

韓國統計의 現況과 將來—統計와 電算

許 文 烈*

통계적 방법론이 실용성(적용성)을 상실할 때 수학과 다른 점이 없을 것이고 방법론이 실용성을 가진 훌륭한 이론이라 할 지라도 계산이 불가능하다면 이는 호화로운 자동차를 가지고 있으면서 연료가 없어 운전치 못하는 것과 다를 바가 없겠다. 이제 한국 통계 학회가 창립 10주년을 맞아 발돋움 단계에 접어들 시점에서 그 동안 특히 부진하고 미비했던 전산통계(혹은 계산 통계 statistical computing) 분야에 대해 통계인들 사이에 과거, 현재와 그리고 미래의 나아갈 방향에 대해 의견을 교환하게 됨은 뜻 깊은 의의가 있다고 하겠다.

I. EDP의 발전과 통계의 역할

Electronic data processing (EDP)는 1960년대에 접어들면서 급격한 발전을 가져 왔다. 대개 EDP가 발전하는 단계를 4단계로 보고 있으나(Gibson, *et al*) 필자는 이를 3단계로 구분하여 단계별 통계의 역할에 대해 기술코자 한다.

초기 단계 :

이는 대개가 단순 반복 작업의 컴퓨터화로써 이 단계에서는 복잡한 통계 방법론 보다는 단순 통계—즉 집계를 필요로 하는 통계가 요구된다. 예를 들면 현재 문교부에서 주관하는 예비고사의 컴퓨터 처리와 여기에 부수되는 점수별, 지역별 각종 통계이다. 이 단계에서는 통계의 전문적 지식 보다는 단순 반복적 숙련공의 인력이 필요하다고 보겠으며 이 단계에서는 대부분 컴퓨터의 경제성이 인정 받게되고 따라서 좀 더 복잡한 일에 컴퓨터화를 시도하게 된다.

성장 단계 :

초기 단계에서 인정을 받은 컴퓨터의 위력은 이제 해당 기관의 전략적 문제를 해결하는데 서서히 적용되기 시작하게 된다. 이 변천 과정을 필자는 5~10년의 기간이 소요된다고 추정한다. 성장 과정에서의 대표적인 업무는 재고 관리라든가 판매량 추정 등이 컴퓨터화하며 따라서 각종 Data들이 수집되기 시작한다. 즉 Data Base의 필요성이 대두되고 따라

* 成均館大 統計學科

서 Data의 효율적인 집약, 분석 방법이 요구된다. 이 단계에서는 방법론의 전문 지식을 어느정도 갖춘 훈련된 요원이 요구되며 따라서 EDP의 경제적 효과를 측정하기가 상당히 주관적이라고 할 수있다. 이 단계를 성공시킨다면 상당한 수준에 도달하게되고 차츰 기술적 측면에 컴퓨터의 응용을 생각하게 된다. 이 변천 단계 또한 5~10년의 기간이 요구된다 하겠다.

성숙 단계 :

성장 단계에서 인정을 받게된 EDP의 활약이 이제 그 기관의 기술적 측면에 까지 관계하게 된다. 예를 들어 Wargame simulation을 통한 국방 정책의 결정이라던가 의사 진료를 컴퓨터화하는 문제라던가 하는 고도의 의사 결정 과정에 컴퓨터가 차지하는 비중이 막대해진다. 이 단계에 이르면 각종 Data들이 통신망을 통하여 real-time으로 집결 분석된다. 따라서 이 단계에서는 상당한 수준에 이르는 통계 전문가가 필수적 요원으로 등장하게 된다. 현재 서구의 몇몇 나라들이 이 성숙 단계에 접어들기 시작함으로써 전문 통계 요원(대학원 이상의 교육을 받은 인원)이 급격히 요구되고있는 실정이다.

II. 현재의 실정

필자가 정확하게 조사를 해 보지는 못 했으나 그간의 경험에 비추어 볼 때 우리나라의 대부분 기관(정부 기관에서 개인 기업에 이르기 까지)이 초기 단계와 성장 단계의 사이에 머무르고있는 것으로 판단된다. 따라서 아직 전산 통계 전문 요원에 대한 필요성을 절실히 느끼지 못한다고 하겠다. 그러나 필자의 추정에 의하면 곧 닥아 올 성장 단계와 또 성숙 단계를 내다 볼 때 통계와 전산, 전산과 통계를 동시에 겸비한 전문 요원은 그 수요가 눈에 명확하다고 하겠다. 그러면 전산 통계란 어떤 내용인가에 대해 고찰해 보기로 하자.

III. 전산 통계(statistical computing)

전산 통계, 혹은 계산 통계를 기술하기 전에 통계 분석의 절차를 고려하여 보면 다음과 같은 그림으로 표시할 수 있겠다.

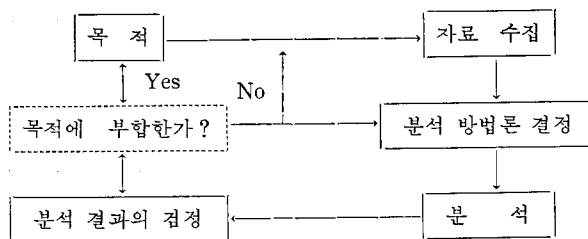


그림 1. 통계 분석 절차의 모형도

윗 그림을 고찰하면 컴퓨터가 통계 분석에 어떻게 이용되어야 할 것인가가 명확히 나타내겠다. 즉 자료 수집 단계에서 data의 효율적 집약 방법과 집약된 data의 정리—그라프,

도표등을 이용한 data display, sort-merge 등을 통한 단순 통계(discriptive statistics)의 작성—이 이뤄진다. 이 다음에 이뤄져야 하는 단계가 문제에 적합한 분석 방법론의 도입이다. 여기서 흔히 제시되는 점은 각 중 컴퓨터 package이다. 이 package의 문제점과 올바른 이용 방법에 관해서는 많은 학자들의 관심과 토론이 있었으며 Bill schucany는 packaged program의 “TV dinner”화를 비판하였다. 여기서는 간략하게 우리나라의 실정에 관해서만 논하기로 한다. 우선 SPSS, BMDP 그리고 SAS 등 널리 쓰이는 통계 package 등은 그 용량이 방대하므로 우리나라 대부분 대학들이 보유하고 있는 컴퓨터에 설치 운영하기가 어려운 점이다. 따라서 이를 보급 교육시킬 책임이 있는 기관이 거의 없는 상태이다. 따라서 전산 통계 요원이 기존 package에 대한 지식이 부족하므로 새로운 package의 개발, 또는 기존 package의 효율화 등에까지는 아직 생각할 단계가 요원하다고 하겠다.

이제 분석 방법이 결정되고 이를 목적에 적합한가를 검토하는 단계(sensitivity analysis)에 컴퓨터가 이용됨으로써 sensitivity 분석이 여러 형태로 가능케 된다.

위에서 기술한 바와 같이 통계 방법론 자체의 계산 뿐만 아니라 자료분석 과정의 제반 절차에 컴퓨터의 효율적인 적용이 필요하며 이렇듯 상호 연관되는 두 분야를 동시에 연구하는 분야가 전산통계라고 불려지고 있다.

IV. 통계 교육을 위한 컴퓨터

컴퓨터를 이용한 교육(computer aided education-CAE)은 요사이 많은 연구가 이뤄지고 있다(Martin, *et al*). 통계 교육에 컴퓨터의 이용은 CAE 뿐만 아니라 이론과 실제를 연결시킬 수 있는 도구로써 큰 기여를 하게 된다. 이에 대해서도 많은 학자들이 관심을 보였다(Thisted, *et al*). 특히 선형 모델(linear model), 다변량 분석(multivariate analysis), Robust estimation 등의 분야에서는 컴퓨터를 이용하지 않고서는 계산하기에 너무 복잡하여 컴퓨터 algorithm을 동시에 교육하여야 효과적이 되겠다. 또한 수리 통계등의 이론 분야에서도 컴퓨터에 의한 시뮬레이션 기법을 교과 과정에 결들인다면 교육효과의 극대화가 이뤄질 것이다.

V. 앞으로의 제언

현재의 각 교육, 연구기관이 가지고 있는 컴퓨터의 능력이 제한되어 있거나, 관심있는 통계인들이 사용하기에 시간과 비용이 너무 많이 드는 실정이므로 정책적 차원에서 정부의 지원이 이뤄져야 하겠다. 즉 연구와 학문을 위한 대형 컴퓨터를 설치하고 여기에 각종 소프트웨어(software)를 보유시키고 각 기관에는 단말 장치를 통하여 학문과 교육의 목적으로 부담을 느끼지 않고 사용할 수 있도록 하여야 하겠다. 그럼으로써 컴퓨터의 산발적 도입으로 인한 인력과 경비의 낭비를 방지하게되고 각 교육기관과 실무기관의 정보유통이 원활하게 되겠다. 또한 이러한 시스템을 위하여 일차적으로 전산 통계에 관심이 있는 학자, 실무 종사자들 간에 그룹을 만들어 이를 통계 학회 내에 소분과위원회 형식으로 설치하기를 제안한다.

끝으로 통계가 그 사회에서 필요성을 인정 받는다면 그 사회는 합리적인 제도에 의해 움직인다는 점을 감안 할 때 통계인 모두 합리적 사회 건설에 합심 노력해야 하겠다.

참고 문헌

- (1) Dixon, W.J., ed (1975), BMDP, *Biomedical Computer Programs*, Los Angeles; University of California Press.
- (2) Gibson, Cyrus F. and Richard R. Nolan (1974), "Managing the four stages of EDP growth," Jan Feb., *Harvard Business Review*
- (3) Martin, James and Adrian R.D. Norman (1970), *The Computerized Society*, chapt, 6 and 23, prentice Hall.
- (4) Nie, Norman and *et al*, (1975), SSPS, *Statistical Package for the Social Sciences*, 2nd ed., Mc Graw-Hill.
- (5) SAS User's Guide (1979 ed.), Statistical Analysis System, SAS Institute.
- (6) Schucany, William R. (1978), Comment on the invited papers, "A Review of the manuals for BMDP and SPSS" by Kenreth N. Berk, *et al*, March, *Journal of the American Statistical Association*.
- (7) Thisted, Ronald A. (1979), "Teaching statistical Computing by computer packages," Feb., *The American Statistician*. Also comments by DuMouchel, *et al*.