that the estimate accuracy of the models used in the early stages of a project is between ±25% and ±50% (Lowe et al., 2006; Schexnayder et al, 2003; Sodikov, 2005).
- The MAPE for the models that use BOQ as independent variables is ranging from 19% to 23%. While the MAPE for the models that use project size (length and width) as independent variables is ranging from 27.6% to 31.4%.
- It can be seen that the models that use bid quantities as independent variables are more accurate than those using project size as independent variables. But it should be noticed that the bid quantities equations cannot be used in the conceptual phase of the project, while they could be used later when design is ready and quantities extracted from the drawings.
- For bid quantities models; using more independent variables resulted models with better accuracy and higher R square value.

**Table 8: Mean absolute percentage error (MAPE) and R square for the developed models**

| Model # | MAPE | R square |
|---|---|---|
| 1 | 27.6% | 0.88 |
| 2 | 31.4% | 0.90 |
| 3 | 28.8% | 0.92 |
| 4 | 23.4% | 0.93 |
| 5 | 19.1% | 0.89 |

## 6. CONCLUSION

This study aims at developing simple mathematical models, which could be handled at the early stages of the project by calculator, to forecast duration of road construction projects using regression analysis. The models were developed based on 112 set of data collected in the West Bank in Palestine. In this study, 5 mathematical models are formulated; 3 of them include road length, road width, and road area as independent variables, while the two remaining include bid quantities as independent variables. The coefficient of multiple determination (R square) of the developed models ranging from 0.88 to 0.93. This confirms the followings:

(1) good correlation between dependent and independent variables, (2) the predicted values from a forecast model fit with the real-life data.

The values of the mean absolute percentage error (MAPE) of the developed regression models are ranging from 19.1% to 31.4%. The results compare favorably with past researches which have shown that the estimate accuracy in the early stages models is between ±25% and ±50%. The findings reveal that the models that include bid quantities as independent variables are more accurate than those use project size, but they require more information.

Based on the study results, the following points are suggested to improve time estimating for road construction projects in Palestine: (1) Equipment and labor productivity should be updated continuously by the Government and industry associations, (2) To develop a well-organized and computerized database in a manner that enables estimators and researchers to use it easily in their estimates and studies.

## REFERENCES

1. Abu Hammad, A., Alhaj Ali, S., Sweis, G., and Bashir, A. (2008). Prediction Model for Construction Cost and Duration in Jordan. *Jordan Journal of Civil Engineering*, 2 (3).
2. Al-Momani, H. (2000). Construction Delay: A Quantitative Analysis. *International Journal of Project Management*, 18: 51-59.
3. Bromilow, F., Hinds, M., and Moody, N. (1980). AIQS Survey of Building Contract Time Performance. *The Building Economist*, 19(2), 79-82.
4. Burrows, T., Pegg, I., and Martin, J. (2005). Predicting Building Construction Duration. *AACE International Transactions*, PS.14.11-PS.14-19.
5. Chan, A. (1999). Modelling building durations in Hong Kong. *Construction Management and Economics*, 17, 189-196.
6. Chan, A. and Chan, D. (2004). Developing a benchmark model for project construction time performance in Hong Kong. *Building and Environment*, 39, 339-349.
7. Chan, D., and Kumaraswamy, M. (1995). A study of the factors affecting construction durations in Hong Kong. *Construction Management and Economics*, 13, 319-333.
8. Lowe, D., Emsley, M., and Harding, A. (2006). Predicting Construction Costs Using Multiple Regression Techniques. *ASCE Journal of Construction Engineering and Management*, 132(7), 750-758.
9. Mahamid, I. (2011). Early Cost Estimating for Road Construction Projects Using Multiple Regression Techniques. *Australasian Journal of Construction Economics and Building*, 11 (4), 78 – 101.
10. Mahamid, I. (2013). Conceptual Cost Estimate of Road Construction Projects in Saudi Arabia. *Jordan Journal of Civil Engineering*, 7 (3), 285 – 294.
11. Nkado, R. (1992). Construction time information system for the building industry. *Construction Management and Economics*, 10, 489-509.
12. Skitmore, R. and Ng, S. (2001). Australian project time-cost analysis: statistical analysis of intertemporal trends. *Construction Management and Economics*, 19, 455-458.
13. Williams, R. (2008). The Development of Mathematical Models for Preliminary Prediction of Highway Construction Duration. PhD thesis, Faculty of Virginia Polytechnic Institute and State University, Virginia, USA.