

시뮬레이션의 통계적 분석을 위한 전문가 시스템*

박영홍**

A Statistical Expert System for Simulation Analysis-Revised

Young Hong Park**

(Abstract)

Simulation is one of the most widely used techniques in operations research and management science, but there are several impediments to even wider acceptance and use of simulation. One of the more significant limitations is the dependence of simulation on statistical methodology. This research identifies eighteen different statistical issues in simulation methodology and develops an expert system which could through interactive dialog with simulation analysts offer advice on statistical approaches which might be used to deal with particular issues and to accomplish required statistical computations. This research revises the previous study published in *Simulation* by incorporating additional statistical issues in the expert system to enhance its performance in analyzing a given simulation problem with statistical methodologies. An overview of the revised system is given and illustrations of the capabilities of the system are presented.

1. 서론

시뮬레이션은 일반적으로 경영과학기법 중 가장 많이 이용되는 최적화 기법 중의 하나로 인식되어왔다. 실제로 미국의 500개 대기업을 대상으로 기업의 제반 의사결정과정에 이용되는 과학적 경영기법의 이용 빈도에 대한 한 설문조사에서 시뮬레이션이 통계적 기법에 이어 두번째로 널리 이용되는 것으로 나타났다 [1]. 또한 최근 Operations Research 관련 학술잡지에 발표된 연구논문에 대한 조사 역시 시뮬레이션이 가

장 자주 인용되는 최적화 기법 중의 하나임을 보여 주고 있다. 더우기 근래에 들어서 컴퓨터 기법의 발달과 보다 편리해진 컴퓨터 그래픽 기능을 통한 자동화된 컴퓨터 코드의 작성(예:Block diagram) 및 애니메이션, 그리고 GUI(Graphic User Interface)를 포함한 사용자 중심의 소프트웨어 개발 및 CAM(Computer-Aided Manufacturing)과의 연계가 가능해짐으로서 시뮬레이션 기법은 모든 분야의 의사결정문제에 보다 널리 이용되게 되었다.

최적의 의사결정을 유도함에 있어서 나타난 현상이

* 이 논문은 1992년도 교육부 지원 한국학술진흥재단의 자유공모(지방대학육성) 과제 학술연구조성비에 의하여 연구되었음.

** 관동대학교 산업공학과

나 주어진 문제를 수리적 모델로 변형시킨 후 난수를 이용하여 확률적 방법에 의해 변형된 수리적 모델을 분석하는 시뮬레이션 기법은 경영분석에서부터 경제 문제, 환경문제, 나아가 우주비행의 모의 실험에 이르기까지 매우 다양하게 이용되면서 최적의 문제 해결 방법을 제공할 수 있는 분석기법으로 그 자리를 굳혔지만 시뮬레이션 기법에 내포된 복잡성 및 통계적, 확률적 전문지식에 대한 요구 때문에 통계적 전문지식을 갖고 있지 못한 현장의 Operations Research 관련업무의 실무자들이 손쉽게 이용할 수 없는 기법이기도 하였다. Law 와 Kelton은 확률적 이론의 적용과 통계적 기법의 사용은 시뮬레이션 분석과정에서 매우 중요하고 중심적인 부분이 되므로 시뮬레이션을 담당하게 되는 모든 분석 팀은 적어도 그 구성원 중에 확률적 이론 및 통계적 기법들에 대하여 철저히 교육된 사람을 한 명 이상 포함시켜야 한다고 주장하였다[8]. 하지만 시뮬레이션 기법을 이용하는 대부분의 사람들은 통계학에 대한 전문지식을 갖지 못하고 있는 경우가 많은 것으로 조사결과 나타났다[13].

〈표 1〉은 일반적으로 받아들여지는 시뮬레이션 분석과정을 단계별로 구분하면서 행으로는 처음 단계인 문제구성 단계에서부터 마지막 단계인 모델의 설치 및 설명서 작성단계까지 매 단계마다 시뮬레이션 분석가가 시뮬레이션 연구를 수행하는데 있어서 반드시 고려해야할 통계적인 쟁점들을 나열하고 있다. 또 행으로는 이들 행에 나타난 통계적 쟁점들을 해결하는데 이용될 수 있는 통계적 기법을 제시하고 있다. 〈표 1〉이 보여 주듯이 대체로 시뮬레이션 분석과정의 각 단계마다 대두되는 통계적 쟁점들은 그 특성에 따라 다시 하나 또는 그 이상의 통계적 기법을 포함하게 된다. 〈표 1〉에 제시된 자료로부터 얻게 되는 두 가지 결론은 첫째 시뮬레이션 기법에는 많은 통계적 요소들이 내재되어 있다는 것과 두 번째 시뮬레이션 분석가가 이들 방법론적인 문제에 깊이 빠져들수록 이런 통계적 문제점들은 보다 복잡해지게 된다는 것이다.

하지만 대부분의 현장의 시뮬레이션 분석가들은 충분한 통계적 지식을 갖추지 못하고 있으므로 실제로 많은 시뮬레이션 분석은 통계적 쟁점들을 전적으로 무시하거나 또는 시뮬레이션의 방법론적인 문제점들

을 충분히 고려하여 분석에 반영하지 못하게 된다. 하나의 예로 1986 년도에 13개 주요 Operations Research 관련 저널에 발표된 354 개의 연구논문 중 시뮬레이션에 관한 내용을 다루고 있는 62개의 연구논문에 대한 비공식 조사는 아래의 내용을 보여 주고 있다.

1. 이들 62 개의 연구논문 중 27 개는 통계적 쟁점에 대하여 전혀 언급이 없었고
2. 나머지 35 개의 연구논문 중에서도 15개는 사용한 통계적 방법에 대한 근거를 제시하지 않았거나 사용한 방법에 포함된 통계적 가정을 설명하지 않고 있었다.

단순히 이러한 비공식적인 조사결과에 근거해서 통계적 방법이 시뮬레이션 연구에 부적절하게 적용되고 있다고 선불리 결론 내릴 수는 없겠지만 적어도 이것은 시뮬레이션 분석과정에서 통계적 고려사항에 대한 만연된 무관심과 그로 인해 과연 타당한 결론이 유도될 수 있는가에 대한 의구심을 제기하게 된다.

최근의 많은 연구는 전문가 시스템 기법이 이러한 문제에 대한 하나의 해결방법을 제공할 수 있을 것이라고 제안했다. Shannon, Meyer, and Adelsberger는 지적하기를 시뮬레이션에 관한 지식이 계속적으로 증가하여 시뮬레이션 전문가들조차 관련된 모든 사항을 파악하기 힘든 상황까지 이르렀다고 하면서 이러한 상황을 해결할 수 있는 한 방법으로 전문가시스템의 사용을 제안하였고[12] O'Keefe는 숙련되지 못한 시뮬레이션 분석가를 위해 그를 도울 수 있는 전문가시스템의 사용을 주장하였다[10].

위에서 언급한 Law 와 Kelton의 제안에 따라 통계적 전문가를 모든 시뮬레이션 분석 팀에 할당하는 것은 불가능하겠지만 통계적 전문가시스템을 통해 이들에게 통계적 자문을 제공하는 것은 가능할 것이다. 본 연구의 목적은 시뮬레이션 분석과정에서 발생할 수 있는 여러 가지 통계적 쟁점들을 밝히고 컴퓨터와의 대화적 접근방법을 통해 시뮬레이션 분석 가에게 특정한 통계적 쟁점과 관련하여 이용할 수 있는 모든 통계적 기법의 종류와 이의 선택 및 적용방법에 대하여 자문을 제공하고 경우에 따라 필요한 모든 통계적 계산을 수행할 수 있는 전문가시스템을 개발하는 것이다.

하므로 통계적 방법을 통해 주어진 시뮬레이션 문제를 분석하는데 그 적용범위를 확대하였고 동시에 전문가 시스템에서 해결할 수 있는 문제의 영역을 확장시키기 위하여 지식베이스에 포함된 통계적 기법의 수도 증가시켰다.

본 연구의 본문에서는 SESSA에 대한 개요를 설명하고 동시에 SESSA가 시뮬레이션 분석과정에서 어떻게 통계적 자문을 제공하는가를 보여 주기 위해 SESSA를 이용하는 과정에서 접하게 되는 여러 가지 문답을 제시하고 결론적으로 SESSA를 이용함으로써 기대되는 효과를 기술한다.

2. 전문가 시스템

가장 널리 이용되고 있는 인공지능의 한 연구분야인 전문가 시스템 기법은 의학 분야에서부터 지질탐사나 기업의 신용평가에 관한 의사결정에 이르기까지 매우 다양한 전문영역에 응용되어 필요한 자문을 제공하고 있다. 간단하게 말하면 전문가 시스템은 관심의 대상이 되는 전문영역에 관한 전문지식을 포함하고 있는 하나의 컴퓨터 프로그램으로서 해당 전문영역에서 필요로 하는 전문지식을 제공할 수 있는 매우 지능적인 컴퓨터 프로그램을 의미하게 된다. 따라서 시뮬레이션의 통계적 분석을 위한 전문가 시스템이라는 의미의 SESSA는 실제로 시뮬레이션과 관련된 일체의 통계적 지식을 지능적인 컴퓨터 프로그램의 형태로 구현하고 있다.

몇몇 연구가들이 지능적인 통계적 소프트웨어와 통계적인 전문가 시스템의 일반적인 사항에 대하여 연구하였지만 이러한 노력들은 이론적인 서술에만 치중하였고 보다 구체적으로 개발된 실제적인 시스템들을 설명하거나 기술하지는 못했다[5][7]. 이들 연구 중에는 특별히 일반인들의 통계적 기법의 사용을 지원하기 위한 전문가 시스템이 고안되었는데 예를 들면 DEXTOR가 실험계획법을 이용하려는 사람들을 돕기 위해 개발되었고[4], REX가 회기분석을 위해[3], 그리고 STATPATH[11] 와 SASS[6]는 주어진 연구목적에 부합하는 적절한 통계적 기법을 선정하기 위해 시도되었다. 하지만 시뮬레이션의 통계적 쟁점과 관련

되어 개발된 전문가 시스템은 SESSA가 최초의 시도이었다.

2.1 SESSA의 구조

SESSA를 위해 선정된 기본적인 구조는 모듈로서 <표 1>에 나타난 18 개의 통계적 쟁점들이 그대로 SESSA의 모듈을 구성하게 된다. SESSA의 지식베이스에 포함된 이들 18 개의 모듈에 대한 개괄적인 설명이 <표 2>에 주어져 있다.

SESSA를 이용할 때 사용자는 먼저 전문가시스템과 그의 처리범위에 대하여 개략적으로 살펴볼 수 있는 여러 개의 기초적인 화면(panel)을 만나게 된다. 그후 SESSA의 주 메뉴가 화면에 나타나면(<표 3> 참조) 사용자는 주메뉴 중에서 자신이 필요로 하는 통계적 자문을 포함하고 있는 항목을 선택하게 된다.

SESSA에 포함된 각각의 모듈은 먼저 모듈에 포함되어 있는 통계적 쟁점들을 소개하고 적절한 방법론적인 접근방법을 제시하며 또 이들과 관련된 참고문헌들에 대한 정보를 제공하고 여러 기법을 비교.평가하여 선택할 수 있는 기준을 설명하는 6 개에서 10 개 정도의 화면으로 구성되는 튜토리얼을 포함하고 있다. 물론 사용자가 원할 때는 이러한 튜토리얼을 건너뛰어 곧바로 자문단계로 넘어갈 수도 있다.

자문단계는 모듈에 따라 그 수는 다르지만 여러 개의 질문들로 구성되어 이를 통해 전문가시스템은 사용자가 안고 있는 통계적 문제의 성격과 내용을 파악하여 적절한 통계적 기법이나 방법론적인 절차를 제시하게 된다. 자문단계에서 묻게 되는 질문들은 이미 튜토리얼에서 설명된 것들이므로 사용자는 대부분의 질문들에 대하여 별 어려움 없이 대답할 수 있게 된다. 이제 모든 질문에 대한 대답이 완료되면 전문가 시스템은 하나 또는 그 이상의 적절한 통계적 기법이나 절차를 제시하게 되지만 만약 사용자가 안고 있는 문제가 매우 복잡하거나 특이하여 SESSA의 지식베이스에 포함되지 못한 경우에는 통계학자와 상의하여 필요한 자문을 얻으라는 제안을 하게된다.

전자의 경우 사용자가 원한다면 전문가시스템은 주어진 문제에 대한 적절한 통계적 해결방법을 제시할

〈표 2〉 SESSA의 모듈에 대한 설명

1. **입력변수들의 값을 추정하기 위한 자료수집**
4 개의 다른 표본추출방법이 각각 사용되어야하는 상황과 이들 방법의 필요한 이용절차에 대한 설명과 함께 제시된다. 여기에 포함된 표본추출방법은 random, stratified, systematic, cluster 방법이다.
2. **입력변수들의 값의 측정**
10 개의 다른 통계적 분포에 대하여 각 분포의 매개변수(parameter)의 값을 결정할 수 있는 절차를 포함하고 있다. 이들 10 개의 통계적 분포에는 정규분포, 지수분포, 이항분포 등이 포함되어 있다.
3. **입력변수들의 확률분포에 대한 판별**
적합도 검정과 가설검증을 이용하여 입력변수들의 값을 정규분포, 지수분포, 이항분포 등을 포함한 10개의 서로 다른 통계적 분포에 대하여 비교하면서 적절한 확률분포를 찾아낸다.
4. **난수의 발생**
가장 많이 사용되는 세가지의 난수 발생방법이 제시된다. 이들은 mixed congruential method, multiplicative congruential method, 그리고 additive congruential method 이다.
5. **확률변수의 발생**
확률변수를 발생시키는 방법으로서 역변환 방법과 합성법 그리고 기각방법이 포함되었다.
6. **난수 발생기에 대한 검증**
난수 발생기에 의해 만들어진 난수의 난수성(randomness), 일양성(uniformity), 그리고 독립성(independence)을 검증하기 위해 2 개의 이론적인 검증방법과 8 개의 실험적인 방법들이 포함되었다.
7. **확률변수발생기에 대한 검증**
두개의 전형적인 적합도 검증방법인 Chi-Square 방법과 Kolmogorov-Smirnov 검증방법들이 확률변수발생기의 적합성을 검증하기 위해 이용되었다.
8. **모델의 반응치와 실제 시스템의 측정치와의 비교**
가설검증을 이용하여 시뮬레이션 모델로부터 얻어진 평균값과 시스템으로부터 직접 측정되거나 가정된 값과의 유의한 차이를 판별한다.
9. **Model Spectra 와 System Spectra의 비교**
Spectra 분석방법이 두 개의 시계열 자료를 객관적으로 비교하고 신뢰구간을 설정하기 위한 방법을 제공하기 위하여 이용되었다.
10. **초기화 조건의 영향에 대한 평가**
초기상태(transient period)와 안정된 상태(steady-state)를 구분할 수 있는 8 개의 방법들이 제시된다. 이들 방법에는 Gafarian's 방법, Conway's 방법, Fishman's 첫째 및 둘째 방법, Ermshoff and Sisson's 첫번째 방법 및 두번째 방법, 그리고 Gordon's 방법과 Tocher's 방법 등이다.
11. **시뮬레이션 분석시간의 결정 및 반복횟수의 결정**
출력변수(response variables)의 측정치에 대하여 일정한 수준 이상의 정확도를 확보하기 위해 필요한 시뮬레이션의 실행시간 및 반복횟수를 결정하기 위한 절차가 22 개의 공식을 중심으로 포함되어 있다.
12. **평가할 대체안에 대한 수립**
6 개의 서로 다른 실험계획법이 소개된다. 이들은 각각 2^k designs, 2^{k-p} designs, random designs, supersaturated designs, grouping-screening designs, and response surface designs 이다.
13. **출력변수의 편기(bias)의 감소**
분산감소기법이 출력변수에 대한 측정치의 편기를 줄이기 위해 제공되는데 7 개의 서로 다른 기법들이 포함되었다.
14. **출력결과로부터 출력변수의 값에 대한 측정**
출력변수의 값을 측정하기 위해 필요한 자료를 생성시키는 방법으로 replications of runs, sub-interval, regenerative method, 그리고 parametric method 의 4가지 방법이 소개된다.
15. **측정값에 대한 신뢰구간의 설정**
측정값에 대한 신뢰구간을 결정하는 20여개의 공식들이 제시된다. 이들 공식은 파라미터(평균, 분산)의 형태와 수, 관찰개수, 및 측정치에 대한 독립성과 정규분포에 대한 가정 여부에 따라 결정된다.
16. **출력변수의 값에 대한 확률분포의 결정**
적합도검증과 가설검증을 이용하여 출력자료를 정규분포, 지수분포 등을 포함한 10 개의 서로 다른 통계적 분포에 대하여 비교하면서 적절한 확률분포를 검증할 수 있는 절차를 포함하고 있다.
17. **출력자료를 이용한 대체안에 대한 비교분석**
평균과 비율, 및 분산에 대한 가설검증을 위한 테스트가 서로 다른 대체적인 시스템 디자인을 비교·평가하기 위하여 포함되어 있다. z-분포와 t-분포를 따라 20 개의 서로 다른 공식들이 주어진다.
18. **출력변수와 입력변수의 연관성 파악**
입력변수와 출력변수와의 관계를 파악하고 추정하기 위해 상관분석과 회귀분석기법들이 이용된다.

〈표 3〉 SESSA의 주메뉴

M E N U	
1.	Collecting data to estimate values for input variables
2.	Estimating values for input variables
3.	Identifying standard distribution for input variables
4.	Generating pseudo random numbers
5.	Generating random variates
6.	Testing random number generators
7.	Testing random variates generators
8.	Comparing model response estimates to system estimates
9.	Comparing model spectra to system spectra
10.	Assessing the impact of initial condition
11.	Determining run length and number of replications
12.	Specifying alternatives to be evaluated
13.	Reducing bias in response variable
14.	Estimating response variable values from output
15.	Constructing confidence intervals for estimates
16.	Associating standard distribution with response variables
17.	Comparing output from alternative designs.
18.	Estimating response/input variable relationships

뿐만 아니라 제시된 통계적 기법이나 절차를 적용하는데 필요한 모든 통계적 계산을 수행할 수도 있다. 이때 대부분의 통계적 기법들은 그 계산을 위해 입력 자료를 요구하게 되는데 전문가시스템은 필요한 모든 자료를 이미 작성되어 저장된 디스크 파일로부터 읽어 들일 수도 있고 자료화일이 준비되지 못한 경우에는 사용자로부터 대화식으로 직접 필요한 자료를 입력 받을 수도 있다. 필요한 모든 자료가 입력되면 최종적으로 SESSA는 사용자가 선택한 통계적 문제에 대한 해결방안을 이에 대한 설명과 함께 사용자가 이해할 수 있는 수준으로 제공하게 된다.

2.2 SESSA의 성격

전문가시스템은 일반적으로 지식베이스와 추론기관을 포함하게 된다. 지식베이스는 전문가시스템의 대상이 되는 전문영역에 관한 지식을 법칙과 사실의 형태로 표현한 것이고 추론기관은 단순히 어떤 의사결정을 내리기 위해 지식베이스에 포함된 지식을 이용하

기 위한 절차를 의미하는 것이다. SESSA는 PC의 DOS 환경에서 "C" 언어를 이용하여 개발되었다.

SESSA에 포함된 지식베이스는 기본적으로 C언어를 이용한 if-then 절의 형태로 구성된 법칙들의 집합으로 이루어져 주메뉴에서 사용자가 선택한 통계적 문제들에 대하여 필요한 기본적인 설명과 튜토리얼 및 해결방법을 화면에 보여 주기 위해 사용된다. 이들 법칙은 다시 TOP-RULE과 SUB-RULE의 두 가지로 나누어져 TOP-RULE은 사용자가 안고 있는 통계적 문제의 성격을 파악하기 위해 이용되고 SUB-RULE은 TOP-RULE을 통해 파악된 통계적 문제를 해결할 수 있는 구체적인 기법이나 절차를 제시하는 역할을 수행한다.

사용자가 한 모듈을 선택해서 이에 대한 통계적 자료를 얻고자 할 때 우선 수많은 법칙 중에서 TOP-RULE을 통해 선택된 메뉴와 관련된 질문을 시작하게 된다. TOP-RULE의 형태를 영어로 쓰면 아래의 〈표 4〉와 같다.

SESSA의 TOP-RULE은 여러 절차를 거치면서 해당법칙에 포함된 기본적인 질문에 대한 사용자의 응답을 검토한 후 필요할 경우 추가적인 질문을 하게 된다. 이와같이 TOP-RULE은 사용자가 안고 있는 통계적 문제의 성격을 파악하기 위해 필요한 질문을 총괄하는 역할을 수행하게 된다. TOP-RULE을 통해 통계

〈표 4〉 SESSA에 포함된 TOP-RULE의 예

<p>If (조건절)</p> <p>the menu selection is 15.</p> <p>Then (실행절)</p> <p>ask the following questions:</p> <ol style="list-style-type: none"> 1. How much variability is present in the sample? 2. What size risks can be taken? 3. What is the type of distribution of the response variable? 4. What is the number of observations in the sample? 5. What is the number of samples, one or two? 6. What is the type of parameter, proportion or mean?

적 문제의 성격이 파악되면 해당 문제를 해결할 수 있는 기법이나 절차를 찾게 되는데 SESSA에서는 SUB-RULE이 이 기능을 수행하게 된다. 즉 사용자가 통계적 문제의 성격을 파악하는데 필요한 모든 질문에 응답하면 이러한 응답들은 적절한 통계적 방법이나 절차를 포함하고 있는 일련의 SUB-RULE 에 자동적으로 연결되어 주어진 통계적 문제에 대한 적절한 해결 방법을 제시하게 된다. SESSA에 포함된 SUB-RULE 의 한 예를 살펴보면 <표 5>와 같다.

현재 SESSA에 포함된 18 개의 모듈은 각각 하나의 TOP-RULE 과 각 모듈 내에 포함된 통계적 기법의 수에 해당하는 SUB-RULE - 각 모듈 당 4 내지 20 개 - 를 갖고 있다. 이렇게 해서 SESSA의 지식베이스에는 위에서 보여진 형태의 법칙이 대략 200 개 정도 포함되어 있다.

SESSA는 추론기관으로 forward-chaining 전략을 채택하고 있다. 즉 TOP-RULE 중의 하나가 항상 사용자가 선택하는 메뉴와 일치하게 되고 이러한 TOP-

RULE은 통계적인 문제들이 서술되고 있는 동안 해당 메뉴의 질문단계를 주관하게 된다. 통계적인 문제를 묘사하는 여러 가지 사실이나 변수의 값들이 지식베이스에 포함된 SUB-RULE 중의 하나와 일치하게 되면 그때 전문가 시스템은 주어진 문제의 해결방법을 제공하게 된다. SESSA의 지식베이스는 두 개 이상의 법칙이 동시에 만족되는 경우를 피하여 구성되었으므로 SESSA의 추론기관은 Conflict Resolutions 을 고려하고 있지 않다.

SESSA 에 포함되어 통계적 계산절차를 수행하는 모든 프로그램은 BASIC을 이용하여 작성되었으며 이들은 SUB-RULE의 명령에 의해 호출되어 필요한 통계적 계산을 수행하게 된다. 현재 SESSA 에 포함되어 통계적 계산을 수행하는 프로그램의 수는 약 180 개 정도가 된다. BASIC으로 작성되어 SESSA에 포함된 모든 통계적 서브루틴은 기존의 통계 패키지인 SPSS 나 SAS를 이용하여 검증절차를 완료하였다.

<표 5> SESSA에 포함된 SUB-RULE의 예

If,	<p>the number of populations(samples) is 1 and the parameter to be estimated is mean</p>
조건절	<p>and the variance is not known and the number of observations of the parameter is greater than 120 and the normality assumption is appropriate</p>
Then,	<p>display the following: Use the following formula</p>
	$\bar{X} \pm t_{1-\alpha/2, n-1} * s / \sqrt{n}$
실행절	<p>where, s = sample standard deviation \bar{X} = sample mean n = number of observations $t_{1-\alpha/2, n-1}$ = tabular value of the Student's-t distribution n-1 degrees of freedom at the 1-α probability level α = type I error rate</p>

3. SESSA의 사용 예

SESSA가 어떻게 시뮬레이션 분석과정에서 나타난 통계적 문제에 대하여 자문을 제공하게 되는가를 예시하기 위해 SESSA를 사용한 예를 들어보자. 가령 시뮬레이션 분석가가 제조공정에 관한 시뮬레이션 모델

을 개발하고 이 모델을 통해 제조공정에 사용되고 있는 특정한 장비에 대한 이용효율을 측정하기 위해 일정한 오차 내에서 필요한 시뮬레이션의 반복횟수를 찾으려 한다고 가정하자.

이러한 문제에 대한 해답을 구하기 위해 분석자는 SESSA의 주 메뉴로 들어가서 "Determining run length

(표 6) Determining Run Length and Number of Replications 에 대한 튜토리얼

Sample size denotes the number of observations of a particular variable collected during a simulation run. The sample size exerts an important effect upon the statistical reliability of estimates of the response variables of the simulated system. Reliability may be measured by the standard deviation of the mean response of the simulated system and can be improved by increasing sample size. But taking more samples usually requires additional cost and time. The choice of an appropriate sample size to achieve a desired level of statistical reliability and to minimize the cost of collecting observations in a simulated system is extremely difficult yet very important.

There are a number of statistical methods to estimate sample size precisely. Most of these are derived from the confidence interval approach. The assumptions required for using confidence interval methods are that the response variable be normally distributed and independent. If the distribution of the response variable violates either of the required assumptions, a confidence interval approach can still be used with a few modifications. For example, if the response variable is not normally distributed, Tchebycheff's theorem can be applied to determine the required sample size. If the response variable is autocorrelated(dependent), a confidence interval approach can be taken with some modifications on the estimate of the variance. To apply a statistical method to determine sample size, the user must be able to specify the following:

1. How large a shift in the response variable(parameter) needs to be detected? Or, how precise an estimate of the population parameter is desired?
2. How much variability is present in the population?
3. What size risks can be taken?
4. What is the type of the distribution of the response variable?
5. Are the response variables autocorrelated?

Sample size can be determined either prior to executing the simulation model or during execution of the model based upon the results generated. The former method is called fixed size sampling because the sample size is fixed prior to the operation and the latter method is called sequential sampling because the sample size is determined on the basis of the precision obtained.

Three different confidence interval procedures-estimation, hypothesis testing, and selection-are included in the expert system. The first question a user is asked by the expert system relates to the purpose of the simulation analysis. If the simulation model is to be used for estimating parameters of a response variable distribution, "estimation" should be selected as the correct confidence interval approach. If the simulation model is designed for the purpose of comparing the value of a parameter with that of a standard or detecting any differences between two alternative systems, "hypothesis testing" should be selected. If the simulation model is intended to evaluate the parameters among three or more alternative systems and to identify the highest(or lowest) value of the parameters, "selection" should be selected.

The second question asked by the expert system is about the type of sampling procedure to be used. This question will be answered with either fixed sampling or sequential sampling. In most cases the answer will be fixed sampling unless collecting samples is very expensive.

The next two questions are asked to verify the assumptions of normality and independence.

The remaining questions asked by the expert system will differ depending on the nature of the problem being investigated by the user. In general, the expert system is attempting to elicit information concerning the number of populations being analyzed, the types of distributions being analyzed, the types of distributions involved, and the specific parameters to be analyzed.

If a series of questions is answered completely, the expert system recommends an appropriate formula to determine the required sample size. If a set of data is submitted, the expert system will perform required calculations.

and number of replications”이라는 열 한 번째 메뉴를 선택하게 된다. 먼저 해당 메뉴에 대한 설명(튜토리얼)이 <표 6>과 같이 한 화면씩 모니터 위에 표시된다.

분석가가 일단 해당 튜토리얼을 이해하고 끝내면

해당 메뉴에 대한 질문단계로 접어들게 된다. 일련의 질문들이 순서에 의해 나타나고 또한 이에 대한 응답이 <표 7>의 예와 같이 주어지게 된다. 모든 질문에 대한 응답이 끝나면 SESSA는 사용자의 요구에 맞는 적절한 공식을 지식베이스에서 찾아내어 화면에 보여

<표 7> Determining Run Length and Number of Replications 을 위한 사용자 문답

<질문 1> What type of problem do you have?

1. Estimation
2. Hypothesis testing
3. Selection

Select a number ⇒ 1

<질문 2> What kind of sampling procedure is appropriate?

1. Fixed sampling size
2. Sequential sampling

Select a number ⇒ 1

<질문 3> Is the normality assumption appropriate or is the number of observations of the parameter greater than 30?

1. Yes
2. No

Select a number ⇒ 1

<질문 4> Is the independence assumption appropriate?

1. Yes
2. No

Select a number ⇒ 1

<질문 5> Can the required precision for an estimate of the parameter be expressed in terms of a multiplier of the standard deviation?

1. Yes
2. No

Select a number ⇒ 2

<질문 6> Is the variance of the population known?

1. Yes
2. No

Select a number ⇒ 2

<질문 7> Is the number of observations of the parameter greater than 30?

1. Yes
2. No

Select a number ⇒ 2

* 밑줄이 그어진 내용은 사용자의 응답에 속하는 것임

준다(〈표 8〉 참조.) 필요한 반복횟수를 결정하기 위하여 SESSA는 α (제 1 종 오차)나, 표준편차, 그리고 오차의 한계 등과 같은 매개변수에 대한 값들을 요구하게 된다(〈표 9〉 참조.) 이들 값들이 명령에 따라 순서대로 입력되면 SESSA는 〈표 10〉에서와 같이 필요한 반복횟수를 계산하여 그 값을 화면에 표시하게 된다.

〈표 8〉 Determining Run Length and Number of Replications 에 대한 SESSA 의 자문

First estimate N. Then use the following formula:

$$N \left(\frac{s}{d} \cdot t_{1-\alpha, N-1} \right)^2$$

where,
 $t_{1-\alpha, N-1}$ = tabular value of the Student's-t distribution with N-1 degrees of freedom at the 1- α probability level
 s = sample standard deviation
 d = allowable error
 α = type I error rate

〈표 9〉 Determining Run Length and Number of Replications 의 입력자료를 얻기 위한 질문

Do you want the expert system to calculate the number of replications?
 Enter Y or N : Y
 Enter a value for α : .05
 Enter a value for d : 9
 Enter a value for s : 12.83811
 Estimate of N : 30

〈표 10〉 주어진 입력자료에 대해 SESSA가 결정한 반복횟수

```
***** A N S W E R *****
*
*      Number of Replications = 9      *
*
*****
```

SESSA 에 포함된 18 개의 모듈 모두가 위의 예와 비슷한 형태로 작동하게 된다. 자료입력을 제외한 모든 사용자 응답은 1,2,3 등의 숫자나 예/아니오 형태의 대답이 된다. 아마 SESSA를 처음 사용하는 경우 SESSA가 제시하는 질문에 대한 대답을 완성해서 원하는 통계적 문제에 대한 자문을 얻어내는데 20-30분 이상 소요되겠지만 익숙해지면 튜토리얼 부분은 생략할 수 있으므로 자료입력을 포함해서 5 분 이내에 원하는 자문을 얻어낼 수 있게 된다.

지금까지 SESSA 에 대한 현장의 시뮬레이션 실무자들의 반응은 SESSA가 제공하는 통계적 자문에 큰 관심을 보였고 또 SESSA가 의외로 사용하기 쉽다는 것이었다. 하지만 SESSA가 해결할 수 있는 통계적 문제의 범위는 SESSA의 지식베이스 내에 포함된 문제에 국한될 수 밖에 없음을 인정하지 않을 수 없다.

4. 결론

본 연구에서 시뮬레이션 기법에 내재된 통계적인 쟁점들이 나열되고 시뮬레이션 분석과정에서 요구되는 통계적 분석을 수행함에 있어 시뮬레이션 분석가를 도울 수 있는 전문가 시스템이 개발되었다. 전체적으로 시뮬레이션 분석단계를 따라 자주 제기되는 18 개의 통계적 쟁점들이 나열되어 SESSA의 지식베이스 내에 포함됨으로서 시뮬레이션 분석단계에서 필요로 하는 통계적 자문을 얻기 위해 전문가 시스템이 개발, 이용될 수 있음을 제시하였다. 더욱이 SESSA는 대화적 접근방식을 통해 매우 상세하게 통계적 자문을 제공함으로써 기본적인 통계적 지식만 갖추고 있으면 누구나 손쉽게 이용할 수 있음을 보여 주었다.

SESSA와 같은 전문가 시스템이 시뮬레이션 분야에서 널리 이용될 때 나타날 수 있는 효과나 장점은 여러 가지가 있을 수 있다. 그 중에서 가장 중요한 것은 전문가시스템을 통해 시뮬레이션 분석과정에 적절한 통계적 분석의 이용이 가능하게 되어 해당 시뮬레이션의 분석결과에 대한 신뢰도를 향상시킬 수 있다는 것이다. 두번째로 전문가 시스템을 이용함으로써 일반 통계 패키지를 이용하는 것보다 시뮬레이션과 연관된 통계적 분석을 수행하는데 소요되는 시간을 5

배 내지 10 배 정도로 단축시킬 수 있고 마지막으로 전문가 시스템은 시뮬레이션에 대한 교육이나 현장실습에 있어 매우 효율적인 교육 및 훈련도구로서의 기능을 수행할 수 있을 것이다.

참고문헌

- [1] Forgiogne, Guisseppi A., "Corporate Management Science Activities-Update," *Interfaces* 13, no. 3, PP. 20-23, June 1983.
- [2] Forgy, Charles L., *The OPS83 User's Manual*, Production System Technologies Inc., 1985.
- [3] Gale, W.A. and Pregibon, D., "An Expert System for Regression Analysis," *Proceedings of the 14th Symposium on the Interface*, PP.110-117, 1982.
- [4] Haaland, P.D., Yen, D., and Liddle, R.F., "Expert System for The Design of Screening Experiments," *Statistical Computing Section Proceedings of The American Statistical Association*, PP.223-228, 1985.
- [5] Hahn, Gerald J., "More Intelligent Statistical Software and Statistical Expert System: Future Directions," *The American Statistician* 39, no.1, PP.1-16, February 1985.
- [6] Hakong, L. and Hickman, E. R., "Expert System Techniques: An Application in Statistics," *Expert Systems 85: Proceedings of the Fifth Technical Conference of the British Computer Society, Specialist Group on Expert Systems*, University of Warwick, Cambridge University Press, NY, PP. 43-63, 17-19 December 1985.
- [7] Hand, D.J., "Statistical Expert System: Necessary Attributes," *Journal of Applied Statistics*, 12, no. 1, PP. 19-27, January 1985.
- [8] Law A.M. and Kelton, W.D., *Simulation Modeling and Analysis*, Mcgraw Hill, New York, 1982.
- [9] Mellichamp, Joseph M. and Park, Young Hong, "A Statistical Expert System for Simulation," *Simulation* 52, no.4, PP.124-139, April, 1989.
- [10] O'keefe, Robert, "Simulation and Expert Systems- A Taxonomy and Some Examples," *Simulation* 46, no. 1, PP.10-15, January 1986.
- [11] Portier, K.M. and Lai, P.Y., "A Statistical Expert System for Analysis Determination," *Proceedings of American Statistical Association, Computing Section*, PP.309-311, 1983.
- [12] Shannon, Robert E., Mayer, Richard, and Adelsberger, Heimo H., "Expert Systems and Simulation," *Simulation* 44, no. 6, PP.275-284, June 1985.
- [13] Thomas, G. and DaCosta, Jo-Anne, "A Sample Survey of Corporate Operations Research," *Interfaces* 9, no. 4, PP.104, August 1979.



박영홍(朴泳弘)
 1954년 8월 7일생
 1979년 강원대학교 경영학과 졸업
 1983년 미국 아칸소대 대학원 졸업(석사: 자료처리전공)
 1988년 미국 알라바마대학 졸업(박사: 경영과학전공)
 현재 관동대학교 조교수