

통계학과에서의 데이터베이스 교육 방안

안정용¹⁾, 한경수²⁾

요 약

통계학과 교육과정의 개선에 관한 연구를 비롯한 몇몇 응용 논문을 통하여 통계학과에서 데이터베이스 교육이 중요하다는 사실은 간혹 언급되어 왔으나, 그 구체적인 교육 방법에 관해 논의한 연구는 찾아보기 힘들다. 본 연구에서는 통계학과에서 데이터베이스 교육의 필요성 및 교육 방안에 관해 생각해보고자 한다. 따라서 본 논문의 목적은 통계학과에서 데이터베이스를 교육할 때 어떻게 통계학이라는 학문과 연관지어 교육할 수 있을 것인가에 대해 살펴보는 것이다.

주요용어 : 통계학과 교육과정, 데이터베이스, 데이터 분석

1. 서론

Friedman(1997)이 지적한 바와 같이 데이터베이스 분야에 대한 통계학자들의 관심은 최근까지도 미미한 실정이며, 통계학과의 정규 교육과정에서도 제외되어 온 것이 사실이다. 물론 통계학과 교육과정의 개선에 관한 연구를 비롯한 몇몇 응용 논문을 통하여 통계학과에서 데이터베이스 교육이 필요하다는 사실은 간혹 언급되어 왔다. 박현진 등(1998)은 통계계산 교과과정을 제언하면서 '전산통계 II' 과목에서 데이터베이스에 대한 내용을 다룰 수 있음을 제안하고 있으며, 조신섭 등(1999)은 정보 관련 통계학과의 교과과정에 '통계 데이터베이스' 과목의 추가를 제안하고 있다. Bryce 등(2001) 또한 데이터베이스 교육의 필요성을 언급하고 있으며, Ritter 등(2001)은 통계학 전공자들의 취업을 하기 위해 요구되는 자질 중 하나로 데이터베이스의 활용 능력을 언급하면서, 통계학의 단편적인 지식뿐만 아니라 종합적인 능력이 필요함을 강조하고 있다.

한편, 최근에 국내의 많은 통계학과들이 정보 관련 학과로 변신을 꾀하면서 데이터베이스를 정규 교육과목으로 추가하고 있다. 이러한 변화는 데이터 수집 및 관리에 대한 관심이 높아져 가고 있음을 반영하고 있다고 하겠다.

과거의 사회와 비교하여 볼 때, 현대 사회에서는 다양한 종류의 데이터를 양산하고 있다. 산업사회에서의 데이터는 인구, 산업 및 제품에 관련된 데이터가 주종을 이루었으나, 현대 사회에서는 컴퓨터 네트워크의 발전으로 전 분야의 데이터가 대량으로 발생하고 있다(물론 이러한 데이터는 우리가 일반적으로 말하는 통계데이터와는 약간 다른 경우가 대부분이다). 데이터의 대용량화는 데이터 관리 및 처리/분석에 있어 기존의 관리 방법과 분석 기법의 적용을 어렵게 만들고 있으며, 데이터의 체계적인 관리와 활용 기법을 요구하고 있다. 데이터베이스는 이러한 요구를 해결하기 위한 적절한 기반을 제공하여 주는 응용 분야이며 데이터와 매우 깊은 관련이 있는 분야이다. 이러한 측면에서 데이터베이스는 전산학 분야뿐만 아니라 통계학, 산업공학, 경영학, 의학 등 데이터와 관련된 많은 학문 분야에서 보다 적극적인 활용이 가능하다. 특히, 테

1) 서남대학교 컴퓨터정보통신학과 조교수, 전북 남원시 광치동 720

2) 전북대학교 수학·통계정보과학부 교수, 전주시 덕진구 덕진동 664-14

통계학과에서의 데이터베이스 교육 방안

데이터베이스 이용 측면만 고려한다면 전산학 분야보다는 오히려 응용 분야에서 데이터베이스 활용에 대한 지식이 더 필요하다고 생각한다. 그 이유는 모든 사람이 한 데이터로부터 같은 정보를 추출하는 것이 아니듯이 전산학 분야와 응용 분야에서 데이터를 바라보는 시각이 분명 다를 것이기 때문이다.

그러나 데이터베이스에 대한 강의는 거의 전산학 분야에서만 이루어져 왔으며, 위에서 언급한 응용 분야들에서 데이터베이스는 거의 다루어지지 않았다. 그 이유는 여러 가지가 있겠지만 다음과 같은 두 가지가 중요한 요인이 아니었나 생각된다. 첫째, 데이터베이스를 강의할 수 있는 인력과 실습 등의 주변 환경적인 여건(하드웨어와 소프트웨어 여건)이 충분하지 못했고, 두 번째로는 데이터베이스의 필요성에 대한 인식의 부족을 들 수 있다. 최근 컴퓨터 환경의 변화는 여러 응용 분야에서도 얼마든지 데이터베이스에 대한 내용을 학습할 수 있도록 도와주고 있으며, 이 분야에 대한 관심이 과거 어느 때보다도 고조되고 있다. 이러한 응용 분야들에서는 데이터베이스를 다룰 때 그 내용에 있어서도 데이터베이스의 이론적인 내용에 치중하기보다는 실제적인 활용 방법에 초점이 맞추어져야 할 것이다.

2. 통계학과에서의 교육 방안

통계학과에서 데이터베이스를 교육할 때 가장 우선적으로 고려해야 할 점은 데이터베이스의 일반적이고 이론적인 내용에 치중하기보다는 통계학이라는 학문과의 관련성을 높일 수 있도록 해야 한다는 것이다. 이것은 데이터베이스를 교육할 때 항상 데이터 처리/분석 및 활용을 염두에 두어야 한다는 의미이며, 실생활에서 데이터베이스를 활용할 수 있는 교육이 되어야 한다는 의미이기도 하다. 여기에서는 교육 방안을 데이터베이스 내용, 통계관련성 고려, 교육효과 개선과 실습 위주 교육을 지원하기 위한 정보기술 활용 등의 세 가지 측면으로 요약하여 살펴보기로 하겠다.

(1) 데이터베이스 내용

데이터베이스를 다루는 여러 교재들을 살펴보면, 구성된 내용을 다음과 같이 데이터베이스에 관한 기초적 내용과 고급/응용적 내용으로 분류할 수 있다.

- 기초내용
 - 데이터베이스 개념 : 파일 처리 시스템, 데이터베이스 시스템
 - 관계형 데이터베이스 : 관계형 데이터 모델, 관계형 데이터베이스 구조, 관계형 데이터 제약(키, 무결성 제약)
 - 관계형 데이터베이스 설계 : 개체-관계 모델, 관계형 스키마, 정규화, 데이터베이스 객체(뷰, 저장 프로시저, 트리거)
 - SQL : 데이터 정의문과 제어문, 데이터 조작문, 함수
- 고급 내용
 - 질의 처리와 최적화
 - 트랜잭션 처리
 - 동시성 제어
 - 보안 및 권한 관리 등

통계학과에서는 데이터베이스를 이론적으로 접근하기보다는 실제적인 활용 측면이 강조될 것

이므로 위의 고급/응용적 내용에서 트랜잭션 처리 부분을 제외하고는 다른 내용 모두를 다룰 필요는 없을 것이다. 트랜잭션 처리 부분은 현재 구축되어 있는 대부분의 데이터베이스 시스템이 트랜잭션 처리를 위하여 구축되어 있기 때문에 간단히 살펴볼 필요가 있겠다. 트랜잭션 처리 위주의 데이터베이스는 데이터 분석적인 측면보다는 데이터의 운영측면에 관심이 있다. 따라서 전산학적인 관점에서는 문제될 것이 없겠으나 통계학적인 관점에서는 많은 문제점이 노출되고 있다. 이 부분에 대한 이해를 통하여 데이터 분석을 고려한 데이터베이스 구축 방향을 설정하는 데 도움이 될 수 있다. 따라서 기본적인 내용들을 실습을 병행하여 교육하고, 데이터베이스를 실제로 활용하는 부분에 더 많은 신경을 써야할 것으로 생각된다.

또 한가지 문제는 응용 분야들에서 데이터베이스를 교육할 때 이용할 수 있는 교재가 거의 없다는 점이다. 기존의 데이터베이스 교재는 전산학 전공자가 전산학과 학생들의 교육에 적합하게 작성한 것이 대부분이다. 따라서 그 내용이 다분히 이론적일 뿐만 아니라 응용 분야의 학과 학생들에게는 부적합하다고 생각된다. 응용 분야의 학과에서 활용할 만한 좋은 교재의 개발에도 많은 관심이 요구된다.

(2) 통계관련성 고려

데이터에 대한 관심과 활용 목적이 다양화되고 있는 현대사회에서 통계분석기법의 활용범위는 한층 더 넓어지고 있다. 데이터를 분석할 때 우리는 보통 EXCEL, SAS, SPSS 등과 같은 응용 프로그램을 이용한다. 어떤 응용 프로그램을 이용하든 간에 데이터 분석은 첫째, 적당한 형식에 맞추어 데이터를 입력하고, 둘째, 원하는 분석 방법을 선택하여 결과를 산출하는 과정으로 이루어진다. 물론 데이터를 꼭 응용 프로그램 안에서 입력할 필요는 없으며, 외부에서 텍스트 파일 형태로 만들어 이용해도 된다. 중요한 것은 첫 번째 단계에서 누군가는 어떤 형태로든지 데이터를 입력해야 된다는 사실이고, 이것이 우리가 지금까지 데이터를 분석해온 전형적인 방법이다.

그러나 이제는 이러한 방법만으로 데이터를 분석하는 것은 학교에서 수업을 목적으로 하는 경우에는 상관이 없을지 몰라도 사회에서 다양하게 나오는 요구를 충족하기에는 부족하다. 현대 사회에서 발생하는 데이터의 대부분은 데이터베이스에 저장되고 있으며, 데이터의 특성에 따라 여러 개의 테이블들에 나뉘어져 보관된다. 따라서 데이터베이스 안에 저장된 데이터에 대한 분석을 생각하여야 하며, 통계학과에서 데이터베이스를 교육할 때 다음과 같은 사항들이 포함되도록 구성할 필요가 있다.

- 데이터 분석 응용 프로그램과의 연결
 - EXCEL, SAS, SPSS 등에서 데이터베이스에 접근하여 데이터를 분석하는 방법 및 과정
- 대용량 데이터라는 특성 감안
 - 데이터베이스에서 표본 추출하는 방법
- 데이터 분석을 고려한 데이터베이스 설계
 - 결측치(missing values) 처리 방법
 - 시계열 데이터의 효과적 활용을 위한 설계

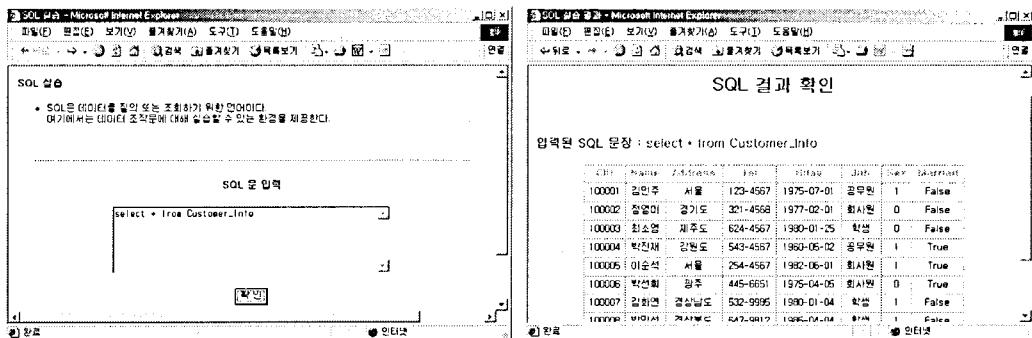
(3) 교육효과 개선과 실습 위주 교육을 지원하기 위한 정보기술 활용

최근의 정보 기술의 발전은 여러 학문 분야의 교육적 활용에 많은 도움을 주고 있다. 데이터베이스 분야도 예외는 아니며, 정보 기술을 이용하면 교육의 효과를 높일 수 있을 것으로 생각

된다. 데이터베이스와 같은 응용 과목은 실습이 뒷받침되어야 그 효과를 배가시킬 수 있기 때문이다. 또한 실생활에 활용할 수 있는 예제를 구축해봄으로서 데이터베이스의 실제적인 활용에 대한 이해를 높일 수 있을 것으로 생각된다.

- Web 기술의 활용

- 교재에서 다루기 어려운 다양한, 대용량 데이터를 활용
- 데이터베이스의 기본적인 내용에 관한 실습 도구 제공(<그림 1> 참조)



<그림 1> SQL 문장 입력 및 실행 결과

- 데이터베이스, 통계 도구(데이터 분석 및 차트 도구) 등을 이용한 실제 Web 사이트 구축
 - 데이터베이스 활용에 대한 이해 증진
 - 사회적 요구 충족

참고문헌

- [1] 박현진, 신봉섭, 심송용, 유종영, 이승천, 이정진 (1998), 변화하는 정보화 사회에 대응되는 통계계산 교과과정의 제언, 한국통계학회 춘계 학술논문발표회 논문집, 75-79.
- [2] 조신섭, 신봉섭, 이상복, 한경수 (1999), 정보관련 통계학과의 교과과정, 응용통계연구, 제12권, 제2호, 683-703.
- [3] Bryce, G. R., Gould, R., Notz, W. I. and Peck, R. L. (2001), Curriculum Guidelines for Bachelor of Science Degrees in Statistical Science, The American Statistician, Vol. 55, No. 1, 7-13.
- [4] Friedman, J. H. (1997), Data Mining and Statistics : What's the Connection?, Proceedings of the International Conference on the Interface : Computing Science and Statistics, <http://www.stat.rice.edu/interface97.html>
- [5] Ritter, M. A., Starbuck, R. R. and Hogg, R. V. (2001), Advice From Prospective Employers on Training BS Statisticians, The American Statistician, Vol. 55, No. 1, 14-18.