

## 통계 패키지를 배우는 효율적인 방법

이 장 택 (단국대학교)

### I. 서론

통계학은 자연과학, 공학을 비롯하여 사회과학을 위하여 제공되는 가장 중요한 수학의 한 분야로 매우 빠른 속도로 인식되고 있으며, 통계교육은 현대 대학수학교육의 필수적인 분야로 자리잡아 가고 있다. 확실히 수학이 자연과학의 바탕이라고 할 수 있다면 통계학은 수많은 정보 속에 파묻혀 사는 현대인에게 각 분야에서 수집되는 많은 자료로부터 의미 있는 결론을 내리는데 도움을 주는 도구라고 할 수 있다. 통계학이 이처럼 급속하게 여러 분야에서 사용되게 된 결정적인 원인은 컴퓨터의 급속적인 발전 때문이라고 할 수 있다. 왜냐하면 컴퓨터는 통계학의 이론에서 나오는 많은 자료의 분석에 필요한 복잡한 계산을 짧은 시간 내에 처리할 수 있게 하여 과거에는 현실적인 응용이 불가능하다고 생각되었던 유용한 통계기법들이 실용화되었기 때문이다. 따라서 대량의 자료처리와 복잡한 계산을 초보자들도 쉽게 하기 위하여 개발된 통계프로그램의 모음인 통계 패키지가 필요하게 되었는데, 최근에 많이 사용되는 통계패키지는 SAS, APAA, MINITAB, BMDP등이 있다. 하지만 컴퓨터의 발전과 비례하여 통계 패키지의 발전도 눈부서서 통계 패키지의 크기나 안내서도 엄청난 분량이 되어 버렸다. 따라서 이러한 결과는 통계학을 아는 사람에겐 더욱 편리하게 되었지만 통계 패키지를 처음 접하는 사람은 사용할 엄두를 낼 수 없게 되어 버렸다.

따라서 본고는 통계학을 전공하지 않았거나 또는 통계 패키지를 사용한 경험이 없는 사람

을 위하여 통계 패키지를 처음 배울 때 필요한 절차를 순서적으로 정리하여 보고자 한다. 아울러 자료를 분석할 경우에 반드시 필요한 통계 분석과정을 간략하게 서술함으로써 통계자료분석에 대한 막연한 공포감을 버리고 자신감을 가질 수 있도록 하였다.

### II. 통계 패키지의 활용을 위한 준비작업

통계 패키지를 활용하는 데는 몇 가지 준비작업이 있다. 예를 들면 근본적으로 컴퓨터를 전혀 사용하여 본 경험이 없으면 컴퓨터를 조작할 수 있는 능력을 어느 정도 배워야 한다. 그리고 평균, 분산, 표본 등에 대하여 전혀 아는 바가 없는 경우에는 통계학개론 책을 이용하여 이러한 개념들을 최소한 알아야 한다. 이와 같은 최소의 요구조건이 만족되는 경우에 우리는 보다 나은 통계 패키지의 활용을 위한 준비작업으로 다음과 같은 몇 가지 관점을 살펴 볼 가치가 있다.

#### 1. 운영체제의 활용

운영체제(Operating System)란 어떤 컴퓨터의 시스템에서도 제일 기본적으로 필요한 소프트웨어이며, 사용자의 명령문을 읽어서 데이터 파일의 관리 등 매우 근본적인 일을 수행한다. 따라서 통계 패키지를 이용하려면 자기가 가지고 있는 운영체제의 사용법을 익혀야만 한다. 특히 요즘은 개인용 컴퓨터에서 가장 많이 사용되는 운영체제는 도스이므로 통계 패키지의 활용을 위하여 도스를 이용하여 명령문을 받고,

화면에 출력하고, 디스크 파일을 관리하는 기능을 수행하는 등의 사용방법을 익혀야 한다.

## 2. 메모리와 수치처리 보조장치

메모리는 컴퓨터가 데이터나 프로그램을 기억하는 장치를 말하는 데, RAM과 ROM을 주메모리라고 하며 하드디스크와 디스켓을 부메모리라고 한다. 통계 패키지는 두 종류의 메모리를 모두 활용하는 데, 결론적으로 주메모리와 부메모리 모두 다다익선이다. 이 경우 주어진 통계 패키지가 주메모리를 어떻게 활용하는지를 매뉴얼을 참고로 확실히 알아야 한다. 수치보조장치란 통계 패키지를 사용할 때 보다 계산을 빨리 정확하게 할 때 필요하다. 하지만 IBM 호환기종 PC를 사용하는 경우에는 486이상 기종은 컴퓨터를 구입할 때 설치되어 있으므로 따로 사용자가 구입하여 설치할 필요가 없다.

## 3. 통계 패키지의 선택

앞에서 언급된 통계 패키지를 활용할 수 있는 기본준비가 되어 있다면 어떠한 통계 패키지를 선택하여 활용하는 것이 좋은가 하는 문제가 그 다음 문제일 것이다. 그런데 통계 패키지가 가져야 할 필수요건을 먼저 고려하여 보면, 분석하고자 하는 데이터를 손쉽게 컴퓨터로 불러 올 수 있는 데이터의 편집기능, 패키지 안으로 읽혀 들어 온 에이터를 사용자의 관심에 따라 자유롭게 변형할 수 있는 데이터의 관리기능, 데이터를 표로 작성하는 기능, 데이터의 평균, 분산 등 기술통계를 구하는 기능, 사용자가 이용하려는 일반적인 통계과정, 사용자가 원하는 다양한 그래프를 지원하는 기능 등이 포함될 수 있다. 이러한 측면에서 살펴본 결과에 의하면 1절에서 소개한 네 가지 패키지는 여섯 가지 측면에서 모두 우수한 것으로 알려져 있다. (김병천, 1987). 하지만 각 패키지마다 장단

점이 있으므로 패키지의 선택은 다른 통계전문가의 조언을 구하는 것이 바람직하다. 예를 들면 컴퓨터의 급속한 발전과 더불어 개발되는 통계그래픽분야를 살펴보면 대부분의 사람이 가장 뛰어나다고 하는 통계 패키지 SAS는 비록 어떠한 통계 분석을 거친 자료라도 그래픽 출력을 할 수 있는 장점은 있지만, 최근 각광을 받는 컴퓨터 그래픽스를 기초로 하는 통계분석을 하는 데는 S-PLUS를 위시한 전문적인 통계그래픽스 분석 소프트웨어에 비해 부족함이 많다. (성내경, 1993).

## III. 통계 패키지의 수행

운영체제의 사용방법을 익히고 통계 패키지를 선택한 뒤에는 사용하고자 하는 통계 패키지를 어떻게 수행할 수 있는가를 알아야만 한다. 이 경우 일반적으로 통계 패키지는 도스용과 윈도우즈용이 있는데, 윈도우즈용은 윈도우즈를 쓸 수 있다는 가정아래에서 마우스를 이용하여 아이콘을 클릭하고 메뉴를 이용하면 어렵지 않게 배울 수 있다. 하지만 윈도우즈를 사용하는 방법을 모르는 경우에는 윈도우즈에 충분히 익숙하여야만 하며 또한 사용하는 컴퓨터가 하드웨어적으로 성능이 뛰어나지 못하는 경우에는 윈도우즈용은 수행이 너무 느린 단점이 있으므로 도스용을 사용하는 것이 바람직하다. 한편 도스용은 반드시 통계 패키지의 수행을 위한 명령어를 알아야 하는 불편함이 있다. 도스용을 사용하는 경우에는 반드시 기억하여야 하는 명령어 또는 방법은 다음과 같은 것이 있다.

(1) 프로그램은 어떻게 입력하는가를 알아야 한다. 이 경우 패키지내에서 프로그램을 어떻게 입력할 수 있는가를 알아야 하며 만일 외부의 편집기를 이용하는 경우엔 작성된 프로그램을 어떻게 주어진 패키지내로 불러 들일 수 있는가를 알아야 한다.

(2) 작성된 프로그램을 어떻게 수행하는가

를 알아야 한다. 이 경우 프로그램을 정상적으로 작성하여 수행하여 나온 출력결과를 어떻게 볼 수 있는가를 알아야 하며 또 만일 프로그램이 에러가 생기면 수정하여 다시 수행을 하는 과정을 반드시 알아야 한다.

(3) 작성된 프로그램의 저장 및 수행결과와의 저장을 어떻게 하는가를 알아야 한다. 작성된 프로그램은 다음에 또 활용될 가능성이 많으므로 저장을 어떻게 하여야 하는지를 알아야 할 것이며 수행결과는 프린터로 보낼 수도 있고 파일로 저장하여 보고서의 한 부분으로 활용할 수 있으므로 각각의 경우에 어떻게 저장하는지를 알아볼 필요가 있다.

(4) 저장된 파일을 통계 패키지내로 어떻게 불러들일 수 있는지를 알아 보아야 한다.

(5) 통계 패키지의 수행을 종료할 때 어떻게 하여야 하는지를 알아 보아야 한다.

#### IV. 통계 패키지 프로그램의 구성

##### 1. 프로그램의 두 단계

통계 패키지를 이용하는 경우에 작성되는 프로그램은 크게 두 가지 단계로 나누어지는 데, 첫째는 데이터단계이며 둘째는 데이터단계에서 만들어진 데이터셋을 이용하여 작업을 수행하는 수행단계이다. 데이터단계의 구성은 패키지마다 약간의 차이가 있지만 대체적으로 시작할 때 데이터셋의 이름, 데이터가 있는 장소를 지정하는 문장들, 데이터의 단계중 각 변수의 입출력 포맷지정, 데이터의 생성, 수정, 치환을 위한 문장, 데이터단계의 끝을 선언하는 문장 등으로 이루어진다. 그리고 일단 데이터셋이 만들어지면 이 데이터셋을 이용하여 통계적인 분석을 하게 되는 데, 이 단계를 수행단계라고 할 수 있다. 따라서 컴퓨터 프로그램에 관하여 비전문가들은 통계 패키지를 잘 익혀두면 한가지 컴퓨터 언어를 배우는 것이 되며, 통계 패키지가 수행단계에서 어떠한 일들을 할 수 있는 지

를 잘 파악하고 있으면 통계분석을 하지 않는 경우에도 매우 다양하게 활용할 수 있다. 예를 들어서 행의 개수가 20, 열의 개수가 20인 행렬의 역행렬, 행렬식의 값 등에 관심이 있다면 이 문제를 프로그램을 작성하여 해결하려면 컴퓨터 언어를 잘 알아야 하며 또 수치해석에 관한 여러 가지 문제점도 파악을 하고 있어야 프로그램을 효율적으로 작성할 수 있다. 하지만 일단 패키지가 주어진 숫자를 행렬의 원소라고 인식하고 각각 1개의 명령으로 행렬의 역행렬과 행렬식의 값을 프린트하라고 했을 때 수행할 수 있다면 얼마나 편리하겠는가? 통계 패키지중 일부는 이러한 문제도 쉽게 해결할 수 있다.

수행단계를 이용하는 경우에는 먼저 사용하는 패키지가 어떤 역할을 감당할 수 있는지를 살펴보는 것이 바람직하다. 예를 들면 자료를 쉽게 출력할 수 있는 지, 자료를 크기순으로 배열할 수 있는 지 등과 같은 기본적인 것부터 자기가 이용하고자 하는 통계분석까지 수행이 가능한지를 먼저 알아보아야 한다. 이 경우 매뉴얼 또는 패키지의 도움말기능을 이용하면 분석 가능한 프로시저의 종류를 알 수 있다. 그리고 만일 어떤 특정한 통계분석을 하는 경우에 있어서 사용하려는 패키지를 이용하여 분석이 가능한 경우에는 단지 그 명령을 어떻게 수행하는지를 알기만 하면 되는 것이다.

##### 2. 데이터셋의 작성

통계 패키지가 인식할 수 있는 데이터셋은 일반적으로 패키지내에 들어가서 데이터셋을 만드는 경우와 외부 데이터 파일을 불러 들여서 만드는 경우가 있다. 외부 데이터 파일은 성격상 크게 두 가지로 나뉘어지는 데 하나는 아스키 코드로 구성된 파일과 로터스, 디베이스와 같은데서 구성된 파일로 나눌 수 있다. 하지만 패키지에 따라서 이와 같은 모든 파일을 모두 취급하지 않을 수도 있음에 유의할 필요가 있

다. 따라서 자기가 분석하고자 하는 자료를 사용하려는 패키지가 취급을 할 수 있는지를 확인한 후 그 다음에는 사용하려는 패키지가 데이터를 읽어 들여서 데이터셋으로 만드는 데이터 프로세스 과정을 알아 두어야 한다. 한편 데이터셋이 다 만들어져 있는 경우에도 주어진 데이터셋을 이용하여 가로병합을 하거나 세로병합 또는 변수의 갯수만 바꾸어 줄 필요가 있는 경우가 종종 있다. 예를 들어서 한국중학교 1학년 학생들의 I.Q.를 알아보기 위하여 1반부터 10반까지 각 반별로 학생의 이름과 I.Q.를 입력한 데이터파일 DAT1.DAT부터 DAT10.DAT까지 10개를 만들었다. 이 자료를 이용하여 한국중학교 1학년 학생들의 I.Q.의 평균, 최대값, 최소값, 중앙값, 분산 등과 같은 기술통계값을 구한다 고하여 보자. 이 경우는 먼저 DAT1.DAT부터 DAT10.DAT까지를 패키지가 알 수 있는 데이터셋으로 변환한다. DAT1.DAT을 이용하여 만든 데이터셋을 DAT1이라고 하고 이와 같이 DAT10.DAT를 이용하여 만든 데이터셋을 DAT10이라고 하면 DAT1부터 DAT10까지를 이용하여 세로병합을 하면 된다. 또 다른 예로써 한국중학교 1학년 학생들의 체력감사를 하는 경우에 오늘은 달리기, 윗몸일으키기, 던지기를 하고 내일은 멀리 뛰기, 오래 달리기, 탁걸이를 한다면 학생들의 체력검사결과를 1개의 데이터셋으로 만들기 위하여 오늘의 자료를 모든 데이터셋과 내일의 자료를 모은 데이터셋을 가로 병합하여야 할 것이다. 또한 기존의 데이터셋을 이용하여 변수중 일부분을 선택하여 창조할 수도 있고, 관측치의 일부분을 선택하여 창조할 수도 있다. 이와 같은 필요성을 만족하기 위하여 통계 패키지에서 수행하여야 하는 작업은 매우 간단하다. 그러므로 자기가 가지고 있는 자료를 직접 수정하려고 하지 말고 패키지상에서 원하는 데이터셋을 만드는 것이 편리하며 따라서 패키지가 어떻게 이 작업을 할 수 있는지를 명확히 알아 둘 필요가 있다.

## V. 기본적인 통계분석의 종류

대부분의 통계 패키지는 굉장히 많은 통계분석기법들을 포함하고 있다. 그러나 이 통계기법들 중에는 기법들이 사용될 때 따라가는 가정이 때로는 모호하거나 현실성이 없을 수도 있다. 따라서 통계학의 비전문가들로서 이러한 기법들을 모두 이해하고 터득하기도 매우 어려운 일이다. 그러나 실제로 자료분석에 많이 활용되는 통계기법은 대체적으로 몇몇 방법에 한정되어 있다. 예를 들어 신한풍 외 4인(1991)이 발표한 논문에서 따르면 생물학에서 예방의학회지 논문(81년부터 90년)에서 사용된 통계방법론의 빈도 수를 조사한 결과 총 465건의 논문중 기술통계량에 관한 논문이 132건으로 가장 많았고, 산포도에 관한 논문이 78건, 카이제곱검정이 50건, t-검정이 32건, 회귀분석이 32건, 상관계수가 29건, 분산분석이 24건의 순서이다. 다행스럽게도 이 내용 모두가 기초통계교과서에서 취급하는 방법론으로 통계학의 기초에만 충실하더라도 통계 패키지를 이용한 대부분의 자료분석을 수행할 수 있다. 이와 같은 경향은 다른 분야에서도 비슷하다고 알려지고 있다. 따라서 너무 복잡하고 어려운 통계기법을 이해 없이 사용하지 말고 기본적인 통계기법을 완전히 소화하면 대부분의 자료분석을 할 수가 있다. 다음은 이와 같은 이유로 기본적으로 꼭 알아 두어야 할 기초적인 통계기법을 간략하게 약속한 것이다.

### 1. 파일 처리에 유용한 절차

#### (1) 소트

우리는 만들어진 데이터셋에서 한 개 이상의 기준변수에 의하여 관측값을 재배열할 필요성을 종종 느낀다. 예를 들면 특정반의 성적울 성적이 좋은 학생순으로 재배열할 경우 등 이러한 필요성은 매우 빈번하게 일어난다.

(2) 프린트

주어진 데이터셋이나 새롭게 만들어진 데이터셋을 이용하여 자료의 값이 정확하게 입력이 되었는지를 확인하거나 원하는 형태로 데이터가 생성되었는지를 확인하거나 분석한 결과를 출력하고 싶은 경우에 프린트한다. 통계 패키지의 가장 기본적인 기능중 하나이다.

2. 기술통계의 여러 가지 절차

주어진 자료를 정리하여 그림으로 나타내거나, 그 자료의 특성을 나타내는 여러 가지 통계량의 값을 나타내는 분야를 흔히 기술통계학이라 하며, 패키지에서 기술통계를 위한 중요절차는 다음과 같은 것이 있다.

(1) 차트

데이터를 가장 잘 파악하는 방법은 여러 형태의 도표를 그려보는 것이라고 할 수 있다. 그러므로 만들어진 데이터셋을 이용하여 가로, 세로 막대그래프, 3차원블록그래프, 원그래프, 별그래프 등을 그려주는 기능이 필요하다.

(2) 플롯

변수사이의 관계를 알아보는 방법 중 가장 중요한 방법중 하나라고 할 수 있는 변수사이의 관계를 산점도로 나타내거나, 제 3의 변수 또는 함수의 값에 따라 결정된 명암으로 등고선을 그릴 수 있어야 한다.

(3) 도수분포표 또는 분할표

도수분포표는 자료정리에 가장 많이 이용되는 방법중 하나이며 도수분포표는 측정값이 나타내는 빈도수를 정리하여 놓은 것이다. 이때 상대도수도 보여주고 측정값 순서대로 누적한 누적상대도수를 같이 보여주기도 한다. 분할표는 두 개의 질적 또는 이산형변수를 정리하여 그 연관된 특성을 연구하는 데 매우 효과적인 도표이다.

(4) 기술통계량

통계량 주어진 정리되지 않은 복잡한 데이터를 압축시켜 쉽게 파악할 수 있는 정보로 변환시키는 것이라고 하면 여러 가지 통계량을 계산하여 주어진 변수의 분포를 수치적으로 묘사하여 볼 필요성이 있다. 따라서 주어진 숫자변수들에 대한 관측값, 평균, 표준편차, 최소값, 최대값, 중앙값, 극한값 등의 기본적인 기술통계량을 제공하는 수행단계가 반드시 필요하다.

(5) 상관계수

두 개 이상의 변수들을 분석하기 위하여 여러 가지 형태의 산포도를 그려보는 것이 효과적이거나 변수 사이의 상관의 관계를 나타내는 측도를 이용하기도 하는 데, 대표적인 경우로 상관계수를 들 수 있다. 하지만 사용자는 일반적인 상관계수는 두 변수의 직선적인 관계를 나타낸다는 사실을 주목하여야만 한다.

3. 기본적인 통계분석의 예

(1) t-검정

t-검정은 한 개의 모집단이나 두 개의 모집단의 미지의 모평균에 대한 가설검정의 대표적인 방법이다. 따라서 패키지에서 다룰 수 있는 기본적인 t-검정은 다음과 같다.

(가설1) 한 개의 모집단에서 모평균이 주어진 값과 같다.

(가설2) 두 개의 모집단에서 독립적인 두 변수들의 모평균들이 서로 같다.

(가설3) 두 개의 모집단이나 한 실험단위에 두 처리를 적용시킨 경우에 두 변수들의 모평균들이 서로 같다.

(2) 분산분석

t-검정은 두 개의 모집단의 미지의 모평균들의 동일성을 검정하는 방법이나 세 개 이상의 집단에 있어서 모평균에 대한 동일성검정을 하는 대표적인 방법은 분산분석이다. 예를 들면

교육학에 있어서 A, B, C 세가지 교수법의 효과를 비교하기 위하여 15명의 성적이 비슷한 학생을 선발하여 5명씩 3그룹으로 나누어 각각 A, B, C 세가지 교수법으로 지도한 다음에 성적을 얻은 경우에 세 가지 방법의 교습효과에 차이가 있는지를 알아보는 경우를 들 수 있다. 따라서 이 문제에 분산분석을 적용하면 일단 세 가지 방법의 교습효과에 대한 차이의 유무를 통계적 가설검정을 이용하여 결정하고 주어진 유의수준에서 만일 세 가지 방법의 교습효과에 차이가 있다고 판정하면 어느 방법이 서로 심각하게 다른지를 알고 싶을 것이다. 이와 같이 어느 처리효과 사이에 차이가 있는지를 알아보기 위하여 검정을 시행하는 방법을 다중비교라고 하는 데, 대부분의 분산분석은 이와 같은 형태로 진행되어진다.

### (3) 범주형 자료분석

일반적으로 범주형 자료란 조사대상이나 관측결과에 따라 몇 가지 범주로 분류되는 자료들을 의미하는 데, 예를 들면 키를 상, 중, 하로 나누면 범주가 3개가 있는 경우이며 각 범주에 대하여 조사대상의 득수를 표시하면 범주형 자료를 완성할 수 있다. 이와 같은 범주형 자료를 분석할 경우에는 일반적으로 적합도 검정, 동일성 검정, 독립성 검정이라는 세 가지 검정의 형태가 나타나며 대표적인 분석기법은 카이제곱 검정을 실시하게 된다. 먼저 적합도 검정이란 다항시행에서 각 범주에 속할 확률들의 참값을 가정하고 실제자료가 가정된 확률모형에 얼마나 적합한가를 검정하는 것을 의미한다. 한 가지 예를 들면 특정호텔에 대하여 하루동안의 예약취소건수가 포아송분포를 따른다는 가정이 적합한지를 검정하는 경우는 바로 적합도 검정이 된다. 분할표를 이용하는 경우에 흔히 다룰 수 있는 두 가지 검정은 독립성 검정과 동일성 검정이다. 독립성 검정이란 두 속성이 서로 독립인가를 검정하는 데 사용된다. 예를 들면 교육수준과 사회경제적 지위라는 두 가지

속성이 서로 독립인가? 하는 종류의 검정이며, 동일성 검정이란 한 성질에 따라 분류되는 몇 개의 부모집단이 어떤 한 특성에 대하여 동일한 분포를 갖는가를 검정하는 데 사용된다. 예를 들면 저장상태에 따라서 범씨의 발아율이 동일한가를 묻는 종류의 검정이다.

### (4) 회귀분석

통계학에서 중요한 문제중의 하나는 변수들의 관련성을 설명하는 것이다. 예를 들어서 위에서 소개한 변수들간의 산점도를 작성하는 플롯이나 상관계수를 구하는 것도 하나의 방법이 될 수 있으며 또 분할표를 이용하더라도 두 범주형 자료의 관련성을 조사할 수 있다. 하지만 분석의 목적이 되는 변수(종속변수)와 수단이 되는 변수(독립변수)들간의 관계를 함수적으로 규명할 수 있다면 예측의 기능을 첨가하여 변수들의 관계를 설명할 수 있는 데 이와 같은 분석방법을 회귀분석이라고 한다. 예를 들면 아버지의 키와 아들의 키, 진통제의 투여량과 진통지속시간의 관계등은 회귀분석으로 설명할 수 있는 대표적인 경우인데, 적합한 모형을 플롯 등을 이용하여 설정하고 통계 패키지를 이용하여 모형을 추정한다. 모형을 추정한 뒤 잔차분석 등을 이용하여 모형의 타당성을 검정한 후에 최종 모형을 결정하고 이 모형을 이용하여 사용자가 원하는 분석을 하는 것이 회귀분석을 하는 일반적인 절차이다.

## VI. 결론

지금 우리는 컴퓨터하드웨어의 눈부신 발전 덕분에 개인용 컴퓨터를 이용하여 자료분석을 충분히 할 수 있는 시대에 살고 있다. 이와 같은 관점에서 본고는 통계학에 비전문가인 사람들이 처음 통계 패키지를 접할 때 알아야 하는 주의점과 배우는 절차를 순서적으로 서술하였다. 하지만 많은 사람들은 통계 패키지를 사용하여 결과를 얻으면 끝이라고 생각하는 데 이

점은 잘못된 생각이다. 현재로는 통계 패키지의 역할은 단지 복잡한 계산을 대신하여 주고 사용자가 판단할 수 있도록 도움을 준다는 차원에서 그치고 있다. 그러므로 통계분석을 하는 경우에는 결과를 해석하는 능력이 반드시 필요하다. 예를 들면 우리가 통계학에서 가장 많이 접하는 분류의 문제중 하나인 가설검정은 일반적으로 교재에서는 유의수준  $\alpha$ 에서 가설의 채택여부를 결정하게 된다. 그러나 패키지상에서는 가설의 채택여부를 결정하는 것이 아니라 유의수준대신 유의확률을 제시함으로써 결정은 사용자의 주관에 맡긴다. 따라서 통계 패키지를 이용하여 자료분석을 행할 때 "분석자가 원하는 것이 무엇이나?"하는 기본철학을 명확히 하여야 하며 또 결과를 분명하게 해석할 수 있어야 한다. 그러나 머지 않은 장래에 사용자가 통계적 문제를 해결하고자 할 때 통계전문가가 이를 해결하는 것과 유사하게 컴퓨터가 수행하도록 만들어 놓은 통계전문가시스템이 널리 보급될 것으로 생각되어진다. 그때까지는 불편하

더라도 통계학의 기본에 충실하지 않으면 자료 분석이 정확하지 않을 것이다.

### 참 고 문 헌

- 신한풍·송해항·김병수·이중협·한원식 (1991). 생물통계의 현황(1981-1990). 통계학연구, 10권, 62-91.
- 성내경 (1993). SAS/GRAPH의 성능과 한계. 응용통계연구, 6권 1호, 13-22.
- 김병천 (1987). 개인용컴퓨터에서의 통계 패키지의 선택과 활용. 응용통계연구, 1권 1호, 75-90.
- 강근석·김성철·김지현·이윤오·이정진·이창수 (1994). PC 통계학. 자유아카데미.
- 최병선 (1995). PC SAS입문. 박영사.
- 연세대학교 응용통계학과 (1991). 미니랩. 자유아카데미.
- 송일성 (1990). SPSS/PC\* 기초통계분석. 자유아카데미.