

## 전화조사에서 재통화 규칙준수와 응답자 임의선택의 영향 - R&R 울산 사례의 통계적 재분석 -

허명희<sup>1)</sup> 임여주<sup>2)</sup> 노규형<sup>3)</sup>

### 요약

우리나라 조사업계에서는 전화조사의 방법론으로 성과 나이, 지역에 표본 수를 사전 지정하는 방식의 할당표집(quota sampling)을 주로 쓰고 있다. 이러한 할당표집은 조사비용과 기간의 단축이라는 이점을 갖지만 이론적 타당성이 결여되어 있어 학문적으로는 받아들이기 어렵다. 때문에, 학계에서는 그 동안 수차례 임의표집(random sampling)에 근거한 전화조사를 조사업계에 요구해 왔다. 이에 응하여, (주)리서치 앤 리서치가 2002년 울산시장 선거예측 조사에 임의표집에 의한 전화조사를 실시하였다. 본 사례연구는 이 자료를 심층적으로 재분석하여 임의표집에서의 재통화 및 응답자 임의선택 절차가 자료 질 및 최종 예측치에 주는 영향에 대하여 살펴볼 것이다.

주요용어: 전화조사, 할당표집, 임의표집, 재통화 규칙, 응답자 임의선택, 가중치, 선거예측.

### 1. 들어가며

전화조사(telephone survey) 만큼 조사업계와 학계가 방법론에 있어 불일치를 보이는 분야는 없을 것이다. 조사업계에서의 일반적인 전화조사 방법론은 1) 전화번호부 또는 전화번호 DB를 표집 틀로 하고, 2) 계통추출된 전화번호에 접촉을 시도하여, 3) 성·나이·지역에 할당된 표본 수 이내인 경우 조사를 수행하여 정해진 총 표본 수를 채워나가는 방식이다. 여기서 문제가 되는 부분은 단계 3에서의 할당표집(quota sampling)인데, 이것은 학계에서 요구하는 재통화(call-back) 규칙준수 및 응답자 임의선택(random selection of respondent within households) 등 임의표집(random sampling)의 원칙들과 배치된다.

때문에, 그 동안 관련학계는 조사업계에 임의표집의 채택을 끊임없이 요구해 왔다 (조성겸, 1997; 홍내리·허명희, 2001; 허명희, 2002). 그러나 조사업계는 늘상 “임의표집을 하게 되면 조사비용과 기간이 수배로 증가하므로 조사업계의 비즈니스적 여건상 불가피하게 할당표집을 채택할 수밖에 없다”는 답변을 하여왔다 (이준웅, 2002; 토론 포함). 마침내 학계의 요구에 응하여 (주)리서치 앤 리서치 (R&R)가 2002년 울산 광역단체장(시장) 선거에

1) (136-701) 서울시 성북구 안암동 5가 1, 고려대학교 통계학과, 교수

E-mail : stat420@korea.ac.kr

2) (136-701) 서울시 성북구 안암동 5가 1, 고려대학교 통계학과, 석사과정 졸업

E-mail : rawoctop@naver.com

3) (주)리서치 앤 리서치(R&R) 대표, 심리학 박사

E-mail : kyuno@randr.co.kr

측 조사 (이하 'R&R 울산 사례'로 약칭)에서 임의표집 원칙을 준수한 표본조사를 실시하였고 이에 대한 1차 연구보고를 한 바 있다 (노규형·강홍수·한철수, 2002). 그들은 5회 재통화 규칙을 준수하고 최근생일(last birthday) 방법에 의한 가구내 응답자 선택 절차를 지켜, 총 1,233 대상가구의 485명으로부터 응답을 얻을 수 있었다. 주요한 연구결과를 요약하면, 1) 제1차 통화에서 조사수행 사례는 총 165건으로 응답률이 13.4%(=165/1,233)이었으나 제5차 통화까지 응답률(누적)이 39.3%(=485/1,233)에 도달하였고, 2) 총 전화회수는 3,751회로 유효응답 1건당 통화회수는 평균 7.73회(=3,751/485)였으며, 3) 나이대별 과거 투표율로 가중치를 두어 각 후보의 득표율을 예측한 결과 당선자와 차점자의 실제 득표율에 근소한 차이로 적중하였다는 것이다.

여기서, 결과적 예측 정확도는 차치하고 조사에 투입된 노력을 비교해 보기로 한다. 제1차 통화시 가구원 중 누군가가 집에 있고 조사에 협조한 사례가 총 294 가구(=응답자 채택 조사수행 165가구 + 응답자 부재 조사약속 129가구)로 그 비율이 23.8%(=294/1,233)로 나타났기 때문에, 만약 할당표집이 적용되었다면 유효응답 1건당 평균 4.20회(=1/0.238) 통화가 필요하였을 것이다. 따라서, 임의표집이 할당표집에 비해 2배 가까운 조사비용(조사원 수당)이 요구됨을 알 수 있다. 조사기간에 있어 차이가 있음은 물론이다.

R&R 울산 사례의 재분석 의도는 다음 세 가지이다. 첫째, 어떤 응답자들이 일찍 응답하는 경향이 있고 어떤 응답자들이 늦게 응답하는 경향이 있는가? 즉, 통화회수의 차이에 영향을 주는 인구·사회적 요인을 밝히고자 한다. 이것을 살펴봄으로써, 재통화 없이 1차 통화로 국한되는 경우 조사 자료가 어떤 편향성을 갖는지를 알 수 있을 것이다. 둘째, 지지 후보에 영향을 주는 인구·사회적 요인은 무엇인가? 만약 이 때의 요인이 통화회수와 관련 있는 요인과 겹친다면 느슨한 규칙의 적용으로 조사결과가 어떤 오류를 범할 수 있는가를 이해하는 데 도움이 될 것이다. 마지막으로, 2002년 울산시장선거에서 임의표집의 방식의 조사가 득표율 예측에 미친 결과에 대하여 토의할 것이다.

## 2. 통화회수와 관련 있는 인구·사회적 요인 분석

임의표집 방식의 전화조사에서는 여러 회에 걸친 재통화 시도가 불가피하다. 외국에서는 3회 또는 5회 통화가 표준적이지만 (Voss et al., 1995; Curtin et al., 2000), 실험적 차원에서 그 이상의 통화시도를 하는 경우도 있다. 그렇다면 어떤 인구·사회적 요인이 통화회수에 영향을 주는가? 외국 문헌에 보고된 바에 의하면, 전화조사에서는 일반적으로 젊은 사람일수록 그리고 교육 수준과 소득수준이 높은 사람일수록 접근하는데 많은 통화회수가 필요하다 (Curtin et al., 2000; Keeter et al., 2000). 그리고 미국의 한 선거 예측을 위한 전화조사에서는 통화회수가 증가함에 따라 젊은 사람, 남성, 보수적 성향자의 비율이 커지는 것으로 나타났다 (Traugott, 1987). 이러한 이유는 남성·젊은 사람들의 채택률이 작고, 고학력·고소득자들의 시간 가치가 상대적으로 크기 때문일 것이다.

R&R 울산 사례의 경우 응답자 분포를 통화회수 별로 살펴보면 표 2.1과 같다. 1차·2차·3차·4차·5차 통화 응답자 구성 비율(총 응답자 대비)이 각각 34.0%, 30.1%, 19.6%, 8.2%, 8.0%로 나타났다. 이에 따라 최대 다섯 차례의 통화 시도에서 3차 통화까지 응답자 구성 누적비

표 2.1: 통화회수에 따른 응답자 분포

	응답자	(비율 %)	누적응답자	(누적 %)
제1차	165명	(34.0)	165명	(34.0)
제2차	146명	(30.1)	311명	(64.1)
제3차	171명	(19.6)	406명	(83.7)
제4차	78명	( 8.2)	446명	(92.0)
제5차	39명	( 8.0)	485명	(100.0)
합계	485명	(100.0)		

을(총 응답자 대비)이 83.7%에 도달한 것을 볼 수 있다.

이제, 면접에 성공하기까지의 통화회수  $J$ 에 대한 통계적 모형을 생각하기로 하겠다. 설명변수로는 응답자의 성, 나이, 거주지역(구), 직업, 교육수준, 소득, 출신지 등 7개 인구·사회적 속성이다. 본 연구에서 설정한 모형은 누적 로짓 모형(cumulative logit model)으로 다음과 같다.

$$\log \frac{P(J \leq j | x_1, \dots, x_p)}{P(J > j | x_1, \dots, x_p)} = \alpha_j + \beta_1 x_1 + \dots + \beta_p x_p, \quad j = 1, \dots, 4$$

여기서,  $P(J \leq j | x_1, \dots, x_p)$ 는 더미 범주  $x_1, \dots, x_p$ 에 해당하는 응답자의 통화회수  $J$ 가  $j (= 1, \dots, 4)$  이하일 확률이고,  $\alpha_j$ 는 종속변수의 누적 범주  $j$ 에 해당하는 기저수준이고 ( $\alpha_1 \leq \dots \leq \alpha_4$ ),  $\beta_k$ 는 더미 변수  $x_k$ 의 모형계수이다 ( $k = 1, \dots, p$ ). [범주형 설명변수의 더미화에 있어 가장 빈도가 큰 범주를 기준으로 하였다. 따라서 최빈 범주의 모형 계수는 0으로 고정된다.] 표 2.2가 이 모형을 485명의 응답자 자료에 적합한 결과이다.

설명변수들에 대한 Wald 통계량과 유의확률을 보면, ‘직업’이 통화회수에 유의한 영향을 주고 있음을 볼 수 있다. 가정주부와 비교하여, 화이트칼라가 접근하기 어렵고 그 다음으로 블루칼라가 어려우며 반면 무직/기타가 접근이 용이한 것으로 나타났다. 직업 이외에 다른 요인들은 통계적으로 유의하게 나타나지 않았다.

상식적으로 남성일수록 그리고 나이가 젊을수록 재택률이 작으므로, 성과 나이가 통화회수에 영향을 주는 중요한 변수일 것으로 기대할 수 있으나 여기서는 그렇게 나타나지 않았다. 이것은 5차에 걸친 시도에도 불구하고 남성과 20대 나이층이 결국 접촉되지 않았기 때문인 것으로 생각된다. 통화회수별 누적응답자 표본의 인구·사회적 특성 분포를 표 2.3에 정리하였다. 모집단에서 남성 비율은 49.6%이나 최종 응답자 표본에서 남성 비율은 37.3%에 불과하다. 또한 20대, 30대, 40대, 50대 이상 비율이 모집단에서 23.8%, 31.5%, 24.1%, 20.6%이나 최종 응답자 표본에서는 각각 13.2%, 28.9%, 25.4%, 32.6%이다. 뿐만 아니라, 통화회수가 1차에서 5차로 증가하면서도 이들 비율들은 거의 변화하지 않았다.

표 2.2: 통화회수에 대한 누적로짓모형의 적합결과

변수	빈도	계수	Wald	유의확률	오즈비
절편1		-0.62	-	.001*	-
절편2		0.68	-	<.001*	-
절편3		1.77	-	<.001*	-
절편4		2.59	-	<.001*	-
성별	-	-	1.44	.230	-
- 여자	304	-	-	-	-
- 남자	181	0.14	-	.232	1.32
나이대	-	-	1.25	.741	-
- 50대이상	158	-	-	-	-
- 20대	64	-0.16	-	.461	0.79
- 30대	140	0.13	-	.401	1.05
- 40대	123	-0.05	-	.737	0.88
거주지	-	-	5.42	.247	-
- 남구	156	-	-	-	-
- 중구	89	-0.09	-	.607	1.23
- 동구	100	0.21	-	.229	1.66
- 북구	53	0.16	-	.453	1.59
- 을주군	87	0.02	-	.933	1.37
직업	-	-	15.66	.016*	-
- 가정주부	200	-	-	-	-
- 농/임/어업	17	0.13	-	.762	0.93
- 자영업	63	-0.11	-	.671	0.74
- 블루칼라	51	-0.36	-	.167	0.57
- 화이트칼라	79	-0.67	-	.004*	0.42
- 학생	10	0.25	-	.667	1.05
- 무직/기타	65	0.55	-	.026*	1.43
학력	-	-	2.57	.277	-
- 고졸	218	-	-	-	-
- 중졸이하	155	-0.25	-	.142	0.77
- 대재이상	112	0.25	-	.131	1.27

표 2.2: 통화회수에 대한 누적로짓모형의 적합결과 (계속)

변수	빈도	계수	Wald	유의확률	오즈비
소득	-	-	0.86	.829	-
- 100-199만원	137	-	-	-	-
- 100만원미만	116	-0.10	-	.574	0.81
- 200-299만원	131	-0.05	-	.136	0.85
- 300만원이상	101	0.05	-	.080	0.94
출신지	-	-	1.79	.970	-
- 부산/울산/경남	243	-	-	-	-
- 서울/인천/경기	22	0.12	-	.750	1.13
- 대구/경북	125	0.00	-	.987	1.00
- 광주/전남	25	0.06	-	.863	1.06
- 대전/충청	27	-0.13	-	.699	0.88
- 강원	21	-0.02	-	.956	0.98
- 제주/이북	15	-0.50	-	.250	0.61
- 기타	7	0.46	-	.464	1.59

표 2.3: 통화회수별 누적 응답자 표본의 인구·사회적 특성 분포(%)

	모집단	통화회수				
		1차	2차	3차	4차	5차
성별						
- 여자	49.6	63.0	61.7	62.1	62.1	62.7
- 남자	50.4	37.0	38.3	37.9	37.9	37.3
나이대						
- 20대	23.8	13.3	12.5	13.3	13.7	13.2
- 30대	31.5	27.9	30.2	30.0	29.4	28.9
- 40대	24.1	23.0	23.5	25.4	24.9	25.4
- 50대이상	20.6	35.8	33.8	31.3	32.1	32.6
거주지						
- 중구	23.0	17.6	19.3	18.5	18.4	18.4
- 남구	31.9	25.5	30.2	30.5	31.2	32.2
- 동구	17.7	25.5	20.9	20.7	20.9	20.6
- 북구	11.0	12.1	11.6	11.3	11.2	10.9
- 울주군	16.4	19.4	18.0	19.0	18.4	17.9

표 2.3: 통화회수별 누적 응답자 표본의 인구·사회적 특성 분포(%) (계속)

	모집단	통화회수				
		1차	2차	3차	4차	5차
직업						
- 농/임/어업		4.8	2.9	3.2	3.6	3.5
- 자영업		12.1	13.8	13.3	13.2	13.0
- 블루칼라		9.1	9.0	9.9	10.5	10.5
- 화이트칼라		10.3	12.9	16.0	15.9	16.3
- 가정주부		44.2	42.8	40.9	40.8	41.2
- 학생		2.4	2.6	2.5	2.2	2.1
- 무직/기타		17.0	16.1	14.3	13.7	13.4
학력						
- 중졸 이하		32.7	28.3	28.6	29.6	29.9
- 고졸		41.8	45.7	46.3	45.7	44.9
- 대재이상		23.6	24.4	23.2	22.9	23.1
- 모름/무응답		1.8	1.6	2.0	1.8	2.1
소득						
- 100만원미만		21.8	19.6	19.5	19.7	20.4
- 100-199만원		30.3	30.5	27.3	28.0	28.2
- 200-299만원		26.1	26.7	27.1	27.4	27.0
- 300만원이상		19.4	19.0	21.9	21.1	20.8
- 모름/무응답		2.4	4.2	4.2	3.8	3.5
출신지						
- 서울/인천/경기		4.2	5.1	4.7	4.9	4.5
- 부산/울산/경남		51.5	49.5	49.5	49.3	50.1
- 대구/경북		23.6	25.7	26.8	26.5	25.8
- 광주/전라		5.5	4.8	5.4	5.4	5.2
- 대전/충청		6.1	5.5	5.2	5.4	5.6
- 강원		6.1	4.5	3.7	3.8	4.3
- 제주/이북		1.8	2.9	3.0	3.1	3.1
- 기타		1.2	1.9	1.7	1.6	1.4
누적 응답자 수		165	311	406	446	485

따라서, 조사현업에서 관행대로 지역별로 성, 나이 등에 표본 수를 할당한 1차 조사만을 수행하게 되면 직업 분포에 있어 주된 편향이 나타날 것으로 보인다. 표 2.3에서 통화회수 별 누적응답자 표본의 직업 분포를 보라. 1차에서 5차로 가면서 가정주부의 비율이 3.0%P 감소하였고 무직/기타가 3.6%P 감소한 반면, 화이트칼라의 비율은 6.0%P 증가한 것으로 나타났다.

### 3. 지지후보 예측 모형

R&R 울산 사례에서는 광역단체장 선거예측이 목적이므로, 이 절에서는 지지후보 예측 모형을 다루기로 하겠다. 2002년 울산시장 선거에서 후보자는 박맹우(한나라당), 송철호(민주노동당), 안승천(사회당) 등이었으나 여기서는 실제로 경합이 된 박맹우(1) 후보와 송철호(2) 후보만 생각하기로 한다. 따라서 지지후보를 묻는 원질문 또는 1차 캐어묻기 질문에 서 양 후보 중 1인을 지지한다고 응답한 317명 자료가 모형 구축에 사용될 것이다.

종속변수는 지지후보  $Y$  (=1 송철호, 0 박맹우)이고 이에 대한 설명변수로는 2절에서 고려한 성과 나이, 거주지역(구), 직업, 교육수준, 소득수준, 출신지를 고려하였다. 그 외에 실사기록 변수로서 통화회수와 응답 시간대를 고려하여 추가적인 설명이 가능한가를 살펴보았다. 여기서 고려할 모형은 로지스틱 회귀모형이다. 즉,

$$\log \frac{P(Y = 1|x_1, \dots, x_p)}{P(Y = 0|x_1, \dots, x_p)} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

를 고려할 것이다. 여기서  $x_1, \dots, x_p$ 는 설명 범주를 나타내는 더미(dummy) 변수이고,  $\beta_0$ 는 절편항,  $\beta_k$ 는 범주 더미  $x_k$ 에 대한 모형 계수이다 ( $k = 1, \dots, p$ ). [범주형 설명변수의 더미화에 있어 가장 빈도가 큰 범주를 기준범주로 하였다. 기준 범주에 대한 모형 계수는 0으로 고정된다.]

표 3.1이 2항 반응 로지스틱 회귀 모형을 적합한 결과이다 (송철호 비교, 박맹우 기준). 통계적으로 유의한 설명변수는 나이 및 일부 직업(화이트칼라 대 가정주부)으로 나타났다. 즉 화이트칼라가 가정주부에 비하여 송철호를 지지하는 경향이 뚜렷하였다. 통화회수의 영향 패턴은 통계적으로 유의하지는 않지만 1차에 비해 2차, 3차, 5차 등에서 송철호 지지율이 떨어지는 것을 볼 수 있다. 즉, 늦게 접촉되는 응답자일수록 박맹우 지지자일 가능성이 높음을 뜻한다.

표 3.1: 지지후보에 대한 로지스틱 모형의 적합 결과 (송철호 비교, 박맹우 기준)

변수	빈도	계수	Wald	유의확률	오즈비
상수		-1.79	6.99	.008*	0.17
성별					
- 여자	190	-	-	-	-
- 남자	127	.36	.86	.354	1.43
나이대		-	23.07	<.001*	-
- 50대이상	93	-	-	-	-
- 20대	38	1.53	7.17	.007*	4.62
- 30대	93	2.29	22.92	<.001*	9.91
- 40대	93	1.41	10.16	.001*	4.09
거주지		-	4.49	.344	-
- 남구	98	-	-	-	-
- 중구	61	.56	1.89	.170	1.75
- 동구	70	.67	3.14	.076	1.95
- 북구	38	.77	2.68	.101	2.16
- 울주군	50	.37	.72	.397	1.44
직업		-	9.47	.149	-
- 가정주부	129	-	-	-	-
- 농/임/어업	11	.15	.03	.859	1.17
- 자영업	39	-.03	.00	.953	.97
- 블루칼라	36	.05	.01	.923	1.05
- 화이트칼라	57	1.15	5.70	.017*	3.17
- 학생	6	-.56	.25	.621	.57
- 무직/기타	39	.01	.00	.988	1.01
학력		-	.44	.804	-
- 고졸	148	-	-	-	-
- 중졸이하	90	.13	.10	.749	1.14
- 대재이상	79	-.19	.28	.599	.82
소득		-	1.19	.756	-
- 100-199만원	99	-	-	-	-
- 100만원미만	59	-.25	.31	.577	.78
- 200-299만원	93	-.02	.00	.965	.99
- 300만원이상	66	-.37	.84	.359	.69



표 3.1: 지지후보에 대한 로지스틱 모형의 적합 결과 (송철호 비교, 박맹우 기준) (계속)

변수	빈도	계수	Wald	유의확률	오즈비
출신지		-	7.17	.305	-
- 부산/울산/경남	164	-	-	-	-
- 서울/인천/경기	13	.03	.00	.960	1.04
- 대구/경북	88	.11	.11	.741	1.11
- 광주/전라	16	1.10	3.00	.083	3.00
- 대전/충청	15	-.47	.55	.458	.63
- 강원	11	.68	.82	.366	1.98
- 제주/이북	10	-1.35	2.26	.133	.26
통화회수		-	3.57	.467	-
- 1차	113	-	-	-	-
- 2차	91	-.45	1.27	.260	.64
- 3차	67	-.45	.97	.325	.64
- 4차	25	.18	.11	.741	1.20
- 5차	21	-.86	1.95	.163	.42
조사 시간대		-	.98	.613	-
- 18~21시	63	-	-	-	-
- 11~13시	122	-.11	.07	.797	.90
- 13~18시	132	-.36	.90	.344	.70

#### 4. 득표율 예측

아무리 최선을 다한 임의표집 조사라고 하더라도 응답자 표본과 모집단 사이에는 크고 작은 괴리가 불가피하게 발생한다. 이를 가중치로 보정하는 것은 합당한 예측을 위한 필수적 절차인데, 선거 예측의 경우 투표의향 또는 투표 가능성의 처리가 덧붙여지게 된다. 본 사례연구에서는 2단계에 걸쳐 가중치 부여 작업을 하였다. 제 1단계에서는 지지후보 응답자(=지지후보를 묻는 원질문과 한 차례의 캐어문기 질문에 지지후보를 밝힌 응답자) 표본과 모집단의 지역(구), 성, 나이대 분포가 일치되도록 가중치를 적용하였다. 제 2단계에서는 성, 나이대별 과거 투표율(1998년 지방선거 투표율)을 가중치로 하였다. 표 4.1이 그 결과이다. 당선자(박맹우)에 대하여 예측값이 51.6%, 실제 득표율이 53.1%로 -1.5%P (=51.6-53.1)의 오차가 발생하였다. [노규형, 강홍수, 한철수 (2002)는 선행연구에서 나이대별 투표율만을 가중치로 하였는데 그 결과 당선자(박맹우)에 대한 예측득표율이 55.3%로 2.2%P (=55.3-53.1)의 오차를 기록하였다.]

표 4.1은 5차례 통화 절차에 따라 얻어진 표본자료로부터 산출된 것이다. 만약 5차례가 아니라 4, 3, 2, 1차례로 통화를 제한하였었다면 어떤 예측값과 오차가 나타났을까? 원 자료에서 지정 통화회수를 초과하는 케이스들을 없애고 남은 자료에 동일한 가중치 절차를

표 4.1: 울산 시장 후보자에 대한 지지율 예측값: 통화회수 = 5

후보자	가중치 적용 예측값	실제 투표 득표율	오차
박맹우 (한나라)	51.6%	53.1%	-1.5%P
송철호 (민주노동당)	45.4%	43.6%	1.8%P
1,2위 후보간 차이	6.2%P	9.5%P	-3.3%P

표 4.2: 각기 다른 통화회수 적용에 따른 울산 시장 후보자에 대한 지지율 예측값

통화회수 = 4

후보자	가중치 적용 예측값	실제 투표 득표율	오차
박맹우 (한나라)	51.3%	53.1%	-1.8%P
송철호 (민주노동당)	45.5%	43.6%	1.9%P
1,2위 후보간 차이	5.8%P	9.5%P	-3.7%P

통화회수 = 3

후보자	가중치 적용 예측값	실제 투표 득표율	오차
박맹우 (한나라)	52.0%	53.1%	-1.1%P
송철호 (민주노동당)	44.8%	43.6%	1.2%P
1,2위 후보간 차이	7.2%P	9.5%P	-2.3%P

통화회수 = 2

후보자	가중치 적용 예측값	실제 투표 득표율	오차
박맹우 (한나라)	52.6%	53.1%	-0.5%P
송철호 (민주노동당)	44.4%	43.6%	0.8%P
1,2위 후보간 차이	8.2%P	9.5%P	-1.3%P

표 4.2: 각기 다른 통화회수 적용에 따른 울산 시장 후보자에 대한 지지율 예측값(계속)

통화회수 = 1

후보자	가중치 적용 예측값	실제 투표 득표율	오차
박맹우 (한나라)	51.8%	53.1%	-1.3%P
송철호 (민주노동당)	46.0%	43.6%	2.4%P
1,2위 후보간 차이	5.8%P	9.5%P	-3.7%P

적용하여 득표율 예측값들을 산출하였다. 표 4.2가 그 결과이다.

연구자의 기대와는 달리, 통화회수에 따른 차이가 득표율 예측값에 별 차이를 주지 않았던 것을 볼 수 있다. 다시 말하여, 통화회수를 1, 2, 3, 4, 5중 어느 것으로 하였더라도 유사한 예측값을 만들게 되었을 것이라는 것을 의미한다. 왜 이런 결과가 나타났을까?

2절에서 통화회수가 늘어남에 따라 화이트칼라가 증가하는 경향을 보았는데, 3절에서는 화이트칼라가 송철호 지지율을 높이고 통화회수가 박맹우 지지율을 높이는 효과가 있음을 본 바 있다. 따라서 2개 요인효과가 상쇄되는 결과가 된다. 때문에, 이 사례에서는 득표율 예측이 통화회수에 별 영향을 받지 않은 것으로 나타나지 않았을까? 그러나 2개 요인이 같은 방향으로 작용하는 상황에서는 통화회수가 득표율 예측에 미치는 영향이 증폭될 것이다.

### 5. 맺음 말

이 사례에서는 5회 통화 규칙을 적용하였다. 그러나, 표 2.3에서 3차 통화이후 응답자 표본의 인구·사회적 속성 분포가 5차 통화이후 것과 별 차이가 없음을 볼 수 있다. 따라서 3회 통화 규칙으로써 5회 통화 규칙과 견줄만한 질의 조사자료 생산이 가능할 것으로 보인다. 한편, 이 사례에서는 표본가구내 응답자 선택을 위하여 최근 생일(last birthday) 방법을 채택하였으나 우리의 정서가 생일을 묻고 답하는 데 익숙하지 않아서 조사 진행이 쉽지 않았으므로, 최근 생일 방법보다는 조성겸 (1991)이 제안한 바 있는 CNU 선정표, 또는 그것의 원형인 Kish 표를 활용하면 조사 진행이 원활할 것으로 생각된다.

전화조사에서 재통화 규칙 준수 및 가구내 응답자 선택 등 임의표집을 채택하게 되면 조사 비용이 증가하고 조사기간이 늘어나지만, 조사업계에서 흔히 생각하듯이 4-5배의 부담이 되는 것은 아니다. 이 시험적 사례에서 드러났듯이, 5회 통화 기준으로 부담은 2배 정도로 추정된다. 따라서 조사업계가 의지를 갖고 조사용역 의뢰자와 협의를 한다면 임의표집에 의한 전화조사가 충분히 가능하리라고 본다.

## 참고문헌

- [1] 노규형, 강홍수, 한철수 (2002). “임의표집법에 의거한 전화조사의 시도: 2002년 울산시장선거의 경우,” 「조사연구」 3권 2호. 77-90.
- [2] 이준웅 (2002). “선거여론조사의 방법 및 응용 (토론: 안부근, 강미은, 이동관),” 「관훈저널」 2002년 여름호. 45-82.
- [3] 조성겸 (1991). “무선확률에서 가구별 응답자의 무선확률표집에 관한 연구: CNU 선정표의 타당성과 실용성의 검토,” 「신문학보」 26호. 189-214.
- [4] 조성겸 (1997). “대통령 선거 여론조사와 할당표집 방법의 문제점,” 「언론과 사회」 18호. 29-53.
- [5] 허명희 (2002). “선거여론조사의 방법론 소고,” 「한국분류학회지」 5권. 109-119.
- [6] 홍내리, 허명희(2001). “제 16대 국회의원 선거의 예측조사에 대한 사후적 검증,” 「조사연구」 2권 1호. 1-35.
- [7] Curtin, R., Presser, S. and Singer, E. (2000). “The effects of response rate changes on the index of consumer sentiment,” *Public Opinion Quarterly*, 64. 413-428.
- [8] Keeter, S., Miller, C., Kohut, A., Groves, R.M. and Press, S. (2000). “Consequences of reducing nonresponse in a national telephone survey,” *Public Opinion Quarterly*, 64. 125-148.
- [9] Traugott, M.W. (1987). “The importance of persistence in respondent selection for pre-election surveys,” *Public Opinion Quarterly*, 51. 48-57.
- [10] Voss, D., Gelman, A. and King, G. (1995). “Pre-election survey methodology: Details from eight polling organizations, 1988 and 1992,” *Public Opinion Quarterly*, 59. 98-132.

[ 2003년 5월 접수, 2003년 6월 채택 ]

## Effects of Call-back Rules and Random Selection of Respondents: Statistical Re-analysis of R&R's Ulsan Survey Data.

Myung-Hoe Huh <sup>1)</sup> Yo-Joo Lim <sup>2)</sup> Kyu-Hyung No <sup>3)</sup>

### ABSTRACT

In Korea, quota sampling is mainly adopted in telephone surveys, instead of random sampling which requires call-back procedure and random selection of respondent within households. The contact mode based on the sex\*age quotas is economically more advantageous and less time-consuming. However, it lacks theoretical ground for valid statistical inference, so that it is hardly accepted in academic circles despite of widely spread practice. Subsequently, survey theoreticians argued that random sampling-based telephone surveys should be tried. In response, Research & Research (R&R), a private research company in Seoul, executed a telephone survey by random sampling mode for the prediction of 2002 Ulsan City Mayor Election.

The aim of this case study is to find out various effects of the call-back rule with random selection of respondents by statistically re-analyzing R&R's Ulsan Survey Data.

*Keywords:* Telephone survey; quota sampling; random sampling; call-back rules; random selection of respondents; weighting; election prediction.

---

1) Professor of Statistics, Korea University. Anam-Dong 5-1, Seoul 136-701, Korea.

E-mail : stat420@korea.ac.kr

2) Graduated with Master's degree in Statistics, Korea University.

E-mail : rawoctop@naver.com

3) Ph. D. in Psychology, President of Research & Research.

E-mail : kyuno@randr.co.kr