

# 統計的方法을 利用한 SOFTWARE 品質評價에 관한 研究

## A Study of SOFTWARE Quality Evaluation by the Use of Statistical Methods.

金正子\*  
趙盛健\*\*

### ABSTRACT

The quality of a final SOFTWARE PACKAGE depends on many complicated factors in the software development process. This paper describes statistical methods for establishing relationships between final quality and development process factors. The final software quality is represented by the number of errors through the system test phase. The data presented here were gathered during the course of a real 18-month development project. Regression theory is used for data analysis. Some of the interesting results include the observation that specification changes during the development process have an adverse effect on final software quality.

### I. 序 論

#### 1. 研究目的

Software의 品質特性에는 信賴性 操作性, 保守性, 擴張性, 互換性 등의 많은 要因이 있다.

이러한 Software의 品質特性은 Software의 開發段階에서 고려해야만 高品質의 Software를 生産할 수 있을 것이다.

그러나 Software는 製品自体가 人間에 의하여 만들어지므로 最終적으로 製品에 영향을 주는 要因은 역시 Software 開發에 종사하는 要員의 熟練度, 開發期間과 開發量, 開發 Software

의 難易度 등에 의하여 많은 영향을 받게 된다.

結果적으로, 하나하나의 Software는 이런 多種多樣的 要因들이 複雜하게 作用되어서 하나의 製品이 만들어지며, Software의 品質도 決定된다. 本 研究에서는 Software의 開發段階에서 發生될 수 있는 各種要因과 最終 製品과의 사이에 어떤 相關關係가 存在하는지를 Software開發에서 發生하는 要因들을 變數로 하여, 統計的方法을 利用해서 解明해 보고, Software 開發時 適用 가능한 方法을 제시하고자 하는데 目的이 있다.

\*東亞大學校 産業工學科

\*\*慶南專門大學 電子計算科

## 2. 研究方法

Software의品質評價는生産되는製品の評價와는 다르다고 볼 수 있다.

따라서, Software의品質을評價하기 위해서는開發者의特性, Computer動作上의特性等の質的 data와 Error數, 命令文의使用數等の量的 data가 모두變數로使用되어야 한다.

그러나, 本研究에서는量的인 Data만을 취급하여 Software의品質을評價해 보고자 한다.

變數로使用하는實測值 Data는開發期間이約1年6個月程度되는學校의業務處理 System에서使用하고 있는 Software의 RUN에서 얻는實際의 Data를利用해서分析한다.

分析에利用하는統計的方法是回歸分析의方法을利用하며 Software는163個의 Program中에서30個를임의로추출하여조사한 Data를變數로使用한다.

業務處理時發生된 Error와選擇變數사이의關係를 나타내는  $r$ 을 구하여變數關係를 규명하고, 分散分析에 의하여 검정을 하고, 기존 System의改善方向을 제시하는 것으로 한다.

## II. 品質評價 Model

Software의信賴性이란綜合 RUN에서各 Software에 의해發生되는 Error發生件數에 의해決定된다.

따라서, 量的으로測定 가능한 Data로變數를構成하여品質評價 Model을作成한다.

製品에 영향을 주는要因들은 Software의 Module構成에 관한要因과 Computer의動作特性에 관한要因 및 System分析和設計에 의한要因 등으로大別할 수 있으며, 本研究에서는 Module構成에 관한要因을變數로하여分析하고자 한다.

따라서, 信賴性에 관한 Data는 Error發生件數로 하고, 變數는 Module構成에 관한 Data로해서 조사한다.

### 1. 信賴性에 관한 Data

System 전체를 RUN하면서選擇된 Software에서發生되는 Error件數를 조사한다.

System을構成하는全體 Program은 COBOL Language로作成된 Program들이며, Error件數는綜合 RUN에서發生하는 Error件數를 조사하여 Y로 한다.

### 2. Model構成에 관한 Data

Model構成에 영향을 주는要因은 많이 있지만品質에 영향을 주는要因들중에서 5個를變數로選擇하여使用한다.

#### 1) Module의綜合度 - $X_1$

Program을構成하는 Module相互間的結合은同種의 Language나異種의 Language나結合은可能하며, 이結合은 CALL이라는命令에 의하여 이루어지므로, 各 Program內的 CALL命令文의數를獨立變數  $X_1$ 으로 한다.

本研究에서의評價對象이 되는 System은 COBOL과 COBOL의結合 및 COBOL과 Assembly의結合만이存在하며 Subprogram內的 CALL命令文의數도 하나의 CALL命令文으로 한다.

#### 2) Program의 크기 - $X_2$

Program 그自体는獨立性を 갖고 있으며, 때로는 Computer System의 제약에 따라야 하는 문제점도 있다.

따라서, Program의 크기도品質에 영향을 주는要因의 하나가 된다.

여기서 Program의 크기는 번역된 상태인 Object에서 Module의 크기를 System의 크기단위인 Page단위로 하며, 獨立變數  $X_2$ 로 한다.

#### 3) IF文의使用數 - $X_3$

Program內에서의論理는 가능한한 단行하게 하는 것이 좋다.

따라서, 論理의 간결화를 위하여 IF文의使用를可能하면억제하는 것이 Program의理解나 읽기를 용이하게 하는 것이므로 IF文의使用數를獨立變數로 하며  $X_3$ 으로 한다.

#### 4) PERFORM命令文의使用數 - $X_4$

Program의 Error發生時 Program Error修正을 위한 추적作業을 지연시키는要因이 되며, 특히作業中에 Logic Error가發生되었을 때, Error 추적이 어려운命令文으로서實際 Soft-

ware開發時 많이 사용되는 命令文의 하나다.

따라서 Perform 命令文의 數를 獨立變數  $X_4$ 로 한다.

5) Computer 命令文의 使用數 —  $X_5$

Computer 命令文은 연산작업을 총칭하는 연산 分의 代表的인 命令文으로서 使用은 편리하지만 作業時間의 지연과 특히 연산문의 혼합사용시 Error를 發生시키는 要因이 되는 命令文이다.

따라서, Computer 命令文의 數를 獨立變數  $X_5$ 로 한다.

이밖에도 많은 要因이 存在하지만 變數選擇上의 問題點 等を 고려하여 위의 5個를 獨立變數로 使用한다.

이것을 整理하여 보면 Table 1과 같다.

要 因	記號	Data의 說明 設
信 賴 性	Y	綜合RUN時 發生된 Error件數
Module의 構成要素	$X_1$	Program에 使用된 CALL文의 數
	$X_2$	Program Module의 Page數
	$X_3$	Program內的 IF文의 數
	$X_4$	Program內的 PERFORM文의 數
	$X_5$	Program內的 COMPUTE文의 數

Table 1. Module에 使用하는 變數

### III. 分析方法

開發된 Program에서 추출한 實測值 Data를 利用하여 回歸分析의 方法으로 分析한다.

#### 1. 變數 Data의 表現

綜合RUN에 의해서 얻은 Error件數와 開發된 Program에서 選擇된 變數 Data를 조사한 結果는 Table 2.와 같다.

Data의 表現은, 分析의 對象이 Software를 利用하여 System을 評價하기 위한 것이므로 信賴度의 對象이 되는 Error發生件數를 Y로 나타냈다.

여기서, 綜合RUN時에 使用한 Data는 1,912件의 實際 Data를 使用했다.

No.	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	Y
1	2	1	1	1	4	3
2	3	3	2	3	6	4
3	3	5	2	2	8	2
4	4	4	5	5	5	9
5	5	6	4	4	7	8
6	5	4	3	6	4	3
7	6	7	2	7	6	1
8	4	5	1	9	5	7
9	7	8	3	11	3	5
10	9	2	4	15	4	3
11	10	10	2	17	6	9
:	:	:	:	:	:	:

Table 2. 實測值 DATA

### 2. Model의 構成

Table 2.에 定義된 Data群을 利用할 수 있는 品質評價의 Model을 構成하면 다음과 같다.

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + a_4 X_4 + a_5 X_5 + \epsilon_i$$

上記의 Model을 利用하여 實際 Data의 適用은 Computer를 利用해서 解決한다.

## IV. 結果分析

Computer處理의 結果를 利用하여  $r$ 값을 보고 검정을 하여 回歸 Model의 精度를 分析한다.

### 1. 處理結果

實測值 Data를 利用하여 Computer處理한 結果는 Fig. 1과 같다.

Model의 妥當性을 判定하기 위하여 回歸分析 相關係數와 分散分析에 의하여 고찰한다.

回歸 Model에서 구한 相關係數는 變數  $X_1, X_2, X_3, X_4, X_5$ 를 選擇했을 때의  $r$  값이 0.85358로서 Model의 精度가 높은 것으로 判明되었다.

따라서 選擇變數는 Error發生의 要因이 되므로 使用上 問題點이 있다고 볼 수 있다.

그리고, 分散比  $F$ 는 12.8878로서 이것은 自由度 (5, 24 : 0.05) = 3.90 < 12.8878이 되므로 有意性이 있다.

이 有意性은 위험을 1.0%에서 有意하다고 할 수 있다.

MULTIPLE REGRESSION...SOFTWARE

VARIABLE NO.	MEAN	STANDARD DEVIATION	STD. ERROR OF REG. CDEF	COMPUTED T VALUE
1	20.79999	15.20301	0.13076	7.11173
2	14.53333	9.05816	0.07526	2.21015
3	6.56667	4.46197	0.16965	3.25343
4	22.26666	13.74129	0.10747	7.15258
5	5.26667	1.61743	0.10465	25.58456
DEPENDENT				
6	4.76667	2.87298		
INTERCEPT		->1.83603		
MULTIPLE CORRELATION		0.85358		
STD. ERROR OF ESTIMATE		1.64526		

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREE OF FREEDOM	MEAN SQUARES	F VALUE
ATTRIBUTABLE TO REGRESSION	5	34.88025	12.88578
DEVIATION FROM REGRESSION	24	2.70688	
TOTAL	29		

Fig. 1 Computer 處理結果

2. 改善方案

Software의 品質을 나쁘게 하는 要因이 되는 것은 물론 分析에서 使用한 選擇된 量的인 變數만은 아니다.

그러나, 本 研究에서 選擇한 變數들이 Software의 品質을 나쁘게 하는 要因이 되는 것은 事實이므로 Software의 開發 및 作業中에 發生하는 Error를 修正하는 경우에는 다음 事項을 고려한다면 Software의 品質을 向上시킬 수 있을 것으로 期待할 수 있다.

1) CALL 命令文의 使用을 억제한다.

꼭 必要한 CALL 命令이 아니면 Main Program에 내장시킨다.

2) Program의 크기를 줄인다.

Program의 基本 크기를 設定하여 그 크기를 초과하면 處理分別로 Program을 나누어서 處理 step을 늘린다.

3) 論理命令文의 使用을 改善한다.

IF 命令文의 自体가 論理를 構成시키면서 복잡하게 될 수 있으므로 可能하다면 간결하게 論理를 構成한다.

4) PERFORM 命令文의 使用을 억제한다.

꼭 使用해야 하는 Perform 文이 아니면 다른 命令文으로 대치해서 使用하므로 Logic Error의 修正을 용이하게 한다.

5) Compute 命令文의 使用을 억제한다.

Compute 命令文은 使用이 편리하다는 이유로 많이 使用하게 되면 Error發生의 原因이 되므로 可能하면 自己 연산문으로 환원하여 使用한다.

여기서는 Software의 Model構成에 관한 要因中에서 變數를 選擇해서 使用했으므로 結果的으로는 Software의 部分的인 評價밖엔 할 수가 없는 것이다.

그러나, 本 研究에서 알 수 있는 것은 Software의 開發時는 경험에 의존하는 것보다는 正確한 System 分析과 設計가 必要하다는 것을 알 수 있다.

V. 結 論

Software의 品質을 評價해 본 결과, 回歸分析에 의한  $r=0.85358$  이고, 分散 分析에서 분

散比  $F$  가 12.88578 로서 有意性이 있다는 結論  
이므로 選擇된 變數에 의해서도 Data Error의  
發生은 당연하다고 본다.

따라서, Software의 信賴性을 높이기 위해서  
는 Software의 開發時點에서 全体 System을 正  
確하게 分析하고 設計해야만 하며, 命令文 하나  
의 使用에도 신중을 기해야 한다고 본다.

결국, Software는 開發이 끝나면 Error의 發  
生없이 作業을 修行해야만 信賴性이 높은 것이

되는 것이다.

이렇게 되기 위해서는 계속해서 Software의  
品質評價 Model을 研究해야 하며, 이렇게 될때  
高品質의 Software를 낮은 開發費로 開發하고,  
管理해 나갈 수 있을 것으로 생각된다.

앞으로는 變數를 量的 Data만 處理하지 말고  
質的 Data도 함께 處理하므로써 보다 高信賴度  
의 一般的인 Model을 開發할 必要가 있을 것으로  
判斷된다.

### 參 考 文 獻

1. G. J. Myers : 高信賴性 Software複合設計,  
近代科學社, 1976
2. 小林 : 數量化理論入門, 日科技連, 1981
3. 河口 : 多變量解析入門 I, 森北出版, 1973
4. P. G. Hoel(淺井訳) : 入門數理統計學, 培風  
館, 1978
5. 竹内啓 : 數理統計學, 東洋經濟新報社, 1968
6. 刈屋武昭 : 回歸分析의 理論, 岩波書店, 1979
7. 佐和隆光 : 回歸分析, 朝倉書店, 1980
8. Plackett, R.L.: *Regression Analysis*,  
Clarendon Press, Oxford, 1960.
9. Draper, N. and H. Smith: *Applied Regres-  
sion Analysis*, John Wiley & Sons. New  
York, 1966.