

SAS/GRAPH의 성능과 한계¹⁾

- S-PLUS의 기능과 대비하여 -

성 내 경²⁾

요 약

SAS/GRAPH는 SAS 시스템을 구성하는 하나의 모듈로서 SAS 시스템에 뒤늦게 추가된 소프트웨어로, SAS 시스템 내의 어떠한 통계 분석을 거친 자료라도 그래프스 출력시킬 수 있는 장점이 있는 반면, 최근 각광을 받는 컴퓨터 그래프스를 기초로 하는 통계 분석을 시도하기에는 상당히 불편한 단점이 있다. 이 점에 있어서는 S-PLUS를 위시한 전문적인 통계 그래프스 분석 소프트웨어에 못 미친다.

1. SAS/GRAPH의 기능

이하에서 언급되는 SAS 시스템은 6.04 판, S-PLUS는 2 판이 기준이다.

SAS/GRAPH는 SAS 시스템을 구성하는 하나의 모듈(module)로서 통계 자료를 도표, 그래프, 슬라이드 형태의 높은 해상도 그래프스(high resolution graphics)로 출력시킬 수 있는 통계 그래프스 시스템(statistical graphics system)이다. SAS/GRAPH는 SAS 시스템에서 출력되는 데이터 포맷을 공유하며 SAS 언어의 사용법을 그대로 답습하므로 SAS를 이용한 통계 분석에 익숙한 사용자라면 사용상 별다른 어려움은 없다.

SAS/GRAPH는 최종 보고서나 연구 논문에 삽입될 세밀한 그래프, 세미나 등에서 발표용으로 사용할 슬라이드나 OHP용 투명지 제작 등에 활용되면 진가를 발휘하는 소프트웨어이다.

¹⁾ 한국과학재단(과제 번호: 913-0105-003-2)과 1992년도 이화교수연구기금에 의하여 일부 지원된 연구임.

²⁾ (120-750) 서울특별시 서대문구 대현동 11-1, 이화여자대학교 통계학과
Bitnet : NKSUNG@KREWHACC

(1) SAS/GRAPH의 최소 요구 제원

SAS/GRAPH는 최소 640 KB의 재래식 메모리를 장착한 80286 AT 급 이상의 PC에서 운용 가능하다. 그러나, 그래픽스 화일의 처리는 상당한 내부 작업 시간을 요구하므로, 실용적인 관점에서 보면 적어도 2 MB의 EMS 메모리를 확보한 80386 급 이상의 PC가 필요하다. 수치 보조 처리기(math-coprocessor)가 반드시 필요하지는 않지만 탑재하면 현저히 처리 속도가 증진된다. 출력 장치로는 가급적 VGA 급 이상의 천연색 모니터와 레이저 프린터를 갖추는 편이 유리하다.

그리고, 가용 메모리를 상당히 소모하는 인터랙티브 모드(interactive mode)의 사용을 지양하고, 대신 배치 모드(batch mode)에서 SAS/GRAPH를 운용하면 처리 속도가 배가된다.

SAS/GRAPH에서는 시장에 나와있는 거의 모든 종류의 출력 장치를 지원하며, 필요한 경우 GDEVICE 절차를 이용하여 사용자 전용의 장치 구동기를 정의할 수 있다.

어떠한 그래픽스 소프트웨어에든 공통적으로 적용되는 출력 속도 증진 방법으로는 수치 보조 처리기를 장착하고, 모니터를 제어하는 그래픽스 카드에 전용 비디오 메모리를 1 MB까지 최대로 올리고, 시스템에 디스크 캐쉬(disk cache) 프로그램과 프린터 캐쉬(printer cache, 또는 spooler) 프로그램을 상주시켜야 한다. 그리고, 레이저 프린터의 출력 크기는 프린터 메모리에 의존하므로 이것 역시 최소 1.5 MB의 메모리를 갖추어야 통상의 8"×10"의 출력을 얻을 수 있다.

운용상의 최소 요구 제원만을 비교하면 SAS/GRAPH가 S-PLUS보다 더 싸게 운용할 수 있다고 하겠다. 그러나, S-PLUS의 요구 제원이 더 가혹한 이유는 그래픽스 화면 상에서 실시간 처리(real-time processing)를 이용한 통계 분석이 가능하기 때문이다.

지원하는 출력 장치를 비교하면 S-PLUS는, 김성수와 한경수 [1]가 요약한대로, SAS/GRAPH와는 비교도 안 될 정도로 제한적이다. S-PLUS의 장점으로 과학자들이 많이 사용하는 문서 컴파일러 TeX과의 호환성을 들기도 하나, 이는 PostScript 포맷으로 레이저 프린터를 가동시키는 중에 자연스럽게 결합된 성질일 뿐 그 자체를 일반적인 장점으로 열거할 수는 없다.

(2) 본문 출력(text output) 제어 기능

본문 출력 제어라 함은 출력되는 문서에서 사용될 문자의 자형(font), 크기, 색상 등의 조절을 의미한다. SAS/GRAPH에서 본문 제어는 거의 무제한이라고 할만큼 융통성이 있다.

가장 중요한 자형을 보면, 기본형 30여종과 변형 70여종을 포함하여 모두 백여종에 이르는 자형을 윤곽선 자형(outline font) 형식으로 지원하므로, 글자의 확대 및 축소에 아무런 제한이 없다. 게다가 GFONT 절차를 사용하면 한글을 포함한 특수 문자

등의 사용자 정의 자형을 만들 수 있다.

색상(color)은 Tektronix사의 천연색 표준 256가지를 지원하지만, 색상의 구현은 전적으로 출력 장치의 성능에 달려있다. VGA급 모니터에서는 256가지 중 골라서 16가지까지의 지정이 가능하다. 글자 단위의 색상 지정과 배경색의 지정이 가능하다.

글자 단위의 회전, 밑줄, 상자 둘러치기 등을 할 수 있으며, 한 줄의 문장에 대하여도 동일한 기능이 가능하며 정렬 방식의 지정도 자유롭다. 그리고, 주어진 이차원 좌표들을 연결하여 직선들을 그을 수도 있다. 밑줄이나 상자, 직선 등에는 선의 굵기 등의 보조 지정을 지원한다.

GMAP 절차를 사용하면 아무리 복잡한 형태의 이차원 도형이라도 함수식으로 좌표를 주거나, 또는 사용자가 직접 좌표를 주는 방식으로 그릴 수 있다.

S-PLUS에는 SAS/GRAPH만큼의 다양성은 부족하다. 그러나, 통계 그래픽스에 아름다운 장식이 반드시 필수불가결하다고 할 수는 없으며, 또한 S-PLUS에는 기본적으로 필요한 몇 가지 도형의 출력 등은 이미 매크로(macro)화 되어 있는 장점이 있다.

(3) 그래픽스 출력(graphics output) 제어 기능

그래픽스 출력의 제어는 출력되는 선(line), 면(area), 심볼(symbol), 범례(legend), 축(axis), 보간법(interpolation) 등의 조절을 의미한다. 그래픽스 출력의 제어 형태는 SAS/GRAPH 내의 특정 절차와 출력 장치에 의존하므로 일반적인 기술은 어렵다.

후백 인쇄를 기준으로 할 때, SAS/GRAPH에서는 모두 46가지의 서로 다른 형태의 선을 출력할 수 있으며 선의 굵기 조절이 자유롭다. 면의 종류는 면을 채우는 방식에 따라 구분된 17가지의 패턴을 기본으로 하여 면을 채울 때 사용하는 사선의 각도 지정이 자유로운 관계로 이 역시 무제한이라고 할 수 있다.

산점도 등에서 데이터를 나타내기 위하여 사용되는 심볼은 고유 이름을 갖는 특수 문자 34종 외에 SAS/GRAPH에서 지원하는 모든 종류의 자형 내의 임의의 문자를 포함한다. 크기의 지정도 자유롭기 때문에 이것 역시 지정에 제한이 없다.

보간법으로는 삼차 스플라인(cubic spline) 평활법을 필두로 라그랑주 보간법 등 몇 가지의 평활 알고리즘을 내장하고 있다. 그러나, 데이터가 지나치게 많은 경우 DOS하에서 스플라인 방법은 메모리 부족으로 실행되지 않을 수 있다.

축과 범례의 조절은 그래프의 출력 형태를 좌우하는 중요한 요소로, 사용된 심볼의 의미, 레이블(label), 틱 표시(tick mark) 등에 관련된 것들이다. 이런 사항들도 출력 공간에 제한만 없다면 자유롭게 지정할 수 있다.

이러한 제반 기능 역시 SAS/GRAPH가 S-PLUS보다 더 다양하다. 그러나, SAS/GRAPH에서 제공되는 많은 기능에도 불구하고, 일반 사용자의 입장에서 보면 그래픽스 출력의 제어는 가능한 한 최소화하는 편이 사용하기에 편하므로, 반드시 SAS/GRAPH가 S-PLUS보다 낫다고는 할 수 없다.

2. SAS/GRAPH의 부속 절차

여기서는 SAS/GRAPH에서 호출할 수 있는 주요 절차들의 기능을 요약한다. S-PLUS와의 기능 비교는 김성수와 한경수 [1]를 보라.

(1) 그림 그리기

① GCHART 절차

기본 SAS 소프트웨어에 포함된 CHART 절차의 그래픽스 출력에 해당하며 여러가지 유형의 도표(chart)를 제작하는 절차이다. 이하 다섯 가지의 도표를 생성할 수 있다.

- 수직 막대 도표(vertical bar chart) - 수직 히스토그램
- 수평 막대 도표(horizontal bar chart) - 수평 히스토그램
- 블록 도표(block chart)
- 파이 도표(pie chart)
- 별 도표(star chart)

② GPLOT 절차

기본 SAS 소프트웨어에 포함된 PLOT 절차의 그래픽스 출력에 해당한다. 기본적으로 두 변수 간의 산점도(scatter plot)를 그린다. 데이터 점을 연결하는 방식이 수십 가지나 되며, 산점도에 직접 회귀선을 적합시키고 신뢰대(confidence band)를 출력할 수 있다. 분산 분석 자료 등에서 요인의 수준별 반응값들의 추이를 한 눈에 볼 수 있는 다양한 그래프 형식을 제공한다. 로그 척도의 그래프, 비례원도(proportional circle graph) 등을 그린다. 둘 이상의 그래프들의 겹쳐 그리기가 가능하다.

③ G3GRID 절차와 G3D 절차

삼차원도(3-dimensional plot)를 그리는 절차들이다. 두 변수들의 값을 조절하면서 얻은 반응값들을 삼차원 공간 상에서 일일이 찍으면 삼차원 산점도(3-dimensional scatter plot)가 되며, 삼차원 산점도를 기초로 드문드문 산재되어 있는 반응값들을 스플라인 평활법(spline smoothing)을 포함한 각종 보간법(interpolation method)을 사용하여 매끄럽게 연결한 다음 입체화하여 삼차원 표면도(3-dimensional surface plot)를 그릴 수 있는데, G3D 절차로 이런 그림들을 그릴 수 있다. G3GRID 절차는

그림을 그리는 절차가 아니라 G3D 절차로 삼차원 표면도를 그리는데 필요한 격자 좌표 데이터를 만드는 보조 절차이다. 데이터가 $z=f(x,y)$ 형태의 함수식으로 생성되는 경우에는 G3GRID 절차없이 G3D 절차로만 삼차원 표면도를 그릴 수 있다.

④ GCONTOUR 절차

삼차원 등고선도(3-dimensional contour plot)를 그린다. 기본 SAS 소프트웨어의 PLOT 절차에서 선택사항 CONTOUR를 지정한 것과 같다. 등고선도를 그리는데 필요한 정보가 G3GRID 절차에서 주어질 수 있다.

⑤ GPRINT 절차

때때로 일반 SAS 절차의 문서 출력을 그래픽스 출력으로 변화시킬 필요가 있을 수 있다. GPRINT 절차는 이런 목적에 사용한다. 출력을 더 보기 좋게하기 위하여 본문 제어 기능을 덧붙일 수 있다.

⑥ GSLIDE 절차

이 절차는 오로지 문서 출력만을 담당한다. 보고서나 슬라이드에서 말로 설명할 부분을 제작할 때 사용한다. SAS/GRAPH에서 가장 많이 호출되는 절차이다.

(2) 유틸리티

① GDEVICE 절차

출력 장치 구동기(output device driver)를 관리한다. 즉, 출력 장치를 제어하는 파라미터(parameter)의 값을 변화시켜 출력의 크기라든가 질을 변경할 수 있다.

② GFONT 절차

출력용 자형(font)을 제작한다.

③ GOPTIONS 절차

현재 설정된 시스템 디폴트 값들을 출력한다.

④ GTESTIT 절차

SAS/GRAPH의 설치 상태의 양호를 확인한다.

⑤ GREPLAY 절차

SAS/GRAPH에서 출력된 그래픽스 카탈로그를 관리하며, 재생(replay)과 판형(template) 출력 기능을 지원한다. SAS/GRAPH에서 일차 출력되는 그래픽스 화일은 장치 독립(DVI: device independent) 화일 포맷으로 저장되며, 여기에 출력 장치의 제어 명령이 덧붙여 실제 출력 장치로 나간다. GREPLAY 절차는 간단히 말해서 중간 단계의 그래픽스 출력 화일을 관리하는 절차이다. 그래픽스 소프트웨어에서 필수 불가결한 관리 절차이다.

(3) 지도

① GMAP 절차

이차원 혹은 삼차원 지도를 제작한다. 지도 제작의 기법을 활용하여 다양한 형태의 기하 도형들을 제작할 수 있다.

② GPROJECT 절차

지도를 그리는데 사용될 사영(projection) 방법을 지정한다.

③ GREMOVE 절차

같은 속성을 갖는 지역 간의 경계선을 제거하여 더 큰 단위의 지역을 만든다. 결과를 새로운 지도 자료로 출력할 수 있다.

④ GREDUCE 절차

지도 좌표 자료를 처리하여 지도를 그리는데 필요한 점의 수를 줄인다. 점의 수가 줄면 정밀도는 떨어지나 처리 시간은 빨라진다.

3. SAS/GRAPH의 출력 관리

SAS/GRAPH에서 그래픽스 출력은 다음과 같은 세 가지의 형태로 내보낼 수 있다.

- ① DVI 화일로 출력하여 카탈로그에 보관
- ② GSF (Graphics Stream File) 컴퓨터 화일로 출력
- ③ 출력 장치로 직접 출력

이상의 세 가지 방식을 모두 지원하는 것이 SAS/GRAPH의 강점이다. 일반적으로 그래픽스 처리는 많은 시간을 요한다. 따라서, 같은 그림이 필요할 때마다 매번 프로그램을 다시 실행하는 것은 극히 비효율적이다.

DVI 화일을 만들어 놓으면 후에 단순히 출력 장치 제어 코드만 추가하여 어떠한 장치로도 출력이 가능하므로 다시 원시 프로그램을 실행시킬 필요가 없다. 그리고, DVI 화일들이 집적된 그래픽스 카탈로그를 관리하는 GREPLAY 절차를 이용하면 일정 시간마다 자동적으로 모니터 등에 출력을 내보내는 등 일체 사람의 손을 빌지 않는 자동 재생, 또는 무인 전시를 시킬 수가 있다. 덧붙여 단일 출력 공간을 적당히 분할하여 여러 장의 그림을 동시에 출력시키는 판형 출력도 가능하다.

GSF 화일은 DVI 화일에 출력 장치 제어 코드가 삽입된 컴퓨터 화일로서, 그래픽스 출력을 직접 출력 장치로 보내지 않고 컴퓨터 화일로 저장하는 방식이다. 일단 GSF 화일을 만들어 놓으면 더이상의 SAS 시스템의 호출이 없이 아무 때나 지정된 출력 장치에서 곧바로 하드카피(hard copy)를 얻을 수 있다.

4. SAS/GRAPH의 한계

근차 Amstat News [2]에 실린 통계 소프트웨어 광고들을 보면 어느 소프트웨어나 화려한 통계 그래픽스 기능을 과시하고 있다. PC의 환경이 평범한 DOS에서 그래픽스 위주의 Windows로 넘어가듯, 숫자를 다루는 통계 분석도 분야에 따라 그래픽스 기법의 대두가 현저하다. 특히 다변량 분야에서 브러싱(brushing), 스피닝(spinning) 등은 컴퓨터 그래픽스 없이는 나타낼 수 없는 기법들이다. 또한, pdf도(density plot), 변수 변환 등의 결과를 즉각 화면에서 볼 수 있는 기능들도 컴퓨터 그래픽스의 이점이다.

S-PLUS에는 이러한 최신 그래픽스 분석 기법들이 갖추어져 있지만 SAS/GRAPH에는 결여되어 있다. 이러한 기능 상의 차이는 두 소프트웨어의 역사와 개발 방식의 차이에서 연유한다.

SAS 시스템은 DOS의 초창기에 탄생하였기 때문에 기본적으로 문서 화면 위주로 프로그램 개발이 진행되었고, SAS/GRAPH는 5 판에서 비로소 삽입되었다. 따라서, SAS/GRAPH에는 다른 모듈의 소프트웨어에서 이송되는 SAS 자료를 받아 그래픽스로 처리해야 한다는 피할 수 없는 제한이 뒤따르게 되었고, 이런 점 때문에 그래픽스의 실시간 처리는 SAS/GRAPH에서는 불가능한 과제가 되고 말았다. 즉, 예를 들어, 다변량 분석 따로, 그림 따로가 되는 셈이다. 그러나, 실시간의 제한이 없다면 SAS/GRAPH에서도 보이는 각도 조절, 스피닝 등을 구현할 수 있기는 하나 처리 속도가 늦어 실용성은 없다.

반면에 S-PLUS는 SAS 시스템에 비하여 상당히 제한된 통계 분석 절차에 국한하여 소프트웨어를 개발하였기 때문에 분석 절차마다 직접 그래픽스 루틴의 삽입이 가능하였고, 최근 개발된 그래픽스 기법의 삽입이 SAS/GRAPH에 비하여 용이하였다.

S-PLUS와 비견될만한 SAS 모듈로 SAS/INSIGHT가 있다. 이 모듈은 최근 선보인 SAS 6.07 판부터 삽입되었으며, SAS/GRAPH의 한계를 인식하고 새로이 제작된 전문 통계 그래픽스 분석 소프트웨어로 마우스를 지원하며 S-PLUS 이상의 기능을 갖고 있다. SAS/INSIGHT는 그래픽스 화면 상의 윈도우에서 모든 자료 입출력과 처리를 인터랙티브하게 실시간 처리한다. 따라서, 산점도 출력의 회전, 밀도 추정, Q-Q도에서 관측값들에 대한 브러싱, 모형에 대한 진단 통계량들의 그래픽스 처리가 가능하며, 다변량 자료 해석에 관련된 다양한 시각적 자료 분석 도구들을 제공한다. 그러나, SAS/INSIGHT는 DOS에서는 사용이 불가능하며 OS/2, 또는 Windows와 같은 운영체제 하에서만 운용되는 단점이 있다.

SAS/GRAPH는 완전한 WYSIWYG (What You See Is What You Get: 위지윅)을 구현하지 못한다. 위지윅이란 모니터 화면에 보이는 그대로 인쇄기로 출력되는 형식의 고급 그래픽스 소프트웨어가 구현하는 기능을 의미한다.

그러나, SAS/GRAPH가 완전한 WYSIWYG은 아니지만, 글자체나 크기 등은 모니터 화면에 나타나는 것과 인쇄기로의 출력이 동일하다. 단지, 모니터 화면에 나타나는 출력의 배치와 실제 인쇄한 결과가 반드시 일치하지 않는다. 다시 말해서 출력 장치에 따라 출력에 차이가 있을 수 있다.

이 점 역시 S-PLUS에 비하면 약점이라 할 수 있는데, 이렇게 된 이유는 SAS/GRAPH에서 지나치게 많은 출력 장치를 지원하고 있는데다가 PC 계열에서의 소위 표준형 출력 장치들을 무시하기 때문이다. 이는 SAS 시스템의 주된 시장이 대형 컴퓨터를 갖춘 기관들이지 PC를 몇 대 갖고있는 개인이 아닌데 근본적인 원인이 있다.

대학 교육의 관점에서 보면, S-PLUS나 SAS 시스템이나 입장일단이 있기 때문에 어느 것이 더 낫다고 단정지을 수 없다.

무엇보다도 S-PLUS를 설치하여 한 학급의 강의를 진행하려면 최소 4 MB의 메모리와 수치 보조 처리기가 장착된 80386 급 PC들과 레이저 프린터가 필요한데, 현재 상당수의 대학 여건에서 이 정도의 장비를 확보할 수 있을지 의문이다. 또한, 흔히 사용되는 14" 모니터로는 화면이 작아 명확히 관측점을 보기가 쉽지 않다. 통계 그래픽스 분석을 제대로 하려면 적어도 17" 이상의 모니터가 필요하나, 현재 대형 모니터의 가격은 지나치게 높다. 이 점에서 SAS 시스템은 비교적 저가의 장비로도 운용할 수 있으므로 S-PLUS에 비하여 유리하다.

S-PLUS에 컴퓨터 언어의 표준인 C와의 직접적인 접속 기능이 있는 것은 장점이라 하겠다. SAS 시스템에도 자체 개발된 SAS/C 컴파일러가 있으나, 일반 통계 분석에서 사용하려면 전문적인 지식이 있어야 한다.

제공되는 통계 분석 방법의 다양성을 보면 S-PLUS는 SAS 시스템에 비교가 되지 못한다. 그렇기 때문에, S-PLUS는 자체 언어, 즉, 명령어가 많지 않으며, SAS보다 사용법을 배우기가 더 쉽다.

문서 화면 위주의 DOS 하의 SAS 시스템은 6.04 판으로 끝나는 것으로 알려져있다. 그리고, 곧, Windows 하에서 구동되는 6.07 판이 보급되는데, 그때까지는, 컴퓨터 그래픽스 분석 기법에 관한 한 SAS/GRAPH가 S-PLUS에 비하여 기능상 미흡하고 결여된 점이 많다고 하겠다.

참고 문헌

- [1] 김성수, 한경수 (1992). S-PLUS의 소개 및 SAS와의 그래픽 비교.
- [2] Amstat News (1992). No. 191, American Statistical Association
- [3] Statistical Sciences Inc. *S-PLUS for DOS: User's manual* (1991).
- [4] SAS Institute Inc. *SAS/GRAPH User's guide* (1989). Release 6.03 edition.

SAS/GRAPH: Its Capabilities and Limitations

Sung, Nae Kyung¹⁾

Abstract

SAS/GRAPH is a part of the SAS System which generates information and presentation color graphics. It is able to import any SAS dataset from other statistical data analysis procedures and produce sophisticated graphics output. It also supports most output devices on the market and offers various tools enhancing graphics output. In this regard SAS/GRAPH outclasses its competitors. However, it does not support an interactive tool for data visualization and graphical data analysis. As far as interactive statistical graphics is concerned, SAS/GRAPH is behind in features and functions, compared to newly emerged statistical graphics software such as S-Plus.

¹⁾ Statistics Department, Ewha Womans University, Seoul 120-750, Korea
BITNET : NKSUNG@KREWHACC