

## Internet Poll System<sup>1)</sup>

Yon Hyong Kim<sup>2)</sup> Min Gweon Oh<sup>3)</sup>

### Abstract

In this paper we propose a poll system on the internet. This system expects to increase the confidence of the internet poll results by sampling theory(proportional allocation). This system provides a cross-table and result of hypothesis test which plays an important role for decision making. These results do offer a few statistical packages(such as SAS, SPSS) in the world wide web.

*Keywords* : cgi, package, sampling, hypothesis test

### 1. 서 론

오늘날은 인터넷을 통한 정보화, 세계화 시대라고 해도 지나치지 않을 정도로 인터넷을 이용하여 방대한 량의 자료를 쉽게 열람하거나 원하는 자료를 신속하게 얻을 수 있다. 따라서 이제는 연구자가 연구 대상의 자료를 얼마만큼 신속하게 분석하여 원하는 결과를 얻을 수 있는지가 주된 관심의 대상이 될 수 있다. 특히, 국내 인터넷 사용자수가 이미 1천만명을 넘어섰고 컴퓨터 보급률과 인터넷 사용률이 지속적으로 증가하고 있어 인터넷을 학문분야에 접목시키려는 많은 연구들이 시도되고 있다. 이러한 사회 전반적인 추세에 의해 통계학의 한 분류인 표본론의 여론조사 방법도 기존의 면접조사, 우편조사, 전화조사, 직접조사 등의 방법과 더불어 인터넷 여론조사라는 새로운 조사방법이 등장하게 되었다.

인터넷 여론조사는 전 세계로 연결되어 있는 인터넷 망을 이용하여 시간이나 장소의 제한에서 벗어나 신속하게 많은 응답자를 확보할 수 있다는 장점 때문에 국내 유관기관 및 여론조사 기관에서 인터넷 여론조사를 시행하고 있다.

현재 국내 정부기관이나 투자기관, 일반회사 및 언론사 등이 실시하고 있는 인터넷 여론조사 방법 및 제공되는 결과는 대체적으로 비슷하고 다음과 같은 몇 가지 문제점들이 있다.

첫째, 단순히 전체 응답자에 대한 각 범주의 빈도수와 비율만을 제시될 뿐 다른 설문 문항과의 비교 분석이 이루어지지 않는다.

둘째, 웹 브라우저 화면 상단의 「새로고침」 버튼을 누르면 총 응답자수가 증가하는 경우도 있어 표본(sample)에 대한 신뢰성이 결여된다.

셋째, 최종적인 의사결정자에게 제공되는 통계적 자료분석 결과가 제공되지 못하고 SAS, SPSS,

1) This research was supported by the Jeonju University Research Fund in 1998.

2) Professor, Department of information statistics, Jeonju University, Jeonju, 560-759, Korea.

E-mail : yhkim@jeonju.ac.kr

3) Additional Instructor, Dept. of information statistics, Jeonju University, Jeonju, 560-759, Korea.

Minitab 등의 통계 소프트웨어를 추가로 사용해야 된다.

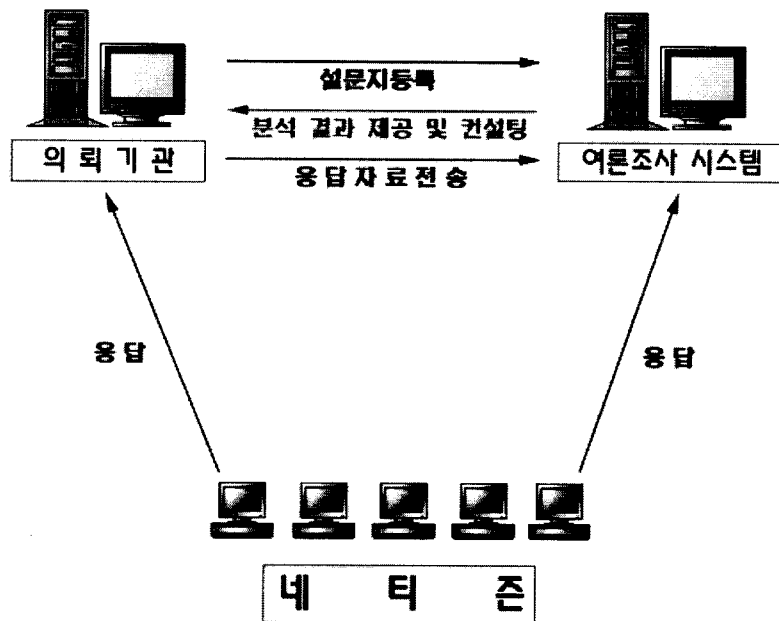
네째, 여론조사를 실시하는 목적에 부합하는 설문 문항이 결여되어 있으며, 기초항목(성별, 연령별, 지역별 등등)에 따른 분석이 제공되지 않는다.

이러한 제반 문제점들은 인터넷 여론조사 결과에 대한 신뢰도를 저하시키는 직접적인 원인이며, 특히 표본에 대한 대표성 문제가 심각하게 제기되고 있다.

본 연구에서는 인터넷 여론조사의 여러 문제점들을 수정 보완한 인터넷 여론조사 시스템을 제안하고자 한다. 구현된 시스템은 설정된 모집단에 대해 표본오차와 허용오차를 이용하여 표본수를 결정한 후 인터넷 홈페이지를 통해 응답한 표본들 중에서 비례배분 방법으로 배정된 표본들만 분석 자료로 활용함으로써 여론조사에 대한 신뢰성 및 표본의 대표성 문제를 보완하였다. 또한, 설문문항들에 대한 2차원 분할표와 독립성 검정결과를 실시간에 제공하며, 3차원 이상의 자료는 홍종선, 김동욱, 오민권(1999)이 제안한 범주형자료분석(홍종선(1995), 남궁 평, 홍종선(1996), Agresti, A. (1996)) 시스템을 이용할 수 있도록 하였다.

## 2. 여론조사 시스템

본 연구에서는 인터넷 여론조사의 여러 제반 문제들을 보완하고 특히 표본의 대표성 및 여론조사 결과에 높은 신뢰성을 부여할 수 있는 실시간 여론조사 시스템을 제안하고자 한다. 개발된 시스템은 널리 이용되고 있는 Linux 서버환경에서 MySQL 데이터베이스와 웹 프로그래밍 언어인 Perl와 자바(java)를 사용하여 구현하였다. 사용자의 설문 입력과 응답자료에 대한 2차원분석 그리고 표본추출 과정은 Perl로 구현된 CGI가 담당하고, 3차원 이상의 범주형자료에 대한 분석은 자바 애플릿(applet)에서 처리되도록 설계하였다.



<그림 1> 인터넷 여론조사 시스템 구성도

<그림 1>은 개발된 시스템의 구성도로서 인터넷 여론조사 시스템은 크게 여론조사 의뢰기관과 여론조사 시스템이 탑재된 서버 운영기관 그리고 설문 응답자인 네티즌들로 구성된다.

<그림 1>에 나타나 있는 것과 같이 개발된 시스템은 Perl로 구현된 CGI(Common Gateway Interface, 이규석 (1997))를 통해 여론조사 의뢰기관에서 설문문항을 직접 등록할 수 있으며, 네티즌들의 누적된 응답결과가 실시간에 분석되어 응답자 및 의뢰기관에 제공된다. 또한, 의뢰기관의 홈페이지를 방문하는 네티즌들이 많을 경우 의뢰기관의 홈페이지와 여론조사 시스템을 운영하는 서버를 서로 연결(link)하여 보다 많은 응답자를 빠른 시간 안에 확보할 수도 있다.

개발된 인터넷 여론조사 시스템은 크게 8개의 윈도우(설문지등록, 응답하기, 응답결과, 자료분석, 표본추출, 패널현황, DataBase, 여론조사신청)로 구분되는데, 각 윈도우에 대한 세부적인 내용은 다음과 같다.

### 2.1 설문지등록 및 응답하기

설문지명	종교와 혈액형과의 관계
의뢰기관	한국정보통계
조사기간	2000년 2월 7일부터 2000년 2월 29일까지
의뢰담당자	이름: 오민권    전화번호: 0652-288-04    E-mail: mgoh@iands.cc

<그림 2> 설문 의뢰 기관에 대한 정보를 입력하는 형식

<그림 2>는 [설문지 등록] 버튼을 클릭했을 때 나타나는 결과로 여론조사를 의뢰하는 기관에 대한 정보를 입력하는 여러 개의 텍스트 필드(Text field)로 구성되어 있다. [설문지 등록] 윈도우의 하단부에는 설문문항을 입력하는 폼(form)이 나타나는데 본 논문에서는 언급을 생략하고자 한다.

설문지를 작성 후 [전송하기] 버튼을 누르면 CGI에 의해 설문지 리스트에 추가되는데 이는 응답하기 버튼을 클릭하여 확인할 수 있다. 네티즌들은 해당 리스트를 클릭하여 나타나는 설문지에 응답 후 [결과보기] 버튼을 클릭하면 현재까지 응답한 자료에 대한 빈도수 및 2차원 분석결과를 열람할 수 있다.

설문지의 하단부에 있는 설문 응답자의 이름과 전화번호를 기입하는 두 개의 텍스트 필드는 해당 기관의 홈페이지에 자주 접속하는 인터넷 이용자가 여러 번 설문에 응하는 것을 배제하기 위한 것으로, 구현된 시스템에서는 표본에 대한 신뢰성을 높이기 위해 같은 이름과 전화번호로 한번 이상 설문에 응답할 수 없도록 구현하였다.

또한, 시스템은 특정 문항에 대한 무응답이 발생했을 때 시스템에서 자동으로 응답자에게 안내 메시지를 제공하여 실질적인 무응답인지 아니면 실수로 누락한 것인지를 체크하도록 구현하였다.

### 2.2 결과 보기

<그림 3>은 응답자가 설문에 응답한 후 『분석결과보기』 버튼을 눌렀을 때와 분석 결과를 원하는 네티즌이 [응답결과] 버튼을 클릭 후 나타나는 목록에서 특정 여론조사(혈액형과 종교와의

관계)을 클릭했을 때 나타나는 결과로 오른쪽에 전체응답자수 및 각 문항에 대한 빈도수와 비율이 제공된다.

1. 성별은?	94명	100%
① 남자	67명	71.27%
② 여자	27명	28.72%
2. 혈액형은?	94명	100%
① A	42명	44.68%
② B	20명	21.27%
③ AB	12명	12.76%
④ O	20명	21.27%
3. 종교는?	94명	100%
① 기독교	32명	34.04%
② 불교	12명	12.76%
③ 천주교	13명	13.82%
④ 무교	37명	39.36%
4. 종교와 혈액형이 관계있다고 생각하는가?	94명	100%
① 있다	5명	5.31%
② 없다	89명	94.68%

<그림 3> 누적된 설문 결과 예제

응답자 및 여론조사 의뢰기관은 누적되는 설문 자료들에 대해서 언제든지 <그림 4>에 나타나있는 기초문항(1번과 2번)과 설문문항(3번과 4번)에 대해 비교 분석할 수 있다.

<그림 4>에서 1번과 3, 4번 그리고 2번과 3, 4번이 선택 가능하도록 나뉘어져 있는 것은 설문을 입력할 때 1번과 2번 문항은 기초문항으로 3번과 4번 문항은 설문문항으로 구분하여 입력했기 때문이다.



<그림 4> 범주형 자료분석을 실시하기 위해 두 변수를 선택하는 형식

<그림 5>은 사용자가 <그림 4>에서 2번 문항(혈액형)과 3번 문항(종교)의 분석을 선택한 경우에 나타나는 결과로 혈액형과 종교에 대한 2차원 분할표이다. 일반적으로 이러한 분할표는 누적된 설문 자료를 코딩한 후 통계 소프트웨어(SAS, SPSS, Minitab 등)를 이용하여야 얻을 수 있는 결과이다. 시스템에서 제공되는 2차원 분할표의 내용은 많이 사용되는 SAS의 분석결과와 동일하게 설계하였다.

당신이 선택한 두 변수에 대한 2차원 분할표 (Contingency Table)					
	범주 1	범주 2	범주 3	범주 4	전체
범주 1	14명 14.89 % 43.75 % 33.33 %	5명 5.31 % 41.66 % 11.9 %	8명 8.51 % 61.53 % 19.04 %	15명 15.95 % 40.54 % 35.71 %	42명 44.68 % 100.00 %
범주 2	7명 7.44 % 21.87 % 35 %	3명 3.19 % 25 % 15 %	4명 4.25 % 30.76 % 20 %	6명 6.38 % 16.21 % 30 %	20명 21.27 % 100.00 %
범주 3	3명 3.19 % 9.37 % 24.99 %	3명 3.19 % 25 % 24.99 %	0명 0 % 0 % 0 %	6명 6.38 % 16.21 % 49.99 %	12명 12.76 % 100.00 %
범주 4	8명 8.51 % 25 % 40 %	1명 1.06 % 8.33 % 5 %	1명 1.06 % 7.69 % 5 %	10명 10.63 % 27.02 % 50 %	20명 21.27 % 100.00 %
전체	32명 34.04 % 100.00 %	12명 12.76 % 100.00 %	13명 13.82 % 100.00 %	37명 39.36 % 100.00 %	94명 100.00 %

<그림 5> 혈액형과 종교에 대한 2차원 분할표

<그림 6>은 사용자가 선택한 두 변수(혈액형과 종교)에 대한 독립성 검정 결과이다. 사용자는 분석을 원하는 설문문항에 대해 언제든지 분할표와 검정결과를 얻을 수 있으므로 보다 신속하게 여론 수렴 및 의사결정에 반영할 수 있을 것이다.

당신이 선택한 두 변수의 독립성 검정(χ<sup>2</sup> 검정) 결과

지금까지 응답한 누적된 자료를 토대로 당신이 선택한 [변수1]와 [변수2]의 각 항목[범주]에 대한 비율이 같은지를 판단하기 위해 다음과 같은 가설을 설정할 수 있습니다.

귀무가설 : [변수1]와 [변수2]의 각 항목에 대한 비율이 같다( 또는 통계적으로 독립).  
 대립가설 : [변수1]와 [변수2]의 각 항목에 대한 비율이 같지 않다(또는 통계적으로 종속).

위의 가설에 대한 자유도(Df: Degrees of Freedom)는 9 이고, χ<sup>2</sup>은 8.63424032714328 이며, 이에 대응되는 p-값은 0.976743207151882 이다.

따라서 유의수준(Alpha)를 0.05로 정하는 경우

유의수준(0.05) < p-값(0.976743207151882) 이므로 귀무가설을 채택합니다.

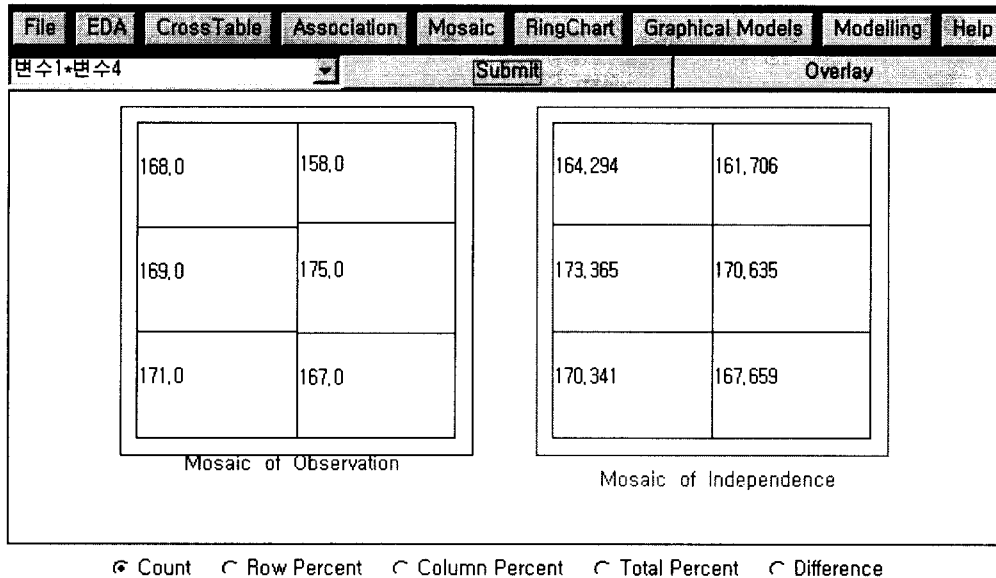
즉, 두 변수의 각 항목에 대한 비율이 다르다고 할 수 없습니다[통계적 독립].

<그림 6> 혈액형과 종교에 대한 독립성 검정 결과

### 2.3 고차원 자료에 대한 범주형 자료 분석

여론조사의 세문항 이상 고차원자료에 대한 분석은 홍종선, 김동욱, 오민권(1999)이 제안한 <그

림 7>의 자바로 구현된 범주형자료 분석 시스템을 이용할 수 있다. 이 시스템은 자바로 구현되어 O/S에 영향을 받지 않는다는 특징이 있으므로 UNIX, LINUX, NT 등 어떤 서버운영체제 내에서도 제안된 여론조사 시스템과 홈페이지 내에서 상호 연결이 가능하다. 특히, 현재 많이 사용되는 통계 패키지들에서 제공하지 못하는 모자의 그림과 연관그림 등이 동적그래픽스로 구현되어 있어 연관성을 측정하거나 모형을 설정하는데 매우 유용한 시스템이다.



<그림 7> 자바로 구현된 범주형 자료분석 시스템

## 2.4 표본설계

일반적으로 여론조사의 모집단은 전국에 거주하는 인구수로 하지만 인터넷 여론조사는 인터넷이라는 특수한 매체를 통해 실시되므로 연령, 지역, 직종에 따른 범주를 고려하는 경우 상대적으로 인터넷 이용률이 낮은 층에 대한 응답자 확보가 어렵고, 특히 표본의 대표성 문제가 심각하게 제기되고 있다. 본 논문에서는 보다 신뢰성 있는 여론조사를 수행하기 위해 다음과 같은 네 가지를 고려한 응답자를 표본으로 활용하는 방안을 제시하고자 한다.

### ◆ 추출틀

현재 인터넷 여론조사는 해당 홈페이지에 접속한 모든 네티즌을 표본으로 이용하고 있는데, 제안된 시스템에서는 각 도의 주민등록 인구통계자료를 추출틀로 한다.

### ◆ 표본크기 결정

표본의 크기는 목표정도를 달성할 수 있는 범위에서 최소로 하는 것을 원칙으로 하며, 여론조사 결과를 평가하는데 있어 지표로 사용되는 표본오차를 다섯 가지(10%, 7.5%, 5%, 2.5%, 1%) 수준, 허용오차 범위를 표본오차의 1/2 수준으로 하는 표본수를 사용한다.

◆ 표본배분

결정된 표본의 크기를 조사 대상인 모집단의 각 지역별로 배분하기 위해서 각 지역별 인구 수 비율에 따라 각 시군의 표본으로 할당한다.

◆ 표본추출

홈페이지를 접속한 네티즌들의 응답 결과는 일단 모두 표본으로 받아들여지지만 분석을 하는 단계에서 비례추출 방법으로 할당된 각 지역의 표본수 만큼만 랜덤화 방법으로 다시 표 본추출을 한 후 해당되는 표본들만 분석에 사용한다.

조사 지역 선택	<input type="text" value="전라북도"/>	
시군 이름	<input type="text" value="전주시"/>	※ 조사 지역의 시군이름을 기입하십시오
조사 대상 성별	<input type="text" value="전체"/>	
조사 대상 선택	<input type="text" value="전체"/>	
표본오차	<input type="text" value="5"/>	%

<그림 8> 표본추출 예제

<그림 8>은 사용자가 [표본추출] 버튼을 클릭한 경우 나타나는 윈도우로 사용자는 조사지역, 시군, 조사대상 성별, 조사대상 연령, 표본오차를 선택한 후 [확인 버튼]을 클릭하면 <그림 9>가 나타난다.

구 분	10세 이하		10 - 19		20 - 29		30 - 39		40 - 49		50 - 59		60 - 69		70 - 79		80세 이상	
	전체	표본	전체	표본	전체	표본	전체	표본	전체	표본	전체	표본	전체	표본	전체	표본	전체	표본
중앙동	617	1	1060	0	1528	3	926	2	1050	2	888	2	513	1	260	0	90	0
풍남동	811	1	1392	0	1884	3	1090	2	1168	2	1052	2	678	1	320	1	125	0
교동	653	1	929	0	1342	2	1211	2	929	2	1013	2	732	1	317	1	75	0
태평동	769	1	1166	0	1658	3	975	2	993	2	908	2	573	1	141	0	28	0
중노송1동	914	2	1498	0	1818	3	1175	2	1127	2	1008	2	647	1	285	1	126	0
중노송2동	566	1	793	0	1214	2	692	1	607	1	594	1	414	1	182	0	44	0
남노송동	694	1	1057	0	1339	2	849	2	773	1	730	1	530	1	248	0	86	0

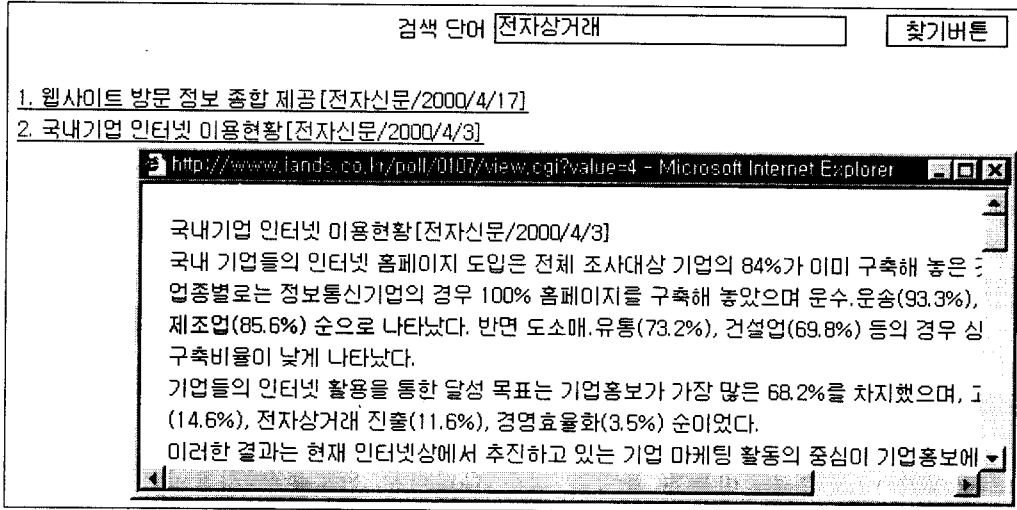
중략

호성동	3846	7	3960	0	3215	6	4465	8	2940	5	1304	2	919	2	482	1	194	0
송천1동	5241	9	4222	0	3847	7	5741	10	2952	5	1240	2	810	1	507	1	157	0
송천2동	4393	8	4371	0	3987	7	4900	9	3239	6	1721	3	1064	2	554	1	193	0
조촌동	1874	3	1635	0	2569	5	2222	4	1285	2	946	2	881	2	393	1	130	0
동산동	3014	5	3201	0	3275	6	3594	6	2284	4	1422	3	1033	2	529	1	165	0

<그림 9> 선택한 모집단 정보 및 비례추출 방법에 의한 표본수

<그림 9>는 <그림 8>의 선택 사양에 대한 결과로 전체 모집단 600,343명에 대해 허용오차 ± 0.025 수준에서 표본수 1,072명을 각 동별로 비례배분한 할당된 표본수와 모집단 인구수가 나타나 있다. 실제로 구현된 시스템은 모든 네티즌들의 응답은 받아들이지만 분석할 경우에는 <그림 9>에 할당된 표본들만 분석 자료로 이용된다.

### 2.5 여론조사 데이터 베이스



<그림 10> 여론조사결과 검색 예제

제안된 시스템에서는 각 종 여론조사 결과들을 데이터베이스로 구축하여 사용자들에게 검색기능을 제공한다. <그림 10>은 사용자가 키워드 “전자상거래”를 검색한 경우의 예로 두 개의 검색 결과 목록과 국내인터넷 이용현황에 대한 내용을 열람한 상태의 예제 그림이다.

### 2.6 회원 현황 보기

구분	서울	부산	대전	대구	인천	광주	경기	강원	충북	충남	전북	전남	경북	경남	제주	전체
일반인	205	14	9	9	17	4	37	2	4	3	49	5	4	10	3	375
공무원	3	1	0	1	0	1	3	0	1	0	8	0	1	0	0	19
교육관련	9	1	0	3	1	0	4	1	0	0	16	0	0	0	0	35
자영업	3	0	0	0	0	0	5	0	0	0	5	0	0	0	0	13
대학생	187	28	19	23	39	23	104	7	17	26	685	37	11	23	3	1232
고등학생	1	1	1	0	1	0	1	0	0	1	25	0	0	1	0	32
중학생	1	0	0	0	0	0	1	0	0	0	10	0	0	1	0	13
초등학생	0	0	0	0	0	1	0	0	0	0	13	0	0	0	0	14
유치원생	2	0	0	0	0	0	0	0	0	0	3	0	2	0	0	7
전체	411	45	29	36	58	29	155	10	22	30	814	42	18	35	6	1740

<그림 11> 패널 현황



구현된 시스템에서 [패널현황] 버튼을 클릭하면 <그림 11>와 같이 직업과 지역별 회원수 현황을 파악할 수 있다. 이는 신속한 설문조사를 요하는 경우 해당되는 패널들에게 전자우편(e-mail) 등을 발송하여 여론조사에 응답할 것을 요청하거나, 보다 신뢰성있는 여론조사를 수행하기 위해 전체 응답자 자료를 모집단으로 하여 층화추출이나 네이만 최적배분 등의 방법으로 새로운 표본을 추출할수 있도록 하기 위함이다. 인터넷 여론조사는 신속하게 많은 응답자를 확보하는 것이 최대의 장점이므로 신속하게 많은 응답자를 확보하기 위해서는 인터넷 이용률이 매우 높은 포털사이트 및 유관기관의 홈페이지와 연계하거나 경품행사를 개최하는 등의 노력이 필요하다.

### 3. 결 론

인터넷 여론조사는 다른 여론조사 방법에 비해 신속하게 많은 자료를 수집할 수 있다는 장점 때문에 많은 관련 기관에서 시행하고 있으며, 앞으로 발전 가능성이 매우 높은 여론조사 방법이다. 그러나 인터넷이라는 특수한 매체를 통해 조사가 이루어지므로 지역별, 연령별, 직업별 등의 여러 범주들에 대한 분석을 하기에는 표본의 대표성 문제가 제기된다.

본 연구에서는 기존의 표본조사 방법에 입각한 인터넷 여론조사 시스템을 제안하였다. 제안된 시스템은 표본추출방법에 의한 응답자 확보, 이차원 자료에 대한 분할표 및 독립성 검정결과 제공, 고차원 자료에 대한 범주형 자료분석이 실시간에 이루어지므로 의사결정에 많은 도움이 될 것이다.

그러나 인터넷 여론조사는 인터넷 이용자수가 1천만명이 넘어섰다고는 하지만 아직까지는 인터넷이라는 한정된 범위 안에서 조사가 이루어지므로 연령별에 대한 분석이나 인터넷의 이용률이 상대적으로 낮은 지역을 포함해야 하는 여론조사인 경우에는 기존의 조사방법(전화조사, 면접조사 등)들과 병행해야 표본오차를 줄일 수 있을 것이다.

본 논문에서 제안한 여론조사 시스템은 <http://www.iands.co.kr>로 접속하여 직접 사용할 수 있다.

### 참고문헌

- [1] 남궁 평, 홍종선 (1996). 범주형 자료분석, 탐진.
- [2] 이규석 (1997). CGI 프로그래밍 언리쉬드, 정보문화사.
- [3] 홍종선 (1995). 대수선형모형, 자유아카데미.
- [4] 홍종선, 김동욱, 오민권(1999), 인터넷에서 범주형자료분석 시스템개발, 응용통계연구, 제12권 1호, 83-95.
- [5] Agresti, A. (1996), *An Introduction to Categorical Data Analysis*, John Wiley and Sons.