

## 웹기반 임상자료의 동적 통계분석 시스템 개발<sup>†</sup>

신임희<sup>1</sup> · 곽상규<sup>2</sup> · 박전우<sup>3</sup>

<sup>123</sup>대구가톨릭대학교 의과대학 의학통계학교실

접수 2013년 10월 14일, 수정 2013년 10월 25일, 게재확정 2013년 12월 5일

### 요약

많은 응용분야에서 통계분석을 이용하여 의사결정을 뒷받침하는 정보를 얻는다. 그러나 PC용 통계분석 프로그램은 경제적인 부담과 시간 및 위치의 제한을 받는다. 이를 최소화하기 위하여 인터넷을 이용하여 서버 PC의 통계분석 프로그램을 사용하거나, 웹 브라우저를 이용하여 통계분석 프로그램을 사용할 수 있는 웹기반 시스템이 개발되어 왔다. 그러나 기존 웹기반 시스템의 연구는 특정 통계분석 프로그램을 사용하여야 하거나 서버에 저장된 자료에 대해서만 이루어 졌다. 자료가 수정되거나 새로 생성되면 서버관리자가 다시 자료를 탑재하여야만 통계분석이 가능하였다. 이를 개선하기 위하여 웹에서 사용되어 지는 HTML, java, JSP 등의 언어를 사용하여 동적(動的) 자료에 대해서도 통계분석이 가능한 웹기반 시스템을 개발하였다.

주요용어: 동적 자료, 웹, 웹프로그래밍 언어, 통계분석.

### 1. 서론

자료를 수집하고 요약 및 분석하여 의사결정에 필요한 정보를 산출해 내는 일련의 과정을 통계분석이라 한다. 여러 많은 응용분야에서 통계분석을 이용하여 의사결정을 뒷받침하는 정보를 얻는다. 특히 치료법이나 의약품 등 환자를 대상으로 하는 임상연구에서는 통계분석으로 산출한 정보가 환자를 치료하는데 있어서 의사결정에 결정적인 역할을 한다. 그러나 통계분석을 위해서는 복잡한 수식에 대한 계산이 필요하다. 이를 해소화하기 위하여 컴퓨터를 이용한 통계분석 프로그램 (통계패키지)이 여러 가지 개발되었다. 예를 들어 SAS, SPSS, S-LINK, MINITAB, Stata, MedCalc, R, Excel 등의 개인 컴퓨터용 통계분석 프로그램이 있다. 이런 개인 컴퓨터용 통계분석 프로그램을 사용하기 위해서는 필요한 만큼 개인용 PC에 설치되어야 하고 각각 고가의 사용권을 구입하여 사용하여야 하므로 이로 인해 사용자가 경제적인 부담을 느끼게 된다. 또한 유지보수에도 많은 시간과 인력을 투자하여야 한다. 이는 통계분석이 물리적·공간적으로 시간과 위치의 제한을 받는 것을 의미한다.

공간과 시간의 제약을 최소화하기 위하여 인터넷을 이용하여 통계분석 프로그램을 사용하거나, 웹 브라우저를 이용하여 통계분석 프로그램을 사용할 수 있는 통계분석 모듈의 개발을 위하여 다양한 노력과 연구가 수행되어져 왔다. Lee 등 (2004)는 S-LINK를 웹에서 사용할 수 있는 시스템을 작성하여 이에 관한 실용화 계획을 발표하였다. 기존의 웹기반 분석도구에 관한 연구로 Hwang과 Uhm (1999), Oh 등

<sup>†</sup> 본 연구는 보건복지부 통합의료 진료지침 개발과 연구사업의 지원 (과제고유번호 : 3033-302)에 의하여 이루어진 것임.

<sup>1</sup> (705-718) 대구광역시 남구 대명4동 3056-6, 대구가톨릭대학교 의과대학 의학통계학교실, 교수.

<sup>2</sup> 교신저자: (705-718) 대구광역시 남구 대명4동, 대구가톨릭대학교 의과대학 의학통계학교실, 연구강사.  
E-mail: sanggyu39@knu.ac.kr

<sup>3</sup> (705-718) 대구광역시 남구 대명4동, 대구가톨릭대학교 의과대학 의학통계학교실, 연구원.

(2006), Ryu와 Park (2006) 등을 들 수 있다. 이는 대부분 특정 통계분석 프로그램을 이용하여야 한다. 즉 서버 PC에는 특정 사용권이 구입된 통계분석 프로그램이 설치가 되어 있어야 한다는 한계점이 있다. 이를 개선하기 위하여, Lee 등 (1999)은 해마다 발간되는 창원시의 통계정보를 웹 상에서 java와 오라클 데이터베이스를 이용한 통계정보시스템을 구현하였다. 그리고 Kim 등 (2011)는 웹에서 사용되어 지는 HTML, java, JSP 등의 언어만 사용하여 특정 통계분석 프로그램의 설치 없이 통계분석이 가능한 시스템을 개발하였다. 그러나 자료가 수정되거나 새롭게 생성되면 서버관리자가 다시 자료를 탑재하여야만 통계분석이 가능하였다. 즉, 자료에 대하여 동적이 아닌 정적 (靜的)인 시스템이었다. 본 연구는 이를 개선하기 위하여 웹프로그램을 사용하여 동적 (動的) 자료에 대해서도 통계분석이 가능한 웹기반 시스템을 개발하였다. 웹에서 원하는 형태의 자료를 사용자가 직접 서버에 저장할 수 있도록 하였고 이를 통하여 통계분석기법에 따라 버튼을 클릭하면 서버에 저장한 자료의 변수명 및 입력값으로 통계분석이 실행되도록 하였다. 이 논문은 다음과 같이 구성하였다. 2절에서는 웹기반 시스템의 구성을 소개한다. 그리고 3절에서는 웹기반 동적 자료 분석 시스템을 구현한 방법을 소개하고, 4절에서 결론을 기술한다.

## 2. 웹기반 시스템 구성

### 2.1. java 및 JSP 언어

본 연구에서는 java와 JSP 언어를 기초로 시스템을 개발하였다. java언어는 객체지향적이고 독립적 플랫폼을 가지며 쉽고 간단하게 멀티스레드를 적용 가능하다는 장점을 가지고 있다. 독립적 플랫폼은 java API, java 가상머신 (JVM)으로 구성된다. Java 가상머신은 자바 바이트 코드를 수행할 수 있는 환경을 의미하며 대부분의 운영체제나 웹브라우저 등 여러 가지 플랫폼에 설치되어 사용될 수 있고 휴대 전화나 가전기기에도 설치가 가능하다.

JSP는 서블릿 (servlet; java기반의 웹프로그래밍 기술로 스레드 기반의 빠른 처리속도를 자랑하나 프로그램 내에서 화면코딩을 제어해야 하는 문제점을 가지고 있음) 기반의 서버 스크립트 언어로 오늘날 가장 대표적인 웹프로그래밍 언어이다. 또한 JSP는 서블릿의 장점을 모두 수용하면서 단점을 보완하였다.

### 2.2. 웹기반 시스템 설치

웹기반 시스템 개발을 위하여 서버 PC에 JDK (java development kit; java 프로그램을 개발할 수 있게 해주는 개발도구) JRE (java runtime environment; java 언어로 개발된 소프트웨어를 실행하기 위해 필요한 플랫폼), Tomcat 6.X (해당 PC를 서버 PC로 사용하게 해주는 도구), Oracle 11g Express Edition (데이터베이스)을 설치하였다.

또한 서버 PC에 자료를 저장하고 분석하기 위해서 본 연구에서는 Oracle 데이터베이스 관리시스템을 사용하였다 (Diana, 2005).

Oracle은 미국 Oracle사의 관계형 데이터베이스 관리 시스템의 이름이다. Oracle 데이터베이스 시스템은 다양한 운영시스템 (windows, unix, linux)에 호환성이 우수하며, 탁월한 RAC (real application cluster)기능을 가지고 있다. 그리고 안정성 및 보안 자료 암호화 기능이 탁월하다는 장점을 가지고 있다.

### 3. 웹기반 동적 자료 분석 시스템 구현

#### 3.1. 자료

예제로 사용된 자료는 Table 3.1에 주어져 있으며 모 대학병원의 소아과 환자 데이터베이스에서 15명을 단순임의 추출하여 구성하였다 (Shin, 2008). 자료는 성별, 나이, 몸무게, 수축기 혈압, 심박동수, 호흡수, 혈당 변수로 이루어져 있다. 이 자료를 환자 자료 (성별, 나이, 몸무게)와 혈액학 자료 (수축기 혈압, 심박동수, 호흡수, 혈당)로 구분하여 2가지의 자료로 사용하고자 한다.

**Table 3.1** Data of pediatrics patients

No.	Sex	Age	Weight	SBP	HR	RR	Glucose
1	1	3	15.3	95	120	23	75
2	2	7	23.2	105	93	17	66
3	1	2	10.4	100	132	19	70
4	2	5	20.5	95	100	18	72
5	1	10	45.8	115	74	14	77
6	1	1	9.3	90	122	30	84
7	2	4	18.0	100	115	17	81
8	2	7	20.4	113	78	16	80
9	1	7	28.3	107	92	19	93
10	2	9	25.9	97	84	14	80
11	1	3	16.2	88	127	23	93
12	2	11	37.2	106	78	18	82
13	1	8	22.2	101	80	16	71
14	1	17	54.8	122	72	13	66
15	2	6	24.3	100	90	20	91

#### 3.2. 동적 자료 탑재 시스템

동적 자료로 통계분석을 하기 위해서는 우선 동적 자료를 서버에 저장하여야 한다. 이를 위해서 개인 컴퓨터에서 새롭게 작성한 자료나 수정한 자료를 서버의 데이터베이스에 탑재 하여야 한다. 임상연구에서 연구자들은 다양한 프로그램을 사용하여 임상자료를 수집한다. 예를 들어, 메모장, 한글, 엑셀 등 자료를 수집하는데 특별히 정해진 기준은 없다. 본 연구에서는 엑셀 프로그램을 이용하여 수집된 엑셀파일 형태의 자료를 서버에 탑재하여 저장하는 것으로 시스템을 개발하였다. 엑셀 파일의 확장자명 '.xls'와 '.xlsx'에 관계없이 즉 엑셀 프로그램의 버전에 관계없이 모두 탑재가 가능하도록 하였고 각 자료에 따른 변수명과 값이 다르더라도 각각의 자료에 형태와 특성에 따라 서버에 저장되도록 하였다. 또한 숫자형 자료, 문자형 자료, 날짜형 자료 그리고 수식에 의해 생성된 자료를 자동적으로 구분하여 서버에 탑재가 가능하도록 하였다. 탑재된 자료는 사용자가 바로 접근하여 사용할 수 있으며, Oracle 데이터베이스를 사용하여 시스템을 구축하였으므로 탑재 할 수 있는 모든 자료의 합한 용량크기는 최대 4G byte이며 자료의 개수는 제한이 없다. 이에 대한 시스템 소스는 jsp, HTML, java 웹프로그래밍언어를 사용하여 다음과 같이 작성하였다. 이때 사용한 라이브러리는 'servlet-api'를 사용하였고 화면출력을 위한 소스코드, 자료 조회를 위한 소스 코드 그리고 자료 탑재 및 저장을 위한 소스코드로 구분하여 제시하였다.

---

**Interface output page code**


---

```

<html><head> <title> Apache ServletFileUpload
example </title></head><body><center>
<% boolean isMultipart =
ServletFileUpload.isMultipartContent(request)
String fileName=null;
if (!isMultipart){}else {
    FileItemFactory factory = new DiskFileItemFactory();
    ServletFileUpload upload = new ServletFileUpload(factory);
    List items = null;
    try { items = upload.parseRequest(request);
    } catch (FileUploadException e) {
    out.println("에러 1: "+e); }
    Iterator itr = items.iterator();
    while (itr.hasNext()) {
    FileItem item = (FileItem) itr.next();
    if (item.isFormField()) {
        if(item!=null && item.getFieldName().equals("name")) {
            }else if(item!=null && item.getFieldName().equals("desc"))
            {} }else {
            try {
                String itemName = item.getName();
                if(itemName==null || itemName.equals("")) continue;
                fileName = FilenameUtils.getName(itemName);
                File savedFile = new File("C:/upload/"+fileName);
                item.write(savedFile);
            } catch (Exception e) {
                out.println("서버에 파일 저장중 에러: "+e); } } }
    ExcelUpload EU = new ExcelUpload();
    String[] var = EU.getVar(fileName);
    String[] data = EU.getData(fileName); %>
    <table border="1" cellspacing="0" cellpadding="7" >
    <tr align="center"><% for(int i=0;i<data.length;i++)
    {for(int j=0;j<var.length;j++) { %> <td><%=data[i][j]%></td><%
    %></tr><% %></table></center></body></html>

```

---

**Class resource code for table inquiry using java**


---

```

public String[] TableName(){
    String sql = "select table_name from user_tables
    order by TABLE_NAME";
    Connection conn = getConn();
    PreparedStatement pstmt = null;
    ResultSet rs = null; try{ pstmt =
    conn.prepareStatement(sql,ResultSet.TYPE_SCROLL_
    INSENSITIVE, ResultSet.CONCUR_READ_ONLY);
    rs = pstmt.executeQuery();
    String[] result = new String[500];
    String tn="TABLE_NAME";
    result[0] = "select";
    rs.first(); result[1]= rs.getString(tn); int i=2;
    while(rs.next()){result[i]= rs.getString(tn); i++;
    }return result;
}

```

---

**Class resource code for data upload using java**


---

```

for (int sheetNumber = 0; sheetNumber <
excelWB.getNumberOfSheets(); sheetNumber++){
    Sheet oneSheet = excelWB.getSheetAt(sheetNumber);
    int rows = 1;
    for (int rowNumber=0; rowNumber < rows; rowNumber++){
        iRow oneRow = oneSheet.getRow(rowNumber);
        iif (oneRow == null){continue;}
        iint cells = oneRow.getPhysicalNumberOfCells();
    }or (int cellNumber = 0; cellNumber < cells; cellNumber++){
        iCell oneCell = oneRow.getCell(cellNumber);
        iif (oneCell == null){continue;}
        iswitch (oneCell.getCellType()) {
        icase Cell.CELL_TYPE_STRING:
            ivar[cellNumber]=oneCell.getStringCellValue(); break;
        icase Cell.CELL_TYPE_ERROR:break;
        ...
    }for (int sheetNumber = 0; sheetNumber <
excelWB.getNumberOfSheets(); sheetNumber++){
    Sheet oneSheet = excelWB.getSheetAt(sheetNumber);
    int rows = oneSheet.getPhysicalNumberOfRows();
    for (int rowNumber=0; rowNumber < rows; rowNumber++){
        Row oneRow = oneSheet.getRow(rowNumber);
        if (oneRow == null){continue;}
        int cells = oneRow.getPhysicalNumberOfCells();
        for (int cellNumber=0; cellNumber <= cells; cellNumber++){
            Cell oneCell = oneRow.getCell(cellNumber);
            if (oneCell == null) {continue;}
            switch (oneCell.getCellType()) {
            case Cell.CELL_TYPE_NUMERIC:
                if(HSSFDateUtil.isCellDateFormatted(oneCell)) {
                    SimpleDateFormat f=
                    new SimpleDateFormat("yyyy-MM-dd");
                    data[rowNumber][cellNumber]
                    =f.format(oneCell.getDateCellValue());}
                else {data[rowNumber][cellNumber]=
                    Double.toString(oneCell.getNumericCellValue());}
                break;
            case Cell.CELL_TYPE_STRING:
                data[rowNumber][cellNumber]=
                oneCell.getStringCellValue();break;
            case Cell.CELL_TYPE_FORMULA:
                int type=oneCell.getCachedFormulaResultType();
                if(type==0){
                    data[rowNumber][cellNumber]=
                    Double.toString(oneCell.getNumericCellValue());}
                else if(type==1){
                    data[rowNumber][cellNumber]=
                    oneCell.getStringCellValue();} break;
            case Cell.CELL_TYPE_ERROR:
                data[rowNumber][cellNumber]="Error";break;
            case Cell.CELL_TYPE_BLANK:
                data[rowNumber][cellNumber]="."; break;
            }
        }
    }
}

```

---

### 3.3. 동적 자료 통계분석 시스템

자료가 서버에 탑재 되고 난 후에는 통계분석 시 사용자가 원하는 자료를 선택할 수 있어야 비로소 자료를 동적으로 통계분석을 할 수 있게 된다. 이를 위해서 분석에 원하는 자료를 동적으로 선택하면 통계분석 메뉴에서, 선택된 자료에 따른 변수가 화면에 출력되고 출력된 변수들을 사용하여 통계분석이 가능하게 하였다. 본 연구에서 사용한 자료로 살펴보면 환자 자료를 선택하면 성별, 나이, 몸무게의 변수를 사용하고 혈액학 자료를 선택하면 수축기 혈압, 심박동수, 호흡수, 혈당 변수를 사용하여 화면에 출력하고 각 자료에 대한 변수로 통계분석이 가능하게 하였다. 이런 일련의 과정은 자료 선택, 통계분석 메뉴, 변수선택, 통계분석 소스 프로그램, 통계분석 결과 그리고 결과 출력으로 진행되며 Figure 3.1와 같이 진행된다. 또한 자료를 선택하면 서버에 탑재하였거나 탑재되어 있는 자료를 조회할 수 있는 데이터조회 시스템도 구축하였다. 이 과정 또한 자료선택, 자료조회 메뉴, 자료조회 소스 프로그램 그리고 결과 출력으로 진행되면 Figure 3.1에 제시하였다.

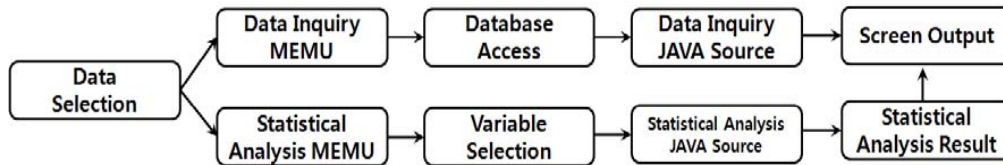


Figure 3.1 Processing from data selection to screen output for data inquiry and statistical analysis

동적 자료에 대한 분석 시스템을 구축하기 위하여 <select>과 <option> 구문을 이용하였고 서버에 탑재된 자료의 리스트 중 통계분석을 원하는 자료를 선택할 수 있도록 하였다. 또한 자료의 수가 항상 변할 수 있기 때문에 반복문을 사용하여 자료명이 나타날 수 있도록 시스템을 구축하였다. 개발된 동적 자료 분석 시스템은 웹브라우저에서 실행이 되며 Figure 3.2과 같이 하나의 화면에서 자료 선택 메뉴, 통계분석 메뉴, 변수선택 메뉴, 결과화면으로 구성된다. 본 연구에서 사용한 자료는 2가지로 환자 자료와 혈액학 자료이며 자료선택 콤보 박스에서 원하는 자료를 선택할 수 있으며 이에 대한 자료를 살펴볼 수 있는 자료 조회 기능도 구축하였다.

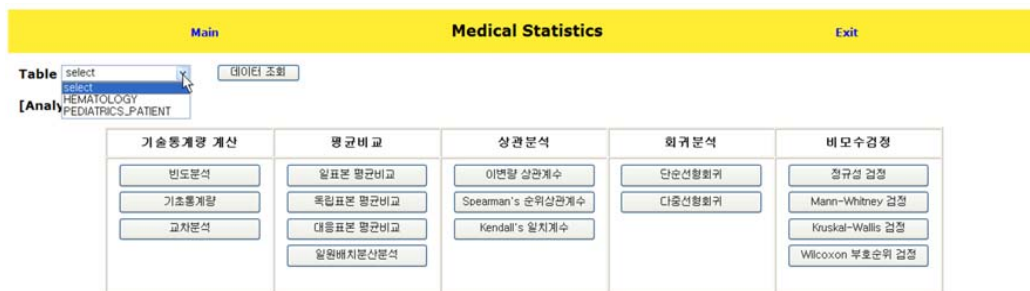


Figure 3.2 System interface for web-based statistical analysis

원하는 자료를 선택하고 각각의 통계분석 메뉴를 선택하면 그에 따른 변수선택 메뉴에서 선택된 자료에 대한 변수 리스트를 시스템 화면에 나타내게 하였다. 즉, Figure 3.3과 같이 환자 자료를 선택하고 빈도분석 메뉴를 선택하면 환자자료에 있는 변수인 성별, 나이, 몸무게가 변수 선택 메뉴로 나타나고 혈액학 자료를 선택하면 수축기 혈압, 심박동수, 호흡수, 혈당이 변수 선택 메뉴로 나타난다. 즉, 자료에 따라 변수명이 나타나도록 하여 데이터에 대한 동적 통계분석이 가능하도록 하였다.

Table <span>PEDIATRICS_PATIENT</span> <span>데이터 조회</span>	Table <span>HEMATOLOGY</span> <span>데이터 조회</span>
<b>[Variable Selection]</b>	<b>[Variable Selection]</b>
<input type="checkbox"/> NO <input type="checkbox"/> SEX <input type="checkbox"/> AGE <input type="checkbox"/> WEIGHT	<input type="checkbox"/> NO <input type="checkbox"/> SBP <input type="checkbox"/> HR <input type="checkbox"/> RR <input type="checkbox"/> GLC
<input type="button" value="빈도분석 실행"/>	<input type="button" value="빈도분석 실행"/>

Figure 3.3 Variable selection interface of frequency analysis according to data set

자료가 변함에 따라 동적으로 통계분석이 가능하려면 자료명과 변수명이 모두 선택되는 자료에 따라 변하여야 한다. 많은 종류의 동적 자료 통계분석 프로그램을 개발하였지만 본 논문에서는 범주형 변수에 대한 빈도표를 작성하는 빈도분석과 독립적으로 존재하는 두 집단 간의 평균 차이를 검증하는 통계분석 기법인 독립표본 t-test에 대해 제시하였다. 우선 빈도분석을 위한 변수선택 화면, 결과출력 화면, JAVA소스를 살펴보면 다음과 같다.

Variable selection interface for frequency analysis	
<pre>&lt;form name="form2" method="post"&gt; &lt;h4 align="left"&gt;[Variable Selection] &lt;br&gt; &lt;/br&gt; &lt;% String TB= request.getParameter("tableName"); ShowData SD = new ShowData(); String[ ] var = SD.AllVar(TB);%&gt; &lt;input type="hidden" name="tname" value="&lt;%=TB%&gt; size=30"&gt; &lt;/input&gt; &lt;% int leng=0; for(int i=0;i&lt;var.length;i++){</pre>	<pre>if(var[i]!=null) leng+=1; }%&gt; &lt;table border="0" cellspacing="0" cellpadding="7"&gt; &lt;tr&gt; &lt;% for(int i=0;i&lt;leng;i++){%&gt; &lt;td&gt; &lt;input type="checkbox" name="colname" value="&lt;%=var[i]%&gt;" /&gt;&lt;%=var[i]%&gt; &lt;/td&gt; &lt;% } %&gt;&lt;/tr&gt;&lt;/table&gt; &lt;input type="button" value="빈도분석 실행" onClick="javascript:test(1);" /&gt;</pre>
Interface output page code for frequency analysis	
<pre>&lt;%String tname = request.getParameter("tname"); String[ ] colname = request.getParameterValues("colname"); BasicFreqEmp BF= new BasicFreqEmp();%&gt;&lt;html&gt; &lt;head&gt;&lt;title&gt; Statistical Analysis in Web&lt;/title&gt;&lt;/head&gt; &lt;body&gt;&lt;center&gt; 기초통계학 분석 결과 &lt;br&gt;&lt;/br&gt; &lt;% for(int i=0;i&lt;colname.length;i++){%&gt; &lt;table border="1" cellspacing="0" cellpadding="7"&gt; &lt;tr align="center"&gt; &lt;td&gt;변수&lt;/td&gt;&lt;td&gt;값&lt;/td&gt;&lt;td&gt;케이스수(N)&lt;/td&gt; &lt;td&gt;퍼센트(%)&lt;/td&gt;&lt;/tr&gt;&lt;%&gt;</pre>	<pre>String[ ][ ] result = BF.getFreqNum(tname,colname[i]); int row=Integer.parseInt(result[0][3]);%&gt; &lt;tr&gt;&lt;td align="center" rowspan="&lt;%=row%&gt;&gt; &lt;%=colname[i]%&gt;&lt;/td&gt;&lt;% for(int j=0;j&lt;row;j++){%&gt; &lt;td align="center"&gt;&lt;%=result[j][0]%&gt;&lt;/td&gt; &lt;td align="center"&gt;&lt;%=result[j][1]%&gt;&lt;/td&gt; &lt;td align="center"&gt;&lt;%=result[j][2]%&gt;&lt;/td&gt;&lt;/tr&gt;&lt;% } %&gt; &lt;tr&gt;&lt;td align="center" colspan=2&gt;합계&lt;/td&gt; &lt;td align="center"&gt;&lt;%=result[row][2]%&gt;&lt;/td&gt; &lt;td align="center"&gt;&lt;%=result[row][2]%&gt;&lt;/td&gt;&lt;/tr&gt;&lt;% } %&gt; &lt;/center&gt;&lt;/body&gt;&lt;/html&gt;</pre>
Class resource code for frequency analysis using java	
<pre>public String[ ][ ] getFreqNum(String tname, String colName){ String sql = "select "+colName+" from "+tname+" where "+colName+" is not null order by "+colName; conn = getConn(); PreparedStatement pstmt = null;ResultSet rs = null; NumberFormat nfa = NumberFormat.getInstance(); nfa.setMaximumFractionDigits(0); NumberFormat nfb = NumberFormat.getInstance(); nfb.setMaximumFractionDigits(3); try{ pstmt = conn.prepareStatement(sql,ResultSet.TYPE_SCROLL_INSENSITIVE, ResultSet.CONCUR_READ_ONLY); rs = pstmt.executeQuery(); int cnt=0;while(rs.next()){cnt++; rs.first();</pre>	<pre>String[ ][ ] result = new String[cnt+5][4]; double subcnt=1.0;int t=0;String a="",b=""; for(int i=0;i&lt;cnt;i++){ a=rs.getString(colName); if(i&lt;cnt-1){rs.next();b=rs.getString(colName);} boolean eq=a.equals(b); double percent=0.0; if(i==cnt-1    eq==false){ result[t][0]=a; String subcntS = nfb.format(subcnt); result[t][1]=subcntS; percent=subcnt/cnt*100; String percentS = nfb.format(percent); result[t][2]=percentS; subcnt=1; t++; }else if(eq==true){subcnt++;} result[0][3]=Integer.toString(t); result[t+1][2]=Integer.toString(cnt);}return result;</pre>

위의 소스코드와 화면출력 코드를 사용하여 본 연구에서 사용하는 두 가지의 자료에 대해 개발된 동적 시스템에서 빈도분석을 실행한 결과는 Figure 3.4와 같다. Figure 3.4의 왼쪽 그림은 환자 자료를 선택하고 빈도분석 변수선택 메뉴에서 성별을 선택하여 빈도분석을 실행한 화면을 캡처한 그림이다. 또한 Figure 3.4의 오른쪽 그림은 혈액학 자료를 선택하고 빈도분석 변수선택 메뉴에서 호흡수를 선택하여 빈도분석을 실행한 화면을 캡처한 그림이다. 따라서, 각 자료에 따른 변수명이 변수선택 메뉴에 나타나게 되고 사용자가 선택하는 변수에 따라 분석결과가 나타나게 된다. 빈도분석에 따른 통계분석 결과는 각 변수의 값에 따른 케이스 수와 퍼센트로 구성되며 또한 합계도 화면에 나타나도록 하였다.

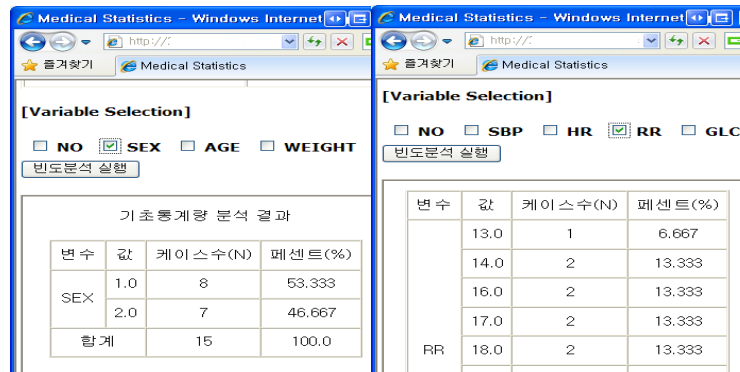


Figure 3.4 Statistical analysis result interface of frequency analysis according to data set

다음으로 독립표본 t-test를 위한 변수선택 화면, 결과출력 화면, JAVA소스를 살펴보면 다음과 같다. 독립표본 t-test는 표준편차가 같은 경우와 다른 경우를 모두 동적 시스템으로 분석하여 한 화면에 나타내도록 하였다.

```

Variable selection interface for two sample t-test
<form name="form2" method="post">
  <h4 align="left">[Variable Selection] <br></h4><br><%
String TB= request.getParameter("tableName");
ShowData SD = new ShowData(); String[] var = SD.AllVar(TB);
%><input type="hidden" name="tbname" value="<%=TB%>
size=30"></input><input type="text" name="colname"
for(int i=0;i<var.length;i++){if(var[i]==null) leng+=1;}%>
  Var : < select id="sel1" name="colname">
  <option value="select"> 변수선택 </option> <%
  for(int i=0;i<leng;i++){
  <option value="<%=var[i]%"><%=var[i]%"> </option> <% } %>
  </select> Test Value : <input type="text" name="colname"
  size="10"> <br> </br> <input type="button" value="일표본 평균비교
  실행" onClick="javascript:test(1);"/>

Interface output page code
<%@ page contentType="text/html; charset=euc-kr" import="java.sql.*"
%><%@ page import="analysis.*" %>
<%@ page import="java.io.*" %>
<%@ page import="java.text.*" %>
<% request.setCharacterEncoding("euc-kr");%>
<%String tbname = request.getParameter("tbname");
String[] colname = request.getParameterValues("colname");
BasicTestEmp BT= new BasicTestEmp();
BasicIndepTtestEmp BIT= new BasicIndepTtestEmp();
BasicIndepUTtestEmp BIUT= new BasicIndepUTtestEmp();
BasicFTtestEmp BFT= new BasicFTtestEmp();
BasicStatEmp BS= new BasicStatEmp();%><html>
<head><title>Statistical Analysis in Web</title></head>
<body><center>독립표본 평균비교 결과<br></br>
<%double[] result = new double[8];
result = BT.getTtest(tbname, colname);
double[] result2 = new double[8];
result2 = BIT.getIndepTtest(tbname, colname);
double[] result3 = new double[8];
result3 = BIUT.getIndepUTtest(tbname, colname);
double[] result4 = new double[8];
result4 = BFT.getFtest(tbname, colname);
double mean1=result[0];double std1=result[1];
double var1=std1*std1; double cnt1=result[2];
double mean2=result[3]; double std2=result[4];
double var2=std2*std2; double cnt2=result[5];
double meand=mean1-mean2; double pF=result4[0];
double df1=cnt1+cnt2-2;
double sp=(var1*(cnt1-1)+var2*(cnt2-1))/df1;
double sd1=sp*Math.sqrt(1/cnt1+1/cnt2);
double tvalue1=meand/sd1;double p1=result2[0];
double temp1=(var1*var1/cnt1+var2*var2/cnt2)*(var1*var1/cnt1+
var2*var2/cnt2);
double temp2=(var1*var1/cnt1)*(var1*var1/cnt1)/(cnt1-1);
double temp3=(var2*var2/cnt2)*(var2*var2/cnt2)/(cnt2-1);
double df2=temp1/(temp2+temp3);
double sd2=Math.sqrt(var1*var1/cnt1+var2*var2/cnt2);
double tvalue2=meand/sd2;
double p2=result3[0]; %>
<table border="1" cellspacing="0" cellpadding="7">
<tr align="center">
<td width=100px>변수</td><td width=100px><%=colname[1]%"> </td> <td
width=100px>평균 </td> <td width=108px>표준편차 </td>
<td>평균 차이 </td> </tr> <tr align="center">
<td rowspan="2"><%=colname[0]%"> </td> <td>1 </td>
<td><%=BS.getCap3(mean1)%"> </td> <td><%=BS.getCap3(std1)%"> </td>
<td rowspan="2"><%=BS.getCap3(meand)%"> </td> </tr>
<tr align="center">
<td>2 </td> <td><%=BS.getCap3(mean2)%"> </td> <td>
<%=BS.getCap3(std2)%"> </td> </tr>
<table border="1" cellspacing="0" cellpadding="7">
<tr align="center">
<td>Test for equal variance</td> <td rowspan=2 width=75px>Variance</td> <td
colspan=4>t-test for equality of means</td> </tr>
<tr align="center">
<td>p-value</td> <td width=55px>df</td> <td width=55px>t-value</td> <td
width=55px>SD</td> <td width=55px>p-value</td> </tr>
<tr align="center">
<td rowspan=2><%=BS.getCap3(pF)%"> </td>
<td>Equal </td>
<td><%=BS.getCap0(df1)%"> </td>
<td><%=BS.getCap3(tvalue1)%"> </td>
<td><%=BS.getCap3(sd1)%"> </td>
<td><%=BS.getCap3(p1)%"> </td>
</tr> <tr align="center">
<td>Unequal </td>
<td rowspan=2><%=BS.getCap0(df2)%"> </td>
<td rowspan=2><%=BS.getCap3(tvalue2)%"> </td>
<td rowspan=2><%=BS.getCap3(sd2)%"> </td>
<td rowspan=2><%=BS.getCap3(p2)%"> </td>
</tr> </table> </center></body></html>

```

```

Class resource code for two sample equal variance test using java
public double[] getFTest(String tbname, String[] colName){
String sql = "select STATS.F_TEST(" + colName[1]
+ "," + colName[0] + ") P from " + tbname; Connection conn = getConn();
PreparedStatement pstmt = null; ResultSet rs = null;
try{ pstmt = conn.prepareStatement(sql,ResultSet.TYPE_SCROLL_
INSENSITIVE, ResultSet.CONCUR.READ_ONLY);
rs = pstmt.executeQuery(); double[] result = new double[2];
String P="P"; while(rs.next()){result[0]= rs.getDouble(P); }return result;
}

Class resource code for two sample t-test (same variance) using java
public double[] getIndepTtest(String tbname, String[] colName){
String sql = "select STATS.T_TEST_INDEP
(" + colName[1] + "," + colName[0] + ") P from " + tbname;
Connection conn = getConn();
PreparedStatement pstmt = null; ResultSet rs = null;
System.out.println(sql);
NumberFormat nfa = NumberFormat.getInstance();
nfa.setMaximumFractionDigits(0);
NumberFormat nfb = NumberFormat.getInstance();
nfb.setMaximumFractionDigits(3);try{
pstmt = conn.prepareStatement(sql,ResultSet.
TYPE_SCROLL_INSENSITIVE, ResultSet.CONCUR.READ_ONLY);
rs = pstmt.executeQuery();
double[] result = new double[2];String P="P";while(rs.next()){
result[0]= rs.getDouble(P);
System.out.println(" result[0]=" +result[0]);}return result;
}

Class resource code for two sample t-test (unequal variance) using java
public double[] getIndepTtest(String tbname, String[] colName){
String sql = "select STATS.T_TEST_INDEPU
(" + colName[1] + "," + colName[0] + ") P from " + tbname;
Connection conn = getConn();
PreparedStatement pstmt = null; ResultSet rs = null;
System.out.println(sql);
NumberFormat nfa = NumberFormat.getInstance();
nfa.setMaximumFractionDigits(0);
NumberFormat nfb = NumberFormat.getInstance();
nfb.setMaximumFractionDigits(3);try{
pstmt = conn.prepareStatement(sql,ResultSet.
TYPE_SCROLL_INSENSITIVE, ResultSet.CONCUR.READ_ONLY);
rs = pstmt.executeQuery();
double[] result = new double[2];String P="P";while(rs.next()){
result[0]= rs.getDouble(P);
System.out.println(" result[0]=" +result[0]);}return result;
}
    
```

본 연구에서 사용한 환자 자료에 대해 개발된 동적 시스템에서 독립표본 t-test를 실행한 결과는 Figure 3.5와 같다. 종속변수를 선택할 수 있는 메뉴와 그룹을 선택할 수 있는 메뉴로 구성하였다. Figure 3.5의 왼쪽 그림은 종속변수를 선택하는 화면, 오른쪽 그림은 그룹을 선택하는 화면을 캡처한 그림이다. 여기에 나타나는 변수는 선택한 자료에 따라 동적으로 나타나게 된다.

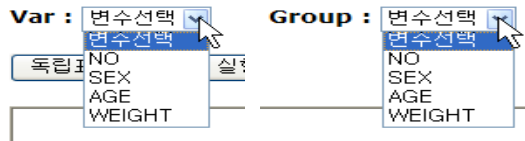


Figure 3.5 Variable and group selection interface of two sample t-test

또한 Figure 3.6은 성별에 따른 나이의 평균 차이를 검정하기 위해서 종속변수에는 AGE를 선택하고 그룹에는 SEX를 선택하고 나서 독립표본 평균비교를 실행한 화면을 캡처한 그림이다. 각 그룹에 따른 평균, 표준편차 값이 나타나고 평균차이값이 나타나고 또한 분산 동질성에 대한 검정결과에 따른 유의확률 값, 표준편차가 같은 경우와 다른 경우에 따른 각 자유도, 검정통계량값인 t-value, 합동표준편차 그리고 유의확률 (p-value) 값이 나타난다.

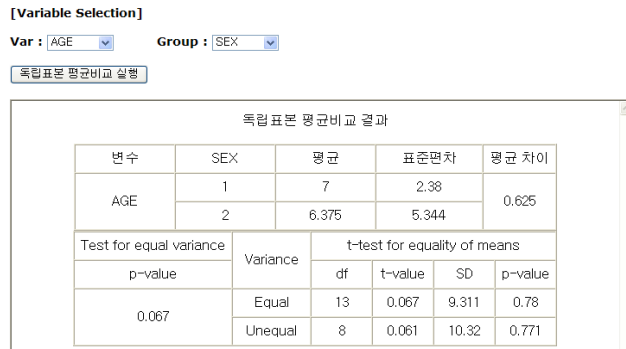


Figure 3.6 Statistical analysis result interface of two sample t-test



#### 4. 결론

개인 또는 서버 컴퓨터 통계분석 프로그램인 SAS, SPSS, S-LINK, MINITAB, Stata, MedCalc, R, Excel 등은 모두 동적 자료에 따른 통계분석 프로그램으로 개발되어 있다. 그러나 몇몇의 프로그램은 개인용 컴퓨터에 설치가 되어야 하고 R을 제외하고는 각각 고가의 사용권을 구입하여 사용하여야 하므로 통계분석이 물리적·공간적으로 시간과 위치의 제한을 받는 것을 의미한다. 이런 제한을 해결하기 위하여 Kim 등 (2011)가 웹에서 사용되어 지는 HTML, java, JSP 등의 언어만 사용하여 특정 통계분석 프로그램의 설치 없이 통계분석이 가능한 시스템을 개발하였으나 이는 동적이 아닌 정적 시스템이었다. 본 연구는 기존의 연구의 한계점을 극복하였다는 점에서 큰 강점을 지닌다고 할 수 있다. 첫째, 웹 브라우저에서 원하는 형태의 자료를 사용자가 직접 서버에 저장할 수 있도록 하였고 둘째, 웹프로그래밍 언어만을 사용하여 여러 가지 통계분석 시스템을 개발하였다. 마지막으로 사용자가 선택한 자료를 이용하여 통계분석을 할 수 있도록 하였다. 아울러 본 연구에서는 엑셀 프로그램만을 탑재하는 시스템을 개발하였는데 향후 다른 파일 형태도 웹기반으로 탑재할 수 있는 시스템을 적용하면 더 많은 자료에 대한 접근이 가능할 수 있을 것이다.

#### References

- Diana, L. and Lex, H. (2005). *Oracle database SQL reference 10g release 2 (10.2)*, Oracle, California, USA.
- Hwang, J. S. and Uhm, D. H. (1999). Statistical analysis on the web using PHP3. *Journal of the Korean Data & Information Science Society*, **10**, 501-510.
- Kim, D. H., Shin, I. H., Choe, J. Y., Kim, S. G., Park, C. W. and Kwak, S. G. (2012). Development of web based system for statistical analysis of clinical data. *Journal of the Korean Data & Information Science Society*, **23**, 191-198.
- Lee, K. J., Park, H. C. and Park, J. P. (1999). An implementation of statistical information system on the database interface using java on the www. *Journal of the Korean Data & Information Science Society*, **1**, 261-269.
- Lee, J. J., Lee, T. R., Kim, S. S. and Lee, G. H. (2004). Development of statistical education system for remote education. *Proceedings of the Spring Conference of the Korean Statistical Society*, 163-166.
- Oh, C. H., Lee, S. H. and Lee, H. J. (2006). A web-based SAS system for lab statistics. *Journal of the Korean Data & Information Science Society*, **17**, 395-400.
- Ryu, K. H. and Park, H. C. (2006). Web-based DNA microarray data analysis tool. *Journal of the Korean Data & Information Science Society*, **17**, 1161-1167.
- Shin, I. H. (2008). *Solution medical statistics series 1*, Koonja, Seoul.

## Development of web-based system for dynamic statistical analysis of clinical data<sup>†</sup>

Im Hee Shin<sup>1</sup> · Sang Gyu Kwak<sup>2</sup> · Jun Woo Park<sup>3</sup>

<sup>123</sup>School of medicine, Catholic University of Daegu

Received 14 October 2013, revised 25 October 2013, accepted 5 December 2013

### Abstract

Statistical analysis provides information that can be applied to draw final decisions in many fields. However, statistical analysis program for PC (personal computer) is yet restricted by time and space. To minimize this issue, a server based PC statistic analysis program using internet in addition to web based system allowing statistical analysis have been continually developed. However, the current web based analysis system is limited to the data that is saved on the server. Data that is modified or newly inserted must go through a server administrator before its use in analysis. In order to solve this problem, we have developed a web based system using HTML, java, JSP scripts to incorporate dynamic data without much restriction.

*Keywords:* Dynamic data, statistical analysis, web, web-programing language.

---

<sup>†</sup> This study was supported by the grant of Korean Ministry of Health & Welfare, Republic of Korea (Project No : 3033-302).

<sup>1</sup> Professor, Department of Medical Statistics, School of Medicine, Catholic University of Daegu, Daegu 705-718, Republic of Korea.

<sup>2</sup> Corresponding author: Research fellow, Department of Medical Statistics, School of Medicine, Catholic University of Daegu, Daegu 705-718, Republic of Korea. E-mail: sanggyu39@knu.ac.kr

<sup>3</sup> Researcher, Department of Medical Statistics, School of Medicine, Catholic University of Daegu, Daegu 705-718, Republic of Korea.