

통계패키지와 Active Server Page를 이용한 통계 분석 웹 콘텐츠 개발

(Development of Web Contents for Statistical Analysis Using Statistical Package and Active Server Page)

강태구*, 이재관**, 김미아***, 박찬근****, 허태영*****

(Taegu Kang, Jae-Kwan Lee, MiAh Kim, Chankeun Park, Tae-Young Heo)

요약 본 논문에서는 통계패키지와 Active Server Page(ASP)를 이용하여 통계분석을 위한 웹 콘텐츠를 개발하였다. 통계패키지는 통계비전공자에게 사용하기 어렵고 배우기도 매우 어렵지만, 통계비전공자들은 SAS, S-plus, R 등과 같은 통계패키지에 대한 학습 없이 자료를 분석하기를 원하고 있다. 따라서 본 연구에서는 통계패키지로 많이 활용되고 있는 S-plus와 ASP를 이용하여 통계분석 웹 콘텐츠를 개발하였다. 실제 응용으로, 수질오염자료에 대하여 웹 상에서 탐색적 자료 분석, 분산분석, 시계열 분석 등과 같은 다양한 분석에 대한 웹 콘텐츠를 개발하였다. 개발된 웹 통계분석은 공무원, 연구원 등과 같은 통계 비전문가들에게 매우 유용한 도구이다. 결과적으로 웹 기반의 통계분석 콘텐츠를 통하여 인터넷으로 하여금 사용자들로 하여금 자료 분석을 쉽게 빠르게 할 수 있다.

핵심주제어 : Active Sever Page, 자바, 통계소프트웨어, 월드 와이드 웹

Abstract In this paper, we developed the web content of statistical analysis using statistical package and Active Server Page (ASP). A statistical package is very difficult to learn and use for non-statisticians, however, non-statisticians want to do analyze the data without learning statistical packages such as SAS, S-plus, and R. Therefore, we developed the web based statistical analysis contents using S-plus which is the popular statistical package and ASP. In real application, we developed the web content for various statistical analyses such as exploratory data analysis, analysis of variance, and time series on the web using water quality data. The developed statistical analysis web content is very useful for non-statisticians such as public service person and researcher. Consequently, combining a web based contents with a statistical package, the users can access the site quickly and analyze data easily.

Key Words : Active Server Page, JAVA, Statistical Software, World Wide Web

1. Introduction

Statistical software packages have been used for decades to perform statistical analyses. The emergence of the World Wide Web (WWW) has expanded the potential for these statistical packages and the recent rapid growth of information technology has changed

* 국립환경과학원 물환경연구부 물환경제어연구과 연구관
** 국립환경과학원 낙동강물환경연구소 연구소장
*** 국립환경과학원 물환경연구부 물환경제어연구과 전문위원
**** 한국해양대학교 데이터정보학과 부교수
***** 한국해양대학교 데이터정보학과 조교수, 교신저자

statistical software. Nakano (1997) developed the statistical software with a user interface with WWW [2, 7, 8].

The web-based statistical analysis system is useful because it provides a framework for the creation of collaborative web-based statistical applications. The system includes a statistical calculation and analysis engine, a graphical presentation engine, input and output components, a communications component, and a framework for storing data sets, information about the data sets, analyses, and information about these analyses [1, 3, 9].

We used statistical package S-plus to analyze the data and link with ASP on the web. There are several advantages using a web as an interface for statistical analysis system, first, the ability to add documents, graphical objects and easy link to other web sites on the web. We can also develop an interface based on user's needs and modify easily. Also, anyone can customize the interface for their mother language without special programming skills [5, 9].

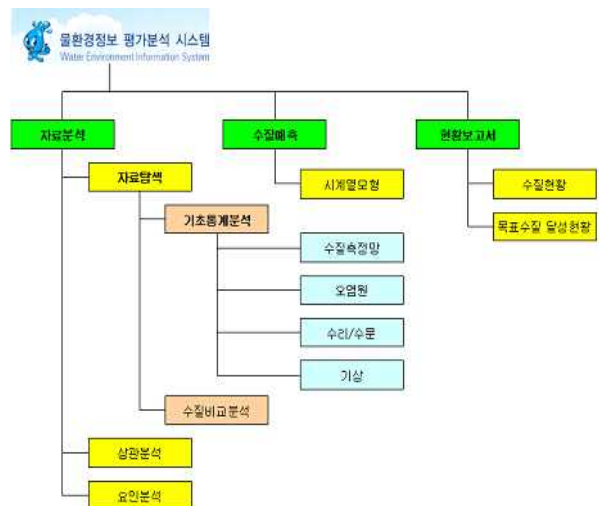
With these advantages, we developed the web content for various statistical analyses such as exploratory data analysis, analysis of variance, and time series on the web using water quality data. The developed statistical analysis web content is very useful for non-statisticians such as public service person and researcher. Consequently, combining a web based contents with a statistical package; the users can access the site quickly and analyze data easily.

2. Web-based Statistical Analysis System

Statistical analysis system with web interface is constructed by a web-based

statistical engine such as S-plus and R allows users to employ tools such as ASP, HTML and JAVA [2, 9, 10]. The statistical engine like S-plus is responsible for performing statistical calculations and creating graphs based on stored data, user parameters, and information about the stored data. The results of these calculations are stored in a separate set of tables, for use by other parts of the system. Finally, the import and export component allows the data and analyses used by the web-based system created by ASP to be augmented by resources that are available locally, as files stored on a user's hard drive [5, 9].

We developed the web-based statistical analysis system with S-plus and ASP using our mother language. Figure 1 shows that the layout of statistical analysis of water environmental information system to analyze the water quality data. In this system, using S-plus and link with ASP, we do advanced analysis such as factor analysis, cluster analysis and time series analysis to show on the web. The ASP can recall the S-plus function to analyze data and create the plot on the web.



<Figure 1> Statistical analysis of water environmental information system layout

The basic web link provides access to a web page in Figure 2 that is compatible with any browser that can handle forms and graphics. The page provides links to an online S-plus using ASP. The main features of the basic page in Figure 2. Once you have the login with ID and password, click on the submit button to execute it. You may enter into the main page to analyze data for user friendly system.



<Figure 2> Web content of water environmental information system

3. Data Import and Basic Analysis

The data import tool allows you to bring either numerical or coded data into the system from a plain text file or various Excel file format. You can either create a new data set, or add entries to an existing data set. There are two steps to importing data into the system.

First, select the time period for data. Second, select the variable name. You can do this by clicking on the "Browse" button, and navigating to the proper file. This can be a plain text file or Excel file with the data

separated by either commas or tab characters. Once we import data, we do basic analysis such as box-plot, density plot, histogram and normal Q-Q plot for any data set stored in system using S-plus. Figure 3 shows that web content of data import and basic analysis.

Once you select data, you can view data on the web. The data set viewer allows you to view the data contained in a data set which is stored by the system. First you may choose your interest variable and click on the line on the left bar. To view a data set, click on the link on the side bar, select the data set with period button. You can download with various format of excel. Figure 4 is the result of data import.



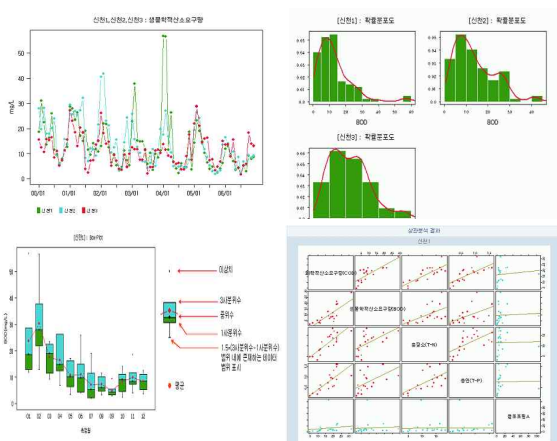
<Figure 3> Web contents of data import and basic analysis

측정소	측정연도	측정항목	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
신원1	2000	BOD	18.6	31.1	22.5	13.6	16.3	26	11.5	9.7	5.5	7.8	11.4	12.7
신원1	2001	BOD	28.7	28	23.5	26.4	17.1	15	19.1	11.6	9.5	14.1	11.8	10.9
신원1	2002	BOD	13	21.9	21.8	12.4	6.5	9.7	5.3	3.3	4.6	11.2	7.4	5.3
신원1	2003	BOD	22.9	37.9	15.2	14.8	14.9	11.7	5.9	10.2	3.5	9.9	8.4	11.1
신원1	2004	BOD	56.9	58.7	11.6	26.3	6.8	4	2.2	6	4.4	4.6	18.6	3.7
신원1	2005	BOD	14.2	23.3	19	14.8	10.2	6.5	3.9	4.8	3.4	2.4	4.1	6.2
신원1	2006	BOD	11.9	13	9.4	7	3.6	4.5	2.1	5.4	3	9.1	7.8	8.6
신원2	2000	BOD	28	19.7	28.1	17.9	18	20.2	23.9	9.1	6.1	7.2	15.7	13.7
서해2	2001	BOD	29.5	25.6	28.9	25.5	27.2	32.1	8.2	6.9	12.1	11.8	9.1	21.5

<Figure 4> Result of data import

4. Visualization

We will use visualizations as responses to queries, to reduce risk by conveying qualitative information without disclosing numerical details. This web-based statistical analysis system produce statistics plot allows you to determine which statistical procedure is most appropriate for your research, based on the research goal and type of used for the study. The data set that the data for the analysis will be drawn from. The variable that you are interested in analyzing; once you select a data set, the variables listed in this menu will correspond to the columns in the selected data set. he stats operation that you would like to perform on that variable; this may be either an operation which creates a table, or a descriptive operation. When you are viewing an analysis, the results may be shown in one of two ways. If the analysis includes one or more graphs, then the results will be shown on a graph, as shown in the Figure 5. If the analysis consists entirely of calculations, then the results will be shown in text format, as shown in the Figure 6.



<Figure 5> Example of visualization and exploratory

Figure 5 shows that time series plot (Top left), histogram with density plot (Top right), box plot (Bottom left) and scatter plot (Bottom right) using S-plus on the web.

모의통계						
측정소 / 계수	최소값	1사분위수	중위수	평균	3사분위수	최대값
관인	26	9.11	9.8575	10.62	11.7338	11.725

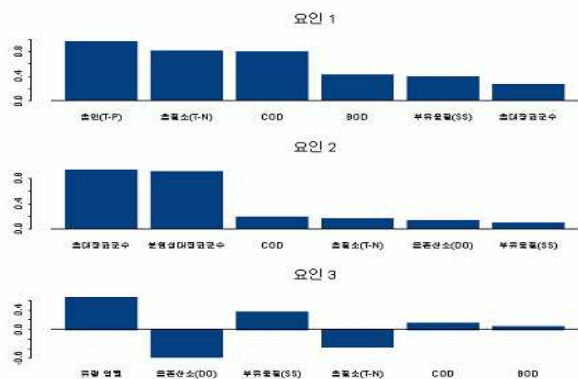
<Figure 6> Example of the text format results on the web

5. Advanced Analysis

The show advanced analyses tool such as factor analysis, cluster analysis and time series analysis allows you to analyze sophisticated method in S-plus stored within the system. In this paper, we only showed the interface and results of factor analysis in Figure 7 and 8. Web based statistical analysis system can be analyzed the sophisticated procedures whatever S-plus can do. It will be very useful to analyze for non-statistician.



<Figure 7> Web content of factor analysis



<Figure 8> Result of factor analysis

6. Results

The web-based statistical analysis system was described in this paper. It includes a statistical analysis engine, data set management tools, an analysis storage framework and a communication component to facilitate information dissemination. Particularly, we describe an outline of the web based statistical analysis contents using S-plus which is the popular statistical package and ASP. We illustrate the power of web-based software combining S-plus and ASP as an approach to statistical computing. This system include more sophisticated procedures such as multiple regression, analysis of variance, factor analysis, cluster analysis, time series analysis, and graphical visualization. With this set of components, it will be possible to construct customized statistical applications which can be tailored according to individual preferences. Therefore, the user can get access to statistical results and easily understand the results.

References

- [1] C. Asano and H. Kimura, "A study on system construction of general purpose statistical package for microcomputers Micro NISAN system", Computational Statistics Quarterly, 6, 1990, pp. 65-77.
- [2] C. Asano and A. Takeuchi, "Web-based statistical system by using the DLL", Computational Statistics and Data Analysis, 4, 2003, pp. 409-418.
- [3] E. Cramer, K. Cramer, U. Kamps "e-stat: A web-based learning environment in applied statistics", Compstat 2002.
- [4] N. Holmberg, B. Wünsche, E. Tempero, "A Framework for Interactive Web based Visualization", Seventh Australasian Interface Conference, 2006.
- [5] M. A. Mullee, "Web-based resources to assist the statistical analysis and preparation of data", Pharmaceutical Statistics, 4, 2005, pp. 129-139.
- [6] P. Myllymäki, T. Silander, H. Tirri, P. Uronen, "A Web based tool for bayesian and causal data analysis." International Journal on Artif. Intel. Tools, 11, 2002, pp. 369-387.
- [7] J. Nakano, "An internet user interface for statistical software", Bulletin of the International Stat. Inst., 51st Session Contributed Paper Book 2, 1997, pp. 35-36.
- [8] J. Nakano, "Graphical user interface for statistical software using internet", Proceedings of Computational Statistics, 98, 1998, pp. 407-412.
- [9] R. W. West and R. T. Ogden, "Statistical Analysis with Webstat, a Java applet for the World Wide Web", Journal of Statistical Software, 1997, pp. 2-3.
- [10] H-K. You, T-Y, Heo, W-S, Cho, J-H, Na, "Web-based Water Quality Data Monitoring and Analysis", Proceedings of open learning society.



강 태 구 (Taegu Kang)

- 비회원
- 1988년 2월 : 고려대학교 농화학 (학사)
- 1990년 2월 : 고려대학교 농화학과(석사)
- 2005년 2월 : 고려대학교 농화학과(박사)
- 1994년 9월 ~ 현재 : 국립환경과학원 물환경 제어연구과 환경연구관
- 관심분야 : 물환경정보시스템, 수질평가기법



이 재 관 (Jae-Kwan Lee)

- 비회원
- 1986년 2월 : 부경대학교 환경공학과 (학사)
- 1989년 2월 : 부경대학교 환경공학과 (석사)
- 1994년 2월 : 부경대학교 환경공학과 (박사)
- 1994년 9월 ~ 현재 : 국립환경과학원 낙동강물 환경연구소장
- 관심분야 : 물환경종합평가, 유역관리

김 미 아 (MiAh Kim)

- 비회원
- 2005년 1월 ~ 현재 : 국립환경과학원 물환경제어연구과 전문위원



박 찬 근 (Chankeun Park)

- 비회원
- 1991년 2월 : 영남대학교 통계학과 (학사)
- 1996년 8월 : 미국 노스다코다주립대학교 통계학과 (석사)
- 1999년 10월 : 미국 노스다코다주립대학교 통계학과 (박사)
- 2000년 3월 ~ 현재 : 한국해양대학교 데이터정보학과 부교수
- 관심분야 : 응용통계, 회귀분석, 비모수통계



허 태 영 (Tae-Young Heo)

- 종신회원
- 1999년 8월 : 충북대학교 통계학과 (학사)
- 2001년 5월 : 미국 노스캐롤라이나주립대학교 통계학과 (석사)
- 2005년 8월 : 미국 노스캐롤라이나주립대학교 통계학과 (박사)
- 2005년 10월 ~ 2007년 8월 : 한국전자통신연구원 선임연구원
- 2007년 9월 ~ 현재 : 한국해양대학교 데이터정보학과 조교수
- 관심분야 : 응용통계, 공간통계, 교통통계, 환경통계, IT통계

논문접수일 : 2009년 10월 17일
 1차수정완료일 : 2009년 11월 18일
 게재확정일 : 2009년 11월 20일