

전화 선거여론조사에서 무응답률 증가로 인한 편의와 응답률 제고 방안[†]

허순영¹ · 이수철²

¹창원대학교 통계학과 · ²창원대학교 수학과

접수 2016년 2월 4일, 수정 2016년 3월 4일, 게재확정 2016년 3월 10일

요약

조사결과의 신속성이 요구되는 선거여론조사의 특성상 짧은 시간에 유권자들의 의견을 파악할 수 있고 저렴한 조사비용이라는 장점 때문에 낮은 응답률이라는 심각한 결함에도 불구하고 대부분 선거 여론조사는 전화조사로 이루어진다. 현재 국내에서 실시되는 많은 선거여론조사는 연령별, 성별, 지역별 인구수 또는 유권자수에 비례하여 할당된 수만큼 조사하는 방식을 취하고 있다. 무응답이 발생하여 초기에 할당된 수만큼 조사가 이루어지지 않은 경우, 응답자들의 가중치를 조정하여 추정하는 방법을 사용한다. 본 연구에서는 무응답이 존재하는 경우, 현재 국내 여론조사에서 사용하는 추정량의 편의와 분산의 편의를 수리적으로 도출하였다. 이 편이는 응답집단과 무응답집단 간의 의사 차이와 무응답률의 크기에 비례한다. 또, 2012년 대선 선거여론 조사 자료를 기초로 전화 선거여론조사에서의 응답률을 산출하였는데, 평균 약 14.4%에 불과하였다. 한편, 자체적으로 실시한 ‘전화 선거여론조사에 대한 의식조사’에서, 응답자들이 전화 선거여론조사에 응답하지 않는 가장 큰 이유로 “귀찮아서”를 가장 많이 선택하였고, 여론조사 응답률이 가장 낮은 20대가 가장 많이 선택한 이유도 “귀찮아서”였다. 따라서 응답률을 높이기 위해서는 인센티브를 지급하는 등 조사대상자들의 응답 동기를 높이기 위한 보다 적극적인 노력이 필요하다.

주요용어: 무응답, 인센티브, 전화조사, 제18대 대선 선거여론조사, 편의.

1. 서론

조사연구의 초기에는 높은 응답률, 면접자와 응답자 간의 좋은 관계의 발전, 측정과정에서 시각적 도구의 사용가능성 등으로 의해 직접면접에 의한 자료수집방법이 선호되었으나, 직접면접조사의 막대한 비용증가는 대체방법에 대한 많은 연구들을 촉발시켰다. 그 결과 전화조사의 질이 직접면접조사와 비교될 만큼 좋다는 연구결과들에 기초하여 1970년대 이후부터 최근까지 전화조사는 가장 많이 사용되는 자료수집방법이 되었다 (Chang과 Korsnick, 2009).

우리나라에서 전화조사를 가장 많이 사용하는 분야는 선거여론조사이다. 조사결과의 신속성을 필요로 하는 선거여론조사는 저렴한 비용으로 짧은 시간에 유권자들의 의견을 파악할 수 있다는 장점 때문에 낮은 응답률이라는 심각한 결함을 가지고 있음에도 불구하고 전화조사로 실시된다 (Park, 1978). 우리나라는 1987년 제13대 대통령선거에서 처음 시도되어 최근 2014년 6.4지방선거까지 선거여론조사 기법

[†] 이 논문은 2015~2016 창원대학교 자율연구과제 연구비 지원으로 수행된 연구결과임.

¹ 교신저자: (51140) 경남 창원시 의창구 창원대학교 20, 국립창원대학교 통계학과, 교수.

E-mail: syheo@changwon.ac.kr

² (51140) 경남 창원시 의창구 창원대학교 20, 국립창원대학교 수학과, 교수.

이 꾸준히 발전해 오고 있다. 2016년 4월 13일은 제20대 대한민국 국회의원선거가 예정되어 있고 변함 없이 많은 여론조사가 실시될 것이다. 그러나 전화조사에 의한 선거여론조사의 가장 큰 문제점 중의 하나는 낮은 응답률 또는 높은 무응답률이다.

무응답이 존재하는 경우, 단순히 무응답을 무시하거나, 무응답을 대체하거나, 가중치 보정을 하거나 또는 재조사를 하는 등의 방법들이 있다. Park과 Na (2015)는 무응답이 존재하는 경우 재조사를 실시하고 보조변수 정보를 사용하여 추정량의 효율을 높이는 방법을 제시하였다. Kwak과 Choi (2014)는 모형에 기반을 둔 무응답대체방법에서 무응답체계 (nonresponse mechanism)에 대한 가정의 정확성을 비교하였다. Lee와 Kang (2012)은 사례를 통하여 무응답률에 따른 분석결과의 차이를 비교하였다. Cho 등 (2007)은 2006년 6.13지방선거에서 실시된 전화여론조사를 중심으로 전화조사의 오차요인인 표집방법, 표집틀, 유선전화 보급의 감소와 무선전화 가입자수의 증가, 무응답자의 처리, 연령대별 투표율 차이 등을 지적하면서 선거여론조사의 발전방향을 제시하고 있다. 그들은 무응답자를 처리하는 방법으로 성별이나 연령 등 무응답자의 특성에 따른 관별분석을 이용한 무응답추정의 필요성을 제안하였다.

Lee 등 (2012)는 유선전화의 모집단 포함률 및 젊은 층의 낮은 채택률과 투표자 선정에서 발생하는 문제 등의 해결방법으로 유무선 RDD의 혼합조사를 제안하고 있으며, Cho와 Cho (2010)는 무선표집 방법으로 전화조사를 실시하여 조사성공률, 성공 및 실패사례의 응답자의 인구학적 특성, 응답단계별 이슈에 대한 응답특성 등의 분석결과를 제공하였다. 이 외에도 전화 선거여론조사의 정확성을 높이기 위한 많은 연구들이 수행되고 있으나 대부분 이론적 접근을 통해 추정치의 정확도를 높이는 데 초점이 맞추어져 있고, 보다 근본적인 문제인 무응답률을 낮추려는 연구들은 찾아보기 어렵다.

본 연구는 우리나라에서 현재 많은 조사회사들이 선거여론조사에서 모수추정방법으로 사용하고 있는 추정량과 그 분산의 추정량에 대해 무응답이 존재하는 경우 발생하는 편의 (bias)의 크기를 수리적으로 정리하였다. 이어서, 2012년 대선 선거여론조사 자료를 기초로 유형별 전화여론조사의 응답률을 산출하였고, 2014년 지방선거의 사례를 통해 연령대별 응답률을 살펴보았다. 마지막으로 저자들이 자체적으로 실시한 ‘전화 선거여론조사에 대한 의식조사’를 통해 응답자들이 응답하지 않는 이유를 살펴보고 응답률 제고를 위한 방안을 제안한다.

2. 선거여론조사 후보자 지지율 추정

현재 국내에서 실시되는 많은 전화 선거여론조사는 연령별, 성별, 지역별 인구수 또는 유권자수에 비례하여 표본을 할당할 후, 표집틀로 사용하는 전화번호목록에서 무작위로 번호를 선정하여 할당된 수만큼 조사하는 방식을 취하고 있다. 그 과정에서 각 그룹별 할당된 인원을 조사하기 위해 조사초기에 연령이나 성별을 확인하고 이미 할당된 표본수의 조사가 이루어진 경우에는 조사를 종료하는 방법으로 조사가 이루어지는 것이 일반이나, 경우에 따라서는 표본수 만큼 조사한 후 각 그룹별 가중치를 재조정하는 방식으로 진행되기도 한다. 어떤 방법이든지 조사결과는 응답자들의 응답결과만을 가지고 분석될 수밖에 없고 응답률이 낮은 경우에는 조사결과의 분석에 상당한 주의가 필요하다.

2.1. 후보자 지지율의 추정과 편의

크기 N 인 모집단을 각각 크기 N_h 인 H 개의 층으로 나누고 ($h = 1, \dots, H$), 크기 n 인 표본을 각 층의 모집단 크기에 비례할당하여 h 번째 층의 표본수 $n_h = (N_h/N)n$ 가 결정되었다 하자. 이 후, 각 층에서 할당된 표본수 n_h 만큼 전화번호목록에서 비복원 단순임의무작위추출 (simple random sampling

without replacement)하여 후보A의 지지율을 조사한다. 이 때, 모집단에서 후보A의 지지율 p 는

$$p = \frac{1}{N} \sum_{h=1}^H \sum_{i=1}^{N_h} y_{hi} = \frac{1}{N} \sum_{h=1}^H A_h \quad (2.1)$$

이고,

$$\hat{p} = \frac{1}{N} \sum_{h=1}^H N_h \hat{p}_h = \frac{1}{N} \sum_{h=1}^H N_h \left(\frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} \right) \quad (2.2)$$

에 의해 추정된다. 여기서, A_h 는 h 번째 층에서 후보A의 지지자 수이고, h 번째 층에서 후보A의 지지율은 $p_h = A_h/N_h$ 이다. $\hat{p}_h = \sum y_{hi}/n_h$ 로 h 번째 층의 표본 지지율이고, y_{hi} 는 h 번째 층의 i 번째 조사단위의 응답으로 후보A를 지지하면 1이고 지지하지 않으면 0이다.

층 h 의 모집단이 크기 $N_{h(R)}$ 인 응답자그룹과 크기 $N_{h(NR)}$ 인 무응답자그룹으로 나누어지고, 각 그룹에서 후보A의 지지자수는 $A_{h(R)}$ 과 $A_{h(NR)}$ 이라 하자. 이 때, $N_h = N_{h(R)} + N_{h(NR)}$ 이고 $A_h = A_{h(R)} + A_{h(NR)}$ 이므로, 식 (2.1)의 후보A의 지지율은

$$p = \frac{1}{N} \sum_{h=1}^H \{A_{h(R)} + A_{h(NR)}\} = \frac{1}{N} \sum_{h=1}^H N_{h(R)} p_{h(R)} + \frac{1}{N} \sum_{h=1}^H N_{h(NR)} p_{h(NR)}$$

로 표현된다. 여기서, $p_{h(R)} = A_{h(R)}/N_{h(R)}$ 이고 $p_{h(NR)} = A_{h(NR)}/N_{h(NR)}$ 로 응답자그룹과 무응답자그룹의 후보A 지지율이다.

또, h 번째 층의 표본 s_h 는, 응답자그룹에서 $n_{h(R)}$ 명이 추출되고 무응답자그룹에서 $n_{h(NR)}$ 명이 추출되어 $s_h = s_{h(R)} \cup s_{h(NR)}$ 라 하자. 이 때, $s_{h(R)}, s_{h(NR)}$ 는 층 h 에서 응답자그룹과 무응답자그룹의 표본이고 $s_{h(R)} \cap s_{h(NR)} = \emptyset$ 이다. 식 (2.2)의 후보A 지지율 추정량은

$$\begin{aligned} \hat{p} &= \sum_{h=1}^H \left(\frac{N_h}{N} \right) \frac{1}{n_h} \left(\sum_{i \in s_{h(R)}} y_{hi} + \sum_{i \in s_{h(NR)}} y_{hi} \right) \\ &= \sum_{h=1}^H \left(\frac{N_h}{N} \right) \frac{n_{h(R)}}{n_h} \hat{p}_{h(R)} + \sum_{h=1}^H \left(\frac{N_h}{N} \right) \frac{n_{h(NR)}}{n_h} \hat{p}_{h(NR)} \end{aligned}$$

로 나타낼 수 있다. 여기서, $\hat{p}_{h(R)} = \sum_{i \in s_{h(R)}} y_{hi}/n_{h(R)}$ 이고 $\hat{p}_{h(NR)} = \sum_{i \in s_{h(NR)}} y_{hi}/n_{h(NR)}$ 이다. 이 \hat{p} 의 기대값은

$$\begin{aligned} E[\hat{p}] &= \sum_{h=1}^H \left(\frac{N_h}{N} \right) \frac{1}{n_h} E[n_{h(R)} \hat{p}_{h(R)}] + \sum_{h=1}^H \left(\frac{N_h}{N} \right) \frac{1}{n_h} E[n_{h(NR)} \hat{p}_{h(NR)}] \\ &= \sum_{h=1}^H \frac{N_{h(R)}}{N} p_{h(R)} + \sum_{h=1}^H \frac{N_{h(NR)}}{N} p_{h(NR)} = p \end{aligned}$$

로, 식 (2.2)의 \hat{p} 은 모집단에서 후보A 지지율 p 의 불편추정량이다.

한편, 표본에 무응답자가 존재할 때 $\hat{p}_{h(NR)}$ 은 추정할 수 없으므로, 많은 여론조사기관은 후보A 지지율의 추정량으로 응답자들의 가중치를 조정하여

$$\tilde{p} = \sum_{h=1}^H \left(\frac{N_h}{N} \right) \frac{1}{n_h} \sum_{i \in s_{h(R)}} \frac{n_h}{n_{h(R)}} y_{hi} = \sum_{h=1}^H \left(\frac{N_h}{N} \right) \hat{p}_{h(R)} \quad (2.3)$$

을 사용한다. 이 때, $E[\tilde{p}] = \sum_{h=1}^H (N_h/N) p_{h(R)}$ 이고, \tilde{p} 의 편의 (bias)는

$$bias(\tilde{p}) = E[\tilde{p}] - p = \sum_{h=1}^H \frac{N_{h(NR)}}{N} (p_{h(R)} - p_{h(NR)})$$

이다. 만일, 응답자그룹과 무응답자그룹의 후보A 지지율이 $p_{h(R)} = p_{h(NR)} = p_h$ 로 동일하다면, $bias(\tilde{p})$ 가 0이 되어 \tilde{p} 는 불편추정량이 된다. 그러나 두 그룹의 지지율이 크게 다르다면 \tilde{p} 의 편이는 커진다. 또, 모집단에서 무응답그룹의 비율 $N_{h(NR)}/N_h$ 이 클수록 편이는 더 커진다.

2.2. 후보자 지지율 추정량의 분산과 편의

무응답자가 존재하지 않는 경우, 후보A 지지율의 추정량 \hat{p} 의 분산과 분산추정량은 각각

$$Var(\hat{p}) = \sum_{h=1}^H \left(\frac{N_h}{N} \right)^2 \frac{N_h - n_h}{N_h - 1} \frac{p_h(1 - p_h)}{n_h} \quad (2.4)$$

$$\widehat{Var}(\hat{p}) = \sum_{h=1}^H \left(\frac{N_h}{N} \right)^2 \frac{N_h - n_h}{N_h} \frac{\hat{p}_h(1 - \hat{p}_h)}{n_h - 1} \quad (2.5)$$

으로 $\widehat{Var}(\hat{p})$ 는 $Var(\hat{p})$ 의 불편추정량이다 (Cochran, 1977; Scheaffer 등, 1990). 한편, 무응답자가 존재할 때, 추정량 \tilde{p} 의 분산과 분산추정량은 각각

$$Var(\tilde{p}) = \sum_{h=1}^H \left(\frac{N_h}{N} \right)^2 \frac{N_h - n_{h(R)}}{N_h - 1} \frac{p_{h(R)}[1 - p_{h(R)}]}{n_{h(R)}} \quad (2.6)$$

$$\widehat{Var}(\tilde{p}) = \sum_{h=1}^H \left(\frac{N_h}{N} \right)^2 \frac{N_h - n_{h(R)}}{N_h} \frac{\hat{p}_{h(R)}[1 - \hat{p}_{h(R)}]}{n_{h(R)} - 1} \quad (2.7)$$

이다. 표본조사 결과 $n_{h(R)}$ 가 결정되었을 때, $\widehat{Var}(\tilde{p})$ 는 식 (2.6)의 $Var(\tilde{p})$ 에 대해 불편추정량이나 반드시 식 (2.4)의 $Var(\hat{p})$ 에 대해 불편추정량이라고 할 수 없다.

무응답이 존재하고 모집단의 응답그룹과 무응답그룹의 후보A 지지율이 동일할 때, 즉, $p_{h(R)} = p_{h(NR)} = p_h$ 일 때, $\widehat{Var}(\tilde{p})$ 는 식 (2.4)의 $Var(\hat{p})$ 에 대한 불편추정량이지만 실제 표본수는 감소했기 때문에 식 (2.5)의 $\widehat{Var}(\hat{p})$ 에 비해 분산은 과대추정 된다. 그러나 $p_{h(R)} \neq p_{h(NR)}$ 라면 $\widehat{Var}(\tilde{p})$ 는 분산을 과대추정할 뿐 아니라, $\hat{p}_{h(R)}$ 이 p_h 의 편의추정량으로 $\hat{p}_h(R)$ 의 분산은 응답그룹의 분산만을 반영하기 때문에 모집단 전체에 대한 분산을 설명하지 못하고 편의 추정량이 된다.

3절에서는 실제적 사례를 통해 전화 선거여론조사의 응답률과 응답률을 높이기 위한 방안들을 살펴본다.

3. 전화 선거여론조사 응답률

3.1. 2012년 대통령선거

Table 3.1과 Table 3.2는 2012년 제18대 대선 D-100일을 전후한 2012년 9월부터 공직선거법 제108조에 의거한 선거여론조사 결과공표금지 기간 시작 전일인 12월 13일까지 언론에 공표된 전화여론조사 결과들로부터 응답률을 정리한 것이다. 한국갤럽이 이 기간 동안 각종 언론매체를 통해 공개된 내용을 정리하여 회사 홈페이지에 게시한 내용과, 한국갤럽이 정리한 내용에 포함되지 않은 일부 조사결과들을 저

자들이 직접 조사하여 추가하였다 (Korea Gallup, 2012). Table 3.1은 그 결과를 기초로 조사의뢰처와 조사회사, 표집틀, 조사방법 별로 응답률을 정리한 것이다. 이 기간 동안 실시되어 언론에 공표된 주요 조사들이 대부분 포함되었지만 일부조사는 이 표에서 누락되어 있을 수도 있다. Table 3.2는 Table 3.1을 표집틀과 조사방법별로 재정리한 것이다.

Table 3.1 The response rates of the public opinion polls on the 18th presidential election by client and survey organization, which were conducted from September 1 to December 13, 2012.

Frame	Survey method	Client	Survey organization	No. of polls providing response rate	Average response rate	
Wire or mobile phone subscriber	Telephone Interviewer	SBS	TNS	8	15.2%	
		The Hankyoreh	KSOI or Research Plus	6	15.7%	
		Joongangilbo	Joongangilbo	8	30.8%	
		The Seoul Shinmun	Embrain	5	29.6%	
		The Munhwa Ilbo	Korea Research Center	3	18.4%	
		The Dong-A Ilbo or The Dong-A Ilbo+Channel A	Research & Research	4	23.4%	
		The Kukmin Ilbo	World Research or Global Research	2	18.1%	
		MBC	Korea Research	7	21.8%	
		KBS	Media Research	4	16.9%	
		Korean Local Press Association	Korea Gallup	3	29.7%	
		The Segye Times	World Research	2	19.0%	
		ARS	JTBC+Joongangilbo	Realmeter	81	9.2%
		Mobile phone subscriber only	Telephone Interviewer	Korea Gallup	Korea Gallup	16
Ohmynews	Research View			16	10.2%	
Wire phone subscriber only	Telephone Interviewer	View & Poll	Research View	7	8.2%	
		Korean Local Press Association	KM Survey Research	3	16.8%	

* The numbers indicate the number of polls, which were conducted and released response rate to the press from September 1 to December 13, 2012.

Table 3.2 The response rates of the public opinion polls on the 18th presidential election by frame and survey method, which were conducted from September 1 to December 13, 2012.

Frame	Survey method	No. of survey organization	No. of polls	No. of polls providing response rate	Average response rate
Wire or mobile phone subscriber	Interviewer	11	71	52	21.7%
	ARS	1	85	81	9.2%
Mobile phone subscriber only	Interviewer	1	16	16	22.6%
	ARS	2	25	23	9.2%
Wire phone subscriber only	Interviewer	1	3	3	16.8%

Table 3.1과 Table 3.2에 의하면, 이 기간 동안 많은 조사회사들이 유선 전화가입자와 무선 전화가입자 모두를 대상으로 조사원인터뷰방식으로 조사를 실시하였는데, 언론에 공표된 71개 조사들 중 응답률을 제공한 52개 조사의 평균 응답률은 약 21.7%에 불과하다. 한편, 한국갤럽이 휴대전화가입자만을 대상으로 조사원인터뷰방식으로 진행한 총 16개 조사의 평균 응답률은 22.6%로 앞의 유무선혼합조사에 비해 약간 높으나 응답률은 여전히 5명당 1명 꼴에 불과하다. 자동응답시스템 (automated response system; ARS)에 의한 조사의 경우 응답률은 훨씬 낮다. JTBC와 중앙일보가 공동으로 의뢰하고 리얼미터에서 유무선 전화가입자들을 대상으로 실시한 ARS조사의 경우, 이 기간 동안 실시한 85회의 조사 중 응답률을 제공한 81개 조사의 평균 응답률은 9.2%에 불과하다. 또 오마이뉴스와 뷰엔폴의 의뢰로 리서치뷰에서 휴대전화가입자만을 대상으로 실시한 25개 조사 중 응답률을 제공한 23개 조사의 응답률 역시 9.2%에 불과하다. 결국 ARS 전화조사의 경우 10명당 1명 미만이 응답하였고, 조사원인터뷰방식으로 진행된 조사에 비해 응답률은 절반수준도 되지 못한다. 한편, 한국지방신문협회가 의뢰하고 케이엠조사연구에서 유선전화 가입자만을 대상으로 조사원인터뷰방식으로 실시한 3개 조사의 평균 응답률은 16.8%이나, 이 경우 유선전화가입자수가 감소하는 추세를 고려할 때 심각한 포함오차를 우려할 수 있다. 이 기간 동안 실시된 전화 선거여론조사 중 응답률을 제공한 175건의 응답률을 단순평균하면 약 14.4%에 불과하다.

3.2. 2014년 6.4지방선거

Table 3.3은 2014년 6.4지방선거에서 한 여론조사기관이 수행한 경상남도 지역의 후보자선호도조사 자료를 기초로 작성된 것이다. 두 번째 열은 2013년도 경상남도의 연령대별 주민등록인구수이고 세 번째 열은 각 연령대별 모집단의 인구비율이다. 이 조사는 무선전화목록을 표집틀로 하여 전화번호를 무작위추출하였고, 연결된 통화수는 총 10,614건 중 유효응답자는 1,789건이었다.

표집틀이 모집단을 잘 반영하지 못할 때 발생하는 오차를 포함오차 (coverage error)라고 한다 (Couper, 2000). 미래창조과학부가 통신사업자의 제출 자료에 기초하여 작성한 2014년도 무선 통신서비스 통계현황에 따르면 2013년 12월 기준 이동전화 가입자수는 약 5,468만 명이고 2013년도 대한민국 주민등록인구는 약 5,114만 명으로 한 사람당 평균 이동전화 한 대를 이용하고 있었다 (http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=2755; KOSIS, 주민등록인구 현황). 한편, 미래창조과학부가 주관하고 한국인터넷진흥원에서 전국의 만12~59세 2500명을 대상으로 가구방문 면접조사로 실시한 「2014년 정보보호실태조사」에 의하면 각 연령대별로 휴대전화를 보유하고 있지 않다는 응답비율이 10대는 0.8%, 20대와 30대는 0% 40대와 50대는 각각 0.1%와 0.2%에 불과했다 (Korea Internet & Security Agency, 2015). 이 조사는 60대 이상을 조사대상에 포함시키지 않아 60대 이상의 휴대전화 보유율은 제공하지 못한다. 단지, 다른 연령대에 비하면 휴대전화 보유율이 낮고 그 결과 포함오차는 상대적으로 더 클 것으로 예상된다. 그러나 미래창조과학부가 주관하고 한국인터넷진흥원에서 조사한 또 다른 조사인 「2013년 모바일인터넷이용실태조사」에 의하면 60대 이상 인구의 휴대폰 보유율도 크게 낮지는 않을 것으로 예측가능하다 (Korea Internet & Security Agency, 2013).

따라서 Table 3.3의 표본크기 (n_h , sample size)는, 표집틀로 사용한 무선전화목록의 모집단에 대한 포함오차 (coverage error)를 무시할만하다는 전제 하에 전화가 연결된 10,614명을 연령대별 모집단의 크기에 비례하여 배분한 결과이다. 다섯 번째 열은 각 연령대별로 조사를 완료한 유효표본수 ($n_{h(R)}$)이다. 마지막 열에는 응답률이 주어졌다. 이 응답률은 $N_{h(R)}/N_h$ 의 추정값이다. Table 3.3에 의하면 20대와 30대의 응답률은 100명당 2명 또는 4명으로 매우 낮다. 이 경우 식 \hat{p} 에 의해 후보자A 지지율을 추정할 때, 20대의 각 응답자는 표본에서 응답하지 않은 1749/38 = 46.0명을 대표하는 것이다. 즉, 응답하지 않은 45명의 의견이 응답한 한명의 응답자와 동일하다는 가정이 필요하다. 이 경우, 추정량과 그 분산 추정량은 심각한 편의를 가져올 수 있다. 이러한 무응답으로 인해 야기되는 문제를 이론적으로 해결하기 위해 무응답 보정을 위한 다양한 통계적 방법이 연구되고 있다. 그러나 보다 근본적으로 응답률 자체를 높이려는 노력이 필요하다. 그러기 위해서는 응답자들이 응답하지 않는 이유에 대해 관심을 가질 필요가 있다.

참고로, Table 3.3에서 60대 이상의 응답률이 50대에 비해 약 9%정도 낮게 나타났는데, 이는 실제 응답률이 낮은 것일 수도 있지만 이 연령대의 낮은 이동전화 보유율과 그에 따른 모집단의 포함오차에 기인한 것일 수도 있다. 면밀한 원인을 파악하기 위해서는 연령대별 이동전화 보유율에 대한 보다 정확한 정보가 필요하다.

Table 3.3 Population of Gyeongsangnam-do, sample size, number of respondents of an election poll on Gyeong-nam area local government head election in 2014

Age group	Population* (N_h)	Population proportion (N_h/N)	Sample size (n_h)	No. of respondents ($n_{h(R)}$)	Respondent rate ($n_{h(R)}/n_h$)
19+20s	435,876	16.5%	1,749	38	0.022
30s	499,756	18.9%	2,006	85	0.042
40s	582,410	22.0%	2,337	253	0.108
50s	529,511	20.0%	2,125	766	0.360
60s+	597,113	22.6%	2,397	647	0.270
Total	2,644,666	100.0%	10,614	1,789	0.169

* Source: The 2013 Gyeongsangnam-do resident registration population statistics (Statistics Korea, KOSIS)

Table 3.4 Distribution of respondents by age, educational attainment, and sex groups in the survey on the attitude toward the telephone election polls

Age group	Educational attainment*			Total	Sex		
	High school graduates	Collage graduates	More than master		Male	Female	Total
19+20s	6 (7.50%)	72 (90.00%)	2 (2.50%)	80	37 (46.25%)	43 (53.75%)	80
30s	5 (15.63%)	26 (81.25%)	1 (3.13%)	32	9 (28.13%)	23 (71.88%)	32
40s	29 (46.03%)	27 (42.86%)	7 (11.11%)	63	24 (38.1%)	39 (61.9%)	63
50s	26 (40.00%)	28 (43.08%)	11 (16.92%)	65	42 (63.64%)	24 (36.36%)	66
60s+	6 (33.33%)	11 (61.11%)	1 (5.56%)	18	17 (94.44%)	1 (5.56%)	18
Missings	-	-	-	2	-	-	1
Total	72	164	22	260	129	130	260

* Each category of educational attainments includes dropouts.

Table 3.5 Distribution of responses by age group to the question “What are the biggest reasons that you do not answer surveys? Select two”. (unit: %)

Categories	Age group					Total
	19+20s	30s	40s	50s	60s+	
1. Feel bothered	28.57	14.04	21.50	23.64	13.79	22.89
2. Do not know the survey topics	8.84	3.51	12.15	7.27	6.90	8.44
3. Feel worthless of answering	6.80	5.26	5.61	7.27	24.14	7.56
4. Be not in the situation to answer the phone	15.65	19.30	20.56	25.45	20.69	20.00
5. Do not trust the survey company	5.44	15.79	10.28	14.55	17.24	10.89
6. Worry about disclosure of personal information	12.93	17.54	10.28	9.09	6.90	11.56
7. Worry about voice-phishing pretending surveys	21.09	22.81	18.69	10.00	10.34	17.33
8. Etc	0.68	1.75	0.93	2.73	0.00	1.33
No. of total responses	147	57	107	110	29	450

*The numbers, except numbers in the last row, are column percents of each category. The last row gives the numbers of responses by age group with total responses last.

3.3. 조사단위의 무응답

여론조사 무응답률을 줄이기 위해서는, 무엇보다도 여론조사에 대한 응답자들의 태도를 아는 것이 중요하지만 지금까지 이러한 조사가 실시된 대한 기록을 찾기 어렵다. 우리는 2014년 4월 13일부터 29일 까지 창원대학교 학부학생들과 평생교육센터 교육생들 중 260명을 편의추출하여 「전화 선거여론조사에 대한 의식조사」를 실시하였다.

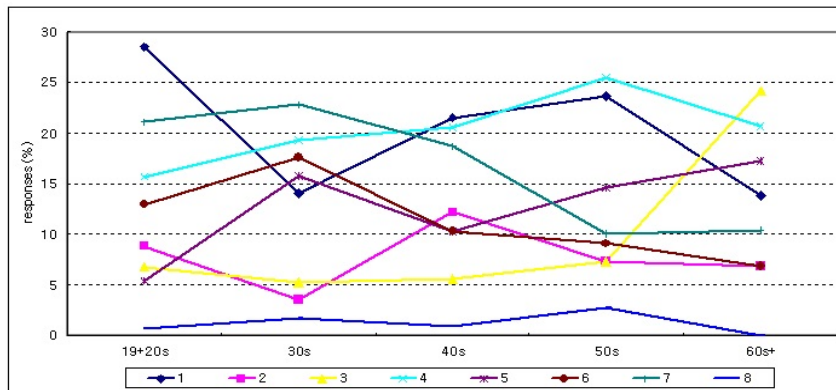


Figure 3.1 Distribution of responses by age group to the question “What are the biggest reasons that you do not answer surveys? Select two”.

Table 3.4는 이 조사 응답자들의 연령·교육수준·성별 응답자 분포를 정리한 것이다. 각 범주별 응답자 수 아래 제시한 백분율은 각 연령대별 행백분율이다.

우리는 응답자들에게 전화 선거여론조사에 응답하지 않는 가장 큰 이유에 대해 중복응답할 수 있도록 질문하였다. Table 3.5는 이 질문에 대한 연령대별 응답분포를 보여준다. Figure 3.1는 이 분포를 도표로 나타낸 것이다. Figure 3.1에서 범례의 각 숫자는 Table 3.5의 Category에서 각 응답범주들의 숫자에 대응한다. 전체 응답자가 가장 많이 선택한 이유는 “1. 귀찮아서”였고, 그 다음이 “4. 전화받을 상황이 아니라서”였다. 연령대별로 살펴보면, 선거여론조사 응답률이 가장 낮은 20대의 경우, 가장 많이 선택한 이유는 “1. 귀찮아서”, 그 다음이 “7. 조사를 사칭한 보이스피싱 등이 염려되어”였고, 30대는 “7. 조사를 사칭한 보이스피싱 등이 염려되어”가 가장 높았고, 그 다음은 “4. 전화받을 상황이 아니라서”였다. 흥미로운 것은 60대 이상의 경우, “3. 응답할 가치를 못 느껴서”라는 응답이 다른 연령대에 비해 월등히 높았다.

이 결과에 의하면 연령대별 응답률을 높이기 위한 전략을 달리할 필요가 있다. 특히 응답률이 낮은 20대와 30대의 응답률을 높이기 위해서는, 젊은 층으로 하여금 조사에 참여하려는 동기를 부여하는 것이 필요하고 조사를 사칭한 보이스피싱에 대한 우려를 불식시키려는 노력 역시 필요하다.

Heo와 Chang (2014)은 2012년 대선 여론조사 자료를 기초로 응답자 기회비용을 산정하였다. 전국의 5인 이상 사업체 월평균임금을 기준으로, 2012년 대선 전화여론조사의 응답자 기회비용은 조사당 평균 약 1,115원 정도 발생한 것으로 산정되었다. Tourageau 등 (2010)은 선거참여 여부를 주제로 한 전화여론조사에서 \$5의 인센티브의 제공한 집단의 조사응답률이 인센티브가 없는 집단에 비해 약 20% 이상 증가했음을 보여주었다. 이 외에도 외국에서는 인센티브 지급이 응답률을 높였음을 보여주는 많은 연구들이 있다. Carlson (1996)은 미국 Idaho주에 새로 이주한 사람들을 대상으로 하는 우편조사에서 \$2의 인센티브를 지급한 그룹의 응답률은 73%로 인센티브를 지급하지 않은 그룹의 응답률 58%보다 크게 높았음을 보여준다. Johnson과 McLaughlin (1990)은 인센티브의 지급시점에 대한 연구에서, \$5의 인센티브를 선지급한 그룹의 응답률은 83%로 10의 후지급을 약속한 그룹의 응답률 72%보다 높았음을 보여준다. 이 외에도 Dillman (2000)은 인센티브가 응답률을 높인 많은 사례들을 제시하고 있다.

여론조사를 포함한 각종 설문조사는 정보를 가지고 있는 응답자들의 시간과 노력이라는 비용을 대가로 정보를 획득하는 연구방법이다. 조사를 완료한 응답자들에게 그들의 시간과 노력에 상응하는 금액을 직접 지불하거나, 불우이웃에게 기부할 기회를 주거나, 이동전화요금에서 차감해주거나, 또는 각종 마일리지 제공하는 등의 방법으로 지불함으로써 젊은 층의 응답률을 높이는 방안을 모색할 필요가 있다.

4. 결론

조사연구의 초기에는 직접면접조사가 조사방법을 선도했으나 막대한 비용증가 등의 문제로 인해 전화 조사가 1970년대 이후부터 최근까지 가장 많이 사용되는 자료수집방법이 되었다 (Chang과 Korsnick, 2009). 우리나라에서는 조사결과의 신속성을 요하는 선거여론조사의 특성상 짧은 시간에 유권자들의 의견을 파악할 수 있다는 점과 저렴한 비용이라는 장점 때문에 낮은 응답률이라는 심각한 결함에도 불구하고 선거여론조사는 대부분 전화조사로 실시된다.

현재 국내에서 실시되는 많은 전화 선거여론조사는 연령별, 성별, 지역별 인구수에 비례하여 표본을 할당할 후, 표집틀로 사용하는 전화번호목록에서 무작위로 번호를 추출하여 할당된 수만큼 조사하는 방식을 취하고 있다. 무응답이 발생하여 초기에 할당된 수만큼 조사가 이루어지지 않은 경우, 응답자들의 가중치를 높여서 추정하는 방법을 사용한다. 이 경우, 모집단의 무응답 집단과 응답 집단 간의 응답이 다르면, 추정량과 추정량의 분산에 편의가 발생한다. 이 편의는 두 집단 간의 응답 차이의 크기와 무응답률의 크기에 비례한다.

2012년 대선 선거여론 조사에서, 조사원인터뷰방식의 응답률이 유무선통합조사는 약 21.7%, 무선전화조사는 약 22.6%였고, ARS조사는 유무선통합조사와 무선조사에서 응답률이 모두 약 9.2%였다. 유선전화조사의 경우, 응답률은 약 16.8%였다. 2012년 대선 선거여론 조사의 평균 응답률은 약 14.4%에 불과하다.

2014년 지방자치단체장 선거여론조사를 실시한 한 무선전화조사에서 연령대별 응답률을 살펴본 결과 20대와 30대의 응답률은 각각 약 2.2%와 4.2%에 불과했다. 20~30대의 낮은 응답률은 정도의 차이는 있지만 거의 모든 전화 여론조사에서 문제점으로 지적되고 있다.

여론조사 무응답률을 줄이기 위해서는 여론조사에 대한 응답자들의 태도를 아는 것이 중요하다. 우리가 자체적으로 실시한 「전화 선거여론조사에 대한 의식조사」에서 우리는 응답자들에게 전화 선거여론조사에 응답하지 않는 가장 큰 이유에 대해 중복응답할 수 있도록 질문하였다. 그 결과, 전체 응답자가 가장 많이 선택한 이유가 “귀찮아서”였고, 여론조사 응답률이 가장 낮은 20대가 가장 많이 선택한 이유도 “귀찮아서”였다.

외국의 경우 인센티브 지급이 응답률을 높였음을 보여주는 많은 연구들이 있다. Tourageau 등 (2010)은 선거참여여부를 주제로 한 전화여론조사에서 \$5의 인센티브를 제공한 집단의 응답률이 인센티브가 없는 집단에 비해 약 20%이상 높았음을 보여준다. 또, Dillman (2000)은 인센티브가 응답률을 높인 많은 사례들을 제시하고 있다.

국내에서 선거여론조사의 정확성에 대한 문제는 반복적으로 지적되고 있다. 젊은 층의 지나치게 낮은 응답률은 선거여론조사의 정확성에 문제가 될 수 있고, 따라서 응답률 자체를 높이려는 보다 근본적인 노력이 필요하다. 조사대상자들의 시간과 노력이라는 비용에 상당하는 대가를 지불하고 응답자들에게 동기를 부여하여 응답률을 높이려는 근본적이고 적극적인 노력이 필요하다. 지금까지 국내에서는 인센티브 제공이 응답률에 미치는 효과에 관한 연구가 활발하게 이루어지지 못했다. 향후 이에 대한 다양한 실험적 연구들이 필요하다.

References

- Carlson, J. (1996). Results of personalization experiments. *Paper presented at the meeting of W-183, Regional Research Committee, Tucson, AZ.*
- Chang, L. and Korsnick, J. A. (2009). National Surveys via RDD telephone interviewing versus the internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly*, **73**, 641-678.
- Cho, S. K. and Cho, E. H. (2010). Application of random sampling procedures in a telephone survey. *Survey Research*, **11**, 141-160.
- Cho, S. K., Kim, J. Y., Na, Y. J. and Lee, M. J. (2007). How to improve the electoral polls? The case of the 2006 local elections. *Survey Research*, **8**, 31-54.
- Cochran, G. W. (1977). *Sampling techniques*, 3rd, John Wiley & Sons, New York.
- Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, **64**, 464-494.
- Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method*, 2nd ed., John Wiley & Sons, New York.
- Johnson, T. and McLaughlin, S. (1990). *GMAT registrant survey design report*, Los Angeles Graduate Admission Council, Los Angeles.
- Korea Gallup (2012). *Data comparing results from election polls to the press (after D-100)*, Gallup Report (<http://www.gallup.co.kr/gallupdb/report.asp>), Seoul.
- Korea Internet & Security Agency (2013). *2013 Survey on the mobile internet usage*, Korea Internet & Security Agency, Seoul.
- Korea Internet & Security Agency (2015). *2014 Survey on the information protection (personal)*, Korea Internet & Security Agency, Seoul.
- Kwak, J. and Choi, B. (2014). A comparison study for accuracy of exit poll based on nonresponse model. *Journal of the Korean Data & Information Science Society*, **25**, 53-64.

- Lee, H. J. and S. B. Kang (2012). Handling the nonresponse in sample survey. *Journal of the Korean Data & Information Science Society*, **23**, 1183-1194.
- Lee, K., Jang, D. H. and Hong, Y. T. (2012). A case study of mixed-mode design incorporated mobile RDD into telephone RDD. *The Korean Journal of Applied Statistics*, **25**, 153-162.
- Heo, S. and Chang, D. (2014). Opportunity cost of survey respondents: Based on 2012 Presidential election poll. *Statistics Korea*, **19**, 147-159.
- Park, H. and Na, S. (2015). Estimation to improve survey efficiency in callback. *Journal of the Korean Data & Information Science Society*, **26**, 377-385.
- Park, M. I. (1978). *Public opinion survey in Gallup*, Korea Gallup, Seoul, 168-172.
- Scheaffer, L. R., Mendenhall, W. and Ott, L. (1990). *Elementary survey sampling*, 4th ed., PWS-KENT Publishing Company, 117, Boston,
- Tourangeau, R., Groves, R. M. and Redline, C. D. (2010). Sensitive topics and reluctant respondents: Demonstrating a link between nonresponse bias and measurement error. *Public Opinion quarterly*, **74**, 413-432.

Bias caused by nonresponses and suggestion for increasing response rate in the telephone survey on election[†]

Sunyeong Heo¹ · Sucheol Yi²

¹Department of Statistics, Changwon National University

²Department of Mathematics, Changwon National University

Received 4 February 2016, revised 4 March 2016, accepted 10 March 2016

Abstract

Thanks to the advantages of low cost and quick results, public opinion polls on election in Korea have been generally conducted by telephone survey, even though it has critical disadvantage of low response rate. In public opinion polls on election in Korea, the general method to handle nonresponses is adjusting the survey weight to estimate parameters. This study first drives mathematical expression of estimator and its bias with variance estimators with/without nonresponses in election polls in Korea. We also investigate the nonresponse rate of telephone survey on 2012 Korea presidential election. The average response rate was barely about 14.4%. In addition, we conducted a survey in April 2014 on the respondents's attitude toward telephone surveys. In the survey, the first reason for which respondents do not answer on public opinion polls on election was "feel bothered". And the aged 20s group, the most low response group, also gave the same answer. We here suggest that survey researchers motivate survey respondents, specially younger group, to participate surveys and find methods boosting response rate such as giving incentive.

Keywords: Bias, incentive, nonresponse, telephone survey, the 18th Korea presidential election survey.

[†] This research is financially supported by Changwon National University in 2015~2016.

¹ Corresponding author: Professor, Department of Statistics, Changwon National University, Changwon 51140, Korea. E-mail: syheo@changwon.ac.kr

² Professor, Department of Mathematics, Changwon National University, Changwon 51140, Korea.