

A statistical model for Estimating Size and Productivity Ratio in a SDLC

Srinivasa Gopal

Abstract: During a software development life cycle, one has to estimate the effort and schedule required producing a unit of software. Estimates for The effort and schedule are derived from other measurements such as size and productivity Ratio. Size may be estimated using methods such as Function Point, WBS, LOC etc., Productivity Ratio is expressed as Size / Time to complete the Unit of Software. Such estimates are static and do not consider real time Parameters which cause variations from the estimate. In this paper. A statistical model for Size and Productivity Ratio are derived using Historical values after taking into account several factors which would Influence the size and the productivity ratio of produced software. Typically an estimate of the effort and schedule is required to complete a software work product or an enhancement request or a software fix. We use Function Points or WBS to estimate the size of the software In consideration. We use baseline values for estimating productivity Ratios. Real time values however differ from Baseline values as they Get influenced by several factors. Here are some factors which would influence the Actual Size of the software Later on we will use Statistical techniques of Multivariable Parameter estimation and logistic Regression to derive run time equations of Size and Productivity Ratio

I. FACTORS WHICH INFLUENCE SIZE ARE

A. Scope Creep: If there is a scope creep in the project, ie., If the requirements phase has a deviation or defects, this would affect the actual size.

B. Code Review: This would have an effect on the size as a Peer review of the code Can ensure that the minimum size is applied to the functions required to generate the code

C. Developer Skill Level: The skill level of the developers would definitely have an effect on the size.

D. Usage of Model: If a model is used for estimating size it makes a lot of assumptions. For example the Function Point methodology is platform independent, but, some times there may be a loss of accuracy in predicting the actual Size when this factor is used for conversion from Function Point to LOC

E. Object code: Use of Object code to nest functionality. If the size estimation Model uses linear supposition to add size to functions then Using Object code in the form of Dynamic Link Libraries etc., will reduce the size of the software.

Let us also take into account some factors, which do not have an impact and measure the co-relation with the size and the productivity ratio

1. Highest degree obtained
2. Gender of the developer

II. REGRESSION ANALYSIS

Here is a sample of tabulated values for the different factors taken into consideration.
 Size of the code is in LOC, the other values are in percentage

| Actual Size | Skill Level | RSI | Code Review | Machine Automation |
|-------------|-------------|--------|-------------|--------------------|
| 4,000.00 | 1.00 | 100.00 | 0.00 | 0.00 |
| 3,400.00 | 2.00 | 100.00 | 0.00 | 0.00 |
| 3,300.00 | 3.00 | 100.00 | 0.00 | 0.00 |
| 3,200.00 | 4.00 | 100.00 | 0.00 | 0.00 |
| 3,000.00 | 5.00 | 50.00 | 0.00 | 0.00 |
| 6,000.00 | 5.00 | 0.00 | 0.00 | 0.00 |
| 3,500.00 | 5.00 | 100.00 | 0.00 | 0.00 |
| 3,750.00 | 5.00 | 75.00 | 100.00 | 0.00 |
| 3,200.00 | 5.00 | 100.00 | 50.00 | 0.00 |
| 3,217.00 | 5.00 | 100.00 | 100.00 | 0.00 |
| 3,200.00 | 5.00 | 100.00 | 100.00 | 10.00 |
| 3,100.00 | 5.00 | 100.00 | 100.00 | 20.00 |
| 3,000.00 | 5.00 | 100.00 | 100.00 | 50.00 |

Table 1.0 : Shows the Actual LOC as a function of the different factors which would influence the size

III. REGRESSION EQUATION

The regression was run using an excel tool assuming linear relations and the result obtained was

$$\text{Actual Size} = -4.39 \times \text{Skill Level} + 21.64 \times \text{RSI} + -0.38 \times \text{Code Review} + -11.53 \times \text{Machine Automation} + 1378.40 (+/- 1569.33)$$

For the same example the Predicted Size using model such as FP is 3500 LOC

IV. SOME FACTORS WHICH INFLUENCE PRODUCTIVITY RATIO ARE

A. REUSABLE CODE: If the entire application can be automatically engineered this would make the application skill independent and would enhance productivity

B. SKILL LEVEL: An experienced developer would produce better than in an experienced developer.

C. *SHIFT TIMINGS*: Day time work is more productive than night time work.

Some sample values for the different parameters under which real time Productivity Ratio would vary were taken into consideration

| Skill Level | Reusability | Pr Ratio |
|-------------|-------------|----------|
| 4.00 | 0.00 | 60.00 |
| 3.00 | 0.00 | 40.00 |
| 4.00 | 3.00 | 100.00 |
| 4.00 | 4.00 | 120.00 |
| 4.00 | 5.00 | 130.00 |

The equation obtained due to multi variable regression analysis using an Excel tool

Was (Appendix A) shows detailed calculations

$$\text{Productivity Ratio} = 20.00 * \text{Skill Level} + 14.23 * \text{Reusable code} + -20.13 (+/- 2.45)$$

The baseline value which was taken in this company was 80 tested lines per day

But in reality due to factors of influence and positive observed correlation

The Actual productivity ratio varies as the linear equation suggests

CONCLUSIONS

Projects derive estimates using baseline values. But frequently these baseline values are correlated with other measures. In this paper two measurements such as Size and Productivity ratio, which are used to derive measurements such as Effort and Schedule, have been taken into consideration. Using sample values a real time regression equation was derived to show that the actual values vary with respect to a number of parameters.

REFERANCES

- Websites : www.softwametrics.com - Estimating Data
 Books : Quality By Design - Juran
 Software and SPC : www.qimacros.com

APPENDIX

A. Regression Analysis

| Equation Parameters | |
|---------------------|----------|
| R Square | 0.9970 |
| Adjusted R Square | 0.9951 |
| Standard Error | 2.4460 |
| F - Statistic | 506.8929 |

| Multiple Regression Equation | | | Independent Analysis | | |
|------------------------------|--------------|----------------|----------------------|----------|-----------|
| | Coefficients | Standard Error | R Squared | Gradient | Intercept |
| Intercept | -20.128 | 7.056 | | | |
| Skill Level | 20.000 | 1.730 | 13.15% | 20.00 | 8.33 |
| Reusable code | 14.231 | 0.480 | 86.55% | 14.23 | 59.87 |

Auto Corelation Statistics

| |
|---------|
| DI=1.08 |
| Du=1.36 |
| DW-Stat |
| 2.25 |
| 2.53 |

AUTHOR

The author obtained a B.Tech degree in Mechanical Engineering From the Indian Institute of Technology at Madras in 1999 And a MS by research degree from the University of Regina in Canada in 1991. The author has more than 15 years of experience In the Information Technology Industry in designing and managing Diverse applications, and has also been part of many quality initiatives such as CMM, ISO, and PCMM.