MULTIPLE LINEAR REGRESSION MODEL FOR SHIP REPAIR TIME ESTIMATION

Dr. Zobair Ibn Awal*, Ahammad Abdullah[2]

[1]Associate Professor, Department of Naval Architecture and Marine Engineering, Bangladesh University of Engineering and Technology, Dhaka.
[2]Department of Naval Architecture and Marine Engineering, Bangladesh University of Engineering and Technology, Dhaka.

*Email: namebuet07@gmail.com

ABSTRACT

Ship repair time indicates the duration of stay of a ship in a dockyard for maintenance services. Docking of ships for maintenance and repair are highly sophisticated works and simultaneously it is labor-intensive with strict time constraints. In this research, a mathematical model has been developed for estimating the repair time of cargo ships using multiple linear regression modeling techniques by the least square method. The paper aims to investigate the dependency among a wide range of parameters that are associated with ship repair time. In this particular model, ship repairing time is the function of the ship's gross tonnage, sandblasting, hull painting, structural steelworks, tank cleaning, valve works and hull cleaning works. The results obtained in this study suggest that some independent variables (e.g. amount of plate works) are more influential/dominating than the other variables (e.g. beam of the ship) in determining the ship repair time. It is believed that this model may support the management to estimate repairing time for providing better service.

Keywords: Mathematical model, least square method, repair time, linear regression.

1. INTRODUCTION

Ship repairing is a complicated engineering process involving extensive usage of human and natural resources that has high environmental and economic impacts on society. Repairing time is composed of two parts, the first part is the time inside the dock (docking time) and the second part is the time at the jetty side (berthing time). During docking time, the main concentration is on the bottom surveys and underwater jobs, which cannot be done in afloat conditions. During berthing time, all other jobs, except underwater jobs, are conducted. It is, however, a very complex management and technical challenge to place the vessel in and out of dock while multiple ships are waiting in a queue, and resources like the dry dock space are limited. It becomes a challenging chess game like a problem for the management to make critical decisions associated with estimating the individual ship repair time.

This paper describes the development of a multiple linear regression model for ship repair time estimation. The following sections of the paper presents a brief literature review on ship repair and describes the model development technique in detail.

2. LITERATURE REVIEW

Ship building and repair is perhaps one of the eldest industries known to human civilization. However, over years the process of ship repairing has evolved significantly with the advancement in industrialization. It is indeed surprising to note that although a vast amount of published literature is available on ship building, yet published work dedicated to ship repairing alone is rare. Perhaps, one of the most significant literature can be found in Butler [1] where the author documented the procedures for estimating the ship repair time in term of 'man-hour'. The author utilized his experience and documented evidences in deducing the required man-hour for various jobs associated with ship repair. The fundamental advantage of this literature of using man-hour is that it can be used even if there are changes in labor cost which varies from country to country. Also the author suggested modification of 'man-hour' for various weather and climate conditions. It is however, the only limitation of this literature is that calculation of man-hour is time consuming and highly dependent on expert judgment.

Nevertheless, over the years several research works have been published on the advancement of ship repair techniques with respect to optimization of the overall process. For example, a goal programming model for vessel dry-docking was developed [2]. The objective of the study was to balance the time that the vessel spends in dry dock and the corresponding costs concerning three goals: survey time, crashing cost and daily cost. The study revealed that the goal programming method can deliver several options for the management to choose for the best economic approach to ship repair.

Apostolicism *et al.* studied the dry-docking cost of four hundred and four tankers which were repaired in the Persian Gulf region over the time span of 4 years [3]. The study developed a mathematical relationship between dry-docking cost and the size of the ships, age of the ships, and repair time in days. Based of the study, the authors suggested incorporation of further independent variables (such as steel work, piping, mechanical work, and others) to increase the model accuracy.

Surjandari and Novita utilized Classification and Regression Tree (CART), a data mining approach, to estimate the duration of ship maintenance for dry docking [4]. The author developed a linear model to estimate the dry docking time by using three different types of repair works (propeller works, washing work and plate work) as input data The research showed that the model is prone to errors due to significant variation in data. However, the authors suggested that CART may predict ship repair duration with reasonable accuracy.

Dev and Saha explored and identified the possible independent variables responsible for influencing ship repairing time for crude oil tankers [5]. The study suggested the possible relationship among various variables in the form of a mathematical equation using multiple linear regression. The study concluded the ship repair time as a function of hull coating area only. The authors recommended that for future studies other types of ships may be incorporated in a single equation.

Another study by Dev and Saha highlighted the labor cost, which directly contributes to the man-hour or duration of ship repair works [6]. In this article, the authors have attempted to identify independent variables that influence ship repairing person-days (the dependent variable) and their interrelationship. The authors proposed a mathematical model for the estimation of the ship repairing labor cost.

The literature review on ship repair time estimation suggests that the advancements in this subject area are fairly new compared to the discipline of shipbuilding itself. Over the years, a few research works have been conducted on repair time estimation particularly focusing on specific ship types and specific region of the globe. A general methodology, overcoming various limitations that have been discussed in this literature review, is yet to be established. Hence, this paper attempts to develop a generalized, efficient and simple technique for ship repair time estimation.

3. THEORETICAL BACKGROUND

In most research problems where regression analysis is applied, more than one independent variable is needed in the regression model. The complexity of most scientific mechanisms is such that in order to be able to predict a critical response, a multiple regression model is needed [7]. This is precisely the case for ship repairing time estimation. It is quite difficult to relate the influence over one particular parameter (e.g. repair time) due to changes in various other kinds of parameters (e.g. engineering, economic, social, etc.). Similar problems have been dealt, in multiple disciplines of science and engineering, in the past using regression modelling. Regression modelling is a powerful yet simple tool for solving complex multidisciplinary problems with unknown interactions [8].

This research attempts to construct a multiple linear regression model using the least-square technique. The dependent variable parameter is the ship repair time, and the independent variables are gross tonnage, principle dimensions, various repair works, etc. The independent variables are elaborated later in this paper. The following sections describe the theoretical backgrounds of regression modelling.

3.1 The Regression Equation

The regression model for this study may be considered as the following:

$$y = f(x_1, x_2, x_3 \cdots, x_k) \quad (1)$$

Where, y is the ship repair time, k is the number independent variables and $x_1, x_2, x_3, \ldots, x_k$ are the independent variables (such as gross tonnage, principle dimensions, etc.

In case of n number of observations of ship repair time (y), which is the dependent-variable, there will be n times k number of independent-variable observations. These observations can be represented as following:

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki} \quad (2)$$

Where, $i = 1, 2, 3, \ldots\ldots n$ and $b_1, b_2, b_3, \ldots, b_k$ are the regression coefficients which need to be determined to construct the multiple linear regression model.

The multiple linear regression model can be expressed with a residual term as [7]:

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki} + e_i \quad (3)$$

This residual term (e_i) need to be as low as possible in order to have best prediction model. Therefore, the idea of using the least square techniques require the sum of squares of errors (SSE) to be minimum about the regression line. The expression of SSE is given as follows:

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} \cdots - b_k x_{ki})^2 \quad (4)$$

Differentiating SSE in turn with respect to $b_0, b_1, \ldots, b_k$ and equating to zero results a set of $k + 1$ number of normal equations for multiple linear regression. These equations can be solved for $b_0, b_1, \ldots, b_k$ by any appropriate method for solving systems of linear equations. Most statistical software can be used to obtain numerical solutions of the above equations.

3.2 Coefficient of Determination (R^2)

The coefficient of determination or the coefficient of multiple determination represents the portion of the variation in dependent variable (repair time) that is a function of a set of independent variable (repair works and various other parameters). According to Levine *et al.* (1997) the coefficient of multiple determination is expressed as following [9]:

$$R^2 = \frac{SSR}{SST} \quad (5)$$

Where SSR is the regression sum of squares, and SST is the total of SSR and SSE. SSR is equal to the sum of the squared differences between the predicted value of y and, the mean value of y. On the other hand, SSE is equal to the sum of the squared differences between the observed value of y and the predicted value of y [9].

In addition to R^2, another essential parameter regarding multiple linear regression is the adjusted R^2, which is utilized for comparing two or more regression model that predicts the same dependent variable. The adjusted R-squared increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor enhances the model by less than expected by chance [9] and [10]. The expression of adjusted R^2 is as follows:

$$R^2(adj) = 1 - \left[(1 - R^2) \frac{n-1}{n-k-1} \right] \quad (6)$$

Where,

R^2 = Coefficient of multiple determination
k = Number of variable
n = Number of sample

3.3 Mean square error (MSE)

The mean squared error (MSE) is a self-explanatory term used to indicate the deviation of observations from the mean. The errors with respect to the mean can be both be positive and negative; addition of errors, therefore, will not indicate the magnitude of deviation from the mean since positive and negative errors will continue to nullify each other. Thus, errors are squared to remove the negative and then added together. Finally, the sum is squarely rooted to obtain the MSE.

3.4 Mallow's C_P Statistic

Mallow's C_p statistic is used to assess the fit of a regression model that has been estimated using ordinary least squares. It is applied in the context of model selection, where a number of predictor variables are available for predicting some outcome, and the goal is to find the best model involving a subset of these predictors. A small value of C_p means that the model is relatively precise. The C_p statistic is computed as,

$$C_p = \frac{SS(RES)_p}{MSE_{all}} + 2(p+1) - n \quad (7)$$

Where, p is the number of independent variables included in a particular regression model. $SS(Res)_p$ is the residual sum of square from p number of variable subset. MSE_{all} is the mean square error including all independent variable. Finally, n is the number of sample [11].

3.5 Standard Error

The standard error is a measure of statistical accuracy of an estimate, equal to the standard deviation of the theoretical distribution of a large population of such estimates. In this study the following is considered

$$s = \sqrt{\frac{SSE}{n-k-1}} \quad (8)$$

The symbols have the usual meaning as explained above.

3.6 F-Statistics

The F-statistic is the test statistic for F-tests. In general, an F-statistic is a ratio of two quantities that are expected to be roughly equal under the null hypothesis, which produces an F-statistic of approximately. For passing the model *f* must be greater than $f_{critical}$.

$$f = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} = \frac{\frac{SSR}{k}}{s^2} \tag{9}$$

3.7 Variance Inflation Factor

Multiple linear regression assumes that there is no multicollinearity in the data. Multicollinearity occurs when the independent variables are too positively correlated with each other. When correlation exists among predictor's the standard error of predictors coefficients will increase and consequently, the variance of the predictor's coefficients is inflated. The VIF is a tool to measure and quantify how much the deviation is grown. VIF is calculated as follows:

$$VIF = \frac{1}{1-R^2} \tag{10}$$

Where, R^2 is the coefficient of determination for a regression model. If the value of VIF is 5 (five) or more then the independent variables are highly correlated with each other and it should be rejected from the regression model [12]

4. DATA COLLECTION AND RELEVECY CHECK

4.1 Data Collection

In this study the data on ship repair were collected from the Chittagong Dry Dock Limited (CDDL), Bangladesh which is one of the leading dry docks in Bangladesh. The dock is located near Chittagong Port and has direct access to the Indian Ocean through the Bay of Bengal. A brief introduction on the dock is presented in Table 1.

Table 1: Dockyard Characteristics

Item	Description
Location	Latitude: 22.269077 Longitude: 91.816474
Land Area	Factory area: 36.46 acres
	Residential area: 11.58 acres
	Ship repair area: 15.28 acres
	Jetty for outfitting: 342 m long
Dry Dock Capacity	Maximum ship size: 22,000 DWT
	Maximum length: 175 m
	Maximum breadth: 24 m

Data base was formed for this study considering 36 years (1980 to 2016) of ship repair data. The data contains various types of ship repair information including gross tonnage, repair time, principal particulars and others. A summary of the database is shown in Table 2. Also,

Table 2 shows the number of different types of ships that had been repaired at CDDL over the years.

Table 2: Brief summary of different types of ship repair times at CDDL (1980 – 2016).

Types of ship	Number of ships	Repair time (Day)		
		Min.	Avg.	Max.
Cargo vessel	191	3	19	57
Oil tanker	61	4	38	50
Fishing vessel	248	2	14	43
Barge	20	2	56	65
Tug boat	5	9	23	29
Frigate	61	2	19	51
Total	586			

The original dry-dock database contains information on various parameters of the ship. However, in this study the following nine parameters were considered:

- Gross tonnage of ship in tons
- Length of ship in meters
- Breadth of ship in meters
- Amount of plate repair work in tons
- Amount of sand blasting work in square meters
- Amount of painting work in square meters
- Amount of hull cleaning work in cubic meters
- Amount of tank cleaning work in cubic meters
- Number of valve replacements

4.2 Model Development Steps

This section discusses the model development steps. Figure 1 describes the process in a flowchart. At first, independent variables are chosen to include in the model. After that, the multi-collinearity of independent variables are checked by VIF values. Later the relevancy check between the dependent variable and all independent variable by the value of the correlation coefficient. Fourthly checking the linearity of selected independent variables with a scattered plot are examined. Later need to check the correlation among independent variables each other. Sixthly, performing the stepwise regression for finding the best subset of independent variables with the regression coefficient of the variables. Seventhly, choose the best model, considering the required parameters. Eighthly, complete the final analysis with the selected model (Residual analysis). Ninthly, transform the model if necessary to deal with the violation of linearity or other model assumptions. Finally, use the model for prediction and inference.

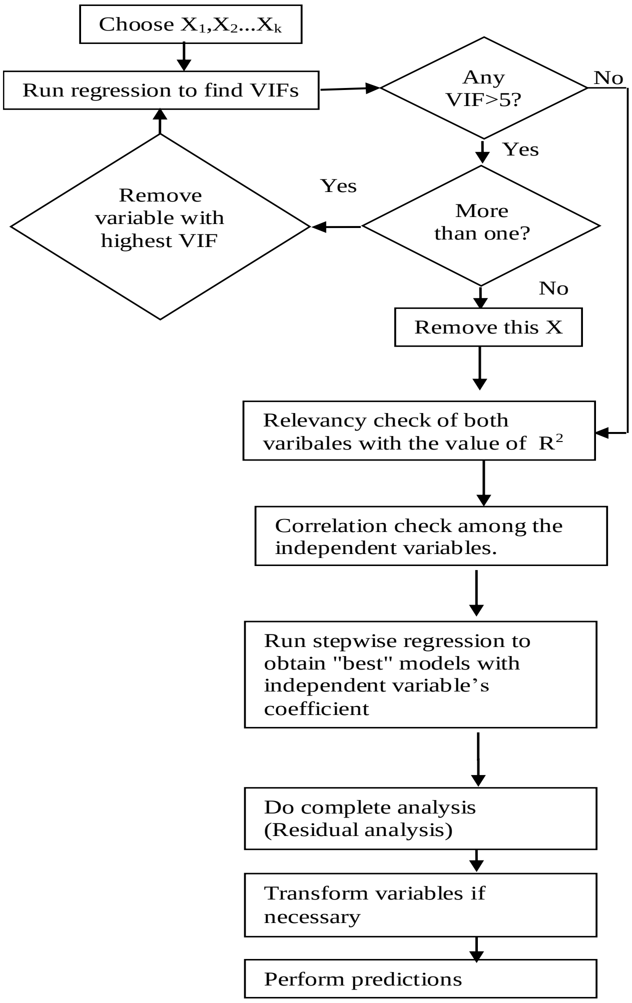

Figure 1: Model development flow-chart

According to the model development step, table 3 and table 4 represents the collinearity statistics (Variation Inflation Factor, VIF) of the independent variables. Table 3 illustrates the length and breadth of VIF values are 13.052 and 14.39, which cross the limit of VIF for multicollinearity.

Table 3: Representation of VIF values of independent variables

Independent Variable	Collinearity Statistics
	VIF (Including All Variable)
Length(L)	13.052
Breadth (B)	14.390
Plate work(PL)	1.895
Sand Blasting(SB)	3.095
Hull Cleaning (HC)	3.110
Tank Cleaning(TC)	2.285
Valve Works (V)	2.039
Painting Works(PA)	2.688
Gross Tonnage	1.822

In line with the flowchart (Remove Variable with highest VIF, figure 1), after removing the highest VIF variable (breadth), table 4 represents the collinearity statistics of independent Variable (VIF value), including and excluding the length variable.

Table 4: Representation of VIF values of independent variables

Independent Variable	Collinearity Statistics	
	VIF(Removing Breadth)	VIF (Removing Length)
Length(L)	2.049	
Plate work(PL)	1.860	1.451
Sand Blasting(SB)	2.701	1.806
Hull Cleaning (HC)	2.588	2.460
Tank Cleaning(TC)	2.149	2.145
Valve Works (V)	2.038	1.806
Painting Works(PA)	2.573	2.522
Gross Tonnage	1.822	1.601

4.3 Dependency Check

This section discusses the dependency check between the dependent variable and the

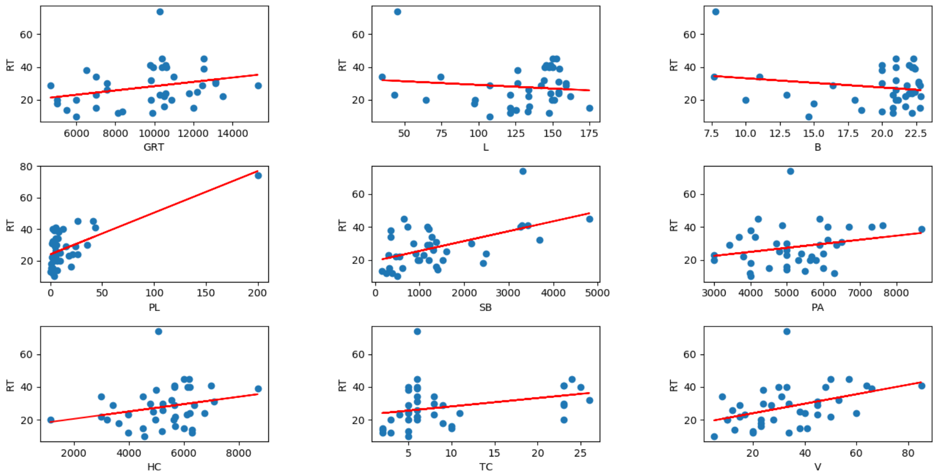

Figure 2: Linearity check of selected independent variables with scattered plot.

independent variables. Figure 2 shows the scatter plot of the dependent variable (repair time) versus independent variables. In this figure, length and breadth show the negative relationship with ship repairing time with a negative slope. According to this, length and breadth must be removed from the model. A summary of the correlation between the dependent variable and independent variables are shown in Table 5. Table 5 represents the correlation value of length and breadth with repair time is very low than other independent variables (scope of repair works). Compared to the value of correlation, length and breadth also must need to be eliminated from the model.

Table 5: Summary of relevancy check between dependent and independent variables.

Dependent Variable vs. Independent Variables			Correlation (r)
Repair Time	vs	Length	0.03
Repair Time	vs	Breadth	0.10
Repair Time	vs	GRT	0.26
Repair Time	vs	Plate Work	0.67
Repair Time	vs	Sand Blasting Work	0.58
Repair Time	vs	Valve Work	0.41
Repair Time	vs	Painting Work	0.24
Repair Time	vs	Tank Cleaning Work	0.30
Repair Time	vs	Hull Cleaning Work	0.24

4.4 Collinearity Check of Independent Variable

The length and breadth of the ship have been eliminated from the model equation. According to the model development flowchart, after removing two independent variables, check the collinearity of remaining independent variables. Table 6 shows the correlation value of remaining all independent variables with each other. According to table 6, most of the correlation values are a weak, weak, moderate range [13]. And these values also checked and passed the criteria of VIF previously, in line with table 4

Table 6: Summary of Collinearity check.

Pearson Correlation Values			
Gross Tonnage	vs	Plate Repair	0.160
Gross Tonnage	vs	Sand Blasting	0.119
Gross Tonnage	vs	Valve Replacements	0.464
Gross Tonnage	vs	Painting	0.422
Gross Tonnage	vs	Tank Cleaning	0.033
Gross Tonnage	vs	Hull Cleaning	0.496
Plate Repair	vs	Sand Blasting	0.376
Plate Repair	vs	Valve Replacements	0.096
Plate Repair	vs	Painting	-0.041
Plate Repair	vs	Tank Cleaning	-0.016
Plate Repair	vs	Hull Cleaning	0.049
Sand Blasting	vs	Valve Replacements	0.458
Sand Blasting	vs	Painting	0.288
Sand Blasting	vs	Tank Cleaning	0.664
Sand Blasting	vs	Hull Cleaning	0.241
Valve Replacements	vs	Painting	0.490
Valve Replacements	vs	Tank Cleaning	0.397
Valve Replacements	vs	Hull Cleaning	0.428
Painting	vs	Tank Cleaning	0.272
Painting	vs	Hull Cleaning	0.742
Tank Cleaning	vs	Hull Cleaning	0.205

4.5 Ship Repair Time Model

The ship repair time model is presented in this section. The statistical relationship between the dependent variable and independent variables are examined and established. The repair time is expressed as a function as shown in equation (10).

$$SR_{TIME} = f\left(R_{PL}, R_{SB}, R_{HC}, R_{GRT}, R_{TC}, R_{V}, R_{PA}\right) \quad (10)$$

Here, SR_{TIME} is ship repairing time, which is dependent on its independent variables. R_{PL} is the quantity of structural steel repair works. This is measured in terms of weight (ton). R_{SB} is the quantity of sandblasting, which is measured in terms of square meters. Sandblasting is used to remove paint and rust from the ship's hull (bottom, ship side, deck plate, etc.). R_{HC} is the quantity of hull cleaning repairing works measured in square meters. R_{GRT} gross tonnage is a measurement of a ship's total capacity expressed in volumetric tons of 100 cubic feet; it calculates by adding the under-deck tonnage and the internal volume of tween-decks and deck space used for cargo. R_{TC} is the quantity of tank waste cleaning repairing works such as mud cleaning. It is measured in terms of weight (ton) as well. R_V is the quantity of repairing the valve. It is counted in the number of valves replaced. R_{PA} is the quantity of painting works. This is measured in terms of area (square meters).

Based on the function as mentioned earlier, the following expression of the linear equation is constructed where the independent variables are linearly associated with the dependent variable:

$$SR_{TIME} = b_0 + b_1 R_{PL} + b_2 R_{SB} + b_3 R_{HC} + b_4 R_{GRT} + b_5 R_{TC} + b_6 R_V + b_7 R_{PA} \quad (11)$$

Where,

b_0, b_1, …, b_7 = Regression coefficients. Other symbols have the usual meaning

Simultaneous equations can be obtained by using the least square method [7], where the collected data of SR_{TIME}, R_{PL}, R_{SB}, R_{HC}, R_{GRT}, R_{TC}, R_V, and R_{PA} are given as input to calculate the regression coefficients. Subsequently, statistical testing verifies the accuracy of the model.

4.6 Best Fit Model

After checking the linearity, stepwise regression is performed to determine the best subset of the regression model. Stepwise regression is a long tiring process because all possible regression models are derived from all possible combinations of the candidate predictors. It can be a large number of possible models. In this study, finally, seven independent variables are considered. The possible combination of one variable, two variables, three variables, four variables, five variables, six variables, and seven independent variables is 2^7 in number. The best subsets are sorted out by considering the model selection criteria mentioned earlier among all the combinations. Finally, checking the F statistics test needs to justify the model. For every model, $f_{critical}$ value (collected from a table) is less than the calculated F statistics (f) value, which is also shown in table 7.

Lastly, completing the selection criteria of the mathematical model residual analysis [9] is done to verify the residual plot. The residual plot is done for each independent variable and predicted value against the residual value (observed value – predicted value). All the graphs show individually irregular shapes that depict that the model is ready to predict or estimate the repair time.

Table7: Different mathematical model with required parameters

SR (TIME) = f(PL, SB, HC, TC, V, PA, GRT)	SR (TIME) = f(PL, SB, HC, TC, V, PA)	SR (TIME) = f(PL, SB, TC, V, PA)	SR (TIME) = f(PL, TC, V, PA)	SR (TIME) = f(PL, TC, V)	SR (TIME) = f(PL,V)	SR (TIME) = f (PL)	Different Mathematical Model
9.659	10.290	10.228	10.075	15.063	15.843	23.91	a
0.248	0.250	0.250	0.258	0.225	0.250	0.263	PL
0.142	0.151	0.150	0.156	0.002	0.247	NA	V
0.279	0.273	0.273	0.328	NA	NA	NA	TC
1.0E-03	1.0E-03	1.0E-03	0.001	NA	NA	NA	PA
6.7E-04	6.2E-04	6.2E-04	NA	0.191	NA	NA	SB
-1.6E-04	-6.6E-05	NA	NA	NA	NA	NA	HC
1.4E-04	NA	NA	NA	NA	NA	NA	GRT
8.292	8.172	8.051	7.948	7.927	8.189	9.272	s
0.636	**0.636**	0.635	0.634	0.626	0.590	0.460	R^2
0.557	0.570	0.583	0.593	**0.595**	0.567	0.446	R^2_{adj}
68.760	66.792	64.831	63.182	**62.845**	67.065	85.982	MSE (residual)
8	6	4.1	2.2	0.9	2.1	11.5	C_P
8.007	9.607	11.877	15.205	20.112	26.637	32.413	f
2.31	2.39	2.49	2.64	2.87	3.25	4.1	$f_{critical(0.05)}$

5. DISCUSSION

In this study, during the multi-collinearity check, length and breadth have been eliminated because of the violation of the VIF criteria in line with table 5 and figure 2. Finally, six repair works and one principal parameter are considered independent variables to develop a different mathematical model. Table 7 represents the developed regression model with coefficients and the required statistical parameter considering the seven independent variables. This statistical parameter needs to justify the passing criteria of the model. One thing is mentionable; this table represents the best subsets which are sorted from 2^7 combinations.

It is easy to understand from the table 7, the inclusion of independent variable successively, the increasing the value of R^2 (multiple determination) which is the main criteria of choosing the best subset of the model simultaneously to justify the adequacy of the model. According to the highest value of adjusted R^2, lower value of mean square error (MSE), and minimum C_P value, the best subset model is SR (TIME)=f(PL, TC, V). But considering the highest value of R^2 with all repair works, selected the final model for estimating the repairing time is as follows:

$$SR_{(TIME)} = 10.290+0.250 *PL+6.2E\text{-}04*SB -6.6E\text{-}05*HC+ 0.273* TC+0.151* V+ 1.0E\text{-}03*PA$$

Between SR (TIME)=f(PL,SB,HC ,TC,V,PA) and SR (TIME)=f(PL,TC,V) model, the difference of the value of adjusted R^2 and mean square error (MSE) is not significant. Another thing for the last model [SR (TIME) = f(PL, SB, HC, TC, V, PA, GRT)] is that the value of R^2 is not improving instead of adding the variable GRT and its coefficient value also very small (1.4E-04). Now, it is clear again GRT has very less effect on repairing time.

Besides these theoretical explanations, some findings of the selected model need to discuss from a practical point of view. Sandblasting and painting works are dependable work for each other. Painting work is done after surface preparation (sandblasting) task. These two works take less time than other repair works because of the sequence of work procedures. It is also reflected in the developed model that the coefficient of these two parameters is 6.2E-04 and 1.0E-03, which is significantly less than other variables. Next for plate and valve, the work sequence is different. Plate works mainly include cutting the old plate, preparing the new plate with the template, fitting the new structure in the previous location, and finally, welding is done at least four times, including the inner and outer side of the frame.

There are different work procedures for valve repair. This work needs first to dismantle the valve from the location then opening the housing. After that, needs to check the inner side for any machining work is required or not. Finally, it becomes ready for pressure test. If it fails, then a new valve is replaced by the old one. For these two works, a long sequence of action is present, that’s why these take more time sandblasting and painting works. The coefficient of these two parameters is (0.250 and 0.151) very high than sandblasting and painting. In Chittagong Dry-Dock, the tank cleaning work is too laborious, and the maximum time it is done by people manually. For this limitation, it takes more time. Sometimes it becomes more difficult for enclosed place in the vessel.

Lastly, the remaining work is hull cleaning. In the developed model (regression equation), all the variables have positive signs except hull cleaning works (HC). Negative (equivalent to zero) impact of hull cleaning work is probably, the result of the fact that hull cleaning work consists of water jetting, light scraping, and finally, if required hard scrapping. If all these jobs include in the hull cleaning works, then it could have an intensive effect on the model, and model fitness value (R^2) will be increased. In this research, only the water jetting area has been considered for hull cleaning.

Pointing on the final regression equation, where the equation passes the F statistic test but the value of multiple determination is in medium range (0.636). This value has significance to justify the model that means 63.6% of the variation in the dependent variable is contributed to the change in independent variables, and remaining is called the error of estimation. This error of an estimate is the result of the absence of one or more influential independent variables responsible for the change in the dependent variable.

For this error estimation, there is some valid reason for this value of R^2. Firstly the variation of repairing works in different types of ships. Type of repair works varies from ship to ship except some typical repair jobs like sandblasting, painting, plate works, etc. For example, some ships have rudder and propulsion system works, but no plating, valve work, and piping works. Some ships have plating, valves, and piping works, but no rudder, propulsion system works. So, there are various combinations of repair categories and quantities.

Secondly, some repair jobs are a combination of two or three steps to do the whole process. If it is considered the entire process of work in one unit, then it will shorten the repairing time. Proper time will not include the developed formula. For example, tank coating work [5] is a combination of grit blasting, cleaning and painting work (touch-up and full of multiple coats). As a result, the effect of painting works is not accounted for in repairing time in the regression equation, which also resulted in lower regression coefficients and R^2.

Thirdly, if a more independent variable (scope of repair work) is considered, then the value of R^2 will be increased, which are shown in table 7. It follows that the lesser value of R^2 indicates the issue of an unsteady or weak relationship and make out of independent variable which influences the dependent variable.

Besides, this repairing time depends on types of docking (routine repair/ damage repair). Routine repair takes more time than damage repair. Regular repair time can be estimated, but damage repair cannot possible to predict.

Another thing is that some jobs are notified after docking of the ship that does not predict before docking. These extra jobs may longer the repair time than the estimated earlier repair time.

However, despite the limitations mentioned, this mathematical model can be useful to the ship's authority, and the Dry-dock can use as a guiding tool to determine the expected ship repairing time against a set of the vessel's particulars and selected scope of repair works for a cargo vessel. This estimated repairing time, which can be translated into total person-days and finally into labor cost for budgetary purposes

6. CONCLUSIONS

This research paper analyzes and picks out the most common, important independent variables which mainly dominate the ship repairing time for selected (cargo vessel) ship. This article also makes out a possible strong relationship between dependent and independent variables in the form of a mathematical equation by using the multiple linear regression, and this equation is verified with required statistical parameters to justify the adequacy of the model for the system. The condition $f > f\alpha$ suggests the rejection of the null hypothesis. It is summarized that there is a convincing amount of variation in their response (the dependent variable) as a result of the differences in independent variables in the developed model. This mathematical model can be used in estimating the repairing time before docking of a vessel based on the independent variable (scope of repair work)

This study is only for a cargo vessel. According to this way the further research is possible to develop a similar type of mathematical model with more sample data for different types of ships and then developing a comparative statement (influencing repair category, quantity, etc.) of the repair time for a different vessel. Not only finding the repair time but also further study could be for detailed working hour estimation for a specific ship or a general formula for all types of vessels.

ACKNOWLEDGEMENTS

The authors would like to thank Chittagong Dry-Dock limited, Bangladesh, for giving the data and other required information, which also helped immensely in this research. The authors also acknowledge the technical support from received from BUET while conducting the research.

REFERENCES

[1] Butler, D. (2000) *A Guide to Ship Repair Estimates in Man-hour*s. Butterworth-Heinemann. ISBN 0-7506-4834-1.

[2] Cristóbal, J. R. S. (2009) A Goal Programming Model for Vessel Dry Docking. *Journal of Ship Production*, 25(2), pp.95-98.

[3] Apostolidis, A., Kokarakis, J. and Merikas, A. (2012) Modeling the Dry-Docking Cost-The Case of Tankers. *Journal of Ship Production and Design*, 28(3), pp.134-143.

[4] Surjandari, I. and Novita, R. (2013) Estimation Model of Dry Docking Duration Using Data Mining. *International Journal of Industrial and Manufacturing Engineering*, 7(7), pp. 1588-1591.

[5] Dev, A. K. and Saha, M. (2015) Modeling and analysis of ship repairing time. *Journal of Ship Production and Design*, 31(2), pp. 129-136.

[6] Dev, A. K. and Saha, M. (2016) Modeling and Analysis of Ship Repairing La*bor*. *Journal of Ship Production and Design,* 32(04), pp. 258-271.

[7] Walpole, R.E., Myers, R.H., Myers, S. and Ye, K. (2012) *Probability & Statistics for Engineers & Scientists*. Prentice Hall. ISBN 10: 0-321-62911-6.

[8] Draper, N.R. and Smith, H. (1998) *Applied Regression Analysis* (Vol. 326). John Wiley & Sons. ISBN 0-471-170482-8.

[9] Levine, D., Stephan, D., Krehbiel, T. and Berenson, M., 2005 *Statistics for Managers using Microsoft Excel,* Prentice Hall. ISBN 0-536-04080 X.

[10] Mendenhall, W., Beaver, R.J. and Beaver, B.M., (2013) *Introduction to Probability and Statistics*. Cengage Learning. ISBN-13: 978-1-133-10375-2.

[11] Rawlings, J.O., Pantula, S.G. and Dickey, D.A., (1998) *Applied regression analysis: a research tool*. Springer Science & Business Media. ISBN 0-387-98454-2.

[12] Jamal I. Daoud (2017) Multicollinearity and Regression Analysis, *ICMAE(2017) Journal of Physics*: Conf. Series 949 (2017) 012009

[13] Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Pacific Grove, CA: Brooks/Cole Publishing.