

# 지식기반 사회에서의 통계학 패러다임의 변화와 데이터 기술의 발전

박 성 현\*

《目 次》

- |                      |                    |
|----------------------|--------------------|
| I. 통계학의 발전과정         | IV. 데이터 기술(DT)의 발전 |
| II. 우리나라 통계학의 현황     | V. 우수한 조직과 데이터 기술  |
| III. 통계학의 새로운 역할과 도전 |                    |

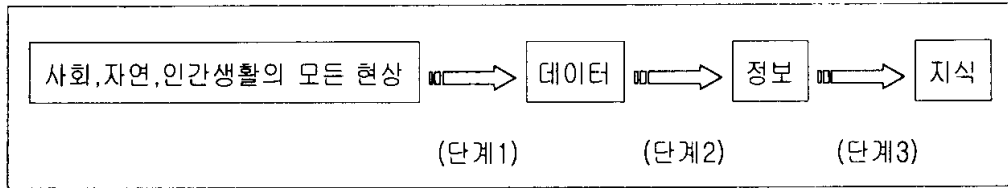
## I. 통계학의 발전과정

통계학(Statistics)은 19세기에 영국을 중심으로 유럽에서 발전한 학문으로, 원래 국가산술(state arithmetic)이란 어원에서 비롯된 학문이다. 국가산술은 정치인들이 국가의 살림을 꾸려나가기 위하여 필요한 숫자 자료를 체계적으로, 과학적으로 산출해 내기 위하여 필요한 수리과학적 분야이다. 대부분의 학문이 그렇듯이 통계학도 역사적으로 발전과 변모를 거듭하고 있다. 19세기 후반에 통계학은 국가산술(정부통계)의 영역을 벗어나 수학자들에 의하여 확률, 확률분포 등을 연구하는 응용수학으로 발전하였다. 20세기 초반이 되면서 모집단과 표본의 개념이 도입되고, 표본 데이터의 정보로부터 모집단의 성질에 관한 검정과 추정을 실시하는 추측학문으로 발전되었다. 20세기 후반에 이르면서 통계학은 그 범위가 매우 넓어져서 의사결정과학으로 발전하였다. 즉, 통계학은 사회, 자연, 인간 생활 등의 온갖 현상을 연구하기 위하여 불확실성이 내포된 자료의 선택, 관찰, 분석, 검정, 추정 등을 통하여 의사결정에 필요한 자료의 획득과 처리방법을 연구하는 학문으로 발전하였다.

21세기에 접어들면서 통계학은 또 한번 새로운 패러다임으로 변화하고 있다. 21세기를 지식기반 정보화 사회라고 한다. 통계학은 데이터로부터 정보를 추출하여 내고, 정보를 적절히 편집, 가공함으로써 지식을 창출하여 내는 정보과학(information science)으로 발전하고 있다. 따라서 통계학은 지식기반 정보화 사회에 필수적인 학문으로 나타나고 있는 것이다. 통

\* 서울대 자연과학대학 통계학과 교수

〈그림 1〉 지식의 생성 과정



계학과 관련하여 지식의 생성과정을 보면 〈그림 1〉과 같다.

위의 각 단계에서 통계학이 하는 역할은 지대하다. 단계 1에서는 유용한 데이터 수집을 위하여 실험계획법, 표본설계 등의 과학적인 데이터 수집계획이 연구 발전되고 있다. 단계 2에서는 데이터로부터 유용한 정보를 다량으로 구하기 위하여 수없이 많은 통계적 분석 방법들(각종의 검정과 추정, 상관분석, 회귀분석, 분산분석, 다변량분석, 비모수 통계분석 등)이 사용되고 있다. 단계 3에서는 얻어진 정보들로부터 유용한 지식을 창출하기 위하여 통계데이터 베이스 활용, 과학기술에 접목시킨 모형의 개발, 시뮬레이션에 의한 가상 지식의 획득 등 여러 가지 연구가 이루어지고 있다. 이러한 발전과정에는 정보기술(IT)의 발전이 큰 몫을 담당하고 있다. 특히 컴퓨터와 통신기술의 발전은 다량 데이터의 수집, 저장, 분석, 해석, 전송 등을 가능하게 하고 있으며, 현대통계학은 IT의 발전과 깊은 관계를 가지고 있다. 즉, 통계학은 IT의 발전과 더불어 정보과학의 핵심을 이루는 학문으로 발전하고 있으며, 지식기반 정보화사회에서 가장 중요한 인프라를 이루는 기초과학으로 자리 매김하고 있다. 통계학 및 국가통계의 발전방안 등에 대해서는 박성현(참고문헌 1, 2)을 참조하여 주기 바란다.

## II. 우리나라 통계학의 현황

우리나라에서 통계학의 도입은 60년대 중반에 고려대, 연세대, 성균관대, 중앙대 등에 통계학과를 설치하면서부터 시작되었다. 그러나 이들 대학이 통계학과를 상경계열대학에 설치하면서 경제·사회 통계를 강조하여 통계학의 전반적인 발전에는 한계를 드러내기도 하였다. 1971년에는 한국통계학회가 창립되어 통계학 연구가 이루어지기 시작하였다. 통계학과와 더불어 확대되기 시작한 것은 1975년에 국립대학인 서울대에 통계학과가 설치되면서부터라고 볼 수 있다. 이후 대부분의 국립대학의 자연과학대학(또는 이과대학)에 설치되면서 의사결정 과학의 성격을 갖는 기초과학으로 발전하게 되었다.

현재 전국에는 70여개 대학에 통계관련학과가 설치되어 있으며, 매년 3,000명 정도의 통

계인력을 배출시키고 있다. 통계관련학과의 이름도 다양하여 통계학과, 응용통계학과, 정보통계학과, 전산통계학과, 데이터 정보과학과, 통계정보학과 등이 사용되고 있으며, 이는 통계학이 정보과학으로 발전되고 있음을 보여주고 있는 증거라고 하겠다.

### Ⅲ. 통계학의 새로운 역할과 도전

통계학이 의사결정과학 그리고 데이터에 근거한 정보과학으로 발전하여 나가면서 새로운 역할이 증대되고 있으며, 또한 새로운 도전을 맞고 있다. 첫째로, 다른 학문의 계량적 발전에 기여하고 있다. 모든 학문은 그의 가설정립에 필요한 데이터 분석이 필요하며, 이러한 계량 분석기법은 기존의 학문과 접목되어 새로운 학문분야를 탄생시키고 있다. 예를 들면, 경제학과 접목된 계량경제학(econometrics), 환경과학과 접목된 환경통계학(envirometrics), 생물학·의학 등과 접목된 생물통계학(biometrics), 품질관리와 접목된 통계적 품질관리(statistical quality control), 심리학과 접목된 계량심리학(psychometrics) 등, 수없이 많은 분야가 탄생되어 각 학문분야를 발전시키고 있다. 또한 컴퓨터의 발전과 더불어 모의 통계 데이터를 발생시켜 연구하는 모의실험 연구는 모든 학문분야에 새로운 깊이를 제공하고 있다.

두 번째로, 인류복지 증진과 공업발전에 기여하고 있다. 예를 들어, 생물통계학과 깊은 관련이 있는 생물정보학(bioinformatics)은 생물학, 의학, 통계학, 수학, 전산과학 등이 어우러져 컴퓨터를 활용하여 생명과학 관련 정보를 규명하여내는 새로운 학문으로, DNA의 규명, 각종 질병의 인과관계 분석, 지놈의 실체를 밝히는 유전체학(genomics), 단백질의 역할을 규명하는 단백질체학(proteomics) 등의 연구를 포함하고 있다. 생물정보학은 인류복지에 큰 영향을 주리라고 예상하고 있으며, 통계학이 생물정보학에서 핵심적 역할을 담당하게 될 것이다. 또한 통계적 품질관리에서 시작된 공업통계학 분야는 최근에 품질과 생산성 향상에 기여하는 핵심 기초학문으로 발전하여 공업발전에 기여하고 있다. 과학적 기업경영 기법으로 각광을 받고 있는 전사적 품질경영(Total Quality Management), 식스 시그마(Six Sigma), 고객관계경영(Customer Relations Management), 품질공학(Quality Engineering) 등은 모두 통계적 정보분석에 바탕을 두고 있다.

세 번째로, 국가운영의 투명성과 선진화에 기여하고 있다. IT의 발전과 더불어 표본설계기법, 조사방법, 통계데이터베이스, 인터넷의 발전 등은 공식통계, 정부통계 등의 신뢰성을 높이고 사용자에게 그 보급을 원활하게 하고 있으며, 국민이 필요한 실생활 관련 자료를 신속 정확히 공개하여 국가운영의 실체를 국민에게 알리고 진정한 국민을 위한 국가발전을 촉진시

키고 있다. 더 나아가서 세계 각 국 및 국제기구(OECD, UN, IMF, UN 등)와 통계정보를 교환함으로써 세계화에도 크게 기여하고 있다.

네 번째로, 지식기반 정보화사회에 필요한 정보 인프라구축에 핵심적으로 기여하고 있다. 이는 앞에서 상세히 설명한 내용으로, 통계학은 데이터에 근거한 정보획득 과정과 지식창출을 연구하는 새로운 패러다임의 학문으로 발전하고 있다.

#### IV. 데이터 기술(DT)의 발전

최근 첨단과학기술 분야로 주목을 받고 있는 것은 IT, BT(생명공학), NT(나노기술), ST(항공우주기술), ET(환경기술), CT(문화기술) 등이다. 지식기반 정보화 사회에서 첨단과학기술 분야로 반드시 끼여야 할 분야로 필자는 데이터로부터 시작되는 데이터 기술(Data Technology: DT)을 꼽고 싶다. DT란 약어는 세상에서 사용되는 용어는 아니지만 앞으로 사용될 것이라고 필자는 확신한다. DT는 통계학 패러다임의 변화와 더불어 통계적 방법으로부터 지원을 받는 새로운 과학기술 분야이다.

DT란 데이터의 측정, 수집, 축적 기술에서부터 시작하여, 데이터의 분석 및 해석 능력, 데이터로부터의 모형화 기술, 데이터로부터의 미래 예측기술을 다루는 과학적 방법론을 말한다. DT는 소프트웨어의 구축과 조직의 인프라를 주로 다루고 있으므로, 그 진행과 결과가 눈에 잘 띄지 않으며, 보통 간과하기 쉽다. 그러나 국가 선진화를 위해서는 DT는 필수적인 요소라고 생각한다. 우리의 산업구조는 DT에 매우 취약하며, 조만간 크게 보완되지 못하면 국가 선진화에 큰 장애 요인이 될 것이다. DT의 미비로 발생하는 몇 가지 손실 사례를 들어 보자.

첫째, 우리나라는 1997년 외환위기 때, 외환보유고를 포함한 각종의 경제지표의 변화를 소홀히 생각하는 가운데 스스로 위기를 자초하게 되었다. 외환보유고라는 데이터의 적절한 수집, 분석 및 예측이 있었다면 IMF 위기를 사전에 대비할 수 있었을 것이다. 둘째, 우리나라 제조기업에서 불량품 발생 등으로 인한 품질비용은 매출액의 20-30% 수준에 이른다는 보고가 있다. 이 품질비용 중에서 예방비용, 평가비용, 내부실패비용, 외부실패비용은 각각 어느 정도인지 객관적 데이터로 평가한 후에, 필요한 데이터의 수집, 분석, 평가, 예측을 통하여 품질비용은 최소화하는 방안을 강구하고 실행한다면, 품질비용은 매출액 대비 10% 수준으로 충분히 낮출 수 있다고 한다. 그러나 우리 기업의 대부분은 아직도 품질비용을 제대로 계산하지 못하고 있다. DT의 적절한 활용은 적자기업을 흑자기업으로 바꾸는 중요한 원인이 될

수 있다. 또 다른 예로, 최근 정계에서 '교원정년 연장안'을 놓고 여당과 야당이 힘 겨루기를 하고 있다. 각 당의 주장은 국민여론은 자기편이며, 자기의 정책이 옳다는 것이다. 만약 DT의 발달과 더불어 국민여론을 정확히 추정할 수 있는 여론조사를 할 수 있다면, 불필요한 힘 겨루기와 소모적인 정쟁은 피할 수 있을 것이다. 이처럼 DT의 적절한 활용은 눈에 안 보이는 막대한 손실비용을 감소시킬 수 있을 것이다.

앞에서 언급된 IT, BT, NT, ST, ET 등은 대부분 눈에 보이는 하드웨어적인 기술과 결과물을 생산해 낸다. 그러나 DT는 눈에 잘 보이지 않으며 밑에 깔려 있는 소프트웨어적인 인프라에 해당하므로 보통 때 무시하기 쉽다. 그러나 데이터에 의하여 현상을 정확히 파악하지 못하고, 문제점을 객관적 데이터에 의하여 찾아내지 못하고, 앞으로 발생할 현상을 수리적 모형을 사용하여 예측할 수 없다면, 위에 열거된 첨단과학기술의 발전에 한계가 있을 수밖에 없다. 즉, DT는 모든 첨단과학기술의 기초에 해당하는 원천적 기초과학 기술인 것이다.

DT의 발전과 밀접한 관계가 있는 학문은 통계학, 응용수학, 전산과학, 정보과학, 산업공학, 경영과학 등이다. 통계학에서 다루는 표본설계, 여론조사, 통계적 공정관리, 시계열분석, 데이터 마인닝 등이 관계가 있고, 응용수학에서 다루는 암호수학, 금융수학, 시뮬레이션 기법 등이 관계가 있다. 전산이나 정보과학에서는 소프트웨어 공학, 뉴럴넷웍 등이 관계가 있으며, 산업공학과 경영과학에서는 품질경영, 생산관리, 시스템 공학적인 접근방법 등이 관계가 있다. 특별히 DT의 발전은 국가 소프트웨어의 발전과 IT의 발전에 심대한 영향을 준다.

우리나라의 주력 수출제품을 보면 반도체, 조선, 자동차, 휴대폰 등 눈에 보이는 제품이 대부분이다. 눈에 잘 보이지 않는 소프트한 제품들은 국제 경쟁력이 미약하다. 예를 들면, 통계분석용 소프트웨어는 SAS, SPSS, Minitab 등 미국제품이 국내 시장을 석권하고 있으며, 이로 인한 외화의 유출은 엄청나다. 심지어 반도체, 조선 등에 사용되는 공정관리용 소프트웨어도 외국제품이다. 우리나라에 수없이 많이 들어와 있는 외국 컨설팅 회사들이 주로 하는 일이 사실상 DT와 관련된 경영자문이 상당부분에 해당한다. 이제 우리산업도 DT에 더욱 눈을 떠서 고부가가치 산업에 투자할 때이다.

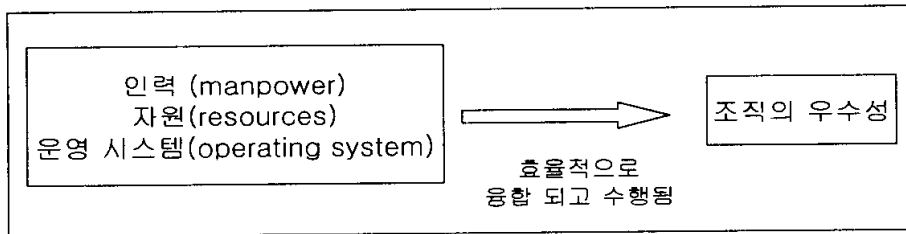
21세기를 지식기반 정보화 사회라고 말하며, 창의적 지식창출이 국가경쟁력의 원천이라고 말한다. 지식의 창출과정을 살펴보면 다음과 같은 과정을 거친다. 먼저 우리 주위의 현상을 정확히 파악할 수 있는 계량화된 데이터가 필요하다. 다음으로 데이터로부터 분석 및 해석을 통하여 정보를 얻는다. 이 정보들을 여러가지 형태로 가공하여 필요한 지식을 얻게 된다. 이러한 지식창출과정은 DT의 도움 없으면 불가능하다. 따라서 DT의 발전은 21세기에 주요 산업국가를 꿈꾸는 우리나라에게 매우 중요하며, 국가 선진화에 핵심적인 요소가 될 것이다.

## V. 우수한 조직과 데이터 기술

우수한 조직의 운영과 데이터 기술과의 관계를 살펴보자. 여기서 조직은 정부, 공공기업, 민간기업, 협회, 대학 등 모든 단체가 포함된다. 우리나라는 개인들은 매우 똑똑하고 영리하나 단체가 취약하다는 말을 가끔 한다. 이는 조직의 품질이 우수하지 못하다는 의미이다. 한 조직(organization)의 우수성이란 이 조직이 가지고 있는 인력과 자원과 운영 시스템이 얼마나 효율적으로 융합되고 수행되어 이 조직이 추구하는 목적을 달성하고 있는가로 판단된다.

우수한 조직은 인력을 100% 잘 활용하고 있고, 구성원의 사기가 높다. 그리고 사용하는 자원이 낭비 없이 목적을 추구하는데 사용되고 있고, 또한 운영 시스템이 간단하고 명료하여 시비 거리가 적다. 이러한 인력, 자원, 운영 시스템이 효율적으로 융합되고 수행될 때 우수한 조직은 유지되고 발전되는 것이다. 이 관계를 간단히 그려보면<그림 2>와 같다.

<그림 2> 우수한 조직의 관계



그러면 이러한 인력, 자원, 운영 시스템의 3 가지가 효율적으로 융합되고 수행되기 위해서는 무엇이 필요한가? 바로 데이터 기술인 것이다. 이 3 가지 요소가 제대로 융합되어 효율성을 가지는가를 모니터링하고 평가하여 주며, 문제가 있을 시에 순발력 있게 조정하여 주기 위해서는, 필요한 데이터의 수집, 분석, 해석, 예측 등이 적기 적소에서 빠르게 이루어져야 가능한 것이다. 데이터 기술은 이러한 막중한 역할을 감당하므로, 우리 기업이 국제경쟁력을 가지려면 데이터 기술의 발전에 각별한 노력을 경주하여야 할 것이다. 데이터 기술의 경제성에 대해서는 박성현(참고문헌 3)을 참조하여 주기 바란다.

### 참 고 문 헌

1. 박성현: "통계학 및 국가통계의 발전 방안", 통계발전 심포지엄 논문집, 한국통계학회 제6회 통계의 날 기념 심포지엄, p. 3-20, 2000. 9.
2. 박성현: "통계의 새로운 지평", 서울경제신문, 오피니언, 2001. 8. 22. 일자.
3. 박성현: "데이터 기술의 경제학", 한국경제신문 다산칼럼, 2001. 12. 3. 일자.