

인터넷 환경에서 븍스트랩 통계 시스템의 개발

- Development of Web-based Statistical System for
Bootstrap on the Internet Environment -

최 성 운 *

Choi Sung Woon

임 인 섭 *

Lim In Sup

Abstract

Recently, growth of internet causes rapid changes in many areas of statistics such as statistical computation and education. Especially, bootstrap is the most interesting statistical methods applying computer resampling simulation. In this study, we try to present how to use a method of bootstrap on the internet. We also develop to user a statistical system which is programed with java applet for user to handle easily.

Keyword : Bootstrap, Java Applet, Internet

1. 서 론

최근 컴퓨터의 발달로 인하여, 특히 인터넷의 발달로 인하여 우리의 생활환경을 바꾸어 놓고 있다. 이러한 발달은 통계학 강의에서도 많은 변화를 가져왔다. 교재와 칠판, 그리고 전자계산기로 진행하던 전통적인 통계학 강의를 컴퓨터와 인터넷을 이용하여 효과적인 통계교육으로 변화를 가져왔다. 기존의 통계분석 패키지인 SAS, S-PLUS, SPSS, MINITAB 등과 같은 패키지들도 이런 변화에 맞춰 다양한 분석과 분석결과를 제시하고 있으나, 일반적으로 통계학을 배우는 학생들의 입장에서는 이런 패키지를 이용한 분석은 그 비용이 매우 많이 들어 이를 활용한 통계학의 교육에는 적합하다고 할 수 없다.

또한 SAS나 S-PLUS와 같은 통계 패키지들은 웹상에서 사용자가 손쉽게 쓸 수 있는

* 경원대학교 산업공학과

프로그램을 개발하여 시판하고 있으나, 이들은 기본적으로 사용자가 해당 소프트웨어를 가지고 있어야 사용할 수 있는 CGI(Common Gateway Interface) 프로그램이며 여전히 해당 소프트웨어를 구입해야 가능하다는 문제점을 가지고 있다.[6]

현재 기존하는 대부분의 통계이론은 계산 능력이 미미했던 시절에 나온 이론들이다. 컴퓨터의 발달로 인하여 계산처리 속도가 빠른 능력을 이용한 여러 가지 기법들이 나왔으며, 그런 통계방법 중 봇스트랩의 방법은 최근 가장 주목을 받고 있는 방법 중 하나이다.

본 연구에서는 컴퓨터의 빠른 계산능력을 이용한 봇스트랩의 방법을 인터넷을 이용하여 다수의 사용자들이 실시간으로 쉽게 접근할 수 있는 방법을 제시하고자 한다. 봇스트랩과 같은 방법을 사용자가 이용하기 위해서는 고가의 통계 패키지를 구입을 해야 가능한 기존의 방법에서, 누구나 쉽게 접근하고 그 사용이 편리한 프로그램을 개발하여 실시간으로 데이터 분석이 가능하고, 능동적으로 통계 방법을 이용할 수 있는 시스템을 구축하고자 한다.

2. 통계 시스템

2.1 통계 시스템의 형태

통계 시스템의 형태는 여러 형태로 나타날 수 있다. 학습 내용을 위주로 한 시스템, 통계분석을 위주로 한 시스템, 그리고 학습과 통계분석을 위주로 한 시스템 등으로 나눌 수 있다. 학습과 통계분석을 위한 시스템은 통계분석 도구에 의한 분류와 통계패키지의 인터페이스에 의한 분류로 나눌 수 있다.

통계분석 도구에 의한 분류는 통계패키지를 이용한 방법과 프로그래밍 언어에 의한 방법 등으로 나눌 수 있다. 통계 패키지를 이용한 통계분석은 인터넷상에서 인터페이스 기법을 활용하여 구현할 수 있다. 서버인터페이스의 경우 통계 학습시스템은 분석하고자 하는 데이터를 클라이언트에서 제시하고 서버에 설치된 통계패키지에 의하여 통계분석을 하여 그 결과를 클라이언트에 전송하는 시스템이며, 클라이언트 인터페이스에 의한 통계학습 시스템은 클라이언트에서 입력한 데이터는 웹 페이지를 거쳐서 클라이언트로 재전송되어 클라이언트에 설치된 통계패키지에 의하여 통계분석을 하는 시스템이다.[2]

프로그래밍 언어에 의한 방법은 기존에 알려진 언어(C, C++, Visual Basic, Fortran IMSL)이나 웹 언어(ASP, PHP, JSP)에 내장된 수학 함수를 이용하여 클라이언트에서 제시한 자료를 가지고 서버에서 연산하여 그 결과를 클라이언트에 제시하는 방법과 Java와 같은 언어를 이용하여 자료의 입력과 연산, 결과의 출력이 클라이언트에서 작동하는 방법이 있다. Java의 경우 클라이언트에서 모든 작업이 수행되므로 서버의 부담이 적은 장점이 있으며, 반복되는 작업에 있어서 작업속도가 빠른 장점이 있으나 최초의 연산을 위한 애플리 실행과정에서 작업속도가 느리다는 단점이 있다.

2.2 통계시스템의 현황

통계 시스템은 현재에도 활발하게 개발이 진행 중이다. XLisp-Stat은 객체지향적인 모듈과 동적 프로그래밍을 구사할 수 있는 공개 소프트웨어로 XLisp-Stat을 이용하여 인터넷상에서 통계시스템을 구현한 사례는 Garcia[8], Leeuw[9], Minnesota 대학 등에서 찾아볼 수 있다.

UCLA의 Leeuw J. D.[9]와 Bond J.는 XLisp-Stat를 사용하여 프로그램으로 작성된 파일을 CGI 프로그램 안에 포함시켜 인터넷상에서 이용할 수 있도록 연구를 하였다. 서버에 설치된 XLisp-Stat를 사용하여 클라이언트에서 데이터의 분석이 가능하도록 하였다.

Minnesota 대학의 Regression Graphics 홈 페이지에서는 XLisp-Stat의 사용 방법을 인터넷을 통하여 이용하도록 하고 있으며, 회귀분석에 대한 소개와 함께 클라이언트에 설치된 응용 프로그램을 통하여 동적인 그래프를 살펴 볼 수 있도록 하였다.

York 대학에서는 SAS를 사용하는 교육과정에서 HTML 문서를 만들고 서버의 SAS 파일과 연결되도록 구축하여 인터넷에서 공유할 수 있도록 구현하고 있다.[5] 또한 West, Ogden과 Rossini[10]는 인터넷에서의 통계 소프트웨어 활용 방안을 논의하였으며, CGI 기법과 Java를 사용하여 개발한 통계 자료 분석 예제들을 URL1에서 찾아 볼 수 있다.[3]

국내의 경우 한경수, 안정용, 강윤비[4]는 Java를 이용하여 통계학 개론을 교육할 수 있는 프로그램(URL2)을 개발하였으며, 전북대학교 수학 · 통계정보과학부 Scientific & Statistical Computing Laboratory 홈페이지에서(URL3)는 Microsoft의 분산 컴퓨팅 모델인 COM 기술을 사용하여 개발한 소프트웨어를 사용할 수 있다. 또한 인터넷에서 통계 프로그램의 사용하기 위한 유료화된 서버는 다수가 개발되었으며, 계속 증가하는 추세로 여겨진다.

3. 븋스트랩 통계시스템 구축

3.1 븋스트랩의 소개

어떤 모분포 F 로부터 크기가 n 인 무작위 표본 $X=(X_1, X_2, \dots, X_n)$ 을 얻어서 표본평균 \bar{X} 으로 추정한다고 하면 가장 문제가 되는 것은 통계적 정확도에 관한 것이다. 표본평균의 정확도는 통상적으로 표준오차에 의해 설명되며, 표본평균의 의미는 바로 표준오차에 의하여 결정된다. 표준오차가 작으면 표본평균은 추정량으로서의 의미를 가지며 그 활용도가 높게 되고, 표준오차가 크면 통계적으로 표본평균값은 의미가 적어진다고 할 수 있다.

Efron B.[7] 의해 제안된 븋스트랩은 $X=(X_1, X_2, \dots, X_n)$ 로 얻어진 자료로부터 복원랜덤추출을 하여 븋스트랩 표본 X^* 을 얻는 것이다. 즉 $P(x_{i^*}=x_j)=1/15$, $i,j=1,2,\dots,n$ 의

확률로 복원추출을 하여 그 표본을 얻는 것이라 할 수 있다.

실제로 븗스트랩은 그리 복잡한 개념이 아니며 잭나이프(jackknife)등을 포함한 비슷한 방법들은 이미 오래 전부터 있었으나, 븗스트랩은 수식을 표현하기에 너무 복잡하여 실용성이 없다고 여겨졌다. 그러나 컴퓨터 능력의 발달로 인하여 븗스트랩의 근사해가 사용가능하며 여러 가지 통계적인 잇점이 있다는 사실을 밝혀냈다.

이와 같은 븗스트랩 방법을 실제로 사용하는 데는 이론적인 정당성 문제가 제기된다. 이론적인 정당성은 ‘일치성(Consistency)’, ‘최적성(Optimality)’등이 있으나 본 연구에서는 븗스트랩 방법에 관한 연구보다는 이를 사용하여 인터넷에서 통계 시스템을 구축하는데 있으므로 생략하기로 한다.[1]

3.2 통계 시스템의 구조

인터넷을 기반으로 하는 계산 환경은 크게 CGI 기법, Java, 컴포넌트 기술 등으로 크게 나눌 수 있다. 사용자는 복잡한 통계전문 소프트웨어를 이용하지 않고도 그들의 데이터를 웹 브라우저에서 입력하여 바로 통계분석 결과를 얻을 수 있다. CGI 기법은 기존에 널리 알려진 언어를 사용하기 때문에 쉽게 시스템을 구축 할 수 있는 장점이 있으나 서버에서 프로그램이 작동하는 방식이므로 여러 사용자가 동시에 접속하는 경우에는 과부하가 발생하여 서버가 다운되거나 클라이언트의 요청만큼 프로그램이 로드가 되어야 하므로 처리 속도가 현저히 떨어지는 문제가 생긴다. 또한 모듈의 재사용이 어려워 서버에 과중한 부담을 주게 되어 고성능 서버의 낭비 요인이 된다. 물론 이런 문제점은 컴퓨터의 발달로 인하여 많이 줄어들었지만 여전히 문제점으로 지적되고 있다.

Java를 기반으로 개발을 하면 운영 차제에 무관하다는 장점을 가진다. Java는 자바 가상머신(Java Virtual Machine)이 자바 바이코드(Java Byte Code)를 실행하는 형태이므로 그 운영체계에 상관없이 독립적으로 실행된다. 하지만 Java로 개발된 모듈은 사용자의 컴퓨터에 설치된 후 사용되므로 수행을 위한 부가적인 시간이 요구되며, 사용자가 보유한 컴퓨터의 성능이 좋지 않다면 계산 성능이 떨어진다는 문제점을 가지고 있다. 하지만 한번 설치된 후에는 그 시간이 매우 빠르며 이것 역시 컴퓨터의 발달로 인하여 많이 줄어들었다고 할 수 있다. 또한 다른 언어로 개발된 그래프나 수치 계산 라이브러리 등을 사용하기 어려워 개발 시간이 오래 걸리는 단점이 있다.

컴포넌트 기술은 하나의 응용 소프트웨어를 개발할 때 이미 개발된 다양한 형태의 소프트웨어 모듈을 블록을 쌓듯이 조립하여 빠르고 쉽게 개발할 수 있게 한다. 이러한 기법은 사용하는 각각의 모듈이 이미 검증된 것이므로 안정적이고 효율적이라 할 수 있다. 특히 Microsoft가 제안한 Microsoft COM은 객체기반 기술인 OLE(Object Linkage and Embedding)를 분산 컴퓨터 환경에 적합하도록 확장시킨 것으로 Windows 환경에서 개발되어 현재 UNIX 환경에서도 운용 가능하도록 개발 중에 있다. Microsoft Windows 자체가 COM 컴포넌트들로 구성된 하나의 응용 프로그램으로 생각할 수 있을 정도로 많은 부분들이 COM을 지원하는 객체 형태로 이루어져 있어 개발하는데 편리성이 있으며 COM 객체의 위치가 클라이언트 컴퓨터, 로컬 서버 또는

원격 서버에 관계없이 동일한 참조 환경을 제공하는 특성을 가진다. 위에서 언급한 개발 환경들의 기술들을 정리하면 < 표 1 >과 같다.[3]

< 표 1 > CGI, Java, 컴포넌트 기술의 비교

	CGI	Java	컴포넌트 기술
플랫폼 의존도	의존적	독립적	기술에 따라 다름
수행 위치	웹 서버	클라이언트	조절 가능
재 사용성	소스의 재사용	클래스의 재사용	객체의 재사용
과부화 가능성	사용자 · 서비스 증가 시	없음	없음

3.3 프로그램의 개발내용

본 연구에서는 다양한 기법 중에서 Java Applet을 이용한 방법을 선택했다. Java Applet을 이용하면 일단 사용자의 컴퓨터 환경에 상관없이 작동한다는 장점을 지녔다. 또한 네트워크의 부하가 없으며, 최근 문제가 되고 있는 보안에도 장점을 보인다. Applet은 서버에서 배포되어 임의의 클라이언트에서 동작하기 때문에 클라이언트의 정보 보호에 강력한 보안성을 가지며, HTML를 제어할 수 있고 반대로 HTML의 스크립트를 사용하여 Applet 메소드를 호출하기가 편리하다는 장점을 가진다.

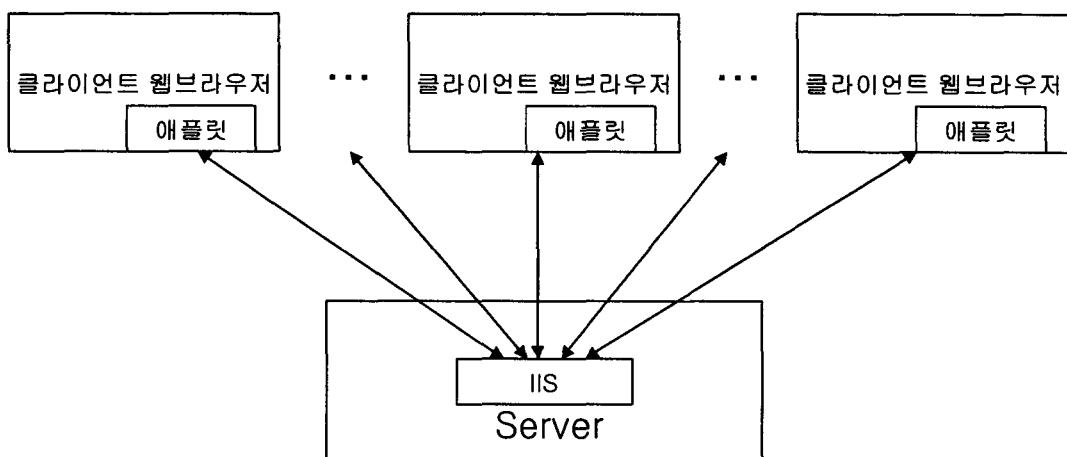
통계 자료를 분석하려는 사용자가 웹 브라우저에 접속하면 Java Applet이 클라이언트에서 작동하여 그 Applet위에 데이터만 입력하면 평균값과 표준편차 값이 연산되며, Resampling 하고자 하는 데이터의 수를 입력하여 봇스트랩을 선택하면 증가된 데이터를 이용하여 평균값과 표준편차 값이 Applet에 출력되게 프로그램을 작성하였다.

실제로 봇스트랩을 이용한 연산을 하기 위해서는 컴퓨터 프로그램을 이용하지 않으면, 낭비된 시간과 노력을 투자해야 하며, 현실적으로는 불가능하다고 할 수 있으며, 전문적인 통계 패키지가 아니면 봇스트랩을 이용한 연산은 불가능하다.

본 연구에서는 개발된 예제는 누구나 웹 브라우저에 자신의 자료만 입력하면 봇스트랩 기법을 이용하여 평균값과 표준편차 값을 얻을 수 있다. 사용자는 별도의 통계 프로그램을 설치할 필요가 없으며, 단지 어떤 종류의 컴퓨터라도 웹 서버에 접속할 수 있는 브라우저만 있으면 원하는 연산을 할 수 있다. 사용자의 컴퓨터 사용이 매우 낮을 경우 Applet이 초기화 되는 시간이 오래 걸리는 단점이 있으나, 최근 컴퓨터의 발달과 초고속 인터넷 환경에서는 문제가 되지 않는다.

개발한 예제의 구조는 < 그림 1 >과 같이 통계 계산에 필요한 Applet 소스를 HTML 코드에 삽입하여 Window 2000 Server에 있는 IIS(Internet Information Server)에 설치하였다. 인터넷 사용자가 브라우저를 열게 되면 IIS에서 Html 페이지

를 클라이언트에 보여주고, 이 과정에서 Applet 프로그램은 HTML 페이지와 함께 클라이언트에 전송되며, 전송이 완료되면 클라이언트에서 사용자가 요청한 데이터를 연산하고 그 결과를 보여주게 된다.



< 그림 1 > 개발된 프로그램의 구조

향후 봇스트랩을 이용한 추가적인 연산이 필요한 경우 기존에 존재하고 있는 Applet 소스에서 제작된 함수를 재사용함으로서 프로그램을 변경할 수 있다. Java의 장점인 Class의 재사용을 이용하면 새로운 프로그램을 작성할 경우 그 적용이 매우 용이하다.

사용한 프로그램 개발 도구와 운영환경은 < 표 2 >와 같다.

< 표 2 > 프로그램 개발에 사용된 도구 및 환경

개발 도구 및 환경	담당부분
Java(J2sdk-1_3_1)	Applet 제작 및 통계 계산 알고리즘 작성
HTML(Hypertext markup language)	클라이언트의 인터페이스와 스크립트를 이용한 Applet 연동
Microsoft Windows 2000 Server	웹 서버를 가동하기 위한 운영체제
Microsoft Internet Information 2000 Server	웹 서버의 운영체제

Html에서 Applet을 사용하기 위해서는 컴파일된 Class파일만 존재하면 되므로 프로그램을 작성하고 컴파일한 파일을 스크립트에 삽입하는 형태로 Html을 작성하였다.

Applet을 보기 위해서는 브라우저에 Java VM이 내장되어 있어야만 한다. 기존의 웹 브라우저에는 Java VM을 내장시켜서 개발하였으나 최근의 웹 브라우저는 Java VM을 여러 가지 복잡한 이유로 인하여 JDK.3x를 내장 Java VM에서 지원하고 있지 않다. 이는 위에서 언급한 클라이언트의 운영체계 및 환경에 상관없이 작동한다고 한 플랫폼 독립성이 보장할 수 없게 된다. 이럴 경우 사용자가 정상적으로 Java Applet을 사용하기 위해서는 브라우저에 VM이 없을 경우 직접 VM을 다운 받아서 실행해야 한다는 번거로움을 가지게 된다. 이를 해결하고자 본 연구에는 자바 폼페이지에서 제공하는 Html 변환기(Html Converter)를 사용하여 해결하였다. 이를 통하여 사용자가 일일이 VM을 다운 받아야 하는 번거로움을 줄이고 플랫폼 독립성을 보장할 수 있게 해결하였다.

*붓스트랩활용예제 1

*붓스트랩활용예제 2

* 자료의 입력시 . , * 자료를 구분합니다.

자료입력	3.5.7.8.9
평균표준편차	
평균	6.4
표준편차	2.4063189157584592
붓스트랩수	100
Bootstrap 평균표준편차	
평균	6.39
표준편차	2.1544277476385003
Clear	

< 그림 3 > 사용자의 입력을 받아 결과가 나오는 Applet 화면

< 그림 3 >은 사용자가 웹 브라우저로 접속을 하면 클라이언트에서 Applet이 작동하는 예제이다. 위의 그림에서와 같이 원하는 데이터를 입력하면 평균값과 표준편차 값을 구할 수 있으며, 붓스트랩의 기법을 이용하기 위하여 Resampling 하고자 하는 데이터의 수를 입력하면 그 데이터를 이용하여 연산된 평균값과 표준편차 값을 얻을 수 있으며, 그 차이를 쉽게 비교 할 수 있다.

*봇스트랩활용예제 1

*봇스트랩활용예제 2

자료입력	3,5,7,8,9,5,4,5,2,3,6,9,5,4,1,2,3.																																																																																																																																																																																																																																																
	<input type="button" value="평균및표준편차"/>																																																																																																																																																																																																																																																
평균	4.764705882352941																																																																																																																																																																																																																																																
표준편차	2.411675034200967																																																																																																																																																																																																																																																
봇스트랩수	1000																																																																																																																																																																																																																																																
	<input type="button" value="Bootstrap후 평균및표준편차"/>																																																																																																																																																																																																																																																
평균	4.718																																																																																																																																																																																																																																																
표준편차	2.3528731398041933																																																																																																																																																																																																																																																
<table border="1"> <tr><td>5.0</td><td>9.0</td><td>5.0</td><td>4.0</td><td>9.0</td><td>4.0</td><td>3.0</td><td>9.0</td><td>1.0</td><td>9.0</td><td>4.0</td><td>4.0</td><td>3.0</td><td>4.0</td><td>2.0</td></tr> <tr><td>9.0</td><td>2.0</td><td>3.0</td><td>5.0</td><td>5.0</td><td>2.0</td><td>3.0</td><td>9.0</td><td>3.0</td><td>5.0</td><td>3.0</td><td>4.0</td><td>5.0</td><td>2.0</td><td>9.0</td></tr> <tr><td>5.0</td><td>1.0</td><td>1.0</td><td>5.0</td><td>5.0</td><td>6.0</td><td>7.0</td><td>5.0</td><td>3.0</td><td>9.0</td><td>9.0</td><td>2.0</td><td>3.0</td><td>5.0</td><td>9.0</td></tr> <tr><td>1.0</td><td>3.0</td><td>4.0</td><td>3.0</td><td>3.0</td><td>5.0</td><td>2.0</td><td>4.0</td><td>6.0</td><td>9.0</td><td>2.0</td><td>5.0</td><td>2.0</td><td>8.0</td><td>3.0</td></tr> <tr><td>1.0</td><td>2.0</td><td>5.0</td><td>5.0</td><td>7.0</td><td>4.0</td><td>5.0</td><td>3.0</td><td>5.0</td><td>3.0</td><td>9.0</td><td>9.0</td><td>3.0</td><td>5.0</td><td>5.0</td></tr> <tr><td>7.0</td><td>9.0</td><td>7.0</td><td>8.0</td><td>1.0</td><td>3.0</td><td>9.0</td><td>4.0</td><td>5.0</td><td>9.0</td><td>8.0</td><td>8.0</td><td>5.0</td><td>2.0</td><td>2.0</td></tr> <tr><td>5.0</td><td>4.0</td><td>5.0</td><td>3.0</td><td>1.0</td><td>3.0</td><td>1.0</td><td>2.0</td><td>5.0</td><td>1.0</td><td>3.0</td><td>6.0</td><td>6.0</td><td>4.0</td><td>3.0</td></tr> <tr><td>3.0</td><td>3.0</td><td>2.0</td><td>5.0</td><td>5.0</td><td>4.0</td><td>2.0</td><td>2.0</td><td>2.0</td><td>5.0</td><td>5.0</td><td>2.0</td><td>9.0</td><td>7.0</td><td>6.0</td></tr> <tr><td>2.0</td><td>9.0</td><td>6.0</td><td>3.0</td><td>8.0</td><td>4.0</td><td>9.0</td><td>1.0</td><td>7.0</td><td>1.0</td><td>3.0</td><td>3.0</td><td>7.0</td><td>5.0</td><td>5.0</td></tr> <tr><td>5.0</td><td>5.0</td><td>2.0</td><td>3.0</td><td>3.0</td><td>5.0</td><td>4.0</td><td>6.0</td><td>5.0</td><td>3.0</td><td>3.0</td><td>5.0</td><td>4.0</td><td>3.0</td><td>4.0</td></tr> <tr><td>5.0</td><td>5.0</td><td>7.0</td><td>4.0</td><td>4.0</td><td>5.0</td><td>2.0</td><td>9.0</td><td>5.0</td><td>9.0</td><td>4.0</td><td>2.0</td><td>3.0</td><td>9.0</td><td>3.0</td></tr> <tr><td>5.0</td><td>3.0</td><td>5.0</td><td>6.0</td><td>3.0</td><td>5.0</td><td>3.0</td><td>1.0</td><td>9.0</td><td>5.0</td><td>2.0</td><td>1.0</td><td>9.0</td><td>3.0</td><td>2.0</td></tr> <tr><td>2.0</td><td>9.0</td><td>4.0</td><td>3.0</td><td>2.0</td><td>3.0</td><td>5.0</td><td>4.0</td><td>3.0</td><td>2.0</td><td>4.0</td><td>9.0</td><td>9.0</td><td>1.0</td><td>2.0</td></tr> <tr><td>4.0</td><td>1.0</td><td>1.0</td><td>5.0</td><td>2.0</td><td>1.0</td><td>7.0</td><td>3.0</td><td>5.0</td><td>1.0</td><td>7.0</td><td>2.0</td><td>2.0</td><td>8.0</td><td>7.0</td></tr> <tr><td>3.0</td><td>9.0</td><td>8.0</td><td>8.0</td><td>5.0</td><td>3.0</td><td>2.0</td><td>3.0</td><td>3.0</td><td>6.0</td><td>5.0</td><td>3.0</td><td>5.0</td><td>2.0</td><td>5.0</td></tr> <tr><td>3.0</td><td>2.0</td><td>8.0</td><td>5.0</td><td>9.0</td><td>1.0</td><td>9.0</td><td>4.0</td><td>5.0</td><td></td><td></td><td></td><td></td><td></td><td></td></tr> </table>		5.0	9.0	5.0	4.0	9.0	4.0	3.0	9.0	1.0	9.0	4.0	4.0	3.0	4.0	2.0	9.0	2.0	3.0	5.0	5.0	2.0	3.0	9.0	3.0	5.0	3.0	4.0	5.0	2.0	9.0	5.0	1.0	1.0	5.0	5.0	6.0	7.0	5.0	3.0	9.0	9.0	2.0	3.0	5.0	9.0	1.0	3.0	4.0	3.0	3.0	5.0	2.0	4.0	6.0	9.0	2.0	5.0	2.0	8.0	3.0	1.0	2.0	5.0	5.0	7.0	4.0	5.0	3.0	5.0	3.0	9.0	9.0	3.0	5.0	5.0	7.0	9.0	7.0	8.0	1.0	3.0	9.0	4.0	5.0	9.0	8.0	8.0	5.0	2.0	2.0	5.0	4.0	5.0	3.0	1.0	3.0	1.0	2.0	5.0	1.0	3.0	6.0	6.0	4.0	3.0	3.0	3.0	2.0	5.0	5.0	4.0	2.0	2.0	2.0	5.0	5.0	2.0	9.0	7.0	6.0	2.0	9.0	6.0	3.0	8.0	4.0	9.0	1.0	7.0	1.0	3.0	3.0	7.0	5.0	5.0	5.0	5.0	2.0	3.0	3.0	5.0	4.0	6.0	5.0	3.0	3.0	5.0	4.0	3.0	4.0	5.0	5.0	7.0	4.0	4.0	5.0	2.0	9.0	5.0	9.0	4.0	2.0	3.0	9.0	3.0	5.0	3.0	5.0	6.0	3.0	5.0	3.0	1.0	9.0	5.0	2.0	1.0	9.0	3.0	2.0	2.0	9.0	4.0	3.0	2.0	3.0	5.0	4.0	3.0	2.0	4.0	9.0	9.0	1.0	2.0	4.0	1.0	1.0	5.0	2.0	1.0	7.0	3.0	5.0	1.0	7.0	2.0	2.0	8.0	7.0	3.0	9.0	8.0	8.0	5.0	3.0	2.0	3.0	3.0	6.0	5.0	3.0	5.0	2.0	5.0	3.0	2.0	8.0	5.0	9.0	1.0	9.0	4.0	5.0						
5.0	9.0	5.0	4.0	9.0	4.0	3.0	9.0	1.0	9.0	4.0	4.0	3.0	4.0	2.0																																																																																																																																																																																																																																			
9.0	2.0	3.0	5.0	5.0	2.0	3.0	9.0	3.0	5.0	3.0	4.0	5.0	2.0	9.0																																																																																																																																																																																																																																			
5.0	1.0	1.0	5.0	5.0	6.0	7.0	5.0	3.0	9.0	9.0	2.0	3.0	5.0	9.0																																																																																																																																																																																																																																			
1.0	3.0	4.0	3.0	3.0	5.0	2.0	4.0	6.0	9.0	2.0	5.0	2.0	8.0	3.0																																																																																																																																																																																																																																			
1.0	2.0	5.0	5.0	7.0	4.0	5.0	3.0	5.0	3.0	9.0	9.0	3.0	5.0	5.0																																																																																																																																																																																																																																			
7.0	9.0	7.0	8.0	1.0	3.0	9.0	4.0	5.0	9.0	8.0	8.0	5.0	2.0	2.0																																																																																																																																																																																																																																			
5.0	4.0	5.0	3.0	1.0	3.0	1.0	2.0	5.0	1.0	3.0	6.0	6.0	4.0	3.0																																																																																																																																																																																																																																			
3.0	3.0	2.0	5.0	5.0	4.0	2.0	2.0	2.0	5.0	5.0	2.0	9.0	7.0	6.0																																																																																																																																																																																																																																			
2.0	9.0	6.0	3.0	8.0	4.0	9.0	1.0	7.0	1.0	3.0	3.0	7.0	5.0	5.0																																																																																																																																																																																																																																			
5.0	5.0	2.0	3.0	3.0	5.0	4.0	6.0	5.0	3.0	3.0	5.0	4.0	3.0	4.0																																																																																																																																																																																																																																			
5.0	5.0	7.0	4.0	4.0	5.0	2.0	9.0	5.0	9.0	4.0	2.0	3.0	9.0	3.0																																																																																																																																																																																																																																			
5.0	3.0	5.0	6.0	3.0	5.0	3.0	1.0	9.0	5.0	2.0	1.0	9.0	3.0	2.0																																																																																																																																																																																																																																			
2.0	9.0	4.0	3.0	2.0	3.0	5.0	4.0	3.0	2.0	4.0	9.0	9.0	1.0	2.0																																																																																																																																																																																																																																			
4.0	1.0	1.0	5.0	2.0	1.0	7.0	3.0	5.0	1.0	7.0	2.0	2.0	8.0	7.0																																																																																																																																																																																																																																			
3.0	9.0	8.0	8.0	5.0	3.0	2.0	3.0	3.0	6.0	5.0	3.0	5.0	2.0	5.0																																																																																																																																																																																																																																			
3.0	2.0	8.0	5.0	9.0	1.0	9.0	4.0	5.0																																																																																																																																																																																																																																									
<input type="button" value="Clear"/>																																																																																																																																																																																																																																																	

< 그림 4 > 봇스트랩의 데이터를 출력하는 화면

< 그림 4 >은 봇스트랩 방법을 이용하여 얻어지는 데이터를 보여주는 화면이다. 화면에 데이터를 보여주는 시간이 오래 걸리므로 데이터를 원하지 않는 경우에는 이 화면을 선택하지 않는 것이 좋다.

소프트웨어 개발 과정을 단계별로 간략히 설명하면 다음과 같다.

1. Java Applet을 사용하여 데이터 및 봇스트랩의 수를 입력받는 GUI를 만든다.
2. 연산을 위한 평균 및 표준편차, 봇스트랩 함수를 작성한다.
3. Html 페이지를 디자인하고 Java Applet 삽입한다.
4. 웹 서버에 Html페이지를 등록한다.

4. 결 론

본 연구에서는 컴퓨터의 계산능력이 요구되는 봇스트랩 방법을 인터넷을 통하여 쉽게 접근하고 사용하기 편리하며 실시간으로 분석이 가능한 통계 시스템을 구축하였다. 봇스트랩 방법은 컴퓨터의 계산능력이 없이는 비효율적인 방법이라 할 수 있다. 또한 값비싼 통계 패키지를 이용하여야 그 연산이 가능하였다. 이와 같은 방법을 Java Applet을 이용하여 웹 브라우저를 통하여 누구나 접근이 가능하고 봇스트랩의 방법을 알지 못해도 사용이 가능하도록 구현을 하였다.

봇스트랩 방법에 의한 모수의 신뢰구간이나 표준편차의 추정 등의 다양한 봇스트랩 분석이 있으나, 본 연구에서는 적절한 추축통계량의 표본을 봇스트랩 방법으로 얻어진 표본을 이용하는 데 중점을 두었다. 이런 봇스트랩 방법에서 언급해야하는 주요 문제들로는 봇스트랩 방법의 최적성, 봇스트랩 표본의 크기에 관한 문제, 다변량 자료 분석에의 적용 등등의 다수가 있다.

본 연구에서는 봇스트랩 방법보다는 그 방법을 인터넷을 이용하여 활용하는 점에 중점을 두었다. 또한 Java Applet을 이용하여 손쉽게 접근하고 활용할 수 있는 시스템을 제시하는데 그 의의가 있다고 할 수 있다.

추후 연구에는 위에서 언급한 다양한 문제들을 적용한 분석을 통계 학습시스템에 적용하는 것과 Applet의 예외 처리, 속도 향상 문제와 다양한 통계 분석이 연구하는 것이 향후연구과제로 남아있다.

본 연구에서 제시된 소프트웨어는 URL (<http://mis.kyungwon.ac.kr/SD/boot.htm>)에서 확인할 수 있으며, 다양한 통계분석이 가능하도록 지속적인 개발을 계획중이다.

5. 참 고 문 헌

- [1] 전명식, “통계적 데이터 분석방법을 위한 컴퓨터의 활용 I : 봇스트랩 이론과 응용”, 응용통계연구 pp 121 - 140, 한국통계학회, 1990.
- [2] 정남철, “인터넷 상에서 CGI와 Xlisp-Stat를 이용한 통계학습시스템의 구축”, 한국정보처리학회 논문지 Vol. 5, NO. 12, 한국정보처리학회, 1998. 12.
- [3] 한경수, 김석기, 최숙희 “On the Developnet of Statistical Software Using Microsoft COM Technology on the Internet Environment”, 한국통계학회 추계학술 발표 논문집, Vol. 6, No. 2, pp 533 - 542, 1999
- [4] 한경수, 안정용, 강윤비, “통계학 교육을 위한 전자교재의 활용”, 응용통계 연구, Vol. 11, No. 1, pp 5 -12, 한국통계학회, 1998.
- [5] 허문열, 정남철, 임경철, 김종수, “Statistics on the Internet”, 1996 Proceedings of the Autumn Conference Korean Statistical Society, pp 50 - 54, 한국통계학회, 1996.

- [6] 황진수, 엄태호, “PHP3를 이용한 웹상에서의 통계분석”, 한국 데이터 정보과학회지, Vol. 10, NO. 2, pp 501- 510, 한국 데이터 정보과학회, 1999.
- [7] Efron B., "Bootstrap Methods ; Another Look at the Jackknife", Ann. Statist., 7 pp 1 - 26, 1979.
- [8] Garcia J. L. "Submit Xlip-Stat",<http://www.stat.ucla.edu/cg-bin/Xlip-stat.cgi>, 1996
- [9] Leeuw J. D. "Regression", i, 1995
- [10] West, R.W., Ogden, R.T. and Rossini, A.J., "Statistical Tools on the World Wide Web", The American Statistician, Vol. 52, No. 3, pp 257-262. 1998.
- [11] <http://stat.sc.edu/rsrch/gasp/>
- [12] <http://compstat.chonbuk.ac.kr/cyberstat/>
- [13] <http://compstat.chonbuk.ac.kr/>

저 자 소 개

최 성 운 : 현 경원대학교 산업공학과 교수 재직 중. 한양 대학교 산업공학과에서 공학사, 공학석사, 공학박사 학위를 취득하고, 1994년 한국과학재단 지원으로 University of Minnesota에서 1년간 Post-Doc을 수행했으며, 2002년부터 1년 반동안 University of Washington에서 Visiting Professor를 역임하였음. 주요 관심분야는 자동화 생산 및 장치 산업에서의 품질관리이며, 컴퓨터 · 정보통신시스템의 신뢰성 설계 및 분석, RFID 시스템에도 관심을 가지고 있음

임 인 섭 : 경원대학교 산업공학과 졸업, 경원대학교 산업공학과 석사과정 재학 중. 관심분야는 컴퓨터 정보통신 프로그래밍, 자동화 생산 시스템, 데이터 베이스, 시뮬레이션에 관심을 가지고 있음.