

인터넷 설문조사의 검증에 관한 사례연구

류귀열¹ · 문영수²

¹서경대학교 컴퓨터학과 · ²한국과학기술정보연구원

접수 2013년 8월 20일, 수정 2013년 9월 9일, 게재확정 2013년 9월 23일

요약

연구의 목적은 인터넷 설문조사에서 자료의 정확성을 데이터베이스의 자료와 설문조사 자료의 비교를 통해 검증하는 것이다. 설문조사는 2012년 8월 한국과학기술정보연구원의 NDSL 원문이용자를 대상으로 실시하였으며, 검증에 이용된 변수는 나이, 소속기관의 인구통계적 변수와 이용빈도, 이용기간의 행동 변수이다 1차 검증결과 불일치 비율은 나이가 7.5%, 소속기관 15%, 이용빈도 92%, 이용기간 55%로 높게 나타났다. 상세검정을 통해 나이에 대한 불일치 비율을 비관적으로 3%, 낙관적으로 1%로, 소속기관에 대한 불일치 비율을 비관적으로 4.5%, 낙관적으로 2%로 추정할 수 있었다. 불일치비율을 1/4 이하로 대폭 축소시켰다. 이용빈도와 이용기간은 측정오차와 응답자의 기억력 문제, 인터넷 이용행태 등으로 불일치 비율이 높게 나타났다. 이러한 현상은 전통적 설문조사에도 간접적으로 적용될 수 있을 것이다. 본 연구의 결과는 인터넷 설문조사가 믿을만한 자료라는 점을 시사하고 있으며, 보다 정확한 검증을 위해 많은 후속 연구가 필요하다.

주요용어: 불일치, 온라인 설문조사, 인터넷 설문조사, 측정오차.

1. 서론

우리나라에서 유선인터넷은 1982년 서울대와 한국전자기술연구소와의 시범망을 구축함으로써 시작되었다. 1994년에는 상용 인터넷 서비스 제공자 (internet service provider)인 한국통신, 데이콤, 아이넷 등이 텍스트 기반의 인터넷서비스를 제공하기 시작하였다. Korea Internet & Security Agency (2013)에 의하면 우리나라는 2001년 다운로드 속도가 초당 1 Mbps (Mbyte per second)를 넘는 초고속 망 네트워크에서 세계 1위를 차지하였고, 인터넷 이용자도 2,000만 명을 돌파하여 기술적 측면이나 이용자 측면에서 비약적으로 발전하였다. 지속적인 성장을 통해 2011년 인터넷 이용자는 3,718만여 명으로 이용률은 78%에 이르고 있는 것으로 나타났다. Korea Communications Commission (2013)에 의하면 2012년 12월말 이동전화 가입자 5,362만여 명으로 나타났는데, 두 개 이상의 이동전화를 가입하는 사람들이 증가함으로써 보급률이 100%를 상회하였다. 이 중 무선인터넷을 원활하게 사용할 수 있는 스마트폰 가입자는 3,272만여 명에 이르는 것으로 나타났다. 따라서 우리나라에서는 유선인터넷이나 무선인터넷으로 설문조사를 쉽게 할 수 있는 환경이 구축되어 있다.

온라인 설문조사는 이메일을 이용한 설문조사와 인터넷을 이용한 조사로 나눌 수 있다. Ryu (2001), Park 등 (2005) 등에 설명되어 있는 바와 같이, 2000년부터 인터넷 인프라의 비약적인 발전과 더불어 인터넷 설문조사도 활발하게 도입되기 시작하였다. 인터넷 설문조사는 웹브라우저로 설문조사하는 방

¹ (136-704) 서울시 성북구 정릉동 산16-1, 서경대학교 컴퓨터학과, 부교수.

² 교신저자: (130-741) 서울시 동대문구 회기로 66, 한국과학기술정보연구원, 책임연구원.
E-mail: youngsum@kisti.re.kr

식으로 전통적인 설문조사와 상당히 유사하나 다양한 디자인의 설문지를 구축할 수 있으며, 즉각적인 응답을 받아 데이터베이스에 실시간으로 저장하고 분석할 수 있기 때문에 시간과 비용을 획기적으로 줄일 수 있다는 장점이 있다. 그리고 멀티미디어 등 부가적인 기능을 도입해 응답자의 이해를 높일 수 있다. 그러나 Couper (2000), Gim과 Kim (2004) 등이 설명하듯이 인터넷 설문에는 대표성 오류, 표본 추출 오류, 무응답 오류, 측정 오류 등이 있다. 온라인 설문조사에서 발생하는 무응답 오류나 측정 오류는 전통적 설문조사에서도 발생하는 오류로서 정도의 차이가 존재한다. 표본추출 오류는 표본을 추출함으로써 발생하는 오류로서 전통적 설문조사에서도 존재하는 오류이다. 대표성 오류는 Gim과 Kim (2004)이 말 한 바와 같이 전통적 설문조사 방법과 비교하여 가장 비판 받는 것으로 연구하고자 하는 목표 모집단과 표본추출에서 정의되는 틀 (frame)과의 불일치로 나타나는 오류이다. 이는 인터넷 사용자들이 계속 증가하고 인터넷 사용 환경이 끊임없이 변하여 목표 모집단이 끊임없이 변함으로써 발생하는 오류이다. 따라서 대표성 오류는 온라인 설문조사의 확대 및 일반적 적용에 걸림돌로 작용하고 있다.

인터넷의 기술적인 발전은 온라인 설문조사에서 측정 오류를 줄이는 데 많은 기여를 하고 있다. 측정 오차를 줄이기 위해서 Gim과 Kim (2004)은 다양한 형태의 디자인과 개인화된 인터넷 설문조사를 이용하였다. 또한 Couper 등 (2001)은 인터넷 설문에서 선택태그 (select tag) 방식과 라디오 버튼 (radio button) 방식을 비교하여 라디오 버튼 방식이 설문 완성률이 높다는 사실을 실증적으로 보였다. 온라인 설문에서 응답자들이 고의적으로 사실을 숨기는 응답오차를 줄이기 위해 Park과 Pi (2013)은 신뢰도를 반영한 온라인 설문을 제시하였으며, 민감한 정보를 얻기 위해 Park 등 (2001)은 응답자의 비밀을 노출시키지 않으면서 보다 정확한 정보를 얻을 수 있도록 양적 확률화 응답기법을 도입하였다. Heo와 Chang (2011)은 인터넷 설문조사에서 익명성이 훼손됨에 따라 응답이 달라지는 문항들을 조사하였으며, Jeon (2006)은 인터넷 설문조사 시 응답자 반응에 영향을 미치는 요인을 연구하였다. 온라인 설문조사의 정확성을 높이기 위해 기술적인 방법과 방법론적인 연구를 병행하여 지속적으로 노력하고 있다.

온라인으로 수집된 자료의 질을 비교하기 위해 Tse (1998)은 무응답 비율을 조사하였는데 우편조사가 1.15%, 이메일조사가 3.0%로 우편조사가 이메일조사 보다 우수하다고 발표하였으나 그 외 다수의 연구에서 우편조사가 이메일조사보다 무응답 측면에서 우수하다고 발표하였다. 그러나 개방형 질문에서는 Mehta와 Sivadas (1995)는 우편조사보다 이메일조사가 완성률이 높다고 발표하였다. 온라인 설문조사에 관한 연구는 많은 부분 전통적인 설문조사와 비교연구, 효율적인 온라인 설문조사 등에 관한 연구이다. 온라인 설문조사 특히 인터넷 설문조사에서 자료의 정확성에 관한 연구는 미진한 실정이다.

본 논문의 목적은 인터넷 설문조사의 질적인 연구를 위해, 온라인 설문조사에서 응답자료의 정확성을 연구하는 것이다. 연구방법으로 인구통계학적 변수와 이용기간, 이용횟수 등 응답자의 행동패턴을 조사하는 행동변수를 이용하여 인터넷 설문조사 응답과 실제 데이터베이스의 자료와 비교분석함으로써 정확성을 추정하였다. 2절은 1차검증과 상세검증을 통해 정확성을 추정하였고 3장은 결론을 기술하였다. 우리가 이용한 자료는 한국과학기술정보연구원 (KISTI)의 NDSL (national discovery for science leaders) 사이트 회원들 중 원문서비스를 이용하는 회원들을 대상으로 실시한 설문조사이다.

2. 인터넷설문 자료의 검증

2.1. 모집단 및 표본설계

본 연구에 이용되는 인터넷 설문조사는 KISTI의 NDSL 사이트 회원들 중 원문서비스를 이용하는 회원들을 대상으로 실시하였다. 조사모집단은 2012년 10월 1일 기준으로 NDSL 회원인 7,254명으로 정의하였으며, 표본은 층화추출법으로 200명을 추출하였다. 층화방법은 Ryu와 Moon (2013)이 제안한 정규분포를 이용한 세분화를 이용하였으며, 표본 분포는 Table 2.1에 나와 있다. 첫 번째에서 네 번째 그룹은 최근 1년간 NDSL의 원문서비스를 이용한 모집단을 세분화한 그룹이며 표본은 150명을 모집단

크기에 비례할당하여 추출하였다. 다섯 번째 그룹은 최근 1년간 NDSL의 원문서비스를 이용하지 않은 그룹으로 이 그룹에서 50명의 표본을 추출하였다.

Table 2.1 Sample design

segmentations	populations	samples	characteristic
1st	952	33	active groups
2nd	1,071	38	
3rd	618	22	
4th	1,634	57	
5th	2,979	50	inactive group
total	7,254	200	

조사는 Figure 2.1에 나와 있는 바와 같이 HTML을 기반으로 하는 인터넷 설문조사를 사용하였다. 설문지는 주로 라디오 버튼을 이용하여 응답을 쉽게 할 수 있도록 설계하였다. 조사방법은 이메일로 인터넷 설문조사를 알리고 세분그룹별 응답표본 수까지 조사하였으며, 조사기간은 2011년 10월 8일부터 8월 10까지 3일간 이었다.

SQ1. Age <input type="radio"/> Teenage <input type="radio"/> Twenties <input type="radio"/> Thirties <input type="radio"/> Forties <input type="radio"/> Fifties <input type="radio"/> over Sixties	SQ2. E-mail <input type="text"/>	SQ3. Telephone <input type="text"/>
SQ4. Occupation <input type="radio"/> Administrative <input type="radio"/> Student <input type="radio"/> Professor <input type="radio"/> Worker <input type="radio"/> Librarian <input type="radio"/> Researcher <input type="radio"/> Owner-operator <input type="radio"/> Other ()		
SQ5. Organization <input type="radio"/> Small business <input type="radio"/> Research <input type="radio"/> University <input type="radio"/> Government <input type="radio"/> Non-profit organization <input type="radio"/> Big business <input type="radio"/> Other()		
SQ6. Employment period <input type="radio"/> Under 5years <input type="radio"/> 5-10years <input type="radio"/> 10-15years <input type="radio"/> 15-20years <input type="radio"/> over 20years		

Figure 2.1 Example of display for on-line survey

2.2. 1차검증

이번 조사에서 인구통계변수는 Figure 2.1에 나와 있듯이 연령, 이메일, 전화번호, 직업, 소속기관, 근무경력이다. 이들 변수 중 소속기관과 근무경력 등은 데이터베이스가 가지고 있는 자료와 상이한 자료이므로, 검증에 사용할 수 있는 인구통계변수는 연령과 직업이다. 또한 행동변수는 이용빈도와 이용기간이다. 이용빈도와 이용기간에 관한 질문은 Figure 2.2에 나와 있다.

Q2) How many time do you use KISTI copy service of paper?			
<input type="radio"/> almost everyday	<input type="radio"/> 2~3 times in a week	<input type="radio"/> 1 time in a week	<input type="radio"/> 2~3 times in a month
<input type="radio"/> 1 time in a month	<input type="radio"/> 1 time in a quarter	<input type="radio"/> 2~3 times in a year	<input type="radio"/> 1 time in a year
Q6) How long do you use KISTI copy service of paper?			
<input type="radio"/> under a year	<input type="radio"/> 1~3 years	<input type="radio"/> 3~5 years	<input type="radio"/> 5~7 years
<input type="radio"/> over 7 years			

Figure 2.2 Questions for activities

우리는 인구통계변수인 연령과 소속기관, 행동변수인 이용빈도와 이용기간으로 온라인 설문 자료의 불일치를 검증할 것이다. 나이에 대해 데이터베이스와 온라인 설문조사의 결과가 불일치한 응답자는 15명으로 7.5%, 소속기관에 대해 불일치한 응답자는 30명으로 15%, 이용횟수에 대해 불일치한 응답자는 184명으로 92%로 거의 대다수를 차지하였고, 이용기간에 대해 불일치한 응답자는 110명으로 55%로 과반수를 넘었다. 전체적으로 불일치 비율은 42.38%로 높게 나타났다. 또한 각 변수에서 차이에 대한 검정을 대응표본 t-검정으로 실시하였다. 결과는 Table 2.2에 나와 있다. 양측검정에 대한 유의확률을 보면 나이에 대해서는 유의한 차이가 나지 않지만 소속기관, 이용빈도, 이용기간에 대해서는 유의한 차이를 보이고 있다.

Table 2.2 Mismatch percentages and correlations

Questions	Mismatch cases	Mismatch percentages	P-values of two sided paired t-test
Age	15	7.50%	0.862
Organization	30	15.00%	<0.001
Frequency of utilization	184	92.00%	<0.001
Period of utilization	110	55.00%	0.04
Average	84.75	42.38%	

2.3. 상세검증

나이에 대해 불일치한 비율은 7.5%로 높게 나타났다. 나이에 대해 상세검증을 위한 자료는 Table 2.3에 있다. 표에 의하면 데이터베이스에는 29살, 39살, 49살, 59살인데 응답자가 한 살 올려서 응답한 경우가 8명으로 나타났는데 이는 데이터베이스에서 나이를 만나이로 계산하였으나 응답자가 일반나이를 응답하여 발생한 불일치로 판단된다. 데이터베이스에서는 60세인데 한 살 내려서 50대라고 응답한 응답자는 1명으로 나타났다. 이는 응답자가 나이를 착각하여 발생한 불일치로 판단된다. 그 외에도 2살을 올려서 응답한 응답자가 1명, 2살을 내려서 응답한 응답자가 1명, 3살을 올려서 응답한 응답자가 1명, 3살을 내려서 응답한 응답자가 1명, 30살을 내려서 응답한 응답자가 1명, 43살을 내려서 응답한 응답자가 1명으로 나타났다. 나이에서 보수적 관점에서 기준을 잘못 판단하거나 착각을 제외한다면 잘못된 응답한 경우는 6명으로 불일치 비율을 3%로 추정할 수 있고, 낙관적 관점에서 악의적으로 숨기려고 한 응답만을 불일치로 판단한다면 나이차가 -30, -43인 두 명으로 불일치 비율을 1%로 추정할 수 있다.

Table 2.3 Reasons of mismatch in age

Database	Survey	Difference	Cases	Reason
29	the thirties	+1	5	conducted by illusion
39	the forties	+1	2	
49	the fifties	+1	1	
60	the fifties	-1	1	
48	the fifties	+2	1	
51	the forties	-2	1	
27	the thirties	+3	1	
42	the thirties	-3	1	
59	the thirties	-30	1	
62	the twenties	-43	1	
total			15	

소속기관에 대해 불일치 비율은 15%로 높게 나타났다. 소속기관에 대해 상세검증을 위한 자료는 Table 2.4에 나와 있다. 자세히 살펴보면 데이터베이스에서는 일반가입자라고 소속기관을 밝히지 않았는데 인터넷 설문에서 소속기관을 밝힌 응답자가 21명으로 오류라기보다는 잘못된 데이터베이스 정보를

바로 잡아 주는 경우이다. 데이터베이스에서 대학이라고 기재하였는데 인터넷설문조사에서 중소기업이라고 응답한 응답자가 2명, 공공기관이라고 응답한 응답자가 2명, 대기업이라고 응답한 응답자가 1명, 기타라고 응답한 응답자가 3명으로 나타났다. 데이터베이스에 중소기업이라고 기재한 응답자가 기타라고 응답한 응답자가 1명으로 나타났다. 따라서 소속기관 문항에서 보수적 관점에서 소속기관을 데이터베이스와 다르게 응답한 경우를 모두 포함한다면 9명이 잘못 응답하였으므로 불일치 비율을 4.5%로 추정할 수 있다. 낙관적인 관점에서 데이터베이스 가입 시 소속과 인터넷 설문 시 소속기관이 달라 질 수 있으므로 이를 제외한다면 데이터베이스에서는 소속기관이 있는데 인터넷 설문 시 기타로 응답한 경우만을 불일치로 판단할 수 있으며, 이 경우 4명이 잘못 응답하였으므로 불일치 비율을 2%로 추정할 수 있다. 그리고 나이와 소속기관 두 문항 모두 불일치한 응답자는 1명으로 0.5%를 차지하였다.

Table 2.4 Reasons of mismatch in organization

Database	Survey	Cases	Total	Reason
general	small businesses	13	21	verified by on-line survey
	research institute	3		
	university	2		
	public institute	1		
	non-profit organization	1		
university	small businesses	2	8	hidden in on-line survey
	public institute	2		
	big businesses	1		
	general	3		
small businesses	general	1	1	
Total			30	

Table 2.5는 나이와 소속기관 두 문항에 대해 불일치 비율과 상세검증 후 불일치 비율들을 종합한 표이다. 최초의 불일치 비율에서 상세검증 통해 불일치 비율이 1/4 이하로 대폭적으로 낮아 졌음을 알 수 있다.

Table 2.5 Mismatch percentages for age and organization

Questions	First verification	Detail verification	
	Mismatch percentages	Pessimistic	Optimistic
Age	7.5%	3%	1%
Organization	15%	4.5%	2%
Age and organization	0.5%	0.5%	0.5%

이용빈도에 대해 질문은 “고객께서는 KISTI의 원문제공서비스를 얼마나 자주 이용하십니까?”이었다. 응답자들은 기준 기간 등 설명 도구가 없기 때문에 이용빈도의 기준을 최근 일주일, 한달 등 주관적으로 설정하였으리라 추정된다. 그리고 데이터베이스에서 사용한 기준은 2011년 1월 1일부터 12월 31일까지 1년을 기준으로 이용빈도를 추정하였으므로 측정오차를 피할 수 없었다. 또한 응답자들은 이용정보를 정확히 기억하기 어렵고, 인터넷 이용자들은 깊이 생각하지 않고 즉흥적으로 응답하는 경향이 높기 때문에 깊은 생각과 계산이 요구되는 이용빈도에서 불일치 비율은 92%로 설문조사 응답자의 거의 대다수가 데이터베이스의 정보와 다르게 응답하였다. 이러한 문제를 줄이기 위해서 추가 설명을 위한 팁이나 기준을 명확히 설명할 필요가 있다.

세부검증을 위해 내용을 살펴보면 Table 2.6에 나와 있듯이 거의 매일 사용한 이용자는 일주일에 2~3번 이용하였다고 응답하였으며, 일주일에 2~3번 이용한 이용자 2명은 거의 매일과 일주일에 1번으로 응답하는 등 한달에 1번 이상 이용하는 이용자들은 대체적으로 데이터베이스 결과의 주위의 항목

에 응답을 하였다. 그러나 이용빈도가 낮은 이용자들은 매우 다르게 응답하는 경향을 보였다.

Table 2.6 Detail table for frequency of utilization

Database \ Survey	almost everyday	2~3 times in a week	1 time in a week	2~3 times in a month	1 time in a month	1 time in a quarter	2~3 times in a year	1 time in a year
almost everyday	0	1	2	0	0	3	0	9
2~3 times in a week	1	0	2	1	0	6	2	24
1 time in a week	0	1	4	2	1	2	4	9
2~3 times in a month	0	0	4	1	1	16	5	23
1 time in a month	0	0	0	0	1	6	3	16
1 time in a quarter	0	0	0	0	2	1	5	15
2~3 times in a year	0	0	0	0	1	3	5	13
1 time in a year	0	0	0	0	1	0	0	4
total	1	2	12	4	7	37	24	113

이용기간에 대한 결과는 Table 2.7에 나와 있다. 이용기간 불일치 비율은 55%로 나타났다. 이용기간에 대한 질문은 기준도 명확하고, 이용빈도 질문보다 많은 생각을 요구되지 않기 때문에, 이용기간 오류의 비율이 이용빈도 보다 낮으나, 이 질문 역시 응답자들이 정확하게 기억하지 못하는 문제를 가지고 있는 것으로 나타났다. 세부 내용을 보면 데이터베이스에서 1년 미만 이용자가 7년 이상이라고 응답한 응답자가 3명 존재하고 있으나, 1년 미만 이용자에서 7년 미만 이용자는 대체적으로 데이터베이스 결과의 주위의 항목에 응답을 하였다. 그러나 7년 이상 이용자는 1년 미만이라고 응답한 응답자가 2명, 1~3년 미만이 5명, 3~5년 미만이 11명, 5~7년 미만이 16명으로 널리 분포하고 있다. 이용기간이 길수록 분산되는 현상을 보이고 있다. 이러한 현상은 이용빈도와 반대 현상을 보이고 있다.

Table 2.7 Detail table for period of utilization

Database \ Survey	under a year	1~3 years	3~5 years	5~7 years	more than 7 years
under a year	14	3	2	1	2
1~3 years	17	21	10	2	5
3~5 years	2	7	9	13	11
5~7 years	0	1	3	10	16
more than 7 years	3	0	6	6	36
Total	36	32	30	32	70

3. 결론

우리는 웹브라우저로 설문조사를 하는 인터넷 설문조사에서 자료의 정확성을 데이터베이스의 자료와 설문조사 자료의 비교를 통하여 검정하였다. 인터넷 설문조사는 2012년 8월 KISTI의 원문이용자를 대상으로 실시하였으며, 검증에 이용된 변수는 인구통계적 변수인 나이와 소속기관과 행동변수인 이용빈도와 이용기간이었다. 검정방법은 데이터베이스 자료와 인터넷설문조사의 응답결과를 비교하는 방법이었다.

1차 검증결과로서 데이터베이스 자료와 설문조사 자료의 불일치 비율은 나이가 7.5%, 소속기관이 15%, 이용빈도가 92%, 이용기간이 55%로서, 평균 42.38%로 높게 나타났다. 상세검정을 위해 불일치 원인을 분석하였다. 나이에 관한 질문의 불일치 원인의 반 이상이 만 나이와 일반 나이의 차이로부터 발생하였으며 이러한 불일치를 제거한다면 나이에 관한 불일치 비율을 비관적으로 3%, 낙관적으로 1%로 추정할 수 있었다. 소속기관에 관한 질문에 대해서도 상세검증을 통해 데이터베이스에서는 소속기관을 밝히지 않았으나 설문 시 소속기관을 밝힌 경우가 70%를 차지하였다. 이러한 불일치를 제거한

다면 소속기관에 관한 불일치 비율을 비관적으로 4.5%, 낙관적으로 2%로 추정할 수 있었다. 이는 상세 검증 전의 불일치 비율을 1/4 이하로 대폭 축소된 규모이다. 또한 데이터베이스의 자료와 비교분석을 통해 불일치 원인을 단순한 실수인지, 의도적인 것인지도 확인할 수 있었다. 이용빈도 질문에 대해 응답자의 기준과 데이터베이스에서의 기준의 차이, 정보를 기억하기 어려운 문제, 인터넷 이용자들의 이용행태 등 여러 가지 문제로 불일치 비율이 92%로 높게 나왔다. 상세검증을 통해 추가적으로 얻은 정보는 미미했다. 이용기간 질문에 대해서는 이용빈도의 불일치 비율보다 낮게 나왔지만 과반수가 넘는 52%로 나타났다. 상세검증을 통해 얻은 정보도 미미했다. 행동변수에 대해 이렇게 높은 불일치 비율은 전통적 설문조사에서도 일어날 수 있는 문제로 추정된다. 따라서 행동변수에 대해 전통적 설문조사에서는 자료의 정확성을 검증할 수 없으므로 본 연구의 결과가 간접적 자료로 활용될 수 있을 것이다.

우리는 인구통계적 변수를 통해 자료의 정확성을 검증할 수 있었지만 행동변수를 통해 자료의 정확성을 추정하기 어려웠다. 인구통계적 변수를 통해, 인터넷 설문조사 자료를 단순히 불일치율 방식으로 평가한다면 정확성이 낮다고 생각할 수 있으나, 상세검증을 통해서 인터넷 설문조사가 전통적 설문조사에 비해 자료의 정확성이 낮지 않다는 점과 인터넷 설문조사가 믿을 만한 자료라는 점을 시사하고 있다. 본 연구는 인터넷 설문조사에서 자료의 검증을 위한 사례연구이다. 인터넷 설문조사의 유효성과 확장성을 높이기 위해 관련 연구들이 지속적으로 요구된다.

References

- Couper, M. P. (2000). Web survey: A review of issues and approaches. *Public Opinion Quarterly*, **64**, 464-494.
- Couper, M. P., Traugott, M. W. and Lamias, M. J. (2001). Web surveys: Review of issues and approaches. *Public Opinion Quarterly*, **65**, 230-253.
- Gim, G. and Kim, G. (2004). Methodological issues in internet survey and development of personalized internet survey system using data mining techniques. *Journal of the Korean Society for Quality Management*, **32**, 93-108.
- Heo, S. and Chang, D. (2011). Effect of threats to anonymity on data reliability in internet survey. *Journal of the Korean Data & Information Science Society*, **22**, 785-794.
- Jeon, J. (2006). A study on the influencing factors of respondent's reaction to internet survey. *Journal of Marketing Management Research*, **11**, 95-113.
- Korea Communications Commission (2013). *2012 statistics for subscribers of wired or wireless telecommunications*, Korea Communications Commission, Kyeonggi.
- Korea Internet & Security Agency (2013). *2012 Korea internet white paper*, Korea Internet & Security Agency, Kyeonggi.
- Mehta, R. and Sivadas, E. (1995). Comparing response rates, response content and mail versus electronic mail surveys. *Journal of the Marketing Research Society*, **37**, 429-446.
- Park, H., Nam, K., Kim H. and Lee, K. (2001). Implementation of probabilistic response system for achieving sensitive information. *Proceedings of the Korean Data & Information Science Society*, Oct. 30, 79-86.
- Park, H., Nam, K., Kim H. and Myung, H. (2005). Implementation of on-line survey and analysis system based on database structure. *Journal of the Korean Data & Information Science Society*, **13**, 1-16.
- Park, H. and Pi, S. (2013). Study on the K-scale reflection the confidence of survey responses. *Journal of the Korean Data & Information Science Society*, **24**, 41-51.
- Ryu, G. and Moon, Y. (2013). A study on proposing a method for grouping R, F, and M in RFM model. *Journal of the Korean Data & Information Science Society*, **24**, 245-255.
- Ryu, J. (2001). Current state of implementation of on-line survey system. *Proceedings of Korea Association for Survey Research*, June 01, 141-156.
- Tse, A. (1998). Comparing the response rate, response speed and response quality of two methods of sending questionnaires: E-mail vs mail. *Journal of the Marketing Research Society*, **40**, 353-362.

A case study on verification of internet survey

Gui-Yeol Ryu¹ · Young-Soo Moon²

¹Department of Computer Science, SeoKyeong University

²Department of ASTI Business Team, KISTI

Received 20 August 2013, revised 9 September 2013, accepted 23 September 2013

Abstract

The object of study is to verify the accuracy of internet survey by comparing database data and internet survey. Internet survey was conducted on August, 2012. Respondents were subscribers of KISTI NDSL. Variables were age, organization as demographic variables, number of use, and period of use as attitude variables. Mismatch rates of age, organization, number of use, and period, are 7.5%, 5%, 92%, and 55% respectively. We could estimate the mismatch rate for age as 3% as a pessimistic point of view, and 1% as an optimistic point of view by detail verification. The mismatch rates of organization are 4.5% as a pessimistic point of view, and 2% as an optimistic point of view. The mismatch rates for the frequency of use, the period of use are very high, because measurement error, problems in memory, and internet attitude, etc. Implication of this study is that data of internet survey could be reliable. Many further researches are needed for verification of internet survey.

Keywords: Internet survey, measurement error, mismatch, on-line survey.

¹ Associate professor, Department of Computer Science, SeoKyeong University, Seoul 136-704, Korea.

² Corresponding author: Principal researcher, Department of Performance Diffusion, KISTI, Seoul 130-741, Korea. E-mail: youngsum@kisti.re.kr