

제3장

생활시간조사방법론 개선결과 심층검토 및 실무적용방안 연구

제3장



최재혁 · 김의영 · 최종희

제1절 서론

1. 연구배경

생활시간조사는 국민들이 주어진 하루 24시간을 어떤 형태로 보내고 있는지를 파악하여 국민의 생활방식(life style)과 삶의 질을 측정할 수 있는 기초자료를 제공하는 것을 목적으로 실시되고 있다. 생활시간조사 결과는 무급 가사노동에 소요된 시간을 파악하여 가사노동의 경제적 가치를 분석하고 국민계정체계에 가계위성계정을 편입시키기 위한 정보를 제공하는 동시에 각종 노동, 복지, 문화, 교통 관련 정책수립이나 학문 연구를 위한 기초자료로 사용되고 있다.

우리나라 통계청에서는 1999년 생활시간조사가 도입되어 2004년과 2009년에 조사가 수행되었고 2014년에 4차 생활시간조사가 시행될 예정이다. 2009년의 경우 850개 조사구의 8,100가구에 거주하는 10세 이상 20,657명을 대상으로 2일간 조사가 이루어졌으며, 최종적으로 8,090가구의 20,263명에 대한 응답이 완료(가구원 기준 응답률 98.1%) 되었다. 2009년 생활시간조사는 시간일지를 10분 단위로 기입하는데 따른 사생활 노출과 시간적 부담에도 불구하고 양질의 자료가 모아진 것으로 평가되고 있다(김영원 외,

2010). 1999년과 2004년 생활시간조사는 연 1회(9월) 실시되었고 2009년 생활시간조사는 연 2회(3월 및 9월)에 걸쳐 조사가 수행되었으며, 조사한 결과를 평균해 연간통계를 생산하였다.

반면 생활시간조사의 한계점을 지적하는 연구도 진행되어 왔는데, 1년 중 2회에 걸쳐 조사가 진행되어 국민들의 계절별 생활실태를 정확하게 반영하기 어렵다는 의견이 있었다. 우리나라와 같이 사계절이 뚜렷한 경우 계절별로 생활시간 사용실태가 달라질 가능성이 높아 2회에 걸친 조사로는 국민들의 시간활용 패턴을 정확하게 파악하는데 한계를 가질 수 있음을 강조하고 있다(김영원 외, 2010). 생활시간조사의 시간일지 자료는 매우 풍부한 내용을 담고 있음에도 불구하고 조사 횟수가 제한적이라는 점 등의 한계 때문에 이용도가 떨어지는 자료라는 평가도 받고 있다(이윤석 외, 2008). 따라서 현재 상용되고 있는 표본추출방법이나 조사시점 결정 및 조사기간 배정방법이 적절한 것인지 보다 심층적인 분석이 필요한 시점이다.

본 연구 보고서의 구성은 다음과 같다. 2절에서 2009년 생활시간조사의 기초분석과 생활시간 활용의 기상효과를 분석하고, 3절에서 조사 횟수 및 조사 시기를 탐색하였다. 4절에서 가중값 및 분산추정방법을 제시하고 5절에서 연구결과를 정리, 결론을 제시하였다.

2. 연구방법

국민들의 생활양식이 계절에 따라 큰 변화가 있다는 점을 고려하여 2014년에 실시될 조사에서의 적정 조사 횟수를 검토하고, 결정된 조사 횟수에 따른 적정 조사 시기 및 조사기간의 결정뿐 만 아니라 표본가구에 조사 날짜(요일)를 배정하는 효율적인 방안에 대한 심층적인 연구를 실시하였다. 생활시간조사의 표본설계 방식은 시간과 가구라는 두 가지 모집단으로 설명되는 2차원 공간상의 표본추출이므로 두 가지 요소를 동시에 고려하는 표본설계 방식에 따라 연간통계를 산출하기 위한 효율적인 가중값 작성방안을 연구하였다. 또한 동시에 2차원 표본추출이라는 특성을 반영한 분산추정방법에 대한 연구도 실시하였다.

따라서 본 연구의 연구방법은 다음과 같이 진행되었다.

첫째, 조사 횟수 및 조사시점에 대한 연구를 수행하였다. 2009년 생활시간조사 자료를 이용하여 요일 및 월에 따른 생활시간 활용의 변동을 분석하고 2008년 이후의 기상자료를 활용하여 우리나라의 계절성을 파악하여 이에 따른 조사 횟수 및 조사시점을 탐색하였다. 탐색된 조사시점을 기반으로 7, 10, 14일의 기간을 선정하고 표본가구를 날짜(요일)별로 효과적으로 배분하는 방법을 살펴보았다.



둘째, 가중값 및 추정량에 대한 연구를 수행하였다. 가구추출 가중값과 시간 가중값을 이용한 설계가중값 및 무응답 보정과 벤치마크(benchmark) 또는 갈퀴비(raking ratio) 등 최종 calibration 보정 방법에 대해 살펴보았다. 2차원 표본추출이라는 특성을 반영한 분산추정방법과 조사시점 변경이 있을 경우의 추정값의 시계열 단절 처리방안에 대해 연구하였다.

제2절 2009년 생활시간조사 분석

1. 2009년 생활시간조사 기초분석

본 내용은 2009년 생활시간조사보고서(통계청, 2010)와 생활시간조사 방법론 개선 연구(김영원 외, 2010)를 참조하여 정리하였다.

2009년 생활시간조사의 1차 조사는 3월 12일부터 23일까지 12일간 실시되었고 2차 조사는 9월 9일부터 22일까지 14일간 실시되었다. 준비조사(1일 또는 2일), 조사표 정리(2일)를 제외하고 실제 가구조사 기간은 1, 2차 모두 10일간 실시되었다. 시간일지 작성은 표본으로 추출된 각 조사구 내 15가구를 대상가구로 선정하여 3가구씩 분할한 5개조별로 2일 동안 정해진 날짜에 작성하도록 하였고, 그 요일별 구성은 다음 [그림 3-1]과 같다.

2009년 3월							2009년 9월						
일	월	화	수	목	금	토	일	월	화	수	목	금	토
				12	13	14				9	10	11	12
15	16	17	18	19	20	21	13	14	15	16	17	18	19
22	23						20	21	22				

※ 12일: 준비조사 / 23일: 조사표정리 ※9,10일: 준비조사 / 21,22일: 조사표정리

[그림 3-1] 시간일지 작성일자 분할

조사대상은 전국에 거주하는 만 10세 이상 가구원이고 조사 참조기간은 1차의 경우 3월 11일, 2차의 경우 9월 9일 현재 거주로 조사되었다. 다만 조사표 배부일부터 회수일까지

가구의 조사대상자 전원이 부재중인 경우나 조사표 작성이 불가능하다고 인정되는 경우 (가구원이 상을 당한 경우, 수재, 주택이 차압당하는 등 가구원 전체의 주거가 매우 불안정한 경우 등)의 가구와 친척방문, 출장, 병원 입원 등의 사유로 조사표 배부일부터 회수일까지 기간 중 부재중인 가구원, 심신장애, 치매 등의 사유로 본인 스스로 시간일지 작성이 불가능한 가구원(단, 함께 동거하는 타 가구원이 그 사람의 생활을 객관적으로 대신 기입할 수 있는 경우는 조사대상자로 포함)은 제외하였다.

본 절에서 살펴볼 생활시간조사 분석을 위한 주요변수는 수면시간, 근로시간, 가사노동시간, TV시청시간으로 근로시간과 가사노동시간은 다음과 같은 행동을 포함하고 있다.

- 근로시간: 주업/부업/일 중 휴식/일 관련 연구/집에 가져와서 일함/그 외 일 관련 행동
- 가사노동시간: 식사준비/설거지, 식후정리/간식 및 비일상적 음식 만들기/세탁 및 세탁물 넣기/옷 정리/다림질, 바느질, 의류손질/의류수선, 세탁 서비스 받기/재봉, 뜨개질/방·물품 정리/집안 청소(쓸기, 닦기)/그 외 청소 및 정리/가재도구·집 직접 손질하기/가재도구·집 손질 관련 서비스 받기/세차, 차량 직접 관리하기/세차, 차량 관련 서비스 받기/그 외 집 직접 관리하기/그 외 집 서비스 받기/생활용품 관련 직접 쇼핑/내구재 관련(생활용품 제외) 직접 쇼핑/생활용품 관련 무점포 쇼핑(인터넷·TV 홈쇼핑 등)/내구재 관련(생활용품 제외) 무점포 쇼핑(인터넷·TV 홈쇼핑 등)/가계부 정리/가정계획/은행 및 관공서 일 보기/기타 가사일

요일은 일요일부터 토요일을 각각 분석하되, 주중/주말을 구분하여 분석하기 위해 주말을 토·일 또는 금·토·일로 구분하여 시간활용 형태를 각각 살펴보았다. 수면시간은 요일별로 차이가 없는 반면 근로시간은 주말에 사용시간이 감소하고 가사노동시간과 TV시청시간은 주말에 사용시간이 증가하는 것으로 나타났다.

〈표 3-1〉 주요변수에 대한 주중/주말 시간활용 형태

요일 (건수)	수면시간		근로시간		가사노동시간		TV시청시간	
	mean	s.e	mean	s.e	mean	s.e	mean	s.e
주중(월-금) (24,256)	438.10	0.614	214.21	1.852	82.27	0.831	98.37	0.776
주말(토-일) (16,270)	480.02	0.916	101.68	1.731	99.95	1.018	143.38	1.054
주중(월-목) (16,135)	438.12	0.720	215.97	2.188	81.81	0.980	98.22	0.914
주말(금-일) (24,391)	466.00	0.725	136.95	1.597	94.64	0.827	128.54	0.844

〈표 3-2〉 주요변수에 대한 요일별 시간활용 형태

요일 (건수)	수면시간		근로시간		가사노동시간		TV시청시간	
	mean	s.e	mean	s.e	mean	s.e	mean	s.e
월요일 (4,011)	443.36	1.403	210.64	4.43	84.14	2.005	100.89	1.92
화요일 (4,045)	435.31	1.420	218.27	4.38	82.49	1.991	100.30	1.80
수요일 (4,045)	435.92	1.456	212.26	4.27	79.11	1.852	97.26	1.85
목요일 (4,034)	437.93	1.471	222.60	4.41	81.54	1.983	94.53	1.74
금요일 (8,121)	438.05	1.078	207.31	3.07	84.04	1.404	98.93	1.32
토요일 (8,173)	465.10	1.264	129.40	2.63	97.30	1.435	129.55	1.40
일요일 (8,097)	495.19	1.296	73.53	2.19	102.65	1.445	157.44	1.56

생활시간조사는 2회(3월, 9월) 실시되므로 조사시점의 기상여건 또는 조사대상자의 차이에 따른 주요변수의 차이가 존재할 수 있다. 이를 확인하기 위해 요일별로 조사시점에 따른 주요변수의 차이를 t-검정으로 실시하였다.

수면시간의 경우, 월요일을 제외한 모든 요일에서 3월과 9월의 차이가 존재하고 주말, 주중 모두 3월의 수면시간이 9월보다 많은 것으로 나타났고 근로시간의 경우, 토요일을 제외한 모든 요일에서 3월과 9월의 차이가 없고 주말의 경우 3월의 근로시간이 9월보다 많았다. 가사노동시간의 경우, 금요일을 제외한 모든 요일에서 3월과 9월의 차이가 없었고 주말과 주중으로 구분했을 경우 모두 9월의 가사노동시간이 3월보다 많았고 TV시청시간의 경우, 모든 요일, 주말, 주중에서 3월의 TV시청시간이 9월보다 많은 것으로 나타났다.

이는 3월의 조사기간보다 9월의 조사기간이 기온(평균기온, 최저기온, 최고기온)이 더 높고 강수량은 더 적으며 일조량도 더 많아 가사활동 및 야외활동이 상대적으로 증가하여, 수면시간은 더 적고 가사노동시간 및 TV시청시간이 더 많았을 것으로 예상할 수 있다. 추가적으로 9월의 조사기간은 추석 2주 전으로 성묘나 벌초와 같은 야외활동이 증가할 요인도 가지고 있었다. 조사기간의 기상여건은 다음 절에서 더 자세히 살펴보았다.

〈표 3-3〉 요일별 조사시점에 따른 수면시간의 차이검정

요일	월	mean	difference	s.e	t-value
일요일	3월	502.91	15.165	1.6414	6.77***
	9월	487.75		1.5271	
월요일	3월	445.49	4.223	1.8563	1.71
	9월	441.26		1.6267	
화요일	3월	440.56	10.529	1.8072	4.28***
	9월	430.04		1.6720	
수요일	3월	440.38	8.930	1.8089	3.53***
	9월	431.45		1.7647	
목요일	3월	441.06	6.199	1.8775	2.45*
	9월	434.87		1.7013	
금요일	3월	440.98	5.877	1.3415	3.19**
	9월	435.10		1.2627	
토요일	3월	468.19	6.157	1.5318	2.87**
	9월	462.04		1.5042	

p-value: ***<.001 **<.01 *<.05

〈표 3-4〉 주중/주말별 조사시점에 따른 수면시간의 차이검정

요일	월	mean	difference	s.e	t-value
주중 (월~금)	3월	441.68	7.127	0.7545	6.93***
	9월	434.55		0.6996	
주말 (토~일)	3월	485.30	10.425	1.1383	6.64***
	9월	474.88		1.0810	
주중 (월~목)	3월	441.86	7.446	0.9189	5.96***
	9월	434.42		0.8472	
주말 (금~일)	3월	470.35	8.624	0.9003	6.96***
	9월	461.73		0.8522	

p-value: ***<.001 **<.01 *<.05

〈표 3-5〉 요일별 조사시점에 따른 근로시간의 차이검정

요일	월	mean	difference	s.e	t-value
일요일	3월	75.45	3.763	2.6988	1.01
	9월	71.69		2.5949	
월요일	3월	213.94	6.558	5.4730	0.87
	9월	207.38		5.2170	
화요일	3월	217.18	- 2.191	5.4583	- 0.29
	9월	219.37		5.1618	
수요일	3월	210.83	- 2.866	5.3185	- 0.39
	9월	213.69		5.0772	
목요일	3월	227.87	10.408	5.4139	1.38
	9월	217.47		5.2211	
금요일	3월	208.46	2.310	3.7314	0.44
	9월	206.15		3.6688	
토요일	3월	134.84	10.833	3.2882	2.40*
	9월	124.01		3.0898	

p-value: ***<.001 **<.01 *<.05



〈표 3-6〉 주중/주말별 조사시점에 따른 근로시간의 차이검정

요일	월	mean	difference	s.e	t-value
주중 (월~금)	3월	215.63	2.821	2.1974	0.93
	9월	212.81		2.1129	
주말 (토~일)	3월	105.58	7.697	2.1579	2.60**
	9월	97.88		2.0378	
주중 (월~목)	3월	217.47	2.977	2.7084	0.80
	9월	214.49		2.5850	
주말 (금~일)	3월	140.28	6.607	1.9551	2.44*
	9월	133.67		1.8818	

p-value: ***<.001 **<.01 *<.05

〈표 3-7〉 요일별 조사시점에 따른 가사노동시간의 차이검정

요일	월	mean	difference	s.e	t-value
일요일	3월	101.28	- 2.696	1.7695	- 1.06
	9월	103.98		1.8134	
월요일	3월	82.87	- 2.510	2.4492	- 0.72
	9월	85.38		2.4990	
화요일	3월	80.11	- 4.769	2.4434	- 1.35
	9월	84.88		2.5354	
수요일	3월	77.86	- 2.522	2.2969	- 0.77
	9월	80.38		2.3284	
목요일	3월	79.22	- 4.594	2.4020	- 1.32
	9월	83.81		2.5108	
금요일	3월	81.39	- 5.321	1.6858	- 2.16*
	9월	86.71		1.7918	
토요일	3월	94.84	- 4.898	1.7599	- 1.93
	9월	99.74		1.8233	

p-value: ***<.001 **<.01 *<.05

〈표 3-8〉 주중/주말별 조사시점에 따른 가사노동시간의 차이검정

요일	월	mean	difference	s.e	t-value
주중 (월~금)	3월	80.29	- 3.953	0.9767	- 2.80**
	9월	84.24		1.0140	
주말 (토~일)	3월	98.01	- 3.841	1.2483	- 2.10*
	9월	101.85		1.2859	
주중 (월~목)	3월	80.01	- 3.612	1.1991	- 2.10*
	9월	83.62		1.2349	
주말 (금~일)	3월	92.41	- 4.442	1.0064	- 3.10**
	9월	96.85		1.0469	

p-value: ***<.001 **<.01 *<.05

〈표 3-9〉 요일별 조사시점에 따른 TV시청시간의 차이검정

요일	월	mean	difference	s.e	t-value
일요일	3월	166.34	17.462	2.0566	6.51***
	9월	148.88		1.7197	
월요일	3월	106.98	12.099	2.5298	3.64***
	9월	94.89		2.1502	
화요일	3월	109.09	17.618	2.4208	5.55***
	9월	91.48		2.0512	
수요일	3월	105.27	16.050	2.4975	4.98***
	9월	89.22		2.0367	
목요일	3월	100.45	11.701	2.3313	3.78***
	9월	88.75		2.0384	
금요일	3월	105.26	12.699	1.7522	5.50***
	9월	92.56		1.5031	
토요일	3월	134.57	9.994	1.8183	4.11***
	9월	124.58		1.6141	

p-value: ***<.001 **<.01 *<.05

〈표 3-10〉 주중/주말별 조사시점에 따른 TV시청시간의 차이검정

요일	월	mean	difference	s.e	t-value
주중 (월~금)	3월	105.41	14.035	1.0002	10.70***
	9월	91.37		0.8499	
주말 (토~일)	3월	150.22	13.512	1.3819	7.43***
	9월	136.71		1.1868	
주중 (월~목)	3월	105.44	14.374	1.2231	8.98***
	9월	91.07		1.0349	
주말 (금~일)	3월	135.06	12.942	1.1071	8.86***
	9월	122.12		0.9555	

p-value: ***<.001 **<.01 *<.05

2. 생활시간 활용의 기상효과 분석

본 연구에서는 생활시간 활용이 기후환경에 따라 변화가 있다고 전제하여 기상자료를 활용하여 비교·분석하였다. 기상자료는 기상청 홈페이지(<http://www.kma.go.kr>)에서 제공하는 평균기온(°C), 최저기온(°C), 최고기온(°C), 강수량(mm), 적설여부, 풍속(m/s), 습도(%), 일조시간(hours), 운량(할) 등을 사용하였으며, 연도별 기후변화의 변동을 줄이기 위해 2008년부터 2012년까지 각 일별 5년 평균값을 사용하여 분석하였다. 기상자료는 자료의 척도(scale)를 표준화(standardization)하기 위해 4분위수(quartile)를 이용하여 4개 범주(순서형 변수)로 범주화하였는데, 각 범주의 질량(밀도; probability, density, mass)을 동일하게 하기 위해 4분위수를 이용하였다.

〈표 3-11〉 기상자료의 범주화 기준

범주	평균기온	최저기온	최고기온	강수량	풍속	습도	일조시간	운량
1	3.6 미만	-0.1 미만	7.9 미만	0.1 미만	2.3 미만	53.1 미만	4.5 미만	3.6 미만
2	3.6~13.6	-0.1~9.8	7.9~19.0	0.1~1.1	2.3~2.5	53.1~59.1	4.5~6.0	3.6~4.7
3	13.6~22.8	9.8~19.1	19.0~26.9	1.1~4.5	2.5~2.9	59.1~65.9	6.0~7.3	4.7~6.1
4	22.8 이상	19.1 이상	26.9 이상	4.5 이상	2.9 이상	65.9 이상	7.3 이상	6.1 이상

〈표 3-12〉 기상자료의 범주화 기준

범주	평균기온		최저기온		최고기온		강수량		풍속		습도		일조시간		운량	
	3월	9월	3월	9월	3월	9월	3월	9월	3월	9월	3월	9월	3월	9월	3월	9월
1	2	-	3	-	2	-	6	8	1	8	2	2	5	3	3	6
2	5	-	5	-	6	-	1	-	-	1	2	2	1	-	-	-
3	3	10	2	10	2	9	-	-	2	1	1	2	-	1	4	2
4	-	-	-	-	-	1	3	2	7	0	5	4	4	6	3	2

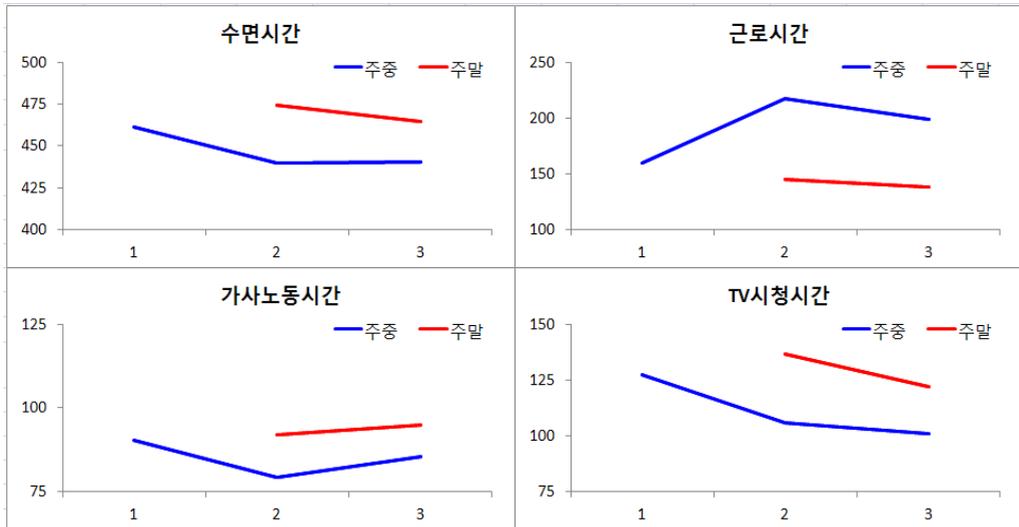
기상효과 분석을 위해 2009년 생활시간조사 자료와 2009년 조사일자의 기상자료를 결합하여 기상변화에 따른 주요변수(수면시간, 근로시간, 가사노동시간, TV시청시간)의 차이검정을 실시하였다. 주중/주말 효과와 교호작용이 존재할 수 있으므로 이를 포함하여 분석하였다. 이때 주말의 정의를 금요일 포함하는 경우와 그렇지 않은 경우를 모두 고려하였다.

기상변화에 따른 2009년 생활시간조사 조사일수를 살펴본 결과, 기온(평균기온, 최저기온, 최고기온)의 경우 3월 조사기간은 9월 조사기간에 비해 변동이 많았다. 강수량의 경우는 3월 조사기간 10일 중 4일간(1일은 강수량이 미비함) 비가 왔고 9월 조사기간 10일 중 2일간 비가 왔으며, 풍속의 경우는 3월 조사기간에는 바람이 강하게 부는 날이 많았고 9월 조사기간에는 바람이 약하게 부는 날이 많았다.

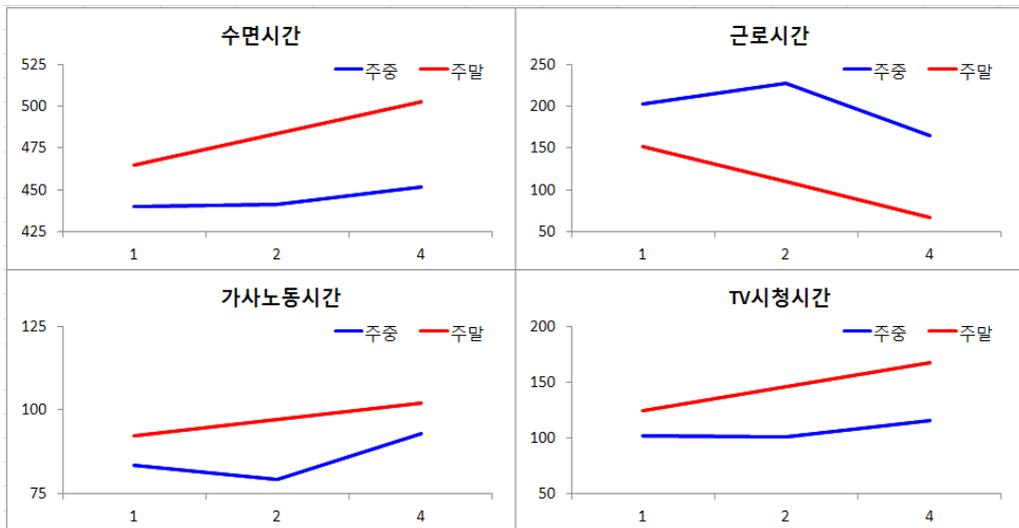
평균기온에 따른 시간활용을 분석한 결과, 수면시간과 TV시청시간은 평균기온이 상승할수록 시간활용이 줄어드는 효과가 있고 근로시간은 평균기온이 상승할수록 시간활용이 증가하는 효과가 있으며, 가사노동시간은 평균기온에 영향을 받지 않는 것으로 나타났다. 가사노동시간을 제외한 근로시간, 수면시간, TV시청시간은 주말/주중과 기상변수 간의 교호작용이 존재하고 있었다.

강수량에 따른 시간활용을 분석한 결과, 수면시간, 가사노동시간, TV시청시간은 강수량이 많아질수록 시간활용이 증가하는 효과가 있고 근로시간은 강수량이 많아질수록 시간활용이 감소하는 효과가 있는 것으로 나타났다. 강수량의 경우도 평균기온과 마찬가지로 가사노동시간을 제외한 근로시간, 수면시간, TV시청시간이 주말/주중 변수와 교호작용이 존재하고 있었다. 주요변수에 대해 기온(평균기온, 최저기온, 최고기온)과 강수량을 제외한 나머지 기상자료에 대한 효과는 존재하지 않는 것으로 나타났다.





[그림 3-2] 평균기온에 따른 평균 시간활용



[그림 3-3] 강수량에 따른 평균 시간활용



제3절 생활시간조사 조사 횟수 및 조사 시기 탐색

1. 해외사례

세계 여러 나라들이 주로 5년마다 생활시간조사를 실시하고 있다. 본 연구의 해외 사례는 생활시간조사 방법론 개선방안(이윤석 외, 2008), 생활시간조사 2010년 정기 통계품질진단(김태현 외, 2010), 생활시간조사 방법론 개선연구(김영원 외, 2010)와 미국의 US Bureau of Labor Statistics(<http://www.bls.gov>), 호주의 Australian Bureau of Statistics (<http://www.abs.gov.au>), 일본의 Statistics Japan(<http://www.stat.go.jp>)에서 제공되는 자료를 정리하였다.

우리나라의 생활시간조사는 5년마다 실시되는 조사로 최근 2009년 실시되었다. 조사는 3월, 9월 각 10일간 총 2회로 실시되었고 표본단위의 조사일수는 2일이며, 10분 간격의 시간일지를 작성하고 있다. 조사 횟수 측면에서 우리나라와 유사하게 진행되고 있는 나라는 존재하지 않으며, 미국, 캐나다, 영국, 프랑스 등이 연중 매일 조사를 실시하고 있고 호주는 사계절을 반영하여 4회 실시하고 있는 반면 일본의 경우 특이하게 연 1회만 실시하고 있다.

〈표 3-13〉 생활시간조사의 주요국 사례

국 가	통 계 명	조사주기 및 최근조사	조사 횟수 (조사기간 및 배분일자)	개인별 조사일수 및 시간일지 간격
한 국	Time Use Survey	5년/2009	2회 (3, 9월 각 10일간)	2일/10분 간격
미 국	American Time Use Survey	매년	연중 매일 (12개월, 월별)	1일/15분 간격
영 국	The Time Use Survey	5년/2010	연중 매일 (15개월 ^{a)} , 월별)	2일 ^{b)} /10분 간격
캐나다	General Social Survey - Time Use	5년/2010	연중 매일 (12개월, 월별)	1일/5분 간격
프랑스	Time Use Survey	10년/2009	연중 매일 (12개월 ^{c)} , 각 6주간의 8개 집락 ^{d)})	1일/10분 간격
호 주	Time Use Survey	5년/2011	4회 (2, 4, 6, 10월 각 13일간)	2일/5분 간격
일 본	Survey on Time Use & Leisure Activities	5년/2011	1회 (10월 9일간)	2일/30분 간격

a) 전년도 6월~당해년도 9월

b) 개인별 주중 1일 주말 1일, 총 2일 조사

c) 당해년도 2월 중순~익년도 2월 중순

d) 2.16-3.29/3.30-5.10/5.11-6.21/6.22-8.2/8.17-9.2/9.28-11.8/11.9-12.20/1.4-2.14

연중 매일 조사를 실시하는 나라의 경우 표본단위의 배정을 시점으로 구분하여 실시되고 있는데, 미국, 캐나다, 영국의 경우는 매월 표본을 결정하고 일자를 배정하는 방법을 택하고 있고 프랑스의 경우는 계절성을 반영하여 1년을 8개의 집락(각 6주¹⁾)으로 구분하여 표본을 결정하고 일자를 배정하는 방법을 사용하고 있다. 연중 매일 순환표본으로 생활시간조사를 실시하는 나라들의 조사일수와 시간일지 간격을 살펴보면, 매년 실시하고 있는 미국(American Time Use Survey)의 경우는 표본단위의 조사일수는 1일이며, 15분 간격의 시간일지를 작성하고 있다. 5년마다 실시하고 있는 나라로 캐나다, 영국, 프랑스의 경우는 캐나다(General Social Survey - Time Use)가 최근 2010년에 조사되었고 표본단위의 조사일수는 1일이고 5분 간격의 시간일지를 작성토록 하고 있다. 영국(The Time Use Survey)은 최근 2010년에 조사되었고 다른 나라와 달리 총 조사기간이 15개월(2009.6.~2010.9.)로 이루어져 있으며, 표본단위의 조사일수를 주중 1일(월~금 중 1일), 주말 1일(토, 일 중 1일)로 배정하고(총 10개의 경우의 수) 10분 간격의 시간일지를 작성토록 하고 있다. 최근 2009년에 조사된 프랑스(Time Use Survey)는 표본단위의 조사일수가 1일이고 10분 간격의 시간일지를 작성하고 있다.

계절을 반영하여 연 4회로 실시하고 있는 호주(Time Use Survey)를 살펴보면, 5년마다 실시되는 조사로 최근 2011년에 조사되었다. 사계절을 반영하여 2월, 6월, 9월, 11월에 각 13일간 4회²⁾에 걸쳐 조사하고 있으며, 표본단위의 조사일수는 2일로 10분 간격의 시간일지를 작성하고 있다. 일본(Survey on Time Use & Leisure Activities)의 경우는 5년마다 실시하는 조사로 최근 2011년 조사되었고 10월 9일간 1회로 실시하고 있으며, 표본단위의 조사일수는 2일로 30분 간격의 시간일지를 작성하도록 되어 있다.

2. 시간모집단 총화

일반적인 통계조사는 특정 시점의 모집단의 특성을 파악하는 것이 목적이지만, 생활시간조사는 특정시점이 아닌 1년 동안의 모집단 특성(생활 활용시간)을 파악하는 조사로 공간적인 모집단(개인 또는 가구)과 시간적인 모집단(1년 365일)을 동시에 고려해야 하는 2차원 모집단의 특성을 파악하는 조사이다. 따라서 가구 또는 개인의 표본추출과 동시에 조사시점을 선택(추출)하는 방법에 대한 고민이 필요하다. 특히 대부분의 연구 결과, 사람들의 생활양식이 계절에 영향을 받는다고 알려져 있기 때문에 적절한 조사시점을 선택하는 것은 생활시간조사의 대표성에 매우 중요한 부분이다.

앞 절에서 살펴본 바와 같이, 계절의 특성인 기상효과에 따른 생활 활용시간의 차이가

-
- 1) 2009년의 경우, 2009년 2월 16일부터 2010년 2월 14일까지 실시되었고 6개의 집락은 2.16.-3.29./3.30.-5.10./5.11~6.21./6.22~8.2./8.17.~9.2./9.28.~11.8./11.9.~12.20./1.4~2.14.로 구분하여 실시되었다.
 - 2) 2011년의 경우, 2.21.~3.5./4.25.~5.7./6.27.~7.9./10.24.~11.5. 동안 조사가 실시되었다.



존재하기 때문에 이를 이용하여 시간모집단을 층화(군집)하여 조사시점을 선택할 필요가 있다. 생활시간조사가 연속된 10일 또는 14일간 실시되기 때문에 본 연구에서는 시간모집단을 365일이 아닌 51주³⁾로 정의하였다. 시간모집단을 365일로 정의하는 경우는 개별 날짜의 기상자료에 지나친 영향을 받으므로 연속된 날짜 중심으로 층화가 형성되지 않아 조사기간을 선택하는데 어려움이 있을 수 있다.

생활시간조사 방법론 개선방안 연구(김영원 외, 2010)에서는 시간모집단을 365일로 정의하고 기상자료⁴⁾와 함께 월변수를 순서형으로 간주하여 분석하였기 때문에 적정 수준 정도의 연속된 날짜가 층화되었다. 그러나 계절의 영향성을 직접 변수에 삽입하기 위해 9월을 '1', 10월을 '2', 8월을 '12'로 변환하였기 때문에 봄/가을을 명확하게 구분되었고 9월은 7, 8월 등 여름에 속할 수 없다는 기본가정을 가지고 분석된 결과라고 볼 수 있다. 최초 층수를 4개로 고정하고 군집하여 봄(3월 1일~6월 17일), 여름(6월 18일~8월 31일), 가을(9월 1일~11월 30일), 겨울(12월 1일~2월 28일)로 층화하였다(<표 3-14> 참조). 층화에 가장 큰 영향력을 미치는 변수는 기온과 강수량인 것으로 나타났고 대학생 방학, 고등학생 방학, 휴가기간 등을 더미 변수로 삽입하여 분석하였기 때문에 이에 대한 영향이 매우 큰 결과로 층화되었다.

<표 3-14> 365일 모집단의 4개 층화 결과(김영원 외, 2010)

1군집(층): 86일		2군집(층): 113일		3군집(층): 77일		4군집(층): 89일	
월	날짜수	월	날짜수	월	날짜수	월	날짜수
12	25일	3	29일	6	14일	9	29일
1	31일	4	30일	7	31일	10	31일
2	28일	5	31일	8	31일	11	29일
3	2일	6	16일	9	1일		
		11	1일				
		12	6일				

↓
겨울

↓
봄

↓
여름

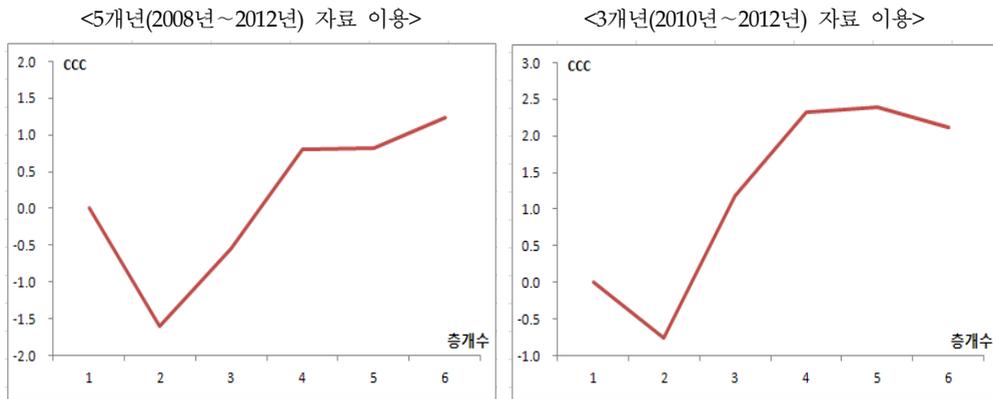
↓
가을

본 연구에서는 연도별 기후변화의 변동을 줄이기 위해 2008년부터 2012년까지 5년 평균을 기상자료로 활용하였고, 최근 기후의 변화가 심한 것을 고려하여 하여 추가적으로 2010년부터

- 3) 생활시간조사는 2014년에 실시될 예정이므로 2014년 달력의 일~토를 한 주로 정의하고 연초, 연말에 한 주가 부족한 부분에 대해서는 제외하여 365일을 51주로 정의하였다.
- 4) 급격한 연도별 기후 변화에 따른 변동을 줄이기 위해 2005년부터 2009년까지 5년 동안의 기상자료의 평균을 구해 이용하였다.

2012년까지 3년 평균 자료를 활용하였다. 사용변수는 평균기온(°C), 최저기온(°C), 최고기온(°C), 강수량(mm), 적설여부, 풍속(m/s), 습도(%), 일조시간(hours), 운량(할) 등을 사용하였고 자료 척도의 표준화를 위해 연속형 변수인 경우 표준화 변환을 이용하였다.

시간모집단(51주)에 대한 층화방법은 계층적 군집방법(ward방법)으로 층화하였다. 계층적 군집방법의 다른 알고리즘이나 비계층적 군집방법인 K-means방법 모두 비슷한 결과를 가지며, 평균에 대한 오차를 최소화한다는 의미에서 평균제곱오차(mean square error; MSE)를 사용하는 ward방법을 최종방법으로 결정하였다. 앞서 언급한 생활시간조사 방법론 개선방안 연구(김영원 외, 2010)는 최적의 층수를 4개로 고정하고 분석한 반면, 본 연구에서는 CCC(cubic clustering criterion)와 고유값(eigen-value) 및 분산변동비율을 이용하여 최적의 층 수(군집 수)를 탐색하여 기상효과를 반영할 수 있는 최적의 조사 횟수를 검토하였다.



[그림 3-4] 층 개수에 따른 CCC

<표 3-15> 층 개수에 따른 분산변동비율의 증감을

조사 횟수	5개년 기상자료 이용 기상변화 변동량 설명		3개년 기상자료 이용 기상변화 변동량 설명	
	설명비율	증감	설명비율	증감
1	63.7%		62.9%	
2	82.6%	18.9%p	81.4%	18.5%p
3	92.7%	10.1%p	91.4%	9.9%p
4	96.9%	4.2%p	96.0%	4.6%p

CCC를 이용하여 최적의 층 개수(조사 횟수)를 탐색한 결과, 2개 지점에서 elbow-point가 나타났고, 5개년 자료에서는 3회 지점, 3개년 자료에서는 4회 지점에서 양수로 전환되는 것으로



나타났다. 고유값을 통해 살펴본 결과, 5개년, 3개년 자료 모두 2회 지점에서 고유값이 1 이상으로 나타났으나 3회 지점 역시 각각 0.912, 0.895로 1에 가까운 값을 가진 것으로 분석되었다. 분산변동 설명비율은 두 자료 모두 2회 지점에서는 80% 이상, 3회 지점에서는 90% 이상, 4회 지점에서는 95% 이상인 것으로 나타났다. 그 증감은 2회에서 3회로 증가했을 경우 10%p 정도를 추가적으로 설명할 수 있고, 3회에서 4회로 증가했을 경우 4%p 정도를 추가적으로 설명할 수 있는 것으로 나타났다. 따라서 조사 횟수를 2회로 결정할 경우, 기상자료에 대한 변동 설명이 부족할 가능성이 존재하여 3회 이상의 조사 횟수가 필요할 것으로 분석되었다.

〈표 3-16〉 층 개수에 따른 고유값과 분산변동비율

〈5개년(2008년~2012년) 자료 이용〉				〈3개년(2010년~2012년) 자료 이용〉			
층 개수	고유값	변동설명		층 개수	고유값	변동설명	
		비율	누적비율			비율	누적비율
1	5.733	63.7%	63.7%	1	5.662	62.9%	62.9%
2	1.699	18.9%	82.6%	2	1.667	18.5%	81.4%
3	0.912	10.1%	92.7%	3	0.895	9.9%	91.4%
4	0.376	4.2%	96.9%	4	0.417	4.6%	96.0%
5	0.180	2.0%	98.9%	5	0.231	2.6%	98.6%
6	0.071	0.8%	99.7%	6	0.098	1.1%	99.7%

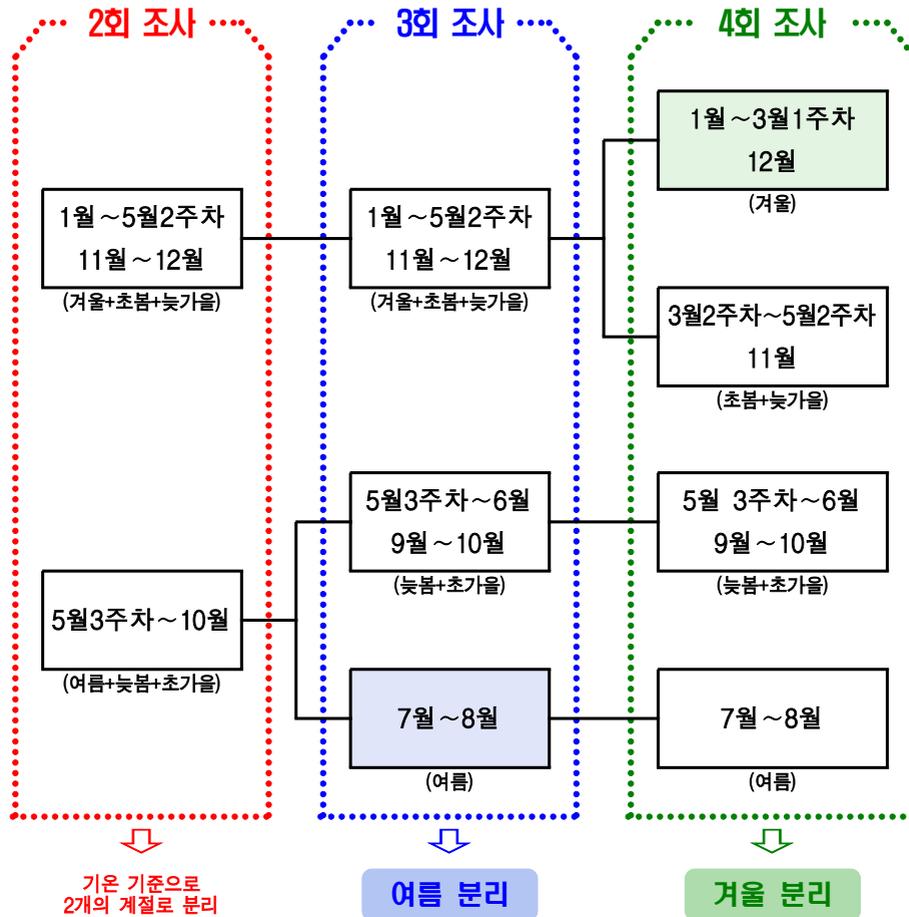
다른 유사연구와는 다르게 본 연구에서는 획일적인 계절 구분에 따른 조사 시기 대표기간(층)을 탐색하는 것이 아닌, 기상자료에 의해 1년 365일을 51주로 하여 이를 군집화하여 대표기간을 설정하였다. 과거 2009년 조사에서는 2번의 조사시점(3월, 9월)이 6개월씩 대표기간(3월; 1월16일~7월15일, 9월; 7월16일~익년 1월15일)을 갖는 것으로 간주하여 조사하여 분석하였으나, 계층적 군집분석으로 조사 시기 대표기간을 설정한 결과는 다음과 같은 덴드로그램(dendrogram)을 통해 확인할 수 있다([그림 3-5], [그림 3-6]).

2008년부터 2012년까지 5개년 기상자료를 이용하여 조사 시기 대표기간을 탐색한 결과, 최초 2개의 조사 시기 대표기간 구분은 기온의 영향력을 가장 크게 받아 5월 3주차부터 10월(‘늦봄+여름+초가을’ 그룹으로 정의)까지와 1월부터 5월 2주차까지, 11월부터 12월까지(‘초봄+늦가을+겨울’ 그룹으로 정의)로 구성되었다. 하나의 조사 시기 대표기간을 추가할 경우, ‘초봄+늦가을+겨울’ 그룹은 그대로 유지한 상태에서 ‘늦봄+여름+초가을’ 그룹에서 여름(7월에서 8월까지)이 분리되는 것으로 나타났다. 4개의 조사 시기 대표기간 구분은 ‘초봄+늦가을+겨울’ 그룹에서 겨울(1월에서 3월 1주차까지, 12월)이 분리되는 것으로 나타났다.

2010년부터 2012년까지 3개년 기상자료를 이용하여 조사 시기 대표기간을 탐색한

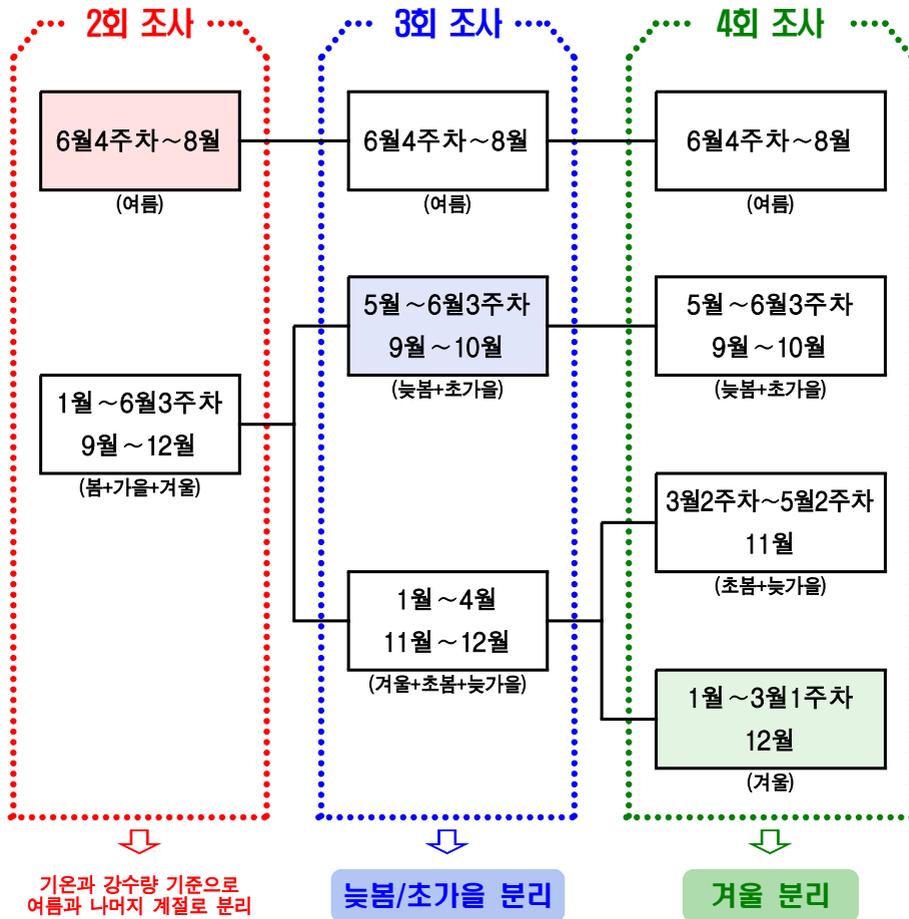
결과, 최초 2개의 조사 시기 대표기간 구분은 5개년 자료와는 다르게 기온과 더불어 강수량의 영향력을 크게 받아 여름(6월 4주차부터 8월까지)과 나머지 계절(‘봄+가을+겨울’ 그룹으로 정의)로 구성되었다. 하나의 조사 시기 대표기간을 추가할 경우, 여름은 그대로 유지한 상태에서 ‘봄+가을+겨울’ 그룹이 5월부터 6월 3주차까지, 9월부터 10월까지(‘늦봄+초가을’ 그룹으로 정의)와 1월부터 4월까지, 11월부터 12월까지(‘초봄+늦가을+겨울’ 그룹으로 정의)로 분리되는 것으로 나타났다. 4개의 조사 시기 대표기간 구분은 ‘초봄+늦가을+겨울’ 그룹에서 겨울(1월에서 3월1주차까지, 12월)이 분리되는 것으로 나타났다.

<5개년(2008년~2012년) 자료 이용>



[그림 3-5] 5개년 자료(2008년~2012년)를 이용한 51주 군집분석(51주)의 덴드로그램

<3개년(2010년~2012년) 자료 이용>



[그림 3-6] 3개년 자료(2010년~2012년)를 이용한 군집분석(51주)의 덴드로그램

5개년 자료와 3개년 자료 모두 동일하게 계절적 분리는 매우 유사한 결과를 가지고 있지만, 기존의 생활시간조사와 같이 2회 조사를 실시할 경우의 조사 시기 대표기간은 강수량의 영향력에 따라 구성이 달라짐을 알 수 있다. 이는 최근의 기상이 국지성 폭우로 인해 여름철의 강수량이 증가하고 봄과 가을의 기간이 짧아져 다른 계절과의 경계가 모호해짐에 기인한다고 볼 수 있다. 또한 4개의 조사 시기 대표기간을 구성하는 경우, 계절 경계가 모호한 '초봄+늦가을' 그룹과 '늦봄+초가을' 그룹이 추가된다는 점에서 조사시간 선택이나 대표기간을 반영한 가중값 계산단계에서 비효율적인 방법이 될 가능성이 높다. 따라서 봄과 가을의 기상효과가 유사하다는 가정 하에서 3개의 조사

시기 대표기간으로 구분하거나 또는 일자의 연속성을 반영하여 4개의 조사 시기 대표기간에서 연속된 일자별로 재구성하는 방법이 시간 및 비용 대비 매우 효과적인 가능성이 높다. 다시 말해 ‘초봄+늦가을’ 그룹과 ‘늦봄+초가을’ 그룹을 봄과 가을로 재구성하는 것이 효율적일 수 있다. 최종적으로 선택 가능한 조사 시기 대표기간은 장마기간을 비롯한 강수량이 반영된 3개년 기상자료를 활용하여 3개의 조사 시기 대표기간을 선택하는 것이 국지성 폭우 및 강수량의 증가, 겨울이 상대적으로 길어지는 추세, 봄과 가을의 계절적 기간이 줄어드는 추세 등을 반영할 수 있는 방법으로 판단된다. 방학, 휴가 등이 생활시간에 영향을 줄 것으로 판단되나, 더미변수(dummy variable)가 가지는 한계점으로 영향력이 지나치게 커질 수 있어 고려하지 않았다.

<표 3-17> 조사 시기 대표기간 구성

조사 시기 대표기간	<5개년(2008년~2012년) 자료 이용>	<3개년(2010년~2012년) 자료 이용>
2 개	[기온높음] 5월 3주차~10월	[여 름] 6월 4주차~8월
	[기온낮음] 1월~5월 2주차, 11월~12월	[봄가을/겨울] 1월~6월 3주차, 9월~12월
3 개	[봄 ^{a)} /가을 ^{b)}] 5월 3주차~6월, 9월~10월	[봄 ^{a)} /가을 ^{b)}] 5월~6월 3주차, 9월~10월
	[여 름] 7월~8월	[여 름] 6월 4주차~8월
	[겨 울 ^{c)}] 1월~5월 2주차, 11월~12월	[겨 울 ^{c)}] 1월~4월, 11월~12월
4 개	[봄 ^{a)}] 3월 2주차~6월	[봄 ^{a)}] 3월 2주차~6월 3주차
	[여 름] 7월~8월	[여 름] 6월4 주차~8월
	[가 울 ^{b)}] 9월~11월	[가 울 ^{b)}] 9월~11월
	[겨 울 ^{c)}] 1월~3월 1주차, 12월	[겨 울 ^{c)}] 1월~3월 1주차, 12월

- a) 봄은 늦봄만을 의미함
- b) 가을은 초가을만을 의미함
- c) 겨울은 초봄과 늦가을을 포함하고 있음

3. 조사 시기 및 조사기간 탐색

각 조사 시기 대표기간에서 기상자료의 중심이 되는 주간을 탐색하기 위해 기상자료의 유클리디안 거리(euclidean distance)를 이용하여 중앙값을 최적의 조사 시기라고 판단하였다. 조사 시기 대표기간(조사 횟수)을 2개, 3개, 4개로 산정하여 5개년, 3개년 자료를 활용하여 각각 조사 시기의 최적 시점을 탐색하고 예산과 조사환경에 따라 선택하는 방법으로 판단하였다. 탐색된 조사 시기는 기상자료의 중심주간으로 선택된 것이며,



예산 및 조사환경에 따라 대표기간에서 임의로 한주를 선택하는 방법도 고려할 수 있다. 5개년, 3개년 자료의 활용에 따른 조사 시기 대표기간의 개수별 최적의 조사 시기는 중앙값을 중심으로 다음 <표 3-18>, <표 3-19>, <표 3-20>과 같이 세 개의 조사 시기가 각각 선택되었다.

세부적으로 살펴보면, 2회 조사의 경우 5개년 기상자료를 활용하면 기존 조사 시기인 3월 2주차와 9월 2주차가 선택되었다. 3회 조사의 경우는 5개년, 3개년 기상자료를 활용한 결과에 공통으로 나타나는 조사 시기는 5월 3주차, 7월 2주차, 11월 3주차이고 4회 조사의 경우는 9월 2주차가 추가되는 것으로 분석되었다. 4회 조사의 경우는 군집분석에서 층수를 4개로 결정한 것이 아니라 3회 조사에서 ‘봄+가을’ 그룹을 연속된 일자로 층을 추가하였기 때문에 나머지 층의 조사 시기는 동일하게 나타나게 된다.

<표 3-18> 2회 조사의 경우 조사 시기 선택

기상자료	조사 시기 대표기간		조사 시기 선택
5개년 (2008년~2012년)	기온높음	5월 3주차~10월	9월 2주차 , 9월 1주차, 6월 4주차
	기온낮음	1월~5월 2주차, 11월~12월	3월 2주차 , 3월 4주차, 11월 3주차
3개년 (2010년~2012년)	여 름	6월 4주차~8월	8월 4주차, 7월 2주차, 7월 3주차
	나머지	1월~6월 3주차, 9월~12월	11월 2주차, 4월 2주차, 4월 1주차

<표 3-19> 3회 조사의 경우 조사 시기 선택

기상자료	조사 시기 대표기간		조사 시기 선택
5개년 (2008년~2012년)	봄 ^{a)} /가을 ^{b)}	5월 3주차~6월, 9월~10월	5월 3주차 , 6월 1주차, 6월 2주차
	여 름	7월~8월	7월 2주차 , 8월 2주차, 7월 1주차
	겨 울 ^{c)}	1월~5월 2주차, 11월~12월	3월 2주차, 3월 4주차, 11월 3주차
3개년 (2010년~2012년)	봄 ^{a)} /가을 ^{b)}	5월~6월 3주차, 9월~10월	5월 1주차, 5월 2주차, 5월 3주차
	여 름	6월 4주차~8월	8월 4주차, 7월 2주차 , 7월 3주차
	겨 울 ^{c)}	1월~4월, 11월~12월	11월 3주차 , 12월 2주차, 1월 1주차

a) 봄은 늦봄만을 의미함

b) 가을은 초가을만을 의미함

c) 겨울은 초봄과 늦가을을 포함하고 있음

〈표 3-20〉 4회 조사의 경우 조사 시기 선택

기상자료	조사 시기 대표기간	조사 시기 선택	
5개년 (2008년~2012년)	봄 ^{a)}	5월 3주차~6월	5월 3주차 , 6월 1주차, 6월 2주차
	여름	7월~8월	7월 2주차 , 8월 2주차, 7월 1주차
	가을 ^{b)}	9월~10월	9월 2주차 , 9월 1주차, 10월 3주차
	겨울 ^{c)}	1월~5월 2주차, 11월~12월	3월 2주차, 3월 4주차, 11월 3주차
3개년 (2010년~2012년)	봄 ^{a)}	5월~6월 3주차	5월 1주차, 5월 2주차, 5월 3주차
	여름	6월 4주차~8월	8월 4주차, 7월 2주차 , 7월 3주차
	가을 ^{b)}	9월~10월	9월 3주차, 9월 2주차 , 10월 3주차
	겨울 ^{c)}	1월~4월, 11월~12월	11월 3주차 , 12월 2주차, 1월 1주차

a) 봄은 늦봄만을 의미함

b) 가을은 초가을만을 의미함

c) 겨울은 초봄과 늦가을을 포함하고 있음

조사 시기는 시간모집단의 구성의 특성상 1주 단위로 선택되었지만 1주의 조사기간은 충분한 자료를 획득하기에 너무 적기 때문에 10일 또는 14일(2주)에 대한 고민이 필요하다. 생활시간조사는 추출단위 및 분석단위가 다른 조사로 추출단위는 가구, 주(일간)이고 조사단위는 가구원과 2일이며, 분석단위는 가구원별 1일로 매우 복잡한 구조를 가지고 있다. 앞서 언급한 대로 생활시간조사는 특정시점이 아닌 1년 동안의 모집단 특성(생활 활용시간)을 파악하는 조사로 공간적인 모집단(개인 또는 가구)과 시간적인 모집단(1년 365일)을 동시에 고려해야 하는 2차원 모집단의 특성을 파악하는 조사이다.

앞 절에서 논의한 조사 횟수 및 조사 시기에 관련된 내용은 시간적인 모집단을 고려한 것이고 공간적인 모집단의 경우는 2단 집락추출을 통해 조사대상이 선택된다. 먼저 공간적인 모집단을 층화(지역 및 가구특성)하여 조사구를 추출하고 조사구 내 가구를 추출한다. 실제 조사는 추출된 조사구 내 가구를 5개의 조로 나누어 가구 내 전체 가구원에 대해 2일씩 조사하는 순환조사(rolling survey) 방법으로 실시되며, 실제 분석에서는 가구원별 1일을 분석단위로 사용하게 된다. 가구원별 1일이 최종 분석단위이므로 요일별 변동을 고려해야 정도(precision) 높은 통계를 생산할 수 있다. 일반적으로 월요일부터 금요일까지의 생활시간 사용에 대한 변동은 크지 않기 때문에 상대적으로 변동이 큰 토요일과 일요일의 추출률을 높이는 것이 더 효율적일 가능성이 높다.

2009년 생활시간조사에서는 월요일부터 금요일까지 12일(365일 중 4.6%; 이하 추출률), 토요일과 일요일의 8일(365일 중 7.7%; 이하 추출률)을 조사하였다. 만약 요일별의 추출률의

차이 없이 1차당 14일을 조사하는 경우, 월요일부터 금요일까지는 20일(추출률 7.7%)을 조사하게 되고 토요일과 일요일은 8일(추출률 7.7%)을 조사하게 된다. 세부적으로 월요일부터 목요일까지의 생활시간 사용에 대한 변동은 크지 않고 금요일의 변동이 상대적으로 크다고 본다면 1차당 10일 조사가 비용대비 더 효율적일 것으로 판단된다. 다시 말해 금요일, 토요일, 일요일을 묶어 주말로 간주하여 월요일부터 목요일까지는 8일(추출률 3.8%)을 조사하고 금요일부터 일요일까지는 12일(추출률 7.7%)을 조사하는 방식이다. 이러한 불균등 추출방식에 대한 보정은 다음 절에서 설명할 가중값 조정 단계에서 해결할 수 있다.



제4절 가중값 작성 및 분산추정방법

1. 가중값 작성방법

앞 절에서 살펴본 바와 같이 생활시간조사는 특정시점이 아닌 1년 동안의 모집단 특성(생활 활용시간)을 파악하는 조사로 공간적인 모집단(가구원)과 시간적인 모집단(365일)을 동시에 고려해야 하는 2차원 모집단의 특성을 파악하는 조사이다. 따라서 가중값을 두 개 작성하여 두 가중값의 곱의 형태를 최종 가중값으로 사용해야 한다. 공간적인 모집단에 대한 가구원의 가중값은 이중추출에 따른 설계가중값을 추계인구로 사후층화(post-stratification) 보정하고 시간적인 모집단에 대한 일자의 가중값은 조사 시기 대표기간과 요일을 고려하여 계산한다. 이러한 방식으로 계산된 두 가중값의 곱을 다시 추계인구로 사후층화 보정하여 최종 가중값을 작성하게 된다.

가중값에 사용되는 기호는 다음과 같다.

- $N_{h,sa}$: h 층(지역⁵⁾)의 s 성별의 a 연령대⁶⁾의 추계인구
- $n_{h,sa}$: h 층의 s 성별의 a 연령대의 응답자 수
- w_{hki}^s : h 층의 k 번째 조사구의 i 번째 가구에 대한 사회조사 설계가중값⁷⁾
- m_h^s : h 층의 사회조사 조사구 수
- m_h : h 층의 생활시간조사 조사구 수

5) 지역 25개 층(7대 대도시, 9개 도의 읍·면/동부)으로 세분화한다.

6) 연령대는 10대부터 60세 이상까지 10세 단위를 기본으로 하고 동부지역의 경우만 10대를 5세 간격(10~14세, 15~19세)으로 세분화하여, 동부는 총 7개 그룹, 읍·면부는 총 6개 그룹으로 구성된다.

7) 설계가중값은 사후층화 보정하기 전 가중값으로 조사구 내 가구는 모두 동일한 가중값을 가진다.

- r_{hk} : h 층의 k 번째 조사구의 응답 가구 수
- D_{ot} : o 층(조사 시기 대표기간)의 t 요일의 총 일자 수
- d_{ot} : o 층(조사 시기 대표기간)의 t 요일의 조사일자 수

생활시간조사는 사회조사 표본 조사구를 대표본(large sample)으로 사용하는 이중추출(double sampling; two phase sampling) 방법으로 설계되었으므로 사회조사 설계가중값을 사용하여 생활시간조사 설계가중값을 작성한다. 대표본으로 사용되는 사회조사는 2단 집락추출(two-stage cluster sampling)이므로 불균등 확률로 추출되었고 이를 고려하기 위해 설계가중값을 그대로 사용해야 한다. 생활시간조사 설계가중값은 최초 배정된 표본크기와 조사가 완료된 유효 표본크기를 고려한 응답률(response rate)을 통해 작성되며, h 층의 k 번째 조사구의 i 번째 가구의 가구원에 대한 설계가중값은 다음과 같다. 다만 향후 생활시간조사에서 이중추출 방법을 사용하지 않을 경우는 사회조사 설계가중값 대신 1을 대입하고 h 층의 사회조사 조사구 수인 m_h^s 대신 h 층의 전체 조사구 수인 M_h 을 대입하여 계산할 수 있다.

$$w_{hki}^d = w_{hki}^s \times \frac{m_h^s}{m_h} \times \frac{15}{r_{hk}} \quad (3.1)$$

대표본인 사회조사 표본 조사구는 15개 가구로 구성되어 있고 생활시간조사는 사회조사 표본 조사구 중 일부 조사구를 추출하여 15개 가구를 모두 조사하는 방법으로 1단 집락추출(one-stage cluster sampling)이기 때문에 조사구 내 가구에 대한 추출률을 적용할 필요가 없다. 또한 표본가구의 전체 가구원을 조사하는 방식이므로 설계가중값은 가구 가중값과 가구원 가중값이 동일한 값을 가진다. 작성된 가구원에 대한 설계가중값을 첫 번째 사후층화 보정한 가중값인 h 층의 s 성별, a 연령대의 j 번째 가구원의 1차 사후층화 보정 가중값은 다음과 같다.

$$w_{hj,sa}^p = N_{h,sa} \times \frac{w_{hj,sa}^d}{\sum_{j=1}^{n_{h,sa}} w_{hj,sa}^d} \quad (3.2)$$

여기서 $w_{hj,sa}^d$ 는 앞서 계산한 w_{hki}^d 의 그룹 속성이 바뀐 값이다. 두 번째로 일자에

대한 가중값은 각 조사 시기 대표기간과 실제 조사기간의 상대적 크기를 이용하여 작성하며, o 층(차 조사) t 요일에 조사된 일자 가중값은 다음과 같다.

$$w_{ot} = D_{ot}/d_{ot} \quad (3.3)$$

가구원에 대한 최종 가중값은 앞에서 작성된 1차 사후층화 보정 가중값과 일자 가중값의 곱을 다시 추계인구로 2차 사후층화 보정하여 작성한다. 앞서 언급한 대로 최종 분석단위가 가구원당 1일이므로 조사된 일자에 따라 각각 최종 가중값을 계산하게 된다. 따라서 2차 사후층화 보정된 o 층(차 조사) t 요일에 조사된 h 층의 s 성별, a 연령대의 j 번째 가구원의 최종 가중값은 다음과 같다.

$$w_{ot,hj,sa} = N_{h,sa} \times \frac{w_{ot,hj,sa}^p w_{ot}}{\sum_{j=1}^{n_{ot,h,sa}} w_{ot,hj,sa}^p w_{ot}} \quad (3.4)$$

여기서 $w_{ot,hj,sa}^p$ 는 앞서 계산한 $w_{h,sa}^p$ 을 가중값으로 가지는 분석단위의 o 층(차 조사) t 요일의 그룹속성을 반영한 값이다.

2. 분산추정방법

앞에서 계산된 최종 가중값 $w_{ot,hj,sa}$ 을 이용하여 모집단 추정량을 구하고 이에 대한 분산추정량을 계산해야 한다. 추정해야 할 항목은 평균시간, 행위자 비율, 행위자 평균시간, 10분 간격별 행위자 비율, 30분 간격별 행위자 비율로 다음과 같은 정의로 계산되어 진다(통계청, 2010).

- 평균시간: 행동을 하지 않은 사람을 포함한 전체 응답자의 행동분류별 평균시간으로 주 행동만 집계한 경우 1일 전체 평균시간의 합은 24시간이다.
- 행위자 비율: 대상자 중 하루 24시간 동안 각 행동을 10분 이상 한 사람의 비율
- 행위자 평균시간: 각 행동을 하루 24시간 동안 10분 이상 한 사람들의 행동별 평균시간
- 10분 간격별 행위자 비율: 하루를 10분 간격으로 구분하여 10분 동안 각 행동을 한 사람의 해당인구 전체에 대한 비율
- 30분 간격별 행위자 비율: 각 30분 간격 내의 3개의 10분 간격 행위자 비율의 평균값



첫 번째로 l 번째 행동분류의 평균시간(\widehat{Y}_l)은 다음과 같은 식으로 계산된다.

$$\widehat{Y}_l = \sum_{h=1}^{25} \sum_{k=1}^{m_h} \sum_{i=1}^{n_{hk}} \sum_{q=1}^{Q_{hki}} \sum_{u=1}^2 \frac{w_{hkij,u} y_{hkij,u,l}}{w_{\dots}} \quad (3.5)$$

여기서 $y_{hkij,u,l}$ 는 h 층의 k 번째 조사구의 i 번째 가구의 q 번째 가구원의 u 일자의 l 번째 행동분류의 관측값(분 단위), $w_{hkij,u}$ 는 h 층의 k 번째 조사구의 i 번째 가구의 q 번째 가구원의 u 일자의 최종 가중값으로 $w_{ot,hj,sa}$ 의 그룹속성이 바뀐 값, Q_{hki} 는 h 층의 k 번째 조사구의 i 번째 가구의 총 가구원 수이고, w_{\dots} 은 가중값의 총계로 다음과 같이 구한다.

$$w_{\dots} = \sum_{h=1}^{25} \sum_{k=1}^{m_h} \sum_{i=1}^{n_{hk}} \sum_{q=1}^{Q_{hki}} \sum_{u=1}^2 w_{hkij,u} y_{hkij,u,l} \quad (3.6)$$

l 번째 행동분류의 평균시간의 분산추정량($\widehat{Var}(\widehat{Y}_l)$) 및 표준오차($\widehat{s.e.}(\widehat{Y}_l)$)와 상대 표준오차($\widehat{CV}(\widehat{Y}_l)$)는 다음과 같은 식으로 계산된다.

$$\widehat{Var}(\widehat{Y}_l) = \sum_{h=1}^{25} \left[\left(\frac{m_h}{m_h - 1} \right) \left(\frac{M_h - m_h}{M_h} \right) \sum_{k=1}^{m_h} (e_{hk\dots} - \bar{e}_{h\dots})^2 \right]$$

$$\widehat{s.e.}(\widehat{Y}_l) = \sqrt{\widehat{Var}(\widehat{Y}_l)}$$

$$\widehat{CV}(\widehat{Y}_l) = \frac{\widehat{s.e.}(\widehat{Y}_l)}{\widehat{Y}_l} \times 100(\%) \quad (3.7)$$

여기서 $e_{hk\dots} = \sum_{i=1}^{n_{hk}} \sum_{q=1}^{Q_{hki}} \sum_{u=1}^2 w_{hkij,u} (y_{hkij,u,l} - \widehat{Y}_l) / w_{\dots}$, $\bar{e}_{h\dots} = \sum_{k=1}^{m_h} e_{hk\dots} / m_h$ 이다.

행위자 비율에 대한 추정값과 분산추정에 관련된 식은 $y_{hkij,u,l}$ 대신 h 층의 k 번째 조사구의 i 번째 가구의 q 번째 가구원의 u 일자의 l 번째 행동분류가 10분 이상 존재하면 1,



존재하지 않으면 0인 지시변수(더미변수) $z_{hkiq,u,l}$ 을 대입하여 계산한다.

세 번째로 l 번째 행동분류의 행위자 평균시간(\widehat{X}_l)은 평균시간과 행위자 비율의 비추정(ratio estimator) 형태로 계산되는 추정량으로 다음과 같이 계산된다.

$$\widehat{X}_l = \frac{\sum_{h=1}^{25} \sum_{k=1}^{m_h} \sum_{i=1}^{n_{hk}} \sum_{q=1}^{Q_{hki}} \sum_{u=1}^2 w_{hkiq,u} y_{hkiq,u,l}}{\sum_{h=1}^{25} \sum_{k=1}^{m_h} \sum_{i=1}^{n_{hk}} \sum_{q=1}^{Q_{hki}} \sum_{u=1}^2 w_{hkiq,u} z_{hkiq,u,l}} \quad (3.8)$$

따라서 분산추정과 관련된 식도 비추정량의 분산추정을 구하는 방식을 이용하여 동일한 형태로 다음과 같이 계산된다.

$$\widehat{Var}(\widehat{X}_l) = \sum_{h=1}^{25} \left[\left(\frac{m_h}{m_h - 1} \right) \left(\frac{M_h - m_h}{M_h} \right) \sum_{k=1}^{m_h} (b_{hk\dots} - \bar{b}_{h\dots})^2 \right] \quad (3.9)$$

$$\text{여기서 } b_{hk\dots} = \frac{\sum_{i=1}^{n_{hk}} \sum_{q=1}^{Q_{hki}} \sum_{u=1}^2 w_{hkiq,u} (y_{hkiq,u,l} - z_{hkiq,u,l} \widehat{X}_l)}{\sum_{h=1}^{25} \sum_{k=1}^{m_h} \sum_{i=1}^{n_{hk}} \sum_{q=1}^{Q_{hki}} \sum_{u=1}^2 w_{hkiq,u} z_{hkiq,u,l}}, \quad \bar{b}_{h\dots} = \sum_{k=1}^{m_h} b_{hk\dots} / m_h \text{이다.}$$

행위자 평균시간 추정량에 대한 표준오차와 상대표준오차는 위의 평균시간의 표준오차, 상대표준오차와 동일한 방법으로 계산된다.

10분 간격별 행위자 비율은 24시간을 10분 간격으로 144개의 시간대(c)로 구성하여 h 층의 k 번째 조사구의 i 번째 가구의 q 번째 가구원의 u 일자의 l 번째 행동분류가 c 시간대에 존재하면 1, 존재하지 않으면 0인 지시변수(더미변수)을 대입하여 계산하고 분산추정에 관련된 식을 동일한 방법으로 계산할 수 있다. 30분 간격별 행위자의 비율은 계산된 10분 간격별 행위자 비율의 $c-1$, c , $c+1$ 시간대의 추정값의 평균을 이용하여 계산하며 분산추정에 관련된 식 또한 독립사건으로 간주하여 분산추정값의 평균을 이용하여 계산할 수 있다.

본 절에서 계산된 추정량 및 분산추정량은 이중추출 방법을 사용하지 않고 생활시간조사 자체의 표본추출방법을 사용할 경우에 해당되는 방법을 가정하고 있고 이는 SAS와 같은 통계프로그램에서 계산이 가능하다. 그러나 사회조사 표본 조사구를 이용한 이중

추출 방법을 사용하는 경우 추정량은 동일한 방법으로 계산되지만 분산추정량의 계산은 대표본과 소표본(second sample)의 추출방법에 따라 추정 방법이 매우 복잡하게 결정되며, 기존의 통계패키지를 이용하여 적절한 추정량의 계산이 불가능하다. 이중추출 하에서의 분산 추정방법에 관한 내용은 별도의 연구를 통해 구현해야 하며, 다각도로 검토가 필요한 사항이므로 본 연구에서는 언급하지 않았다.

제5절 제언

생활시간조사에서 주요 활동에 대한 시간 활용을 분석한 결과, 주말여부별로 조사 시점에 따라 차이가 존재함을 확인할 수 있었다. 이는 시간활용의 형태가 평균기온, 강수량 등의 기상효과에 직접적인 영향을 받기 때문인 것으로 생각된다. 따라서 국민들의 다양한 행동을 가능한 한 많이 조사하기 위해 연중 12개월에 걸쳐 조사를 시행하는 것이 가장 바람직할 것으로 판단되나, 면대면 조사를 실시하는 현실을 고려할 때 연 2회 보다는 연 3회 이상 실시하여 계절효과를 파악할 수 있도록 하는 것이 적절할 것이다.

2014년 생활시간조사를 위한 예산의 부족 등 현실적 제약으로 인하여 다양한 조사 시점을 적용하기 어려운 실정이므로 현실적으로 가능한 조사 횟수는 3회 또는 4회가 가장 적당할 것으로 생각되며, 추후 예산과 일정 및 조사환경 등을 고려하여 연중 조사로 전환하는 방향으로 연구가 계속되어야 할 것이다.

조사가 연속된 10일 또는 14일로 실시된다는 점을 고려하여 기상자료를 1주 형태(평균)로 구성하고 시간모집단을 51주로 정의하여 총화하였다. 연도별 기후변화의 변동을 줄이기 위해 2008년부터 2012년까지 5년 평균, 2010년부터 2012년까지 3년 평균 자료를 각각 활용하여 최적의 층 개수 및 대표기간을 탐색한 결과, 3개로 총화하고 5월, 7월, 11월의 조사시기가 가장 적절한 것으로 나타났다. 다만 기존의 생활시간조사와의 시계열 단절을 방지하기 위해 봄과 가을의 대표주간인 5월 조사를 9월 조사로 전환하여 7월, 9월, 11월 각각 10일간 조사하는 것이 가장 효과적일 것으로 판단된다. 2009년 조사와는 다르게 각 조사 시기가 대표하는 기간이 서로 다르므로 가중값 계산단계에서 조정이 필요하다. 최근 3년 기상평균은 다소 다른 결과를 가지며, 이는 최근의 기상변화 중 여름철 강수량 증가, 특히 국지성 폭우가 반영된 것으로 판단된다. 따라서 여름기간의 조사는 반드시 필요할 것이다.

생활시간조사는 추출단위(가구원, 일자) 및 분석단위(가구원별 1일)가 다른 조사이므로 이를 반영하는 조사기간이 필요하다. 따라서 가구원별 1일을 최종 분석단위로 정의할

경우, 월요일부터 금요일까지의 생활 활용시간에 대한 변동은 크지 않기 때문에 상대적으로 변동이 큰 토요일과 금요일의 추출률을 높이는 것이 더 타당할 가능성이 높다. 다만 금요일이 다른 주중 요일과 시간활용 측면의 차이점이 있다는 측면에서 금요일의 추출률을 높여 한 시기에서 10일간(월~목 각 1일, 금~일 각 2일) 조사하는 것이 가장 효율적일 것이다.

2009년 생활시간조사는 사회조사 표본 조사구를 이용한 이중추출 방법을 사용했으나 분산추정 단계에서 이를 반영하지 않았기 때문에 분산이 과소 추정되었을 가능성이 매우 높다. 따라서 2014년 생활시간조사는 조사의 관리 측면에서 독립된 표본 조사구를 사용하여 이를 해결하는 것이 가장 좋은 방법이다. 다만 가구 접촉의 어려움을 가만하여 이중추출방법을 그대로 사용하는 경우, 분산 추정에 대한 방법론적 추가 연구가 계속 되어야 할 것이다.



참고문헌

김영원 · 박민규 · 정재기(2010), 『생활시간조사 방법론 개선 연구』, 통계청
김태현 · 김정석 · 이기재 · 고영미(2010), 『생활시간조사 2010년 정기통계품질진단』, 통계청
이윤석 · 김규성 · 이희길 · 심수진(2008), 『생활시간조사 방법론 개선방안』, 통계개발원
통계청(2010), 『2009년 생활시간조사 보고서』

기상청, <http://www.kma.go.kr>

Australian Bureau of Statistics, <http://www.abs.gov.au>

Statistics Japan, <http://www.stat.go.jp>

US Bureau of Labor Statistics, <http://www.bls.gov>