

인터넷 조사에서 무응답 오차에 관한 연구

A Study on Nonresponse Errors in the Internet Survey

남궁 평*, 김 민정**
Pyong Namkung · Min Jung Kim

인터넷 조사는 전통적인 조사방법에 비해 신속하고 저렴하며 멀티미디어를 이용한 고도화된 설문을 사용할 수 있다는 장점이 있는 반면 표본을 확률 추출하는 것이 어렵고, 대표성, 무응답 등의 비표본 오차가 심각하다. 본 연구에서는 비표본 오차 중 무응답 오차를 사례와 함께 정리하고, 인터넷 조사가 새로운 조사 방법으로서 활용될 수 있는 대안을 제시한다.

I. 서 론

통계조사에서 모든 대상으로부터 모든 필요한 정보를 얻을 수 있다면 모집단의 특성을 추정하거나 추론하는 과정에서 발생하는 비표본 오차는 발생하지 않을 것이다. 특히, 사회가 개인주의화되고 복잡해지면서 표본으로 선정된 대상으로부터 완벽한 응답을 얻기가 어려워지면서 비표본 오차 중 무응답의 연구가 활발해지고 있으며 다양한 무응답의 대체(imputation) 방법들이 제시되고 있다.

최근 통계 조사방법 중 컴퓨터와 통신 기술의 발달하면서 신속성, 정확성, 경제성이라는 새로운 생활환경에 맞춰 컴퓨터를 이용한 인터넷 조사(internet survey)방법이 발전하고 있다. 그러나 인터넷 조사는 인터넷 사용자들을 대상으로 웹 또는 전자메일(E-mail)을 이용하여 설문을 진행하는 자료를 수집하는 방법으로써 전통적인 방법에 비해 자료수집이 신속하고, 비용이 저렴하며, 멀티미디어를 활용한 고도화된 설문을 설계, 자료처리와 분석이 조사와 동시에 진행될 수 있다는 장점이 있다(Batagelj 와 Vehovar, 1998). 반면, 인터넷 조사를 수행함에 있어 시스템의 안정성과 보안성 등의 기술적인 문제와 함께 표본추출, 응답률, 중복응답, 설문지 설계, 조사결과의 일반화 등의 비표본 오차의 문제가 심각하다. 비표본 오차의 문제 중 인터넷 조사로 포함오차(coverage error)가 점차 해결됨에 따라 무응답 오차가 심각하게 대두되고 있다(Mick P. Couper, 2000).

* 성균관대학교 통계학과 교수, namkung@skku.ac.kr

** Internet Metrix 마케팅팀 연구원, mjkim@internetmetrix.com

본 연구에서는 전통적인 조사방법에 대한 대체방법으로서의 인터넷 조사방법이 신뢰성과 타당성을 갖추고 있는지 알아보려고 한다. 문제점으로 지적 받는 사항 중 무응답 오차와 관련하여 인터넷 조사를 실시한 후, 실제 조사 상황에서 발생할 수 있는 기술적인 문제점 및 항목 무응답률, 단위 무응답률 등을 알아본다.

II. 인터넷 조사와 무응답 오차

국내 인터넷 사용자수가 2,412만 명을 넘는 것으로 보고되고 있으며 해마다 폭발적인 증가 추세에 있다. 설문조사에 있어서 인터넷 인구는 접근하기 어려운 전문가 집단이나 지역적으로 먼 거리에 있는 집단 등에 쉽게 접근이 가능하다. 또, 인터넷 조사가 전화를 걸거나, 면접을 통해 설문을 받는 것보다 노력이 적게 들고, 인터넷은 개별 응답자의 응답결과에 따라 필요 없는 질문은 자동적으로 생략하는 제어 기능이 가능하며, 멀티 미디어를 이용한 자료 제시 등이 가능하다. 이러한 장점하에 적절한 표본 추출틀의 확보와 응답자들의 응답률만 높아진다면 보다 신뢰할 수 있는 조사방법으로써 구현이 기대된다.

그러나 인터넷의 특성상 응답 자료의 질(quality)은 조사결과의 적용과 수집정보의 신뢰성이 가장 큰 걸림돌이 되고 있다. 인터넷 설문문사는 역사가 짧고 기술적, 방법론적으로 정확한 통제가 불가능할 뿐만 아니라 실제로 이러한 기술적 통제가 수집자료를 왜곡시키기도 한다.

이 장에서는 인터넷 조사도구로서 CGI(Common Gateway Interface)와 ASP(Active Server Page)를 통한 인터넷 설문조사 시스템의 운영원리에 대해 알아본다. 응답 데이터의 질을 측정할 수 있는 척도와 단위 무응답과 항목 무응답의 발생원인과 해결방안을 선행 연구로부터 알아보고, 설문지 디자인이 응답자에게 어떤 영향을 미치는지 연구해보고자 한다.

1. 인터넷 조사의 선행 연구

Schaefer와 Dillman(1998), Couper, Blair and Triplett(1999)의 연구에서는 전자우편 조사(e-mail survey)는 우편조사(mail survey)의 응답률 수준에 도달하는데 실패하였다. 우편조사에서와 같이 응답 재 요청, 회사명, 소재지, 전화번호 등의 인쇄된 편지지, 동기유발 등의 동기부여를 하였지만 실패하였다. 그 이유로는 우편 조사는 종이와 펜을 이용해 완성하기가 쉽지만 전자우편조사는 경험적인 양식이 적으며 사용자가 browser들이 동일하지 않고 컴퓨터와 속도와 인터넷 접속 등에서 인터넷과 친숙하지

않은 응답을 완성할 수 없다. 또, 얼마나 많은 사람들이 웹 설문지를 작성하다가 완성
에 실패했는지는 파악이 불가능하다. 몇몇 조사기관에서 수신 또는 발신 기록 등을
보관함으로써 조사 주제가 민감할수록 설문완성이 힘들어진다고 하였다.

한편, Metha와 Sivada(1995)의 연구에서는 1994년 2월 1일에서 1994년 4월 1일까지
20개의 인기 있는 newsgroup에서 기사를 쓴 6,000명의 전자우편주소를 확보하여 663
명의 표본을 계통추출법으로 추출하였다. 추출된 표본을 다섯 그룹으로 나누어 인터
넷의 상업화의 태도에 대한 내용의 5페이지의 19개 문항으로 구성된 설문지를 우편과
전자우편 두 가지 방법으로 발송하였다. 그 결과 우편은 45%, 전자우편으로는 40%의
응답률을 얻었고, 사전통보, 응답 재 요청, 인센티브를 제공한 그룹의 응답률은 우편
조사가 80%, 전자우편 조사가 63%의 응답률을 얻었다.

Nichols와 Sedivi(1998)은 미국 Census 국이 시행한 조사에서 Java를 이용한 설문
디자인은 아직 인터넷에 익숙하지 않은 응답자에게 응답률을 저해하는 요인으로 작용
한다고 지적하였다.

그 외 인터넷 설문조사와 관련한 연구로 객관적인 조사방법으로서의 위치정립을 위한
전통적인 조사방법과의 다양한 비교를 통해 인터넷 설문조사의 신뢰성과 타당성을 제시
하는 연구(Tes et al, 1995 ; Metha & Sivadas, 1995 ; Cooper et al, 1997 ; Dillman, 1998 ;
Stanton, 1998)와 인터넷 설문조사 내에서 표본의 문제를 개선하고, 수집된 자료의 질을
높이려는 연구(Comley, 1998; Couper & Triplet, 1997)가 활발히 진행되고 있다.

2. 인터넷 설문조사 방법

인터넷상에서 설문조사를 실시하게 위한 온라인 리서치 솔루션을 이용한 방법과 웹
프로그래밍을 이용한 방법이 있다. 리서치 솔루션은 리서치와 마케팅, 인터넷 솔루션
업체 등에서 앞다퉈 출시되고 있지만, 소프트웨어를 구입해야 한다는 단점이 있다. 반
면, CGI, Windows NT서버 환경에서 새로운 웹 테크놀로지로 각광받고 있는 ASP 등
을 이용한 웹 프로그래밍은 비용이 전혀 들지 않으면서, 간단한 프로그래밍만으로 인
터넷 설문조사가 가능하다. 사례연구에서 이용한 ASP 프로그래밍 인터넷 설문조사에
대해 주로 살펴보고자 한다.

1) CGI와 ASP

인터넷 설문조사에 있어 가장 중요한 데이터베이스와의 연동을 하는 기능을 제공하
여야 한다. CGI는 인터넷의 성장과 더불어 동적이고 상호 작용하는 웹서비스를 제공
하는 강력한 수단으로 오래 전부터 사용되어온 도구이다. CGI는 웹 서버와 외부 프로

그램을 통신하게 하거나, 또는 그 반대로 통신하게 하는 일종의 HTML로 외부 실행 파일을 실행해 그 결과를 볼 수 있다. CGI의 가장 일반적인 방식은 사용자가 "Submit"버튼을 누르면서 발생시키는 방식과 CGI를 HTML자체에 내장시켜 처리하는 방식으로 나눌 수 있는데, 인터넷 설문에서는 두 가지 방법을 복합적으로 사용하여 처리하게 된다.

ASP는 인터넷 서비스를 지원하는 IIS(Internet Information Sever), Windows NT에 포함된 IIS를 토대로 서버군의 운영체제 시장이 상당부분 Unix에서 Windows NT로 전환하는 시점에서 상당히 많은 사용자층을 확보하게 되었다. ASP는 WWW을 비롯하여 Gopher와 FTP 서비스를 지원하는데, 초창기 서버는 스크립트 소스자체를 클라이언트에게 전송해 주고, 클라이언트가 모두 분석하는 형태였다. 이때, 특히 문제가 되었던 것은 VBScript의 경우, Microsoft 익스플로러만 처리할 수 있었기 때문에, 웹 개발에 있어 상당히 제한적이었다. ASP는 이러한 문제를 해결하는 Server Side Script방식을 지원하고 보다 정교한 제어가 가능하게 되었다. ASP는 Active X의 서버 컴포넌트로서 IIS의 요청에 의해 서비스를 실행하는 기능을 담당한다. ASP가 Perl이나 Tcl/Tk등의 스크립트 언어와 가장 큰 차이점은 Active X 서버 컴포넌트를 실행시킬 수 있기 때문에 제공되는 강력한 확장성이다.

2) 인터넷 설문조사 시스템의 운영원리

인터넷 설문조사 시스템을 구축하기 위하여 ASP기술을 기반으로 인터넷에서 설문 조사를 하기 위해서는 기본적으로 Windows NT(IIS포함), DBMS(database management system), 데이터베이스 연동(聯動)을 위한 ASP, 홈페이지 작성을 위한 도구(Text편집기, 또는 웹에디터 프로그램)등이 필요하다.

설문조사 과정을 살펴보면 게시판에 게시된 설문사이트 주소와 발송된 전자우편을 통해 설문 사이트에 도달한 응답자의 응답은 웹사이트에서 "POST"³⁾되어, 서버와의 연결통로인 ASP프로그래밍을 거쳐, DBMS에 저장되어진다. DBMS저장결과는 다시 ASP를 통해 웹 브라우저에서 확인할 수도 있도록 만들었다. 최종적으로 DBMS에 저장된 데이터는 통계프로그램을 통해 분석되어진다.

3) 응답자들이 설문조사를 마치면 "Submit"버튼을 누르게 됨과 동시에 데이터는 서버로 전송되어진다. 이 때 데이터를 전송하는 방식은 "GET"과 "POST"방식을 따르게 되는데, GET방식은 프로그램이 환경변수를 사용하여 데이터를 서버로 전달한다. 입력된 내용과 접속경로 및 서버종류 등의 정보가 Query_String이라는 환경변수로 전달된다. POST방식은 데이터를 환경변수가 아닌 표준입력으로 처리하여 프로그램의 인수로 처리하게 된다. 여기서 입력길이에 제한을 받지 않으며, 많은 정보를 포함하기 위해서는 POST방식을 사용하는 것이 필수적이다. 대부분의 인터넷 설문조사의 경우 데이터의 전송방식은 POST방식으로 이루어지고 있다.

3) 설문지 디자인

인터넷 조사의 설문지 디자인은 Html로 이루어지는데, 가장 많이 사용되는 디자인은 라디오 버튼(Radio Button)([그림 2-2])이다. 라디오 버튼(Radio button)은 설문지에서 한 문항을 선택할 경우 사용하지만 선택 문항이 많은 경우에 설문지를 길게 할 뿐만 아니라 브라우저나 화면크기 설정에 따라 형태가 가장 많이 달라질 수 있다는 단점이 있다. 따라서 선택문항이 적은 경우에 많이 사용한다. 팝업은 무응답이 적게 나타나고 설문지가 긴 경우 공간을 줄일 수 있다는 장점이 있다. 그러나 선택문항을 모두 보기 위해 클릭한 후 스크롤 해야하는 번거로움이 있고, 응답이 위쪽 선택문항이 많이 치우칠 수 있다는 단점이 있다. 체크 박스는 중복 응답을 요구할 때 많이 사용되는 태그이다.

1. 귀하는 **본교에 지원한 가장 큰 이유**는 무엇이었습니까?

- 학업성적에 준해
 교수진이 좋아서
 사회평판이 좋아서
 주변의 권유로
 뚜렷한 이유없이
 기타

[그림 2-2] 라디오 버튼(Radio Button)

7. 귀하는 **성균관대학교 다니는 것에** 대하여 스스로 얼마나 만족하고 계십니까?

- 매우 만족 약간 만족 보통 약간 불만족 매우 불만족

[그림 2-3] 라디오 버튼(Radio Button) : 서열 척도

1. 귀하는 **본교에 지원한 가장 큰 이유**는 무엇이었습니까?

- 학업성적에 준해
 교수진이 좋아서
 사회평판이 좋아서
 주변의 권유로
 뚜렷한 이유없이
 기타

[그림 2-4] 체크 박스(Check box)

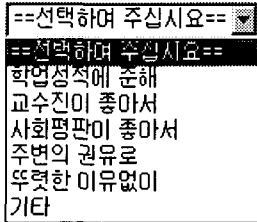
1. 귀하는 **본교에 지원한 가장 큰 이유**는 무엇이었습니까? 1순위() 2순위() 3순위()

- (1) 학업성적에 준해 (2) 교수진이 좋아서 (3) 사회평판이 좋아서 (4) 주변의 권유로
 (5) 뚜렷한 이유없이 (6) 기타

[그림 2-5] 엔트리 박스

인터넷 조사에서 무응답 오차에 관한 연구

1. 귀하는 본교에 지원한 가장 큰 이유는 무엇이었습니까?



[그림 2-6] 팝업

4. 귀하의 나이는? 만 세

[그림 2-7] 단 답형

1. 귀하는 본교에 지원한 가장 큰 이유는 무엇이었습니까?



[그림 2-8] 서술형

1. 귀하는 본교에 지원한 가장 큰 이유는 무엇이었습니까?



[그림 2-9] 자유 서술형

3. 항목 무응답

1) 설문지

질문지와 관련해 질문지의 구성, 문항의 복잡성, 질문의 형태 및 내용, 보조 자료 활용 등에서 여러 가지 항목 무응답 요인들이 있다. 특히 설문 디자인에 대한 비교에서 3가지 다중응답 질문에 팝업[그림 2-6], 엔트리 박스[그림 2-5], 체크박스[그림 2-4] 3가지 형태에서는 응답을 3개 이상 못하게 미리 제어한 팝업과 엔트리 박스는 응답오류가 거의 발생하지 않았지만 체크박스는 응답오류가 상당수 발생하였다. 팝업과 엔트리 박스에서는 클릭해서 기록해야하는 엔트리 박스보다는 클릭만으로 응답할

수 있는 팝업의 응답오류가 더 적게 나타났다. 또, [그림 2-5]의 단 답형(Short text)의 경우는 무응답률이 적었지만 [그림 2-8]의 서술형이나 [그림 2-9]의 자유서술형의 문항은 높은 무응답률로서 효율적인 조사 디자인 방법이 아니었다 (윤은성, 2000).

2) 응답자 인터페이스

인터넷 조사는 기본적으로 자계식(self-administered) 조사이기 때문에 응답자들로 하여금 긴 질문에 답하도록 설득하거나 혼란스러운 부분을 설명해 줄 면접자가 존재하지 않는다. 따라서 인터넷 조사에서는 응답자들이 흥미를 가지고 쉽게 편하게 응답할 수 있도록 구성하는 것이 무엇보다도 중요하다.

인터넷 조사에서는 질문지의 형식이나 제시되는 방법이 응답과정에 영향을 줄 수가 있다. 동일한 질문 내용이라고 하더라도 텍스트나 전자설문의 형태로 제시되는 경우와 웹문서로 제시되는 경우는 무응답 상황을 만들어 낼 수 있다.

3) Html Form-based 설문조사

인터넷 설문조사에 관련한 연구 중에서 특히 Html Survey Mode 비교 연구를 정리한 [표 2-1]를 보면, Web에서 조사가 전통조사 방법보다 데이터의 질(Data Quality)이 같거나 우수하며, 화려한 Html 설문보다는 단순한 것이 응답률이 좋다는 결론을 얻고 있다. [표 2-1]에서 HTML Survey Mode 비교에 관련한 선행 연구를 살펴보면 다음과 같다.

[표 2-1] HTML Survey Mode 비교

연구자	연구대상	연구 결과
Batagelj & Vehovar(1998)	Text(Single, Multiple) Graphic(Single, Multiple)	설문지 길이와 Layout에 따른 완성률(Completion Rate)에 차이없음
Dillman et al(1998b)	Plain Html Fancy Html	Plain Version이 Fancy Version 보다 높은 응답 비율(response rate)와 완성률을 나타냄
Comley(1998a)	E-mail Postal-website	E-mail이 Postal website를 이용한 조사보다 평균 응답시간이 짧고, Data Quality가 좋다.
Stanton(1998)	Paper Web Survey	Web Survey와 Paper Version은 항목간 분산과 요인구조에 있어 차이가 없음
김광용 외(1999)	Single Mode Multiple Mode	Single mode가 Multiple mode보다 완성률이 높음

Batagelj와 Vehovar(1998)는 설문지의 Layout비교를 완성률(Completion Rate)에 한정하여 비교하여 유의한 차이가 없다는 결론을 얻었다.

Dillman 외(1998)가 실시한 평이한(plain)설문지와 장식적인(fancy) 설문지를 비교하여 실험하였는데 장식적인 디자인은 959k, 단순한 디자인은 317k로 컴퓨터 메모리를 더 많이 차지했다. 따라서 모뎀을 이용한 경우 전송시간이 각각 345초, 120초로 평이한 설문지에 비해 장식적인 설문지가 682초가 더 걸린다고 하였다. 이때 응답률은 평이한 설문지가 93.1%이었고, 장식적인 설문지가 82.1%였다. 이는 응답자의 소프트웨어가 불일치하고, 진보된 프로그래밍 기술을 이용한 화려한 장식은 긴 전송시간으로 응답률을 저해하는 요인으로 작용할 수 있다고 지적하였다.

김광용 외(1999)는 웹 설문지의 Multiple Mode와 Single Mode의 설문 완성율에 대한 비교에서 각각 62.6%와 97.5%로 Single Mode에서 높은 응답률을 나타내었다. Multiple Mode의 경우 전송속도는 최적화 하였으나, 마지막 페이지까지 응답하는 경우가 적어 여러 장의 설문은 연결되는 구조는 전반적인 응답률을 저하시키는 결과를 가져온다고 하였다.

4. 단위 무응답

1) 전자우편조사에서 리스트 결함

사용자명(user id)이나 메일 서버가 잘못 입력된 것은 전송이 불가능하다. 제대로 입력되었다고 하더라도 해당 사용자가 더 이상 그 메일 서버를 사용하지 않거나, 조사 당시에 서버에 일시적인 장애가 발생할 경우에도 무응답이 발생할 수 있다. 이것은 무응답률을 낮추는 주된 원인으로 김광용 외 (1999)는 무응답률을 다음과 같이 계산하였다.

$$\text{무응답률} = \frac{\text{타당한 응답수} \times 100}{\text{발송 } E\text{-mail수} - (\text{주소불명} + \text{부적절응답})}$$

사용자나 서버에 불명확으로 되돌아오는 경우는 응답률 계산에서 제외할 수 있으며, 서버의 일시적인 장애인 경우에는 오류 메시지를 보내오기 때문에 재시도를 해 볼 수 있다.

2) 사적 정보의 기록

인터넷 조사를 하는 몇몇 기관은 수신 또는 발신 메시지 기록을 보관함으로써 조사의 주제가 민감할 경우 설문을 완성하는데 힘들어지며 무응답을 증가시킨다. 민감한 사안이나 응답자의 사적인 정보의 유출과 관련하여 쿠키(cookie)정보의 사용에 유의할

필요가 있다. 쿠키는 응답자의 확인, 구역간의 이동, 응답유형의 분석 등에 필요한 중요한 정보를 담고 있기 때문에 인터넷 조사에서 중요한 가치를 가지고 있다. 그러나 이러한 정보가 조사이외의 목적으로 전용되는 경우에는 응답자의 프라이버시를 침해할 경우 응답을 거부하는 경우가 증가할 수 있다. 따라서 쿠키 정보를 최소한으로 한정시켜야 한다.

3) 중복 응답

인터넷상에서는 사용자의 신분을 확인할 방법이 제한되어 있기 때문에 자격의 미비나 불일치, 또는 중복응답에 따른 자료의 타당성이 문제가 원인이 될 수 있다. 특히 중복응답자의 처리는 인터넷 조사에서 해결되어야 할 중요한 과제 중의 하나이다. 회원조사나 전자설문조사의 경우에는 조사대상자에 대한 DB와 연동되어 중복응답은 사전에 선별될 수 있으나 사전 선별이 불가능한 자발적 참여 조사의 경우에 중복응답은 중대한 문제가 될 수 있다. 전자우편조사의 경우 특히 인터넷상에서는 사용자들이 복수의 전자우편 계정을 가질 수 있으며 그에 대한 판정 여부가 어렵기 때문에 중복응답의 문제가 제기된다. 또 응답자가 설문을 전송한 후 다시 전송 버튼을 눌러 중복 전송되는 경우도 있다(Johnson 외 1999). 이때 중복 응답을 막기 위해 전송이 완료된 후 확인시켜 주는 메시지를 응답자에게 보여주어야 한다.

5. 무응답 오차의 척도

무응답 비율, 응답률, 응답밀도, 사후 추출률로 계산이 가능한데 이것은 조사의 질을 보여주는 중요한 척도로서 높을수록 조사의 질이 높아진다. 무응답 오차는 관심변수 상에서 무응답 비율과 응답자 사이의 차이에 대한 비율의 함수이다.

$$\text{무응답 비율} = \frac{\text{전체모집단} - \text{응답자}}{\text{전체모집단}} \quad (2-1)$$

그러나 무응답 비율에 정의는 정의할 수 없는 경우가 대부분이다. 특성화 오차에서도 언급했듯이, 전체 모집단이 알려지지 않은 경우가 대부분이기 때문에 무응답 오차에 대한 측정은 추출률과 확률추출을 기반으로 한 조사에만 계산이 가능하다. 따라서 온라인 조사에서는 무응답에 대한 정보가 적기 때문에 조사의 가치를 측정하기 위해서 무응답 비율보다는 응답률을 계산하는 경우가 많다.

전자우편조사의 경우에는 응답률의 계산은 다음과 같다.

인터넷 조사에서 무응답 오차에 관한 연구

$$\text{응답률} = \frac{\text{회신된 메일}}{\text{전체 발송 메일수}} \quad (2-2)$$

응답률의 계산에서는 일반적으로 응답자의 적합성이 중요한 기준이 되는데, 전자우편 주소 목록에서 적합한 표본을 판별해 낼 방법이 없기 때문에 이와 같은 계산 방법은 엄밀한 것이 아니며 실제보다 낮은 응답률이 나타날 수 있다. 전자우편을 통해서 응답자들의 참여를 유도하는 데에는 한계가 있기 때문에 전자우편 조사에서 응답률은 대체로 낮은 편이다.

실제로 편의추출을 하는 경우가 많기 때문에 응답률을 대신할 수 있는 사후 추출률(post sampling rate)을 사용하는 경우가 많다.

$$\text{사후추출율} = \frac{\text{일정한 기간동안 조사에 참여한 수}}{\text{목표 모집단}} \quad (2-3)$$

사후추출율은 일정 기간동안 조사에 참여한 모든 사례를 표본으로 삼는 것이 아니라 그 중의 일부(흔히 목표표본 수)를 사후적으로 추출하여 자료에 포함시키는 경우에 적용되는 개념이다.

이와 함께 응답밀도(response density)라는 개념도 유용할 수 있다.

$$\text{응답밀도} = \frac{\text{응답횟수}}{\text{일정 조사 기간}} \quad (2-4)$$

응답밀도는 주어진 기간 동안 응답자들의 참여가 얼마나 집중적으로 이루어지는가에 대한 지표로서 일간 응답밀도, 시간당 응답밀도를 계산해 볼 수 있다. 주중과 주말에 따라 일간 응답밀도의 차이가 나타날 수 있고, 시간당 응답밀도는 하루 중 주간과 야간의 응답자 수를 차이로 보여줄 수 있다.

Ⅲ. 인터넷 조사 사례분석

1. 조사목적

인터넷 조사는 개인조사나 여론조사의 경우 수집한 자료의 처리와 분석이 조사와 동시에 진행될 수 있으므로 조사기법으로서 충분한 활용가치가 있다(Swoboda 외 1997). 그러나 인터넷 조사는 조사의 신뢰성을 저해하는 여러 요인이 있다. 예를 들어,

확률표본 확보의 어려움, 목표모집단과 표본추출모집단의 차이에서 발생하는 포함 오차, 무응답, 설문지 설계 문제, 그 밖의 다양한 기술적인 문제 등이 있다. 따라서 이런 다양한 요인 중에서 어떻게 조사를 시행해야 오차를 최소화시키고 자료의 질을 높이는지에 대해 연구해야 한다.

인터넷 조사 중 대표적인 방법은 게시판에 조사참여를 공지해 자발적인 참여 (self-selection)방법과 전자우편조사를 실시하여 두 가지 방법에 따른 응답률을 비교해 보고 인터넷 조사에서의 표본추출의 필요성을 검증한다. 또한, 인터넷 설문지를 작성해 설문지 디자인에 따른 항목 무응답률을 살펴보고 응답률에 영향을 미치는 요인들을 사례 분석을 통해 알아보고자 한다.

2. 조사 방법

본 연구에서는 Windows NT 2000, IIS, Database server(ODBC)의 운영체제를 사용하였다. 웹설문을 통해 응답자의 응답은 서버와의 응답통로인 ASP 프로그래밍을 거쳐, 데이터 베이스 Microsoft Access에 저장되어 진다. 조사 결과는 데이터베이스에 저장된 데이터를 SPSS로 분석하였다.

조사 제목은 『성균관대학교 재학생의 학교 만족도와 인터넷 사용실태』이며, 조사 대상은 성균관대학교 재학생으로 2001년 10월 20일부터 2001년 11월 15일까지 27일간 조사를 실시하였다. 조사 형태는 자발적인 참여와 전자우편의 두 가지이며, 설문문항은 동일하게 29문항으로 학교 생활 만족도에 관한 문항 13문항, 인터넷 사용실태 11문항, 분류용 질문 5문항으로 크게 3가지 주제로 이루어졌다.

설문지와 별도로 안내문을 설문지 앞에 위치하게 해 조사의 목적이나 취지, 분석결과 제공 여부를 충분히 설명하였으며, 설문 전송 후에는 ‘설문에 응해주셔서 감사합니다’ 라는 메시지가 있는 페이지를 응답자에게 보여줌으로써 전송이 완료되었음을 확인시켜 설문을 여러 번 전송하는 것을 예방하였다.

자발적인 참여 조사에서는 학부와 학과의 홈페이지 게시판과 Daum 의 성균관대학교 재학생을 대상으로 하는 카페의 게시판에 설문참여 안내문을 공지하여 참여를 유도하였다. 전자우편 조사는 e-mail의 확보 문제로 학교에서 제공하는 메일 서비스 (mail. skku.ac.kr)를 사용하는 재학생을 표본틀(frame)로 국한시켰으며, e-mail을 확보 하였더라도 그 전자메일을 사용하지 않는 경우에 무응답이 발생할 수 있으므로 접속 시간을 고려해 장기간 접속이 없는 경우는 표본에서 제외시켰다.

3. 조사 결과

1) 표본 설계와 응답자 구성 비율

성균관대학교 재학생 총 16,371명 중 1,000명을 층화추출법으로 추출하였다. 학년별 구성은 1학년 25.6%, 2학년 24.7%, 3학년 24.2%, 4학년 25.5%로 [표 3-1]서 보면 전자우편조사의 경우 1학년의 응답률이 높은 것으로 나타났고, 자발적 참여조사에서는 응답자 구성이 모집단 구성비와 유사하게 나타났다.

[표 3-1] 학년별 응답자 구성비율 비교

학년	모집단	전자우편조사(%)	자발적 참여(%)
1학년	4199(25.6)	115(50.88)	18(25.71)
2학년	4039(24.7)	34(15.04)	12(17.14)
3학년	3964(24.2)	42(18.58)	17(24.29)
4학년	4169(25.5)	16(7.08)	16(22.86)
응답총합	16371	207(91.59)	63(90.00)
무응답		19(8.41)	7(10.00)
총합		226(100)	70(100)

[표 3-2] 학부별 조사 참여율

학부	총합(%)	전자우편 조사 표본빈도(%)	자발적 참여 표본빈도(%)
유학 동양 학부	278(1.61)	2(0.93)	2(3.03)
어문학부	1242(7.22)	14(6.51)	2(3.03)
인문학부	504(2.93)	7(3.26)	2(3.03)
계열학부	3091(17.96)	69(32.09)	8(12.12)
법과대학	1075(6.24)	19(8.84)	5(7.58)
사회과학부	863(5.01)	16(7.44)	3(4.55)
경제학부	928(5.39)	7(3.26)	8(12.12)
경영학부	1351(7.85)	9(4.19)	10(15.15)
생활과학부	367(2.13)	1(0.47)	0(0.00)
사범대학	529(3.07)	3(1.40)	1(1.52)
예술학부	546(3.17)	3(1.40)	1(1.52)
자연과학부	678(3.94)	10(4.65)	2(3.03)
전기전자및 컴퓨터공학부	1059(6.15)	22(10.23)	5(7.58)
응용화학부	683(3.97)	1(0.47)	2(3.03)
금속·재료공학부	461(2.68)	7(3.26)	3(4.55)
기계공학부	625(3.63)	8(3.72)	2(3.03)
건축 조경 및 토목공학부	647(3.76)	9(4.19)	4(6.06)
시스템경영공학부	285(1.66)	2(0.93)	4(6.06)
약학부	319(1.85)	3(1.40)	1(1.52)
의과대	194(1.13)	0(0.00)	0(0.00)
스포츠과학부	247(1.43)	3(1.40)	1(1.52)
총합	17214	215	66

[표 3-2] 학부별로 응답 참여율을 살펴보면 자발적인 참여에서는 의학부, 생활과학부의 응답이 전혀 나타나지 않았으며, 전자우편 조사에서도 의학부는 응답이 나타나지 않았으며, 생활과학부도 조사 참여율이 저조하였다.

[표 3-3] 성별 조사 참여율은 성대의 남녀 구성비는 10990명과 4882명으로 약 7:3인데 비해 응답률은 전자우편 조사, 자발적 참여 조사 모두 남성의 참여율 80%가 넘어 성별 응답률 차이를 나타내기도 하였다.

[표 3-3] 성별 조사 참여도

	남녀 구성비(%)	전자우편	자발적 참여
남 성	10990(68.81)	186(82.30)	56(80)
여 성	4982(31.19)	32(14.16)	10(14.29)
합 계	15972(100)	218(96.46)	66(94.29)
결 측		8(3.54)	4(5.71)
합 계		226(100)	70(100)

2) 응답률

일반적으로 전자 우편을 발송한 후 반송되어 되돌아오는 경우가 많다. 그 이유로는 'user known', 'messagebox full', 'send block'라는 메시지와 함께 되돌아오는 경우가 많은데 그 원인으로서는 당시 서버의 일시적인 장애 또는 사용자가 메일을 사용하지 않는 경우, 서버가 할당된 메일 저장 공간이 모두 소비한 경우이다. 이런 경우는 전자우편조사에서 단위 무응답 오차를 높이는 주요한 문제이다. 이런 경우를 고려해 학교 메일을 사용하는 사람들 중 접속시간을 고려, 최근 3개월 동안 로그인 정보가 있는 메일 주소만을 선택하여 프레임에 포함하였다.

표본으로 추출된 1,000명 중 전체 응답자는 226명으로 응답률 22.6%이었으며, 메일 서버의 문제로 82개의 메일이 돌아왔으며, 2개의 메일은 부재중이라는 응답이 되돌아왔다. 응답대상자 중에서는 본인이 대학원생이며 응답 대상자가 아님을 알려주는 응답도 있었다. 또한, 응답자중 전자우편조사의 노출에 불만을 가진 메일을 보낸 경우 정중한 사과와 함께 조사의 목적을 자세히 설명해 답장을 보내기도 하였다.

응답률 계산은 정확히 조사대상자에게 전자우편이 정확히 전달되었는지 확인하기 어려우므로 응답률 계산에서 리턴메일(return mail)도 전체 응답 대상자에 포함시켜

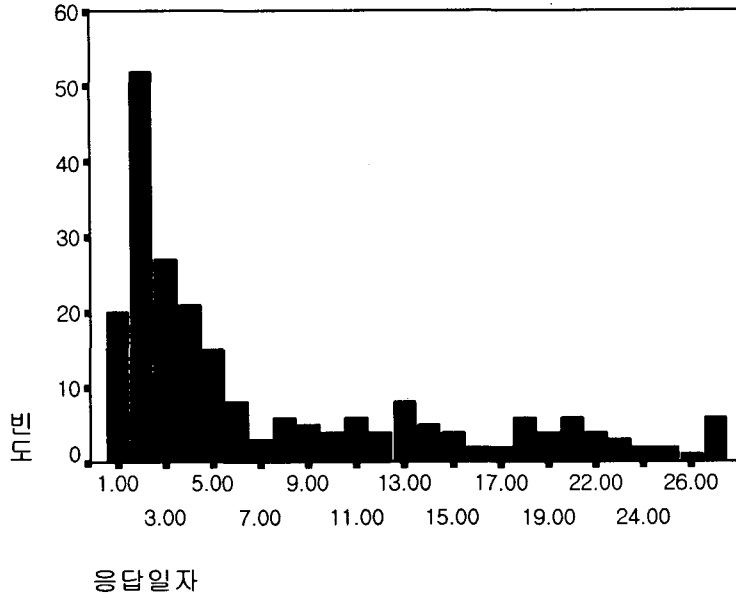
인터넷 조사에서 무응답 오차에 관한 연구

계산하였다. 따라서 본 사례분석에서 제시된 응답률은 실제 전자우편이 도착된 응답 대상자만을 고려한다면 응답률은 더 높아질 수 있다. [표 3-4]는 전자 우편 조사에서 응답 날짜별 응답 빈도와 응답률을 나타내며, 전자 메일 발송 후 4일 동안 전체 응답자중 53.1%를 차지하였으며, [그림 3-1]에서는 메일 발송 후 2일, 3일째에 응답이 가장 많았다. 자발적인 참여 조사인 경우 통계학과 홈페이지, 학부별 게시판에 설문에 응할 수 있도록 하였다. [표 3-5]는 자발적인 참여에 의한 조사로서 게시판에 설문 조사 공고를 낸 후 4일간 전체 응답의 78%가 응답을 하였으며, 게시판 공고 2일, 3일째에 가장 응답이 많았고, 조사 시작 일주일 후부터는 응답률이 급격히 떨어졌다.

또, 조사 결과 자발적 참여 조사에서 한 문항도 응답하지 않고 설문을 전송한 경우도 있었다. 이런 경우 항목 무응답이 높아질 수 있으므로 응답에서 제외하였다.

[표 3-4] (전자메일 조사)응답 일자별 응답 빈도와 응답률

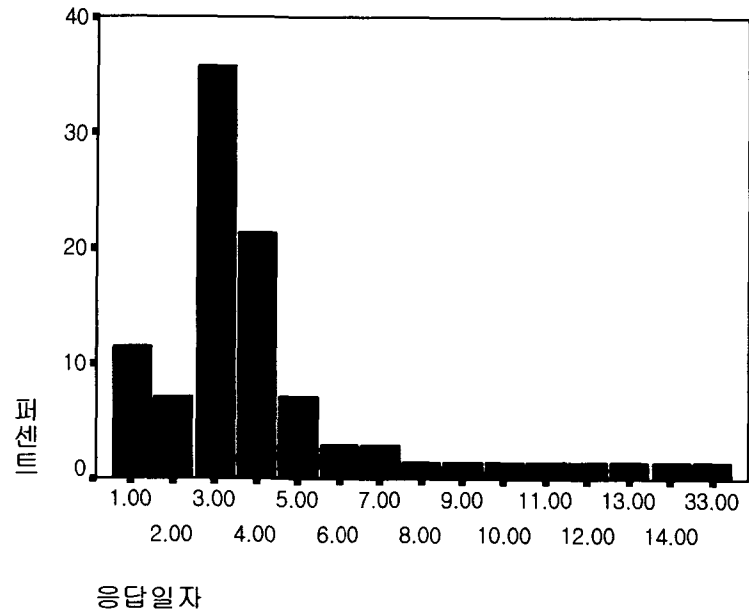
응답일자	빈도	응답률(%)	누적응답률	응답일자	빈도	응답률(%)	누적응답률
1	20	8.85	8.85	14	5	2.21	81.42
2	52	23.01	31.86	15	4	1.77	83.19
3	27	11.95	43.81	16	2	0.88	84.07
4	21	9.29	53.10	17	2	0.88	84.96
5	15	6.64	59.73	18	6	2.65	87.61
6	8	3.54	63.27	19	4	1.77	89.38
7	3	1.33	64.60	21	6	2.65	92.04
8	6	2.65	67.26	22	4	1.77	93.81
9	5	2.21	69.47	23	3	1.33	95.13
10	4	1.77	71.24	24	2	0.88	96.02
11	6	2.65	73.89	25	2	0.88	96.90
12	4	1.77	75.66	26	1	0.44	97.35
13	8	3.54	79.20	27	6	2.65	100.00
				합계	226	100	



[그림 3-1] (전자 메일 조사)일간 응답 밀도

[표 3-5] (자발적인 참여)응답 날짜별 응답 빈도와 응답률

응답 날짜	빈도	응답률(%)	누적응답률
1	8	10.96	10.96
2	5	6.85	17.81
3	28	38.36	56.16
4	16	21.92	78.08
5	5	6.85	84.93
6	2	2.74	87.67
7	2	2.74	90.41
8	1	1.37	91.78
9	1	1.37	93.15
10	1	1.37	94.52
11	1	1.37	95.89
12	1	1.37	97.26
13	1	1.37	98.63
14	1	1.37	100.00
합계	73	100	



[그림 3-2] (자발적 참여)일간 응답 밀도

3) 항목 무응답

본 조사에서는 전자우편 조사와 자발적인 참여의 항목 무응답률과 라디오 버튼, 엔트리 박스, 팝업의 3가지 설문디자인에 따른 응답 결과를 비교하였다.

설문 참여 안내문을 게시해 자발적으로 참여하여 설문을 응할 수 있도록 한 방법과 성균관대 재학생의 전자우편주소 목록에서 1,000명에게 전자우편 설문지를 전송해 조사를 실시하는 두 방법에 따라 항목응답률의 차이가 있는지 살펴보고자 한다.

[표 3-6] 전체적인 항목 무응답률을 살펴보면 자발적인 참여에 응하여 설문을 한 경우가 평균 항목 무응답률이 5.72%로 전자우편조사의 3.42%보다 높았다. 특히, 응답자의 의견, 또는 신상에 관한 문항에서 특히 무응답률 높은 것으로 나타났다. 설문 의 응답률을 높이기 의견을 직접 서술하는 텍스트 형식의 설문 디자인은 제외하였으며, 단 답형 형식이 무응답률이 높을 것이라고 예상했으나 두 조사 모두 높지 않았다.

디자인에 따른 무응답률은 [표 3-7]과 같은 결과를 보였다. 라디오 버튼 중에서도 응답하기 쉬운 서열 척도의 무응답률은 낮았으며 엔트리박스 디자인도 마우스가 아닌 키보드를 이용한 입력으로 낮은 응답률을 예상했으나 결과는 그렇지 않았고 반면 마우스 동작만으로 입력이 가능한 팝업 메뉴는 오히려 무응답률이 높았다.

표 3.6 항목무응답률 비교

			전자우편조사			자발 참여(self selection)		
			유효	결측	무응답률	유효	결측	무응답률
1	학교 생활 만족도에 관한 질문	라디오버튼	219	7	3.1	68	2	2.94
2			220	6	2.7	68	2	2.94
3			221	5	2.2	69	1	1.45
4_1		라디오버튼 (서열척도)	224	2	0.9	69	1	1.45
4_2			223	3	1.3	69	1	1.45
4_3			222	4	1.8	69	1	1.45
4_4			223	3	1.3	69	1	1.45
4_5			218	8	3.5	68	2	2.94
5		라디오	217	9	4.0	69	1	1.45
7		라디오버튼 (서열척도)	223	3	1.3	67	3	4.48
8			221	5	2.2	67	3	4.48
9		라디오버튼	222	4	1.8	67	3	4.48
10		라디오버튼 (서열척도)	223	3	1.3	67	3	4.48
11		라디오버튼	217	9	4.0	66	4	6.06
12_1		엔트리	221	5	2.2	66	4	6.06
12_2			220	6	2.7	66	4	6.06
12_3			219	7	3.1	66	4	6.06
13		라디오버튼	218	8	3.5	66	4	6.06
1		인터넷 사용 실태	라디오버튼 (서열척도)	222	4	1.8	66	4
2	라디오버튼		215	11	4.9	66	4	6.06
3			214	12	5.3	64	6	9.38
4			220	6	2.7	64	6	9.38
5	라디오버튼 (서열척도)		221	5	2.2	66	4	6.06
6			213	13	5.8	66	4	6.06
7	라디오		218	8	3.5	65	5	7.69
8			218	8	3.5	66	4	6.06
9			213	13	5.8	62	8	12.90
11			219	7	3.1	64	6	9.38
1	분류용 질문		팝업	218	8	3.5	66	4
2		215		11	4.9	66	4	6.06
3_1		207		19	8.4	63	7	11.11
3_2		214	12	5.3	64	6	9.38	
4		text	217	9	4.0	65	5	7.69
5		라디오	220	6	2.7	64	6	9.38
평균 무응답			3.24			5.72		
전체 응답자			226			70		

인터넷 조사에서 무응답 오차에 관한 연구

[표 3-7] 설문지 디자인에 따른 무응답률

	라디오(서열)	라디오	팝업	엔트리
전자우편조사	2.132%	3.512%	6.195%	2.655%
자발적참여조사	3.961%	6.119%	8.849%	6.061%

4) 문항별 결과

본 조사는 학교 생활 만족도에 대한 만족도를 묻는 15문항으로 전반적인 학교 생활에 관한 문항과 관련된 문항으로 구성되어 있다. 조사 방법별 차이가 나타난 문항도 있었지만 각 문항의 응답차이가 있는지 알아보는 [표 3-8]와 [표 3-9]을 보면 조사방법에 따른 특별한 응답차이를 발견하기 어려웠다.

[표 3-8] 설문 문항에 따른 조사 방법의 차이 비교(t-test)

		n	평균	t	자유도	유의확률
전반적인 학교 생활	전자우편조사	224	2.72	3.98	291.00	0.00
	자발적 참여	69	2.23			
학과내 활동 상황	전자우편조사	223	3.29	1.20	290.00	0.23
	자발적 참여	69	3.12			
전공공부	전자우편조사	222	3.05	3.57	289.00	0.00
	자발적 참여	69	2.51			
학과 동료 선후배 사이	전자우편조사	223	3.00	1.03	290.00	0.30
	자발적 참여	69	2.84			
동아리 활동	전자우편조사	218	3.17	-0.28	284.00	0.78
	자발적 참여	68	3.22			
성대에 다니는 것에 대한 만족도	전자우편조사	223	2.27	3.51	288.00	0.00
	자발적 참여	67	1.81			
소속학부에 대한 만족도	전자우편조사	221	2.51	1.87	286.00	0.06
	자발적 참여	67	2.21			
학부 행사에 참여 정도	전자우편조사	223	3.51	0.02	288.00	0.98
	자발적 참여	67	3.51			

[표 3-9] 설문 문항에 따른 조사 방법의 차이 비교

	카이값	자유도	유의 확률
전공외 가장 중점을 두는 부분	10.54	6	0.10
강의 없는 시간의 있는 곳	3.44	7	0.84
한달 용돈	7.89	7	0.34

IV. 결 론

설문조사를 위해 Window NT운영 체제에서 적합한 ASP 프로그래밍을 이용한 설문 조사를 실시하였으며 결과는 다음과 같다.

인터넷 설문 조사의 결과 장점으로 알려진 것들을 확인할 수 있었다. 전혀 설문 비용이 들지 않았으며, 빠른 데이터 수집, 코딩과정의 생략과 더불어 코딩 입력과정의 오류가 발생하지 않았다.

전자우편 조사의 경우 응답률 22.6%, 항목 무응답률 4.34%로 나타나 일반적인 우편 조사와 큰 차이가 나타나지 않았다. 조사방법별로 표본 구성비율을 모집단과 비교해 볼 때 다소 차이가 있었지만 모집단 구성비와 큰 차이는 없었으며, 문항별 비교에서도 통계적으로 유의한 차이가 없었다.

자발적 참여 조사에서는 응답자와 의사 교환이 된다는 점과 익명성이 보장된다는 점에서 솔직한 응답을 얻을 수 있다는 장점이 있었다. 그러나 저조한 응답률과 함께, 자료의 질을 판단할 수 있는 항목 무응답의 문제점이 나타났다.

전자우편조사의 경우 자발적인 참여 조사보다 높은 응답률과 낮은 항목 무응답률이 나타났으나, 단위 무응답에 따른 표본의 대표성 문제가 드러났다.

인터넷 설문조사가 전통적인 방법을 대체하기 위해서는 과학적인 확률추출을 근거로 조사가 실시되어야 한다. 따라서 표본추출틀 획득이 가능한 특정 집단을 대상으로 설문이 이루어 져야하며, 표본으로 선정된 응답단위들의 성실한 응답을 유도할 수 있어야 한다. 또, 높은 응답률을 갖는 설문 디자인, 충분한 사전준비를 하고, 시스템 안정성을 갖춘다면 전통적인 조사방법을 대체할 수 있을 것이라고 생각된다.

<참 고 문 헌>

- 김광용, 김기수,(1999), 다양한 인터넷 설문방식 비교와 설문만족에 관한 실증연구, 춘계 경영정보학회 학술대회
- 김영원, 변종석,(2000), 인터넷 조사에서 표본추출 동향 및 문제점, Internet Survey 워크샵 논문집
- 김준한,(2000) 동일 주제에 대한 인터넷 조사와 기존 조사 결과 비교, Internet Survey 워크샵 논문집(2000)

인터넷 조사에서 무응답 오차에 관한 연구

- 김종호, 남궁 평, 박홍래, 이계오, 이상우, 함종욱, (1990), 통계조사에 있어서 효율적 표본설계와 오차에 관한 연구, 한국통계학회
- 변종석, 남궁 평, (1995), 표본조사론 관점에서의 인터넷조사 방법에 대한 비교 연구, 한신정보과학연구소 논문집 제 1호
- 정재구, (1997), 통계조사에서 비표본오차의 최소화방안 연구, 한국전통상학회제 10집 한국통계학회 조사통계 연구회, (2000), 무응답 오차, 자유아카데미
- 한국 인터넷 정보센터. <http://stat.nic.or.kr/>
- Don A. Dillman, Robert D. Tortora, and Dennis Bowker, (1999), Principle for Constructing Web Surveys, *SESRC technical Report 98-50, Pullman*
- Don A. Dillman, Tortora and Bowker, (1998), Influence of Plain vs. Fancy design on response rates for web surveys, *Proceeding Of SMS, 1998 annual meeting of ASA*
- Don A. Dillman, (1999), Mail and Other Self-Administered Surveys in the 21st Century, *The Beginning of a New Era. Revised version available in Gallup Research Journal*
- Judith T. Lessler and Willam D. Kalsbeek, (1992), "Nonsampling error in Survey", John Wiley & Sons, Inc..
- M. P. Couper, (2000), Web Survey : the Good, the Bad, and the Ugly, *Survey Research Center, University of Michigan.*
- Mehta and Sivadas, (1995), Comparing response rates and response content in mail vs. electronic mail survey, *Journal of the market Research Sociery*
- Nichols and Sedivi, (1998), Economic data collection via the web, A census bureau case study, *Proceedings of the American Statistical Association, Section on Survey Research Methods*
- Swoboda et al., (1997), Internet surveys by direct mailing, *Social Science Computer Review*
- Zenel Batagelj, Katja Lozar, Vasja Vehovar, (1998), Who are nonrespondents in Web Surveys?, *the 9th international Workshop on household survey nonresponse.*