

## An Empirical Analysis of Information Technology Operations Cost

Masateru Tsunoda<sup>1</sup>, Akito Monden<sup>1</sup>, Ken-ichi Matsumoto<sup>1</sup>, Akihiko Takahashi<sup>2</sup>,  
and Tomoki Oshino<sup>2</sup>

<sup>1</sup>Graduate School of Information Science, Nara Institute of Science and Technology  
{masate-t, akito-m, matumoto}@is.naist.jp

<sup>2</sup>Economic Research Institute, Economic Research Association  
{er102, er352}@zai-keicho.or.jp

### **Abstract:**

*When outsourcing IT (Information Technology) operations, it is important to estimate operations cost accurately. To help IT operations cost estimation, we analyzed cross-organization data which includes 73 IT operation cases collected from Japanese companies/organizations. Operations cost is roughly decided by the number of operations staff and unit cost of staff. So we clarified attributes which have relationships to the number of staff or unit cost of staff to help analogous estimating of them. The result showed that the number of computer terminals and network range are major factors of the number of staff. Also, the number of UNIX server and process standardization are major factors of unit of staff. In addition, we tried to build estimation models for the number of staff and total cost of operations. The number of staff estimation model is used when IT operations service providers estimate operations cost, and the operations cost estimation model is assumed to be used by IT operations service purchasers. In the experiment, the number of staff estimation model showed moderate accuracy, but the total operations cost estimation model showed low accuracy.*

### **Keywords**

*Estimation, staff, cost, cross organization data, linear regression*

## **1 Introduction**

Recently, IT (Information Technology) operations (or software system operations) are often outsourced. After software is released, IT operations are necessary to sustain availability of the software system. Activities of IT operations consist of manipulations such as daily backup, monitoring software and hardware status, supporting software and hardware recovery when incidents occurred, and so on.

When outsourcing IT operations, it is important to estimate operations cost accurately. For IT operations service providers, accurate cost estimation is the basis for generating appropriate profits. For IT operations service purchasers (system users), accurate cost estimation by them is useful to validate operations cost offered by IT operations service providers.

Operations cost is roughly decided by the number of operations staff and unit cost of the staff (per hour). To help analogous estimating of the number of staff and unit cost of staff, we clarify attributes which have relationships to them. Also, we built linear regression models to estimate the number of staff and total cost of operations (per year). The number of staff estimation model is used when IT operations service providers estimate operations cost. By the model, operations cost can be decided because IT operations service providers know unit cost of staff. The operations cost estimation model is assumed to be used by IT operations service purchasers.

In the analysis, we used 73 cases of cross-organization data collected from Japanese companies/organizations. Analysis results using cross-organization data is expected to have more generality than using single-organization data.

In what follows, Section 2 explains dataset used in the analysis. Section 3 describes procedure of the analysis. Section 4 shows results of the analysis and discussion of them. Section 5 introduces related works. In the end, Section 6 concludes the paper with a summary.

## 2 Dataset

Dataset includes 108 IT operations cases which were collected from 107 Japanese companies/organizations in 2007 by Economic Research Association [6]. In the dataset, the scope of IT operations is defined as operation process of ISO/IEC 12207 [8], and operations staff means people who engage in both service support and service delivery defined in ITIL (Information Technology Infrastructure Library) version 2 [3]. Each case is the most typical IT operations case in a company/organization. Although the cases were collected from both IT operations service purchasers and IT operations service providers, we selected 73 cases collected from only the purchasers, because total cost of operations was not collected from the providers. Note that process standardization, ITIL status, and SLA (Service License Agreement) status signify the status of each company/organization, not each IT operations case (For example, process standardization indicates whether the company's process is standardised or not).

Attributes of the cases are as follows.

### Numerical attributes

- **The number of {mainframes, Windows servers, UNIX servers, computer terminals}**: the number of equipments included in the system.
- **The number of users**: the number of people using the system.
- **The number of program**: the number of program of software on the system. The attribute was used instead of software size (function point) in the analysis.

- **Operation time:** total system operation time per week.
- **Unit cost of staff:** unit cost of providers' operations staff per hour (Japanese yen).
- **The number of staff:** total number of staff of the IT operations service provider and the purchaser engaged in IT operations.
- **Total cost:** total contract cost for IT operations per year (Japanese yen).
- **The percentage of {service request, incident, problem, change, release, configuration, service level, capacity, IT service continuity, availability} management staff, the percentage of financial management staff for IT services:** assignment of operations staff for each management area. The classifications of management areas are based on ITIL version 2 [3].

#### Nominal attributes

- **Process standardization:** denote operation process is standardized, standardization is work in progress, or operation process is not standardized.
- **ITIL status:** denote ITIL is applied whole operations, ITIL is applied part of operations, adoption of ITIL is under review, or ITIL is not applied.
- **SLA status:** denote SLA is applied all agreements, SLA is applied part of agreements, adoption of SLA is under review, or SLA is not applied.
- **Business sector:** business sector of the company/organization where IT operations are performed.
- **System service time:** system service providing time represented by 5 days, 5 days and 24 hours, 7 days, or 7 days and 24 hours.
- **Network range:** denote the system is connected to other companies/organizations' system, the system is connected to branch offices' system, or the system is connected to local systems only.
- **Using {Microsoft SQL Server, Oracle, open source database}:** The database is used or not on the system.

### 3 Analysis procedure

First, we focus on unit cost of staff and the number of staff, and analyze relationships to other attributes, to help analogous estimation of unit cost of staff and the number of staff. When analyzing relationships between them and other numerical attributes, we use Spearman's rank correlation coefficient, to avoid influences of outliers. When the analysis results are suspected to include spurious relationships, we apply partial correlation to correct bias.

We use adjusted variance explained ( $\omega^2$ ) of one-way ANOVA [17] when analyzing relationships between them and other nominal attributes. The value range of adjusted variance explained is [0%, 100%], and a large value indicates there is a strong relationship between attributes. The value of adjusted variance explained is calculated using the following equation.

$$\omega^2 = \frac{SSB - (k - 1)MSE}{SST + MSE} \quad (1)$$

When the analysis results are suspected to include spurious relationships, we apply analysis of covariance (ANCOVA) to correct bias.

We draw boxplots when strong relationships are observed by adjusted variance explained, because it does not denote whether there is positive relationship or negative relationship, unlike correlation coefficient. In boxplots, the bold line in each box indicates the median value. Small circles indicate outliers, that is, values that are more than 1.5 times larger than the 25%-75% range from the top of the box edge. Stars indicate extreme outliers, whose values are more than 3.0 times larger than this range.

Next, we build the number of staff estimation model and the operations cost estimation model. The operations cost estimation model does not include unit cost of staff and the number of staff as independent variable, because the model is assumed to be used by IT operations service purchasers, and they do not know actual values of them.

To build the models, we use linear regression model and stepwise variable selection. Attributes which show some relationship to unit cost of staff or the number of staff are regarded as major factors of operations cost, and they are used as independent variables for the operations cost estimation model. Also, attributes which show some relationship to the number of staff are used as independent variables for the number of staff estimation model. Note that stepwise variable selection with all attributes did not work well, and consequently low accuracy models were made.

Before building the model, log transformation is applied to numerical attributes, because distributions of numerical attributes are skewed. Also, nominal attributes are transformed into dummy variables (A nominal attribute is transformed into multiple variables whose variables are 0 or 1). Leave-one-out cross-validation is applied when building the models.

As evaluation criteria of the models, we use average and median of *AE* (absolute error), *MRE* (magnitude of relative error) [5], and *MER* (Magnitude of Error Relative to the estimate) [10]. These evaluation criteria are calculated as following equations (where  $X$  is actual value, and  $\hat{X}$  is estimated value):

$$AE = |X - \bar{X}| \quad (2)$$

$$MRE = \frac{|X - \bar{X}|}{X} \quad (3)$$

$$MER = \frac{|X - \bar{X}|}{\bar{X}} \quad (4)$$

*MRE* and *MER* are criteria to evaluate overestimation and underestimation respectively [7]. Using either of them is insufficient because even if the *MRE* of a model is small, the model might overestimate if *MER* is much greater than *MRE*.

## 4 Results

### 4.1 Attributes related to the number of staff

Table 1 shows correlation coefficient of the number of staff and other (numerical) attributes (Table 1 shows attributes whose *p*-value of correlation coefficient is smaller than 10%). Considering tasks of operations staff, it is reasonable that the number of computer terminals has the highest correlation coefficient to the number of staff.

Also, as shown in Table 1, the number of computer terminals has high correlation coefficient to other attributes, except for the number of mainframes. There is a probability that only the number of computer terminals has actual relationship to the number of staff, and other attributes have spurious relationships. So we applied partial correlation to eliminate influence of the number of computer terminals. Table 2 shows partial correlation of attributes included in Table 1. The number of Windows servers, mainframes, and UNIX servers have moderate relationships to the number of staff. So in addition to the number of computer terminals, the number of server machines should be considered when the number of staff is estimated. It is reasonable that server machine operations needs additional staff.

Other observations are described below.

- Although the number of users seems to have moderate relationship to the number of staff in Table 1, it has weak relationship to the number of staff when applying partial correlation (Table 2). Hence, correlation of the number of users and the number of staff shown in Table 1 is regarded as spurious correlation. A large number of users may not increase workload of operations staff directly.

Attributes	# of staff		# of computer terminals	
	Correlation coefficient	<i>p</i> -value	Correlation coefficient	<i>p</i> -value
# of computer terminals	0.69	0%	-	-
# of users	0.63	0%	0.89	0%
# of Windows servers	0.55	0%	0.80	0%
# of mainframes	0.47	2%	0.27	17%
% of financial management staff for IT services	0.32	8%	0.52	0%
# of UNIX servers	0.29	9%	0.74	0%

**Table 1:** Correlation coefficient to the number of staff ( $p$ -value < 10%) and the number of computer terminals

- Percentages of operations staff assignment (That is, the percentage of service request management staff and the percentage of other management staff) do not have relationship to the number of staff. The result suggests that when estimating the number of staff, it is not necessary to consider operations staff assignment, and modification of operation task balance would not need to change the number of staff.
- The percentage of financial management staff for IT service has moderate relationship to the number of computer terminals (Relationship to the number of staff is spurious correlation). It may be because when operations cost is higher (That is, the number of computer terminals is larger), more financial management staff is required.
- The number of program has low correlation coefficient to the number of staff (Correlation coefficients is 0.28). Roughly, the number of program denotes software size. It is natural that the number of program (software size) does not affect the number of the staff, because operations staff do not modify software on the system, and therefore workload of operations staff is not affected by software size (Large size software does not always need many staff).
- Operation time has low correlation coefficient to the number of staff (Correlation coefficients is 0.14). Although it seems plausible that long operation time

needs shift work and it increases the number of staff, operation time does not affect the number of staff. So basically, we do not have to care operation time when estimating the number of staff.

Attributes	Partial correlation coefficient	<i>p</i> -value
# of Windows servers	0.44	1%
# of mainframes	0.44	4%
# of UNIX servers	0.44	1%
% of financial management for IT service	-0.24	18%
# of users	-0.13	43%

**Table 2:** Partial correlation eliminating influence of the number of computer terminals

Attributes	# of staff		# of computer terminals	
	Adjusted variance explained	<i>p</i> -value	Adjusted variance explained	<i>p</i> -value
Network range	35%	0%	0%	44%
ITIL status	28%	0%	7%	10%
SLA status	25%	0%	8%	8%
Using MS SQL Server	7%	5%	13%	0%

**Table 3:** Adjusted variance explained to the number of staff (*p*-value < 10%) and the number of computer terminals

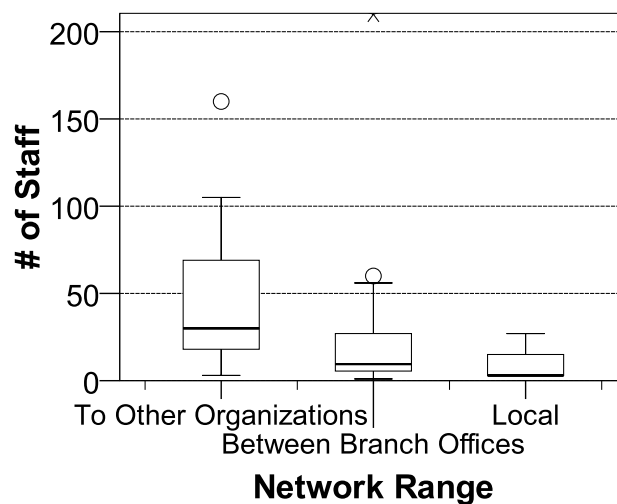
Table 3 shows adjusted variance explained (of nominal attributes) to the number of staff (Table 3 shows attributes whose *p*-value of adjusted variance explained is smaller than 10%) and the number of computer terminals. Although adjusted

variance explained to the number of computer terminals is not large, their  $p$ -value are almost less than 10%, except for network range. There is a probability that only the number of computer terminals has actual relationship to the number of staff, and other attributes have spurious relationships.

We applied ANCOVA to eliminate influence of the number of computer terminals. The result showed that only network range has statistically significant relationship to the number of staff. As shown in Figure 1, when network range is broader, more operations staff is needed. The reason would be that operation tasks for network needs additional staff, separately from operations staff for computer terminals.

ITIL status, SLA status, and using Microsoft SQL Server have some relationships to the number of computer terminals (Relationships to the number of staff are spurious). It may be because when operations cost is higher (That is, the number of computer terminals is larger), ITIL and SLA are needed. Microsoft SQL Server seems to be used when the number of computer terminals is large.

From the results, we conclude that the number of staff has relationships to network range, and the number of computer terminals, Windows servers, mainframes, and UNIX servers.



**Figure 1:** Relationship between the number of staff and network range

#### 4.2 Attributes related to unit cost of staff

Table 4 shows correlation coefficient of unit cost of staff and other (numerical) attributes. (Table 4 shows attributes whose  $p$ -value of correlation coefficient is smaller than 10%). In the analysis, we assumed that allocation of staff of the IT operations service provider is almost same as the purchaser. Although there is a



probability that all configuration management staff was the purchasers' one for example, almost in half cases, 80% of staff was the providers' one, and therefore we regarded the probability as low.

Observations are described below.

- The number of UNIX server and the percentage of configuration management staff have moderate relationships to unit cost of staff, and the results suggest that unit cost of UNIX server operations staff and configuration management staff are higher than other operations staff. The reason of higher price would be that these staff are required certain skill level, and therefore their market value is higher in Japan. Note that the results may be different from countries due to supply and demand of the market.
- The percentage of staff of financial management for IT services has positive correlation to unit cost of staff. The result would not suggest that unit cost of staff of financial management for IT services is high, but suggest that when unit cost of staff is higher, financial management staff is needed.
- The number of computer terminals has low correlation coefficient to unit cost of staff (Correlation coefficients is -0.06). Hence, we do not have to care the number of computer terminals when estimating unit cost of staff.

Attributes	Correlation coefficient	<i>p</i> -value
# of UNIX server	0.53	1%
% of configuration management staff	0.50	2%
% of financial management staff for IT service	0.37	9%

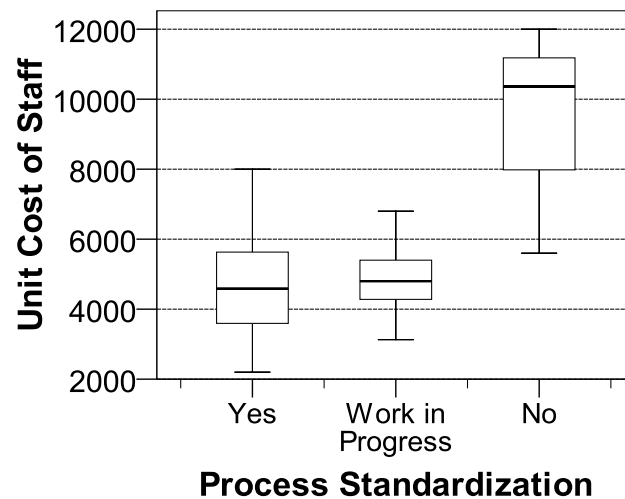
**Table 4:** Correlation coefficient to unit cost of staff (*p*-value < 10%)

Attributes	Adjusted variance explained	<i>p</i> -value
Process standardization	41%	0%
Business sector	38%	1%
Using open source database	21%	1%
System service time	17%	8%

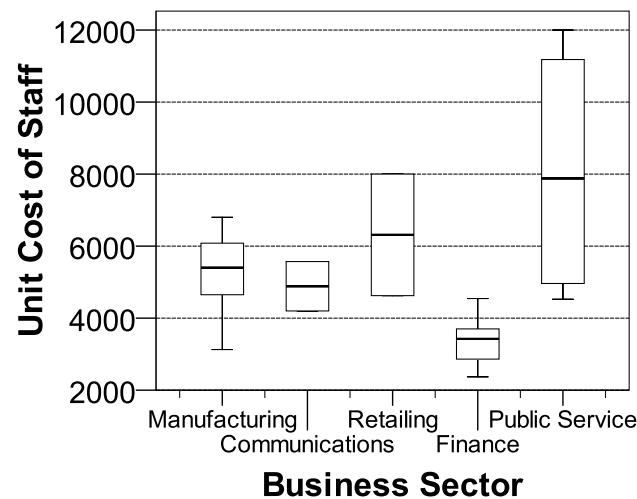
**Table 5:** Adjusted variance explained to unit cost of staff (*p*-value < 10%)

Table 5 shows adjusted variance explained (of nominal attributes) to unit cost of staff (Table 5 shows attributes whose  $p$ -value of adjusted variance explained is smaller than 10%). The relationships between unit cost of staff and other attributes are shown in Figure 2-5. Descriptions of the observations are as follows.

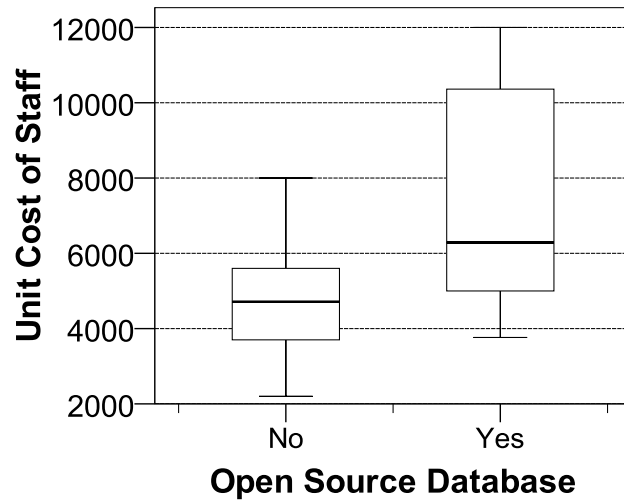
- As shown in Figure 2, when process is not standardized, unit cost of staff is higher. The reason would be that operations staff who have higher skills are needed when process is not standardized, and that makes unit cost of staff higher.
- Unit cost of staff is different by business sector as shown in Figure 3. Intensity of competition of IT operations service may be different from each business sector, and that would make the differences of unit cost of staff.



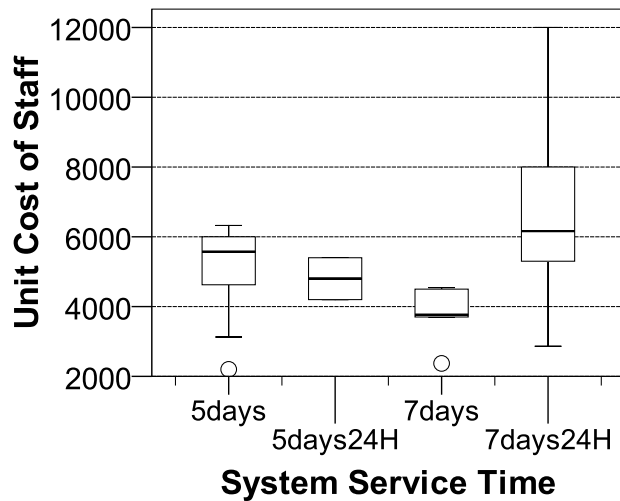
**Figure 2:** Relationship between unit cost of staff and process standardization



**Figure 3:** Relationship between unit cost of staff and business sector



**Figure 4:** Relationship between unit cost of staff and using open source database



**Figure 5:** Relationship between unit cost of staff and system service time

- When using open source database, unit cost of staff is higher as shown in Figure 4. This would be that the staff are required certain skill level, and therefore their market value is higher in Japan.
- As shown in Figure 5, when system service time is 7 days and 24 hours, unit cost of staff is highest (Note the number of cases of 5 days and 24 hours is only two, so the cases are better to omit). The reason would be that 24 hours operation needs shift work, and that makes unit cost of staff higher.

From the results, we conclude that the number of UNIX server, the percentage of configuration management staff, process standardization, business sector, using

open source database, and system service time have relationships to unit cost of staff.

### 4.3 Staff and cost estimation models

We built the staff estimation model using the number of computer terminals, the number of Windows servers, the number of mainframes, the number of UNIX servers, and network range (attributes which have relationships to the number of staff as shown in section 4.1). Accuracy of the model is shown in Table 6. The model has moderate accuracy, and therefore the number of staff can be estimated by system configurations such as the number of computer terminals.

The total cost estimation model was built using the number of computer terminals, the number of Windows servers, the number of mainframes, the number of UNIX servers, network range, the percentage of configuration management staff, process standardization, business sector, using open source database, and system service time (attributes which have relationships to the number of staff as shown in section 4.1 or unit cost of staff as shown in section 4.2). Accuracy of the model is shown in Table 6. Accuracy of the model is very low. This result suggests that it is difficult to estimate total cost by mathematical models using attributes shown in this paper.

From the results, we conclude that IT operations service providers can estimate operations cost by mathematical models only, using estimated number of staff by the models and actual unit cost of staff which they already know. That would help enhancing reliability of cost estimation. On the contrary, it is difficult for IT operations service purchasers to estimate operations cost by mathematical models only. Operations cost estimation by the purchasers needs both mathematical models for number of staff estimation and analogous estimation of unit cost of staff. The analogous estimation is performed based on similarity of major factors of unit cost shown in section 4.2.

Dependent variable	Mean <i>AE</i>	Median <i>AE</i>	Mean <i>MRE</i>	Median <i>MRE</i>	Mean <i>MER</i>	Median <i>MER</i>
Number of staff	18.3	5.4	76%	57%	116%	47%
Total cost	242.4	62.5	221%	65%	172%	80%

**Table 6:** Accuracies of staff estimation and cost estimation

## 5 Related work

While there are some studies treating IT operations, they mainly focus on the process of IT operations. For instance, Brooks [4] showed metrics for evaluation of IT operations process, but the metrics are not effective for IT operations cost estimation. Pollard et al. [12] performed case study of the United States and Australian companies' IT operations, and identified some critical success factors (CSF) for successful ITIL implementations. The research also focused on IT operations process, and did not treat cost estimation.

Mannino et al. [11] analyzed efficiency of data warehouse operations by data envelopment analysis (DEA). Although data warehouse operations is one part of IT operations, their results cannot apply to staff or cost estimation.

To our knowledge, there is no study which treats quantitative IT operations staff (cost) estimation. ITIL explains cost estimation based on cost types such as equipment cost or organization (staff) cost [3]. However, it does not tell how to estimate the number of staff quantitatively. The concept of our research is similar to software development effort (cost) estimation [2, 13, 14, 15, 16], or software maintenance effort estimation researches [1, 9].

## 6 Conclusion

To help estimating IT (Information Technology) operations cost, we analyzed cross-organization dataset of IT operations, focusing on the number of operations staff and unit cost of the staff, because IT operations cost is basically decided by them. Our findings include the followings:

- The number of staff has relationships to network range, and the number of computer terminals, Windows servers, mainframes, and UNIX servers.
- Unit cost of staff has relationships to the number of UNIX Servers, the percentage of configuration management staff, process standardization, business sector, using open source database, and system service time.

The findings are useful to perform analogous estimation of the number of staff and unit cost of staff.

Also, we tried to build the number of staff estimation model and the IT operations cost estimation model. As a result, the number of staff model showed moderate accuracy, but the IT operations cost model showed low accuracy. The result suggests that although IT operations service providers can estimate operations cost by mathematical models only, IT operations service purchasers should combine both mathematical estimation and analogous estimation to speculate IT operations cost.

Our future work is collecting more data and analyzing it to enhance reliability of the analysis results.

## **Acknowledgement**

This work is being conducted as a part of the StagE project, The Development of Next-Generation IT Infra-structure, and Grant-in-aid for Young Scientists (B), 22700034, 2010, supported by the Ministry of Education, Culture, Sports, Science and Technology.

## **References**

1. Y. Ahn, J. Suh, S. Kim, and H. Kim: The software maintenance project effort estimation model based on function points, *Journal of Software Maintenance: Research and Practice*, vol.15, no.2, pp.71-85, 2003.
2. B. Boehm: *Software Engineering Economics*, Prentice Hall, 1981.
3. J. Bon, ed.: *Foundations of IT Service Management: based on ITIL*, Van Haren Publishing, 2005.
4. P. Brooks: *Metrics for IT Service Management*, Van Haren Publishing, 2006.
5. S. Conte, H. Dunsmore, and V. Shen: *Software Engineering, Metrics and Models*, Benjamin/Cummings, 1986.
6. Economic Research Association.  
On-line < <http://www.zai-keicho.or.jp/about/english.php>>
7. T. Foss, E. Stensrud, B. Kitchenham, and I. Myrtveit: A Simulation Study of the Model Evaluation Criterion MMRE, *IEEE Transactions on Software Engineering*, vol.29, no.11, pp.985-995, 2003.
8. ISO/IEC 12207:1995: *Information technology - Software life cycle processes*, International Organization for Standardization, 1995.
9. M. Jørgensen: Experience With the Accuracy of Software Maintenance Task Effort Prediction Models, *IEEE Transactions on Software Engineering*, vol.21, no.8, pp.674-681, 1995.
10. B. Kitchenham, S. MacDonell, L. Pickard, and M. Shepperd: What Accuracy Statistics Really Measure, *In Proc. of IEE Software*, vol.148, no.3, pp.81-85, 2001.
11. M. Mannino, S. N. Hong, and I. J. Choi: Efficiency evaluation of data warehouse operations, *Decision Support Systems*, Vol.44, no.4, pp.883-898, 2008.
12. C. Pollard, and A. Cater-Steel: Justifications, Strategies, and Critical Success Factors in Successful ITIL Implementations in U.S. and Australian Companies: An Exploratory Study, *Information System Management*, vol.26, no.2, pp.164-175, 2009.
13. L. Putnam: A general empirical solution to the macro software sizing and estimating problem, *IEEE Trans. on Software Eng.*, vol.4, no.4, pp.345-361, 1978.

- 14.R. Selby, and A. Porter: Learning from examples: generation and evaluation of decision trees for software resource analysis, IEEE Trans. on Software Eng., vol.14, no.12, pp743-757, 1988.
- 15.M. Shepperd, and C. Schofield: Estimating software project effort using analogies, IEEE Trans. on Software Eng., vol.23, no.12, pp.736-743, 1997.
- 16.K. Srinivasan, and D. Fisher: Machine Learning Approaches to Estimating Software Development Effort, IEEE Trans. on Software Eng., Vol.21, No.2, pp.126-137, 1995.
- 17.B. Winer, D. Brown, and K. Michels: Statistical Principles in Experimental Design (3rd edn.), McGraw-Hill, 1991.

