

통계처리용 소프트웨어 패키지의 품질 비교에 관한 연구

이상석

강남대학교 경영학부

윤민석

여수대학교 멀티미디어학부

A Study on The Quality Comparison of Software Packages for Statistical Analyses

Sang-Suk Lee

School of Business Administration, Kangnam University

Min-Suk Yoon

School of Multimedia, Yosu National University

Abstract

The multiplicity of software packages brings about a choice problem to the end-user and a multi-fold intensity to the competing developer. This study intends to provide guidelines solving the problems for both the end-user and the developer of packages for automatic statistical analyses. For the purpose, we adopt an appropriate evaluation model in terms of user-perceived quality through the applicable criteria and an evaluation method. The result of empirical study is described, which shows that the weight of each criterion is dynamic along usage levels. Briefly, functionality, usability, and reliability are key factors to contribute to the perceived quality. The empirical study also illustrates a comparative analysis on the dominant products, SAS and SPSS, in Korea.

1. 서론

통계학의 적용 범위가 사회과학 및 자연과학을 비롯하여 인문과학에까지 확대됨에 따라 통계처리의 자동화를 위한 소프트웨어가 질적·양적으로 괄목할 만큼 확대되었다. 이러한 양상은 80년대 이후 PC(Personal Computer)의 성능 향상과 보급 확대 그리고 소프트웨어 패키지 개발 분야의 급속한 성장에 따른 것이다. 패키지는 시장생산(make-to-market)형 소프트웨어의 대표적인 형태로 그 종류의 증가는 최종 사용자에게 제품 선택이라는 의사결정 문제를 유발시켰고, 개발자에게는 자사제품의 경쟁 요인의 파악이라는 과제를 남겼다[3, 14].

지금까지 알려진 통계 처리용 소프트웨어 패키지(이하 통계패키지라 칭함)로는 SAS(Statistical Analysis System)와 SPSS(Statistical Product & Service Solutions)가 널리 알려져 있으며 이외에도 MINITAB이나 SYSTAT 등과 같은 패키지들도 상당한 기능을 가지고 있다. 특정 목적을 위주로 한 통계패키지로는 다차원척도를 위한 KYST 및 시계열 분석을 위한 TSP 등이 있으며 최근 계산 전용 소프트웨어인 스프레드쉬트에서도 일정 수준의 통계 처리를 지원하고 있다[2]. 이들 외에도 Statistica, Sigma Plot 그리고 RATS 등이 사용되고 있다. 한편, PC 성능의 현저한 향상과 윈도(Windows) 운영시스템(Operating System)으로의 전환은 소프트웨어의 기능확대 및 사용자 인터페이스(User Interface) 향상의 계기가 되었으며, 이와 같은 현상은 통계패키지 영역에서도 예외가 아니다. 기능확대의 일 예로 현재의 SAS와 SPSS는 공히 각종 통계 처리와 관련된 모듈은 물론, Operations Research 및 품질관리 기법과 데이터베이스 관리시스템과의 연계를 지원하는 모듈 등이 추가되어 있다 [9, 10]. 사용자 인터페이스는 그래픽 기반으로 전환되어 다양한 형태의 표현 및 보고서 양식 등을 지원하고 있다. 더 나아가 각종 패키지는 다양한 형태의 플랫폼(platform)을 지원하도록 코드를 전환하고 있다. 예로 SAS는 초기에 메인프레임용의 PL/1과 IBM 어셈블리어의 조합으로 개발되었으나 이후 마이크로컴퓨터 및 PC와 해당 운영체제도 지원하기 위하여 적응성이 좋은 C 언어로 모든 코드를 전환한 바 있다[5].

통계패키지의 기능 확대와 인터페이스 향상은 필수 불가결한 프로그램의 볼륨 증가와 신뢰성 문제를 초래하여 제품평가를 더욱 어렵게 하였다. 이로 인하여 최종사용자들은 다양한 제품들 가운데서 제품선택에 더욱 어려움을 겪고 있는 실정이다. 이에 본 연구는 최종사용자에게 통계패키지 제품선택에 지침이 될 수 있는 정보의 제공과 함께, 제품개발자 또는 판매대행자(Sales Agent)에게는 자사 제품의 전략수립에 활용될 수 있는 통계패키지 품질 평가모형을 제시하는데 그 목적이 있다.

이를 달성하기 위한 본 연구는 절차는 다음과 같은 체계로 구성된다.

첫째, 평가모형의 설계는 통계패키지를 위한 평가기준을 정하고 이에 적절한 평가기법을 결합하여 완성한다. 최근 소프트웨어와 종합적 품질경영(Total Quality Management)의 중요성을 동시에 반영하고 있는 소프트웨어 품질 지향적 평가기준을 고찰하고, 최종사용자의 판단에 의한 세심한 품질평가를 수행하기 위한 평가기법을

탐색한다.

둘째, 제시된 모형의 적용 가능성을 실사를 통하여 보인다. 주관적 판단의 계량화를 통하여 각 품질 평가기준에 대한 구체적인 정보로서 중요도를 산출하고 사용 수준에 따른 중요도의 변화를 비교·분석한다. 아울러, 대표적 통계패키지인 SAS와 SPSS를 대상으로 본 연구의 평가모형을 적용하고, 제시한 평가모형이 제품 경쟁하에서 제품 간 장점과 약점을 파악할 수 있는 현실적인 모형임을 예시한다.

2. 선행연구의 검토 및 평가기준의 설계

2.1 소프트웨어 품질의 선행연구

소프트웨어 평가기준은 Fritz & Carter(1994)가 1980년대 이후의 연구에 대해서 탐색적 연구를 시도하였으며, <표 1>에서 보는 바와 같이 그 이후에 추가적인 연구가 진행되었다 [23]. 주요 연구들을 살펴보면, Borwnstein & Lerner는 우선적으로 평가계획 지침이 사용자 요구사항을 평가하고, 필요활동을 분류하고, 요구된 자원을 추정하기 위해서 검토되어야 한다고 제안하였다. Kuan은 PC에 대한 소프트웨어 제품의 평가기준으로 사용의 용이성, 이용자 친밀성, 판매업자 지원, 하드웨어와의 적합성 등이 포함되어야 한다고 말하고 있다. Adeli & Wilcoski는 마이크로컴퓨터를 위한 도스용 소프트웨어의 구조적 설계를 위한 평가기준을 연구하였다. Zahedi는 DBMS 선택기준으로 기능적 요인, 물리적 요인, 비용, 이익 등의 요인으로 분류하여 제안하였으며, Davis는 정보기술과 소프트웨어를 위한 사용의 용이, 유용성 등의 특성을 제안하였다. Subramanian & Gershon은 용량, 조직과 융화, 시스템 품질, 사용자 인터페이스, 가격 등의 요인을 제안하였다. Kekre 등(1995)은 소비자 만족을 위한 7개 요인을, Jung & Yoon(1996)은 ISO 9126 품질모델에 근거한 기준을 제안하였다.

이상의 선행연구에서 보이듯이 소프트웨어 평가는 평가 목적에 따라 상이한 기준이 적용되며 본 연구에서는 품질 지향적 목적에 따른 평가 기준을 적용하고자 한다. 소프트웨어 품질은 “명시적이거나 묵시적인 필요를 만족시키는 소프트웨어 제품의 능력과 관련된 소프트웨어 제품의 특징(features) 및 특성(characteristics)의 전체”로 정의된다[11, 12]. 지금까지 소프트웨어 품질의 이해 및 측정은 소프트웨어 ‘품질모델’을 통하여 논의되어 왔다[4, 6, 7, 12, 16]. 소프트웨어 품질모델은 “품질 요구사항 및 품질 평가에 기초를 제공하는 품질특성 및 그들간의 관계들의 집합”이며, 이 때 품질특성은 “소프트웨어 제품의 품질이 기술되고 평가될 때 적용되는 소프트웨어 제품 속성들의 집합”으로 각 품질특성은 여러 개의 하위 특성으로 더 세분화 될 수 있다 [12]. 제안된 여러 품질모델 중 ISO 9126 모델이 소프트웨어가 실제 작동되는 환경에서 고려되는 사용자 관점을 잘 반영한다. Boehm(1978)과 McCall(1977)의 모델에는 소프트웨어 사용상의 품질 요인과 소프트웨어 내부 속성(internal properties)이 혼재하고 있다. 반면, ISO 9126은 양자를 구분하여 내부 속성들은 소프트웨어 제품 개발과정에 적용되

는 품질특성으로 규정한다[15]. Dromey(1995)의 모델은 개발자 관점으로부터 출발하는 품질모델이므로 사용자 관점을 포괄적으로 반영하기에는 한계가 있는 것으로 판단된다.

< 표 1 > 소프트웨어 평가기준에 관한 연구

연구자	내 용
Brounstein & Lemer(1982)	기능적 요구사항, 설계개념, 비용, 판매인, 지원
Sanders, Ghandforoush & Austin(1983)	주요요인, 주관적 요인, 객관적 요인
Tally(1983)	일반적 특성: 문서, 투입/산출, 운영 전반적 특성: 워드프로세스, 스프레드시트
Zahedi(1985)	기능적 관점, 물리적 관점, 비용관점, 이익관점
Frankel(1986)	사용자인터페이스, 기술적 특성, 문서, 훈련
NASA & Kuan(1986)	사용의 용이, 사용자 친밀감, 판매자 지원 및 갱신, 문서, 기계적합성
Visker & Bree(1987)	용이성, 시간유용성, 재정적 측면, 조직적 적응
Lucas, Walton & Ginzberg(1988)	소프트웨어 품질, 이용자 요구, 조직적 특성
Davis(1989)	유용성, 사용의 용이성
USACERL, Meier & Williamson(1989)	시스템 요구사항, 하드웨어 요구사항, 기능적 엔지니어링요구, 판매자 지원, 비용
Subramanian & Gershon(1991)	용량, 조직과 융화, 시스템품질, 사용자인터페이스, 가격
Adeli & Wilcoski(1993)	사용자인터페이스, 그래픽호환, 자료입력장치, 이동성, 재설계관리, 프로그래밍환경, 특정 고려사항
Kekre, Krishnan & Srinivasan(1995)	신뢰성, 용량, 유용성, 설치용이성, 유지보수, 성능, 문서
Jung & Yoon(1996)	기능성, 신뢰성, 사용성, 효율성, 유지보수성, 이식성

자료: Yoon M.S.(1997), *S/W Quality Evaluation Model Using the AHP*, Thesis for The Degree of Doctor, Korea University, p. 17.

2.2. 통계패키지 품질 평가기준의 설계

본 연구에서는 ISO 9126 품질모델을 기반으로 하고 이를 통계패키지에 적합하도록 수정하여 통계패키지 품질 평가기준으로 이용하고자 한다. ISO 9126 품질모델을 기반으로 최종사용자 지각에 의한 통계패키지 품질 평가를 수행하기 위해서는 다음 두 가지 사항이 선행되어야 한다. 첫째, ISO 9126 품질모델은 범용 목적이므로 본 연구의 대상인 통계패키지를 평가하는데 적용되는 품질특성이 확인되어야 한다. 둘째, 각 품질특성은 다수의 소프트웨어 공학 용어로 기술되어 있으므로 이를 최종사용자가 이해하기 쉽게 표현하여야 한다.

첫 번째 요건을 위하여 기존의 연구결과를 활용하고자 한다. ISO 9126 품질모델은 소프트웨어 사용상의 품질을 반영하는 6개의 품질 주특성과 각 특성의 하위에 2~5개

씩 총 21개의 품질 부특성으로 구성되어 있다. 이에 관하여 Yoon(1997)은 소프트웨어 패키지에 적용될 품질모델을 제안한 바 있다. 그의 연구절차를 간략히 언급하면, 우선 소프트웨어 패키지의 구성영역을 프로그램, 매뉴얼, 시스템 요구사항, 그리고 고객 서비스로 구분하였고 프로그램은 기능을 중심으로 인터페이스, 기본기능, 핵심기능, 그리고 지원기능으로 세분하였다. 그리고 구분된 각 영역과 21개의 소프트웨어 품질 부특성간 상호 관련성을 규명하고 재구성함으로써 소프트웨어 패키지를 위한 품질 주특성 및 부특성을 결정하였다. 본 연구의 대상이 소프트웨어 패키지 범주에 속하는 통계패키지이므로 위 연구결과의 품질모델 체계를 수용하여 평가의 기준으로 확정하였다.

한편, ISO 9126의 각 품질특성에 대한 정의는 내용의 모호성을 피하기 위하여 소프트웨어 공학용어로 기술되어 그 내용을 이해하는데 소프트웨어 공학에 관한 상당한 수준의 지식이 요구된다. 이러한 사실은 본 연구의 최종사용자에 의한 평가 방법에 배치된다. 이를 해결하기 위하여 본 연구는 ISO 9126에서 예시하고 있는 각 품질특성과 관련된 Metrics & Indicators를 충분히 검토하여 최종 사용자가 이해할 수 있도록 최대한 통계패키지와 관련된 용어로 전환하였다.

<표 2>에 제시한 품질모델에 관하여 간략히 언급하면, 품질 주특성의 기능성(Functionality)은 소프트웨어 존재 목적에 부합되는 기능의 구비 정도와 기능의 지원 수준을 의미한다. 신뢰성(Reliability)은 명시된 기간에 명시된 조건에서 소프트웨어의 성능수준을 유지하는 능력이다. 소프트웨어는 마모나 고령화가 일어나지 않으므로 신뢰성의 한계는 요구사항 파악, 설계, 또는 구현 단계상의 결함에 기인하게 된다. 사용성(Usability)은 소프트웨어 사용을 위해 요구되는 노력의 정도로 정의되며, 효율성은 규정된 조건에서 소프트웨어 성능 수준과 사용된 자원의 양 사이에 관계된 속성들의 집합이다. ISO 9126 품질모델에서는 유지보수성(Maintainability)과 이식성(Portability)을 따로 정의하고 있으나 이는 주로 주문형 소프트웨어에 적용된다. 따라서 본 연구의 모형에서는 패키지에 적용 가능한 유지보수성 및 이식성의 하위 특성만을 고려하여 설치·변경성으로 통합하여 정의하였다. 아울러 5개의 각 품질 주특성은 관련된 15개의 하위 품질 부특성으로 세분하고 각각을 정의하였다.

3. 통계패키지 품질 평가모형의 설계

3.1. 평가기법의 선택

본 연구에서 제시될 평가모형의 평가기법은 다음 세 가지 사항을 잘 반영할 수 있어야 한다. 첫째, 앞서 제시된 통계패키지 품질 평가기준은 계층적 분화에 의한 다기준 모형이므로 평가기법은 각 기준에 대한 대안들의 평가 내용을 잘 반영하여 종합할 수 있어야 한다. 둘째, 본 연구에서 제시된 평가 기준은 ISO 9126의 표준에 근거하고 있다. 따라서 제시될 평가기법의 가정 또는 공리(Axioms)는 ISO 9126 제정 과정에 수립되었던 원칙에 부합되어야 한다. 셋째, 본 연구의 목적이 통계패키지 사용자의 지

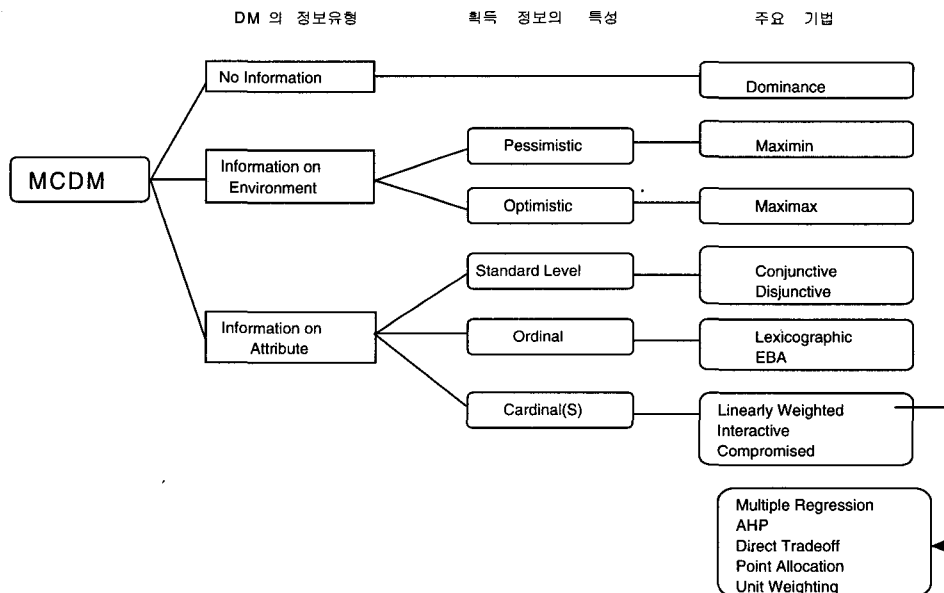
각에 의한 품질 평가모형의 제시이므로 평가자료는 사용자의 주관적 판단에 의존할 수밖에 없다. 따라서 평가기법은 주관적 판단을 정량화 하는데 이론적이며 과학적인 근거를 확보한 기법이어야 한다.

< 표 2 > 통계패키지 품질특성 및 내용

		1 단계 주특성	2 단계 부특성	내 용
통 계	기능성	적절성 (suitability)		<ul style="list-style-type: none"> ○ 통계패키지의 목적상 필요한 기능의 구비 정도 ○ 사용자가 원하는 기능의 제공 정도 ○ 제공되는 기능 수준의 적당성
		정밀성 (accuracy)		<ul style="list-style-type: none"> ○ 수행된 작업 결과의 정확성·정밀성 정도 ○ 매뉴얼의 설명과 프로그램 동작의 일치 정도
		상호 운영성 (interoperability)		<ul style="list-style-type: none"> ○ 다른 종류의 통계패키지 파일의 입·출력 능력 ○ 다른 종류의 소프트웨어에서 작성된 파일(워드프로세서 데이터베이스, 스프레드시트 등)의 입·출력 능력
		보안성 (security)		<ul style="list-style-type: none"> ○ 프로그램과 데이터에 허가없이 접근하는 것을 방지하는 능력 ○ 암호화 기능의 존재와 수준
패 키 지	신뢰성	성숙성 (maturity)		<ul style="list-style-type: none"> ○ 프로그램의 결함에 따른 고장/장해 빈도와 심각성 ○ 프로그램의 결함에 영향을 받는 작업 결과물의 오류 정도
		고장 허용 (fault tolerance)		<ul style="list-style-type: none"> ○ 오조작/오입력의 경우에도 프로그램과 데이터를 보존하는 능력 ○ 오조작/오입력의 검출 능력 또는 안내 메시지 제공 수준
		회복성 (recoverability)		<ul style="list-style-type: none"> ○ 고장/장해의 경우 원래 성능 수준으로 회복하는 능력 및 그 수준 ○ 고장/장해로 인해 직접적으로 영향을 받는 데이터를 본 상태로 되돌려 놓는 능력과 그 수준 ○ optional install, undo, shut-down 경우 자동 back-up 등
품 질	사용성	이해성 (understandability)		<ul style="list-style-type: none"> ○ 각 기능과 제반 관련사항의 개념 및 적용과 관련된 명확성 ○ On-line tutorial과 같은 데모 기능의 명확성
		학습성(learnability)		<ul style="list-style-type: none"> ○ 쉽게 배울 수 있는 가?
		운용성 (operability)		<ul style="list-style-type: none"> ○ 프로그램 운용이나 운용제어가 용이한가? ○ 메뉴, On-line help, 단축키 등 주로 user interface와 관련
효 율 성	시간행동 (time behavior)		<ul style="list-style-type: none"> ○ 작업 처리시간 ○ 프린터 출력시간, 파일 입·출력 시간 ○ 프로그램 loading/unloading 시간 	
	자원행동 (resource behavior)		<ul style="list-style-type: none"> ○ CPU 또는 MPU 요구사항 ○ 메모리 요구사항 및 프로그램의 크기(Hard disk 점유량) 	
설 치 · 변 경 성	변경성 (changeability)	변경성 (changeability)		<ul style="list-style-type: none"> ○ H/W 및 OS 환경변화에 대한 프로그램 환경설정 변화의 용이성 ○ 통신서비스 또는 우편 서비스를 이용하여 손상된/수정된 프로그램 파일 획득의 편리성
		적응성(adaptability)		<ul style="list-style-type: none"> ○ 적용 가능한 H/W 및 OS 종류
		설치성(installability)		<ul style="list-style-type: none"> ○ 프로그램을 설치하기에 필요한 노력의 정도

지금까지 많은 다기준 의사결정(Multi-Criteria Decision Making: MCDM) 기법과 관련된 연구 중 Yoon & Hwang(1995), Schoemaker & Waid(1982), 그리고 Saaty(1980)의 연구를 본 연구에서 재구성하여 <그림 1>에 제시하였다. 구분의 절차는 의사결정(Decision Making: DM)에 주어지는 정보의 유형, 획득되는 정보의 특성, 그리고 방법론 분류로 하였다. 이들 중 정보의 특성이 기수적일 때는 최적 대안의 선택이 가능하며 그 외의 경우는 휴리스틱(Heuristic)에 의한 대안 선택이 제시된다. 최적 대안의 선택 방법 중 비교적 쉬우며 일반적인 형태는 기준의 중요도에 따른 대안의 우선순위를 평가하고 가중 합을 통하여 최적 대안을 구하는 방법이다. 이 때 기준의 중요도를 구하는 방법 중 계층적 분화를 이용하는 방법이 계층분석과정(Analytic Hierarchy Process: AHP)이다.

계층분석과정의 공리 및 기법 이용상의 특징은 ISO 9126 제정 원칙과 잘 부합된다. 계층분석과정이 계층적 분화에 의한 의사결정의 이점을 배경으로 한다는 이면에는 구성된 계층의 완전성을 전제로 한다. 계층의 완전성이란 해결하고자 하는 문제에 필요한 기준들이 계층에 모두 포함되어야 함을 의미하는데 이를 달성하는 데는 축적된 경험과 지식이 요구된다[1]. ISO 9126 품질모델의 범용성과 모델 수립 과정은 계층분석과정의 완전성을 만족시킨다. 계층분석과정과 ISO 9126은 모두 차상위 계층의 동일한 요소에 속하는 차하위 요소들의 수를 인지적 판별이 가능한 수 이내로 제한하고 있다. 계층분석과정은 7 ± 2 이내의 수를 그리고 ISO 9126은 6~8 이내의 수를 제시하였다.



< 그림 1 > 다기준 의사결정 기법의 분류도

3.2. 계층분석과정에 의한 품질 평가모형

3.2.1. 계층분석과정

Saaty(1977)에 의해서 개발된 계층분석과정은 ①의사결정을 위한 요소들을 분해하는 단계, ②요소들을 쌍대비교(pairwise comparison)하는 단계, ③고유치(eigenvalue) 산출 및 일관성 검정 단계, ④산출된 가중치를 종합하여 각 대안들에 대한 복합가중치를 구하는 단계 등 네 단계로 구성된다. 계층분석과정의 방법론은 지금까지 광범위한 의사결정분야에 적용되어 왔으며 이론의 보완을 위해서도 많은 연구가 진행되어 왔다[19, 20].

계층분석과정에 의한 계층별 각 요소의 중요도는 동일한 차상위 요소에 속하는 차하위 요소들 간의 쌍대비교(pairwise comparisons)에 의해 산출된다. 이 때 산출된 요소들의 중요도 합이 요소들이 속한 차상위 요소의 중요도가 되도록 하여 중요도의 계층적 분화를 이루어지게 한다. 쌍대비교할 요소들의 수가 n 이라 하고 상대적 비율척도에 의한 요소들간의 쌍대비교로 이루어진 쌍대비교 행렬을 $A_{n \times n}$ 이라 하면 행렬 A 는 역수행렬(reciprocal matrix)이 된다. 행렬 A 의 고유방정식 $A \cdot W = \lambda \cdot W$ 의 최대 고유치(λ_{\max})에 대응하는 고유벡터 $W^T = (w_1, \dots, w_n)$ 를 $\sum w_i = 1$ 로 정규화 하면 W 는 n 개 요소의 중요도 벡터가 된다. 쌍대비교 행렬에 존재하는 서수적·기수적 불일치 정도를 지수화 한 일관성 지수(consistency index: CI)는 $(\lambda_{\max} - n)/(n - 1)$ 로 정의된다. 동일한 차원의 무작위 행렬로부터 구한 일관성 지수들의 평균을 무작위 지수(random index: RI)라 하며 행렬의 차원에 따른 무작위 지수는 <표 3>과 같다. 쌍대비교 행렬의 일관성 비율(consistency ratio: CR)은 CI/RI 로 정의되며 이 값이 10% 이하이면 바람직한 수준으로 받아들여지며 문제에 따라 20%까지는 허용할 만한 수준으로 간주된다. 또한 계층 전체에 대한 계층일관성 지수(HCI) 및 계층 무작위 지수(HRI)를 이용하여 계층 일관성 비율(consistency ratio of an entire hierarchy: CRH)도 계산이 가능하다 [18]. 본 연구의 계층구조를 고려한 CRH는 다음과 같다.

$$HCI = CI_1^{(0)} + (W_1^{(0)})^T \cdot CI_i^{(1)} \quad \text{그리고} \quad HRI = RI_1^{(0)} + (W_1^{(0)})^T \cdot RI_i^{(1)},$$

$$\therefore CRH = HCI/HRI,$$

여기서 $C_i^{(h)}$ 를 계층 수준 h 의 i 번째 기준에 속하는 차하위 계층 기준들의 집합이라 할 때,

$CI_i^{(h)}$: $(C_i^{(h)})$ 의 요소간 비교에서 산출된 일관성 지수,

$RI_i^{(h)}$: $(C_i^{(h)})$ 의 요소간 비교에서 산출된 무작위 지수,

$W_1^{(0)}$: $(C_1^{(0)})$ 의 중요도 벡터.

집단 의사결정에 있어 개별적인 평가자들로부터 얻은 쌍대비교 값들의 대표 값으로 는 기하평균이 추천되는데 이는 기하평균들로 구성된 쌍대비교 행렬만이 역수행렬 조 건을 만족하기 때문이다. 즉, $1/\left(\prod_{m=1}^M a_m\right)^{1/M} = \left(\prod_{m=1}^M 1/a_m\right)^{1/M}$, 여기서 a_m 은 개별치이며 M 은 그룹의 구성원 수이다[20].

< 표 3 > 무작위 지수

차원(n)	2	3	4	5	6	7	8	9	10
RI	.00	.58	.90	1.12	1.24	1.32	1.41	1.45	1.47

자료: Saaty T.L.(1980), *The Analytical Hierarchy Process*, McGraw-Hill, New York, p. 21.

3.2.2. 품질 평가모형

통계패키지 품질 평가를 위해 구성한 계층적 평가기준에 계층분석과정 기법을 결합 한 본 연구의 품질 평가모형은 다음과 같이 정식화된다.

$$Q_{ik}^{(h)} = \sum_{j \in ID(C_i^{(h)})}^{j_{\max}} w_j^{(h+1)} \cdot Q_{jk}^{(h+1)}, \text{ for } h=0, 1 \left[\begin{array}{l} \text{여기서 } \sum_{j=1}^{j_{\max}} w_j^{(h+1)} = 1, \\ j_{\max} = n(C_i^{(h)}) \end{array} \right.$$

- $Q_{ik}^{(h)}$: 계층 수준 h의 i번째 기준에 대한 k제품의 품질 수준,
- $w_i^{(h)}$: 계층 수준 h의 i번째 기준의 중요도,
- $C_i^{(h)}$: 계층 수준 h의 i번째 기준에 속하는 차하위 계층 기준들의 집합,
- $ID(*)$: 집합 *를 구성하는 원소들의 인덱스(index) 집합,
- $n(*)$: 집합 *를 구성하는 원소들의 수.

위 식에서 $Q_{ik}^{(0)} = Q_k$ 가 k통계패키지 제품의 품질지수가 된다. 실제 각 통계패키지 제품의 품질측정은 계층의 최하위 단계를 구성하는 품질 평가기준에 대해서만 이루어 진다. 이 때 각 기준에 대한 지각된 품질 수준의 측정에는 리커트 척도나 상대적 비 율척도를 적용할 수 있으나, 본 연구에서는 계층분석과정 적용에 있어 이론적 타당성 이 더 있는 상대적 비율척도를 채택하고자 한다(부록 참조, 각 척도의 이론적 차이는 [19] 참조). 상대적 비율척도는 계층분석과정에서 사용하는 기본척도로 척도의 단위 (scaling unit)를 명시하지 않고, 비교되는 두 대안 중 한 쪽의 수준을 기저(base)인 1 로 하고 다른 대안의 수준을 비율로 표시하는 방법이다. 계층의 각 기준에 따른 통계 패키지 제품 대안의 품질수준들 간에는 다음과 같은 조건이 성립하게 된다.

$\sum_k Q_{jk}^{(h)} = 1, k \in ID(P), \text{ for } \forall h, j,$ 여기서 $ID(P)$ 는 제품 대안들의 인덱스 집합.

4. 평가모형의 적용

4.1. 표본의 특성

본 연구의 설문 대상은 통계학을 강의하거나 통계적 방법론을 자주 사용하는 사회과학 분야의 대학 교수로 한정하였다. 이는 본 연구에서 다루어지는 소프트웨어 품질 평가 항목이 통계패키지에 대한 전반적인 지식을 요구함과 동시에 일부 항목은 통계적 지식과 관련된 평가 항목이기 때문이다. 부가적으로 표본 크기에 비하여 신뢰성 있는 응답비율을 높이고자 하는 조사의 효율성 때문이기도 하다.

자료의 수집을 위해서 전국에 소재하고 있는 각 대학에 대표자를 선정하고 설문에 답할 수 있는 대상자를 문의하여 대표자들에게 총 110매의 설문을 우송하였다. 각 설문자의 문의에 답할 수 있도록 대표자에게 설문의 구성을 상세히 설명하고 협조를 요청하였다. 그 결과 110매 중 48매(43.6%)의 설문이 약 1개월 여만에 도착하였다. 다른 설문과는 달리 전문성을 내포하고 있다는 점과 설문에 응답할 수 있는 응답자가 제한되어 있기 때문에 적은 표본을 대상으로 분석하였다.

본 연구에서는 전술한 바 있는 계층분석과정의 CRH가 0.2 이하인 응답을 신뢰성 있는 응답으로 간주하였다. 이에 따라 총 48매의 응답 중 21매(43.8%)의 응답이 신뢰성이 있는 것으로 나타났으며 이들의 기하평균으로 구성된 쌍대비교 행렬로 평가기준인 품질특성의 중요도를 산정 하였다.

통계패키지 품질 평가기준의 중요도가 사용하는 통계패키지에 대한 숙련도에 따라 어떻게 변화되는가를 파악하고자 숙련도를 다음 <표 4>와 같이 정의하였다. 분석에 사용된 응답자의 사용수준은 모두 초급 이상인 것으로 나타났다.

< 표 4 > 통계패키지 사용수준 정의

숙련도 :	정	의
① 초심 (0):	기본적인 사항(입·출력, 기본 명령어, 키 조작 등)도 매뉴얼 없이는 사용이 곤란	
② 초급 (4):	기본적인 사항(입·출력, 기본 명령어, 키 조작 등)은 매뉴얼 없이 사용	
③ 중급(11):	기본적 data-handling 및 각종 통계분석의 기본 명령어 정도는 매뉴얼 없이 사용	
④ 고급 (6):	data-handling 및 원하는 table 또는 graph의 출력이 자유로우며 각종 option에 관하여 상당히 알고 있다.	

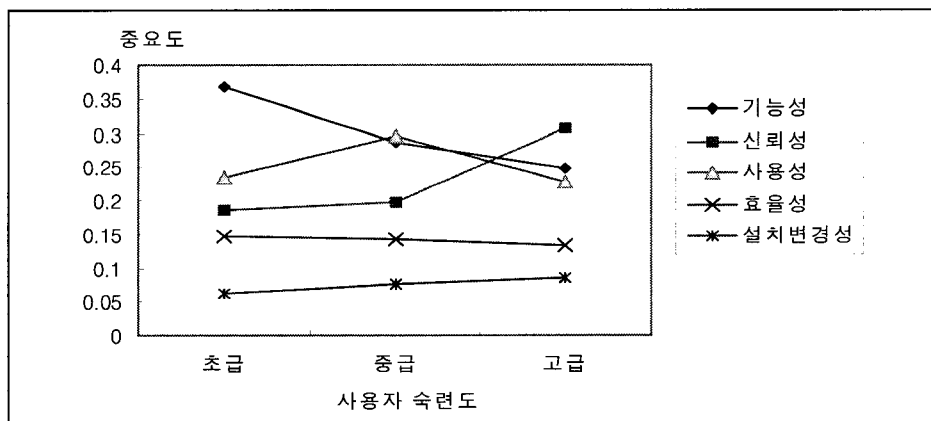
※괄호안은 응답자 수

4.2. 결과 분석

4.2.1. 사용수준에 따른 품질특성의 중요도 변화

숙련도에 따른 각 품질특성의 중요도 변화는 다음 <그림 2>와 <표 5>에 정리되어 있다. 사용자 숙련도로 볼 때 초·중급 사용자의 경우 기능성과 사용성이 중요한 것으로 나타났으며 고급 사용자는 사용성 보다는 신뢰성이 중요한 것으로 나타났다. 이는 대량의 데이터를 통계적으로 처리하는데 있어 경험적으로 소프트웨어 시스템의 안정성 여부가 중요함을 시사하는 것으로 판단된다. 플랫폼의 발달 및 소프트웨어 기능 확대에 따른 프로그램 볼륨의 증가 추세를 고려하면 신뢰성의 중요도는 더욱 높아질 것으로 사료된다. 소프트웨어에 익숙할수록 사용성에 대한 중요도는 다소 감소하는 것으로 나타났는데 그 논리적 개연성이 인정될 만하다. 응답자 전체 평균으로는 기능성, 사용성, 신뢰성, 효율성 그리고 설치·변경성 순서로 중요한 것으로 나타났다. 설치·변경성의 중요도가 가장 낮게 나온 것은 윈도우환경에서 소프트웨어 설치 및 변경의 자동화가 상당히 이루어져 큰 문제점이 야기되지 않기 때문인 것으로 판단된다.

품질 부특성의 중요도에 관하여 특징적인 면만을 간략히 언급하면 다음과 같다. 기능성의 하위특성 중 필요한 기능 범위 및 수준의 구비 여부를 나타내는 적절성의 중요도가 전체 품질 부특성 중 가장 중요한 것으로 나타났으며 통계 처리용이라는 소프트웨어 목적상 정밀성이 상당히 중요한 것으로 산출되었다. 신뢰성의 하위 특성에서는 고장/장애로 인한 시스템 종료상황 하에서 데이터 및 프로그램의 보호 또는 복원을 나타내는 회복성이 중요한 것으로 나타났다. 사용성의 하위 특성으로는 학습성이 중요하며 효율성의 경우 자원행동 보다는 시간행동이 더 중요한 것으로 확인되었다.



< 그림 2 > 숙련도에 따른 중요도의 변화

< 표 5 > 품질특성의 중요도

주특성	부특성	숙련도에 따른 중요도			
		초급사용자(4)	중급사용자(11)	고급사용자(6)	전체(21)
기능성		0.3673	0.2865	0.2481	0.2906
	적절성	0.1857 (0.5055)	0.1088 (0.3798)	0.0981 (0.3954)	0.1195 (0.4112)
	정밀성	0.1037 (0.2824)	0.0810 (0.2825)	0.0652 (0.2630)	0.0814 (0.2802)
	상호 운영성	0.0393 (0.1069)	0.0730 (0.2547)	0.0556 (0.2241)	0.0615 (0.2116)
	보안성	0.0386 (0.1052)	0.0238 (0.0830)	0.0292 (0.1175)	0.0282 (0.0969)
신뢰성		0.1871	0.1980	0.3071	0.2242
	성숙성	0.0852 (0.4552)	0.0570 (0.2878)	0.1285 (0.4185)	0.0795 (0.3546)
	고장 허용	0.0504 (0.2692)	0.0553 (0.2795)	0.0838 (0.2730)	0.0627 (0.2797)
	회복성	0.0516 (0.2757)	0.0857 (0.4326)	0.0947 (0.3085)	0.0820 (0.3657)
사용성		0.2354	0.2962	0.2275	0.2661
	이해성	0.0851 (0.3613)	0.0856 (0.2891)	0.0870 (0.3826)	0.0875 (0.3286)
	학습성	0.0834 (0.3542)	0.1137 (0.3837)	0.0862 (0.3787)	0.1008 (0.3786)
	운용성	0.0670 (0.2845)	0.0969 (0.3272)	0.0543 (0.2387)	0.0779 (0.2928)
효율성		0.1472	0.1422	0.1332	0.1421
	시간행동	0.0763 (0.5180)	0.0731 (0.5142)	0.0787 (0.5905)	0.0763 (0.5369)
	자원행동	0.0710 (0.4820)	0.0691 (0.4858)	0.0545 (0.4095)	0.0658 (0.4631)
설치·변경성		0.0629	0.0770	0.0842	0.0770
	변경성	0.0340 (0.5403)	0.0295 (0.3826)	0.0369 (0.4383)	0.0331 (0.4293)
	적응성	0.0124 (0.1969)	0.0274 (0.3552)	0.0245 (0.2910)	0.0233 (0.3031)
	설치성	0.0165 (0.2628)	0.0202 (0.2622)	0.0228 (0.2707)	0.0206 (0.2676)

※괄호안은 그룹내 중요도

4.2.2. 제품 비교의 예시

앞서 제시한 품질 모형에 따라 대표적인 통계패키지인 SAS와 SPSS를 대상으로 지각된 품질 비교를 예시하고자 한다. 총 48명의 응답자 중 SAS를 주로 사용하는 응답자의 수는 17명(35.4%), SPSS의 경우는 26명(54.2%) 그리고 기타 5명(10.4%)으로 조사되었다. 또한 SAS와 SPSS를 모두 사용해 본 경험이 있는 응답자는 15명(31.3%)이었으며 이들로부터 상대적 비율척도에 의한 두 제품의 품질 우수성을 조사하였다. 평가 대상의 동질성을 고려하여 두 제품 모두 PC 윈도우 버전(version)으로 한정하였다. 비교에 사용된 척도는 계층분석과정에서 사용하는 척도이며 응답자들의 자료는 기하 평균을 통하여 제품간 상대적 우수성으로 산출되었고 그들의 합이 1이 되도록 정규화 되었다.

<표 6>에 나타난 바와 같이 두 제품의 상대적 품질 지수를 산출한 결과, SPSS가 SAS에 비하여 다소 우수한 것으로 평가되었다. 1단계 품질 주특성을 기준으로 볼 때, 기능성과 신뢰성에서는 SAS가 근소한 차이로 우수하고 사용성, 효율성 그리고 설치·변경성에서는 SPSS가 우수한 것으로 평가되었다. 일반적으로 기능성의 향상은 프로그램의 불륨의 증가를 유발할 가능성이 높고 이는 시간 및 자원 행동을 포함하는 효율성과는 음의 상관관계(negative correlation)를 갖는다.

본 연구의 제품 비교에는 상대적 비율척도에 의한 자료와 이들의 기하평균을 이용한 관계로 제품간 차이에 대한 t-검정과 같은 통계적 검정은 이론적으로 성립되지 않는다. 따라서 단순한 우열만을 중심으로 결과를 해석하였다. 아울러 본 연구의 표본은 통계와 통계패키지에 관한 전문가로서 대학 교수들로 하였으나 사회과학 분야에 치중 되었으므로 제품 비교 결과를 일반화하는 데는 무리가 따를 수도 있음을 밝혀둔다.

< 표 6 > 품질 비교 결과

주특성	부특성 (중요도)	상대적 우수성 (15)*	
		SAS	SPSS
기능성	(0.2906)	0.5172 ⁺	0.4827
	적절성 (0.4112)	0.4573	0.5427 ⁺
	정밀성 (0.2802)	0.6465 ⁺	0.3535
	상호 운영성 (0.2116)	0.4494	0.5506 ⁺
	보안성 (0.0969)	0.5457 ⁺	0.4543
신뢰성	(0.2242)	0.5081 ⁺	0.4919
	성숙성 (0.3546)	0.5413 ⁺	0.4587
	고장 허용 (0.2797)	0.4561	0.5439 ⁺
	회복성 (0.3657)	0.5157 ⁺	0.4843
사용성	(0.2661)	0.3185	0.6815 ⁺
	이해성 (0.3286)	0.2975	0.7025 ⁺
	학습성 (0.3786)	0.2378	0.7622 ⁺
	운용성 (0.2928)	0.4469	0.5531 ⁺
효율성	(0.1421)	0.4096	0.5904 ⁺
	시간 행동 (0.5369)	0.3982	0.6018 ⁺
	자원 행동 (0.4631)	0.4229	0.5771 ⁺
설치·변경성	(0.0770)	0.3299	0.6701 ⁺
	변경성 (0.4293)	0.3034	0.6966 ⁺
	적응성 (0.3031)	0.3751	0.6249 ⁺
	설치성 (0.2676)	0.3214	0.6786 ⁺
품질 지수 (Q_k)		0.4327	0.5673

* : 괄호안은 응답자 수
 + : 상대적 우수

5. 결론

본 연구에서는 사용자에게 의해 지각된 통계패키지 품질 비교 방안에 관한 연구를 수행하였다. 이를 위하여 우선 ISO 9126 품질모델에 근거한 통계패키지 소프트웨어 평가기준을 계층적으로 분류하고, 계층분석과정(AHP)을 통한 평가모형을 제시하였다. 또한 제시한 평가모형을 실제로 SAS와 SPSS에 적용하여 예시적으로 분석하였다.

본 연구의 결과 및 의의 그리고 앞으로의 연구방향을 제시하면 다음과 같다.

첫째, 사용자의 통계패키지에 대한 숙련도에 따라 평가기준의 중요도가 달라짐을 밝혔다. 그 결과 사용자 숙련도로 볼 때 초·중급 사용자의 경우 기능성과 사용성이 중요한 것으로 나타났으며 고급 사용자는 사용성 보다는 신뢰성이 중요한 것으로 나타났다. 응답자 전체 평균으로는 기능성, 사용성, 신뢰성, 효율성 그리고 설치·변경성 순서로 중요한 것으로 나타났다.

둘째, 본 연구의 의의는 통계패키지의 사용자 측면과 공급자 측면으로 나누어 파악할 수 있다. 사용자 측면에서 볼 때, 본 연구의 결과는 향후 통계패키지 사용자에게 제품선택의 지침이 될 수 있다는 점에서 의의가 있다. 따라서 세부적 평가 항목의 중요도 변화는 최종사용자로 하여금 미래 지향적인 제품선택을 가능하게 하며 각 평가기준에 따른 제품간 우열은 사용목적에 보다 적절한 제품을 선택하는데 도움이 될 것이다. 공급자에게 제공하는 의미로서는 현재 시장에서 통계패키지에 대한 품질측면의 만족을 높이기 위한 핵심요소 및 향후 자사 제품의 개선점 파악에 본 연구의 결과가 도움이 될 것으로 사료된다.

셋째, 지금까지 소프트웨어 품질 평가를 위한 객관적 지표(objective indicator)의 개발에 관한 연구는 시도되고 있으나 최종사용자의 지각된 품질과의 관련성에 관한 연구가 미흡한 실정이다. 향후, 최종사용자의 지각된 품질과 높은 상관관계가 있는 객관적 지표의 확인과 이를 이용한 소프트웨어 품질 평가가 이루어져야 할 것이다. 또한, 품질 이외에 가격, 광고, 제품 충성도 그리고 개발자의 명성을 등을 종합적으로 고려한 제품 선택에 관한 연구도 요구된다.

넷째, 본 연구의 표본이 사회과학 분야의 교수들에 치중되었으므로 제품 비교 결과를 일반화하는 데는 다소 무리가 따를 수 있음을 밝힌 바 있다. 따라서 이공계 교수들을 포함한 폭넓은 SAS/SPSS 이용자들의 의견을 반영한 좀 더 세밀한 연구는 추후 연구과제로 남긴다.

참고문헌

- [1] 이상석(1996), "항공서비스 품질의 경쟁력," 「품질경영학회지」, 24권, 4호, pp. 124-140.
- [2] 이호배, 이성근, 배수현(1994), 「SPSS/PC+를 이용한 기초 통계 분석」, 데이터리서치.
- [3] Anderson, E.E.(1990), "Choice Models for The Evaluation and Selection of Software Packages," *Journal of Management Information Systems*, Vol. 6 No. 4, pp. 123-138.
- [4] Boehm, B.W., J.R. Brown, M. Lipow, G.J. MacLeod, and M.J. Merritt(1978), *Characteristics of Software Quality*, North-Holland, New York.
- [5] Cody, R.P., and J.K. Smith(1991), *Applied Statistic and the SAS Programming Language*, Prentice-Hall, New Jersey.

- [6] Dromey, R.G.(1995), "A Model for Software Product Quality," *IEEE Transactions on Software Engineering*, Vol. 21, No.2, pp. 146-162.
- [7] Evans, M. and J. Marciniak(1987), *Software Quality Assurance and Management*, John Wiley & Sons, New York.
- [8] Fritz, C.A. and B.D. Carter(1994), *A Classification, and Summary of Software Evaluation and Selection Methodologies*, Computer Science Technical Report No. 940823, Mississippi State University.
- [9] [http\www.sas.com](http://www.sas.com)
- [10] [http\www.spss.com](http://www.spss.com)
- [11] ISO 8402(1994), *Quality Management and Quality Assurance - Vocabulary*, ISO.
- [12] ISO 9126(1991), *Information Technology - Software Product Evaluation - Quality Characteristics and Guidelines for Their Use*, ISO.
- [13] Jung H.W. and Yoon, M.S.(1996), "A Software Product Quality Evaluation and Resource Allocation Model," *5th. European Conference on Software Quality, Conference Proceedings*, pp. 286-294.
- [14] Kekre, S., M.S. Krishnan and K. Srinivasan(1995), "Drivers of Customer Satisfaction for Software Products: Implications for Design and Service Support", *Management Science*, Vol. 41, No. 9, pp. 1456-1470.
- [15] Kitchenham, B. and Pfleeger, S.L.(1996), "Software Quality: The Elusive Target," *IEEE Software*, January, pp. 12-21.
- [16] McCall, J.A., Richards, P.K. and Walters, G.F.(1977), *Factors in Software Quality (3 Vol.)*, RADC-TR-77-369.
- [17] Saaty, T.L.(1977), "A Scaling Method for Priorities in Hierarchical Structures," *Journal of Mathematical Psychology*, Vol. 15, pp. 234-281.
- [18] Saaty, T.L.(1980), *The Analytic Hierarchy Process*, McGraw-Hill, New York.
- [19] Saaty, T.L.(1994), "Highlights and Critical-Points in the Theory and Application of the Analytic Hierarchy Process," *European Journal of Operational Research*, Vol 74, No. 3, pp. 426-447.
- [20] Saaty, T.L. and Vargas, L.G.(1982), *The Logic of Priorities*, Kluwer-Nijhoff Publishing, London.
- [21] Schoemaker, P.J.H. and Waid, C.C.(1982), "An Experimental Comparison of Different Approaches to Determining Weights in Additive Utility Models," *Management Science*, Vol. 28, No. 2, pp. 182-196.
- [22] Yoon, K.P. and Hwang, C.L.(1995), *Multiple Attribute Decision Making*, Sage Publications, London.
- [23] Yoon M.S.(1997), *S/W Quality Evaluation Model Using the AHP*, Thesis for The Degree of Doctor, Korea University.

부록

품질특성의 중요도 평가와 제품 비교를 위한 설문 예

예) 통계패키지 품질특성의 중요도 평가																		
특성 A	A가 절대로 중요	A가 확실히 중요	A가 매우 중요	A가 약간 중요	A가	서로 비슷한 정도	B가 약간 중요	B가 매우 중요	B가 확실히 중요	B가 절대로 중요	특성 B							
	9	8	7	6	5	4	3	2	1	1/2	1/3	1/4	1/5	1/6	1/7	1/8	1/9	
기능성					√													신뢰성
기능성										√								사용성
기능성						√												효율성
<p>☞ 응답 예의 설명</p> <p>첫째 행: 특성 A란의 기능성이 특성 B란의 신뢰성 보다 매우 중요하다고 생각하는 경우</p> <p>둘째 행: 특성 B란의 사용성이 특성 A란의 기능성 보다 약간 중요하다고 생각하는 경우</p> <p>셋째 행: 약간 중요와 매우 중요한 사이인 경우</p>																		

예) 2가지 통계패키지의 우수성 비교																			
품질특성	제품	제품 A SAS						제품 B SPSS											
		A가 절대 우수	A가 확실히 우수	A가 매우 우수	A가 약간 우수	A가	서로 비슷한 정도	B가 약간 우수	B가 매우 우수	B가 확실히 우수	B가 절대 우수								
	품질 주특성	품질 부특성	9	8	7	6	5	4	3	2	1	1/2	1/3	1/4	1/5	1/6	1/7	1/8	1/9
기능성	적절성						√												
	정밀성												√						
	상호운영성										√								
<p>☞ 응답 예의 설명</p> <p>첫째 행: 적절성 측면에서 제품A가 제품B 보다 매우 우수할 때</p> <p>둘째 행: 정밀성 측면에서 제품B가 제품A 보다 약간 우수할 때</p> <p>셋째 행: 상호운영성 측면에서 두 제품이 서로 비슷할 때</p>																			