

In a study to examine the relationship between the time required to complete a construction project and several pertinent independent variables, an analyst compiled a list of three variables that might be useful in predicting the time to completion. These three variables were size of the contract, x_1 (in \$1000 unit), number of workdays adversely affected by the weather, x_2 , and the number of subcontractors involved in the project, x_3 . Ten construction projects were randomly chosen, and each of the three variables as well as the time to completion were measured. It was realized that the previous 10 projects were completed when there was no workers' strike. Knowing that workers' strike may adversely affect the timeline of the projects, an additional eight construction projects were also selected randomly where workers' strike was reported during the construction phase. The data is presented in the following table:

Obs	x_1	x_2	x_3	y
During no workers' strike				
1	60	7	7	29
2	80	10	8	15
3	50	14	5	10
4	50	4	3	15
5	75	5	6	30
6	750	10	10	45
7	70	5	3	7
8	80	3	6	21
9	300	8	8	28
10	110	7	4	30
During workers' strike				
11	100	8	10	60
12	200	12	11	70
13	500	15	12	75
14	1200	20	12	90
15	250	8	12	62
16	400	14	14	72
17	800	17	12	82
18	2600	14	13	50

OBJECTIVES OF THE PROJECT

1. There an impact of workers' strike on construction project completion time? Perform a relevant statistical analysis and comment on the same.
2. Is there any association between various predictor variables and the time required to complete a construction project? Also present a descriptive summary of the data used in this analysis.
3. Fit the first-order model with four predictors (including the workers' strike/no strike as one of the variables) and assess its utility. Do all the predictors appear to be important? Does the sign of the coefficient of x_1 make sense? How much variability in y is not explained by the model?
4. Since the size of the contract, x_1 contains relatively large range of data, it was suggested that x_1 should be natural-log transformed and the model should be re-run. How does the new model perform compared to the previous one? Check the residual plot. Do you find any observation(s) having extremely large residual? What does that suggest?
5. It was realized that the observation 18 in the dataset was incorrect and cannot be trusted for the size of the contract as well as the time required for completion. Hence, re-run the model (in 4, above), excluding this observation and assess its utility. Write the model expression and estimate the time required to complete a construction project during non-strike period when $x_1 = \$65,000$, $x_2 = 7$ days, and $x_3 = 5$.
6. Based on your analysis of this data so far, which predictor variables would you must have to predict the construction project completion time? Write the model expression and state how much useful this model will be for prediction.