Master Thesis
Software Engineering
Thesis no: MSE- 2010:03
January 2010

The Evaluation of Well-known Effort Estimation Models based on Predictive Accuracy Indicators

Khalid Khan

School of Computing
Blekinge Institute of Technology
Box 520
SE – 372 25 Ronneby
Sweden

This thesis is submitted to the School of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Software Engineering. The thesis is equivalent to 15 weeks of full time studies.

Contact Information:
Author(s):
Khalid Khan
E-mail: khalidkhan79@gmail.com

University advisor(s):
Dr. Cigdem Gencel
Assistant Professor, Dept. of Systems and Software Engineering
Soft Centre, Ronneby
Tel: 0457 385736

School of Engineering
Blekinge Institute of Technology
Box 520
SE – 372 25 Ronneby
Sweden

Internet : www.bth.se/tek
Phone : +46 457 38 50 00
Fax : + 46 457 271 25

ABSTRACT

Accurate and reliable effort estimation is still one of the most challenging processes in software engineering. There have been numbers of attempts to develop cost estimation models. However, the evaluation of model accuracy and reliability of those models have gained interest in the last decade. A model can be finely tuned according to specific data, but the issue remains there is the selection of the most appropriate model. A model predictive accuracy is determined by the difference of the various accuracy measures. The one with minimum relative error is considered to be the best fit. The model predictive accuracy is needed to be statistically significant in order to be the best fit. This practice evolved into model evaluation. Models predictive accuracy indicators need to be statistically tested before taking a decision to use a model for estimation.

The aim of this thesis is to statistically evaluate well known effort estimation models according to their predictive accuracy indicators using two new approaches; bootstrap confidence intervals and permutation tests. In this thesis, the significance of the difference between various accuracy indicators were empirically tested on the projects obtained from the International Software Benchmarking Standard Group (ISBSG) data set. We selected projects of Un-Adjusted Function Points (UFP) of quality A. Then, the techniques; Analysis Of Variance ANOVA and regression to form Least Square (LS) set and Estimation by Analogy (EbA) set were used. Step wise ANOVA was used to form parametric model. *K-NN* algorithm was employed in order to obtain analogue projects for effort estimation use in EbA.

It was found that the estimation reliability increased with the pre-processing of the data statistically, moreover the significance of the accuracy indicators were not only tested statistically but also with the help of more complex inferential statistical methods. The decision of selecting non-parametric methodology (EbA) for generating project estimates in not by chance but statistically proved.

Keywords: Measurement, models evaluation, global accuracy measures, bootstrap confidence intervals, ISBSG data set.

ACKNOWLEDGEMENT

I would like to extend my gratitude to my supervisor, Dr. Cigdem Gencel for putting her endless efforts into my thesis work. Her kind, caring and supporting behavior supported me throughout my work. At last but not least I am very thankful to my wife, Sadia who supported and encouraged me throughout my thesis work. Without her prayers and support I would have never been able to complete my work. I am thankful to all my well wishers who prayed for my success and may God succeed them in their endeavor.

CONTENTS

ABSTRACT ..... **I**

ACKNOWLEDGEMENT ..... **II**

CONTENTS ..... **III**

1 INTRODUCTION **1**

2 PROBLEM DEFINITION/ GOALS **3**

2.1 PROBLEM DEFINITION 3
2.2 GOALS 4

3 RESEARCH METHODOLOGY **5**

3.1 QUALITATIVE PART OF RESEARCH 5
3.2 QUANTITATIVE PART OF RESEARCH 6

4 BACKGROUND **8**

4.1 SOFTWARE COST AND EFFORT ESTIMATION 8
4.1.1 *Reasons Which Make Software Cost Estimation Challenging* *8*
4.1.2 *Estimation Process* *9*
4.1.3 *Software Cost Estimation Techniques and Tools* *10*
4.1.4 *Strengths and weaknesses of cost estimation models* *13*
4.2 SIZE MEASUREMENT 13
4.3 MODEL CALIBRATION USING LINEAR REGRESSION 14
4.4 MEASUREMENT OF MODEL PERFORMANCE 14
4.4.1 ***Measures of Accuracy:*** *15*
4.4.2 ***Measure of Consistency:*** *15*
4.4.3 ***Regression Criteria:*** *15*
4.5 CRITERIA FOR MODEL EVALUATION 15
Functionality: *15*
Reliability: *15*
Usability: *16*
Portability: *16*

5 RELATED WORK **17**

6 EMPIRICAL STUDY **19**

6.1 DATA COLLECTION 19
6.2 STATISTICAL MODELING TECHNIQUES AND TOOLS USED 21
6.2.1 *EbA* *21*
6.2.2 *Ordinary Least Squares (OLS)* *21*
6.2.3 *Linear Regression* *21*
6.2.4 *Analysis of variance (ANOVA)* *21*
6.2.5 *Analysis of covariance (ANCOVA)* *22*
6.2.6 *Stepwise ANOVA* *22*
6.2.7 *MS Excel* *22*
6.2.8 *Statistical package for social sciences (SPSS)* *22*
6.2.9 *MatLab 7.0* *22*
6.2.10 *Angel Tool* *22*
6.3 EMPIRICAL STUDY CONDUCT - PHASE I: 22
6.3.1 *Formation of the LS and EbA sets* *23*
6.4 EMPIRICAL STUDY PHASE-II 41
6.4.1 *Global Accuracy Measures* *42*
6.4.2 *Parametric and non-parametric tests* *42*
6.4.3 *Constructing Confidence Intervals* *43*

6.5 SUMMARY OF THE RESULTS 44
7 CONCLUSION AND FUTURE WORK 46
APPENDIX I 48
APPENDIX II 60
REFERENCES: 85

1 INTRODUCTION

Estimation of effort and duration of software development is a field of significant research for both researchers and practitioners. The process of estimating software cost is the procedure to predict the effort which is required to develop a certain software system [1].

Software cost estimation models have their roots in the early 1950s [2] and by the time several models have been developed [23, 24, 25, 26, 27] for the reason to effectively and efficiently allocate the scarce resources to fulfill the organization needs [4]. However, still the field of estimation is not mature [5] and no model or method is stated as standardized and reliable for estimation by any organization [6, 7, 8, 17].

Unfortunately, although easy to be used and affective in certain cases, these cost models produces highly inaccurate and unreliable estimates whenever they do not fit into the specific environment in which they will be used [9, 10]. It is shown by [11] that less than one quarter of the projects is estimated accurately. Heemstra [8] showed in his study on 598 Dutch organizations that 35% of the organizations do not make an estimate, 50% of the responding organizations record no data on an ongoing project, 80% of the projects of the responding organizations have overruns of budgets and duration due to inaccurate estimation. The 2009 Standish Group CHAOS report indicates that 24% of the projects fail on the average [13].

Various reasons have been explored which makes software cost estimation difficult such as estimates are often done hurriedly and usually done by inexperienced estimators and large number of cost factors contributing to the estimation problem. However reliable software estimates are required for generating requests for proposals, contract negotiations, scheduling, monitoring and control [1, 20]. It also helps manager in better resource planning [19].

Software engineering society has developed and has been using a number of cost estimation tools and models for the last three decades. A model is a mathematical object using mathematical formulae or intrinsic algorithms capable of estimating human resources and duration required for the development of a software project [48] while a software tool is a utility which is based on the notion of a specific model [48].COCOMO [23] is a model while basic COCOMO, Intermediate COCOMO and COCOMO II are the tools based on COCOMO model, similarly Putnam [24] central part of the model is called Software Equation and SLIM is the tool based on Software Equation. The model hidden inside a tool is not completely known; therefore the choice of selecting a right tool is very difficult [35]. For example PRICE-S [41] and ESTIMACS [43] are software tools but the inter alia mathematical formulae and algorithms are not publicly available.

These models and tools can be broadly categorized as empirical models and analytical models [1, 20]. An empirical model requires data from the past projects to evaluate the current project and derives the basic function for the current project from the analysis of the particular projects from the database. On the other hand, analytical models uses formulae on global assumptions like the rate at which developer solve a problem and the number of available problems [20].

The most common prediction methods used for software effort prediction are Regression [84] and Analogy [85] based. In regression-based estimation the model must be specified in advance in which; one dependent variable is plotted against several independent variables in order to from a linear relationship [84]. On the other hand, in analogy based estimation; analogues which are completed projects that are most similar to the new project are identified with the help of expert or through automated tools. Estimation by analogy is free of distribution assumption and outlier problem therefore; it produces more realistic estimates [85].

The most common measure used for the evaluation of model accuracy is Mean of Relative Error (MRE) [86]. Since MRE is sensitive to outliers therefore single values of MRE have been used to measure the estimation accuracy that is Prediction at level *l* (PRED *l*)) [86]. Median of MRE (MdMRE) or Q_2 (second quartile) was proposed by Miyazaki et al [87] in order to overcome the problem of outliers in MRE.

According to Stensrud et al [88] the common criteria for model evaluation is MMRE, MdMRE and PRED (*l*) and can be used for the comparison of two cost prediction models, the one with lower MRE is preferred. Another use of MRE is to identify quantitatively the measure of the uncertainty of the prediction model. MRE is the basic measure in all three metrics, the reason of its popularity is that it is independent of units and type of prediction systems, that is why it can be used for both regression and analogy based systems [88].

Performance of the cost estimation models has always been the point of interest in research community and in industry. Models are usually evaluated for accuracy and consistency measurement [7, 28]. Keemer performed an empirical validation of four algorithmic models (SLIM, COCOMO, FPA and ESTIMACS) [7] without recalibration of the model and kept only the point of view of fidelity [48]. M.Boraso et al [30], O.Benediktsson et al [31] and T.Menzies et al [32] performed studies on model evaluation on the basis of the difference in global accuracy measures. All of the above studies are performed on model evaluation using measures of accuracy indicators and none of them evaluated the significance of these measures which are the functions of the model relative error.

Almost all models are evaluated for estimation accuracy and consistency. Several measures are proposed in literature which are the function of the predictive error [89] measured by appropriate method. Single measure of error is just a statistic and may contain significant variability. Another study from Shepperd [56] on regression and analogy based techniques proved that twelve out of twenty studies proved the superiority of EbA technique and only six papers used hypothesis testing instead of using conventional accuracy measures was used to evaluated the model predictive accuracy. Four out of six proved the superiority of regression.

The formal testing is overlooked in favor of simplicity and models are solely evaluated in the light of a single measure with risk. Therefore models should be evaluated with more inferential statistics such as confidence intervals and hypothesis testing [88]. Most oftenly the evaluation of the effort estimation models is done without prior concise data analysis with a particular framework and the model is stated significant on the basis of single measure of accuracy.

In this study we will use more robust techniques in order to identify the significance of the accuracy indicators and use inferential statistics to evaluate the cost estimation models. A new model will be developed for effort estimation. The results will be compared with other researchers work on model evaluation and the differences if any shall be identified and discussed.

Rest of the thesis is organized as follows: Chapter 2 discusses the problem being addressed, the research questions and the outcomes of the thesis work. Chapter 3 discusses the methodology adopted to find solutions to the problem discussed in the previous chapter. Chapter 4 is the literature review of the cost estimation process, various techniques of model evaluation and the detail discussion of the various accuracy measures. Chapter 5 presents the current state of the art and chapter 6 presents the empirical investigation of the research questions and conclusion is provided at the end.

2 PROBLEM DEFINITION/ GOALS

A brief description of the current state of the art is presented along with the challenges encountered in the evaluation of effort estimation models. The prospect goals are discussed which will be achieved upon the successful answering of the research questions posted here.

2.1 Problem Definition

In the cost estimation process the choice of selecting the right methodology is a challenging task. Several studies have been performed on the evaluation of model performances. Estimation models are always evaluated for accuracy and consistency [12, 28]. There is a need to evaluate estimation models quantitatively [28].

Wohlin and Milicic [12] observed various patterns in data representation and showed that these patterns can help estimator to observe various trends in data distribution. They concluded that this can help in selecting the right model for effort estimation and hence can be used to improve the accuracy of the estimation. Keemer [7] and M. Boraso [30] showed in the study of model evaluation that cost estimation models perform best in the environment in which they are built. Faren and Gurner [36] evaluated various cost estimation models in the light of model accuracy and consistency using MRE, MdMRE and PRED (*l*) and found that the estimation error is too large. Jeffery and Low [33] and Jeffery et al. [34] showed large estimation errors in model performance evaluation keeping fidelity in view and concluded that there is a need of model recalibration. Miyazaki and Mori [29] demonstrated that even COCOMO which is considered to be the mostly documented and publicly available model, showed high overestimation when used on external data. Various researchers performed study on model evaluation using conventional measures for simplicity and none of them make use of inferential statistics because of its complexity.

The International Software Benchmarking Standard Group (ISBSG) is an organization which has been collecting data on the completed projects and is publicly available for benchmarking purposes. Until now, few studies have been done on the model evaluation using ISBSG data set. Sun-Jen Huang et al [37] performed a comparative study on the evaluation of software effort estimation using only one kind of technique using ISBSG data set. Liu et al [38] performed a study on ISBSG Data R9 and defined a framework for the evaluation of model performance but never applied that framework to any of the available cost estimation tools. Stensrud et al [88] used ISBSG R7 for the evaluation of cost estimation models using inferential statistics without concise data filtration. They identified that relying only a statistical figure in the sense of accuracy measure is risky and the models should be evaluated in more inferential statistics thus using hypothesis testing and use of confidence intervals. Angelis and Mittas [90] in another study used ISBSG R7 for the evaluation of model predictive performance in the light of re-sampling tests and proposed semi-parametric methodology for cost prediction.

A couple of other studies were conducted on the model evaluation using hypothesis testing and confidence intervals instead of the classical accuracy measures such as [88] and [56]. The models being evaluated using these formal testing methods are more reliable and can be used to further data sets [88].

RESEARCH QUESTIONS:

This research is mainly focused on the evaluation of model accuracy and model consistency. The following research questions are answered within this thesis:

1. What are the measures required for the evaluation of model accuracy and consistency?

2. What is the mechanism behind well-known effort estimation models, their calibration, and evaluation?
3. Does the estimation reliability increases with the pre-processing of the project data statistically?
4. Are the accuracy indicators statistically significant?
5. What other measures can be used for the evaluation of effort estimation models?

2.2 Goals

This research is mainly targeted to evaluate the performance of various software estimation models. Well-known effort estimation models will be evaluated for accuracy and consistency. Measure for the consistency and accuracy will be calculated and trends in the data contained in the project repository (ISBSG) will be observed. Moreover it will contribute to:

- Propose a model for effort estimation.
- Use of more inferential statistics will be introduced.
- An understanding of the cost estimation models and their working mechanism.
- An understanding of the various model accuracies.
- Recommendation for the improvement of cost estimation models evaluation.

3 RESEARCH METHODOLOGY

"A good researcher is not 'one who knows the right answers' but 'one who is struggling to find out what the right questions might be". Phillips and Pugh (1994: 48)

Higher Education Funding Council for England defines research as ' original investigation undertaken in order to gain knowledge and understanding' (HEFCE 1998) [40].

Research method is a specific way of data collection and analysis [29]. Creswell narrated these methods as [39].

- ***Quantitative approach*:** The researchers mainly focus on cause and effect thinking, reduction to specific variables and hypothesis and questions, use of measurements and observation and the test of theories. The quantitative research is employed in the form experiment, surveys and collection of data on predetermined instruments that results in statistical data.

- ***Qualitative approach*:** The researcher collects data using case studies, grounded theory, ethnographies or literature review.

- ***Mixed method approach:*** Combination of qualitative and quantitative methods, it allows researcher to collect data both in numeric as well as text form.

The goals of this research are achieved using mixed methodology.

3.1 Qualitative Part of Research

In the qualitative part of this thesis, we aimed to answer the following research questions.

RQ1. What are the measures required for the evaluation of model accuracy and consistency?

RQ2. What is the mechanism behind well-known effort estimation models, their calibration, and their evaluation?

Literature review was adopted as qualitative method of gathering the relevant information about various measures used for the evaluation of cost estimation methodologies. Their strengths and weaknesses were identified in the light of particular metrics which affect their performance. Literature review also provided an insight of the research work which had been done on the usefulness of these cost estimation models. The building block and working mechanism of the well-known effort estimation models was studied. The use of model calibration in order to use the external model according to company own data set was studied.

Various methodologies were explored for the evaluation of cost estimation models. Criterion for model evaluation was identified based on the literature review. Besides a detailed survey of the cost estimation models were conducted in order to identify their strengths and weaknesses. Several case studies conducted were investigated in order to identify metrics for model evaluation. Qualitative part of our thesis helped in identifying a set of metrics which defines criterion for the evaluation of effort estimation models.

Figure 3.1 shows the detailed proposed research methodology adopted for the current research work.

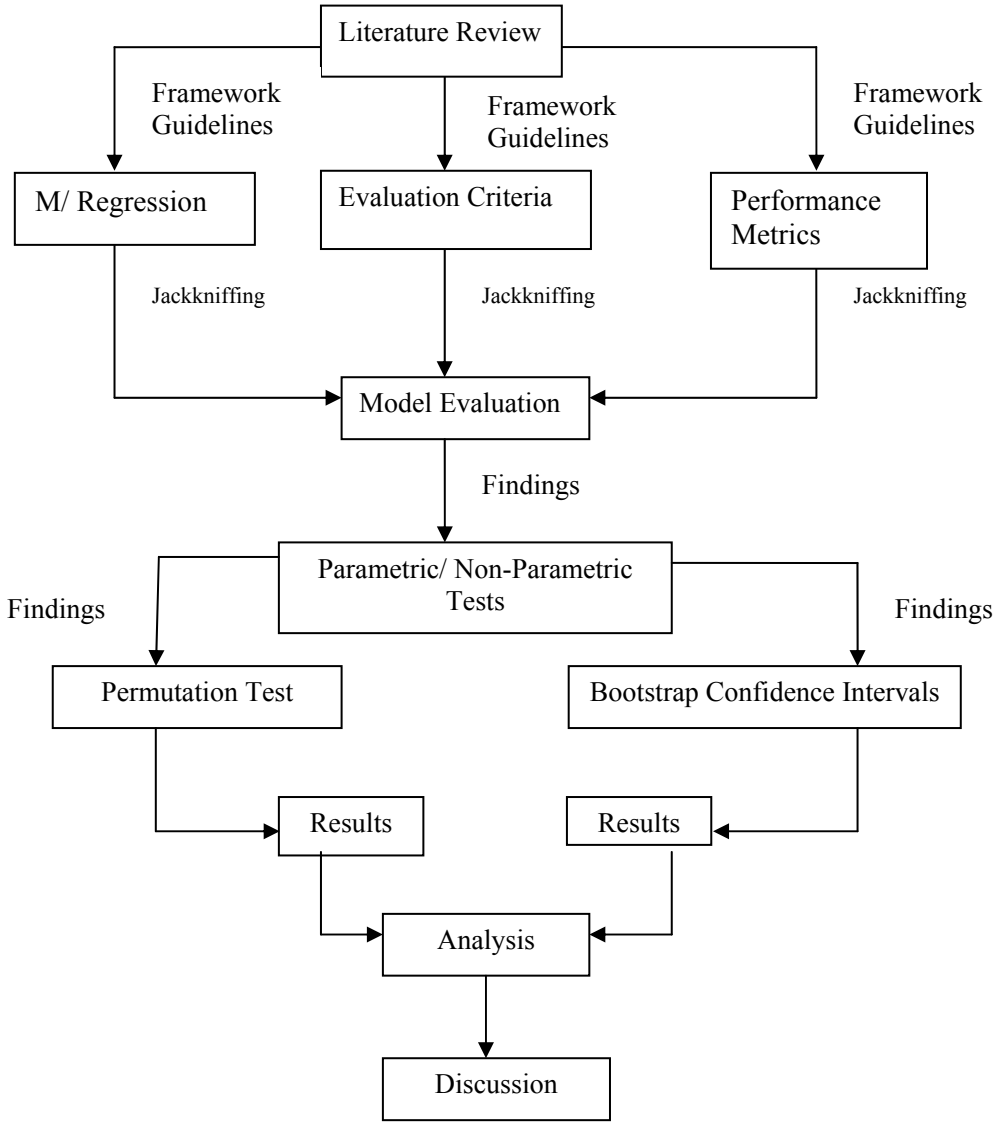


Figure 3.1 Proposed Research Methodology

Above Figure shows that literature review will result in framework guideline for quantitative study. Moreover multiple regression was selected for model development, a model evaluation criterion was defined and performance metrics were selected. Using framework guidelines from Liu et al [38] the data set will be prepared and by applying multiple regression and defining evaluation criterion and giving performance metrics to jackknife procedure the models will be evaluated. The results from model evaluation will be tested statistically through parametric and non-parametric tests. The findings from parametric and non-parametric tests will be then evaluated by two re-sampling tests i.e. permutation tests and bootstrap confidence intervals. The results from both tests will be analyzed and the superiority of regression and estimation by analogy will be observed. Finally conclusion and discussion on results will be performed.

3.2 Quantitative Part of Research

Quantitative part of our research fulfilled the gap identified in qualitative part of research. It helped us in order to answer the following research questions;

RQ3. Does the estimation reliability increases with the pre-processing of the project data statistically.

RQ4. Are the accuracy indicators statistically significant?

RQ5. What other measures can be used for the evaluation of effort estimation models?

The criterion which was identified as a result of qualitative research was put into use on ISBSG data set release 10 in using various statistical techniques. The data set was prepared for empirical studies according to the framework proposed by Liu et al [38]. A step wise ANOVA was used to identify the strong bond between dependent and independent variables in order to build N-variable model. Estimation by Analogy (EbA) was used as non-parametric and Ordinary Least Squares (OLS) as parametric estimation models for comparison. These models were evaluated against different metrics which helped us in identifying various measures for estimation consistency and estimation accuracy.

RQ3, RQ4 and RQ5 are answered in the empirical part of the quantitative research. RQ3 is answered while comparing our results to the previous work done in this area (see Section 6.4 Table 6.19). RQ4 is answered during the process of the statistical significance of the accuracy indicators (see Section 6.4 Table 6.15). The answer for RQ5 is provided in Section 6.4 Table 6.8 while discussing the usage of inferential statistics for the evaluation of effort estimation models.

4 BACKGROUND

This chapter discusses the results of the thorough literature review on the evaluation of effort estimation models as well as the problems associated to evaluation. It discusses the effort estimation mechanism, calibration, and performance measurement and evaluation criteria.

4.1 Software Cost and Effort Estimation

Making an estimate before starting of any project is necessary to be able to plan and manage any project. Estimate is an intelligent guess for the project resources. Nowadays, software has become a major contributor to economic growth for any country. Building software is evolved to an engineering a software product and thus making an estimate before starting any software project vital for the project managers and key stakeholders. Major project milestones like project schedules, budgeting, resource allocation and project delivery dates are set on the effort and cost estimates. Thus the reliability of the estimation leads project to success or otherwise failure.

Yan-Chin and McDevitt [4] stated that; effort estimation models provide managers an estimate of the required development effort and time. This estimate should be obtained at the maximum productivity level of the development team effort and at minimum time to complete the project and thus minimizing the overall cost of the project. Leung and Fan [1] extracted the general predictors of the cost and effort estimation methods as:

- Effort (usually determined in person-months)
- Time (in months)
- Cost (dollars)

The cost estimation models work primarily on an effort estimate, which is later converted into required duration and cost for the project. As the process of estimation is a quantifiable process, these predictors have common measuring units; effort is mainly calculated in person months, time in months and cost usually in money units.

4.1.1 Reasons Which Make Software Cost Estimation Challenging

The future of the project from the business point of view and the integrity of the project manager are closely related to reliable estimation. The project manager has a good reputation if the project completes within budget and time as per estimates provided by him. There are various reasons which make estimation difficult [1, 8, 14].

- Lack of historical data base of cost measurement
- Estimates are often done hurriedly
- Clear, complete and reliable specifications are difficult to formulate at the start of the project
- There are many inter related factors which affect effort and productivity but its relationship is not well understood
- Estimation is usually done by inexperienced estimator which lack expertise knowledge.
- Large number of cost drivers which affect estimation
- Rapid change in the development environment
- A miss-consumption of linear relation between the required capacity per unit time and the available time as stated in Brooks law adding more people to a late project only makes it later" [18].

- Premature requests, dynamic evolution environment and highly complex process make estimation even more difficult [15, 16]

4.1.2 Estimation Process

Once a project has got the approval to move beyond the Concept stage, detailed planning is required to set out the tasks and activities. Cost estimation is usually done in the planning phase. In planning stage the following steps are observed [1].

- Goals are set out which must be fulfilled upon the successful implementation of the required cost estimation model.
- Make feasibility of the project keeping in view the available resources
- Prepare software requirements specification (SRS) document
- Document all hardware and other technical specification of the software system.
- Generate different estimates using multiple cost estimation techniques minimizing the risk of over or underestimation.
- Compare these estimates keeping in view the available resources and iterate them.
- Once the project is started, monitor it in each and every single step according to the planned costs and feedback to the management.

The basic input to any cost estimation model is size. Software size is the most important factor that affects the accuracy of estimation [1]. Most of the software cost estimation models work in two stages [8], the sizing and the productivity adjustment factor stage. Figure 4.1 and Figure 4.2 shows the working mechanism of cost estimation models. Some models like Albrecht's Function Point Analyses (FPA) [25] is more focused on sizing stage while other well-known models like COCOMO [23] and SLIM [24] on productivity stage.

As shown in Figure 4.1 [8], in the first stage an estimate of the size of the product is obtained using any appropriate sizing technique. Usually size is measured in source lines of code (SLOC) or Function Point (FP). Later on in the second stage; time and effort is estimated to develop the software of the estimated size, a number of cost affecting factors usually known as cost drivers, are identified and their weight is calculated, its effect is normally known as productivity adjustment factor. The effect of productivity adjustment factor is added to the estimate to have a more realistic estimate [8].

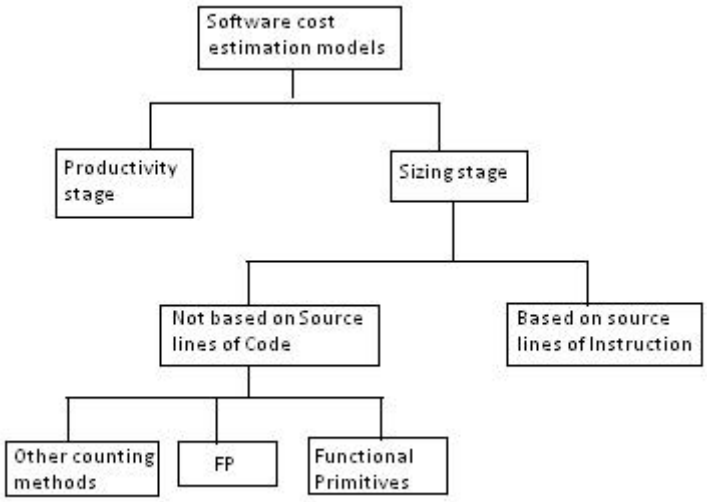


Figure 4.1 Two stages of Software cost estimation models (adapted [8])

Most of the tools implements software cost estimation models in the following seven steps (see Figure 4.2) [8].

- Creation of database of completed projects
- Size estimation
- Productivity estimation
- Phase distribution
- Sensitivity and risk analysis
- Validation
- Calibration

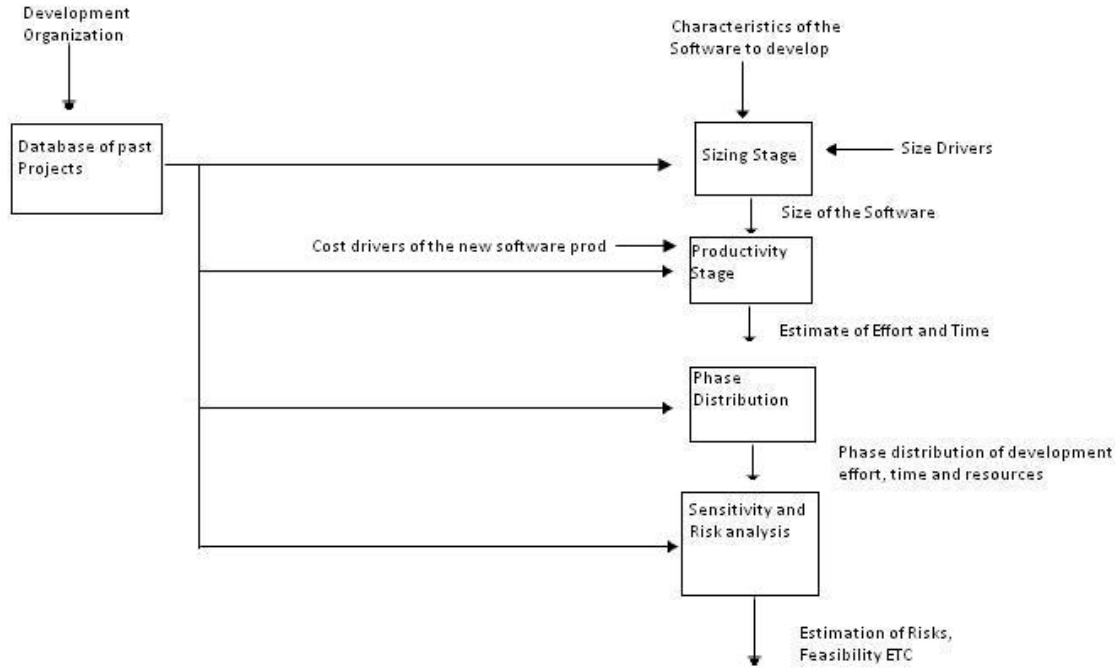


Figure 4.2 Structure of the cost estimation models (adapted [8])

In Figure 4.2 [8] five components of the general structure of the cost estimation model are given. Besides size and productivity stage phase distribution and sensitivity and risk analysis is different. The total effort and duration from the productivity stage are divided in phases and activities of the project in phase distribution. Experience from the past completed projects is very necessary for the phase distribution in which resources are allocated to particular activities and phases of the project.

In sensitivity and risk analysis phase the risk factors of the project and sensitivity of the estimates to the cost drivers is determined. Existence of the past completed project is necessary. Using any model for the first time according to own setting, validation is necessary and calibration if needed. Most oftenly the environment and the project database on which software cost estimation model is being developed are quite different from the environment and project characteristics in which the model is to be used. In order to make validation and calibration database of the past projects is of utmost importance.

4.1.3 Software Cost Estimation Techniques and Tools

There are various cost estimation techniques and tools developed for cost estimation. Some are used individually while in some cases a combination of technique and tools is used for reliable effort estimate. These techniques and tools are [8];

- Expert judgment
- Analogue estimation
- Price-to-win technique
- Estimates based on available capacity
- Estimates based on the use of parametric models

In parametric models the development time and effort is estimated as a function of various variables (cost drivers) using algorithms [8]. Later, these tools and techniques were classified as algorithmic and non-algorithmic methods [9]

A) NON-ALGORITHMIC METHODS

Expert judgment: Estimates are based on the judgment of expert persons of the organization. It is used when project is in the initial phase, planning, to give an estimate of the effort and time keeping in view the completed projects. Despite of its usefulness expert judgment have some drawbacks [1, 8]

- Depends highly on expert opinion.
- Very difficult to reproduce and use the knowledge and experience of an expert
- The estimation is not repeatable and means of deriving an estimate are not implicit.

Estimation by analogy:
As its name indicates, this estimation depends on database of similar historical projects or modules of the projects [8].

Price-to-win technique:
Commercial motives play an important role in this technique. Cost is estimated on customer affordance instead of its functionality [1].

B) ALGORITHMIC MODELS

Algorithmic models and tools can be broadly categorized as empirical models and analytical models [1, 8, 20]. An empirical model requires data from the past projects to evaluate the current project and derives the basic function for the current project by analyzing the particular projects from the database while analytical models uses formulae on global assumptions like the rate at which developer solve a problem and the number of available problems [1, 20].

Some of the well-known algorithmic cost estimation models are described below.

Constructive COst MOdel (COCOMO81): Presented by Barry Boehm [9] and it was considered to be the most transparent and documented model so far. This is an empirical model which is mainly focused on determining the effect of different cost drivers on development effort [8]. COCOMO does not support sizing stage and gives several equations which represent relation between effort and size and between effort and development time [8]. COCOMO is of the exponential form and uses linear regression for calibration. We will develop exponential model and will identify the values of its exponents through linear regression.

The general form of COCOMO model is,

Effort = a * $(size)^{b}$ such that a and b is the complexity or effect of other cost factors

COCOMO works in three modes [9]

a) *Organic:* when the development is stable and less innovative and the size is relatively small, Organic model is used.
b) *Embedded:* when the requirements are evolving and complex and the development is innovative and within tight schedule, embedded version is used
c) *Semi-detached*: when the requirements are between organic and embedded, semi-detached version is used.

COCOMO II: COCOMO II is the newer version of COCOMO 81 with addition of more cost drivers and also support for project planning. The detail of cost drivers is given below.

- Development Flexibility
- Architecture or Risk Resolution
- Team Cohesion
- Process Maturity

Putnam Model: This is analytical model which works on global assumptions [24]. The tool which implements Putnam model is SLIM. SLIM is based on the work of Norden/ Rayleigh [46] manpower distribution curves.

The equation which implements Putnam model is called 'software equation'; it is of the form,

$$S = C_k * (Effort)^{1/3} * t_d{}^{4/3} \quad \text{.....I}$$

Ck is technology constant which shows the effect of language and development platform on effort.

$$Effort = D_0 * (t_d)^3 \quad \text{.....II}$$

Where D_0 is the magnitude of the difficulty gradient or manpower-build up

From I and II

$$Effort = \left(D_0{}^{4/7} * E^{-9/7}\right) * S^{9/7} \text{ And}$$

$$DevTime = \left(D_0^{4/7} * E^{-3/7}\right) * S^{3/7}$$

There are other models such as ***Price-S*** [41] (Programming Review of Information Costing and Evaluation-Software) model which was developed by RCA PRICE systems. Like Putnam SLIM [24] the underlying concepts of the model is not publicly available but the important thing is that US Department of Defense demands a PRICE estimate for all quotations for a software project.

Estimacs [43] developed by H. Rubin takes an input in the form of a set of questioner consisting questions of the complexity of the user-organization and size of the software to be developed. The underlying mathematical assumption of the model to convert input into effort estimates is publicly not available.

SPQR-20 stands for Software Productivity, Quality and Reliability developed by C. Jones [42], claims that the model is enough sufficient to estimate duration, effort and cost to develop the software. It is also capable to estimate maintenance cost. It uses FPA to size the volume of the project.

4.1.4 Strengths and weaknesses of cost estimation models

The table below explains the strengths and Weaknesses of the cost estimation models [1, 8].

Methods	Strengths	Weaknesses
Non-algorithmic models		
Expert judgment	Expert with the relevant experience can provide good estimation; Fast estimation; Estimates are qualitative; Estimate time and effort in early phases	Depends on the knowledge of 'Expert'; Estimation biasness; Can not be iterated;
Analogy	Based on actual project data; Past project experience is needed; Works on reasoning	Difficult to find projects with same characteristics; historical data may not be accurate
Parkinson Price-to-win	Often win the contract; Greatly motivated;	Might experience overruns; No empirical estimation; Poor practice
Top-down	System level focus; Faster and easier; Estimation is derived from the global characteristics of the product; Require minimal project detail	Little detail for justifying estimates; No level of project granularity; Less accurate than other methods
Bottom-Up	Based on detail analysis; Support project tracking better than other estimates; project requirements are defined at low level of granularity	Estimation is difficult at early lifecycles of development; require more estimation effort
Algorithmic	Results are more objective; Can be iterated in different lifecycles; better understanding of the estimation methods	Subjective inputs; calibration problem; Past projects are necessary; Algorithms are company specific; not useful other than its own environment

Table 4.3 Summary of strengths and weaknesses of different cost estimation models.

4.2 Size Measurement

Software size measurement is the first and most complex activity in estimation. It is very difficult to measure the software in early phases of abstraction. Size is the basic input to cost estimation models hence reliable and accurate estimation depends solely on accurate and reliable size measurement. Here we will use UFP with quality rating A [74] projects which ISBSG suggests reliable for estimation. UFP includes IFPUG FP, NESMA FP, Albrecht FP and MARK II. Laird and Brennen [44] showed the importance of measuring accurate size;

- Contracts with customers and employment of contractual staff depends upon size
- Size shows the volume of the software
- Effort is calculated from size

Measuring software size involves either [44];

a) physical measurement of the software
b) measuring the functionality of the software

Physical measurement of software comprises of a) measuring the length of code and b) measuring the length of specification and design [44]. SEI provided a standardized framework for measuring the length of software [44].

Measuring software in terms of the functionalities was first introduced by Albrecht [25]. FPA is a cost estimation model which works on function point as a basic input to measure productivity. Function point is promoted by International Function Point Users Group (IFPUG) [45].

4.3 Model Calibration Using Linear Regression

Calibration is the fine tuning of estimation model according to the local available data and environment. According to [8] software cost estimation model is calibrated according to own development environment and project characteristics and also one of the component of software cost estimation structure.

The general power function model is of the form:

$$Effort = a * (S)^b$$

Taking log on both sides

$$\log(Effort) = \log(a) + b\log(S)$$

Suppose;

$$Y = \log(Effort),\ A = \log(a),\ X = \log(S)$$

$$Y = A + \beta X$$

Such that;

Y = Dependent variable (Effort)
X = Independent variable
B= coefficient of regression

Applying the standard least square method to a set of previous project data {Yi, Xi: i =1, …, k}, we obtain the required parameters b and A (and thus a) for the power function [1].

4.4 Measurement of Model Performance

Estimation models are always evaluated for accuracy and consistency measurement [7, 28]. Modeling is associated with measurement [47]. Measurement of the accuracy and consistency of the models is done for assessment and prediction [49]. Common and agreed standards are necessary for model assessment [28] that is why a model being developed in a particular environment does not show same accuracy and consistency for different development environment [51].

Various statistical methods are used for the measurement of accuracy and consistency of the model.

4.4.1 *Measures of Accuracy:*

Accuracy of the estimation model shows the closeness of the estimated result to its actual value [52]. Accuracy is achieved either by difference of the estimated and actual value or by ratio [28]. Both measures are used for model evaluation.

Difference of measure of accuracy [28];

a. Mean of absolute error (MAE)
b. Root mean of square errors (RMSE)
c. Coefficient of determinant (R^2)
d. Mean of residues (MR)

Ratio of Measures of accuracy [28];

a. Mean (average) of relative errors (ARE)
b. Mean of magnitude of relative errors (MRE)
c. Root mean of squared relative errors (RMSRE)
d. Prediction at level *l* (PRED(*l*))

4.4.2 *Measure of Consistency:*

According to [51] all estimation models are sensitive to productivity measures and the model will always generate unreliable estimates whenever it is used in an organization whose productivity rate is different from the environment where the model is developed. Therefore consistency is another measure the model will possess.

Mukhopadhyay [51] suggested some measure for consistency of the model; Correlation coefficient of estimates and actual values (SDR).
This measure tests the linear association between actual and estimated values. Statistically, standard deviation is used as measure of estimation consistency.

4.4.3 *Regression Criteria*:

Regression criterion is a technique used to determine model parameters [28]. A statistical method least square (LS) [52] is used to determine model parameters or we can say regression is used to determine the exponents (exponential model) of the model being developed during over studies.

4.5 Criteria for Model Evaluation

Barry Boehm [9] and Andolfi [48] defined certain criteria for cost estimation model evaluation. These are discussed in the following paragraphs:

Functionality:

Algorithmic models which use functions to estimate effort and duration are evaluated under functionality tree. The functions are evaluated for;

a. Calibration
b. Risk analysis
c. Constraint
d. Details of estimation (system level and activity level details)
e. Size estimation

Reliability:

Reliability is the behavior of the model which it corresponds to different situations. It is evaluated in terms of;

a. Fidelity (average absolute relative error, AARE, weighted average relative error, WARE, PRED $(k)_3$)
b. Stability

c. Relevance analysis

Usability:

It evaluates the usage of characteristics like;

a. Understandability
b. Cost factors quality
c. Ease
d. Documentation

Portability:

Portability of the model is evaluated as;

a. Domain
b. Operative environment

5 RELATED WORK

The software engineering literature is in rich of the cost and effort estimation techniques. It is extremely difficult to select appropriate model for estimation due its abundance in literature. Further more the mixed nature of the data from the past projects makes estimation extremely difficult and unreliable. Estimator has to deal with both linear and non-linear relationship of the dependent variable and independent variables.

For the last few decades a plethora of studies has been conducted to explore the exact relationship between response and explanatory variables. Numerous accuracy measures are proposed for the measurement of model predictive accuracy. Various methodologies are proposed to evaluate the effort estimation models and tools in order to explore the exact relationship of the nature of the data and the model or tool used to predict effort estimates.

According to Stephen [53], prediction is a probabilistic estimate or description of a future value, while estimation reliability is the probability that the estimation error becomes smaller than a threshold value [54]. The threshold can be given by some standard evaluation criteria such as Mean Magnitude of Relative Errors (MMRE), Median MMRE (MdMRE) etc. Generally the software cost estimation tools results in a single value of estimate which is used as a basis for further resource allocation and project related activities. Risk and uncertainty is inherited with point estimates and a slight change to the most significant variables makes a huge variation in estimates. Therefore, these accuracy indicators need to be statistically tested. Further more new approaches must be used for model predictive accuracy evaluation.

Prediction Intervals (PI) which provides optimistic and pessimistic guess for the true magnitude of the cost is the best alternative [55]. Regression analysis (LS) and analogy based estimation (EbA) are the two most widely used cost estimation techniques [56]. Estimates made by regression face the problem of uncertainty, risk [56] and multicollinearity [57] while estimates made by analogy face the problem of precision [58].Numerous studies [12, 20, 21] showed that PI is more realistic approach for the counter measure of uncertainty and risk analysis.

According to Mair and Shepperd [56], the success of each prediction method is dependent on number of factors such as cost factors (size), adjustment factors (cost drivers), environmental factors and data set characteristics [60]. Myrveit and Stensrud [62] concluded that results are sensitive to experimental design. Similarly, Briand et al. [70] found analogy based prediction systems were less robust than regression models when using data external to the organization for which the model is built. In contrast, Mendes and Kitchenham [61] suggest that CBR predicts better across a large heterogeneous data set, and regression is better for within-company predictions.

Finnie, Wittig and Desharnais [68] claimed that due the complexities involved in software development projects, regression models were less effective than analogy-based prediction systems which benefit from human judgment and intuition. Shepperd and Schofield [63] found that analogy gave better results than regression in terms accuracy. Angelis and Stamelos [64] also found analogy-based methods superior.

Mendes et al. [65] and Jeffery et al. [66] found that stepwise regression outperformed CBR. However, in later work, Mendes [67] claimed that CBR gave better prediction accuracy. Niessink and van Vliet [71] proposed that analogy offers an alternative to regression models when applied to a heterogeneous data set. With the passage of time new approaches are proposed for reliable and consistent predictions. These include machine learning approaches like Artificial Neural Networks (ANN) and Case Base Reasoning (CBR) [69].

Almost all models were evaluated for estimation accuracy and consistency. Several measures are proposed in literature which are the function of the predictive error [88] measured by appropriate method. Single measure of error is just a statistic and may contain significant variability. Study on model evaluation showing one common problem that these evaluations are done without prior concise data analysis and the significance of the accuracy indicators is not tested.

Another study from Shepperd [56] on regression and analogy based techniques proved that twelve out of twenty studies proved the superiority of EbA technique and only six papers used hypothesis testing instead of using conventional accuracy measures was used to evaluated the model predictive accuracy. Four out of six proved the superiority of regression.

The formal testing is overlooked in favor of simplicity and models are solely evaluated in the light of a single measure with risk. Therefore models should be evaluated with more inferential statistics such as confidence intervals and hypothesis testing [88]. Therefore, keeping in view the above problems we will evaluate regression and EbA models in the light of their accuracy indicators, the significance of the accuracy indicators will be tested statistically and more inferential statistics will be used identify the superiority of either methodology.

6 EMPIRICAL STUDY

In the following sections we empirically evaluate effort estimation models. These models are broadly categorized as LS and EbA. We evaluate these models in the light of accuracy indicators. We further identify the significance of the accuracy indicators statistically and will observe the difference between various accuracy indicators.

In this thesis study, we conducted our empirical study in two phases. In phase I, we identified variables which are linearly related to predictor variable, this serves as LS set and the remaining set of variables form EbA set.

6.1 Data Collection

One of the vital issues for the construction of the efficient and reliable cost estimation models is the unavailability of the past project data. International Software Benchmarking Standard Group (ISBSG) [74] is an organization which collects data from almost all over the world for the last 10 years period on the basis to help and improve different project management, software development and engineering research purposes [74].

The ISBSG R10 dataset was released in 2007 and contains data of 4106 completed projects. The project corpus contains data from more than 20 countries across the globe covering a huge cross section of projects from various organizations and of various domains [82]. The project database contains 70% of the projects being less than 6 years old and is of multi-organizational nature [82]. There are round about 22 countries which contribute to ISBSG data set, some of the main contributors are USA 26%, JAPAN 28%, AUSTRALIA 24% NETHERLAND 7% and so on [82]. Figure 6.1 shows the various contributors to the ISBSG repository.

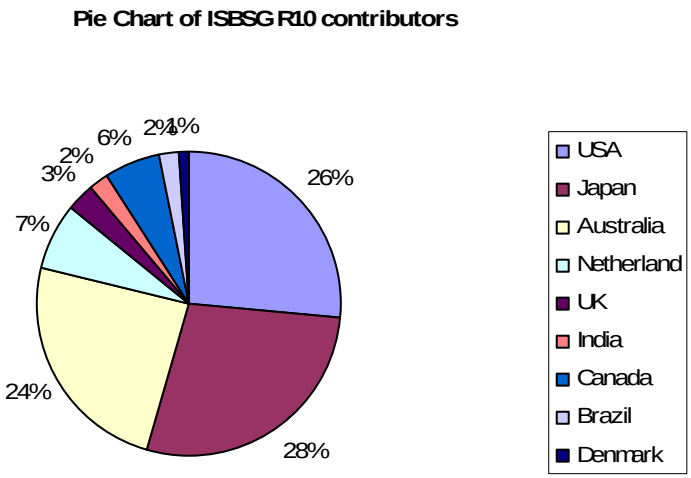


Figure6.1. ISBSG R10 Contributors

Table 6.1 [74] shows the project attributes description which is relevant to this thesis.

Table 6.1 Project attributes description

Project Att:	Description	Type	Detail
Size	Functional size(Function points)	Ratio	The unadjusted function point count. This may be reported in different units depending on the FSM method.
Effort	Summary work effort	Ratio	Provides the total effort in hours recorded against the project
Duration	Project elapsed time	Ratio	Total elapsed time for the project in calendar months
Count App	Count approach	Nominal	A description of the technique used to size the project.
DevType	Development type	Nominal	It describes whether development is a new development, enhancement or re-development
OrgType	Organization type	Nominal	It describes the type of organization that submits the project data.
BusiType	Business type	Nominal	The type of business area being addressed by the project.
AppType	Application type	Nominal	The type of application being addressed by the project
ArchType	Architecture	Nominal	It shows that whether application is stand alone, multi tier or client server
DevPlat	Development platform	Nominal	It defines the primary development platform as PC, Mid range and main frame
LangType	Language type	Nominal	It defines the language type used for the project e.g. 3GL, 4GL, Application Generator (AG)
PriLang	Primary language	Nominal	The primary language used for the development, e.g. JAVA, C++ etc.
MDBS	Main database used	Nominal	It defines the main database technology used for development
Umethod	Used methodology	Nominal	It defines whether a development methodology was used by the development team.
Rmethod	How methodology acquired	Nominal	It shows whether the development methodology was purchased or developed in-house
Rlevel	Resource level	nominal	It defines about the people whose time is included in the work effort reported. 1 = development team effort 2 = development team support 3 = computer operations involvement 4 = end users or clients

Some of the major types of organizations which provide data are; communication 18.4%, banking 10.20 %, Insurance 13.80%, Public administration 15%, manufacturing 11.60% and 25.30% of the overall remaining organizations contribute less than 5%. Figure6.2 [82] shows various distribution of organizations contribution in ISBSG.

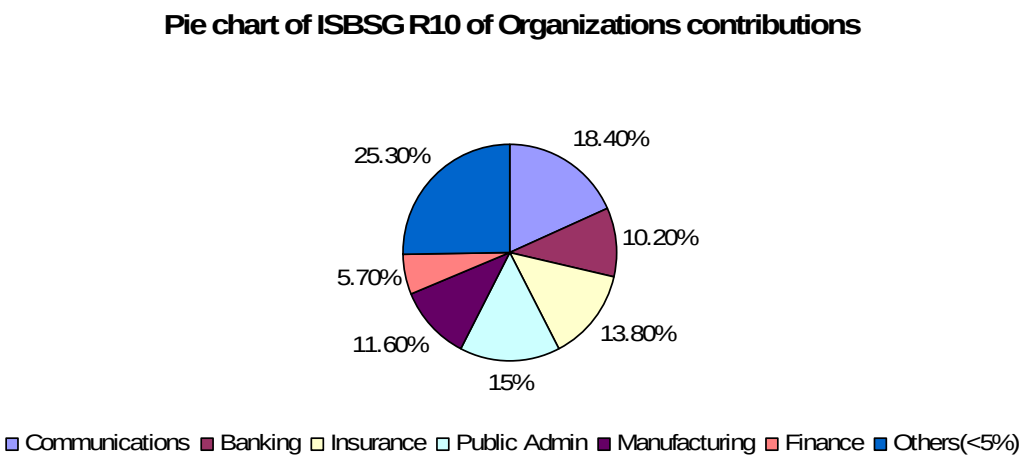


Figure 6.2 ISBSG R10 Organizations Contribution

6.2 Statistical Modeling Techniques and Tools Used

6.2.1 EbA

Estimation by analogy is a non-parametric procedure which is used to estimates the new project cost on the basis of similar projects in the project repository [3]. It uses Euclidean distance matrix to identify cases which have minimum distance between their weights. EbA is free from distribution assumptions and predict the cost of new project on the basis of analogies [3].

6.2.2 Ordinary Least Squares (OLS)

OLS regression analysis is applied in order to fit the data to specified models [70, 91]. OLS is applied only if the data is normally distributed. A linear relationship is established between predictor and predicted variable before applying OLS.

6.2.3 Linear Regression

Parametric estimation techniques assume that the function f (Xi) can be estimated by a linear regression [5],

$f(\mathrm{Xi}) = Xi\ \beta$ or

$Y = a + \beta\ Xi + C$

Where;
Y is dependent variable
X is independent variable
β is unknown parameter, regression co-efficient
a is Y intercept
C is constant

6.2.4 Analysis of variance (ANOVA)

ANOVA will be used to see the effect of independent variable on dependent variable [6] in order if the independent variable is categorical. The level of significance is 0.05. ANOVA can be used in two situations
1. One way ANOVA: ONE dependent and one independent variable
2. Factorial ANOVA: ONE dependent and more than one independent variable

6.2.5 Analysis of covariance (ANCOVA)

It is applied in a situation where we have numerical dependent variable and both categorical and numerical independent variables [6], or if we wanted to see the effect of both categorical and non-categorical independent variables at the same time. To observe the change in variation on response variable simultaneously by categorical and non-categorical variables ANCOVA is the best choice.

6.2.6 Stepwise ANOVA

Stepwise ANOVA technique is used to build multi variable model comprising of both categorical and non-categorical independent variables and non-categorical dependent variable [4].

6.2.7 MS Excel

MS Excel is used for preliminary data analysis. It includes built-in statistical functions like regression, ANOVA, correlation, covariance and much more.

6.2.8 Statistical package for social sciences (SPSS)

We will use SPSS [2] version 17.0 which is a very comprehensive tool for complex statistical analysis.

6.2.9 MatLab 7.0

MatLab 7.0 was used as a simulator to generate different kinds of bootstrap confidence intervals and permutation tests for the comparison of different effort estimation models.

6.2.10 Angel Tool

Angel [10] is a tool based on Case Base Reasoning (CBR) algorithm used to identify "*k*", nearest neighbor on Euclidean distance method applying jackknife procedure.

6.3 Empirical Study Conduct - Phase I:

We conducted the empirical study in two phases. In phase I, we identified two data sets; one for LS and another for EbA. We used Ordinary Least Square (OLS) Model as parametric and EbA as non-parametric model. Statistical tests are performed on ISBSG R10 in order to identify predictor variables which have linear relationship with response variable.

We used project development Effort as dependent variable and all other variables as independent variables. All those variables which show a linear relationship with response variable will make LS set and the remaining will make EbA set.

We used framework defined by Liu et all [38] to process data in order to identify the most significant variables. As mentioned by Angelis [90] that there is a need of statistical framework that identifies the predictor variables which form LS set. The framework is based on Maxwell's [77] work which applies a set of statistical approaches to publicly available data set. According to Liu et al [38] proposed framework, the data being processed by it, is significant at 95% for both parametric and non-parametric models.

6.3.1 Formation of the LS and EbA sets

Step0 Data Preparation

The maximum number of variables was selected initially following to the recommendation of Liu [92] in order to capture as much detail about these factors on response variable as possible. These variables include project effort, size, duration, organization type, development type, business type, application type, required resource level, methodology acquired, language type, primary language, database used, primary database, architecture type, count approach, used methodology and development platform.

Then, we performed a filtration process to prepare the initial set of data (see Table 6.2.).

According to ISBSG recommendation, the projects which are of quality rating[1] A or B were selected. Secondly, we selected projects measured by using IFPUG Unadjusted Function Points (UFP) which have quality rating A[2].

Since OLS cannot cope with missing values, we further eliminated projects with missing values. We eliminated those projects whose Base Functional Components, i.e. External Inputs, External Outputs, External Enquires, Internal Logical Files and External Interface Files are missing. Finally, 287 projects remained with 16 variables. The dependent variable was Effort and the remaining 15 were predictor variables, which are numerical or categorical.

Table 6.2 Filtration of ISBSG R10

Steps	Attributes	Filter	Projects excluded	Remaining projects
1	Data Quality Rating	= A or B	295	3811
2	Count Approach	QA= A	1672	2139
3	Size Attributes (Inputs, Outputs, Enquires, Files, Interfaces)	≠ Empty	867	1272
4	All Variables	≠ Empty	985	287

According to Liu et al [38] the data set will be scanned for both numerical and categorical outliers before using them in further steps. In the following step strategy for removing numerical outliers is defined.

Step1 Outliers Detection in Numerical Variables

There are many methods of outlier detection; Univariate, bivariate or multivariate [93]. We will use Univariate as we have only one predicted variable. The values in the numerical variables are converted to the standard scores with 0 mean and 1 standard deviation. For small samples <80 we will use standard score of 2.5 or 3 when sample size is >80. Those projects will be removed whose standard score is less than 3.

Table 6.3 summarizes variables with numerical values.

[1] According to ISBSG data quality rating is related to the integrity of the data and suggests only data of quality rating A or B to be used in experimentation.

[2] ISBSG suggests using only UFP of quality rating A due to its high integrity

Table 6.3 Detail of the Numerical variables in raw data set

Statistics			
	SIZE	EFFORT	DURATION
N	287	287	287
Mean	540.85	5819.84	8.59
Std. Deviation	1042.478	11889.466	9.149
Minimum	13	26	1
Maximum	13580	150040	84

Table 6.3 shows that project size ranges from 13 function points to 13,580 function points with an average of 541 function points, effort ranges from 26 man-hours to 150,040 man-hours with an average of 5820 man-hours and duration ranges from 1 month to 84 months with an average of 8 months. As we discussed earlier we use a standard score of 3 and will remove those projects whose standard score is less than 3.

According to Liu et al [38] the numerical variables will be checked for linearity in order to use them in further steps. They can be used in further steps only if they are normally distributed.

Step 2 Data Transformation.

We plot the histograms of the numerical variables. None of them were normally distributed (See Figure 6.2, Figure6.3 and Figure6.4).

Figure 6.2 Histogram of the Project Effort

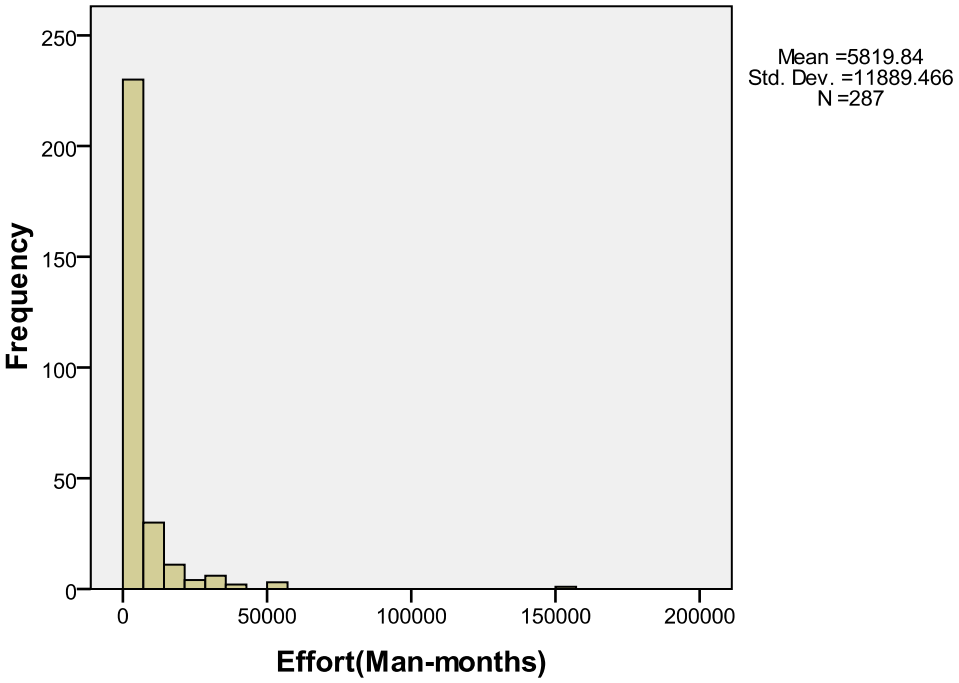


Figure 6.3 Histogram of the Project Size

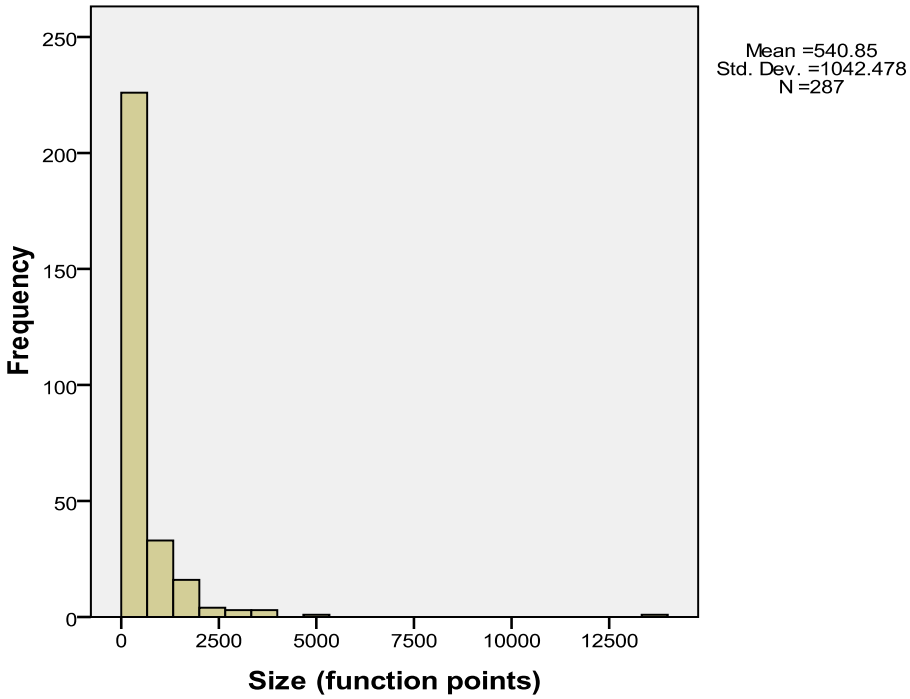
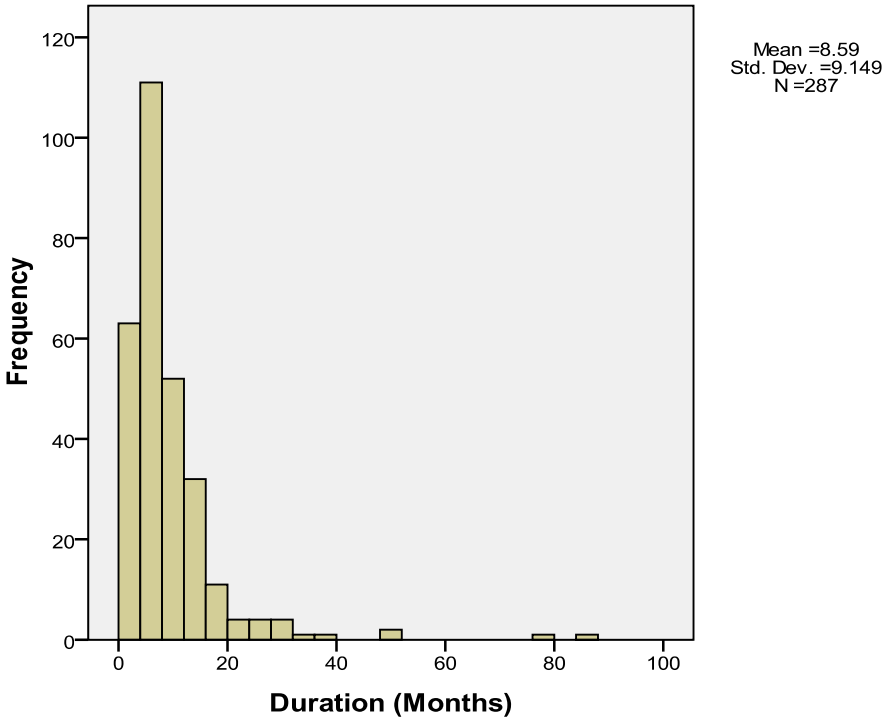


Figure 6.4 Histogram of Project Duration



The distribution shows a very high effort, big size and very long duration also it shows very small size, low effort and short duration. These kinds of distributions are very common for software project databases. The variation is due to very high effort, large size and duration or the project size may be very small with small effort and duration.

In order to approximate normal distribution, these variables are transformed to their natural logarithmic values. The new variables are LnEffort, LnSize and LnDuration. This transformation makes them linearise and bring them closer. The histograms of the

transformed variables are shown in Figure 6.5, Figure6.6 and Figure 6.7 which shows linear nature. We used transformed variables for further analyses.

Figure 6.5 Histogram of Project LnEffort in ISBSG R10 (raw data set)

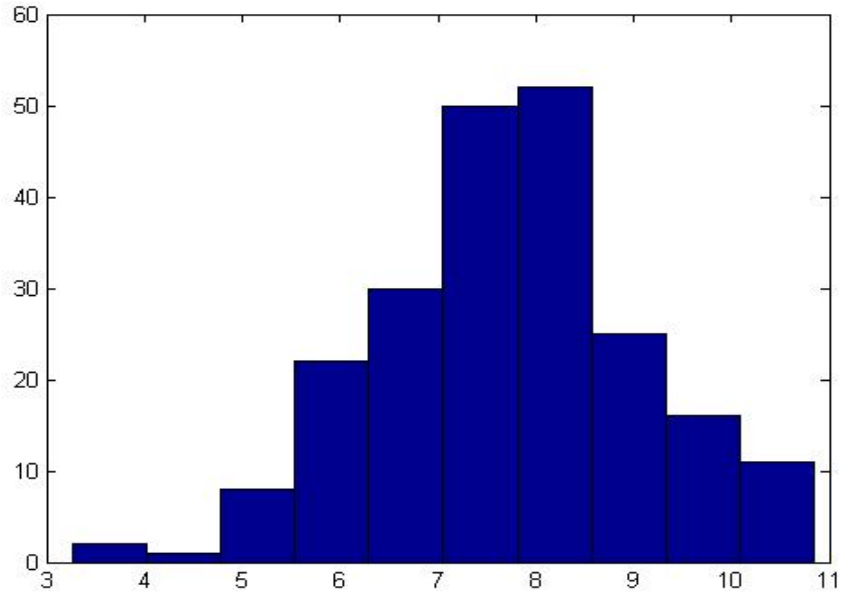


Figure 6.6 Histogram of LnSize in ISBSG R10 (raw data set)

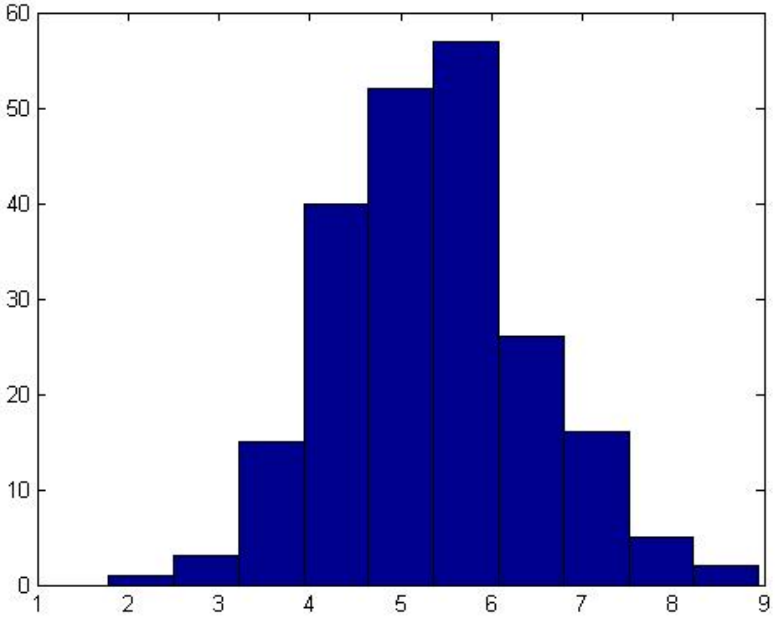
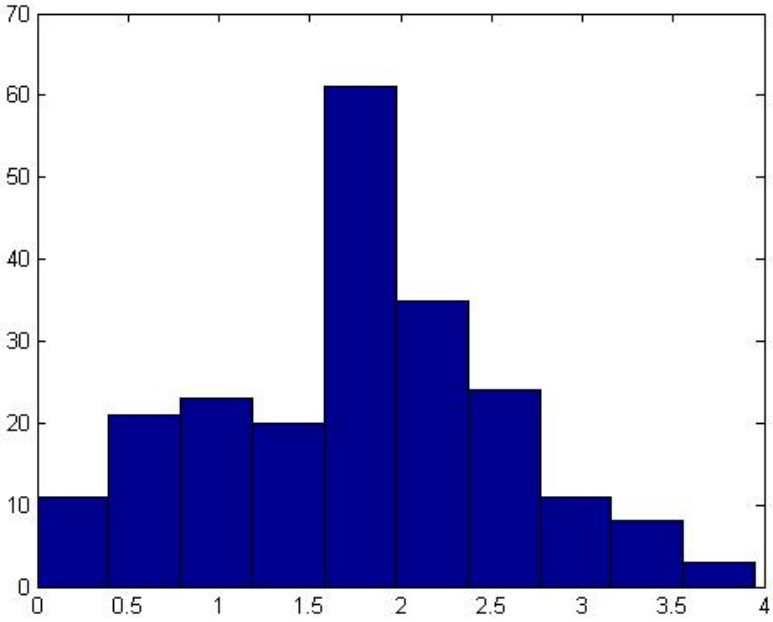


Figure 6.7 Histogram of LnDuration in ISBSG R10 (raw data set)



The normal Quantile – Quantile (Q-Q) plots of the transformed numerical variables are shown in Figure 6.8, Figure6.9 and Figure6.10.

Figure 6.8 Normal Q-Q plot of LnEffort

.

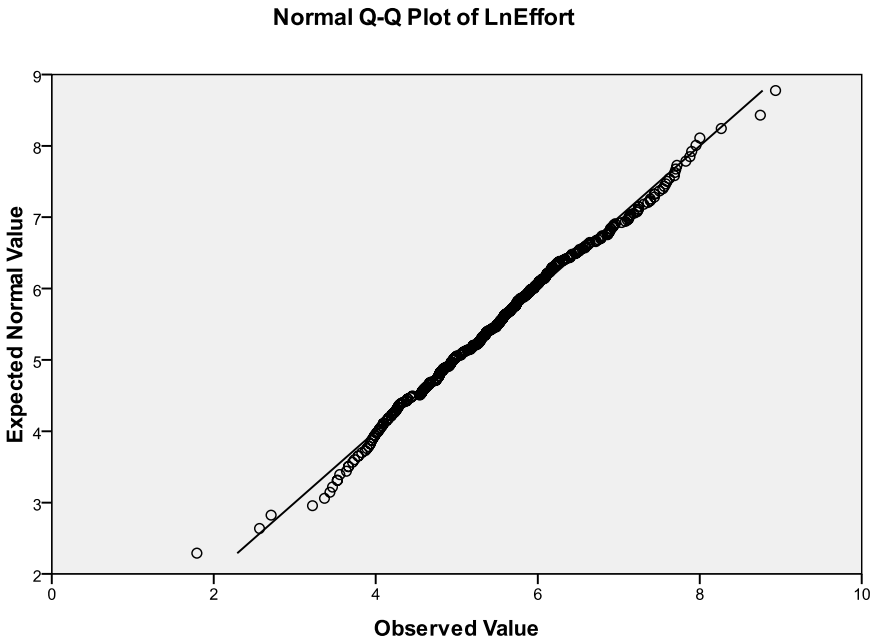


Figure 6.9 Normal Q-Q plot of LnSize

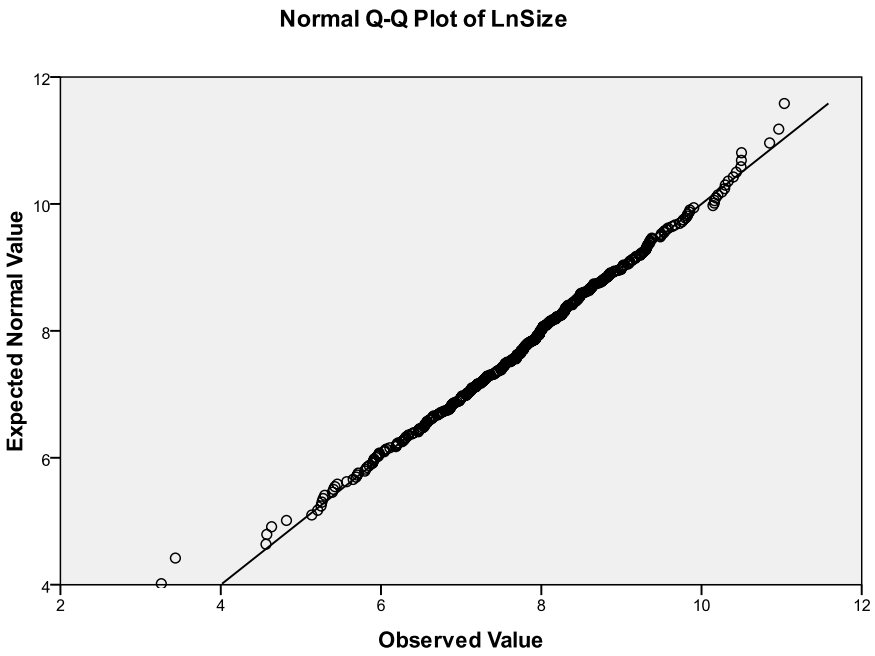
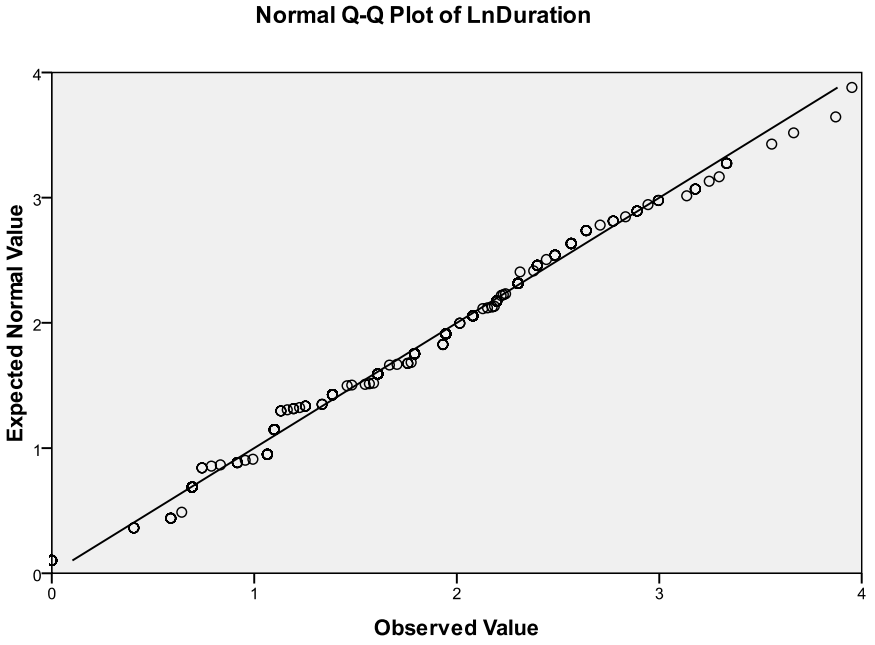


Figure 6.10 Normal Q-Q plot of LnDuration



Categorical outliers must be removed from the dataset in order to get more realistic estimates. The presence of outliers in the data makes estimate unreliable. Liu et al [38] summarized the procedure of categorical outliers based on other researchers work. In the following step the detailed procedure for categorical outlier removal is given.

Step 3 Outliers Detection in Categorical variables

According to [77] categorical variables are transformed to their numbers and each category which appears less than 3 is identified as outlier and should be removed from the observations. The remaining 13 categorical variables including organization type, development type, business type, application type, required resource level, methodology acquired, language type, primary language, database used, primary database, architecture type, count approach, used methodology and development platform are transformed into numbers. These numbers shows the frequency of the particular category in the main attribute. These transformed variables are shown in Appendix I. According to [77] those categories which appear less than three times is identified as outlier and removed from all tables.

Finally, after removing the numerical and categorical outliers, 217 projects remained, which was used for our final model building process in order to predict new effort.

Table 6.4 Filtration of ISBSG R10 based on Categorical variables

Step	Attribute	Filter	Project excluded	Remaining projects
1	13 categorical variables	Frequency<3	70	217

The relationship between numerical variables must be identified in order to use them in one model. In the following step Collinearity between numerical predictor variables is identified.

Step 4 Collinearity between Numerical Predictor Variables

In order to check the independent nature of the transformed numerical variables we used 2-dimensional plot diagram which shows the relationship among numerical predictor variables (see Figure 6.11, figure 6.12 and figure 6.13).

Figure 6.11 Scatter Plot between LnEffort and LnSize

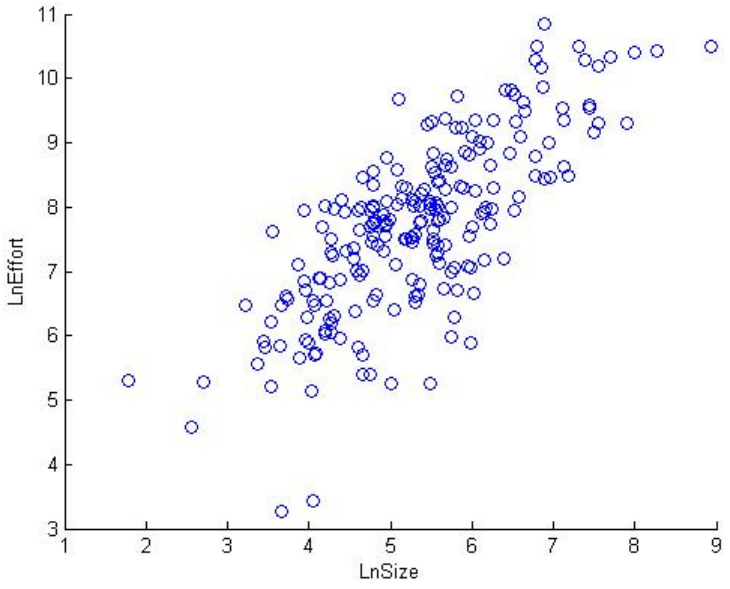


Figure 6.12 Scatter Plot between LnSize and LnDuration

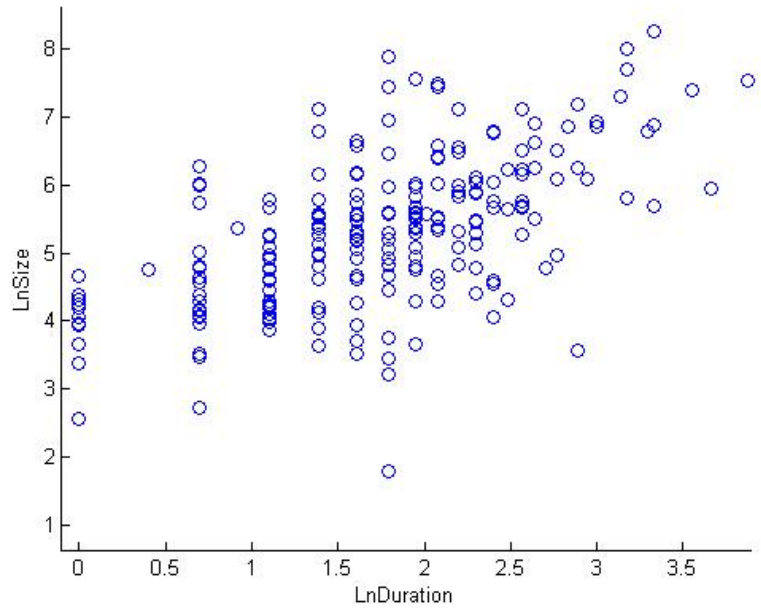


Figure 6.13 Scatter Plot between LnEffort and LnDuration

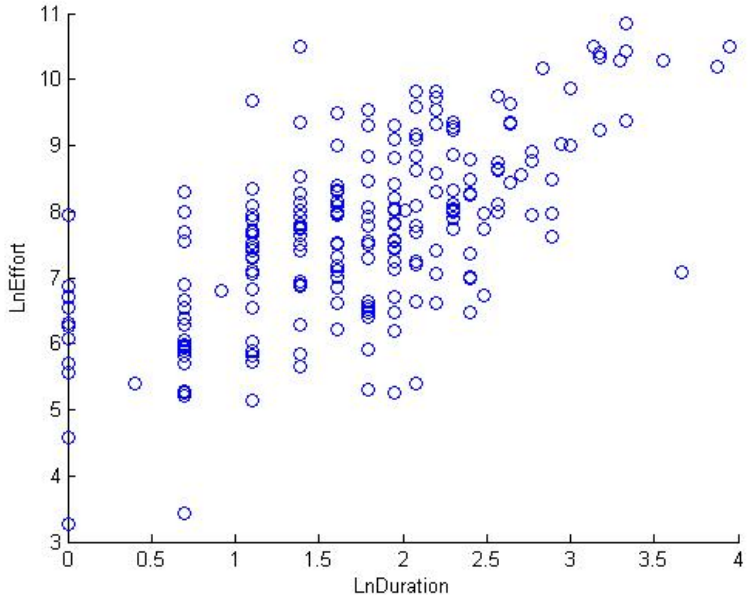


Figure 6.11 shows a strong relationship between LnEffort and LnSize. In Figure 6.12 LnSize and LnDuration shows strong bond. It means a slight change in one variable makes a considerable variation in another therefore both LnSize and LnDuration can not be used in a single model. Figure 6.13 shows relationship among LnEffort and LnDuration.

Collinearity test shows the relationship among predictor variables. A strong relationship among variables shows the dependent nature of the variable and cannot be used in a single model and vice versa. Pearson correlation coefficient is used for correlation if the data is normally distributed. Two variables can be perfectly related to each other if they have weak relationship [77].The correlation coefficient between two numerical variables less than 0.05 shows a weak relationship and can be used in a model [77].

Table 6.5 shows the correlation coefficients between all numerical variables. The correlation coefficient between LnEffort and LnSize is 0.758 and significant and 0.01. This shows strong relationship between LnEffort and LnSize, i.e. LnEffort is dependent on Size. Also, LnSize and LnDuration are strongly related on each other, and therefore, they cannot be used in a single model.

Table 6.5 Correlation coefficients between LnEffort and LnSize and LnDuration

		LnEffort	LnSize	LnDuration
LnEffort	Pearson Correlation	1	.758**	.653**
	Sig. (2-tailed)		.000	.000
	N	217	217	217
LnSize	Pearson Correlation	.758**	1	.598**
	Sig. (2-tailed)	.000		.000
	N	217	217	217
LnDuration	Pearson Correlation	.653**	.598**	1
	Sig. (2-tailed)	.000	.000	
	N	217	217	217
**. Correlation is significant at the 0.01 level (2-tailed).				

Step 5 Regression Analysis

Regression analysis is used in order to find the effect of the predictor variable on predicted variables. It determines the amount of variation caused by independent variables in response variable. In other words we can say it determines the most significant predictor variable. LnSize makes a variation of 57% alone in LnEffort and is the most significant variable. Table 6.5 shows the regression analysis between LnEffort and LnSize.

In the following steps we will use Stepwise ANOVA to identify those variables which cause the most variation in predictor variable.

Step 6 Stepwise ANOVA to build k-variable Model

Stepwise ANOVA is used to build a *k-variable* model. Since we have one predicted variable and 15 predictor variables therefore we will iterate the process of building one variable model up to 15 times until we had significant variable. This variable will be kept in the model and the best two variable model will be identified by iterating the process up to 14 times and the process will be continued until a fair combination of the variables are added and no more improvement is seen. The level of significance will be identified by the adjusted correlation coefficient square (R^2) in the model.

Linear Regression will be used if the predictor variables are numerical; ANOVA will be used in the case of categorical variables and ANCOVA in case of both numerical and categorical independent variables.

I. Identify Best One variable Model.

In order to identify the best one variable model which will cause the most variation in LnEffort, we need to build 15 one variable model and will select the most significant variable which makes the most variation in LnEffort. Table 6.6 is the most significant one variable model with LnSize which made 57% variation in LnEffort. The details of the other one-variable models are given in Appendix II.

Table 6.6 Best One Variable Model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.758[a]	.575	.573	.880023971382517

a. Predictors: (Constant), LnSize

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	224.928	1	224.928	290.439	.000[a]
	Residual	166.505	215	.774		
	Total	391.433	216			

a. Predictors: (Constant), LnSize

b. Dependent Variable: LnEffort

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.855	.291		9.800	.000
	LnSize	.912	.053	.758	17.042	.000

a. Dependent Variable: LnEffort

II. Identify Best Two Variables Model.

Keeping LnSize in the model we identified the second most significant variable which will make the most variation in LnEffort. By running linear regression between numerical variables and ANCOVA, Primary language found to be causing the most variation in LnEffort, i.e. both LnSize and Primary Language caused variation of 68%.

The model is given below in Table 6.7 while rest of the models is provided in Appendix II.

Table 6.7 Best Two Variable Model

Tests of Between-Subjects Effects

Dependent Variable: LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	274.162[a]	16	17.135	29.223	.000
Intercept	53.097	1	53.097	90.555	.000
LnSize	187.972	1	187.972	320.578	.000
PriLang	49.234	15	3.282	5.598	.000
Error	117.271	200	.586		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .700 (Adjusted R Squared = .676)

III. Identify Best Three variable Model

Keeping LnSize and Primary Language into the model we will identify the best three variables Model which will cause the most variation in LnEffort. Since we have already one categorical and one numerical variable in the current model therefore, we will run only ANCOVA tests in order to identify best three variables Model i.e. LnSize, Primary Language and Organization type which caused 70% of the total variation in LnEffort.

Table 6.8 shows the best three variables Model. The rest of the three variable models are given in Appendix II.

Table 6.8 Best Three Variable Model

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	298.658[a]	43	6.946	12.951	.000
Intercept	30.798	1	30.798	57.430	.000
LnSize	118.535	1	118.535	221.034	.000
PriLang	26.181	15	1.745	3.255	.000
OrgType	8.452	11	.768	1.433	.162
PriLang * OrgType	13.886	16	.868	1.618	.068
Error	92.775	173	.536		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .763 (Adjusted R Squared = .704)

IV. Identify the Best Four Variable Model

The previous three variables will be held in the model and we will try to identify the best Four variable Model which will cause the most variation in LnEffort. Table 6.9 shows the best four-variable model with Business Type as the most significant variable which caused 72% variation in LnEffort. The rest of the four variable models are presented in Appendix II.

Table 6.9 Best Four Variable Model

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	306.666[a]	52	5.897	11.410	.000
Intercept	32.845	1	32.845	63.546	.000
LnSize	108.819	1	108.819	210.536	.000
PriLang	8.690	14	.621	1.201	.279
OrgType	9.327	8	1.166	2.256	.026
DBMS	6.246	6	1.041	2.014	.067
PriLang * OrgType	2.727	1	2.727	5.277	.023
PriLang * DBMS	1.147	1	1.147	2.219	.138
OrgType * DBMS	.020	1	.020	.039	.843
PriLang * OrgType * DBMS	.000	0	.	.	.
Error	84.767	164	.517		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .783 (Adjusted R Squared = .715)

V. Identify the Best Five Variable Model

The process of ANCOVA will be iterated for the rest of the variables in order to select the best five variables Models which will make more variation in LnEffort along with the selected four variables.

Table 6.10 shows the best five variables Model with Architecture type as the most significant variable which made 73% variation in LnEffort.

Table 6.10 Best Five Variable Model

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	315.212[a]	60	5.254	10.752	.000
Intercept	32.570	1	32.570	66.661	.000
LnSize	101.960	1	101.960	208.679	.000
PriLang	8.696	14	.621	1.271	.231
OrgType	8.707	8	1.088	2.228	.028
DBMS	3.184	5	.637	1.303	.265
ArchType	1.124	1	1.124	2.301	.131
PriLang * OrgType	2.726	1	2.726	5.579	.019
PriLang * DBMS	1.149	1	1.149	2.351	.127
PriLang * ArchType	4.010	4	1.003	2.052	.090
OrgType * DBMS	.156	1	.156	.319	.573
OrgType * ArchType	.469	1	.469	.961	.329
DBMS * ArchType	3.805	1	3.805	7.787	.006
PriLang * OrgType * DBMS	.000	0	.	.	.
PriLang * OrgType * ArchType	.000	0	.	.	.
PriLang * DBMS * ArchType	.000	0	.	.	.
OrgType * DBMS * ArchType	.000	0	.	.	.
PriLang * OrgType * DBMS * ArchType	.000	0	.	.	.
Error	76.221	156	.489		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .805 (Adjusted R Squared = .730)

No more variables along with best five selected variables made any improvement in the current model.

Finally we obtained the best five variables which made 73% variation in LnEffort. The LnEffort is function of the following five variables;

LnEffort = f (LnSize, PrimLang, OrgType, DBMS, ArchType)

Step 7 Collinearity between Extracted Variables

The relationship between selected numerical and categorical variables must be checked in order to use them in a single model. They can be used only if the relation between numerical and extracted categorical variables is non linear. Also it is important to check whether these extracted categorical variables are independent.

In order to check the relation between numerical and categorical variables we use ANOVA and to check the Collinearity between categorical variables we use Chi square test. Table 6.11 shows the summary of the tests. In order to further eliminate strongly correlated variables, we will apply Univariate Analysis on the extracted four variables in response to LnEffort. Table 6.12 shows Univariate analysis of the four categorical variables.

At this step we are succeeded in selecting variables for LS containing Effort, Size, Duration and business type which will be used for OLS, while the remaining variables will make EbA set.

Table 6.11 Correlation Coefficients

variables	*pr(<F)*	Test method	significant code
LnSize vs. PriLang	0.168	ANOVA	
LnSize vs. OrgType	0.000	ANOVA	**
LnSize vs. ArcType	0.007	ANOVA	**
LnSize vs DBMS	0.066	ANOVA	
PriLang vs OrgType	0.000	Chi-Square	**
PriLang vs ArchType	0.01	Chi-Square	**
PriLang vs DBMS	0.000	Chi-Square	**
OrgType vs DBMS	0.000	Chi-Square	***
OrgType vs ArchType	0.000	Chi-Square	**
ArchType vs. DBMS	0.000	Chi-Square	**

Table 6.12 Univariate Analysis

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	213.252[a]	59	3.614	3.185	.000
Intercept	4238.520	1	4238.520	3734.669	.000
PriLang	22.254	14	1.590	1.401	.159
OrgType	36.249	8	4.531	3.992	.000
DBMS	4.032	5	.806	.711	.616
ArchType	1.392	1	1.392	1.227	.270
PriLang * OrgType	2.218	1	2.218	1.954	.164
PriLang * DBMS	1.786	1	1.786	1.574	.212
PriLang * ArchType	3.264	4	.816	.719	.580
OrgType * DBMS	5.332	1	5.332	4.698	.032
OrgType * ArchType	4.956	1	4.956	4.367	.038
DBMS * ArchType	7.685	1	7.685	6.771	.010
PriLang * OrgType * DBMS	.000	0	.	.	.
PriLang * OrgType * ArchType	.000	0	.	.	.
PriLang * DBMS * ArchType	.000	0	.	.	.
OrgType * DBMS * ArchType	.000	0	.	.	.
PriLang * OrgType * DBMS * ArchType	.000	0	.	.	.
Error	178.181	157	1.135		
Total	13309.405	217			
Corrected Total	391.433	216			

Table 6.12 shows that PriLang and DBMS are not significant, OrgType and ArchType are significant with LnSize also OrgType and ArchType are independent and can be used in a single model. Thus, the final model with three variables is in the form;

LnEffort = f (LnSize, ArchType, OrgType)

Step 8 Extract Intercept for the Linear equation

By passing these extracted variables through linear regression it identified that the equation will be;

Effort = Intercept + LnSize+ ArchType+ OrgType

In order to extract their coefficients and intercept, we will conduct Generalized Linear Model test. The procedure is already installed in SPSS [75]. Table 6.13 Shows parameter estimates of the model.

Table 6.13 parameter estimates of the model

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	2.163	.5909	1.005	3.322	13.403	1	.000
LnSize	.929	.0556	.820	1.038	279.139	1	.000
[ArchType=Client server]	.553	.1365	.286	.821	16.431	1	.000
[ArchType=Stand alone]	0[a]	.	.	.	.	.	.
[OrgType=Aerospace / Automotive;]	.007	.5158	-1.004	1.018	.000	1	.989
[OrgType=Agriculture, Forestry, Fishing, Hunting;]	-.270	.6049	-1.455	.916	.199	1	.656
[OrgType=Banking;]	.828	.4859	-.124	1.781	2.907	1	.088
[OrgType=Communications;]	.347	.4717	-.578	1.271	.540	1	.462
[OrgType=Electricity, Gas, Water;]	-.027	.4996	-1.007	.952	.003	1	.956
[OrgType=Financial, Property & Business Services;]	-.379	.5790	-1.514	.755	.429	1	.512
[OrgType=Government;]	1.130	.5425	.067	2.194	4.343	1	.037
[OrgType=Insurance;]	.507	.4758	-.426	1.439	1.134	1	.287
[OrgType=Manufacturing;]	-.646	.6471	-1.915	.622	.997	1	.318
[OrgType=Public Administration;]	.321	.5217	-.702	1.343	.378	1	.538
[OrgType=Public Administration;Community Services;Insurance;]	.640	.4820	-.305	1.585	1.763	1	.184
[OrgType=Public Administration;Insurance;]	0[a]	.	.	.	.	.	.
(Scale)	.596[b]	.0572	.494	.720			

Dependent Variable: LnEffort

Model: (Intercept), LnSize, ArchType, OrgType

a. Set to zero because this parameter is redundant.

b. Maximum likelihood estimate.

The intercept and coefficients are;
LnEffort = 2.163 +0.929(LnSize) +(ArchType)+(OrgType)

ArchType =1 = Client Server = .553
ArchType = 2 = Stand alone = 0
While;
OrgType = 1 = Aerospace/ Automotive = .007
OrgType = 2 = Agriculture= -.270
OrgType = 3 = Banking = .828
OrgType = 4 = Communications = .347
OrgType = 5 = Electricity = -0.027
OrgType = 6 = Finance= -0.379
OrgType = 7 = Government = 1.13
OrgType = 8 = Insurance = 0.507
OrgType = 9 = Manufacturing = -0.646
OrgType = 10 = Public Administration = 0.321
OrgType = 11= Community services = 0.640
OrgType = 12 = Public administration, Insurance = 0

Step 9 Test Residuals

Residual is the difference between actual effort and predicted effort by the model. In a well-fit model there should be no patterns plotted against the project effort. We plotted 2-dimensional scatter diagram against the project residuals in Figure 6.15 and model predicted effort and will also draw histogram of the residuals Figure 6.14. If it is normal then the model is well-fit or the extracted variables are valid. We also show Q-Q normality plot figure 6.16.

Figure 6.14 Histogram of the Project residuals calculated by Model

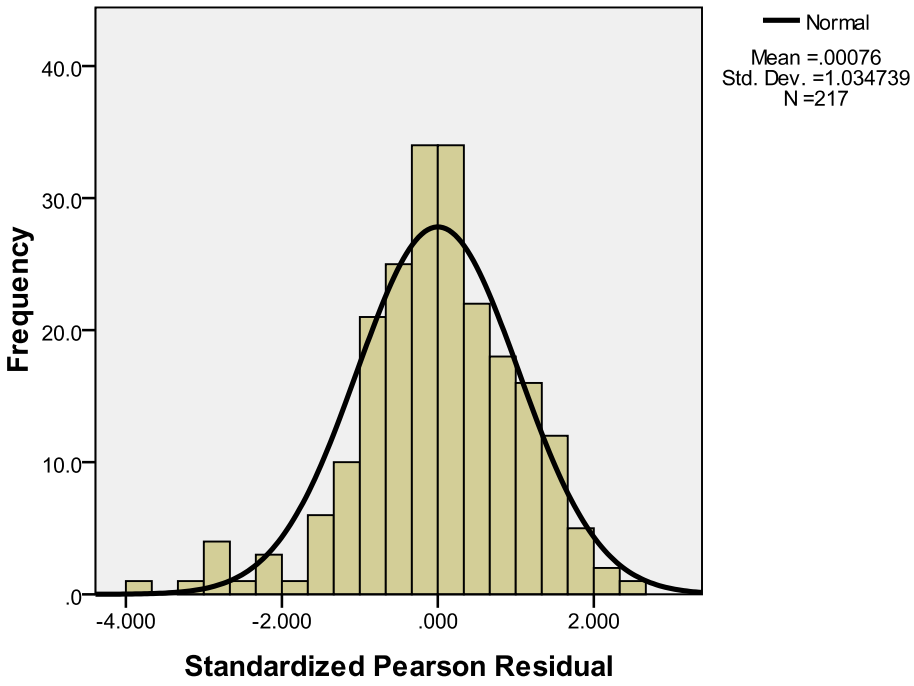


Figure 6.15 Scatter Plot of the residuals against predicted values

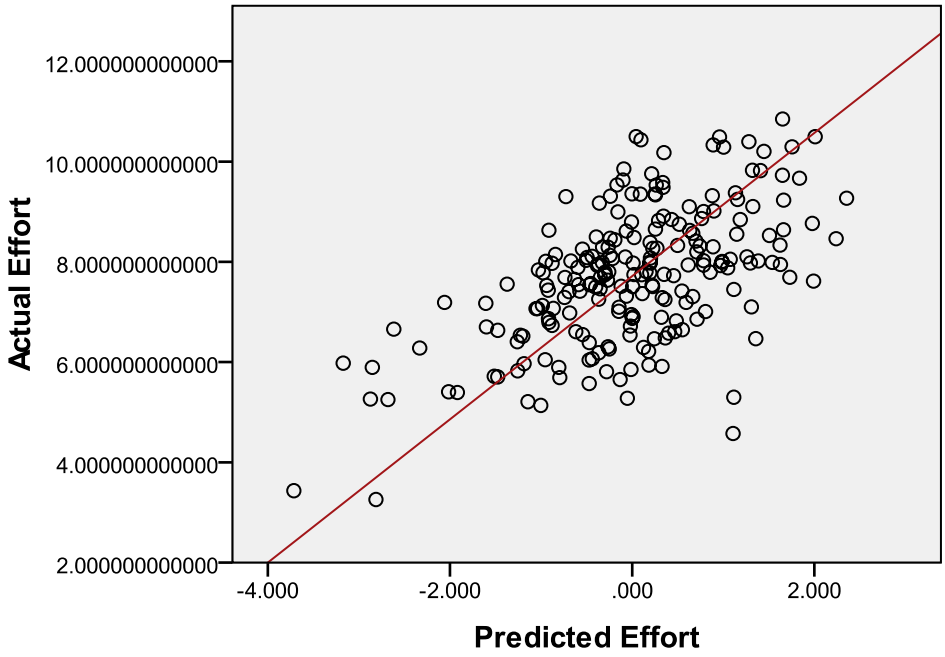


Figure 6.16 Normal Q-Q plot of the values estimated by Model

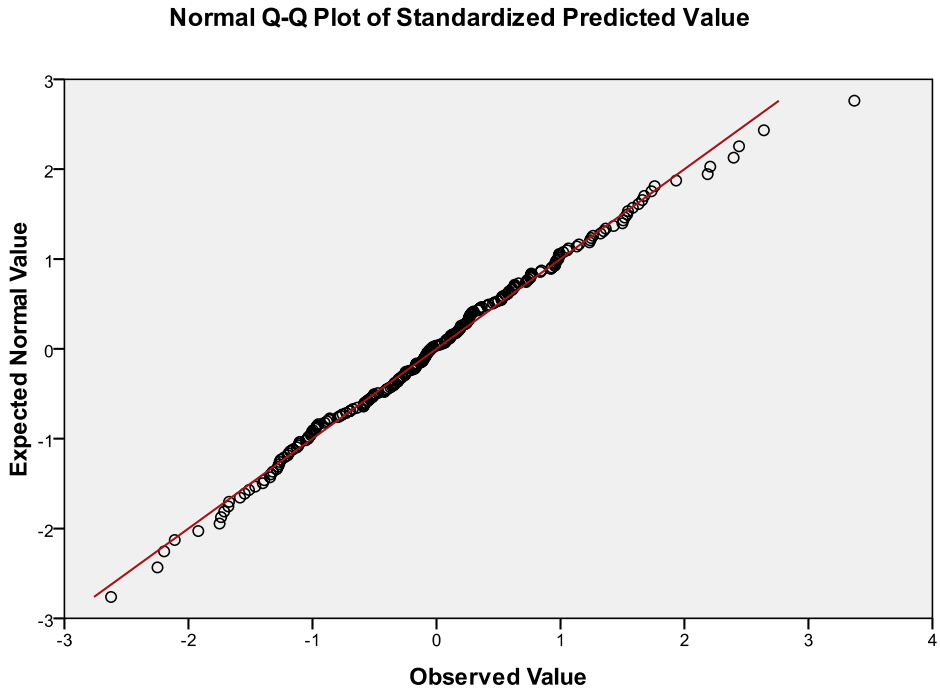


Figure 6.14 shows the histogram of the residuals of the model predicted values, it shows normal distribution but the data is negatively skewed. Figure 6.15 shows the model predicted values against actual values and shows strong relation showing that predictor variables are significant and the model will produce reliable estimates.

6.4 Empirical study Phase-II

In this part of the empirical study, we evaluated the models on the basis of their predictive accuracy. The model predictive accuracy is measured in terms of the model predicted value and actual value.

Two models are evaluated on the basis of their predictive accuracy at various prediction levels. Classical measures used for the model predictive accuracy are; Magnitude of Relative Error (MRE) [7], Mean of MRE (MMRE)[8], Magnitude of Relative Error to Estimate (MER), Mean MER (MMER), Median of MRE (MdMRE), Median of MMER (MdMER) [8], PRED(*l*).

All these accuracy indicators are used for point estimation. The formulae calculated are as follows:

RE = Actual – Estimated/ Actual
MRE = |Actual – Estimated|/ Actual
While MER = |Actual – Estimated| / Estimated
Z ratio is Estimated/ Actual

The term residual is the measure of error between estimated and actual

MMRE = $1/n \sum^{i=1}_{n}$|Actual – Estimated| / Actual
MdMRE = median {|Actual – Estimated|} /Actual
MMER = $1/n \sum^{i=1}_{n}$|Actual – Estimated| / Estimated
MdMER = median {|Actual – Estimated|} /Estimated
PRED (*l*) mre = No. of projects with MRE <=*l*/ No. of projects
PRED (*l*) mer = No. of projects with MER <=*l*/ No. of projects
MSE = $1/n \sum^{i=1}_{n}$ (Actual - Estimated) ^{2}
Median *(z)* = median {Estimated/ Actual}

Evaluating models on the basis of single measure is very error prone as a slight change in data can make a large variation and hence results in inaccurate prediction [89]. In the following work we will make use of re sampling techniques for the comparison of the accuracy measures. We will use OLS as parametric and EbA as non-parametric regression in order to calculate effort from the same data set.

Two popular re sampling methods; bootstrap confidence interval and permutation test will be used in order to form confidence intervals of the calculated residuals. Both of them will be applied through computer aided tools on the calculated results from the classical tests. Results will be compared and appropriate method will be identified.

We used ANGEL [83] tool to identify the accuracy measures for EbA and LS. Categorical variables were recoded and then all variables were passed through the tool. The tool was tuned applying jack-knifing as hold-out strategy, MMRE as performance indicator. Backward sequential search was kept as feature selection strategy and according to Angelis [90] number of analogies was kept to one. Applying Jackknife algorithm [83] the following features were identified significant for EbA set.

Size, Effort, DevType, OrgType, BusType, AppType, DBMS, Umethod. The following measures were recorded.

Accuracy measure shows the model predictive reliability and measured in terms of MMRE, MdMRE and PRED (*l*). The detail of each measure is given in section 6.4. The following sections will provide data for the model evaluation.

Parametric and non-parametric procedures will be used to evaluate the predictive performance of the two models on the basis of their accuracy indicators.

Statistical tests and the use of inferential statistics will be used in order to evaluate the significance of the accuracy measures

6.4.1 Global Accuracy Measures

By applying Jackknife procedure to both LS and EbA set, the following accuracy measures were recorded.

Table 6.14 Descriptive statistics of the global accuracy measures

Measures	OLS	EbA
MMRE	1.183	0.825
MMER	1.15	0.967
MdMRE	0.595	0.470
MdMER	0.570	0.551
PRED mre (25%)	25	31.86
PRED mer (25%)	25.4	34.80
PRED mre (10%)	15.19	19.12
PRED mer(10%)	15.19	18.63

Out of 8 global accuracy measures EbA outperforms OLS but on the basis of PRED (*l*) OLS outperforms EbA. So it needs further analysis.

Now in order to check the significance of the difference between accuracy measures, we will conduct paired sample *t-test* as a parametric test and *Wilcoxon signed rank* as non-parametric test.

6.4.2 Parametric and non-parametric tests

The paired sample *t-test* is a parametric test which compares the difference of the means of two samples and applies to MRE, MER and *z* values. *T-test* follows normal distribution of the difference of errors and is very sensitive to outliers. While *Wilcoxon-signed rank* test is a non-parametric procedure used to identify the significance of the difference between median of the accuracy indicators. It is free from distribution assumptions.

Table 6.15 Significance of the difference between global accuracy measures

Measures	Paired-sample *t*-test	Wilcoxon signed rank test
MMRE	0.043	-
MdMRE	-	0.05
MMER	0.067	-
MdMER		0.059
PRED mre (25)	-	-
PRED mer (25)	-	
Mean(z)	0.04	-
Median(z)	-	0.014

The results from Mean (z), Median (z) and MMRE are significant from the two comparative models.
We used permutation test and confidence intervals for the comparison of accuracy measures. We will construct confidence intervals of the errors in order to identify the "best fit".

6.4.3 Constructing Confidence Intervals

6.4.3.1 Bootstrap confidence Intervals

Since we have three metrics of measuring errors; a) mean b) median and c) percentages therefore, we will construct three bootstrap confidence intervals for mean, median and percentages and will observe the significance.

I. Confidence Intervals for the difference of means

Table 6.16

CI	MMRE		MMER	
	90%	95%	90%	95%
SE CI	-0.045, 0.443	-0.053, 0.422	-0.089, 0.367	-0.358, 0.432

CI	MSE		Mean(*z*)	
	90%	95%	90%	95%
SE CI	-15089, 44501	-53086, 29215	0.003, 0.765	0.051, 0.787

Standard Error confidence interval

The minus sign indicates there is zero difference between means

II. Bootstrap confidence interval for the median of global accuracy measures

Table 6.17

CI	MdMRE		MdMER		Median(*z*)	
	90%	95%	90%	95%	90%	95%
P CI	0.145, 0.753	-0.140, 0.788	-0.087, 0.766	-0.594, 1.879	.880, 1.899	.450, 0.879

The median confidence interval of the global accuracy measures is significant.

6.4.3.2 Permutation tests

Permutation test will be conducted in order to get the significant values and will be compared with paired sample t- test for parametric and Wilcoxon signed rank test for non-parametric. We will also observe that whether the contradictory result from the bootstrap confidence interval is satisfied or not.

Table 6.18 shows results from the permutation tests conducted on various accuracy measures.

Table 6.18

Global Accuracy Measures	Permutation test result (*p*)
MMRE	0.352
MdMRE	0.05
MMER	0.701
MdMER	0.493
Mean(*z*)	0.03
Median(*z*)	0.035

Based on the results from Table 6.15 and Table 6.18, it is concluded that non-parametric tests give significant results. Both MdMRE and Median (z) has given significant results for non-parametric tests, showing that EbA will provide better prediction results.

Table 6.19 shows the comparison of results from the studies performed on model evaluation using ISBSG data set. Mendes performed studies on model evaluation using step wise regression on ISBSG data set in 2005. In 2008 Liu performed studies on model evaluation using four kinds of model out of which two are OLS and K-NN. The results from both studies are in the form of mean magnitude of relative error of the model, median of the magnitude of the relative error and prediction at certain level of the given models. The results from this study both in terms of the MMRE and MdMRE are better from previous studies. While testing the significance of the relative error of the model statistically and evaluated the significance of the results by using more inferential statistics we concluded that estimation by analogy will produce more reliable estimates.

Table 6.19 Comparison of MMRE, MdMRE and PRED (*l*) results Mendes' and Liu et al experiments

Comparison with other research	MMRE	MdMRE	PRED (0.25)	Authors
Step wise regression	1.02	0.52	0.14	Mendes 2005
	1.23	0.61	0.21	Mendes 2005
OLS	0.66	0.44	0.31	Liu 2008
K-NN	2.48	0.81	15.47	Liu 2008
OLS	1.18	.595	25	This study
K-NN	0.825	0.47	0.31	This study

6.5 Summary of the Results

The study was conducted in order to evaluate prediction accuracy of the cost estimation models. We evaluated these models according to their fidelity not only on single point based estimation but with the help of two re-sampling methods. Summary of the results from both phases of the empirical studies are presented below.

Preliminary data analysis was conducted in order to form LS set for parametric cost estimation models and EbA set for non-parametric cost estimation models.

Initially maximum variables participated in order to identify the linear relationship with response variables. A set of variables containing LnSize, ArchType and OrgType were

identified as the most significant variables which caused maximum variation in Effort. They are found to be linearly related with Effort.

LnEffort = *f* (LnSize, ArchType, OrgType)…… (1)
LnEffort is a function of LnSize, ArchType and OrgType. By applying generalized linear method which is already implemented in SPSS, we obtained the parameters of the model. Linear regression was then used to calculate the exponents of the model.

In empirical study Phase II, the most significant variables identified in Phase I were used as input for OLS regression. We used ANGEL which is a case based reasoning tool to identify best features (cost factors) and analogues for EbA models. We adopted backward sequential search strategy to identify best features and kept number of analogies to one in order to identify the best analogues. The following features were selected; Size, Effort, DevType, OrgType, BusType, AppType, DBMS, Umethod. These variables are free from distribution assumption and were used as input for EbA.

The test was conducted and results from both models; parametric and non-parametric were recorded. The results from EbA in the case of MMRE and MdMRE are better than OLS (0.825<1.18) (0.47<.595) respectively. Also mean magnitude of relative error to estimate (MMER) of the EbA is lower than that of OLS (0.967<1.15). We can say that EbA is predicting reliable estimates.

We used parametric and non-parametric tests in order to identify the significance of global accuracy measures. Paired sample *t-test* was used for measures of parametric method and *Wilcoxon signed rank test* was used as non-parametric. The procedures are implemented in SPSS. The test resulted in significance that both MdMRE (0.05), and Median (z) (0.014) are significant ($p<0.05$) at 95 percent confidence of interval for EbA. This shows that EbA will produce more reliable estimates than using regression based models.

We used bootstrap confidence interval and permutation tests as re-sampling methods in order to evaluate the significance of the results from both regression and analogy based models. We used MatLab tool as a simulator to produce Bootstrap confidence intervals of the standard error of both models. The results from bootstrap confidence interval showed that there is no difference between the accuracy indicators. Finally from the permutation test results it is cleared that EbA will result in better prediction than OLS.

7 CONCLUSION AND FUTURE WORK

In this thesis work we evaluated two well known effort estimation methodologies namely regression and EbA with the help of traditional global accuracy measures. All effort estimation models predict the effort or productivity of the new project. Then, the predicted value is compared to the actual value to determine the accuracy of the prediction.

Various methodologies for the evaluation of effort estimation models were explored in the light of their accuracy indicators. The working mechanism of the effort estimation models is studied. Strengths and weaknesses of the well-known effort estimation methodologies were identified. Various performance indicators were explored for the evaluation of well-known effort estimation models. Their usefulness and shortcoming were identified and alternative approaches are introduced. During our investigation we make use of inferential statistics to evaluate the significance of the accuracy indicator.

Two data sets were extracted from ISBSG project repository, one with variables which have linear relationship with response variable and the other set contained variable which were free from distribution assumptions. UFP of quality 'A' projects were selected for further analysis. After a thorough analysis of the data, Project Size, Architecture Type and Organization type were identified as linearly related to Project Effort. These variables were used for regression and the remaining variables were utilized in EbA to predict the cost of new project.

We made our investigation in two phases; empirical study Phase I and Phase II. By conducting the empirical studies, we could have answered the research questions of this thesis study. Our first research question was;

RQ1. What are the measures required for the evaluation of model accuracy and consistency?

We identified the classical accuracy indicators MMRE, MdMRE and PRED mre (*l*) as measures for model evaluation. We also used the most recent measures MMER, MdMER and PRED mer (*l*) to determine the model predictive accuracy. Besides the conventional accuracy indicators test we make use of two complex inferential statistical re-sampling techniques in order to evaluate model predictive accuracy.

Our second research question was;

RQ2. What is the mechanism behind well-known effort estimation models, their calibration, and evaluation?

Research question No. 2 was answered during our literature review in chapter 4. see section 4.1.2 Para 2 page 9 and 10

Our third research question was;

RQ3. Does the estimation reliability increases with the pre-processing of the project data statistically.
After the concise processing of the data set statistically using a newly proposed framework for removing any kind of ambiguities the results are collected. We compared the results Table 6.19**,** with the published results from other researchers using ISBSG data set and noticed that estimation reliability has increased considerably after the pre-processing of the data set statistically. It clearly indicates that checking data for any kind of ambiguities is vital for reliable estimation.

Our fourth research question was;

RQ4. Are the accuracy indicators statistically significant?

We used paired sample *t-test* for parametric regression and *Wilcoxon signed rank test* for non-parametric regression in order to identify the significance of the accuracy indicators. The non-parametric test gave significant results for EbA method. Since non-parametric tests are free from distribution assumptions therefore the significance of the Wilcoxon test goes in favor of the EbA which is non-parametric cost estimation method.

Finally, our fifth research question was;

RQ5. What other measures can be used for the evaluation of effort estimation models?

It is learned from the literature review that there are various re-sampling methods which can be used to evaluate model performance. We used two kinds of re-sampling techniques one: bootstrap confidence interval and second: permutation test, to evaluate performance of the accuracy indicators. Both are complex inferential statistical procedures and were used successfully for the model predictive accuracy evaluation. The results from permutation tests identified the superiority of estimation by analogy model.

The above research questions were answered through a comprehensive literature review and conducting exhaustive empirical studies. We concluded in that before using any model for effort estimation concise prior data analysis is of utmost importance. Rely on a single measure of error for selecting the right model for effort estimation is risky and threat to reliable estimation. The results from various models must be validated through statistical tests and use more inferential statistics to identify the significance of these results. Tests have shown that the use of non-parametric models for effort estimation will provide more realistic estimates than regression based models.

Parametric methods estimate cost only if there is a linear relation between predicted variable and predictor variable while non-parametric techniques are free from distribution assumptions. We can use feature (cost drivers) selection algorithms such as ‘forward selection search’ and ‘backward selection search’ for the selection of best features which will result in more reliable estimation and also biased free. Further investigations are needed to identify the significance of the global accuracy indicators using more inferential statistical techniques. A framework for using both parametric and non-parametric methodologies in a single model will be a field of promising research.

APPENDIX I

Table 1 Frequency of Count Approach

		Frequency	Percent
1.	Albrecht	1	.2
2.	Feature Points	1	.2
3.	IFPUG	546	97.2
4.	Mark II	6	1.1
5.	NESMA	8	1.4
6.	Total	562	100.0

Table 2 Frequency of Architecture type

		Frequency	Percent
1.	Client server	107	19.0
2.	Multi-tier	7	1.2
3.	Multi-tier / Client server	167	29.7
4.	Multi-tier with web public interface	1	.2
5.	Stand alone	280	49.8
6.	Total	562	100.0

Table 3 Frequency of development type

		Frequency	Percent	Valid Percent	Cumulative Percent
1.	Enhancement	244	43.4	43.4	43.4
2.	New Development	303	53.9	53.9	97.3
3.	New Utility	1	.2	.2	97.5
4.	Re-development	14	2.5	2.5	100.0
5.	Total	562	100.0	100.0	

Table 4 Frequency of development platform

		Frequency	Percent	Valid Percent	Cumulative Percent
1.	MF	192	34.2	34.2	34.2
2.	MR	129	23.0	23.0	57.1
3.	Multi	168	29.9	29.9	87.0
4.	PC	73	13.0	13.0	100.0
5.	Total	562	100.0	100.0	

Table 5 Frequency of Application type

	Frequency	Percent	Valid Percent	Cumulative Percent	
1.	Accounting;	12	2.1	2.1	2.1
2.	Actuarial System - calculate employer rates.;	1	.2	.2	2.3
3.	Administration;	1	.2	.2	2.5
4.	Architectural;	1	.2	.2	2.7
5.	Banking;	57	10.1	10.1	12.8
6.	Blood Bank;	1	.2	.2	13.0
7.	Case Management;	1	.2	.2	13.2
8.	Chartered flight operation;	1	.2	.2	13.3
9.	Claims Processing - Product pays claim;	1	.2	.2	13.5
10.	Contract Management;	1	.2	.2	13.7
11.	Dispute Resolution;	1	.2	.2	13.9
12.	Distribution & Transport;	1	.2	.2	14.1
13.	Distribution/Scheduling;	1	.2	.2	14.2
14.	don't know;	169	30.1	30.1	44.3
15.	Energy generation;	1	.2	.2	44.5
16.	Engineering;	18	3.2	3.2	47.7
17.	Financial (excluding Banking);	40	7.1	7.1	54.8
18.	Financial (excluding Banking);Accounting;	1	.2	.2	55.0
19.	Financial (excluding Banking);Personnel;	1	.2	.2	55.2
20.	Financial;	3	.5	.5	55.7
21.	Fine Enforcement;	8	1.4	1.4	57.1
22.	Generate & Distribute Electricity;	1	.2	.2	57.3
23.	Government;	1	.2	.2	57.5
24.	Insurance;	56	10.0	10.0	67.4
25.	Inventory;	3	.5	.5	68.0

26.	Inventory;Sales;	1	.2	.2	68.1
27.	Legal;	5	.9	.9	69.0
28.	Library Management;	1	.2	.2	69.2
29.	Logistics;	3	.5	.5	69.8
30.	Manufacturing;	22	3.9	3.9	73.7
31.	Mining Production Information;	1	.2	.2	73.8
32.	Network Management;	1	.2	.2	74.0
33.	Pension Funds Management;	8	1.4	1.4	75.4
34.	Personnel;	10	1.8	1.8	77.2
35.	Personnel;Banking;	1	.2	.2	77.4
36.	Personnel;Health, Safety & Environment;	1	.2	.2	77.6
37.	Procurement;	1	.2	.2	77.8
38.	Project management & job control.;	1	.2	.2	77.9
39.	Provide computer services and IT consultation;	1	.2	.2	78.1
40.	Provide computer systems and IT Consultation;	1	.2	.2	78.3
41.	Providing IT Companies different clients and vice versa;	1	.2	.2	78.5
42.	Public Administration;	1	.2	.2	78.6
43.	PUBLIC HEALTH & FAMILY SERVICES;	1	.2	.2	78.8
44.	Purchasing;	1	.2	.2	79.0
45.	regulation;	4	.7	.7	79.7
46.	Regulatory agency;	2	.4	.4	80.1
47.	Regulatory/contaminated site management;	1	.2	.2	80.2
48.	Research & Development;	4	.7	.7	81.0
49.	Sales & Marketing;	3	.5	.5	81.5
50.	Sales;	9	1.6	1.6	83.1
51.	Social Services;	4	.7	.7	83.8
52.	Telecommunications;	89	15.8	15.8	99.6
53.	TESTING;	1	.2	.2	99.8
54.	Valuation;	1	.2	.2	100.0
55.	Total	562	100.0	100.0	

Table 6 Frequency of Business type

	Frequency	Percent	Valid Percent	Cumulative Percent	
1.	Advertising/Mailing Campaign;	1	.2	.2	.2
2.	Business enabling service;	1	.2	.2	.4
3.	Card Administration INCAS & Fault Assurance CAPRA;	1	.2	.2	.5

4.	Case Management Study;	1	.2	.2	.7
5.	Corporate Taxation;	1	.2	.2	.9
6.	Customer billing/relationship management;	9	1.6	1.6	2.5
7.	Customer billing/relationship management;CRM;	1	.2	.2	2.7
8.	Data Warehouse;	1	.2	.2	2.8
9.	Decision Support System;	10	1.8	1.8	4.6
10.	Decision Support System;Management Information System;	1	.2	.2	4.8
11.	Document management;	3	.5	.5	5.3
12.	Electronic Data Interchange;	16	2.8	2.8	8.2
13.	Equipment Management;	1	.2	.2	8.4
14.	Executive Information System;	11	2.0	2.0	10.3
15.	Fault Tolerance;Management Information System;Fault Management;	1	.2	.2	10.5
16.	Financial transaction process/accounting;	75	13.3	13.3	23.8
17.	information systems (Web);	1	.2	.2	24.0
18.	information systems;	3	.5	.5	24.6
19.	Inventory Control;	1	.2	.2	24.7
20.	Job, case, incident, project management;	4	.7	.7	25.4
21.	Logistic or supply planning & control;	6	1.1	1.1	26.5
22.	Management Information System;	125	22.2	22.2	48.8
23.	Management Information System;Office Information System;Transaction/Production System;	2	.4	.4	49.1
24.	Mathematical modelling (finance or engineering);	1	.2	.2	49.3
25.	Network Management;	13	2.3	2.3	51.6
26.	Not specified;	1	.2	.2	51.8
27.	Office Information System;	15	2.7	2.7	54.4
28.	Office Information System;Case Management System;	1	.2	.2	54.6
29.	Office Information System;Functional Specification System;	1	.2	.2	54.8

30.	Other;	4	.7	.7	55.5
31.	other: condemnation proceedings;	1	.2	.2	55.7
32.	other: Management system;	1	.2	.2	55.9
33.	other: mission-critical system;	3	.5	.5	56.4
34.	other: personnel system ;	3	.5	.5	56.9
35.	other: production management system;	6	1.1	1.1	58.0
36.	other: Sales contact management;	18	3.2	3.2	61.2
37.	other: Strategic planning;	1	.2	.2	61.4
38.	personal system;	1	.2	.2	61.6
39.	Process Control;	12	2.1	2.1	63.7
40.	Purchasing;	1	.2	.2	63.9
41.	Real-time System;	7	1.2	1.2	65.1
42.	RECONCILIATION;	1	.2	.2	65.3
43.	Scientific;	1	.2	.2	65.5
44.	Software for machine control;	4	.7	.7	66.2
45.	Stock control & order processing;	11	2.0	2.0	68.1
46.	System conversion;	1	.2	.2	68.3
47.	Technical Information System;	1	.2	.2	68.5
48.	tools or system;	2	.4	.4	68.9
49.	Trading;	6	1.1	1.1	69.9
50.	Trading? (procurement management);	1	.2	.2	70.1
51.	Transaction/Production System;	162	28.8	28.8	98.9
52.	Type 1 Function Point Counting Tool;	1	.2	.2	99.1
53.	Vehicle Systems Software;	1	.2	.2	99.3
54.	Workflow support & management;	4	.7	.7	100.0
	Total	562	100.0	100.0	

Table 7 Frequency of language type

		Frequency	Percent	Valid Percent	Cumulative Percent
1.	3GL	233	41.5	41.5	41.5
2.	4GL	296	52.7	52.7	94.1
3.	ApG	32	5.7	5.7	99.8
4.	APG	1	.2	.2	100.0
5.	Total	562	100.0	100.0	

Table 8 Frequency of primary language

	Frequency	Percent	Valid Percent	Cumulative Percent	
1.	4GL	1	.2	.2	.2
2.	ABAP	2	.4	.4	.5
3.	Access	24	4.3	4.3	4.8
4.	BASIC	3	.5	.5	5.3
5.	C	21	3.7	3.7	9.1
6.	C/C++	1	.2	.2	9.3
7.	C++	26	4.6	4.6	13.9
8.	COBOL	102	18.1	18.1	32.0
9.	CSP	1	.2	.2	32.2
10.	DELPHI	2	.4	.4	32.6
11.	EASYTRIEVE	1	.2	.2	32.7
12.	HPS	9	1.6	1.6	34.3
13.	IDEAL	4	.7	.7	35.1
14.	IEF	1	.2	.2	35.2
15.	INGRES	2	.4	.4	35.6
16.	Java	8	1.4	1.4	37.0
17.	LEX	1	.2	.2	37.2
18.	Lotus Notes	13	2.3	2.3	39.5
19.	NATURAL	29	5.2	5.2	44.7
20.	ORACLE	35	6.2	6.2	50.9
21.	Other 4GL	13	2.3	2.3	53.2
22.	Other ApG	9	1.6	1.6	54.8
23.	OutlookVBA	4	.7	.7	55.5
24.	Pl-SQL	1	.2	.2	55.7
25.	PL/I	19	3.4	3.4	59.1
26.	PL/SQL	11	2.0	2.0	61.0
27.	PowerBuilder	22	3.9	3.9	64.9
28.	Pro-C	6	1.1	1.1	66.0
29.	RALLY	2	.4	.4	66.4
30.	SAS	2	.4	.4	66.7
31.	Shell Script	1	.2	.2	66.9
32.	Smalltalk	1	.2	.2	67.1
33.	SQL	41	7.3	7.3	74.4
34.	TELON	14	2.5	2.5	76.9
35.	VB.Net	1	.2	.2	77.0
36.	Visual Basic	129	23.0	23.0	100.0
37.	Total	562	100.0	100.0	

Table 9 Frequency of main data base used

	Frequency	Percent	Valid Percent	Cumulative Percent	
1.	Access ODBC;	1	.2	.2	.2
2.	ACCESS;	27	4.8	4.8	5.0
3.	ACCESS;DB2 & IMS;	1	.2	.2	5.2
4.	ACCESS;DB2;IMS;	1	.2	.2	5.3
5.	ACCESS;Notes;	1	.2	.2	5.5
6.	ACCESS;Other;	1	.2	.2	5.7
7.	ACCESS;SQL SERVER V6;	1	.2	.2	5.9
8.	ACCESS;SYBASE;	1	.2	.2	6.0
9.	Access2000;	1	.2	.2	6.2
10.	ADABAS V5;	4	.7	.7	6.9
11.	ADABAS V5;Other;	1	.2	.2	7.1
12.	ADABAS 2.25;	1	.2	.2	7.3
13.	ADABAS;	21	3.7	3.7	11.0
14.	ADABAS;ACCESS;	1	.2	.2	11.2
15.	CA-IDMS;	1	.2	.2	11.4
16.	CICS;ORACLE 8;	1	.2	.2	11.6
17.	DATACOM;	4	.7	.7	12.3
18.	DB2 V3;IMS;SYBASE;	1	.2	.2	12.5
19.	DB2 2.2;	1	.2	.2	12.6
20.	DB2;	83	14.8	14.8	27.4
21.	DB2;FLAT FILES;	1	.2	.2	27.6
22.	DB2;FLATFILE;	1	.2	.2	27.8
23.	DB2;IMS;	2	.4	.4	28.1
24.	DB2;SYBASE;	3	.5	.5	28.6
25.	DB2/2;	2	.4	.4	29.0
26.	DB2/400;	5	.9	.9	29.9
27.	DB2/VSAM;	1	.2	.2	30.1
28.	DL/1;	18	3.2	3.2	33.3
29.	Domino R5.04a;	1	.2	.2	33.5
30.	Domino R5.05;	2	.4	.4	33.8
31.	ENSCRIBE;	1	.2	.2	34.0
32.	Exchange;	5	.9	.9	34.9
33.	ExchangeServer5.5;	1	.2	.2	35.1
34.	Flat File;	1	.2	.2	35.2
35.	FOXPRO;IMS;DB2;	1	.2	.2	35.4
36.	GUPTA SQL BASIC;	1	.2	.2	35.6

37.	HYDRA;	1	.2	.2	35.8
38.	IBM OS/2 Database Manager;OS/2 Database Manager;	1	.2	.2	35.9
39.	IBM;	1	.2	.2	36.1
40.	IDMS 10.2 Focus 6.8;	1	.2	.2	36.3
41.	IDMS 10.2;	2	.4	.4	36.7
42.	IDMS;	8	1.4	1.4	38.1
43.	IDMSX;	1	.2	.2	38.3
44.	IMAGE;	1	.2	.2	38.4
45.	IMS Database;	1	.2	.2	38.6
46.	IMS;	21	3.7	3.7	42.3
47.	IMS;ACCESS;	1	.2	.2	42.5
48.	IMS;DB2;	1	.2	.2	42.7
49.	IMS;MS SQL Server;	1	.2	.2	42.9
50.	IMS/DB2;OMNIS 7;	1	.2	.2	43.1
51.	IMS/DBDC;	1	.2	.2	43.2
52.	INFORMIX;	2	.4	.4	43.6
53.	INGRES;	3	.5	.5	44.1
54.	ISAM;	16	2.8	2.8	47.0
55.	ISAM;DB2/400;	1	.2	.2	47.2
56.	LOTUS NOTES;	3	.5	.5	47.7
57.	M204;	4	.7	.7	48.4
58.	MS Access;	2	.4	.4	48.8
59.	MS SQL Server;	2	.4	.4	49.1
60.	Notes;	1	.2	.2	49.3
61.	NOTES;	1	.2	.2	49.5
62.	NotesDomino6;	1	.2	.2	49.6
63.	NotesR5.08;	1	.2	.2	49.8
64.	OBJECTSTOR;	6	1.1	1.1	50.9
65.	ORACLE V6;ORACLE V7;	6	1.1	1.1	52.0
66.	ORACLE V7;	5	.9	.9	52.8
67.	Oracle 7;	2	.4	.4	53.2
68.	ORACLE 7;	1	.2	.2	53.4
69.	Oracle 7.3;	6	1.1	1.1	54.4
70.	Oracle 7.3.2;	1	.2	.2	54.6
71.	Oracle 7.3.3;	1	.2	.2	54.8
72.	Oracle 7.3.4;	1	.2	.2	55.0
73.	Oracle 8;	4	.7	.7	55.7
74.	Oracle 8.0;	1	.2	.2	55.9
75.	Oracle 8.0.5;	6	1.1	1.1	56.9
76.	Oracle 8.0.6;	6	1.1	1.1	58.0
77.	Oracle 8.05;	2	.4	.4	58.4
78.	Oracle 8.1.6;	6	1.1	1.1	59.4
79.	Oracle 8.1.7;	1	.2	.2	59.6

80.	Oracle 8i;	3	.5	.5	60.1
81.	Oracle 8i;Sybase ASA;	1	.2	.2	60.3
82.	ORACLE 8i;WGS8.1.7.1;	1	.2	.2	60.5
83.	Oracle 9i;	2	.4	.4	60.9
84.	Oracle 9i;Empress;	2	.4	.4	61.2
85.	ORACLE SQLN / WATCOM;ORACLE;	1	.2	.2	61.4
86.	ORACLE V7;SYBASE V10;	1	.2	.2	61.6
87.	Oracle,SQL Server;	1	.2	.2	61.7
88.	ORACLE;	117	20.8	20.8	82.6
89.	Oracle;Access;	1	.2	.2	82.7
90.	Oracle;SybaseSQL-Anywhere;	1	.2	.2	82.9
91.	Other;	6	1.1	1.1	84.0
92.	Other;RMS;	1	.2	.2	84.2
93.	PARADOX;	1	.2	.2	84.3
94.	Personal Oracle 8i;	1	.2	.2	84.5
95.	PICK;	2	.4	.4	84.9
96.	Proprietery;	1	.2	.2	85.1
97.	Raima;	1	.2	.2	85.2
98.	RDB 4-2-1;	1	.2	.2	85.4
99.	RDB 6.0;	1	.2	.2	85.6
100.	RDB;	7	1.2	1.2	86.8
101.	RDMS;DMS 1100;	1	.2	.2	87.0
102.	RMS;	2	.4	.4	87.4
103.	SAS;	1	.2	.2	87.5
104.	SQL Server 7.0;	1	.2	.2	87.7
105.	SQL SERVER V6;	1	.2	.2	87.9
106.	SQL SERVER;	15	2.7	2.7	90.6
107.	SQL Server6.0;	2	.4	.4	90.9
108.	SQL Server6.5;	1	.2	.2	91.1
109.	SQL Server7.0;	1	.2	.2	91.3
110.	SQL-Server7.0;	1	.2	.2	91.5
111.	SYBASE 4.4;	1	.2	.2	91.6
112.	SYBASE 4.9;	1	.2	.2	91.8
113.	Sybase ASE 11.9.2 for UNIX;	1	.2	.2	92.0
114.	Sybase ASE 11.9.2;	1	.2	.2	92.2
115.	Sybase SQL Anywhere;	1	.2	.2	92.3
116.	SYBASE;	32	5.7	5.7	98.0
117.	VMS;	1	.2	.2	98.2
118.	VSAM;	2	.4	.4	98.6
119.	VSE/1SQL;	1	.2	.2	98.8
120.	WATCOM SQL;	1	.2	.2	98.9
121.	WATCOM;Other;	1	.2	.2	99.1
122.	WGRES;Adaptable to	1	.2	.2	99.3

	Oracle;Informix;				
123.	Yes;	4	.7	.7	100.0
	Total	562	100.0	100.0	

Table 10 Frequency of Used methodology

		Frequency	Percent	Valid Percent	Cumulative Percent
1.	Don't Know	165	29.4	29.4	29.4
2.	No	60	10.7	10.7	40.0
3.	Yes	337	60.0	60.0	100.0
4.	Total	562	100.0	100.0	

Table 10 Umethod

Table 11 Frequency of required methodology

		Frequency	Percent	Valid Percent	Cumulative Percent
1.		232	41.3	41.3	41.3
2.	CICS;	1	.2	.2	41.5
3.	Combined Developed/Purchased;	54	9.6	9.6	51.1
4.	Developed Inhouse;	250	44.5	44.5	95.6
5.	Purchased;	25	4.4	4.4	100.0
6.	Total	562	100.0	100.0	

Table 12 Frequency of organization type

	Frequency	Percent	Valid Percent	Cumulative Percent	
1.	Aerospace / Automotive;	11	2.0	2.0	2.0
2.	Agriculture, Forestry, Fishing, Hunting;	4	.7	.7	2.7
3.	Any organization which counts function points;	1	.2	.2	2.8
4.	Banking;	56	10.0	10.0	12.8
5.	Business Services;	3	.5	.5	13.3
6.	Chemicals;	6	1.1	1.1	14.4
7.	Communications;	104	18.5	18.5	32.9
8.	Community Services;	4	.7	.7	33.6

9.	Computer Consultants;	1	.2	.2	33.8
10.	Computers & Software;	5	.9	.9	34.7
11.	Construction;	9	1.6	1.6	36.3
12.	Consultancy;	1	.2	.2	36.5
13.	Consumer Goods;	2	.4	.4	36.8
14.	Coronial Services;	1	.2	.2	37.0
15.	Defence;	1	.2	.2	37.2
16.	Developing global software solutions;	1	.2	.2	37.4
17.	Distribution;	1	.2	.2	37.5
18.	Electricity, Gas, Water;	30	5.3	5.3	42.9
19.	Electronics;	2	.4	.4	43.2
20.	Energy;	1	.2	.2	43.4
21.	Financial, Property & Business Services;	31	5.5	5.5	48.9
22.	Financial, Property & Business Services;Banking;	1	.2	.2	49.1
23.	Food Processing;	1	.2	.2	49.3
24.	Government;	10	1.8	1.8	51.1
25.	Insurance;	76	13.5	13.5	64.6
26.	Manufacturing;	56	10.0	10.0	74.6
27.	Manufacturing;Consumer Goods;	1	.2	.2	74.7
28.	Manufacturing;Transport & Storage;	6	1.1	1.1	75.8
29.	Manufacturing;Wholesale & Retail Trade;	1	.2	.2	76.0
30.	Manufacturing;Wholesale & Retail Trade;Transport & Storage;	1	.2	.2	76.2
31.	Media;	1	.2	.2	76.3
32.	Medical and Health Care;	9	1.6	1.6	77.9
33.	Medical and Health Care;Public Administration;Insurance;	1	.2	.2	78.1
34.	Mining;	3	.5	.5	78.6
35.	Mining;Manufacturing;Chemicals;	2	.4	.4	79.0
36.	Occupational Health and Safety;	1	.2	.2	79.2
37.	Other;	1	.2	.2	79.4
38.	Professional Services;	4	.7	.7	80.1
39.	Professional Services;Computers & Software;	2	.4	.4	80.4
40.	Public Administration;	33	5.9	5.9	86.3
41.	Public Administration;Aerospace / Automotive;Computers & Software;Insurance;	1	.2	.2	86.5
42.	Public Administration;Community Services;Insurance;	41	7.3	7.3	93.8
43.	Public Administration;Financial, Property & Business Services;	1	.2	.2	94.0

44.	Public Administration;Insurance;	5	.9	.9	94.8
45.	Real Estate & Property Services;	2	.4	.4	95.2
46.	Recreation & Personnel Services;	3	.5	.5	95.7
47.	Transport & Storage;	15	2.7	2.7	98.4
48.	Transport & Storage;Aerospace / Automotive;	1	.2	.2	98.6

APPENDIX II

```
UNIANOVA LnEffort BY DevType WITH LnSize
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize DevType.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	225.614[a]	3	75.205	96.603	.000
Intercept	55.222	1	55.222	70.935	.000
LnSize	186.570	1	186.570	239.656	.000
DevType	.686	2	.343	.441	.644
Error	165.819	213	.778		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .576 (Adjusted R Squared = .570)

```
UNIANOVA LnEffort BY OrgType WITH LnSize
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize OrgType.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	252.284[a]	12	21.024	30.822	.000
Intercept	31.865	1	31.865	46.716	.000
LnSize	184.357	1	184.357	270.278	.000
OrgType	27.356	11	2.487	3.646	.000
Error	139.149	204	.682		
Total	13309.405	217			

Corrected Total	391.433	216			
a. R Squared = .645 (Adjusted R Squared = .624)					

```
UNIANOVA LnEffort BY BusiType WITH LnSize
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize BusiType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	257.313[a]	11	23.392	35.755	.000
Intercept	45.283	1	45.283	69.215	.000
LnSize	194.375	1	194.375	297.100	.000
BusiType	32.385	10	3.239	4.950	.000
Error	134.120	205	.654		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .657 (Adjusted R Squared = .639)					

```
UNIANOVA LnEffort BY AppType WITH LnSize
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize AppType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	228.486[a]	8	28.561	36.457	.000
Intercept	72.919	1	72.919	93.080	.000
LnSize	215.510	1	215.510	275.095	.000
AppType	3.558	7	.508	.649	.715
Error	162.947	208	.783		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .584 (Adjusted R Squared = .568)					

```
UNIANOVA LnEffort BY ArchType WITH LnSize
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize ArchType.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	238.350[a]	2	119.175	166.599	.000
Intercept	86.020	1	86.020	120.251	.000
LnSize	197.868	1	197.868	276.606	.000
ArchType	13.422	1	13.422	18.763	.000
Error	153.083	214	.715		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .609 (Adjusted R Squared = .605)

```
UNIANOVA LnEffort BY Devplat WITH LnSize
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize Devplat.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	235.058[a]	3	78.353	106.725	.000
Intercept	73.982	1	73.982	100.772	.000
LnSize	210.387	1	210.387	286.569	.000
Devplat	10.130	2	5.065	6.899	.001
Error	156.375	213	.734		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .601 (Adjusted R Squared = .595)

```
UNIANOVA LnEffort BY LangType WITH LnSize
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LangType.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	234.026[a]	3	78.009	105.560	.000
Intercept	54.409	1	54.409	73.625	.000
LnSize	227.037	1	227.037	307.222	.000
LangType	9.098	2	4.549	6.156	.003
Error	157.407	213	.739		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .598 (Adjusted R Squared = .592)

```
UNIANOVA LnEffort BY PriLang WITH LnSize
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize PriLang.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	274.162[a]	16	17.135	29.223	.000
Intercept	53.097	1	53.097	90.555	.000
LnSize	187.972	1	187.972	320.578	.000
PriLang	49.234	15	3.282	5.598	.000
Error	117.271	200	.586		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .700 (Adjusted R Squared = .676)

```
UNIANOVA LnEffort BY DBMS WITH LnSize
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize DBMS.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	270.926[a]	16	16.933	28.103	.000
Intercept	54.932	1	54.932	91.169	.000
LnSize	173.687	1	173.687	288.262	.000
DBMS	45.999	15	3.067	5.089	.000
Error	120.506	200	.603		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .692 (Adjusted R Squared = .668)

```
UNIANOVA LnEffort BY Umethod WITH LnSize
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize Umethod.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	224.928[a]	1	224.928	290.439	.000
Intercept	74.376	1	74.376	96.038	.000
LnSize	224.928	1	224.928	290.439	.000
Umethod	.000	0	.	.	.
Error	166.505	215	.774		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .575 (Adjusted R Squared = .573)

```
UNIANOVA LnEffort BY Rmethod WITH LnSize
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize Rmethod.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	225.624[a]	3	75.208	96.613	.000
Intercept	57.415	1	57.415	73.755	.000
LnSize	204.190	1	204.190	262.305	.000
Rmethod	.696	2	.348	.447	.640
Error	165.809	213	.778		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .576 (Adjusted R Squared = .570)

```
UNIANOVA LnEffort BY CountApproach WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration CountApproach.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	249.343[a]	3	83.114	124.592	.000
Intercept	51.949	1	51.949	77.874	.000
LnSize	76.173	1	76.173	114.187	.000
LnDuration	24.357	1	24.357	36.512	.000
CountApproach	.044	1	.044	.066	.797
Error	142.090	213	.667		
Total	13309.405	217			
Corrected Total	391.433	216			

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	249.343[a]	3	83.114	124.592	.000
Intercept	51.949	1	51.949	77.874	.000
LnSize	76.173	1	76.173	114.187	.000
LnDuration	24.357	1	24.357	36.512	.000
CountApproach	.044	1	.044	.066	.797
Error	142.090	213	.667		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .637 (Adjusted R Squared = .632)					

```
UNIANOVA LnEffort BY DevType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration DevType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	249.920[a]	4	62.480	93.601	.000
Intercept	58.181	1	58.181	87.161	.000
LnSize	80.968	1	80.968	121.298	.000
LnDuration	24.307	1	24.307	36.414	.000
DevType	.622	2	.311	.466	.628
Error	141.513	212	.668		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .638 (Adjusted R Squared = .632)					

```
UNIANOVA LnEffort BY OrgType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration OrgType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	277.904[a]	13	21.377	38.225	.000
Intercept	39.767	1	39.767	71.108	.000
LnSize	65.593	1	65.593	117.287	.000
LnDuration	25.620	1	25.620	45.811	.000
OrgType	28.606	11	2.601	4.650	.000
Error	113.529	203	.559		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .710 (Adjusted R Squared = .691)					

```
UNIANOVA LnEffort BY BusiType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration BusiType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	282.459[a]	12	23.538	44.064	.000
Intercept	56.720	1	56.720	106.180	.000
LnSize	60.304	1	60.304	112.888	.000
LnDuration	25.145	1	25.145	47.072	.000
BusiType	33.160	10	3.316	6.208	.000
Error	108.974	204	.534		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .722 (Adjusted R Squared = .705)					

```
UNIANOVA LnEffort BY AppType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
```

```
/CRITERIA=ALPHA(.05)
/DESIGN=LnSize LnDuration AppType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	252.512[a]	9	28.057	41.806	.000
Intercept	85.516	1	85.516	127.424	.000
LnSize	78.305	1	78.305	116.680	.000
LnDuration	24.026	1	24.026	35.801	.000
AppType	3.213	7	.459	.684	.686
Error	138.921	207	.671		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .645 (Adjusted R Squared = .630)					

```
UNIANOVA LnEffort BY ArchType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration ArchType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	263.379[a]	3	87.793	146.032	.000
Intercept	99.541	1	99.541	165.573	.000
LnSize	69.760	1	69.760	116.037	.000
LnDuration	25.030	1	25.030	41.633	.000
ArchType	14.081	1	14.081	23.421	.000
Error	128.054	213	.601		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .673 (Adjusted R Squared = .668)					

```
UNIANOVA LnEffort BY Devplat WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration Devplat.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	257.116[a]	4	64.279	101.455	.000
Intercept	87.494	1	87.494	138.096	.000
LnSize	75.227	1	75.227	118.735	.000
LnDuration	22.058	1	22.058	34.816	.000
Devplat	7.817	2	3.909	6.169	.002
Error	134.317	212	.634		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .657 (Adjusted R Squared = .650)

```
UNIANOVA LnEffort BY LangType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration LangType.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	254.482[a]	4	63.620	98.484	.000
Intercept	66.522	1	66.522	102.975	.000
LnSize	83.488	1	83.488	129.239	.000
LnDuration	20.456	1	20.456	31.665	.000
LangType	5.183	2	2.591	4.012	.019
Error	136.951	212	.646		
Total	13309.405	217			
Corrected Total	391.433	216			

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	254.482[a]	4	63.620	98.484	.000
Intercept	66.522	1	66.522	102.975	.000
LnSize	83.488	1	83.488	129.239	.000
LnDuration	20.456	1	20.456	31.665	.000
LangType	5.183	2	2.591	4.012	.019
Error	136.951	212	.646		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .650 (Adjusted R Squared = .644)					

```
UNIANOVA LnEffort BY DBMS WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration DBMS.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	286.034[a]	17	16.826	31.768	.000
Intercept	64.903	1	64.903	122.541	.000
LnSize	60.794	1	60.794	114.784	.000
LnDuration	15.108	1	15.108	28.525	.000
DBMS	36.736	15	2.449	4.624	.000
Error	105.399	199	.530		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .731 (Adjusted R Squared = .708)					

```
UNIANOVA LnEffort BY Umethod WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration Umethod.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	249.299[a]	2	124.649	187.674	.000
Intercept	87.011	1	87.011	131.005	.000
LnSize	82.139	1	82.139	123.670	.000
LnDuration	24.371	1	24.371	36.693	.000
Umethod	.000	0	.	.	.
Error	142.134	214	.664		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .637 (Adjusted R Squared = .633)					

```
UNIANOVA LnEffort BY Rmethod WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration Rmethod.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	253.417[a]	4	63.354	97.316	.000
Intercept	57.837	1	57.837	88.841	.000
LnSize	79.725	1	79.725	122.463	.000
LnDuration	27.793	1	27.793	42.692	.000
Rmethod	4.119	2	2.059	3.163	.044
Error	138.016	212	.651		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .647 (Adjusted R Squared = .641)					

```
UNIANOVA LnEffort BY PriLang CountApproach WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
```

```
/INTERCEPT=INCLUDE
/CRITERIA=ALPHA(.05)
/DESIGN=LnSize LnDuration PriLang CountApproach PriLang*CountApproach.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	297.749[a]	19	15.671	32.953	.000
Intercept	48.011	1	48.011	100.958	.000
LnSize	55.700	1	55.700	117.126	.000
LnDuration	23.317	1	23.317	49.032	.000
PriLang	47.506	15	3.167	6.660	.000
CountApproach	4.177E-6	1	4.177E-6	.000	.998
PriLang * CountApproach	.028	1	.028	.059	.809
Error	93.684	197	.476		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .761 (Adjusted R Squared = .738)					

```
UNIANOVA LnEffort BY PriLang DevType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang DevType PriLang*DevType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	307.401[a]	30	10.247	22.680	.000
Intercept	44.860	1	44.860	99.294	.000
LnSize	54.568	1	54.568	120.783	.000
LnDuration	20.959	1	20.959	46.391	.000
PriLang	33.488	15	2.233	4.942	.000
DevType	.524	2	.262	.580	.561
PriLang * DevType	9.669	11	.879	1.946	.036
Error	84.032	186	.452		

Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .785 (Adjusted R Squared = .751)					

```
UNIANOVA LnEffort BY PriLang OrgType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang OrgType PriLang*OrgType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	317.653[a]	44	7.219	16.830	.000
Intercept	35.589	1	35.589	82.967	.000
LnSize	46.946	1	46.946	109.444	.000
LnDuration	18.995	1	18.995	44.283	.000
PriLang	22.927	15	1.528	3.563	.000
OrgType	8.527	11	.775	1.807	.056
PriLang * OrgType	10.474	16	.655	1.526	.095
Error	73.780	172	.429		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .812 (Adjusted R Squared = .763)					

```
UNIANOVA LnEffort BY PriLang AppType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang AppType PriLang*AppType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.

Corrected Model	315.874[a]	41	7.704	17.843	.000
Intercept	58.535	1	58.535	135.571	.000
LnSize	54.192	1	54.192	125.513	.000
LnDuration	24.235	1	24.235	56.131	.000
PriLang	31.066	15	2.071	4.797	.000
AppType	3.951	7	.564	1.307	.249
PriLang * AppType	12.333	17	.725	1.680	.050
Error	75.559	175	.432		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .807 (Adjusted R Squared = .762)					

```
UNIANOVA LnEffort BY PriLang ArchType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang ArchType PriLang*ArchType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	303.289[a]	23	13.186	28.873	.000
Intercept	71.114	1	71.114	155.711	.000
LnSize	59.139	1	59.139	129.491	.000
LnDuration	22.397	1	22.397	49.040	.000
PriLang	30.853	15	2.057	4.504	.000
ArchType	2.966	1	2.966	6.495	.012
PriLang * ArchType	2.148	5	.430	.941	.456
Error	88.144	193	.457		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .775 (Adjusted R Squared = .748)					

```
UNIANOVA LnEffort BY PriLang Devplat WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang Devplat PriLang*Devplat.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	305.402[a]	24	12.725	28.399	.000
Intercept	66.358	1	66.358	148.096	.000
LnSize	58.205	1	58.205	129.899	.000
LnDuration	23.521	1	23.521	52.493	.000
PriLang	31.139	15	2.076	4.633	.000
Devplat	1.006	2	.503	1.123	.327
PriLang * Devplat	6.392	5	1.278	2.853	.016
Error	86.031	192	.448		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .780 (Adjusted R Squared = .753)					

```
UNIANOVA LnEffort BY PriLang LangType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang LangType PriLang*LangType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	298.071[a]	18	16.560	35.119	.000
Intercept	65.501	1	65.501	138.914	.000
LnSize	58.745	1	58.745	124.584	.000
LnDuration	23.324	1	23.324	49.466	.000
PriLang	43.589	14	3.114	6.603	.000
LangType	.352	1	.352	.746	.389
PriLang * LangType	.000	0	.	.	.
Error	93.362	198	.472		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .761 (Adjusted R Squared = .740)					

```
UNIANOVA LnEffort BY PriLang DBMS WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang DBMS PriLang*DBMS.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	311.908[a]	36	8.664	19.611	.000
Intercept	49.549	1	49.549	112.152	.000
LnSize	51.167	1	51.167	115.813	.000
LnDuration	21.693	1	21.693	49.100	.000
PriLang	21.296	14	1.521	3.443	.000
DBMS	13.561	14	.969	2.193	.010
PriLang * DBMS	.823	5	.165	.373	.867
Error	79.525	180	.442		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .797 (Adjusted R Squared = .756)

```
UNIANOVA LnEffort BY PriLang Umethod WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang Umethod PriLang*Umethod.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	297.719[a]	17	17.513	37.188	.000
Intercept	65.949	1	65.949	140.041	.000
LnSize	60.281	1	60.281	128.006	.000
LnDuration	23.557	1	23.557	50.024	.000
PriLang	48.421	15	3.228	6.855	.000

Umethod	.000	0	.	.	.
PriLang * Umethod	.000	0	.	.	.
Error	93.714	199	.471		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .761 (Adjusted R Squared = .740)					

```
UNIANOVA LnEffort BY PriLang Rmethod WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang Rmethod PriLang*Rmethod.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	307.445[a]	28	10.980	24.578	.000
Intercept	39.709	1	39.709	88.885	.000
LnSize	60.010	1	60.010	134.327	.000
LnDuration	22.846	1	22.846	51.139	.000
PriLang	38.440	15	2.563	5.736	.000
Rmethod	2.553	2	1.277	2.858	.060
PriLang * Rmethod	8.614	9	.957	2.142	.028
Error	83.988	188	.447		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .785 (Adjusted R Squared = .753)					

```
UNIANOVA LnEffort BY PriLang BusiType CountApproach WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang BusiType CountApproach PriLang*BusiType
PriLang*CountApproach BusiType*CountApproach PriLang*BusiType*CountApproach.
```

Tests of Between-Subjects Effects
Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	318.755[a]	43	7.413	17.645	.000
Intercept	34.037	1	34.037	81.020	.000
LnSize	50.095	1	50.095	119.245	.000
LnDuration	21.503	1	21.503	51.185	.000
PriLang	20.534	15	1.369	3.259	.000
BusiType	12.314	9	1.368	3.257	.001
CountApproach	.056	1	.056	.134	.715
PriLang * BusiType	6.875	15	.458	1.091	.368
PriLang * CountApproach	.000	0	.	.	.
BusiType * CountApproach	.000	0	.	.	.
PriLang * BusiType * CountApproach	.000	0	.	.	.
Error	72.678	173	.420		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .814 (Adjusted R Squared = .768)					

```
UNIANOVA LnEffort BY PriLang BusiType DevType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang BusiType DevType PriLang*BusiType
PriLang*DevType BusiType*DevType PriLang*BusiType*DevType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	325.545[a]	58	5.613	13.460	.000
Intercept	32.758	1	32.758	78.555	.000
LnSize	39.464	1	39.464	94.635	.000
LnDuration	20.770	1	20.770	49.806	.000
PriLang	14.922	15	.995	2.386	.004
BusiType	10.241	10	1.024	2.456	.009
DevType	.571	2	.285	.685	.506
PriLang * BusiType	5.809	11	.528	1.266	.249

PriLang * DevType	3.005	6	.501	1.201	.308
BusiType * DevType	1.000	5	.200	.480	.791
PriLang * BusiType * DevType	.016	1	.016	.039	.843
Error	65.888	158	.417		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .832 (Adjusted R Squared = .770)					

```
UNIANOVA LnEffort BY PriLang BusiType OrgType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang BusiType OrgType PriLang*BusiType
PriLang*OrgType BusiType*OrgType PriLang*BusiType*OrgType.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	320.764[a]	47	6.825	16.321	.000
Intercept	35.318	1	35.318	84.462	.000
LnSize	47.778	1	47.778	114.258	.000
LnDuration	17.061	1	17.061	40.800	.000
PriLang	18.812	15	1.254	2.999	.000
BusiType	1.211	2	.605	1.448	.238
OrgType	2.066	5	.413	.988	.427
PriLang * BusiType	1.484	1	1.484	3.548	.061
PriLang * OrgType	.000	0	.	.	.
BusiType * OrgType	.000	0	.	.	.
PriLang * BusiType * OrgType	.000	0	.	.	.
Error	70.669	169	.418		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .819 (Adjusted R Squared = .769)					

```
UNIANOVA LnEffort BY PriLang BusiType AppType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang BusiType AppType PriLang*BusiType
PriLang*AppType BusiType*AppType PriLang*BusiType*AppType.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	326.487[a]	64	5.101	11.939	.000
Intercept	34.870	1	34.870	81.610	.000
LnSize	47.718	1	47.718	111.680	.000
LnDuration	18.615	1	18.615	43.566	.000
PriLang	12.991	15	.866	2.027	.017
BusiType	5.110	9	.568	1.329	.226
AppType	1.124	7	.161	.376	.915
PriLang * BusiType	1.323	5	.265	.619	.685
PriLang * AppType	2.471	7	.353	.826	.567
BusiType * AppType	1.164	3	.388	.908	.439
PriLang * BusiType * AppType	.000	0	.	.	.
Error	64.946	152	.427		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .834 (Adjusted R Squared = .764)

```
UNIANOVA LnEffort BY PriLang BusiType Devplat WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang BusiType Devplat PriLang*BusiType
PriLang*Devplat BusiType*Devplat PriLang*BusiType*Devplat.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	322.215[a]	49	6.576	15.865	.000
Intercept	33.214	1	33.214	80.135	.000
LnSize	46.791	1	46.791	112.892	.000
LnDuration	21.437	1	21.437	51.720	.000
PriLang	11.057	14	.790	1.905	.029
BusiType	11.940	10	1.194	2.881	.002
Devplat	.326	2	.163	.394	.675
PriLang * BusiType	2.560	10	.256	.618	.797
PriLang * Devplat	2.341	3	.780	1.883	.134
BusiType * Devplat	.112	1	.112	.270	.604
PriLang * BusiType * Devplat	.000	0	.	.	.
Error	69.218	167	.414		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .823 (Adjusted R Squared = .771)

```
UNIANOVA LnEffort BY PriLang BusiType LangType WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang BusiType LangType PriLang*BusiType
PriLang*LangType BusiType*LangType PriLang*BusiType*LangType.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	318.920[a]	43	7.417	17.695	.000
Intercept	39.011	1	39.011	93.072	.000
LnSize	48.867	1	48.867	116.586	.000
LnDuration	21.379	1	21.379	51.007	.000
PriLang	12.355	13	.950	2.267	.009
BusiType	11.388	10	1.139	2.717	.004
LangType	.222	1	.222	.529	.468
PriLang * BusiType	2.133	7	.305	.727	.649

PriLang * LangType	.000	0	.	.	.
BusiType * LangType	.000	0	.	.	.
PriLang * BusiType * LangType	.000	0	.	.	.
Error	72.513	173	.419		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .815 (Adjusted R Squared = .769)					

```
UNIANOVA LnEffort BY PriLang BusiType DBMS WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang BusiType DBMS PriLang*BusiType
PriLang*DBMS BusiType*DBMS PriLang*BusiType*DBMS.
```

Tests of Between-Subjects Effects					
Dependent Variable:LnEffort					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	321.937[a]	51	6.312	14.987	.000
Intercept	35.800	1	35.800	84.996	.000
LnSize	47.303	1	47.303	112.309	.000
LnDuration	17.489	1	17.489	41.522	.000
PriLang	10.101	14	.721	1.713	.057
BusiType	4.871	7	.696	1.652	.124
DBMS	2.415	6	.402	.956	.457
PriLang * BusiType	1.336	1	1.336	3.172	.077
PriLang * DBMS	.162	1	.162	.385	.536
BusiType * DBMS	.359	2	.180	.426	.654
PriLang * BusiType * DBMS	.000	0	.	.	.
Error	69.496	165	.421		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .822 (Adjusted R Squared = .768)					

```
UNIANOVA LnEffort BY PriLang BusiType Umethod WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
```

```
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang BusiType Umethod PriLang*BusiType
PriLang*Umethod BusiType*Umethod PriLang*BusiType*Umethod.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	318.699[a]	42	7.588	18.153	.000
Intercept	37.504	1	37.504	89.719	.000
LnSize	50.058	1	50.058	119.752	.000
LnDuration	21.483	1	21.483	51.394	.000
PriLang	20.521	15	1.368	3.273	.000
BusiType	12.592	10	1.259	3.012	.002
Umethod	.000	0	.	.	.
PriLang * BusiType	7.118	15	.475	1.135	.328
PriLang * Umethod	.000	0	.	.	.
BusiType * Umethod	.000	0	.	.	.
PriLang * BusiType * Umethod	.000	0	.	.	.
Error	72.734	174	.418		
Total	13309.405	217			
Corrected Total	391.433	216			

a. R Squared = .814 (Adjusted R Squared = .769)

```
UNIANOVA LnEffort BY PriLang BusiType Rmethod WITH LnSize LnDuration
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=LnSize LnDuration PriLang BusiType Rmethod PriLang*BusiType
PriLang*Rmethod BusiType*Rmethod PriLang*BusiType*Rmethod.
```

Tests of Between-Subjects Effects

Dependent Variable:LnEffort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	320.263[a]	49	6.536	15.337	.000
Intercept	32.368	1	32.368	75.952	.000

LnSize	49.783	1	49.783	116.816	.000
LnDuration	20.514	1	20.514	48.136	.000
PriLang	14.025	15	.935	2.194	.008
BusiType	4.871	9	.541	1.270	.257
Rmethod	.175	2	.087	.205	.815
PriLang * BusiType	3.403	8	.425	.998	.440
PriLang * Rmethod	.012	1	.012	.027	.869
BusiType * Rmethod	.444	2	.222	.520	.595
PriLang * BusiType * Rmethod	.000	0	.	.	.
Error	71.170	167	.426		
Total	13309.405	217			
Corrected Total	391.433	216			
a. R Squared = .818 (Adjusted R Squared = .765)					

REFERENCES:

[1] H.Leung and Z.Fan. "Software Cost Estimation", *Handbook of Software Engineering and Knowledge Engineering*", Hong Kong Polytechnic University, 2002

[2] Abdel-Hamid, T.K."On the utility of historical project statistics for cost and schedule estimation: results from a simulation-based case study", *Journal of Systems & Software*, Vol.13, No.l, pp.71-82. Sep.1990

[3] C.D. Wrigley and A.S.Dexter. "A model for measuring information system size", *MIS Quarterly*, pp.245-257. Jun.1991

[4] H. Yan-Chin and C. D. McDevitt. " Determination of Optimal Resource Allocation for Software Development an application of a software equation", *Information and Management*, No. 18, PP. 79-85.1990

[5] B.A.Kitchenham et al. "The MERMAID approach to software cost estimation", in *Proceedings of the Annual ESPRIT Conference, Brussels*, Nov. 12-15, 1990, pp.296-314.

[6] M. Jorgensen and M. Shepperd, "A Systematic Review of Software Development Cost Estimation Studies", in *IEEE Transactions on Software Engineering*, 2007, Vol.33, No.1, pp. 33-53

[7] C. F. Kemerer. "An empirical validation of software cost estimation models",in *communication of the ACM, Vol. 30, No.5*, May. 1987, pp. 416-429

[8] F. J. Heemstra. " Software cost estimation", *Information and software technology*, Vol. 34, No. 10, 1992, pp. 627-639

[9] B. W. Boehm. *Software engineering economics*, New Jersey: Prentice –Hall, 1981

[10] M. van Genuchten and H. Koolen. "On the use of software cost models", *Information and Management*, no. 1, Aug.1991 pp. 37-44.

[11] www.isbsg.org, 2004

[12] C. Wohlin and D. Milicic. "Distribution Patterns of Effort Estimations," *30th EUROMICRO Conference*, IEEE, 2004, pp.422-429.

[13] The Standish Group International, The Standish Group CHAOS Report, 2009.

[14] W. Royce. *Software project management: a unified framework*, Addison Wesley, 1998

[15] T.K. Abdel-Hamid, "Adapting, Correcting and Perfecting Software Estimates: A Maintenance Metaphor," in *IEEE Computer*, March 1993, pp. 20–29.

[16] R. E. Fairley, "Recent Advances in Software Estimation Techniques," in *Proceedings of the IEEE 14th International Conference on Software Engineering*, 1992, pp. 382–391.

[17] B.H. Rowlands. "Estimating instructional systems development effort and costs", *Proceedings of the 7th Annual Conf. of the Australian Society for Computers in Learning in Tertiary Education, Gold Coast*, 1989, pp.340-348.

[18] F.P. Brooks. *The Mythical Man-Month: Essays on software engineering*, Addison-Wesley, 1975.

[19] R. Neumann and L.Santillo. "Experience with the Usage of COCOMO II Effort Estimation in Embedded Software Context", In *Proceedings of SMEF* 2006.

[20] K. Jaswinder et al."Comparative analysis of the software effort estimation models", *proceedings of world academy of science, eng. and tech*. Dec. 2008, Vol.36, pp. 485-487

[21] Juan J. Cuadrado-Gallego et al. "Enhancing input value selection in parametric software cost estimation models through second level cost drivers", in *Software quality journal*, 2006, No. 14, pp. 339-357

[22] A Guide to the Project Management Body of Knowledge, Ed. 3rd, 2004

[23] B. W. Boehm. *Software engineering economics* , New Jersey: Prentice –Hall, 1981

[24] L. H. Putnam. "A General Empirical Solution to the Macro Software Sizing and Estimating Problem," in *IEEE Transactions On Software Engineering*, 1978, pp. 345- 360

[25] A. J. Albrecht. "Measuring Application Development Productivity," in *Proc. of the IBM Applications Development Symposium*, Oct.1979, pp. 83-92

[26] *SEER-SEM Users Manual*. El Segundo, CA: Galorath Inc., Mar. 2001

[27] H. A. Rubin. "Macro estimation of software development parameters: The Estimacs system", In *SOFTFAIR Conference on Software Development Tools, Techniques and alternatives*, 1983, pp. 109-118.

[28] Lo. Bruce and G. Xiangzhu. "Assessing Software Cost Estimation Models: Criteria For Accuracy, Consistency And Regression", in *Australasian Journal of Information Systems*, Vol.5, No. 1, 1997
[29] Y. Miyazaki, and K. Mori, "COCOMO evaluation and tailoring", *Eighth Int. Conf. Software Engineering*, 1985, pp. 292-299.
[30] M. Boraso, C. Montangero, and H. Sedehi, "Software cost estimation: An experimental study of model performances," tech. report, 1996.
[31] O. Benediktsson, D. Dalcher, K. Reed, and M. Woodman, "COCOMO based effort estimation for iterative and incremental software development," *Software Quality Journal*, vol. 11, pp. 265–281, 2003.
[32] T. Menzies, D. Port, Z. Chen, J. Hihn, and S. Stukes, "Validation methods for calibrating software effort models," in *ICSE '05: Proceedings of the 27th international conference on Software engineering, (New York, NY, USA)*, pp. 587–595, ACM Press, 2005.
[33]D. R. Jeffery, and G. C. Low, "Calibrating estimation tools for software development", *Software Engineering Journal*, July 1990, pp. 215-221.
[34] D. R. Jeffery, G. C. Low, and M. Barnes, "A comparison of function point counting techniques", *IEEE Trans on Soft. Eng*., vol. 19, no. 5, 1993, pp. 529-532.
[35] M. costamagna et al. "an integrated environment for productivity assesment of telecommunications software", in *proc. of European software cost modeling conference*, May 17-19, 1995, pp. 21.1-21.11
[36] D. V. Ferens, and R. B. Gumer, "An evaluation of three function point models of estimation of software effort", *IEEE National Aerospace and Electronics Conference*, vol. 2, 1992, pp. 625-642.
[37] Sun-Jen Huang, Nan-Hsing Chiu and Yu-Jen Liu. "A comparative evaluation on the accuracies of software effort estimates from clustered data" in *Information and Software Technology*, 2008, pp. 879- 888.
[38] Qin Liu et al. "Evaluation of preliminary data analysis framework in software cost estimation based on ISBSG R9 Data" in *Software Quality Journal*, 2008, pp. 411- 458
[39] W. Creswell. *Research design- Qualitative, Quantitative and Mixed method approaches*, Sage Publication, 2002
[40] C. W. Dawson. *The Essence of Computing Projects: A Student's Guide*.
[41] R. E. Park, "PRICE S: The calculation within and why", *Proceedings of ISPA Tenth Annual Conference, Brighton, England*, July 1988.
[42] C. Jones. *Programming productivity* McGraw-Hill (1986)
[43] Rubin, H A 'Interactive macro-estimation of software life cycle parameters via personal computer: a technique for improving customer/developer communication' Proc. *Symp. on application & assessment of automated tools for software development, IEEE, San Francisco* (1983).
[44] L. M. Laird, and M. C. Brennan. *Software Measurement and Estimation: A Practical Approach*, 2006, Wiley-IEEE Computer Society Pr, ISBN: 0-471-67622-5.
[45] www.ifpug.org, 2006
[46] Norden and V. Peter. *Useful tools for project management, management of production*, M.K. Starr, Penguin Books, Inc., Baltmore, Md., 1970, pp. 71-101
[47] S. H. Kan, V. R. Basili and L. N. Shapiro. "Software quality: an overview from the perspective of total quality management", in *IBM Systems Journal*, Vol.33, No.l, Mar. 1994, pp.4-19.
[48] M. Andolfi et al. " A multicriteria-based methodology for the evaluation of software cost estimation models and tools", in *conference on software measurement and management*, Vol. XXIV, No.4, Aug.1996, pp. 643-659
[49] N. Fenton. "Software measurement: a necessary scientific basis", in *IEEE Transactions on Software Engineering*, Vol.20, No.3, Mar. 1994, pp. 199-206.
[51] T.Mukhopadhyay, S.S.Vicinanza, M.J.Prietula, "Examining the feasibility of a case-based reasoning model for software effort estimation", *MIS Quarterly*, Jun. 1992, pp. 155-170.
[52] E. Deeson, *Collins Dictionary of Information Technology*, Pub. Harper Collins, Glasgow, 1991

[52] T.M. Khoshgoftaar, J.C. Munson, et al. "Predictive modeling techniques of software quality from software measures",in *IEEE Transactions on Software Engineering*, Nov.1992, Vol.18, No.11, pp.979-987.
[53] A. D. Stephen. *Forecasting principles and application*. Irwin: McGraw-Hill, 1997.
[54] Naoki Ohsugi et al "Is This Cost Estimate Reliable? – The Relationship between Homogeneity of Analogues and Estimation Reliability", *IEEE First International Symposium on Empirical Software Engineering and Measurement*, 2007.
[55] M. Nikolaos and L. Angelis, " Bootstrap prediction intervals for a semi-parametric software cost estimation model",*35th Euromicro conf on software engineering and advanced application*, 2009.
[56] C. Mair and M. Shepperd." the consistency of empirical comparissions of regression and analogy based software project cost predictions", *Proc. International symp on emprirical software engineering (ISESE), Noosa Heads, Queensland, Australia*, Nov. 17-18 2005, pp. 509-518.
[57] M. wissmann et al. " Role of catagorical variables in Multicollinearity in the linear regression model," technical report No. 008, 2007, Department of statistic, University of Munich.
[58] P. Anglin and R. Gencay, " Semiparametric estimation of a hedonic price function," *Journal of applied economics*, Vol.11, No. 6, 1996, pp.633-648.
[59] F. Walkerden and R. Jeffery, "Empirical study of analogy-based software effort estimation," *Empirical Software Engineering*, vol. 4, pp. 135-158, 1999.
[60] M. Shepperd and G. Kadoda, "Comparing software prediction techniques using simulation," *IEEE Transactions on Software Engineering*, vol. 27, pp. 1014-1022, 2001.
[61] E. Mendes and B. Kitchenham, "Further comparison of cross-company and within-company effort estimation models for Web applications," *Proceedings International Software Metrics Symposium*, pp. 348-357, 2004.
[62] I. Myrtveit and E. Stensrud, "A controlled experiment to assess the benefits of estimation with analogy and regression models," *IEEE Transactions on Software Engineering,* vol. 25, pp. 510-525, 1999.
[63] M. Shepperd and C. Schofield, "Estimating software project effort using analogies," *IEEE Transactions on Software Engineering*, vol. 23, pp. 736-743, 1997.
[64] L. Angelis and I. Stamelos, "Simulation tool for efficient analogy based cost estimation," *Empirical Software Engineering*, vol. 5, pp. 35-68, 2000.
[65] E. Mendes and N. Mosley, "Further investigation into the use of CBR and stepwise regression to predict development effort for web hypermedia applications," *Proceedings International Symposium on Empirical Software Engineering, ISESE,* 2002,
[66] R. Jeffery, M. Ruhe and I. Wieczorek, "Using public domain metrics to estimate software development effort," *Proceedings of the International Software Metrics Symposium,* 2001.
[67] E. Mendes and S. Counsell, "Early web size measures and effort prediction for web costimation," in *Proceedings of 9th International Software Metrics Symposium, METRICS*, 2003.
[68] G. R. Finnie, G. E. Wittig and J-M. Desharnais, "A comparison of software effort estimation techniques: Using function points with neural networks, case-based reasoning and regression models," *Journal of Systems and Software*, vol. 39, pp. 281-289, 1997.
[69] T. Mukhopadhyay, S. S. Vicinanza, and M. J. Prietula, "Examining the feasibility of a casebased reasoning model for software effort estimation," *MIS Quarterly*, vol. 16, pp. 155-171, 1992.
[70] Briand et al. "An assessment and comparison of common software cost estimation modeling techniques," *presented at The 21st International Conference on Software Engineering, Los Angeles, California*, 1999.
[71] F. Niessink and H. van Vliet, "Predicting Maintenance Effort with Function Points," *presented at International Conference on Software Maintenance (ICSM '97), Bari, ITALY*, 1997.
[72] B.Kitchenhamn and S. Linkmand, " Estimates, Uncertainity and Risks", *IEEE Transactions on software engineering*, Vol.14, No.3, 1997, pp.69-74

[73] M. Jorgensen et al. "Better sure than safe? Overconfidence in judgement based software development effort predictions intervals," *Journal of systems and software*, Vol. 70, No. 1-2, 2004, pp.79-93.
[74] "Repository Data Release 10 - Field Descriptions", ISBSG, 2007. Internet:
[http://www.nesma.nl/download/isbsg/R10%20Field%20Descriptions.pdf]
[75] SPSS 17.0, http://www.spss.com
[76] M. Shepperd and C. Schofield, "Estimating software project effort using analogy", *IEEE Trans. Soft. Eng. SE-23:12*, 1997, pp. 736-743.
[77] K. D. Maxwell, 2002. *Applied Statistics for Software Managers*, Prentice Hall PTR,ISBN 0-13-041789-0.
[78] E. J. Pedhazur, 1997. *Multiple Regression in Behavioral Research: Explanation and Prediction*, Third Edition, Harcourt Brace College Publishers, ISBN: 0-03-072831-2.
[79] S. K. Kachigan, *Statistical Analysis: An Interdisciplinary Introduction to Univariate and Multivariate Methods*, Radius Press, 1986.ISBN 0-942152-99-1.
[80] Conte, S., Dunsmore, H., Shen, V.Y., 1986. *Software Engineering Metrics and Models*. Benjamin Cummings, Menlo Park, CA.
[81] B. Kitchenham, S.G. MacDonell, L. M Pickard, M.J.Shepperd,"What accuracy statistics really measure". IEEE Proc. Softw. 148 (3), 81–85. 2001.
[82] "What you can find in the ISBSG Repository – January 2007", ISBSG, 2007, Internet:
[http://www.nesma.nl/download/isbsg/R10%20Demographics.pdf]
[83] Bournemouth University,UK. archANGEL version 0.91
[84] F. Walkerden and R. Jeffery, "An empirical study of analogy-based software effort estimation," *Journal of Empirical Software Engineering*, vol. 4, pp. 135-158, 1999.
[85] M. Shepperd and C. Schofield, "Estimating software project effort using analogies," *IEEE Transactions on Software Engineering*, vol.13, pp. 736-743, 1997.
[86] S.D Conte, HE Dunsmore, and VY Shen, *Software Engineering Metrics and Models*, Benjamin-Cummings, Menlo Park CA, 1986.
[87] Miyazaki, Y. Takanou, A. & Nozaki, H. "Method to estimate parameter values in software prediction models", *Information & Software Technology*, Apr.1991, Vol.33, No.3, pp.239-243.
[88] E. Stensrud, T.Foss, B. Kitchenham and I. Myrtveit, "An empirical validation of the relationship between magnitude of relative error and project size", *Proceedings of the Eighth IEEE Symposium on Software Metrics*, 2002.
[89] L. Angelis and N.Mitas, "Comparing cost prediction models by re-sampling techniques", in *The Journal of System and Software 81*, 2008, pp. 616-632
[90] L. Angelis and N.Mitas, "Combining regression and estimation by analogy in a semi-parametric model for software cost estimation", in *ESEM' 08*, Kaiserslautern, Oct. 9-12, 2008. pp. 70-79
[91] Marouane, R., & Mili, A.. Economics of software project management in Tunisia: Basic Tucomo. *Information and Software Technology*, 31, 251–257. 1989
[92] Liu, Q. "Optimal utilization of historical data sets for the construction of software cost prediction models". 2005 PhD thesis, School of Computing, Engineering and Information Sciences, Northumbria University, UK.
[93] J. F Hair, R. E Anderson, R. L Tatham, & W. C. Black. *Multivariate Data Analysis* (4th Ed.) Prentice-Hall, Inc. 1995.