

## S-PLUS와 XLISP-STAT의 다이나믹그래픽 기능

김철용<sup>1)</sup>, 서한손<sup>2)</sup>

### 요 약

컴퓨터 기능의 발전과 가격의 하락은 컴퓨터를 이용한 통계그래픽을 과거에는 상상할 수 없을 정도로 발전시키고 여러사람이 손쉽게 쓸 수 있게 되었다. 화면 위에서 직접 자료조작과 그 결과의 순간적 출력으로 요약되는 다이나믹 그래픽의 기능을 최근에 개발된 통계패키지인 S-PLUS와 XLISP-STAT에서 비교하여 보았다. XLISP-STAT은 아주 간단한 프로그램으로 여러 다이나믹 그래픽기능을 수행할 수 있다.

아직 초기단계에 있는 컴퓨터 통계그래픽에는 개선의 여지가 많고 새로운 방법을 창안할 수 있는 가능성도 있어 연구대상으로서도 그 전망이 매우 밝다고 생각된다.

### 1. 머리말

컴퓨터 hardware의 발전과 가격의 하락으로 컴퓨터를 이용한 통계그래픽이 가능하여 짐으로써 통계방법의 새로운 분야가 열리게 되었다. 그래프는 그 정보를 사람의 눈을 통하여 전달하므로 눈의 기능에 일차적으로 영향을 받게 되며 그래프의 내용이 인지되는 과정에서 이차적으로 왜곡되어 질 수 있다. 따라서 통계 그래픽에 관한 연구는 통계학적인 측면은 물론, anchoring, projection등 심리학적인 측면도 함께 연구되어야 할 분야이다 (Simkin 등, 1989; Hoaglin 등, 1989). 통계 그래픽을 통해 얻을 수 있는 대표적인 잇점은 기존의 통계방법에서는 행하기 힘든 탐색기능과 자료의 집약성을 가능하게 하는데 있다. 즉, 그래프의 이용으로 다른 방법으로는 쉽사리 알아볼 수 없는 자료의 구성형태를 알아 볼 수 있으며 (Chambers등, 1983; Tufte, 1983; Tufte, 1990), 그래프 안에 무수히 많은 자료를 집약하여 보여줌으로써 그 전체적인 특성을 한 눈에 파악하도록 할 수 있다 (Celeveland, 1985; Tufte, 1983; Greenacre 등, 1987; Stuetzle, 1987). 통계 그래픽의 탐색 기능이 급격하게 발전하게된 계기는 화면 위에서 자료를 직접 조작할 수 있고, 그 결과가 거의 동시에 화면 위로 나타나게 하는 다이나믹 그래픽 기법의 등장에 있다. S-PLUS에서도 몇 가지의 다이나믹 그래픽

<sup>1)</sup> 120-749 서울시 서대문구 신촌동 연세대학교 응용통계학과 조교수  
BITNET : CEKIM@KRYSUCC1

<sup>2)</sup> 200-701 강원도 원주군 흥업면 연세대학교 문리대학 통계학과 강사,  
BITNET : SEOHS@KRYSUCC1

기능이 있으나 대체적으로 이 분야의 연구는 아직 초기단계에 있는 관계로 개선의 여지가 많다고 할 수 있다.

## 2. 다이나믹 그래프

다이나믹 그래픽은 1970년대 후반에 시작하여 80년대에 컴퓨터 hardware의 발전과 더불어 널리 보급되기 시작했다 (Cleveland, 1987). 처음에는 main frame이나 workstation에서만 가능하던 것이 90년대에 들어와 PC에도 보급되었다 (Thisted, 1986; Workshop committee, 1986). 최근에 개발된 linking방법등 다이나믹 기법을 이용하여 여러 변수 사이의 관계를 탐색하는 것은 종래에는 상상할 수 없는 것이었으며, 과거의 일차원 또는 이차원의 한계를 극복하여 다차원으로의 문을 연 것이다 (Cleveland, 1987). S-PLUS에서 수행 가능한 다이나믹 그래프의 대표적인 예는 spinning기법과 brushing, labeling등이 있지만 이러한 기능은 일정한 범위내에서만 이루어질 수 있다(Cleveland 등, 1988; Huber, 1987). 즉 brushing은 scatter matrix에서만 수행이 가능하다(Carr 등, 1987). 그러나 S-PLUS에서는 불가능한 기능, 예를 들면, 어떠한 그래프 종류이든지 결합하여 수행이 가능하고 특히 자료의 변화를 화면에서 지시하고 그 결과를 즉시 화면에서 볼 수 있는 deleting, linking, slide bar등의 방법(Becker 등, 1987; Tierney, 1990; Weihs 등, 1990) 이 자료탐색에 새로운 영역을 제공하고 있다. XLISP-STAT은 이러한 면에서 가장 유용하게 사용될 수 있는 그래픽 언어라고 할 수 있다. XLISP-STAT은 S-PLUS와 마찬가지로 object-oriented 된 언어이지만 S-PLUS보다 더 graphic tool을 object화 할 수 있게 개발되어서 사용자가 보다 편리하고 자유롭게 이용할 수 있다.

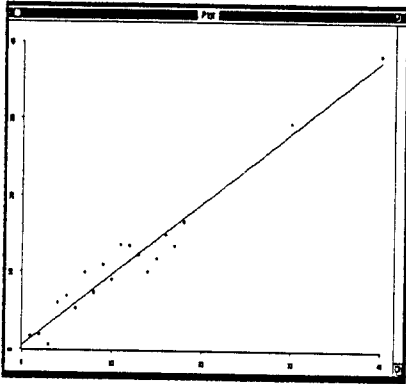
XLISP-STAT (Tierney, 1990; Lisp-Stat: Book Reviews, 1991)의 가장 큰 특징중의 하나는 graphical animation controller를 acting graphic window에 연관 시킴으로써 그래픽의 action parameter를 즉각적으로 변화시키고 그 결과를 눈으로 확인하는데 있다. graphical animation controller는 여러 형태로 역할을 수행할 수 있는데, 대표적으로는 화면상단에 설정한 매뉴를 통한 경우와 소위 slider라는 dialog box를 통한 경우, 그리고 마우스를 이용한 화면에서의 모수(parameter) 변화방법이 있다. 또한 이상의 방법은 혼합되어서 사용될 수 있다. 각각의 경우를 몇가지 예로 들면 다음과 같다.

### (1) 회귀선에서 influential case에 관한 예:

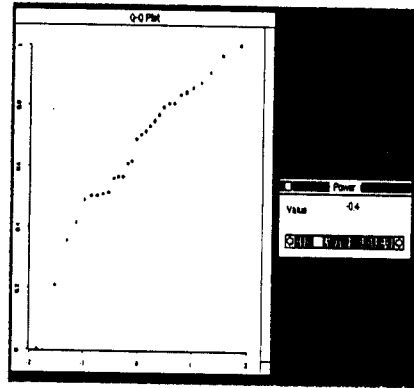
화면에서 각 data point를 이동시켜 변하게되는 회귀선을 직접 눈으로 확인함으로써 어떠한 data point가 얼마만큼 회귀선을 실제로 변화케 하는 지 알 수 있다(그림 1).

(2) Box-Cox변환에 관한 예:

화면에서 조그마한 window안의 값은 변환 승수(power)를 나타내고 scroll bar를 따라 마우스를 이동하여 이 값을 변화시키면 이 승수값에 의한 Box-Cox변환을 거친 자료의 Q-Q Plot이 그려지게 되어 정규성(normality)을 충족케 하는 승수값을 찾을 수 있다(그림 2).



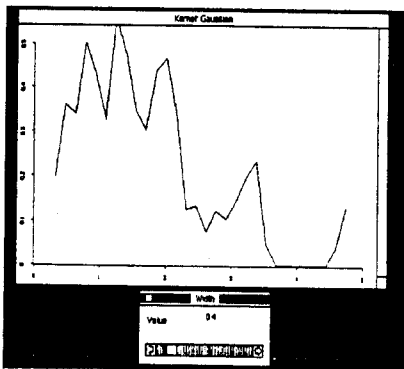
(그림 1)



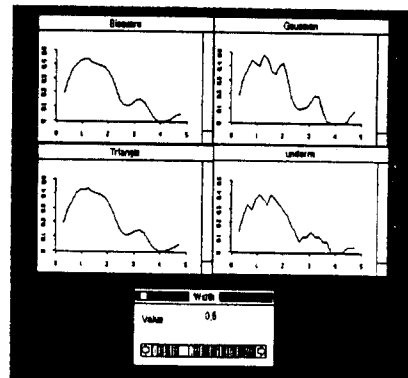
(그림 2)

(3) 밀도함수 추정에 관한 예:

화면은 일정한 커널함수를 정하고 bandwidth를 slider를 통해 변화시킬 때 달라지는 밀도함수 추정식을 나타내고 있다(그림 3). 만일 같은 bandwidth값에서 네개의 커널함수에 의해 추정된 밀도함수식을 모두 한 화면에 그리면, 커널함수에 따른 밀도함수 추정치의 영향을 비교할 수 있다 (그림 4).



(그림 3)



(그림 4)

이상의 예에서 사용되는 메뉴와 slider bar는 사용자 개인이 목적에 따라 만드는 것이므로 XLISP-STAT은 S-PLUS가 갖지 못하는 두 가지 장점을 가진다. 첫째, linking을 통하여 자유롭게 그래프의 결합을 추구할 수 있다. 둘째, 자료의 변화에 따른 그래픽 절차의 반복을 필요로 하지 않는다. 따라서 XLISP-STAT을 통하면 보다 즉각적이고 종합적인 그래픽 분석이 가능하게 된다.

### 3. 맺음말

S-PLUS는 다이내믹 그래픽의 기능이 제한적이지만 다른 다이내믹 그래픽 언어에 비해 편집기능등 많은 장점을 갖고있다. 실제로 사용될 통계그래픽 언어를 선택 할 때는 각종 통계 그래픽 언어의 장단점은 물론 hardware와 software의 구입과 더불어 기존 것의 maintenance, upgrade의 비용을 충당할 비용등 경제적인면도 고려 되어야 할 것이다(표 1). 연구 분야로서의 통계 그래픽은 일반적인 필요에 의한 개발보다는 구체적인 통계문제, 예를 들면, diagnostics등의 해결을 위한 시도가 요구된다. William F. Eddy(Lisp-Stat: Book reviews, 1991)가 예견했듯이 통계그래픽은 연구 분야로서도 미래가 밝다고 하겠다.

표 1. 다이내믹 그래픽 패키지의 비교

	S-Plus for Dos	SAS-DOS	SAS-Windows	Lisp-Stat
CPU	386+387	286	386	Mac SE30+ math
RAM	2MB (8MB)	640K	8MB	5MB
Disk space	15 ~ 25MB	13 ~ 30MB	90MB +	447KB
Graphic	EGA, VGA, Hercules	CGA, EGA, VGA	EGA, VGA	
Price	\$875 + \$150/year per unit	25 ~ 90만원 per site	250만원	free

Mathematica for DOS ...\$895, 3.5MB(8MB) RAM, 7MB disk space +16MB swap space  
 Mathematica for Windows ...\$995, 4MB(8MB) RAM, 7MB disk space +16MB swap space  
 Maple V ...\$695, 2.5MB RAM, 7MB disk space

## 참 고 문 헌

- [1] Becker, R. A., Cleveland, W. S., and Wilks A. R.(1987), "Dynamic graphics for data analysis," *Statistical Science*, vol.2, no.4.
- [2] Carr, D. B., Littlefield, R. J., Nicholson, W. L., and Littlefield, J. S. (1987), "Scatter matrix techniques for large N," *Journal of the American Statistical Association*, **82**, 424-436.
- [3] Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983), *Graphical Methods for Data Analysis* Wadsworth & Brooks/Cole Publishing Co.
- [4] Cleveland, W. S. (1985), *The Elements of Graphing Data*, Wadsworth Advanced Books and Software. (1992 2nd Ed.).
- [5] Cleveland, W. S.(1987), "Research in statistical graphics," *Journal of the American Statistical Association*, **82**, 419-423.
- [6] Cleveland, W. S., and Marylyn, E. M.(1988), *Dynamic Graphics For Statistics*, Wadsworth & Brooks/Cole Publishing Co.
- [7] Diaconis, Persi (1985), "Exploring data tables, trends, and shapes," *Theories of data analysis: From magical thinking through classical statistics*, edited by Hoaglin, D. C., Mosteller F. and Tukey, J. W., John Wiley & Sons. 1-36.
- [8] Greenacre, M. and Hastie, T. (1987), "The geometric interpretation of correspondence analysis," *Journal of the American Statistical Association*, **82**, 437-447.
- [9] Huber, P.(1987), "Experiences with three-dimensional scatterplots" *Journal of the American Statistical Association*, **82**, 448-453.
- [10] Lisp-Stat: Book reviews (1991). *Statistical Science*, vol.6, no.4.
- [11] Simkin, D., and Hastie T.(1987), "An information-processing analysis of graph perception" *Journal of the American Statistical Association*, **82**, 454-465.
- [12] Stuetzle, W. (1987), "Plot windows" *Journal of the American Statistical Association*, **82**, 466-475.
- [13] Thisted, R. A.(1986), "Computing environment," *Statistical Science*, vol.1, no.2.
- [14] Tierney, L.(1990), *Lisp-Stat: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics*, John Wiley & Sons.
- [15] Tufte, E. R.(1983), *The Visual Display of Quantitative Information*, Graphics Press.
- [16] Tufte, E. R.(1990), *Erwisioning Information*, Graphics Press.
- [17] Weihs, C. and Schmidli, H.(1990), "Online multivariate exploratory graphical analysis," *Statistical Science*, vol.5, no.2.
- [18] Workshop Committee(1986), "Computer in statistical research," *Statistical Science*, vol.1, no.4.

## Dynamic Graphic Features in S-PLUS and XLISP-STAT

Chul Eung Kim<sup>1)</sup>, Han Son Seo<sup>2)</sup>

### Abstract

The increase in computing power and the decrease in price of computers has enabled statistical computer graphics to progress tremendously in recent years. Many people can now access to the newly developed computer graphical methods easily. The direct manipulation on screen and the simultaneous realization of the results are two main ingredients of dynamic graphics. We compare the dynamic graphical features in two relatively new packages; SPLUS and XLISP-STAT. XLISP-STAT is very lean packed with powerful dynamic graphical tools.

The statistical computer graphics, being still in the state of infancy, has a lot of room to grow, and is a new research area with a great potential.

---

<sup>1)</sup> Assistant Professor, Department of Applied Statistics, Yonsei University,  
Seoul, 120-749  
BITNET : CEKIM@KRYSUCC1

<sup>2)</sup> Lecturer, Department of Statistics, Yonsei University, Wonjoo, Kangwon-Do,  
200-701  
BITNET : SEOHS@KRYSUCC1