Identification of Factors Driving Crew Production Rate : Methodology and Application

허 영 기*
Huh, Youngki

요 약

For accurate construction contract time estimation, few parameters are more significant than crew production rates and factors affecting the rates. However, statistical analysis techniques for finding such factors are not always simple mainly because there are many factors and the interaction between factors is not well quantitatively understood. This paper presents methodology of identifying factors driving crew production rates. The methodology is further demonstrated with representative data collected by the author from 13 on-going highway constructions. Three factors were identified as statistically significant drivers of Cap crew production rate: 'Cap Size (m3/ea)'; 'Cap Length (m)'; and 'Cap Shape (Rectangle vs. Inverted 'T')'. It was also found that the production rates are best explained by a multiple regression model with two of the drivers; 'Cap Size' and 'Cap Shape'.

키워드 : Factor, Quantitative Analysis, Crew Production Rate, Time Determination

1. Introduction

There are numerous factors reported to affect construction labor productivity. Thomas and his colleagues (1989) listed 42 factors under three categories. Rifat (1996) also summarized 23 similar factors under three categories: Management related; Project related; and Labor related. Furthermore many researchers studied in different ways how those factors affect the productivity and how those are interpreted and/or used for better construction project management. Herbsman and Ellis (1995) found 17 factors affecting overall construction duration of a transportation facility project from a survey: weather and seasonal effects; location of a project; traffic impacts; relocation of construction utility; type of project; letting time; special items; night and weekend work; dominant activities; environmental; material delivery time; conflicting construction operation; permits; waiting & delay time; budget & contract payment control; and legal aspects.

One of problems in such studies is that there are many factors and the interaction between factors is not well quantitatively understood. It is mainly because, to a certain extent, there is no such data that can be analyzed with appropriate statistical confidence in order to find out the relations quantitatively. Another reason might be methodology of its analysis. Analysis techniques are not always simple. Finding proper analysis approach is actually still a topic of ongoing research in construction industry (Sanders et al. 1989).

This paper presents methodology of identifying factors driving crew production rates that is defied as 'Total output per unit time, produced by a crew'. The methodology will also be demonstrated with one of major concrete bridge activities, namely Cap1. The data used for the demonstration was collected by the author from 13 on-going highway projects between February, 2002 and May, 2004 in Texas, USA.

2. Literature Review

This section reviews productivity data analysis methods that have been applied in the industry. In the study of construction productivity, regression analysis, factor models, neural network, and expert systems have been commonly used for data analyses. It is not the intention of this section to provide a theoretical background of such

* 학생회원, 미국 텍사스대, 건설경영학, 박사후 과정

1 The structure on top of column(s) supporting beams.

Table 1. Summary of Studies that Employed Regression Analysis

Author	Year	Title	Type of project/activity
Awad S. Hanna, Pehr Peterson, and Min-Jae Lee	2002	Benchmarking Productivity Indicators for Electrical/Mechanical Projects	28 mechanical and 31 electrical projects
Simon D. Smith	1999	Earthmoving Productivity Estimation Using Linear Regression Techniques	Highway
H. Randolph Thomas and Ivica Zavrki	1999	Construction Baseline Productivity: Theory and Practice	Masonry-23 projects, concrete formwork-8 projects, structure steel erection-12 projects
H. Randolph Thomas and Karl A. Raynar	1997	Scheduled Overtime and Labor Productivity: Quantitative Analysis	Industrial projects: pipe fitting, electricians
Peter F. Kaming, Paul O. Olomolaiye, Gary D. Holt, and Frank C. Harris	1997	Regional Comparison of Indonesian Construction Productivity	Buildings: masonry, carpenter
Steve R. Sanders and Thomas, H. R.	1993	Masonry Productivity Forecasting Model	Commercial projects: masonry
Thomas, H. R., Mathews, C. T., and Ward, J. G.	1986	Learning Curves Models of construction productivity	Housing/ Bridges: frame erection, masonry, formwork
Enno Koehn and Gerald Brown	1985	Climatic Effects on Construction	Excavation, erection, masonry, electrical, carpenter, labor

analysis methods, but to provide an overview of what methods and how they have been applied to what types of productivity data in the industry. (Readers may find detailed theoretical information from references.)

2.1 Regression Analysis and Factor Model

Modeling construction productivity requires quantification of both the effect of each factor on productivity and the interactions among factors. Regression analysis is one of the common methods employed by researchers in exploring such relationships, as summarized in Table 1. Koehn and Brown (1985) developed two nonlinear regression model equations; cold or cool - hot or warm as a function of temperature and humidity, based on 172 data points obtained from previous publications. Sanders and Thomas (1993) developed a regression model to predict masonry productivity. Regression techniques were also employed to quantify effects of overtime work on productivity by Thomas and Raynar (1997). Smith (1999) estimated earthmoving productivity by means of a multiple regression equation. Hanna (Hanna et al. 2002) studied the relationships between project size and duration, average manpower, and peak man power, and estimated S-curve(x; percent time of total duration, y; percent hours of total work hours).

The Factor model, that is, a multivariant approach for modeling of crew-level productivity, was introduced to predict productivity by Sanders and Thomas (Thomas and Yiakoumis 1987; Thomas et al. 1990). The underlying theory of this model is based on factors identified as affecting productivity at the crew-level. According to Thomas, factors can be individually isolated to quantify the effects of each factor on productivity. They developed a Factor model and proved that the model can explain 87.7% of the variability in masonry productivity. Thomas also used a Factor model developed from masonry projects to forecast labor productivity at completion (Thomas et al. 1994).

2.2 Neural Network

Although the regression analysis has been a common tool in construction productivity studies, many researchers have applied alternative methods such as the Neural Network (NN) model. Among several kinds of NN systems, a feed-forward back-propagation network with a limited number of inputs is a typical network employed by researchers in the construction industry (Portas and AbouRizk 1997).

Chao and Skibniewski (1994) presented two feed-forward, multilayered NN models to predict excavator productivity; one for estimating excavator capacity based on job conditions and the other for estimating excavator efficiency based on the attributes of operation elements. Simulating 786 excavator cycles (9.67~30.99 seconds per cycle) they demonstrated that NN was feasible for productivity estimation. However, since the data was drawn from computer simulated excavator under controlled conditions, it hardly reflected operator skills, weather, and various constraints. Portas and AbouRizk (1997) constructed a three-layered feed-forward, back-propagation NN model with a fuzzy output structure to estimate productivity of concrete formwork. A case study showed that the model could predict average labor productivity of formwork activity within 5% accuracy. However, more extensive data collection should be pursued for applications in real life, as the authors stated that 50 sets of data points may be inadequate for just proper training of a NN model. AbouRizk and Hermann (2000) developed a probability inference neural network model and compared it with a feed-forward back-propagation neural network. They obtained 119 historical data points from three different site activities on 66 plant projects. The model incorporated 36 factors ranging from project- to detailed activity-levels and predicted a point value of production rate. Although the study showed that NN models could deal probabilistically with high dimensional input-output mapping with multiple influential factors, its concept may be too complex to be accepted by industry practitioners.

Sonmez and Rowings (1998) compared NN models with regression. They used 112 weekly data points of concrete pouring, 76 of formwork, and 46 of concrete finishing obtained from completed projects' documentation, and showed that NN models had the potential to quantify the effects of complex multiple factors. AbouRizk and Hermann (2001) also constructed a NN model with data obtained from 27 completed projects involving 39 pipe installation activities. However, the data heavily relied on survey-type input.

NN provides an effective tool for complex estimation problems, such as labor production rates, where relationships between inputs and outputs can not be represented by mathematical functions (Portas 1996). There are, however, some limitations to its application to construction productivity studies. First of all, a NN model needs high quality data. Data of poor quality or an insufficient dataset could result in false predictions. As such, researchers have focused on the applicability of neural networks rather than developing actual models that can be applied in the construction industry. Secondly, unlike regression models, NN do not generally require users to establish the class of relationships between factors and production rates as a controlling algorithm accomplishes this relationship (Sonmez 1996). Perhaps justifiably so, industry practitioners tend not to accept and apply a model if its computational concepts are not thoroughly understood.

2.3 Expert System

Expert systems were first used in construction productivity to predict activity duration and productivity for masonry construction (Hendrickson et al. 1987). The expert system, called MASON, was developed based on interviews with one professional mason and one supporting laborer, but was not validated by any technique. Christian and Hachey (1995) also attempted to predict production rates for concrete placement by using an expert system. They developed the system based on 11 days of concrete placement data from 11 projects and found that concrete truck delays were the main reason for unproductive operations. However, there was significant variation in data sources, and sample size was limited (Sonmez 1996).

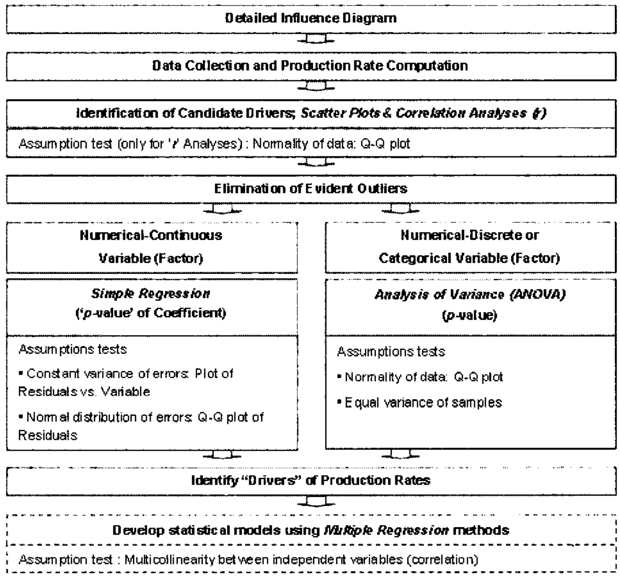


Figure. 1 Methodlogy of Driver Identification

3. Methodology of Driver Identification

Data analyses and results should be presented in a way that industry practitioners can understand them without difficulty. Research experiences have suggested that the construction industry tends to be reluctant to make decisions based upon data unless the organizations understand and are confident about the process of data analysis (Huh 2004). Fig. 1 shows the overall process of applying statistical methods in identifying drivers, statistically significant factors, of production rate. The methodology was developed based on numerous discussions with industry experts and modified by testing with representative data in order to make it the most practical. Unlike some other complex methods, it is easy yet reliable enough to be used in real projects. The detailed discussion on the process of Data collection is excluded in this paper and can be found in the reference.

3.1 Influence Diagram and Data Collection

Influence Diagram is useful tool to identify various possible factors and their interrelationships. A diagram can provide a great opportunity of brain-storming about factors at the early stage of study. It also needs to be refined throughout literature review and data collection for better results, which will lead to collection of all useful information pertaining to various factors and events.

3.2 Correlation Analyses and Scatter Plots

Correlation analysis is to determine if there is a significant linear relationship between a factor and production rates data. While regression, in general, is used to build and test a prediction model, correlation analysis is often applied to test the direction and strength of their linear relationship. The correlation coefficient, r, is a unit-less value that always falls between -1 and +1. A coefficient close to 1 indicates a strong positive linear relationship between the two random variables. A key assumption underlying the analysis is that the data is normally distributed. The results of the analysis should be interpreted with an understanding of how well the data meets this assumption. The assumption can be tested by visually inspecting a 'Q-Q plot', which is believed to be effective enough for real world data (Thomas 2003). It plots the standardized values of the data set versus values that would be expected if the data were perfectly normally distributed. If data tends to be normally distributed, values will fall on or near a straight line (Albright et al. 2003; Thomas 2003). Visual inspection on scatter plot can give sense of production rate data points spreads by a given factor, hence to check if linear relationship is appropriate and/or if the data represents more than one population (Huh 2004).

3.3 Analysis of Variance (ANOVA)

The one-way ANOVA is to test if there are significant differences among population means of more than two groups of data. The null hypothesis (Ho) for this test is that the population means of the groups are all equal, suggesting different groups of data have been drawn from some common population and the observed differences in sample means are attributable to chance fluctuations. On the other hand, the alternative hypothesis (Ha) is that at least one mean of the groups is different from others (Albright 2002). The null hypothesis of equal means is tested to see if the resulting p-value is sufficiently small. A p-value is the probability of wrongly rejecting the null hypothesis if it is in fact true. Hence, if the p-value is small enough, the alternative hypothesis can be accepted, which leads to the conclusion that some of the population means are different (Thomas 2003).

There are two underlying assumptions for this test: (1) the normality of population from which the observed data has been drawn, and (2) the equal variance of populations. Likewise, for the correlation analysis, the assumptions should be checked for gross violations whenever possible, although the assumptions are never satisfied exactly in any application (Albright et al. 2003). The normality assumption can be tested by examining Q-Q plots of the observed data. For the equal variance, if the largest standard deviation is less than twice the smallest standard deviation, equal variance can be assumed (Thomas 2003).

The significance level, that is, α level; probability of wrongly rejecting the null hypothesis that the researcher is willing to accept, used in this study for the one-way ANOVA test is 0.1. If the p-value is smaller than this significance level, then the null hypothesis is rejected. Thus it can be concluded that some of the population means are significantly different.

3.4 Regression Analysis

The simple regression method is to test the statistical significance of continuous numerical candidate drivers at a given significant level. The value of R^2, coefficient of determination, for a regression model, indicates how well the observed data fit into the predicted model. When R^2 is reasonably high enough, the research hypotheses can be tested by comparing the p-value of regression coefficients with the statistically significant α value, that is, 0.1 in this study. If p-value of a regression coefficient is larger than a given α value, it can be concluded that the effect of the independent variable (factor) on dependent variable (production rate) is not statistically significant at the given significance level α (Wonnacott et al. 1986; Sakamoto 2002). Three assumptions to be tested for better interpretation of a simple regression model: Linear model is appropriate; Constant variance of errors; Normal distribution of errors.

Multiple regression is a widely accepted statistical application to build a prediction model. While the simple regression and correlation analysis examine the relationship between two variables only, the multiple regression shows the net effect of each independent variable under controlling all other variables. There is one additional assumption to be tested for a multiple regression model:

Independent variables are not highly correlated. Two things have to be denoted for better application of a multiple regression analysis. First, the number of data points collected and used for developing models has to be at least 30 data points, which is believed to be enough to develop a prediction model with two independent variables, as Stevens (2002) stated that "for social science research,

Table 2. Scope Description for Production Rate Computation

	Scope – Included	Scope – Not Included
Cap	False work, if any / Installation of forms and rebar / Inspection of forms and rebar / Handling and placing of concrete	Site preparation / Preparation of rebar and forms / Rebar fabrication / Curing / All necessary work for the protection of concrete placed under any weather conditions / Removal of forms / Finishing of structure surface / Installation of drainage pipe / Removal of false work / Non – metal form

about 15 subjects (data points) per predictor (independentvariable) are needed for a reliable equation." Secondly, more complicated regression model may need to be developed for better prediction by applying relatively complex statistical methodology such as transformations of data. Those complicated prediction model, however, is very unlikely accepted by industry, in particular, for determining highway bridge construction time that is too much variable to be modeled into an equation. Despite the limitations, the multiple regression is a useful tool for understanding the relationships between production rates and factors.

4. Methodology Application: Cap of Highway Concrete Bridges

4.1 Influence Diagram

Influence diagrams were used to identify factors that are believed to be affecting Cap production rate (Huh 2004). Potential factors were first identified based on an intensive literature review and discussions with Texas Department of Transportation (TxDOT) personnel. Then such factors were refined through the preliminary data collection process, which involved observations, interviews with site personnel, and data analyses. The diagram showing Work item-related factors is presented in Appendix A. Some other diagrams can be found in the reference.

4.2 Data Collection and Production Rates Computation

Based on standardized data collection tool developed, data was acquired from 13 ongoing highway projects between February, 2002 and May, 2004 (Huh 2004). The projects were all located in Texas and had a range of cost between 1 to 261 million dollars. A total of 40 crew production rates (data points) were computed by predefined formula of 'Total Crew Work Days / Total Number of Output (EA)'. The mean of the data was five (crew days/ea) with standard deviation of 2.5. Detailed data collection process can be found from the reference. Scope included and not-include in the production rates computation is presented in Table 2.

4.3 Identification of Candidate Drivers

1) Scatter Plots and Correlation Analysis

Normality test of visual observation on Q-Q plot revealed that the sample is normally distributed. From observation on various scatter plots, three candidate drivers

Table 3. Correlation Matrix; Production Rate vs. Candidate Drivers

	Production Rate	Cap Size (m^3)	Cap Length (m)
Production Rate	1.000	0.716	0.707
Cap Size (m^3)	0.716	1.000	0.807
Cap Length (m)	0.707-	0.807	1.000

Table 4. ANOVA Results; Production Rate vs. Cap Shape

Categories	Descriptive Statistics			ANOVA Results
	N	Mean	Std. Deviation	Sig. (p-value)
Rectangle	22	4.3	2.0	0.026
Inverted 'T'	17	6.0	2.8	

were selected: Cap Size (m3); Cap Length (m); and Cap Shape (Rectangle versus Inverted 'T'). As shown in Table 3 and Table 4, those three are very likely to be the drivers of crew production rate. The table also shows that 'Cap Size' and 'Cap Length' is correlated.

2) Simple Regression Analyses and ANOVA

To test the significance of the candidate drivers, each continuous numerical variable was regressed on production rate, and the results are shown in Fig. 2 and Fig. 3. The middle straight line in the graphs is the linear regression line and the curved two lines represent the 95% confidence interval of the predicted production rate values. For the analysis of the discrete variable, 'Shape of Cap', ANOVA was employed (see Table 4). The results shows that the mean difference is statistically significant as Significant level of 0.026 is smaller than 0.1. Testing all the assumptions needed for the analyses showed that they were met to a reasonable extent.

3) Multiple Regression Analyses

Three multiple regression models were developed with the three drivers and it was found that the production rates are best explained by a multiple regression model with two of the drivers; 'Cap Size' and 'Cap Shape,' giving R^2 of 0.56.

In the above statistical analyses,

- The factor of 'Cap Size (m^3/ea)' has statistically significant impact on Cap crew production rate at a level of 0.1.
- The factor of 'Cap Length (m)' has statistically significant impact on Cap crew production rate at a level of 0.1.
- The factor of 'Cap Shape (Rectangle versus Inverted 'T')' has statistically significant impact on Cap crew production rate at a level of 0.1.
- Consideration of only one of two drivers; 'Cap Size' and 'Cap Length'; is recommended since they are correlated, whichever is more convenient and/or confident.

Findings from this study, including weekly observed crew production rates along with identified drivers will enable highway agencies to enhance accuracy of contract time estimation for highway bridge construction. A practitioner will be able to select an accurate rate from the range of production rate to be used for time estimation of a particular project, referring to the identified drivers as well as using his/her professional experiences.

5. Conclusion

Numerous papers have been reviewed in order to gain quantifiable insight on methods of productivity data analysis. It has not been easy to identify factors driving construction productivity largely because it is very difficult to isolate an effect caused by one factor due to the nature of construction operation. Lack of a representative data is also one of the reasons. This paper presented methodology of identifying factors driving crew production rates. With representative data collected by the author from 13 on-going highway projects between February, 2002 and May, 2004, the methodology was demonstrated. Three factors were identified as statistically significant drivers of Cap crew production rate: 'Cap Size (m^3/ea)'; 'Cap Length (m)'; and 'Cap Shape (Rectangle versus Inverted 'T')'. It was also found that the production rates are best explained by a multiple regression model with two of the drivers; 'Cap Size' and 'Cap Shape'. It is the interests of industry as well as academia to collect a representative data and find drivers of activity production rate, which will lead to better construction time estimation and resource management.

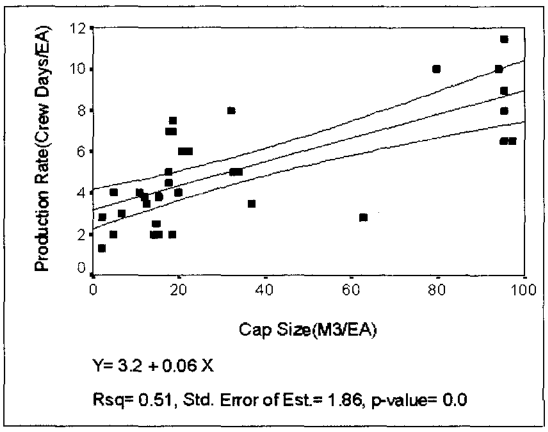


Figure 2. Cap; Scatter Plot & Regression Results (vs. Cap Size)

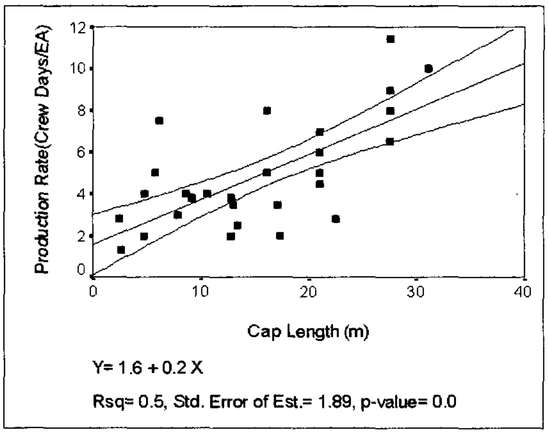


Figure 3. Cap; Scatter Plot & Regression Results (vs. Cap Length)

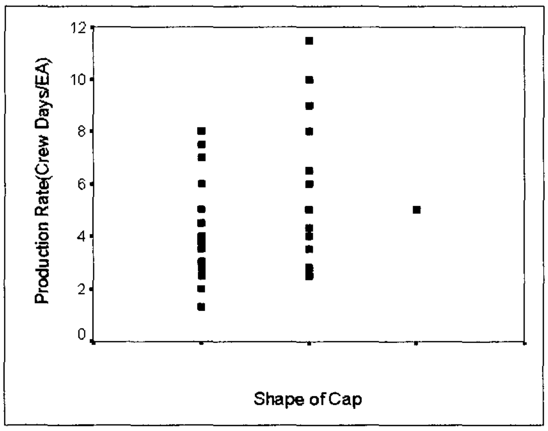


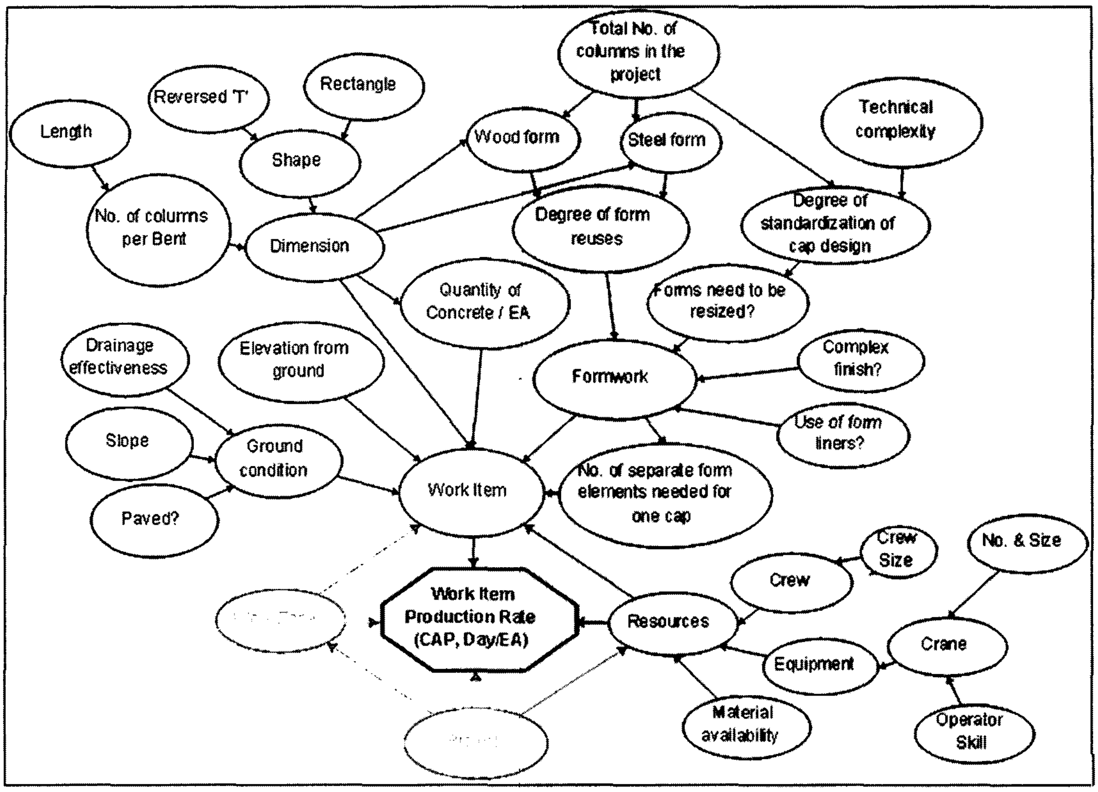
Figure 4. Cap; Scatter Plot (vs. Cap Shape)

Acknowledgements; This paper is written based on a research project data sponsored by the Texas Department of Transportation and special thanks to the inspectors and contractors who helped the data collection for this research.

References

1. Albright et al, 'Data Analysis and Decision Making' , 2nd edition, Thomas Learning, Inc., CA, USA, 2003
2. Chao, L. C., and Skibniewski, M. J., Estimating Construction Productivity: Neural-Network-based Approach, Journal of Computing. in Civil Engineering, Vol. 8, No. 2, 234-251, 1994
3. Christian, J. and Hachey, D., Effects of Delay Times on Production Rates in Construction, Journal of Construction Engineering and Management, Vol. 121, No. 1, pp. 20-26, 1995
4. Hanna, A. S., Peterson P., and Lee, M., Benchmarking Productivity Indicators for Electrical/mechanical Projects, Journal of Construction Engineering and Management, Vol. 128, No. 4, pp. 331-337, 2002
5. Hendrickson, C. et al., Expert System for Construction Planning, Journal of Computing in Civil Engineering, Vol. 1, No. 4, pp. 253-269, 1987
6. Herbsman, Z. J. and Ellis, R., 'Synthesis of Highway Practice 215: Determination of Contract Time for Highway Construction Projects,' Transportation Research Board, National Cooperative Highway Research Program, National Academy Press, Washington, D.C, 1995
7. Huh, Y. K., 'Highway Bridge Construction Production Rates for Contract Time Estimation." Dissertation, Dept. of Civil Eng., University of Texas at Austin, Austin, Texas, USA, 2004
9. Ming, Lu et. al, Estimating Labor Productivity Using Probability Inference Neural Network, Journal of Computing in Civil Engineering, Vol. 14, No. 4, pp. 241-248, 2000
12. Portas, J. and AbouRizk, S., Neural Network Model for Estimating Construction Productivity, Journal of Construction Engineering and Management, Vol. 123, No. 4, December 1997, pp. 399-410, 1997
13. Sakamoto, A. J., ' SOCL. STAT: LIN. MOD./STRC. EQ. SYS.,' Class (SOC 385L) Note, University of Texas at Austin, Texas, USA, 2002
14. Sanders, S. R. and Thomas, H. R., Masonry Productivity Forecasting Model, Journal of Construction Engineering and Management, Vol. 119, No. 1, pp. 163-179, 1993
15. Sanders, S. R., Thomas, H. R., and Smith, G. R., An Analysis of Factors Affecting Labor Productivity in Masonry Construction, PTI Rep. 9003, Pennsylvania Transportation Institute, Pennsylvania State University, University Park, Pa, 1989
16. Smith, S. D., Earthmoving Productivity Estimation Using Linear Regression Techniques, Journal of Construction Engineering and Management, Vol. 125, No. 3, pp. 133-141, 1999
17. Sonmez, R., 'Construction Labor Productivity Modeling with Neural Networks and Regression Analysis,' PhD thesis, Dept. of Civil Eng., Iowa State University, Iowa, USA, 1996
18. Sonmez, R. and Rowings, J. E., Construction Labor Productivity Modeling with Neural Networks, Journal of Construction Engineering and Management, Vol. 124, No 6, pp. 498-504, 1998
19. Stevens, J., 'Applied Multivariate Statistics for the Social Sciences,' eBook (http://www.netlibrary.com), Mahwah, N.J. Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey, 2002
20. Thomas, H. R. and Raynar, K. A., Scheduled Overtime and Labor Productivity: Quantitative Analysis, Journal of Construction Engineering and Management, Vol. 123, No. 2, pp. 181-188, 1997
21. Thomas, H. R. and Yiakoumis, I., Factor Model of Construction Productivity, Journal of Construction Engineering and Management, Vol. 113, No. 4, pp. 623-639, 1987
22. Thomas, H. R., Smith, G. R., and Sanders, S. R., An Exploratory Study of Productivity Forecasting Using the Factor Model for Masonry, PTI Rep. 9005, Pennsylvania Transportation Institute, Pennsylvania State University, University Park, Pa, 1989
23. Thomas, S. R., 'Quantitative Methods for Project Analysis,' Class (CE 395R.6) Note, Dept. of Civil Eng., University of Texas at Austin, TX, USA, 2003
24. Wonnacott, T. H. and Wonnacott, R. J., 'Regression: A Second Course in Statistics,' Krieger Publishing Company, Malabar, Florida, USA, 1986

Appendix A. Influence Diagram; Work item-related



Appendix B: Variables Studied But Not Yet Found Significant

WZ(Work Zone) Related	WZ Accessibility(Easy/Moderate/Difficult), WZ Construction Congestion(Mild/Moderate/Severe), Work Zone Site Drainage Effectiveness(Quickly Drains/Moderate/Easily Flooded), Land Slope(Flat/Moderate/Steep)
Work Item Related	Total CY of Concrete Placed, Crew size (Formwork crew only), Use of Form liners?(Yes/No), Complex finish?(Yes/No), Elevation of cap, No of columns per bent, Forms need to be modified for the operation?(Yes/No)