

연구논문

사회조사에서 조사방법에 따른 가중 칸 설정에 관한 연구:

전화조사, 인터넷 조사, 모바일 조사

A Study on Weighting Cells by Survey Methods for Social Surveys:
Telephone, Internet and Mobile Surveys

허명희* · 강용수** · 손은진***

Myung-Hoe Huh · Yong-Soo Kang · Eun-Jin Sohn

본 연구의 목적은 전화조사, 인터넷 조사와 모바일 조사에서 표본의 대표성을 제고하기 위해 가중 칸 구성을 어떻게 할 것인가를 탐구하는 데 있다. 이를 위하여 2000년 인구주택 총조사 2% 원자료(285,514가구, 20세 이상 624,034명)를 분석하였는데, 그 결과 1인 가구와 2인 이상 가구 사이에는 전화보유율에 있어 상당한 차이가 난다는 사실을 확인하였다. 또한 인터넷 활용과 모바일 보유에 영향을 주는 주요 인구사회적 요인을 C5.0 나무형 모형을 통하여 찾고자 하였다. 주요결론은, 1) 전화조사에서 1인 가구와 2인+가구의 표본 분포가 모집단 분포와 일치하도록 가구에 가중치를 부여함으로써 표본의 대표성을 제고할 필요가 있다는 점, 2) 인터넷 조사에서는 성·나이대를 고려하는 외에 학력과 직업(또는 주택유형, 자동차 보유여부) 등에 의한 가중 칸(weighting cell) 설정 또는 크기비례 할당이 필요하다는 점, 3) 모바일 조사에서도 성·나이대를 고려하는 외에 학력과 직업(또는 자동차 보유여부, 혼인상태) 등에 의한 가중 칸 설정 또는 크기비례 할당이 필요하다는 점 등이다.

주제어: 전화조사, 인터넷 조사, 모바일 조사, 가중 칸, 크기비례 할당, C5.0 나무 모형

The aim of this study lies in answering the question "How to form weighting cells to enhance sample representativeness in telephone, Internet and mobile surveys?". For this, we explored 2% raw data of Year 2000 Population and Housing Census of Korea looking for meaningful patterns for ownership of

* 교신저자(corresponding author): 고려대학교 통계학과 교수 허명희 E-mail: stat420@korea.ac.kr

** 고려대학교 대학원 통계학과 석사과정.

*** 고려대학교 대학원 통계학과 석사과정.

telephones, the usage of Internet and/or mobile phones. We found that telephone coverage rates vary significantly by household size; 84.6% for one member households, contrasting 98.5% for two-or-more member households. Thus, telephone survey samples need to be weighted differently in sub-groups by household size for proportional representation of target population. Searching socio-demographic factors influencing the use of Internet by C5.0 tree models, we found that education levels and the occupation (or housing type, the automobile ownership) are two most important factors in addition to gender and age. Thus, surveyor might form weighting cells by such factors at the stage of post-stratification or set quotas, a priori, proportional to size of the cells by such factors. For mobile surveys, we approached similarly and found that education levels and the occupation (or the automobile ownership, marriage status) are two additional factors that may be used in forming weighting cells or in setting quotas for cells.

Key words: telephone survey, Internet survey, mobile survey, weighting cell, quota sampling, C5.0 tree model.

I. 서론

사회조사가 성공하기 위해서는 표본의 대표성 확보가 필수적이다. 때문에 조사자는 확률적 표본추출의 원칙을 준수하려고 최대한 노력한다. 그러나 모든 응답대상자를 포함하는 것은 불가능에 가깝고 접촉이 된 응답대상자 중 일부는 조사 자체를 거절하는 것이 현실이다. 따라서 표본추출 오차(sampling error)뿐 아니라 비포함 오차(noncoverage error) 및 무응답 오차(nonresponse error)의 결과로, 조사표본은 인구 사회적 특성 측면에서 목표집단으로부터 다소의 편차를 보일 수 있다. 이를 보정하기 위하여 가중법(weighting)이 적용되는데, 가중법으로는 가중 칸(weighting cell, post-stratification) 방법과 반복비례 가중법(iterative proportional weighting, rim weighting, raking)이 대표적이다(Bethlehem 2002, 박진우 2000). '가중 칸'은 상이한 가중치로 보정될 정해진 인구사회적 요인들의 조합을 일컫는다.

우리나라에 인구주택 총조사는 5년 주기로 시행되고 있으며, 주관

부서인 통계청은 2% 원자료를 공개하고 있다. 이 자료는 총가구의 2%에 해당하는 가구들에 대한 모든 조사내용을 담고 있어 통계청에서 공식 발표하는 결과 보고서에서 찾아볼 수 없는 각종 통계적 정보를 사용자가 직접 만날 수 있는 기회를 제공한다(인구주택 총조사의 2% 원자료에 대한 상세한 설명과 구입 및 활용에 대하여는 통계청 홈페이지 <http://www.nso.go.kr>을 참조할 것). 본 연구에서 우리는 각종 사회조사에서 가중 칸 구성시 어떤 변수를 고려하는 것이 합당한가에 관하여 관심을 갖고 2000년 총조사의 2% 원자료에 접근할 것이다. 구체적인 자료탐사의 목적은 전화조사, 인터넷 조사, 모바일 조사에서 표본의 대표성 확보를 위해 필요한 가중 칸 구성 변수를 찾는 데 있다.

첫 번째 연구문제는 전화조사에서 가중 칸 설정에 관한 것으로, 1인 가구의 사회적, 경제적 특성을 탐구하는 과정에서 도출되었다.

1인 가구의 비율이 2000년 인구주택 총조사에서 15.5%로 나타났는데, 1980년, 1985년, 1990년, 1995년 총조사에서 각각 4.8%, 6.9%, 9.0%, 12.7%였음을 참조할 때 꾸준한 증가세에 있다. 미국은 1960년대 이후 30년 간, 일본도 1970년대 이후 30년 간에 걸쳐 1인 가구 비율이 10%대에서 출발하여 25% 수준에 도달하였고, 이제는 증가세가 멈춘 상태이다. 그러므로 우리나라에서도 향후 20년 정도에 걸쳐 1인 가구의 비율이 더욱 증가할 것으로 예상할 수 있다. 따라서 1인 가구의 인구사회적·경제적 특성을 탐구할 필요가 있는데, 이에 관하여는 1990년 총조사 자료를 분석한 권태환 외 2인(1995), 1995년 총조사 자료를 분석한 권태환 외 5인(1997), 2000년 총조사 자료를 분석한 김두섭 외 2인(2002)의 연구가 있다. 이 외에도 변화순 외 4인(2000)은 독자적인 표본조사를 통하여 노인 1인 가구에 대하여 연구한 바 있다.

우리 연구자들은 2000년 총조사의 2% 원자료를 탐색한 결과 1인 가구의 전화보유율이 2인 이상(이하 '2인+'로 표기) 가구의 전화보유율에 비해 뚜렷이 낮다는 사실을 발견하였다. 이것은 전화조사의 경우 1인

4 조사연구

가구가 과소 표집될 가능성이 있음을 시사하기 때문에 조사표본의 대표성 측면에서 이 문제를 탐구하여 가중 칸 구성시 가구크기를 고려하여야 한다는 결론을 얻게 되었다. 2절에서 이에 대하여 기술할 것이다.

두 번째 연구문제는 인터넷 조사에서 가중 칸 설정에 관한 것이다. 인터넷은 최근 10년 동안 강력한 커뮤니케이션 도구로 발전하였으며 우리나라는 인터넷 활용에 있어 세계적 추세를 선도하는 위치에 있다. 이에 따라 인터넷 조사도 매우 유력한 사회조사 방법으로 등장하였고(김종훈 외 1인 2002, Couper 2000) e-마케팅 조사 등 일부 분야에서는 이미 정착단계에 들어서 있다. 그러나 인터넷 사용자 특성이 일반인 특성과는 상당히 다르고(이두희 외 2인 1998, 박종무 외 2인 2001) 인터넷 사용자 비율이 전화 보급률 수준에는 못 미치고 있어 일반 모집단에 대한 추론은 어려운 상태이다. 본 연구에서는 성·나이대 외에 인터넷 사용과 관련된 강력한 사회적·경제적 요인을 2000년 인구주택총조사의 2% 원자료에 대한 나무모형 분석을 통해 찾아내고자 한다. 선행연구들과는 분석자료의 크기 및 신뢰성에서 현격하게 다르다는 점에서 1차적 의미가 있지만 기법 측면에서도 중요한 의미가 있다고 본다. 생성된 나무모형을 사후 가중치 산출 또는 사전적 크기비례 할당에 활용함으로써 인터넷 조사표본의 대표성을 높일 수 있기 때문이다. 3절에서 이에 대하여 탐구하였다.

세 번째 연구문제는 모바일 조사에서 가중 칸 설정에 관한 것이다. 일반(유선) 전화조사가 특정 성·나이대 인구 그룹의 낮은 재택률 때문에 중대한 위협을 받는 반면, 휴대폰은 일반 전화와는 달리 시간대에 제한 없이 응답자 접촉이 가능하다는 점이 장점이다. 그러나 아직 휴대폰의 보유율이 충분히 크지 않아서 조사표본의 대표성이 의문시되고 있다. 4절에서는 나무모형을 통해 휴대폰 보유와 관련된 사회적·경제적 요인을 찾아 가중 칸 설정 또는 사전적 크기비례 할당에 활용할 것을 제안할 것이다.

II. 전화조사에서의 가중 칸 설정

총조사의 2% 원자료에 포함된 285,514 가구를 전화(유선)의 유무에 따라 분류한 결과 전화를 보유한 가구가 전체의 96.3%인 274,901 가구로 나타났고, 전화를 보유하지 않은 가구가 전체의 3.7%인 10,613 가구였다. 이 정도의 전화보유율은 전화조사의 타당성을 확보하는 데 있어 충분한 수치처럼 보인다.¹⁾ 그러나 전체 가구를 2인+가구 층과 1인 가구 층으로 나눈 뒤 전화보유율을 구해 본 결과, 2인+가구 층(240,666 가구)에서 전화보유율이 98.5%인 반면 1인 가구 층(44,848 가구)에서의 전화보유율은 84.6%에 불과한 것으로 나타났다. 이것은 1인 가구가 전화조사에서 과소 표집될 가능성이 있음을 시사한다.²⁾ 앞에서 말한 바, 2000년 현재 우리나라 전체가구 중 1인 가구는 15.5%로 상당한 비중을 차지하고 있다.³⁾

이에 따라 1인 가구의 사회적·경제적 특성을 살펴볼 필요가 있겠다. 만약 2인+ 가구에 대비하여 1인 가구가 뚜렷한 차이를 보인다면, 전화조사로 얻은 사회조사 표본을 가구원수를 기준으로 가중치 보정을 해야 할 필요가 있을 것이다.

〈표 2.1〉은 인구 기준 성·나이대 그룹에서 소속 가구크기의 구성 비율을 본 것이다. 남자의 경우, 1인 가구 비율이 20대 그룹에서 8.9%, 30대에서 6.3%로 다른 나이대 그룹에 비해 높은 편이므로 1인 가구주는 주로 젊은 나이의 미혼자로 생각된다. 여자의 경우는, 1인 가

-
- 1) 2003년 현재 우리나라의 거의 모든 전화조사에서는 전화번호부/전화번호DB를 추출물로 하므로, 비등재, 갱신 등의 문제로 실제 포함률(coverage)은 훨씬 작다. 업계에 의하면 대략 70% 수준이라고 한다.
 - 2) 더욱이 가구 내 최소 1인 재택률에 있어 1인 가구가 상대적으로 낮을 것이기 때문에 재통화 규칙을 철저하게 준수하지 않는다면 1인 가구는 더욱 과소 표집되는 결과가 될 것이다.
 - 3) 2000년 총조사의 2% 원자료에서는 1인 가구의 비율이 15.7% (=44,848/285,514)로 나왔다.

6 조사연구

〈표 2.1〉 성·나이대별 인구그룹에서의 2인+ 가구 비율 및 1인 가구 비율

성 나이대	남 자		여 자	
	2인+ 가구	1인 가구	2인+ 가구	1인 가구
20대	91.1%	8.9%	93.7%	6.3%
30대	93.7%	6.3%	96.4%	3.6%
40대	95.7%	4.3%	95.8%	4.2%
50대	95.6%	4.4%	92.4%	7.6%
60대+	94.6%	5.4%	79.5%	20.5%

* 2% 원자료 분석 결과

〈표 2.2〉 성·나이대별 2인+ 가구 가구원과 1인 가구 가구원의 동·읍·면부 거주 비율

		2인+ 가구			1인 가구		
		동부	읍부	면부	동부	읍부	면부
성	남자	78.3%	8.2%	13.5%	79.9%	7.5%	12.6%
	여자	79.1%	8.0%	13.0%	66.9%	9.0%	24.1%
나이대	20대	84.5%	6.9%	8.6%	91.0%	4.7%	4.3%
	30대	82.6%	8.4%	9.0%	87.7%	5.8%	6.4%
	40대	82.0%	7.4%	10.6%	80.1%	8.3%	11.6%
	50대	75.4%	8.1%	16.6%	68.0%	9.8%	22.2%
	60대+	62.0%	10.2%	27.9%	49.2%	11.9%	38.9%

* 2% 원자료 분석 결과

구 비율이 50대 그룹에서 7.6%, 60대+ 그룹에서 20.5%로, 다른 나이대 그룹에 비해 높은 편으로 나타나 1인 가구주는 주로 나이 많은 사별자임을 짐작할 수 있다. 이와 같이 1인 가구 집단은 인구적 측면에서 2개의 이질적 그룹이 혼합되어 있다.

〈표 2.2〉는 성별·나이대별 동·읍·면부 비율을 2인+ 가구 가구원과 1인 가구 가구원을 대비시켜 본 것이다. 여자의 경우 2인+ 가구

가구원들의 13.0%가 면(面)부에 사는 데 비하여 1인 가구 가구원들의 24.1%가 면부에 사는 것으로 나타난다. 반면 남자의 경우는 2인+ 가구 가구원들과 1인 가구 가구원들의 거주지 분포 사이에 별 차이가 없다. 나이대별로 보면, 20대·30대 1인 가구 가구원은 같은 나이대의 2인+ 가구 가구원에 비해 동(洞)부에 다소 많이 거주하며 50대·60대+ 1인 가구 가구원은 같은 나이대의 2인+ 가구 가구원에 비해 면부에 다소 많이 거주하는 경향이 있다. 앞에서 여자의 1인 가구 가구원들이 주로 나이 많은 사별자임을 밝힌 바 있는데, 이들이 지역적으로는 농촌에 편중적으로 거주하고 있음을 알 수 있다.

<표 2.3>은 2인+가구 가구원과 1인 가구 가구원의 직업 분포를 비교해 본 것이다. 2인+ 가구 가구원 그룹에 비하여 1인 가구 가구원 그룹에서 서비스직, 농림어업직, 단순직 비율이 상대적으로 높고 판매직, 조립직, 고위직, 사무직 비율이 낮은 것이 특징이다.

<표 2.3> 2인+가구 가구원과 1인 가구 가구원의 직업 분포

직업 범주	2인+ 가구	1인 가구
고위직	5.4%	3.4%
전문가	9.1%	10.1%
준전문가	7.7%	7.0%
사무직	13.5%	11.2%
서비스직	9.2%	13.2%
판매직	12.7%	9.8%
농림어업직	13.5%	16.1%
기능직	10.3%	10.3%
조립직	10.9%	8.9%
단순직	7.3%	9.9%
군인	0.4%	0.2%
계	100.0%	100.0%

* 2% 원자료 분석 결과

〈표 2.4〉 성·나이대와 가구 크기별 20세 이상 인구그룹의 경제활동상태 비율

	가구 구분	남 자		여 자	
		활동 중	활동 안 함	활동 중	활동 안 함
20대	2인+ 가구	55.6%	44.4%	44.7%	55.3%
	1인 가구	68.3%	31.7%	68.3%	31.7%
30대	2인+ 가구	90.8%	9.2%	41.1%	58.9%
	1인 가구	83.4%	16.6%	70.7%	29.3%
40대	2인+ 가구	89.7%	10.3%	48.5%	51.5%
	1인 가구	76.8%	23.2%	62.5%	37.5%
50대	2인+ 가구	77.6%	22.4%	39.6%	60.4%
	1인 가구	65.8%	34.2%	53.1%	46.9%
60대+	2인+ 가구	45.5%	54.5%	22.1%	77.9%
	1인 가구	37.3%	62.7%	27.6%	72.4%

* 2% 원자료 분석 결과

〈표 2.4〉는 2% 원자료의 285,514 가구에 속한 20세 이상 가구원 624,034명을 성·나이대와 가구크기로 나누어 가구원들의 경제활동상태 비율을 비교한 것이다. 20대 남자 2인+ 가구 가구원들의 55.6%가 경제활동 중인 데 비하여 20대 남자 1인 가구 가구원들의 68.3%가 경제활동 중임을 볼 수 있다. 20대 여자의 경우는 차이가 더 심하다. 따라서 젊은 나이대에서는 경제활동이 1인 가구를 구성하는 주요 원인일 것으로 생각된다. 다른 나이대에서도, 여자의 경우는 2인+ 가구 가구원들의 경제활동 비율에 비하여 1인 가구 가구원들의 경제활동 비율이 높은 것으로 나타났다. 그러나 남자의 경우는 20대를 제외한 다른 나이대에서 반대적 경향이 나타났다.

〈표 2.5〉는 2인+ 가구와 1인 가구의 자동차 보유비율을 비교한 것인데, 2인+ 가구의 64.7%가 자동차를 보유하고 있는 데 비하여 1인

가구는 19.8%만이 자동차를 보유하고 있는 것으로 나타났다. 1인 가구의 경우에는 가족동반 이동의 기회가 적으므로 자동차를 보유할 욕구도 작을 것이다. 따라서 자동차 보유비율의 차이를 그대로 경제적 소비 수준의 차이로 해석할 수는 없기는 하나 그래도 일반적으로 자동차 보유여부가 소득 수준과 밀접한 관계에 있다는 점을 감안하면 2인+ 가구와 1인 가구 사이에는 다소의 경제력 차이가 있다고 보아야 할 것이다.

〈표 2.5〉 가구의 크기별 자동차 보유비율

	보유 안 함		보유함		계	
	가구 수	비율	가구 수	비율	가구 수	비율
전 체	120,833	42.3%	164,681	57.7%	285,514	100.0%
2인+ 가구	84,880	35.3%	155,786	64.7%	240,666	100.0%
1인 가구	35,953	80.2%	8,895	19.8%	44,848	100.0%

* 2% 원자료 분석 결과

이상을 종합하여 보면, 1인 가구와 2인+ 가구 및 가구원 사이에는 인구사회적·경제적 특성에서 뚜렷한 차이가 있음을 알 수 있다. 그런데 1인 가구의 전화보유율과 2인+ 가구의 전화보유율에 차이가 있으므로, 전화조사로 수행되는 사회조사에서 대표성을 확보하기 위해서는, 사후 층화 가중시 가구크기를 고려할 필요가 있다고 본다. 개인이 추출단위인 여론조사에 있어서는 가구크기 분포가 일치하도록 가구에 1차 가중치를 주고, 가구 내 가구원 수를 2차 가중치로 하여 재보정하는 것이 원칙일 것이다. 〈표 2.6〉은 가구크기 분포에 관한 2000년 인구주택 총조사 결과이다. 가구가 추출단위인 경우에는 가구크기별 가구 비율을, 개인이 추출단위인 경우에는 가구크기별 인구비율을 가중 칸 설정에 활용할 것을 제안한다.

〈표 2.6〉 가구크기 분포에 관한 2000년 인구주택 총조사 결과

		1인 가구	2인 가구	3인 가구	4인 가구	5인+가구	계
총조사	가구비율*	15.5%	19.1%	20.9%	31.1%	13.4%	100.0%
2%	가구비율	15.7%	19.8%	21.4%	30.3%	12.8%	100.0%
원자료	인구비율**	7.1%	17.1%	22.6%	33.5%	19.8%	100.0%

* 통계청 발표, ** 20세 이상

3. 인터넷 조사에서의 가중 칸 설정

2000년 인구주택 총조사의 2% 원자료를 분석한 결과 20세 이상 성인 가운데 35.3%가 인터넷 사용자로, 64.7%가 비사용자로 밝혀졌다.⁴⁾ 〈표 3.1〉은 인터넷 사용실태를 각종 인구사회적 변인에 따라 나누어 본 것이다.

남자가 여자에 비하여 인터넷 사용비율이 높고, 20대가 그 밖의 나이에 비하여, 서울이 기타 지역에 비하여, 전문가·준전문가·사무직이 다른 직업에 비하여, 대졸 및 석사 이상 고학력자가 저학력자에 비하여, 미혼자가 그 밖의 혼인 상태자에 비하여, 아파트·연립·다세대주택 거주자가 그 밖의 거주형태 거주자에 비하여 인터넷 사용자 비율이 높다는 사실을 볼 수 있다. 그러나 이것은 다수의 관련 요인을 1회 1요인씩 주변적으로만 살펴본 것으로 여러 요인이 중첩되어 있는 상황에서는 관련 요인들을 중복 산입하게 된다. 조사자료의 칸 가중치가 급적 기준 변수의 개수를 작게 하는 것이 바람직하므로 본 연구에서 우리는 Quilan(1993)의 C5.0 나무모형을 활용하여 가급적 소수의 관련

4) 2000년 인구주택 총조사에서는 인터넷 사용여부를 ① 매주 사용, ② 1주일에 1번 이상 사용하나 매일 사용하지는 않음, ③ 1달에 1번 이상 사용하나 매주 사용하지는 않음, ④ 2~3달에 1번 이상 사용하나 매달 사용하지는 않음, ⑤ 사용하지 않음 등으로 질문하였다. 본 연구에서는 2~3달에 1번 이상 사용하는 사람을 '인터넷 사용자'로 정의한다.

〈표 3.1〉 인구사회적 변인에 따른 인터넷 사용자 비율

인구사회적 변인		사 용	사 용 안 함
전체		35.3%	64.7%
성별	남자	41.9%	58.2%
	여자	29.3%	70.8%
연령	20대	68.8%	31.2%
	30대	45.8%	54.2%
	40대	29.1%	70.9%
	50대+	6.7%	93.3%
지역	서울	45.2%	54.8%
	부산	33.1%	66.9%
	대구	35.1%	64.9%
	인천	35.2%	64.8%
	광주	41.7%	58.3%
	대전	41.2%	58.8%
	울산	39.8%	60.2%
	경기	38.4%	61.6%
	강원	26.3%	73.7%
	충청	25.9%	74.1%
	전라	24.1%	75.9%
경상	26.9%	73.1%	
제주	29.7%	70.3%	
직업	전문가, 준전문가, 사무직	74.5%	25.5%
	고위관리직, 군인	58.2%	41.8%
	서비스직, 판매직	30.5%	69.5%
	기능원, 장치조립, 단순노무	23.9%	76.1%
	농림어업	3.8%	96.2%
	없음/결측	29.1%	70.9%
학력	중졸 이하	2.9%	97.1%
	고졸	32.5%	67.5%
	대졸	73.3%	26.7%
	석사 이상	87.4%	12.6%
혼인상태	미혼	69.1%	30.9%
	배우자 있음	30.5%	69.5%
	사별, 이혼	4.7%	95.3%
거주형태	단독주택	25.5%	74.5%
	아파트, 연립, 다세대주택	46.0%	54.0%
	기타**	30.4%	69.6%

* 2% 원자료 분석 결과 ** 영업용 건물, 오피스텔, 숙박업소, 기숙사, 비닐하우스 등

변수들을 선별하고자 한다. 그런데 대부분의 사회조사에서 성과 나이대를 통제하므로 본 연구에서의 나무모형도 성·나이대별로 나누어 만들 필요가 있다. <표 3.2>는 성·나이대별 인터넷 사용자 비율을 보여 준다.

<표 3.2> 성·나이대 조합별 인터넷 사용자 비율

나이대	남 자		여 자	
	사용	사용 안 함	사용	사용 안 함
20대	72.4%	27.6%	65.6%	34.4%
30대	53.3%	46.7%	38.2%	61.8%
40대	38.1%	61.9%	20.1%	79.9%
50대+	11.2%	88.8%	2.9%	97.1%
합계	41.9%	58.2%	29.3%	70.8%

* 2% 원자료 분석 결과

각 성·나이대 인구 그룹에서 인터넷 사용을 목표변수로 C5.0 모형을 만들기에 앞서, 전체 자료를 2:1로 임의 분할하여 2/3 자료를 훈련자료(training data)로 구분하여 모형생성에 사용하고 1/3 자료를 테스트 자료로 생성모형에 대한 평가를 위해 사용하였다. 이 때, 나무모형 생성을 위한 소프트웨어로는 SPSS 클레멘타인 버전 7.0을 활용하였다. 그 결과, 20대 남자 그룹에서는 훈련자료로부터 <그림 3.1>의 나무가 형성되었으며, 이 나무를 테스트 자료에 적용하여 평가한 결과가 <표 3.3>의 2원 분류표이다. 테스트 자료의 20,927명에 대한 인터넷 사용여부 예측에서 15,290(=4,202+11,088)명에 대하여 옳은 결과를 넘으로써 73.1%(=15,290/20,927)의 정확도를 기록하였다. <표 3.4>는 설명변수들의 상대적 중요도를 index한 결과인데 학력과 직업이 가장 중요한 요인으로 나타났다.⁵⁾

5) 입력변수의 중요도를 평가하기 위해 사용된 방법은 다음과 같다(허명희·이용구

- Ⓢ 학력 in ["안발있음" "초등학교" "중학교" "고등학교"] [Mode: 미사용]
- Ⓢ 직업 in ["고위관리자" "서비스" "판매" "기능원" "장치, 조립" "단순노무" "missing"] [Mode: 미사용]
 - Ⓢ 지역 in ["서울"] [Mode: 미사용]
 - 차보유수 <= 0 [Mode: 미사용] ⇨ 미사용
 - Ⓢ 차보유수 > 0 [Mode: 미사용]
 - Ⓢ 직업 in ["고위관리자" "서비스" "missing"] [Mode: 사용]
 - 직업 in ["판매" "기능원" "장치, 조립" "단순노무"] [Mode: 미사용] ⇨ 미사용
 - 지역 in ["부산" "대구" "인천" "광주" "대전" "울산" "경기" "강원" "충북" "충남" "전북" "전남" "경북" "경남" "제주"] [Mode: 미사용] ⇨ 미사용
 - Ⓢ 직업 in ["전문가" "준전문가" "사무직" "군인"] [Mode: 미사용]
 - 지역 in ["서울" "부산"] [Mode: 사용] ⇨ 사용
 - Ⓢ 지역 in ["대구" "인천" "광주" "대전" "울산" "경기" "강원" "충북" "충남" "전북" "전남" "경북" "경남" "제주"] [Mode: 미사용]
 - 주력유형 in ["주력" "연립주력" "다세대주력" "영입용 건물 내 주력" "오피스텔" "숙박업소" "기타"] [Mode: 미사용] ⇨ 미사용
 - 주력유형 in ["아파트" "기숙사 및 특수시설"] [Mode: 사용] ⇨ 사용
 - 직업 in ["농임여업"] [Mode: 미사용] ⇨ 미사용
- Ⓢ 학력 in ["대학(2)" "대학(4)" "석사" "박사"] [Mode: 사용]
 - 직업 in ["고위관리자" "전문가" "준전문가" "사무직" "서비스" "판매" "기능원" "군인" "missing"] [Mode: 사용] ⇨ 사용
 - 직업 in ["농임여업"] [Mode: 미사용] ⇨ 미사용
 - Ⓢ 직업 in ["장치, 조립" "단순노무"] [Mode: 사용]
 - Ⓢ 결혼상태 in ["미혼"] [Mode: 사용]
 - 지역 in ["서울" "대구" "대전" "경기" "충남"] [Mode: 사용] ⇨ 사용
 - Ⓢ 지역 in ["부산" "인천" "광주" "울산" "강원" "충북" "전북" "전남" "경북" "경남" "제주"] [Mode: 미사용]
 - 학력 in ["대학(2)" "석사" "박사"] [Mode: 미사용] ⇨ 미사용
 - 학력 in ["대학(4)"] [Mode: 사용] ⇨ 사용
 - Ⓢ 결혼상태 in ["배우자 있음"] [Mode: 미사용]
 - 주력유형 in ["주력" "연립주력" "영입용 건물 내 주력" "오피스텔" "숙박업소" "기숙사 및 특수시설" "기타"] [Mode: 미사용] ⇨ 미사용
 - 주력유형 in ["아파트" "다세대주력"] [Mode: 사용] ⇨ 사용
 - 결혼상태 in ["사별"] [Mode: 미사용] ⇨ 미사용
 - 결혼상태 in ["미혼" "missing"] [Mode: 사용] ⇨ 사용

* 5단계 분리까지만 수록하였음.

<그림 3.1> 남자 20대 그룹의 C5.0 나무모형: 인터넷 사용여부

2003, Huh and Lee 2003). 이 방법은 출력변수가 연속형인 경우뿐만 아니라 범주형인 경우에도 적용가능한데, 적합 모형

$$\hat{y} = f(x_1, \dots, x_p)$$

에서 변수 x_p 의 상대적 중요도에 대한 측도를

$$Imp(x_p) = \frac{\frac{1}{n} \sum_{i=1}^n |f(x_1, \dots, x_{i,p-1}, x_{ip}^*) - f(x_1, \dots, x_{i,p-1}, x_{ip}^{**})|}{\frac{1}{n} \sum_{i=1}^n |y_i^* - y_i^{**}|}$$

으로 정의한다. 이때 x_{ip}^*, x_{ip}^{**} 는 x_p 의 자료값 중 임의로 선택한 두 값이며 y_i^*, y_i^{**} 역시 y 의 자료값 중 임의로 선택한 두 값이다.

〈표 3.3〉 남자 20대 그룹 테스트 자료에서의 예측 분류표: 인터넷 사용여부

실제 \ 예측	사용하지 않음	사용함	합계	예측정확도
사용하지 않음	4,202(51.2%)	4,006(48.8%)	8,208(100.0%)	51.2%
사용함	1,631(12.8%)	11,088(87.2%)	12,719(100.0%)	87.2%
합 계	5,833(27.9%)	15,094(72.1%)	20,927(100.0%)	73.1%

* 2% 원자료 분석 결과

〈표 3.4〉 남자 20대 그룹 나무모형에서 설명변수들의 중요도: 인터넷 사용여부

변수명		평균절대값 차이	상대적 중요도
종속변수	인터넷 사용여부	0.397	---
설명변수	지역	0.048	0.12
	학력	0.364	0.92
	직업	0.121	0.30
	혼인상태	0.005	0.01
	주택유형	0.021	0.05
	주택소유	0.000	0.00
	자동차 보유여부	0.014	0.00
	연건평	0.004	0.01

〈표 3.5〉는 20대 남자 그룹과 그 밖의 인구 그룹에서 생성된 나무 모형의 주요 결과를 요약하여 보여 준다. 모든 인구 그룹에서 학력이 가장 중요한 요인으로 꼽혔으며 2순위로 나온 요인은 대개 직업이었으나 여자 40대 그룹에서는 주택유형, 50대+ 그룹에서는 자동차 보유대수로 나타났다. 모든 성·나이대 그룹에서 사회적 요인인 학력이 1순위로 꼽힌 점, 2순위 변수로 사회·경제적 요인인 직업(남자 전체, 여자 20대와 30대) 또는 주택유형(여자 40대), 자동차 보유여부(여자 50대+) 등이 꼽힌 점이 흥미롭다.

<표 3.5> 중요한 2개의 설명변수와 중요도(괄호) 및 정확도: 인터넷 사용여부

그룹 순위	남 자			
	20대	30대	40대	50대+
1	학력 (0.92)	학력 (0.60)	학력 (0.77)	학력 (0.67)
2	직업 (0.30)	직업 (0.49)	직업 (0.33)	직업 (0.50)
정확도	73.1%	74.9%	77.9%	87.3%

그룹 순위	여 자			
	20대	30대	40대	50대+
1	학력 (0.62)	학력 (0.71)	학력 (0.99)	학력 (0.37)
2	직업 (0.40)	직업 (0.30)	주택유형(0.30)	자동차 (0.23)
정확도	73.2%	71.2%	76.8%	96.4%

따라서 <표 3.5>에서 제시된 관련 요인들에 의거하여 사후적으로 가중치를 적용하든가 사전적으로 크기비례 할당에 의하여 인터넷 조사 표본을 일반 모집단에 맞출 필요가 있다. 부록 1에 성·나이대별 주요 요인에 따른 인터넷 사용분포를 제시하였다. 인터넷 조사를 응답자 패널의 구축 시 활용할 것을 제안한다.

IV. 모바일 조사에서의 가중 칸 설정

앞 절에서와 같이 2000년 인구주택 총조사의 2% 원자료를 분석한 결과 20세 이상 성인 가운데 58.7%가 모바일 사용자로, 41.3%가 비사용자로 밝혀졌다.⁶⁾ <표 4.1>은 모바일 사용자를 각종 인구사회적 변인에 따라 나누어 본 것으로, 남자가 여자에 비하여 모바일 사용비율

6) 2000년 인구주택 총조사에서는 모바일 사용여부를 ① 이동전화기, ② 무선 호출기, ③ 둘 다 ④ 없음 등으로 질문하였다. 본 연구에서는 ①과 ③의 경우를 모바일 사용자로 정의한다.

〈표 4.1〉 인구사회적 변인에 따른 모바일 사용자 비율

인구사회적 변인		사 용	사 용 안 함
전 체		58.7%	41.3%
성별	남자	76.1%	23.9%
	여자	42.9%	57.1%
연령	20대	80.1%	19.9%
	30대	70.2%	29.8%
	40대	64.6%	35.4%
	50대+	29.3%	70.7%
지역	서울	67.7%	32.3%
	부산	59.4%	40.6%
	대구	62.0%	38.0%
	인천	61.7%	38.3%
	광주	64.2%	35.8%
	대전	64.6%	35.4%
	울산	61.4%	38.6%
	경기	63.7%	36.3%
	강원	50.8%	49.2%
	충청	47.5%	52.5%
	전라	43.8%	56.2%
	경상	48.7%	51.3%
제주	55.1%	44.9%	
직업	전문가, 준전문가, 사무직	89.0%	11.1%
	고위관리직, 군인	93.7%	6.3%
	서비스직, 판매직	68.8%	31.2%
	기능원, 장치조립, 단순노무	72.8%	27.2%
	농림어업	21.8%	78.3%
	없음/결측	42.5%	57.5%
학력	중졸 이하	26.7%	73.3%
	고졸	66.6%	33.4%
	대졸	83.8%	16.2%
	석사 이상	90.3%	9.7%
혼인상태	미혼	84.7%	15.3%
	배우자 있음	56.7%	43.3%
	사별,이혼	25.1%	75.0%
거주형태	단독주택	50.4%	49.6%
	아파트, 연립, 다세대주택	50.4%	49.6%
	기타**	67.5%	32.5%

* 2% 원자료 분석 결과 ** 영업용 건물, 오피스텔, 숙박업소, 기숙사, 비닐하우스 등

이 높고, 나이가 낮을수록 모바일 사용비율이 높으며, 기타 지역에 비해 수도권과 광역시에서 모바일 사용비율이 높은 것으로 나타났다. 또, 전문가, 준전문가, 사무직, 고위관리직이 기타 직업에 비해 모바일 사용비율이 높고, 학력이 높을수록 모바일 사용비율이 높으며, 미혼자가 그 밖의 경우보다 모바일 사용비율이 높다. 그러나 이 역시 3절의 인터넷 사용여부에 대한 분석에서와 마찬가지로 1회 1요인만 살펴보는 것이기 때문에, 보다 적절한 다원적 분석을 위하여 모바일 사용여부를 목표변수로 하여 C5.0 나무모형을 성·나이대 그룹별로 만들어 보았다. 참고로, <표 4.2>는 성·나이대별 모바일 사용자 비율을 보여 준다.

그 결과, 20대 남자 그룹에서는 훈련자료로부터 <그림 4.1>의 나무가 형성되었으며, 이 나무를 테스트 자료에 적용하여 평가한 결과가 <표 4.3>의 2원 분류표이다. 테스트 자료의 20,927명에 대한 모바일 사용여부 예측에서 17,646(=848+16,798)명에 대하여 옳은 결과를냄으로써 84.3%(=17,646/20,927)의 정확도를 기록하였다. <표 4.4>는 20대 남자의 모바일 사용여부에 대한 설명변수들의 상대적 중요도를 인덱스(index)한 결과인데, 인터넷 사용에 대한 나무모형에서와 마찬가지로 모바일 사용에 있어서도 학력과 직업이 가장 중요한 2개의 요인으로 나타났다.

<표 4.5>는 각 성·나이대 그룹에서 생성된 나무모형에서 주요 결과를 요약하여 보여 준다. 인구 그룹에 따라 여러 요인들이 중요한 것으로 나타났다. 전술한 대로, 남자 20대에서는 학력과 직업이 중요하게 나타났지만, 남자 30대와 40대, 50대+에서는 직업과 자동차 보유여부가 가장 중요한 2개의 설명변수로 꼽혔다. 여자 20대는 혼인상태와 직업이, 여자 30대, 40대와 50대+ 그룹에서는 학력과 직업이 꼽혔다.

따라서 각 성·나이대별 관련 요인들에 의거하여 사후 가중치를 적

<표 4.2> 성·나이대 조합별 모바일 사용자 비율

나이대	남자		여자	
	사용함	사용 안 함	사용함	사용 안 함
20대	87.7%	12.3%	73.5%	26.5%
30대	89.6%	10.4%	50.8%	49.2%
40대	84.3%	15.7%	44.7%	55.3%
50대+	47.9%	52.1%	14.4%	85.6%
합 계	76.1%	23.9%	42.9%	57.1%

* 2% 원자료 분석 결과

- ⦿ 직업 in ["고위관리자" "전문가" "준전문가" "사무직" "서비스" "판매" "기능원" "장차, 조립" "군인"] [Mode: 사용]
 - ⦿ 학력 in ["안받았음" "초등학교" "중학교"] [Mode: 사용]
 - ⦿ 차보유수 <= 0 [Mode: 미사용]
 - 결혼상태 in ["미혼" "사별" "missing"] [Mode: 미사용] ⇒ 미사용
 - 결혼상태 in ["배우자 있음" "미혼"] [Mode: 사용] ⇒ 사용
 - 차보유수 > 0 [Mode: 사용] ⇒ 사용
 - ⦿ 학력 in ["고등학교" "대학(2)" "대학(4)" "석사" "박사" "missing"] [Mode: 사용] ⇒ 사용
- ⦿ 직업 in ["농림어업" "단순노무" "missing"] [Mode: 사용]
 - ⦿ 학력 in ["안받았음" "초등학교" "중학교" "고등학교"] [Mode: 미사용]
 - ⦿ 직업 in ["농림어업" "missing"] [Mode: 미사용]
 - ⦿ 학력 in ["안받았음" "초등학교" "중학교"] [Mode: 미사용] ⇒ 미사용
 - ⦿ 학력 in ["고등학교"] [Mode: 미사용]
 - 차보유수 <= 0 [Mode: 미사용] ⇒ 미사용
 - 차보유수 > 0 [Mode: 사용]
 - ⦿ 직업 in ["단순노무"] [Mode: 사용]
 - ⦿ 학력 in ["안받았음" "중학교"] [Mode: 미사용] ⇒ 미사용
 - ⦿ 학력 in ["초등학교" "고등학교"] [Mode: 사용] ⇒ 사용
 - ⦿ 학력 in ["대학(2)" "대학(4)" "석사" "박사" "missing"] [Mode: 사용] ⇒ 사용

* 5단계 분리까지만 수록하였음.

<그림 4.1> 남자 20대 그룹의 C5.0 나무모형: 모바일 사용여부

<표 4.3> 남자 20대 그룹 테스트 자료에서의 예측 분류표: 모바일 사용여부

실제 \ 예측	사용하지 않음	사용함	합계	예측정확도
사용하지 않음	848(35.2%)	1,559(64.8%)	2,407(100.0%)	35.2%
사용함	1,722(9.3%)	16,798(90.7%)	18,520(100.0%)	90.7%
합 계	2,570(12.3%)	18,357(87.7%)	20,927(100.0%)	84.3%

* 2% 원자료 분석 결과

<표 4.4> 남자 20대 그룹 나무모형에서 설명변수들의 중요도: 모바일 사용여부

변수명		평균절대값 차이	상대적 중요도
종속변수	인터넷 사용여부	0.215	---
설명변수	지역	0.008	0.04
	학력	0.190	0.88
	직업	0.142	0.66
	혼인상태	0.015	0.07
	주택유형	0.000	0.00
	주택소유	0.006	0.03
	자동차 보유여부	0.032	0.15
	연건평	0.000	0.00

<표 4.5> 중요한 2개의 설명변수와 중요도(괄호) 및 정확도: 모바일 사용여부

그룹 순위	남자			
	20대	30대	40대	50대+
1	학력(0.88)	직업(0.38)	자동차(0.52)	직업(0.38)
2	직업(0.66)	자동차(0.27)	직업(0.46)	자동차(0.33)
정확도	84.3%	87.5%	81.0%	75.2%

그룹 순위	여자			
	20대	30대	40대	50대+
1	혼인 (0.73)	직업 (0.46)	학력 (0.50)	학력 (0.74)
2	직업 (0.44)	학력 (0.38)	직업 (0.27)	직업 (0.38)
정확도	76.9%	66.0%	65.9%	82.1%

용하든가 아니면 사전적으로 크기비례 할당에 의하여 모바일 조사표본을 일반 모집단에 맞출 필요가 있다. 부록 2에 모집단의 성·나이대별 주요 요인에 따른 모바일 사용분포를 제시하였다. 모바일 조사를 위한 응답자 패널의 구축 시 활용할 것을 제안한다.

V. 맺음말

이상으로 2000년 인구주택 총조사의 2% 원자료 분석을 통해 전화 조사, 인터넷 조사, 모바일 조사에 활용 가능한 가중 칸 설정 문제를 다루어 보았다. 전화조사에 관한 2절의 연구에서는, 가구크기에 따라 전화보유율에서 차이가 있다는 사실을 찾아 냈고 가구크기에 따라 인구사회·경제적 특성이 다르다는 사실을 확인하였다. 이에 따라 가구 크기에 의한 가중치 부여를 제안하였다.

인터넷 조사에 관한 3절의 연구에서는, 인터넷 사용여부를 목표변수로 하는 나무모형을 성·나이대별로 만들어 주로 학력과 직업이 주요한 변수임을 밝혔다. 이를 근거로 성과 나이대 외에 학력과 직업, 또는 주택유형과 자동차 보유여부에 의한 사후 가중 칸 설정 또는 사전적인 크기비례 할당을 제안하였다. 일반 모집단에서의 인터넷 조사를 위한 임의표집은 매우 어렵기 때문에 현실적으로 대다수의 조사가 관은 패널회원 모집에 의해 응답자를 확보하고 있으므로 우리가 제안한 방법을 사용하더라도 인터넷 사용자와 비사용자 간의 차이가 완전히 없어지지 않는 것이다(김영원·변종석 2001, Couper 2000). 그러나 인터넷 사용여부에 대하여 중요한 인구·사회적 요인 면에서 대표성 있는 표본을 사전에 확보하거나 대표성 있는 표본이 되도록 가중치로 보정하는 것은 충분히 의미있는 일이라고 생각한다.

최근 모바일 조사는 전화조사에 대한 희망적인 대안으로 생각되고 있다(강남준 2002, 조성겸 2002). 그러나 모바일 보유율이 아직은 전화 보유율만 못하고 모집단을 확보에 문제가 있어 인터넷 조사에서와 같이 패널회원 모집 방식을 쓰는 경우가 많다. IV절의 모바일 조사에 관한 연구에서는 III절의 방법을 취하여 유사한 결과를 냈다.

우리는 전화조사에서의 가중치 조정을 위한 결과물로 <표 2.6>를 제시하였다. 또한 인터넷 조사와 모바일 조사에서의 가중치 조정 또는

사전 크기비례 할당을 위한 결과물로 부록 1과 부록 2를 제시하였다. 전화, 인터넷, 모바일 등 통신 환경이 빠르게 변하고 있는 것은 사실이지만 이상의 3개 표는 인구·사회적 요인에 대한 통계분포이므로 다음 총조사 결과가 발표되기까지 4~5년 정도는 유효할 것이다.

참고문헌

- 강남준. 2002. “전화조사방법의 대안적 조사기법으로 핸드폰 모바일 조사: 방법론적 논의 및 실제 사례분석”, 한국조사연구학회 모바일 서베이 학술 세미나 자료집, 55-84.
- 권태환·김태현·최진호. 1995. 《한국의 인구와 가족》 일신사.
- 권태환 외 5인. 1997. 《한국 인구구조의 변화와 사회정책적 과제》 통계청.
- 김두섭·박상태·은기수. 2002. 《한국의 인구》 통계청.
- 김영원·변종석. 2001. “인터넷 조사에서 표본추출 동향 및 문제점”, 《인터넷 조사》(이계오 편), 서울: 나남 출판사, 36-55.
- 김종훈·류진화. 2002. “전자우편 설문조사 반응에 관한 문헌적 고찰”, 《조사연구》 3권 2호, 91-122.
- 박진우. 2000. “가중값 조정 방법” 《무응답 오차》 (조사통계연구회 편). 3장.
- 박종무·한영춘·이희욱. 2001. “인터넷 이용자의 특성 분석: 탐색적 연구”, 《대구·경북 마케팅 학회 마케팅 논집》 9권 2호, 95-121.
- 변화순 외 4인. 2000. 《최근 가족해체 실태 및 복지대책》, 한국여성개발원.
- 이두희·강태윤·윤원상. 1998. “우리나라 인터넷 이용자의 인구통계적 특성과 사용현황에 관한 연구”, 《경영연구》 32권 2호, 109-128.
- 조성겸. 2002. “새로운 사회여론조사 방법: 모바일 조사방법의 전망”, 한국조사연구학회 2002년 추계 워크숍 자료집, 91-108.
- 허명희·이용구. 2003. “클레멘타인을 활용한 지도학습모형에서의 입력변수 중요도 평가”, SPSS Korea White Paper 0301 (www.spss.co.kr).
- Bethlehem, J.G. 2002. “Weighting nonresponse adjustments based on auxiliary information,” in *Survey Nonresponse* (Edited by R.M.

Groves et al.), Chapter 18. Wiley.

Couper, M. P. 2000. "Web surveys: A review of issues and approaches," *Public Opinion Quarterly*, Vol. 64. 464–494.

Huh, M. H. and Lee, Y. G. 2003. "Input Variable Importance in Supervised Learning Models," *Korean Communications in Statistics*, Vol 10, No. 1. 239–246.

Quilan, J. R. 1993. C4.5: *Programs for Machine Learning*. Morgan Kaufmann, CA: San Mateo.

부록 1: 주요 2개 요인 조합 간별 인터넷 사용자 비율(%)

남자 20대

직업 \ 학력	중졸-	고졸	대졸	석사+
고위관리자, 군인	0.0	0.7	0.8	0.1
전문가, 준전문가, 사무직	0.1	5.9	14.4	1.2
서비스, 판매직	0.3	5.9	4.0	0.0
기능원, 장치, 조립, 단순노무	1.2	15.8	5.1	0.0
농림어업	0.2	1.1	0.3	0.0
없음/결측	1.1	10.2	30.3	1.3

남자 30대

직업 \ 학력	중졸-	고졸	대졸	석사+
고위관리자, 군인	0.2	2.4	4.1	0.5
전문가, 준전문가, 사무직	0.3	7.9	22.3	3.9
서비스, 판매직	0.8	7.6	5.7	0.1
기능원, 장치, 조립, 단순노무	4.4	21.7	5.7	0.1
농림어업	1.0	1.7	0.5	0.0
없음/결측	1.4	4.4	3.0	0.4

남자 40대

직업 \ 학력	중졸-	고졸	대졸	석사+
고위관리자, 군인	0.7	3.9	5.2	0.9
전문가, 준전문가, 사무직	1.1	7.2	11.4	3.8
서비스, 판매직	3.0	7.9	3.4	0.2
기능원, 장치, 조립, 단순노무	12.3	18.6	3.2	0.1
농림어업	4.4	2.3	0.4	0.0
없음/결측	4.0	4.3	1.8	0.1

남자 50대+

직업 \ 학력	중졸-	고졸	대졸	석사+
고위관리자, 군인	0.7	1.7	2.5	0.5
전문가, 준전문가, 사무직	1.3	3.0	3.6	1.3
서비스, 판매직	3.4	3.3	1.3	0.1
기능원, 장치, 조립, 단순노무	11.1	6.9	1.0	0.0
농림어업	17.1	2.4	0.5	0.0
없음/결측	23.0	9.2	5.6	0.5

여자 20대

직업 \ 학력	중졸-	고졸	대졸	석사+
고위관리자, 군인	0.0	0.1	0.1	0.0
전문가, 준전문가, 사무직	0.1	10.4	20.4	1.0
서비스, 판매직	0.3	7.0	3.2	0.0
기능원, 장치, 조립, 단순노무	0.3	2.6	0.5	0.0
농림어업	0.1	0.3	0.0	0.0
없음/결측	1.4	23.7	27.3	1.1

여자 30대

직업 \ 학력	중졸-	고졸	대졸	석사+
고위관리자, 군인	0.0	0.3	0.3	0.0
전문가, 준전문가, 사무직	0.3	5.2	8.6	1.2
서비스, 판매직	2.6	11.4	2.5	0.0
기능원, 장치, 조립, 단순노무	2.7	4.6	0.4	0.0
농림어업	1.3	1.0	0.1	0.0
없음/결측	6.3	34.0	16.6	0.7

여자 40대

주택유형 \ 학력	중졸-	고졸	대졸	석사+
단독주택	25.9	15.1	2.6	0.2
아파트, 연립, 다세대주택	16.8	24.0	9.1	0.9
기타	2.7	2.3	0.5	0.0

여자 50대+

자동차 보유 \ 학력	중졸-	고졸	대졸	석사+
0대	51.9	4.2	0.8	0.0
1대+	32.9	7.2	2.7	0.2

* 2% 원자료 분석 결과

부록 2: 주요 2개 요인 조합 간별 모바일 사용자 비율(%)

남자 20대

직업 \ 학력	중졸-	고졸	대졸	석사+
고위관리자, 군인	0.0	0.7	0.8	0.1
전문가, 준전문가, 사무직	0.1	5.9	14.4	1.2
서비스, 판매직	0.3	5.9	4.0	0.0
기능원, 장치, 조립, 단순노무	1.2	15.8	5.1	0.0
농림어업	0.2	1.1	0.3	0.0
없음/결측	1.1	10.2	30.3	1.3

남자 30대

직업 \ 자동차	0대	1대+
고위관리자, 군인	0.6	6.6
전문가, 준전문가, 사무직	5.1	29.3
서비스, 판매직	2.8	11.5
기능원, 장치, 조립, 단순노무	9.2	22.5
농림어업	0.9	2.3
없음/결측	4.8	4.4

남자 40대

직업 \ 자동차	0대	1대+
고위관리자, 군인	0.8	10.0
전문가, 준전문가, 사무직	3.1	20.3
서비스, 판매직	2.7	11.8
기능원, 장치, 조립, 단순노무	10.7	23.5
농림어업	2.1	5.0
없음/결측	5.6	4.5

남자 50대+

직업 \ 자동차	0대	1대+
고위관리자, 군인	0.7	4.8
전문가, 준전문가, 사무직	2.1	7.1
서비스, 판매직	2.5	5.5
기능원, 장치, 조립, 단순노무	8.3	10.7
농림어업	12.8	7.3
없음/결측	22.0	16.4

여자 20대

직업 \ 혼인상태	미혼	배우자 있음	사별	이혼
고위관리자, 군인	0.1	0.1	0.0	0.0
전문가, 준전문가, 사무직	25.2	6.6	0.0	0.1
서비스, 판매직	7.1	3.2	0.0	0.2
기능원, 장치, 조립, 단순노무	2.2	1.2	0.0	0.0
농림어업	0.2	0.3	0.0	0.0
없음/결측	27.1	26.2	0.0	0.2

여자 30대

직업 \ 학력	중졸-	고졸	대졸	석사+
고위관리자, 군인	0.0	0.3	0.3	0.0
전문가, 준전문가, 사무직	0.3	5.2	8.6	1.2
서비스, 판매직	2.6	11.4	2.5	0.0
기능원, 장치, 조립, 단순노무	2.7	4.6	0.4	0.0
농림어업	1.3	1.0	0.1	0.0
없음/결측	6.3	34.0	16.6	0.7

여자 40대

직업 \ 학력	중졸-	고졸	대졸	석사+
고위관리자, 군인	0.1	0.4	0.2	0.0
전문가, 준전문가, 사무직	1.0	3.2	3.2	0.7
서비스, 판매직	10.2	10.4	1.3	0.0
기능원, 장치, 조립, 단순노무	8.5	3.7	0.2	0.0
농림어업	5.4	0.8	0.1	0.0
없음/결측	20.2	22.9	7.2	0.3

여자 50대+

직업 \ 학력	중졸-	고졸	대졸	석사+
고위관리자, 군인	0.1	0.1	0.1	0.0
전문가, 준전문가, 사무직	0.4	0.4	0.5	0.1
서비스, 판매직	5.8	1.4	0.2	0.0
기능원, 장치, 조립, 단순노무	4.9	0.5	0.0	0.0
농림어업	15.8	0.2	0.0	0.0
없음/결측	57.8	8.8	2.7	0.1

* 2% 원자료 분석 결과