

Establishing Correlation between Size Estimation Metrics and Effort - A Statistical Approach

Meenakshi Kandpal
M.Tech (Computer Science)
Bhubaneswar, India

Anmol Kandpal
Assistant Manager, Project & Technical Department,
NALCO, Bhubaneswar, India

ABSTRACT

For effort estimation, among various inputs, size is the most vital one. The estimation of effort and cost depends on the accurate prediction of the size. Given the central role that software size plays in the Software Cost Estimation, software developers are increasingly focusing on process improvement in the software size estimation area. This demand has spurred the provision of a number of new and /or improved approaches to software size estimation. Among all, Object-oriented technology is becoming popular in today's software development environment. There are various methods available for establishing the relationship between the size estimation metric and effort. This paper aims to establish statistics relation using mathematical approach. Statistical correlation is being calculated between the two variables to measure the dependability of the relationship and determining the precision of a relationship

Keywords

Size, effort, Correlation coefficient, Public methods, Classes

1. INTRODUCTION

For more than a decade, the procedural paradigm is being studied in software metrics. Various traditional software size metrics like heuristics based on experience, Line of Code, Function point, software science, object point etc. used for size estimation. Researchers were trying to link metrics with system maintainability in procedural paradigm. Researchers like Rombach [1][2], Wake & Henry[3] shows that maintainability of the system can be predicted using software metrics..The SLOC and Function point requires dividing the solution space into data and procedure. However, these metrics have major drawbacks when used for object-oriented approaches of software development. Object Oriented Programming was another area of focus which proves to have pace & higher quality of software than procedural paradigm. The difference between the procedural paradigm and the object oriented paradigm is due to the difference between the programming philosophies in two paradigms. The study of the object oriented paradigm results in object oriented concepts such as object, class, attributes, inheritance, method and message passing.

The primary objectives for Object Oriented metrics are no different than those for metrics derived for conventional software and aims at:

- To better understand the quality of the product
- To assess the effectiveness of the process
- To improve the quality of work performed at a project level

Various object oriented metrics have been proposed in literature. Out of these, the metrics proposed by Abreau [4, 5], J. Bansiya et al. [6], Briand et al. [7], Chidamber and Kemerer [8], Lorenz et al. [9], W. Li et al. [10, 11] are mostly referred.

This paper aims to establish a statistical correlation between software size estimation metrics & effort. Mathematical model is being used to establish the correlation.

2 VARIOUS STATISTICAL TOOLS FOR ESTABLISHING CORRELATION

2.1 Least Square Method:

A very popular technique used to compute estimations of parameters and to fit data. It is one of the oldest techniques of modern statistics. Nowadays, the least square method is widely used to find or estimate the numerical values of the parameters to fit a function to a set of data and to characterize the statistical properties of estimates. It exists with several variations: Its simpler version is called ordinary least squares (OLS), a more sophisticated version is called weighted least squares (WLS), which often performs better than OLS because it can modulate the importance of each observation in the final solution.

The most frequently used method of OLS is linear regression, which corresponds to the problem of finding a line (or curve) that best fits a set of data. In the standard formulation, a set of 'n' pairs of observations $\{Y_i, X_i\}$ is used to find a function giving the value of the dependent variable (Y) from the values of an independent variable (X). With one variable and a linear function, the prediction is given by the equation 2.1:

$$Y = a + bX \quad \text{- Eqn. 2.1}$$

This equation involves two free parameters which specify the intercept (a) and the slope (b) of the regression line. The least square method defines the estimate of these parameters as the values which minimize the sum of the squares (hence the name least squares) between the measurements and the model (i.e., the predicted values). This amounts to minimizing the expression:

$$\varepsilon = \sum_i^n ((Y_i - \hat{Y}_i)^2) = \sum_i^n ((Y_i - (a + bX_i))^2) \quad \text{- Eq. 2.2}$$

(where ε stands for "error" which is the quantity to be minimized). This is achieved using standard techniques from calculus, namely the property that a quadratic (i.e., with a square) formula reaches its minimum value when its derivatives vanish. Taking the derivative of ε with respect to 'a' and 'b' and setting them to zero gives the following set of equations:

$$\frac{\partial \varepsilon}{\partial a} = 2na + 2b \sum_i^n X_i - 2 \sum_i^n Y_i = 0 \quad \text{- Eqn. 2.3}$$

$$\frac{\partial \varepsilon}{\partial b} = 2b \sum_i^n X_i^2 + 2a \sum_i^n X_i - 2 \sum_i^n Y_i X_i = 0 \quad \text{- Eqn. 2.4}$$

Solving the above two equations, Eqn. 2.3 & Eqn. 2.4 to obtain gives the value of slope 'b' as:

$$b = \frac{\{\sum(X_i - \bar{Y})(X_i - \bar{X})\}}{\sum(X_i - \bar{X})^2} \quad \text{- Eqn. 2.5}$$

And the value of Y intercept of the line i.e 'a' as:

$$a = \bar{Y} - b\bar{X} \quad \text{- Eqn. 2.6}$$

Where \bar{Y} and \bar{X} denoting the means of X and Y.
The Least Square Method is used for establishing Correlation between Traditional Metrics & Objects Oriented Sizing metrics with respect to Effort.

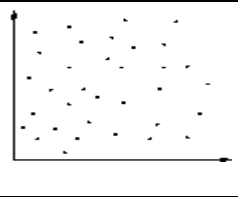
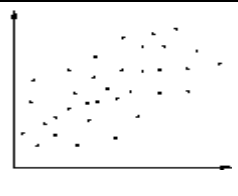
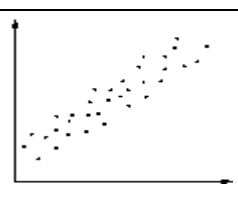
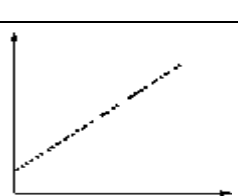
2.2 Scatter Plot

A scatter plot, scatter plot, or scatter graph is a summary of bivariate data used to analyze the relationship between two variables. It displays the relationship between the two variables, in graphical form and aids the interpretation of the correlation coefficient or regression model.

Each unit contributes one point to the scatter plot, on which points are plotted but not joined. The resulting pattern indicates the type and strength of the relationship between the two variables. (Valerie J. Easton and John H. McColl's *Statistics Glossary v1.1*).

Table 2.1 summarizes the type of scatter diagram, the degree of Correlation it indicate & its interpretation.

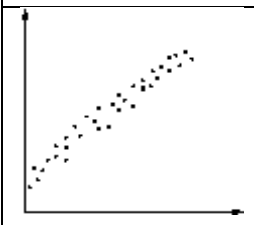
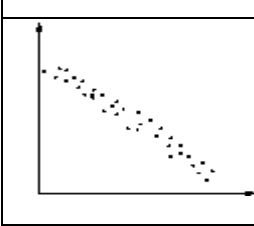
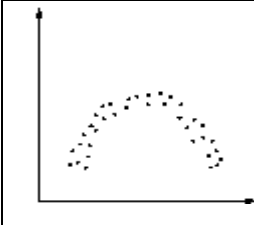
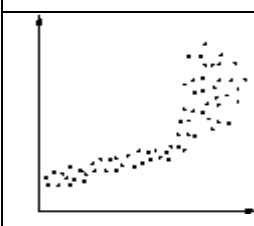
Table 2.1: Degree of Correlation

Scatter Diagram	Degree of Correlation	Interpretation
	None	No relationship can be seen.
	Low	A vague relationship is seen.
	High	The points are grouped into a clear linear shape.
	Perfect	All points lie on a line (which is usually straight). Perfect relationship

Scatter plot also helps in visualizing the type of Correlation. Sets of data are plotted on a graph, with the y-axis being used for the variable to be predicted and the x-axis being used for the variable to make the prediction. Scatter Plot diagrams are useful tools to determine the type of Correlation between the variable. By observing the pattern in the scatter plot, the nature of association between the two variables can be determined.

Table 2.2, indicates the various types of scatter diagram with the type of correlation & their interpretation with respect to each scatter diagram.

Table 2.2: Type Of Correlation

Scatter Diagram	Types of Correlation	Interpretation
	Positive	Straight line, sloping up from left to right.
	Negative	Straight line, sloping down from left to right..
	Curved	Various curves, typically U- or S-shaped.
	Part linear	Part of the diagram is a straight line (sloping up or down).

2.3 Correlation Coefficients

Correlation measures the dependability of the relationship (the goodness of fit of the data to that). It is a measure of how well one variable can predict the other (given the context of the data), and determines the precision of a relationship. There are following two measures for measuring the relationship:

- Coefficient of Correlation
- Coefficient of Determination

2.3.1 Correlation Coefficients:

Coefficient of correlation is a mathematical measure of how much one number can expected to be influenced by changes in another. The linear correlation coefficient is sometimes referred to as the Pearson product moment correlation coefficient in honor of its developer Karl Pearson. This is also known as product moment correlation and simple correlation coefficient. It gives a precise numerical value of the degree of linear relationship between two variables X and Y. The linear relationship may be given by:

$$Y = a + bX \quad \text{Eqn. 2.7}$$

This type of relation may be described by a straight line. The intercept that the line makes on the Y-axis is given by 'a' & the slope of the line is given by 'b'. It gives the change in the value of Y for very small change in the value of X.

Let X_1, X_2, \dots, X_n be n values of X and Y_1, Y_2, \dots, Y_n be the corresponding values of Y. The arithmetic means of X and Y are defined as:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{Eqn. 2.8}$$

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} \quad \text{Eqn. 2.9}$$

And their variances are as follows:

$$\sigma_x^2 = \frac{1}{n} * \sum (X_i - \bar{X})^2 \quad \text{Eqn. 2.10}$$

$$\sigma_y^2 = \frac{1}{n} * \sum (Y_i - \bar{Y})^2 \quad \text{Eqn. 2.11}$$

The standard deviations of X and Y respectively are the positive square roots of their variances.

$$\sigma_x = \sqrt{\frac{1}{n} * \sum (X_i - \bar{X})^2} \quad \text{Eqn. 2.12}$$

$$\sigma_y = \sqrt{\frac{1}{n} * \sum (Y_i - \bar{Y})^2} \quad \text{Eqn. 2.13}$$

Covariance of X and Y is defined as,

$$Cov(X, Y) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n} \quad \text{Eqn. 2.14}$$

Where $(X_i - \bar{X})$ & $(Y_i - \bar{Y})$ are the deviations of the ith value of X and Y from their mean values respectively.

The product moment correlation or the Karl Pearson's measure of correlation is given by

$$r = \frac{Cov(X, Y)}{\sigma_x * \sigma_y} \quad \text{Eqn. 2.15}$$

Substituting values in equation, the correlation coefficient r,

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 * \sum (Y_i - \bar{Y})^2}} \quad \text{Eqn. 2.16}$$

The Interpretation of Coefficient of Correlation is as follows:

- 1) A correlation coefficient of zero means that the two numbers are not related.
- 2) A non-zero correlation coefficient means that the numbers are related, but unless the coefficient is either 1 or -1 there are other influences and the relationship between the two numbers is not fixed. So if you know one number you can estimate the other, but not with certainty. The closer the correlation coefficient is to zero the greater the uncertainty, and low correlation coefficients means that the relationship is not certain enough to be useful.
- 3) A correlation coefficient of 1 means that the two numbers are perfectly correlated: if one grows so does the other,

and the change in one is a multiple of the change in the other.

- 4) A correlation coefficient of -1 means that the numbers are perfectly inversely correlated. If one grows the other falls. The growth in one is a negative multiple of the growth in the other.

The significance of Correlation Coefficient as described above is shown in Fig. 2.1.

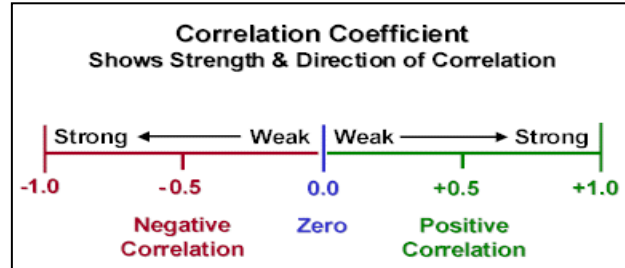


Fig. 2.1, Strength of Coefficient of Correlation

2.3.2 Coefficient of Determination:

Coefficient of Determination, also known as r Squared, is interpreted as the goodness of fit of a regression. The higher the coefficient of determination, the better the variance that the dependent variable is explained by the independent variable. The coefficient of Determination is the overall measure of the usefulness of a regression. The coefficient of determination is the ratio of the explained variation to the total variation. The Coefficient of Correlation can be defined by using following formula

$$r^2 = \frac{(\sum(X_i - \bar{X})(Y_i - \bar{Y}))^2}{\sum (X_i - \bar{X})^2 * \sum (Y_i - \bar{Y})^2} \quad \text{Eqn. 2.17}$$

The Importance of Coefficient of Determination is as follows:

The Coefficient of Determination, r^2 , is useful because it gives the proportion of the variance (fluctuation) of one variable that is predictable from the other variable. The coefficient of determination is such that $0 < r^2 < 1$, and denotes the strength of the linear association between x and y. The coefficient of determination represents the percent of the data that is the closest to the line of best fit.

The coefficient of determination is equal to the percent of variation in one variable that is accounted for (predicted) by the other variable. Though the correlation coefficient is useful to determine the degree of linear relationship between two variables, the coefficient of determination allows us to interpret the relationships in terms of variations, a more familiar term.

3. STATISTICAL RELATION BETWEEN TRADITIONAL SIZING METRICS & OBJECTS ORIENTED SIZING METRICS WITH RESPECT TO EFFORT.

There are various mathematical models are available for establishing the statistical correlation between the two variables. Out of these variables, one variable is dependent & other is independent variable. The independent variable is size & the dependent variable is effort as with increase in the size of the software the effort increases. In order to establish the equation

of the line least square method is used and the r^2 i.e Coefficient of determination is calculated.

3.1 Data Collection

3.1.1 Type for Data:

The data collected is Secondary data.

3.1.2 Source of Data:

Data has been taken from a research paper published in the research paper “An innovative model for object-oriented costs estimating” by Vincenzo Giliberti, Michele Gorgoglione, Raffaele Vitulli [15].

The available data set is for five different projects. The calculated data set with respect to each project for a traditional size metric, KLOC and two Object Oriented Size Metrics i.e No. of Class Method & No. of Public Methods, has been indicated in the table 3.1.

Table 3.1: Data Set for 5 projects

Project	KLOC	Classes	Public Methods
1	16.3	64	972
2	11.5	36	1076
3	26	130	3730
4	12	74	698
5	3.5	42	95

The actual effort with respect to each project as indicated in the paper [15], is listed in table 3.2. The effort with respect to each project has been calculated in man month.

Table 3.2: Actual effort w.r.t five projects

Project	Actual Effort
1	7.5
2	7
3	90
4	3
5	2.5

3.1.3 Limitation:

There are following limitation for the collection of data for establishing a statistical relation between various metrics & effort:

- Collection of primary data, for calculation of effort with respect to various software codes developed is difficult to obtain.
- The codes & actual efforts with respect to each code, Software firms are not willing to share as it their propriety information.
- Software companies are mostly using non Algorithmic methods like Delphi techniques for calculation of efforts.

Considering above limitations, Secondary data has been taken for establishing the correlation between the metrics & actual effort.

3.2 Correlation between KLOC & Actual effort:

In this section, statistical correlation has been established between an independent variable ‘KLOC’, which is a popular traditional size estimation metrics and dependent variable ‘effort’. The KLOC calculated for five different projects & the actual effort required for the five projects is tabulated in Table 3.3.

Table 3.3: KLOC Vs Actual Effort

Project	KLOC	Actual Effort
1	16.3	7.5
2	11.5	7
3	26	90
4	12	3
5	3.5	2.5

Let KLOC be ‘X’ & Actual Effort be ‘Y’

Mean of X (KLOC),

$$\bar{X} = \frac{(16.3 + 11.5 + 26 + 12 + 3.5)}{5} = 13.86$$

Mean of Y (Actual Effort),

$$\bar{Y} = \frac{(7.5 + 7 + 90 + 3 + 2.5)}{5} = 22$$

Table 3.4 summarizes the value of deviation of values of X & Y from the mean & squared & product values of the deviation of X & Y from their respective means.

Table 3.4: Calculation of squared & product values for KLOC & Effort

Project	KLOC (X)	Actual Effort (Y)	(X - \bar{X})	(Y - \bar{Y})	(X - \bar{X}) ²	(Y - \bar{Y}) ²	(X - \bar{X})(Y - \bar{Y})
1	16.3	7.5	2.44	-14.5	5.95	210.25	-35.38
2	11.5	7	-2.36	-15	5.56	225	35.4
3	26	90	12.14	68	147.37	4624	825.52
4	12	3	-1.86	-19	3.45	361	35.34
5	3.5	2.5	-10.36	-19.5	107.32	380.25	202.02

From Table 2.4,

$$\sum(X - \bar{X})^2 = 269.69$$

$$\sum(Y - \bar{Y})^2 = 5800.50$$

$$\sum(X - \bar{X})(Y - \bar{Y}) = 1062.90$$

Calculating the Slope using equation 2.5,

$$b = \frac{\{\sum(Y_i - \bar{Y})(X_i - \bar{X})\}}{\sum(X_i - \bar{X})^2} = \frac{1062.9}{269.69} = 3.94$$

And the intercept on Y axis,

$$a = \bar{Y} - b\bar{X}$$

$$a = 22 - 3.94 * 13.86 = -32.62$$

So, the equation of the line,

$$Y = -32.62 + 3.94 X$$

Calculating the Coefficient of Correlation using equation 2.16,

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2} * \sqrt{\sum(Y_i - \bar{Y})^2}}$$

$$r = \frac{1062.90}{\sqrt{269.69} * \sqrt{5800.50}} = 0.845$$

Coefficient of Determination

$$r^2 = (0.845)^2 = 0.7229$$

3.3 Correlation between No. of Class & Actual Effort:

In this section, statistical correlation has been established between an independent variable 'No. of Class', which is an Object Oriented size estimation metrics and dependent variable 'effort'. The No. of class calculated for five different projects & the actual effort required for the five projects is tabulated in Table 3.5.

Table 3.5: CLASSES Vs Actual Effort

Project	Classes	Actual Effort
1	64	7.5
2	36	7
3	130	90
4	74	3
5	42	2.5

Let No. of Classes be 'X' & Actual Effort be 'Y'

Mean of X (No. of Classes),

$$\bar{X} = \frac{(64 + 36 + 130 + 74 + 42)}{5} = 69.20$$

Mean of Y (Actual Effort),

$$\bar{Y} = \frac{(7.5 + 7 + 90 + 3 + 2.5)}{5} = 22$$

Table 3.6 summarizes the value of deviation of values of X & Y from the mean & squared & product values of the deviation of X & Y from their respective means.

Table 3.6: Calculation of squared & product values for CLASSES & Effort

Project	Classes (X)	Actual Effort (Y)	(X- \bar{X})	(Y- \bar{Y})	(X- \bar{X}) ²	(Y- \bar{Y}) ²	(X- \bar{X})(Y- \bar{Y})
1	64	7.5	-5.2	-14.5	27.04	210.25	75.4
2	36	7	-33.2	-15	1102.24	225	498
3	130	90	60.8	68	3696.64	4624	4134.4
4	74	3	4.8	-19	23.04	361	-91.2
5	42	2.5	-27.2	-19.5	739.84	380.25	530.4

Also from the above table,

$$\sum(X-\bar{X})^2 = 5588.80$$

$$\sum(Y-\bar{Y})^2 = 5800.50$$

$$\sum(X-\bar{X})(Y-\bar{Y}) = 5147$$

Calculating the Slope using equation 2.5,

$$b = \frac{\{\sum(Y_i - \bar{Y})(X_i - \bar{X})\}}{\sum(X_i - \bar{X})^2} = \frac{5147}{5588.80} = 0.92$$

And the intercept on Y axis,

$$a = \bar{Y} - b\bar{X}$$

$$a = 22 - 0.92 * 5588.80 = -41.72$$

So the Equation of the line,

$$Y = -41.72 + 0.92 X$$

Calculation of Coefficient of Correlation using equation 2.16,

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2} * \sqrt{\sum(Y_i - \bar{Y})^2}}$$

$$r = \frac{5147}{\sqrt{5588.80} * \sqrt{5800.50}} = 0.903$$

Coefficient of Determination,

$$r^2 = (0.903)^2 = 0.8172$$

3.4 Correlation between Public Method & Actual Effort

In this section, statistical correlation has been established between an independent variable 'No. of Public Methods', which is an Object Oriented size estimation metric and dependent variable 'effort'. The No. of class calculated for five different projects & the actual effort required for the five projects is tabulated in Table 3.7.

Table 3.7: Public Methods Vs Actual Effort

Project	Public Method*	Actual Effort
1	9.72	7.5
2	10.76	7
3	37.3	90
4	6.98	3
5	0.45	2.5

* - Per Hundred

Let No. of Public method be 'X' & Actual Effort be 'Y'

Mean of X (No. of Public Method),

$$\bar{X} = \frac{(9.72 + 10.76 + 37.3 + 6.98 + 0.45)}{5} = 13.04$$

Mean of Y (Actual Effort),

$$\bar{Y} = \frac{(7.5 + 7 + 90 + 3 + 2.5)}{5} = 22$$

Table 2.8 summarizes the value of deviation of values of X & Y from the mean & squared & product values of the deviation of X & Y from their respective means.

Table 3.8: Calculation of squared & product values for Public Method & Effort

Project	Public Method* (X)	Actual Effort (Y)	(X- \bar{X})	(Y- \bar{Y})	(X- \bar{X}) ²	(Y- \bar{Y}) ²	$\frac{(X-\bar{X})}{(Y-\bar{Y})}$
1	64	7.5	-5.2	-14.5	27.04	210.25	75.4
2	36	7	-33.2	-15	1102.2	225	498
3	130	90	60.8	68	3696.6	4624	4134.4
4	74	3	4.8	-19	23.04	361	-91.2
5	42	2.5	-27.2	-19.5	739.84	380.25	530.4

* - Per Hundred

From the table 3.8,

$$\sum(X-\bar{X})^2 = 800.00$$

$$\sum(Y-\bar{Y})^2 = 5800.50$$

$$\sum(X-\bar{X})(Y-\bar{Y}) = 2092.66$$

Calculating the Slope by using equation 2.5,

$$b = \frac{\{\sum(Y_i - \bar{Y})(X_i - \bar{X})\}}{\sum(X_i - \bar{X})^2} = \frac{2092.66}{800} = 2.61$$

And the intercept on Y axis,

$$a = \bar{Y} - b\bar{X}$$

$$a = 22 - 2.61 * 13.04 = -12.12$$

So, equation of the line,

$$Y = -12.12 + 2.61 X$$

Calculating the Coefficient of Correlation using equation 2.16,

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2} * \sqrt{\sum(Y_i - \bar{Y})^2}}$$

$$r = \frac{2092.66}{\sqrt{800} * \sqrt{5800.50}} = 0.971$$

Coefficient of Determination,

$$r^2 = (0.971)^2 = 0.9437$$

4. CONCLUSION

The accuracy of a software estimate is always questionable and lots of efforts have been given to make an estimate more accurate. Among various factors size is a critical factor in determining cost, schedule, and effort. Poor size estimation may lead to budget overruns and late deliveries.

Traditional methods used for size estimation, requires lot of efforts and also do not give accurate results when used for newer and efficient software development method like Object Oriented technology.

In an Object Oriented Paradigm, several metrics are suggested by various researchers for size estimation. Use of Object Oriented Metrics for size estimation helps in standardizing the results. These metrics gives accurate results with less effort than traditional methods.

A statistical correlation has been established between the effort & size of the metrics. The correlation has been established using

mathematical models. It is evident from the results of the analysis that the variation in the actual effort due to the size can be best described by the Object Oriented metrics in comparison of traditional methods. Hence, The Object Oriented metrics when used for size estimation give more accurate results in comparison to the traditional size estimation methods.

5 FUTURE WORK

Object oriented metrics overcomes the limitation of traditional metrics when used for size estimation. Further studies are required for establishing the statistically correlation to determine the metrics among various available object oriented metrics which can best describe the variation in the effort.

6. REFERENCES

- [1] Rombach, H. Dieter, "A controllable experiment on impact of software structure on maintainability," IEEE Transaction on Software Engineering, Vol. SE-13, No. 3, March 1987.
- [2] Rombach, H. Dieter, "Design Measurement: Some lesson learned," IEEE Software, March 1990.
- [3] Wake, Steve and Sallie Henry, "A model based on software quality factor which predicts maintainability," Proceeding: Conference on Software Maintainability, October 1988.
- [4] F. B. Abreu and R. Carapua, "Candidate Metric for OOS within taxonomy framework, Journal of System & Softwrae, Vol. 26, No. 1, July 1994.
- [5] F. B. Abreu, "The MOOD Metrics Set", In Proc. ECOOP'95, Workshop on Metrics, 1995.
- [6] J. Bansiya and C.G. Davis, "A Hierarchical Model for Object-Oriented Design Quality Assessment", IEEE Transactions on Software Engineering, Vol. 28, No. 1, 2002.
- [7] L. C. Briand, J. W. Daly and J. Wust, "A Unified Framework for Coupling Measurement in Object-Oriented Systems", IEEE Transactions on Software Engineering, Vol. 25, No. 1, 1999.
- [8] S. R. Chidamber and C. F. Kemerer, "A Metrics Suite for Object Oriented Design," IEEE Transactions on Software Engineering, Vol. 20, No. 6, 1994.
- [9] M. Lorenz and J. Kidd, Object-Oriented Software Metrics, Prentice Hall, 1994.
- [10] W. Li, and S. Henry, "Object-Oriented Metrics that Predict Maintainability". Journal of Systems and Software, Vol. 23, 1993.
- [11] W. Li, "Another Metric Suite for Object Oriented Programming", The Journal of Systems and Software, Vol. 44, 1998.
- [12] 'A Toolbook for Quality Improvement and Problem Solving', Author, David Straker.
- [13] "Statistics for Management", Seventh Edition, Author, Levin Rubin.
- [14] Hervé Abdi: Method for least squares, The University of Texas at Dallas (<http://www.utdallas.edu/~herve/Abdi-LeastSquares06-pretty.pdf>).
- [15] Vincenzo Giliberti, Michele Gorgoglione, Raffaele Vitulli, "An innovative model for object-oriented costs estimating", Proceedings of the LMO'97 Conference, Brest (France), 22-24 October 1997.