

## Implementation of Forced Answer Model for Sensitive Information at On-Line Survey

Hee-Chang Park<sup>1)</sup> · Jee-Hyun Ryu<sup>2)</sup> · Sung-Yong Lee<sup>3)</sup>

### Abstract

In this paper we implement the system for the forced answer model which is an indirect response technique on the internet as a way of obtaining much more precise information, not revealing secrets of responders. In this system we consider that respondents are generally reluctant to answer in a survey to get sensitive information targeting employees, customers, etc.

**Keywords** : 확률화응답기법, 강요질문기법, E-R diagram

### 1. 서론

온라인 설문조사는 대표성의 결여 등과 같은 단점이 있음에도 불구하고 오프 라인(off-line)의 조사에서 가지지 못한 많은 장점(비용의 절감, 신속한 응답 및 집계)이 있기에 급속도로 발전하고 있으며, 모집단이 제한적인 회원이나 기업체의 구성원 등을 대상으로 한 조사에서는 온라인 설문조사의 단점(대표성의 결여)을 보완할 수 있기에 현실적으로 많이 사용하고 있다.

최근 들어 기업 등에서 자사의 고객 특히 온라인 회원을 대상으로 혹은 자회사의 사원을 대상으로 기업의 경영 자문을 얻기 위해 온라인 설문조사가 현장에서 많이 행해지고 있으나, 온라인 설문조사 역시 기존의 설문조사와 마찬가지로 조사자가 응답자들의 프라이버시나 사생활과 관련된 민감한 정보를 얻고자 할 경우에 정확한 정보를 얻기란 쉬운 일이 아니다.

한편, 응답자들은 기존의 설문조사와 마찬가지로 온라인 상에서 민감한 질문을 직

---

1) Professor, Department of Statistics, Changwon National University, Changwon, Kyungnam, 641-773, Korea  
E-mail : hcpark@sarim.changwon.ac.kr

2) Graduate Student, Department of Statistics, Changwon National University, Changwon, Kyungnam, 641-773, Korea

3) Associate Professor, Department of Industrial & Systems Engineering, Changwon National University, Changwon, Kyungnam, 641-773, Korea

접적으로 받게 되면 혹시 자신의 비밀이나 사생활이 노출될까 의심하여 혹은 응답 후의 부당한 대우를 받게될까 두려워 정직한 응답을 꺼리게 된다. 따라서, 기업 의사결정에 대한 찬·반 여부, 성적 질문 등과 같은 민감한 정보를 얻고자 할 경우 간접응답기법인 확률화응답기법을 온라인 설문조사에 적용해 볼 수 있다. 그러면, 응답자들의 응답이 확률장치를 이용하여 간접적으로 이루어지게 되므로 응답자가 자신의 신상에 대한 불안이나 개인 정보의 유출을 이유로 정확하지 않은 응답을 할 가능성을 줄일 수 있게 된다. 그러므로 온라인 설문조사를 하는 데 민감한 정보를 얻기 위한 확률화응답기법을 이용할 수 있는 시스템이 필요하다고 생각된다.

이러한 확률화응답기법(randomized response technique ; RRT)은 Warner(1965)에 의해 처음으로 제안되었다. Warner는 응답자들에게 민감한 질문과 민감한 질문에 배반되는 질문으로 구성된 확률장치를 사용하여 민감한 속성에 대한 질적 정보를 얻고자 하였다. Greenberg 등(1969)은 민감한 질문과 배반되는 질문 대신에 민감한 질문과 전혀 무관한 질문을 사용하는 무관질문기법(unrelated question technique)을 제안하였으며, Greenberg 등(1971)은 이를 민감한 변수에 대한 양적 정보를 얻기 위해 양적속성기법으로 발전시켰다. Loynes(1976)는 Warner기법의 민감한 질문과 배반이 되는 질문 대신에 “예”라고 응답하도록 강요하는 강요질문기법(forced answer technique)을 제안하였다. 또한 Fox와 Tracy(1986), Chaudhuri와 Mukerjee(1988)는 확률화응답기법을 정리, 요약하여 체계화하였으며, 최근에는 이들 이론들의 실제적 활용에 많은 관심이 집중되고 있다. 특히 사회학, 경영학, 의학 등 여러 학문분야에서의 조사활동에도 이의 활용이 적극 모색되고 있다.

국내에서는 이기성(1992)이 2단계 확률화응답모형에 관한 연구를 수행한 바 있으며, 류제복 등(1993)이 확률화응답기법에 관한 책을 출간하여 그 중요성을 강조하였다. 류제복 등(1995)은 확률화응답기법이 적용된 사례들을 비교 분석하여 실용화를 위한 방안을 제시하였다. 또한 박희창 등(2001)은 질적 자료의 관련질문기법에 대한 온라인 설문조사시스템을 구현하였으며, Park과 Myung(2002)은 질적 자료의 무관질문기법에 대한 인터넷시스템을 개발하였다.

본 논문에서는 민감한 정보를 얻기 위한 조사에서 응답자들이 정직하게 응답하기를 꺼리는 질문들에 대하여 응답자의 비밀을 노출시키지 않고서 보다 정확한 정보를 얻을 수 있는 확률화응답기법 중에서 강요질문기법을 인터넷상에서 사용할 수 있는 시스템을 구현하고자 한다. 2절에서 강요질문기법에 대해 전반적으로 살펴본 후, 3절에서는 시스템의 개발환경과 구성에 대하여 기술하고, 4절에서는 예제를 통하여 구현된 시스템에 관해 토의하며, 5절에서 결론을 맺는다.

## 2. 강요질문기법

본 절에서는 질적 강요질문기법에 관하여 기술하고자 한다. Loynes(1976)가 제안한 강요질문기법은 응답자들에게 민감한 질문과 무조건 “예”라고 응답하는 질문 중에서 확률장치를 통해 선택된 질문에 응답하게 함으로써 응답자의 신분이나 비밀을 노출시키지 않고서 민감한 질문에 대한 정보를 이끌어 낼 수 있는 확률화응답기법이다.

응답자들은 확률장치에 의해 선택된 질문에 응답하면 되며, 이 때 관리자는 응답자가 어떠한 질문에 응답을 했는지를 알 수 없게 되므로 응답자는 솔직하게 응답할 수

있다. Loynes가 사용한 질적 강요질문기법의 확률장치는 다음과 같은 2개의 설문으로 구성되어 있다.

- 설문 1 : 당신은 민감한 속성 A를 가지고 있습니까?  
설문 2 : “예”라고 응답하십시오.

이 때, 단순임의복원으로 추출된  $n$ 명의 응답자들은 확률장치에 의해서 선택된 질문에 “예” 또는 “아니오”로만 응답한다. 설문 1이 선택될 확률을  $p$ , 설문 2가 선택될 확률을  $1-p$  라 하면 응답자들이 “예”라고 응답할 확률은 다음과 같다.

$$\lambda = p\pi + (1-p) \quad (2.1)$$

여기서,  $\pi$ 는 민감한 속성 A에 대한 모비율이다.

$n$ 명의 응답자들 중에서 “예”라고 응답한 사람의 수를  $n_1$ 이라 하면  $\hat{\lambda} = \frac{n_1}{n}$ 이다. 따라서, 민감한 속성 A에 대한 모비율  $\pi$ 의 추정량과  $\hat{\pi}$ 의 분산추정량은 다음과 같다.

$$\hat{\pi} = \frac{\hat{\lambda} - (1-p)}{p} \quad (2.2)$$

$$\widehat{Var}(\hat{\pi}) = \frac{\hat{\lambda}(1-\hat{\lambda})}{(n-1)p^2} \quad (2.3)$$

### 3. 확률화응답기법의 구현

#### 3.1 시스템 개발 환경 및 시스템 흐름

구현된 시스템의 개발 환경에서 개발 언어는 gnu c compiler, Java, html 등이며, 운영 체제는 Linux, 그리고 데이터베이스는 MySQL-Ver 3.23.39를 사용하였다.

본 연구에서 개발된 시스템은 데이터베이스에 바탕을 두어 기존의 온라인 설문조사 시스템과 더불어 사용할 수 있을 뿐만 아니라 독립된 스팟 서베이(spot survey)가 가능하도록 개발하였다. 그리고 본 시스템은 동일한 응답자가 여러 번 답하는 것을 막기 위해 로그인(log in)을 하는 사이트에서는 동일 아이디에 대하여 중복 응답을 하는 것을 막을 수 있고, 로그인을 하지 않는 사이트에서는 동일 IP에서 중복 응답하는 것을 막을 수 있도록 구현하고자 한다. 또한 방화벽이 구축된 사무실 및 공공 기관에서 IP가 동일시 될 경우 기계주소(Mac Address)로 구분을 하며, 학교 실습실 등과 같이 여러 명이 사용하는 경우에 대비하여 응답에 제한을 두지 않을 수 있도록 구현하고자 한다. 또한 본 시스템은 자료의 입력에서 처리, 결과를 모두 데이터베이스를 바탕으로 이루어져 있다. 이로 인하여 동일 응답자 등의 반복 측정에서도 기존의 설문응답시스

템과 쉽게 합쳐서 사용할 수 있다. 처리과정에서도 데이터베이스를 사용함으로써 인하여 쿼리(query)를 사용하여 수행 속도면에서 파일시스템보다 빠르게 진행할 수 있다. 기본적인 결과를 데이터베이스에 저장함으로써 지속적인 조사에서 추세 분석이 가능하도록 개발하였다.

본 시스템은 관리자(조사자) 모드와 응답자 모드 두 가지 부분으로 구성되어 있다. 관리자 모드에서는 설문을 작성하는 에디터와 확률장치의 선택 및 민감한 설문이 선택될 확률을 입력하는 부분으로 이루어져 있고, 기타 계산에 필요한 정보를 입력하도록 되어 있다. 본 시스템의 구현을 위해 설계한 테이블은 <표 3.1>부터 <표 3.4>와 같다.

<표 3.1> 메인 테이블

메인			
Logical Name	Physical Name	Data Type	비고
인덱스 키	idx	integer	pk, auto_increment
설문 작성일	day	date	
설문 작성시간	time	time	
설문의 주제	subject	varchar(79)	
설문 문항수	number	tinyint	
설문 작성 종료 검사	check	char(2)	
설문의 소개 및 인사말	title	blob	

<표 3.1>의 메인 테이블은 고유한 테이블이며, 여기에 모든 설문 문항들에 대한 정보를 보관하고 있다. 또한 이 테이블은 <표 3.2>의 설문 테이블과 연계하여 분석을 수행하는데, 여기에는 설문 작성일시, 주제, 문항수, 소개 등이 기록되며 고유한 인덱스 키로 설문지 테이블을 참조한다.

<표 3.2> 설문 테이블

설문			
Logical Name	Physical Name	Data Type	비고
인덱스 키	idx	integer	pk, auto_increment
설문 1	q1	varchar(79)	
설문 2	q2	varchar(79)	
확률장치종류	type	varchar(2)	
표본 1에서 설문1이 선택될 확률	p1	float	
참여자	n1	smallint	

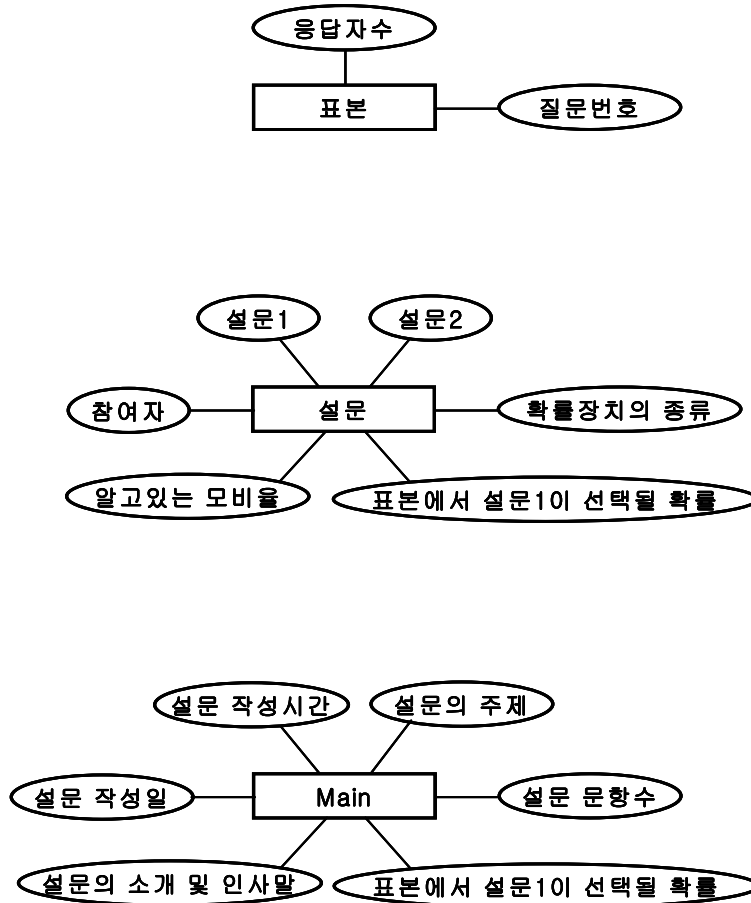
설문 테이블은 실제 설문 문항의 정보를 보유하는 테이블이다. 설문 문항의 내용, 확률장치, 설문 1이 선택될 확률 등 상위 테이블인 메인에서 인덱스 키를 참조하여 설문 생성시 추가로 생성된다.

<표 3.3> 표본 테이블

표 본			
Logical Name	Physical Name	Data Type	비 고
인덱스 키	idx	integer	pk, auto_increment
“예”라고 응답한 응답자 수	s1	mediumint	

<표 3.3>의 표본 테이블은 응답자 집계용 테이블이다. 설문지 테이블에서 넘어온 문항을 응답한 수치가 기록되는 테이블이다.

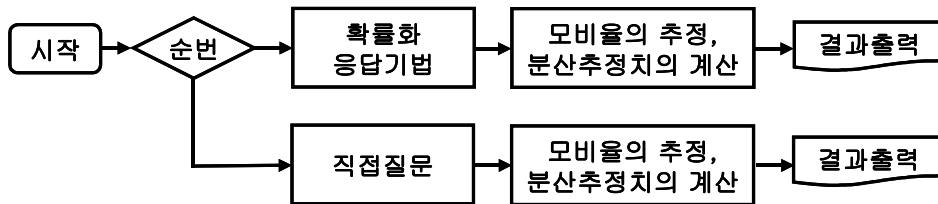
본 시스템에 사용된 데이터베이스 구성은 <그림 3.1>과 같은 구조를 가지고 있다. 메인 테이블은 설문이 선택될 확률 및 주제, 작성일자 등의 자료를 가지고 있으며, 표본 테이블은 설문 테이블에서 응답자들이 응답한 결과를 집계하여 결과를 보여주는 구조를 하고 있다.



<그림 3.1> E-R Diagram

### 3.2 시스템의 구성

본 시스템의 구성은 응답자 모드와 관리자 모드 2개로 구성되며, 시스템의 흐름은 <그림 3.2>와 같다.



<그림 3.2> 시스템 흐름도

관리자 모드는 설문 생성 및 삭제 담당하는 부분이며, 응답자 모드는 실제 응답자들이 질문에 응답을 하는 부분이다. 설문을 생성하기 위해서는 <그림 3.3> 및 <그림 3.4>와 같이 관리자 화면에 접속하여 설문 소개, 설문 주제, 문항의 수 등을 입력한다.



<그림 3.3> 관리자 초기화면



<그림 3.4> 질적 강요질문기법

본 시스템은 4종류의 확률장치 중 하나를 선택하도록 되어 있으며, 확률장치에서 설문 1이 선택될 확률을 입력하면 그 확률에 따라 동전의 앞면 혹은 뒷면, 연필의 오른쪽 혹은 왼쪽과 같이 두 사건 중 하나의 형태로 나타난다.

응답자 화면은 <그림 3.5>와 같으며, 이 화면에서 응답자는 설문 문항에 대하여 응답할 수 있다.



<그림 3.5> 응답자 화면

관리자는 응답자가 설문 1에 대하여 “예” 라고 응답을 하였는지 설문 2에 대하여 “예”라고 응답을 하였는지 알 수 없도록 설계되어 응답자가 진솔하게 응답할 수 있도록 구현하였다.

#### 4. 예제

이 절에서는 강요질문기법 시스템을 이용하여 실시한 인터넷조사로부터 얻어진 결과에 대해 논의하고자 한다. 본 조사는 대상 (주)Esab이라는 회사의 일반 사무직 사원 150명을 대상으로 2002년 1월 4일부터 1월 31일까지 인터넷상에서 실시하였으며, 문항은 다음과 같다.

“귀하의 상사로부터 인사고과에서 불이익을 당한 적이 있습니까?”

응답결과는 응답자용과 관리자용으로 구분되어 있으며, 응답자용은 민감한 속성  $A$ 의 모비율의 추정값만을 <그림 4.1>과 같이 나타낸다.



<그림 4.1> 응답자용 결과

<그림 4.1>은 “귀하의 상사로부터 인사고과에서 불이익을 당한 적이 있습니까”라는 민감한 질문에 대해 29명이 참여하였고 모비율  $\pi$ 의 추정량  $\hat{\pi}$ 은 0.31039임을 보여주고 있다. 관리자는 <그림 4.2>과 같이 확률장치에서 민감한 설문 1이 선택될 확률  $p$ , 표본의 크기(총참여자 수), “예”라고 응답한 응답자수, 모비율  $\pi$ 의 추정량  $\hat{\pi}$ 와  $\hat{\pi}$ 의 분산 추정치 등을 볼 수 있도록 구현하였다.





<그림 4.2> 관리자용 결과 화면

“귀하의 상사로부터 인사고과에서 불이익을 당한 적이 있습니까?”에 대한 직접질문에 대한 결과는 <그림 4.3>과 같다.



<그림 4.3> 직접 질문기법

이 화면에서 나타난 바와 같이 강요질문기법을 이용한 인터넷 설문조사에 참여한 사람의 수는 29명이며, 이들 중 “예”라고 응답한 사람의 수는 15명이었다. 이로부터 인사고과에서 불이익을 당한 적이 있는 사람들의 모비율을 추정해 본 결과 31.03%로 나타났으며, 분산추정치는 0.0182이었다. 그리고 직접질문을 이용한 인터넷 설문조사에 참여한 사람의 수는 59명이며, 이 중 “예”라고 응답한 사람의 수는 10명이었다. 이로부터 인사고과에서 불이익을 당한 적이 있는 사람들의 모비율을 추정해 본 결과

16.95%로 나타났으며, 분산추정치는 0.0024이었다. 이와 같은 결과에서 알 수 있듯이 민감한 질문에 대해 강요질문기법을 이용한 인터넷 설문조사에서 구한 모비율의 추정치가 직접질문을 이용해서 구한 모비율의 추정치보다 높게 나타났다.

## 5. 결론 및 향후과제

본 논문에서는 민감한 정보를 얻기 위한 조사에서 응답자들이 정직하게 응답하기를 꺼리는 질문들에 대하여 응답자의 비밀을 노출시키지 않고서 보다 정확한 정보를 얻을 수 있는 간접응답기법 중의 하나인 강요질문기법을 인터넷상에서 사용할 수 있도록 구현하였다. 본 시스템은 기존의 설문조사 시스템과 연계하여 민감한 질문에만 확률장치를 이용할 수 있도록 하여 다른 속성에 따라 민감한 질문에 대한 차이도 볼 수 있을 뿐만 아니라 독립된 단일문항 질문으로도 사용이 가능하도록 하였다. 본 연구에서 구현된 시스템을 공공기관이나 기업내의 인트라넷시스템에서 사용한다면 민감한 질문에 대해 응답자들의 좀 더 진실한 응답을 기대할 수 있을 것으로 생각된다.

## 참고문헌

1. 김정기, 김희재, 남기성, 박희창, 이성철, 정정현 (1999). 사회조사분석론, 창원대학교출판부.
2. 류제복, 홍기학, 이기성 (1993). 확률화응답모형, 자유아카데미, 서울.
3. 류제복, 이계오, 이기성 (1994). 확률화응답기법의 실용화 방안, 응용통계연구 8(1), 9-26.
4. 박희창, 이기성, 김희재, 남기성 (2001). 인터넷조사와 설문조사시스템, 자유아카데미, 서울.
5. 박희창, 남기성, 이기성 (2001). 인터넷조사에서의 확률화응답기술의 구현, 한국통계학회논문집, 8(3), 731-737.
6. Coomber, R. (1997). Using the Internet for Survey Research, *Sociological Research Online*, 2(2).  
<<http://www.socresonline.org.uk/socresonline/2/2/2.html>>
7. Fox, J. A. and Tracy, P. E. (1986). *Randomized Response : A Method for Sensitive Survey*, Sage Publications.
8. Loynes, R. M. (1976). Asymptotically Optimal Randomized Response Procedures, *Journal of the American Statistical Association*, 71, 924-928.
9. Park, H. C. and Myung, H. M. (2002). Implementation of Qualitative Unrelated Question Model for Obtaining Sensitive Information, *Journal of the Korean Data & Information Science Society*, 13(2), 341-354
10. Schwarz, C. J. (1997). "StatVillage : An On-line, WWW-Accessible, Hypothetical City Based on Real Data for Use in an Introductory Class in Survey Sampling. *Journal of Statistics Education* 5(2).  
<<http://www.amstat.org/publications/jse>>

11. Warner, S. L. (1965). "Randomized Response ; A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association*, 60, 63-69.

[ 2003년 7월 접수, 2003년 8월 채택 ]