

연구보고서 2019-42

이혜정 · 이계오 · 손창균 · 박승환 · 신재동

【책임연구자】

오미애 한국보건사회연구원 연구위원

【주요 저서】

기계학습(Machine Learning)기반 사회보장 빅데이터 분석 및 예측모형 연구
한국보건사회연구원, 2017(공저)

기계학습(Machine Learning)기반 이상 탐지(Anomaly Detection) 기법 연구
-보건사회 분야를 중심으로
한국보건사회연구원, 2018(공저)

【공동연구진】

이혜정 한국보건사회연구원 부연구위원

이계오 한국궐렵조사연구소

손창균 동국대학교 빅데이터·응용통계학과 교수

박승환 강원대학교 정보통계학과 교수

신재동 한국보건사회연구원 전문원

연구보고서 2019-42

2019년 한국복지패널 심층분석

발 행 일 2019년 12월
저 자 오 미 애
발 행 인 조 흥 식
발 행 처 한국보건사회연구원
주 소 [30147]세종특별자치시 시청대로 370
세종국책연구단지 사회정책동(1~5층)
전 화 대표전화: 044)287-8000
홈페이지 <http://www.kihasa.re.kr>
등 록 1994년 7월 1일(제8-142호)
인 쇄 처 (주)현대아트컴

© 한국보건사회연구원 2019
ISBN 978-89-6827-673-6 93510

한국복지패널 조사(Korea welfare panel study, KOWEPS)는 외환 위기 이후 빈곤층, 근로빈곤층, 차상위층의 가구 형태, 소득 수준, 취업 상태가 급격히 변화하고 있는 상황에서 이러한 계층의 규모 및 생활실태 변화를 동태적으로 파악함으로써 정책 형성에 기여하기 위한 조사이다.

2006년 1차 조사에서 7072가구가 조사되었고 6차 연도인 2011년 조사까지 5000가구 이상의 원패널 규모를 유지하여, 원패널 유지율이 약 75% 이상 되었다. 그러나 조사 거부 및 자연 손실 등으로 원표본이 지속적으로 감소할 뿐만 아니라 표본 탈락으로 인한 패널 표본 가구 분포상의 문제점도 제기됨에 따라, 이를 개선하여 최초 구축 당시의 표본 규모와 대표성을 유지할 수 있도록 하기 위하여 7차 연도인 2012년에 1800가구의 신규 패널을 추가하였다.

해마다 동일한 조사를 반복하는 패널 조사의 특성상 이와 같은 패널 마모는 불가피하다. 그러나 원표본 유지율이 60%대로 낮아져 표본의 대표성이 문제가 되고 표본의 보완 필요성이 제기됨에 따라, 표본 관리 및 모집단 추정의 정도를 기하기 위한 마모패턴별 가구 특성 비교 분석 연구가 필요한 상황이다. 이러한 배경에서 원패널과 신규 패널의 가중치 산출, 마모패턴별 가구 특성 비교, 가중치 변화 분석, 가구 소득의 추정 방식에 따른 차이 비교에 관한 연구를 수행하고자 하였다.

본 연구를 위해 조언을 해 주신 많은 전문가들과 원고 집필에 참여해 주신 교수님들께 감사드린다.

끝으로 본 보고서에 수록된 모든 내용은 우리 연구원의 공식적인 견해가 아니며 연구에 참여한 연구진의 의견임을 밝힌다.

2019년 12월

한국보건사회연구원 원장

조 흥 식

목 차

Abstract	1
요 약	3
제1장 서 론	9
제1절 연구 배경 및 목적	11
제2절 연구 내용 및 방법	13
제2장 표본조사의 가중치 개념	15
제1절 표본조사에서 가중치 계산	17
제2절 표본조사에서 자체 가중	20
제3장 패널 조사의 가중치	23
제1절 한국노동패널의 가중치	25
제2절 한국복지패널의 가중치	32
제3절 소결	42
제4장 마모패턴별 특성 비교 분석	45
제1절 패널 마모패턴 분석	47
제2절 패널 마모패턴별 패널 가구 특성 분석	52
제3절 차수별 횡단 가중치 분포 분석	66
제4절 패널 마모가 가구 소득(가처분소득/경상소득)에 미치는 영향 분석	79
제5절 소결	97

제5장 종단면 결측 패턴에 대한 통계적 가설 검정 101

제1절 종단적 결측자료 이슈 103

제2절 결측자료 모형 106

제3절 완전랜덤결측 검정 112

제4절 한국복지패널 사례 분석 119

제5절 결측 패턴에 대한 통계적 검증 143

제6절 소결 166

제6장 결 론 169

참고문헌 175

부 록 179

부록 1. 기존 패널 마모패턴 분석 179

부록 2. 차수별 일반/저소득 가구 총가구원 수 분석 결과 182

부록 3. 분가 가구의 특성 분석 186

표 목차

〈표 1-1-1〉 1~13차 한국복지패널 조사 표본 가구 현황 및 원표본 유지율	12
〈표 3-1-1〉 1998년 노동패널 조사의 시·도별 표본 할당 결과	26
〈표 3-1-2〉 시·도별 동부/읍·면부별 추가 패널 구축 현황	28
〈표 3-2-1〉 일반 가구와 저소득 가구의 분류 기준	33
〈표 3-2-2〉 KOWEPS 1차 조사의 지역별 조사구와 가구 분포 현황	34
〈표 3-2-3〉 KOWEPS 7차 조사 신규패널 지역별 표본 배분 결과	36
〈표 4-1-1〉 기존 패널 마모패턴별 원가구 범주 구분	48
〈표 4-1-2〉 기존 패널 차수별 마모패턴별 원가구 가구 수	49
〈표 4-1-3〉 신규 패널 마모패턴별 가구 수	51
〈표 4-1-4〉 신규 패널 마모패턴별 원가구 범주 구분	51
〈표 4-1-5〉 신규 패널 차수별 마모패턴별 원가구 가구 수	52
〈표 4-2-1〉 연간 자료의 마모패턴별 패널 가구 수	53
〈표 4-2-2〉 기존 패널 마모패턴별 저소득 가구 비율	55
〈표 4-2-3〉 신규 패널 마모패턴별 저소득 가구 비율	56
〈표 4-2-4〉 차수별 총가구원 수	58
〈표 4-2-5〉 차수별 총가구원 수 비율	59
〈표 4-2-6〉 차수별 가구 형태 범주 가구 수	61
〈표 4-2-7〉 차수별 가구 형태 범주 가구 수 비율	62
〈표 4-2-8〉 차수별 주택 유형 범주 가구 수	64
〈표 4-2-9〉 차수별 주택 유형 범주 가구 수 비율	65
〈표 4-3-1〉 차수별 기존 패널 전체 가구 가중치 분포	67
〈표 4-3-2〉 차수별 기존 패널 일반 가구 가중치 분포	69
〈표 4-3-3〉 차수별 기존 패널 저소득 가구 가중치 분포	71
〈표 4-3-4〉 차수별 신규 패널 전체 가구 가중치 분포	73
〈표 4-3-5〉 차수별 신규 패널 일반 가구 가중치 분포	75
〈표 4-3-6〉 차수별 신규 패널 저소득 가구 가중치 분포	77

〈표 4-4-1〉 가구별, 패널별, 차수별 가처분소득 비가중 평균	80
〈표 4-4-2〉 기존 패널 마모패턴별 가처분소득 비가중 평균	81
〈표 4-4-3〉 신규 패널 마모패턴별 가처분소득 비가중 평균	82
〈표 4-4-4〉 가구별, 패널별, 차수별 가처분소득 가중 평균	84
〈표 4-4-5〉 기존 패널 마모패턴별 가처분소득 가중 평균	85
〈표 4-4-6〉 신규 패널 마모패턴별 가처분소득 가중 평균	87
〈표 4-4-7〉 가구별, 패널별, 차수별 경상소득 비가중 평균	90
〈표 4-4-8〉 기존 패널 마모패턴별 경상소득 비가중 평균	91
〈표 4-4-9〉 신규 패널 마모패턴별 경상소득 비가중 평균	92
〈표 4-4-10〉 가구별, 패널별, 차수별 경상소득 가중 평균	93
〈표 4-4-11〉 기존 패널 마모패턴별 경상소득 가중 평균	94
〈표 4-4-12〉 신규 패널 마모패턴별 경상소득 가중 평균	96
〈표 5-3-1〉 개인-시점 자료의 예	117
〈표 5-3-2〉 공변량을 가진 개인-시점 자료의 예	118
〈표 5-4-1〉 KOWEPS의 횡단면 가구원 수	120
〈표 5-4-2〉 KOWEPS와 KOSIS의 가구주 성별 분포	122
〈표 5-4-3〉 KOWEPS와 KOSIS의 가구주 연령 분포 비교(2016년 기준)	123
〈표 5-4-4〉 KOSIS와 KOWEPS의 가구주 혼인 상태 비교	124
〈표 5-4-5〉 KOWEPS와 KOSIS의 경상소득 비교	126
〈표 5-4-6〉 1회 이상 패널에 참여한 가구 수	128
〈표 5-4-7〉 지속적 참여 여부에 대한 패널 가구의 소득 비교	128
〈표 5-4-8〉 가구주 변동 유무와 패널 참여 횟수	129
〈표 5-4-9〉 가구주 변동 유무와 가구 소득	130
〈표 5-4-10〉 가구주 변동 유무와 패널 참여 횟수별 가구 소득 비교	130
〈표 5-4-11〉 패널 참여 횟수별 가구원 수	131
〈표 5-4-12〉 가구원 수 변동 유무와 가구 소득	132
〈표 5-4-13〉 가구주 변동 유무와 패널 참여 횟수별 가구 소득 비교	133
〈표 5-4-14〉 가구원 수 변화에 따른 경상소득 비교	134

〈표 5-5-1〉 웨이브 간 결측 패턴에 따른 가구 수 분포	144
〈표 5-5-2〉 가구원 수에 따른 결측 횟수에 대한 ANOVA	145
〈표 5-5-3〉 가구원 수에 따른 결측에 대한 사후 분석	145
〈표 5-5-4〉 지역에 따른 결측 횟수에 대한 ANOVA	146
〈표 5-5-5〉 지역에 대한 사후 분석	146
〈표 5-5-6〉 가구주 혼인 상태에 따른 결측 횟수에 대한 ANOVA	147
〈표 5-5-7〉 가구주 혼인 상태에 대한 사후 분석	147
〈표 5-5-8〉 가구 소득 따른 결측 횟수에 대한 ANOVA	148
〈표 5-5-9〉 가구 소득에 대한 사후 분석	148
〈표 5-5-10〉 가구주 연령에 따른 결측 횟수에 대한 ANOVA	149
〈표 5-5-11〉 가구주 연령에 대한 사후 분석	149
〈표 5-5-12〉 가구주 교육 수준에 따른 결측 횟수에 대한 ANOVA	150
〈표 5-5-13〉 가구주 교육 수준에 대한 사후 분석	150
〈표 5-5-14〉 가구주 경제활동 상태에 따른 결측 횟수에 대한 ANOVA	151
〈표 5-5-15〉 가구주 경제활동 상태에 대한 사후 분석	151
〈표 5-5-16〉 가구주 건강 상태에 따른 결측 횟수에 대한 ANOVA	152
〈표 5-5-17〉 가구주 건강 상태에 대한 사후 분석	152
〈표 5-5-18〉 분석변수 목록	154
〈표 5-5-19〉 (모형 1)에 대한 모수에 대한 검정	155
〈표 5-5-20〉 (모형 1)에 대한 로지스틱 분석	156
〈표 5-5-21〉 (모형 2)에 대한 모수에 대한 검정	157
〈표 5-5-22〉 (모형2)에 대한 로지스틱 분석	158
〈표 5-5-23〉 (모형 3)에 대한 모수에 대한 검정	161
〈표 5-5-24〉 (모형 3)에 대한 로지스틱 분석	162
〈표 5-5-25〉 (모형 4)에 대한 모수에 대한 검정	163
〈표 5-5-26〉 (모형 4)에 대한 로지스틱 분석	163

부표 목차

〈부표 1-1-1〉 기존 패널 마모패턴별 가구 수	179
〈부표 2-1-1〉 차수별 일반 가구 총가구원 수	182
〈부표 2-1-2〉 차수별 일반 가구 총가구원 수 비율	183
〈부표 2-1-3〉 차수별 저소득 가구 총가구원 수	184
〈부표 2-1-4〉 차수별 저소득 가구 총가구원 수 비율	185
〈부표 3-1-1〉 분가 가구 가구 구분	186
〈부표 3-1-2〉 분가 가구 가구원 수	186
〈부표 3-1-3〉 분가 가구 가구 형태	187

그림 목차

[그림 3-2-1] KOWEPS의 가중치 부여 체계	37
[그림 4-2-1] 마모패턴별 저소득 가구 비율	56
[그림 4-3-1] 기존 패널 전체 가구 1~13차 조사 가중치(10,000까지)	68
[그림 4-3-2] 기존 패널 일반 가구 1~13차 조사 가중치(10,000까지)	70
[그림 4-3-3] 기존 패널 저소득 가구 1~13차 조사 가중치(10,000까지)	72
[그림 4-3-4] 신규 패널 전체 가구 1~13차 조사 가중치(10,000까지)	74
[그림 4-3-5] 신규 패널 일반 가구 1~13차 조사 가중치(10,000까지)	76
[그림 4-3-6] 신규 패널 저소득 가구 1~13차 조사 가중치(10,000까지)	78
[그림 4-4-1] 통합 패널 가구별 가처분소득 가중/비가중 평균	86
[그림 4-4-2] 기존/신규 패널 가구별 가처분소득 비가중 평균	87
[그림 4-4-3] 기존/신규 패널 가구별 가처분소득 가중 평균	88
[그림 4-4-4] 통합 패널 가구별 경상소득 가중/비가중 평균	95
[그림 4-4-5] 기존/신규 패널 가구별 경상소득 비가중 평균 비교	96
[그림 4-4-6] 기존/신규 패널 가구별 경상소득 가중 평균 비교	97
[그림 5-4-1] KOWEPS와 KOSIS의 가구원 수 분포 비교	121
[그림 5-4-2] KOWEPS와 KOSIS의 가구주 연령 분포 비교(2016년 기준)	123
[그림 5-4-3] KOWEPS와 KOSIS의 가구주 혼인 상태 분포 비교	125
[그림 5-4-4] KOWEPS와 KOSIS의 가구 평균 소득 분포 비교	126
[그림 5-4-5] KOWEPS 차수별 유지 현황	127
[그림 5-4-6] 가구 변동 효과에 대한 분석	136
[그림 5-4-7] 웨이브 효과에 대한 분석	136
[그림 5-4-8] 가구주 혼인 상태 변동 효과 분석	137
[그림 5-4-9] 가구주 연령 변동 효과 분석	138
[그림 5-4-10] 가구원 수 변동 효과 분석	139
[그림 5-4-11] 가구 변동 효과 분석	140
[그림 5-4-12] 웨이브 효과에 대한 분석	141



[그림 5-4-13] 혼인 상태 변화 효과에 대한 분석 141

[그림 5-4-14] 가구주 연령 변동 효과 분석 142

[그림 5-4-15] 전체 변동 효과 분석 143

Abstract <<

An In-depth Study of Korea Welfare Panel

Project Head: Oh, Miae

This study attempts to give an overview of the weighting process of the Korean Welfare Panel. We studied several household characteristics related to panel attrition, examining how they may have conditioned the panel data in the Korea Welfare Panel Study (KOWEPS). We also studied the cause of the differences in household income between the original and new panels in KOWEPS.

The panel attrition in the KOWEPS data is monotonous and affects household characteristics and the estimation of household income. The weight of low-income household tends to decrease over the years, presumably a consequence of an overall increase in household income. Such changes are reflected in the pattern in which older panels have higher estimates of household income. The characteristics of longitudinal area missing patterns of Korean welfare panel households were examined.

Based on this in-depth analysis report, the Korea Welfare Panel should continue to study various statistical issues including quality improvement.

*Key words : panel attrition, panel conditioning, design weight, design effect

Co-Researchers: Lee, Hyejung · Lee, Kayo · Son, Changkyoon · Pack, Seunghwan · Shin, Jaedong

1. 연구의 배경 및 목적

한국복지패널 조사(Korea welfare panel study, KOWEPS)는 외환 위기 이후 빈곤층, 근로빈곤층, 차상위층의 가구 형태, 소득 수준, 취업 상태가 급격히 변화하고 있는 상황에서 이러한 계층의 규모 및 생활실태 변화를 동태적으로 파악함으로써 정책 형성에 기여하기 위한 조사이다.

2006년 1차 조사에서 7072가구가 조사되었고 6차 연도인 2011년 조사까지 5000가구 이상의 원패널 규모를 유지하여, 원패널 유지율이 약 75% 이상 되었다. 그러나 조사 거부 및 자연 손실 등으로 원표본이 지속적으로 감소할 뿐만 아니라 표본 탈락으로 인한 패널 표본 가구 분포상의 문제점도 제기됨에 따라, 이를 개선하여 최초 구축 당시의 표본 규모와 대표성을 유지할 수 있도록 하기 위하여 7차 연도인 2012년에 1800가구의 신규 패널을 추가하였다. 2018년 13차 조사 때의 원표본 가구 수는 4266으로 원표본 유지율은 60.3%이며 신규 패널의 원표본 가구 수는 1392로 원표본 유지율은 77.3%이다.

해마다 동일한 조사를 반복하는 패널 조사의 특성상 이와 같은 패널 마모는 불가피하다. 그러나 원표본 유지율이 60%대로 낮아져 표본의 대표성이 문제가 되고 표본의 보완 필요성이 제기됨에 따라, 표본 관리 및 모집단 추정의 정도를 기하기 위한 마모패턴별 가구 특성 비교 분석 연구가 필요한 상황이다. 이러한 배경에서 원패널과 신규 패널의 가중치 산출, 마모패턴별 가구 특성 비교, 가중치 변화 분석, 가구 소득의 추정 방식에 따른 차이 비교에 관한 연구를 수행하고자 하였다.

2. 주요 연구 결과

한국복지패널의 조사 연차가 6년 이상 지나면서 패널 탈락률이 커짐에 따라 표본 대표성을 보완하기 위해 1800가구를 신규 패널로 추가하였다. 이에 따라 처음부터 응답을 계속해 온 원패널과 신규 추가 패널을 합한 통합 패널을 대상으로 가중치 산출과 패널의 변동성에 대한 심층 연구를 실시하였다. 이를 위해 2장에서는 가중치의 개념을 설명하고 표본조사에서 가중치 산출 단계를 자세히 살펴보았다. 한국복지패널은 2012년에 신규 패널을 추가하였는데, 한국노동패널은 복지패널이 신규 표본을 추가하기 이전에 이러한 경험(2009년)을 가지고 있으며 원패널 구성과 통합 패널 구성에 대한 연구 사례가 있다. 이 부분은 3장에서 가중치 산출과 관련하여 한국노동패널과 한국복지패널을 설명하였다.

4장에서는 한국복지패널의 패널 마모패턴에 따른 가구 특성의 차이에 대한 분석, 차수별 가구별 가중치 분포 분석, 가중/비가중 가구 소득 추정 결과 분석을 진행하였다. 분석 결과, 저소득 가구는 패널에 잔류할 가능성이 큰 반면, 소득이 높은 가구 혹은 3인 이상 가구는 패널에서 이탈할 가능성이 컸다. 이러한 시간에 따른 패널 마모로 인하여 가중치 값이 증가함을 알 수 있고, 저소득 가구 가중치 증가량은 일반 가구에 비하여 적게 나타남을 알 수 있다.

이 장에서는 한국복지패널에서 패널 마모패턴에 따른 가구 특성 차이 분석을 통해 패널조건화 현상을 살펴보았다. 가구 소득 추정에 있어서 기존 패널과 신규 패널의 차이 발생 여부를 살펴보고 차이가 발생하는 원인이 패널 마모에 따른 가구 특성 차이 때문일 수 있음을 보였다. 한국복지패널의 패널 마모는 단조마모패턴으로 발생하며, 가구의 특성과 가구 소득 추정에 영향을 미친다. 패널이 오래될수록 기존 패널에서는 일반 가구

비율이 증가하지만 신규 패널에서는 차수가 지나도 그 비율이 유사한 값으로 유지된다. 이로 인하여 기간이 긴 기존 패널의 가구 소득은 기간이 짧은 신규 패널의 가구 소득보다 더 크게 추정되는 경향이 있다고 추론할 수 있다. 패널 조사에서 패널조건화 현상은 불가피하긴 하지만 조사 결과를 바탕으로 한 통계적 추론에 있어 패널조건화로 인한 왜곡이 발생하지 않아야 한다.

5장에서는 패널 가구의 종단적 결측이 어떠한 원인에 의해 발생하는지를 통계적 관점에서 살펴봄으로써 단순히 종단 결측을 다룰 경우 자료의 왜곡이 발생할 수 있다는 사실을 확인할 수 있었다. 즉, 패널 데이터의 분석에서 패널 가구의 결측 패턴에 부합되는 분석이 실시되어야 한다는 것이다. 단순히 종단적으로 결측된 모든 가구를 제거하고 완전하게 응답된 가구만을 대상으로 분석하는 것은 MCAR 결측 가정에서만 가능한 것이다. 따라서, 이 연구에서와 같이 결측 모형에 대한 검증을 실시한 후에 그에 적합한 분석이 이루어질 경우 보다 신뢰성 있는 분석 결과를 도출할 수 있을 것이다.

3. 결론 및 시사점

한국복지패널 데이터 자체에 대한 심층분석은 다양한 이슈를 중심으로 이루어질 수 있다.

한국노동패널의 경우 매해 기초분석보고서, 노동패널자료 연구, 패널자료 품질개선 연구, 워킹페이퍼 시리즈, 패널브리프, 학술대회 논문자료집 등이 연구 결과물로 발간되고 있다. 기초분석보고서는 한국복지패널 기초분석보고서와 동일하게 주요 변수에 대해 조사 차수별 기초 통계

분석 결과가 정리되어 있어 자료를 사용하는 연구자가 수치를 검증하는데 기초 자료로 사용할 수 있다. 노동패널자료 연구는 다른 패널자료와의 비교를 통해서 주제에 따른 수치 비교, 개선 방향 등을 연구한 보고서이며, 패널자료 품질개선 연구는 실험 연구, 표본 이탈, 항목 무응답 처리 방법, 표본 추가 등과 관련된 내용을 다룬 보고서이다.

한국복지패널도 이번 심층분석 보고서를 기점으로 하여, 품질 개선을 포함한 다양한 통계적 이슈를 다룰 수 있는 연구가 지속되어야 하겠다. 앞으로 한국복지패널의 심층연구가 필요한 주제를 몇 가지 제안해 보면 다음과 같다. 한국복지패널은 13년 이상 된 중요한 패널로 세대 간 이전 연구가 가능하다. 이를 위해서는 한국복지패널 분가 가구가 원가구와 비교했을 때 어떠한 특징이 있는지에 대한 연구도 심층적으로 이루어져야 한다(부록 3 참고). 또한, 향후 신규 표본 추가를 위한 준비도 철저하게 되어 있어야 한다. 신규 표본 추가는 이전의 방법처럼 double sampling이 가능하지 않기 때문에, 어떻게 보정해 주어야 하는지에 대한 고민도 필요하다. 그리고, 다른 패널들에서 패널을 운영하면서 겪어온 경험 및 이슈를 공유할 수 있는 네트워크 구성도 필수적일 것이다. 한국복지패널의 품질 개선을 위해서 paradata 관리 틀도 재정립할 필요가 있다. 조사 방법과 관련해서도 시대의 흐름을 따라 웹(Web)을 활용한 조사(CAWI: Computer Assisted Web Interviewing)를 고려해 볼 필요가 있다. 한국노동패널 조사는 CAWI 효과에 대한 실험 연구를 2년에 걸쳐 진행하였고, 2018년 본 조사에서 이메일 주소를 가지고 있는 강력 거절 응답자만을 한정하여 CAWI 조사를 실시하였으며, 점진적으로 CAWI 병행도 고려하고 있다. 한국복지패널 역시 조사 방법의 다양화 문제를 검토해 볼 필요가 있다. 마지막으로, 데이터 연계에 대한 관심이 높아지고 있는 시점에, 한국복지패널과 다른 자료와의 연계로 새로운 가치 창출을 할 수

있는 연구도 필요하다.

이러한 다양한 이슈들에 대한 한국복지패널 연구가 활발히 이루어진다면, 기존 데이터 생산 및 관리 사업에 더해 향후 한국복지패널 조사 연구의 안정적인 수행에 기여할 수 있으며, 데이터 품질을 개선하고 향상시키는 데 중요한 역할을 할 수 있을 것이다.

*주요 용어: 가중치, 통합표본, 패널 마모패턴, 결측패턴, 통계적 검증

제 1 장

서론

제1절 연구 배경 및 목적

제2절 연구 내용 및 방법

제1절 연구 배경 및 목적

한국복지패널 조사(KOWEPS: Korea welfare panel study)는 외환 위기 이후 빈곤층, 근로빈곤층, 차상위층의 가구 형태, 소득 수준, 취업 상태가 급격히 변화하고 있는 상황에서 이러한 계층의 규모 및 생활실태 변화를 동태적으로 파악함으로써 정책 형성에 기여하기 위한 조사이다.

한국복지패널 조사는 2006년에 1차 조사를 시작하여 2018년에 13차 조사를 완료하였다. 2006년 1차 조사에서 7072가구가 조사되었고 6차인 2011년 조사까지 5000가구 이상의 원패널 규모를 유지하여 원패널 유지율이 약 75% 이상 되었다. 그러나 조사 거부 및 자연 손실 등으로 원표본이 지속적으로 감소할 뿐만 아니라 표본 탈락으로 인한 패널 표본 가구 분포상의 문제점도 제기됨에 따라, 이를 개선하여 최초 구축 당시의 표본 규모와 대표성을 유지할 수 있도록 하기 위하여 7차 연도인 2012년에 1800가구의 신규 패널을 추가하였다. 2018년 13차 조사 때의 원표본 가구 수는 4266으로 원표본 유지율은 60.3%이며 신규 패널의 원표본 가구 수는 1392이다.

〈표 1-1-1〉 1~13차 한국복지패널 조사 표본 가구 현황 및 원표본 유지율

(단위: 가구, %)

조사 차수		조사 완료 가구 수	표본 유지 가구 수 (원가구)	원표본 유지율 (%)	신규 패널 조사 완료 가구 수	신규 패널 표본 유지 가구 수
1차	2006년	7,072	7,072	100.00	-	-
2차	2007년	6,580	6,511	92.07	-	-
3차	2008년	6,314	6,128	86.65	-	-
4차	2009년	6,207	5,935	83.92	-	-
5차	2010년	6,034	5,675	80.25	-	-
6차	2011년	5,735	5,335	75.44	-	-
7차	2012년	5,732	5,271	74.53	1,800	1,800
8차	2013년	5,619	5,104	72.17	1,693	1,690
9차	2014년	5,438	4,896	69.23	1,610	1,594
10차	2015년	5,343	4,760	67.31	1,571	1,534
11차	2016년	5,189	4,560	64.48	1,534	1,478
12차	2017년	5,081	4,398	62.19	1,500	1,426
13차	2018년	4,997	4,266	60.32	1,477	1,392

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

패널 조사는 해마다 동일한 가구를 대상으로 동일한 조사를 반복하기 때문에 표본의 마모가 불가피하다. 한국복지패널 11차 조사 기준 원표본 규모는 4560가구로 원표본 유지율이 64.48%로 낮아져 표본의 대표성이 문제가 되고 표본의 보완 필요성이 제기됨에 따라, 표본 관리 및 모집단 추정의 정도를 기하기 위한 가중치 관련 심층 연구가 필요한 상황이다.

2019년 한국복지패널 사업에서는 표본 관리 및 가중치 관련 심층 연구를 수행함으로써 지난 10여 년의 한국사회의 동태적 특성을 파악하고, 한국복지패널 데이터의 통계적 안정성을 제고할 수 있는 방안을 도출하고자 한다. 심층분석 연구의 병행은 기존 데이터 생산 및 관리 사업에 더해 향후 한국복지패널 조사 연구의 안정적인 수행에 기여할 수 있을 것으로 기대한다.

제2절 연구 내용 및 방법

패널 조사는 동일한 가구를 대상으로 계속해서 조사를 수행하는 방식이므로 조사가 거듭될수록 패널 마모가 발생하는 것은 불가피하다. 패널 마모는 이사, 가구 분화 혹은 가구 확장 등과 같은 패널 가구의 변화 때문에 발생하기도 하고, 혹은 장기간 조사 응답에 의한 피로감 때문에 생기기도 한다. 어떤 이유이든 원표본 수의 감소는 패널 표본의 대표성을 저하시키고 또한 표본 수 감소로 인한 통계 결과의 신뢰도 저하를 가져온다. 이러한 점은 패널 조사가 안고 있는 피하기 어려운 본질적인 단점이기도 하다. 이는 가중치와도 밀접한 관련이 있다. 이를 위해 2장에서는 가중치의 개념을 설명하고, 3장에서는 한국노동패널과 한국복지패널의 가중치 산출 방법을 살펴본다.

4장에서는 원패널과 신규 패널 간의 마모패턴별 가구 특성 비교, 차수별 및 두 패널의 가중치 변화 분석, 가구 소득의 추정 방식에 따른 차이 비교에 관한 연구를 수행하였다. 첫째, 1~13차 원패널과 신규 패널 데이터를 대상으로 패널 마모패턴을 분석한다. 두 패널의 마모패턴별 가구 수의 분포를 살펴보고 마모패턴이 패널의 인구·사회학적 변인에 어떤 영향을 주는지 살펴본다. 패널 가구 특성으로 균등화 소득에 따른 가구 구분, 건강보험 가입 여부, 총가구원 수, 가구 형태, 주택 유형, 가구 총소득의 6개 변수에 대하여 패널 마모패턴이 어떤 영향을 주는지 살펴본다.

둘째, 1~13차 원패널과 신규 패널 데이터를 대상으로 패널 마모에 따른 가중치 분포의 변화를 살펴본다. 일반 가구와 저소득 가구를 나누어 차수별 가중치 분포의 변화를 분석하도록 한다.

셋째, 1~13차 원패널과 신규 패널 데이터를 대상으로 패널 마모가 가구 소득 추정에 미치는 영향을 가중치를 중심으로 살펴본다. 분석 대상

변수로 가구 가처분소득과 가구 경상소득을 사용한다. 두 변수에 대하여 가구 소득은 가중 평균과 비가중 평균을 구하여 차이가 발생하는지 살펴 보도록 한다.

5장에서는 종단적 결측 패턴 분석을 위한 통계적 개념 설명과 이론을 다룬다. 결측자료 모형의 정의와 결측자료에 대한 기본 가정을 완전랜덤 결측, 랜덤결측, 비랜덤결측으로 나누어 살펴보고자 한다. 한국복지패널 가구의 횡단적 변화를 객관적 비교를 위해 통계청 자료와 함께 제시한다. 한국복지패널의 종단적 변화는 가구원 변동과 소득 변화를 함께 분석하였다. 한국복지패널 가구원 수 변동의 요인 분석, 가구 소득 변동의 요인 분석을 별도로 실시하고, 결측 패턴에 대한 통계적 검증을 여러 특성을 고려하여 분석하고자 한다.

이 연구를 위해 R, 엑셀, SAS 등 다양한 프로그램을 활용하였으며, 수차례의 전문가 자문회의를 거쳐 연구 방향을 설정하고 연구를 진행하였다.

제 2 장

표본조사의 가중치 개념

제1절 표본조사에서 가중치 계산

제2절 표본조사에서 자체 가중

2

표본조사의 가중치 개념 <<

제1절 표본조사에서 가중치 계산

표본조사는 모집단의 일부를 표본으로 선정하여 조사한 후에 수집된 데이터로부터 모집단의 특성을 파악해야 하므로 조사 단위별로 관찰된 값이 모집단의 몇 개의 조사 단위에 대응되는지를 수치로 나타내고 이를 적용하여 모집단 특성을 나타내게 된다. 이때 표본조사 단위와 모집단 조사 단위들을 연계하는 수치를 가중치(weight)라 한다.

일반적으로 가중치의 산출은 3단계를 통해서 계산하게 되는데 첫 단계는 표본추출 과정을 반영한 설계 가중치로 표본추출률의 역수로 계산하고, 두 번째 단계는 조사 과정에서 응답 상태를 반영하는 무응답 보정의 가중치로 응답률의 역수로 계산한다. 3번째 단계는 사후층화 보정 가중치로 앞선 단계에서 구한 설계 가중치와 무응답 보정 가중치를 곱한 가중치 합계의 분포(성별*연령대별)와 모집단의 분포(성별*연령대별)를 일치시키기 위한 보정계수를 의미한다. 최종 가중치는 위에서 설명한 3단계에서 구한 각각의 값을 곱하여 산출한다.

설계 가중치의 산출은 모집단에서 표본을 추출하는 모든 과정에서 표본으로 추출할 확률(표본추출 확률)의 역수들을 곱해서 계산한다. 예를 들어 도시와 농촌지역으로 이루어진 도·농복합시에서 취업률을 조사한다고 가정하자. 이때 도시지역은 400개 조사구로 구성되고 농촌지역은 150개 조사구로 구성되었으며 또한 도시와 농촌지역에서 각각 20개, 10개의 조사구를 단순 확률로 추출하기로 한다. 선정된 각 조사구 내에서

도시지역은 10가구씩 추출하고 농촌지역은 5가구씩을 선정하여 조사 대상으로 추출된 가구 내의 15세 이상 모든 사람의 취업 여부를 조사한다면 이는 층화2단계집락추출법이 되며 이 경우의 조사된 모든 사람들의 가중치는 아래와 같이 계산할 수 있다.

1차로 도시지역과 농촌지역을 층으로 간주하고 각 층별 가중치는 조사구 추출률의 역수와 각 조사구 내 가구 추출률의 역수를 곱하여 계산한다.

가중치 산출 과정을 수식으로 나타내는 데 필요한 기호를 다음과 같이 정의하자.

- N_h : h 층의 모집단 조사구 수($h = 1$ 은 도시지역, $h = 2$ 는 농촌 지역을 나타냄)
- n_h : h 층의 표본조사구 수
- M_{hi} : h 층 내의 i 번째 조사구의 모든 가구 수
- m_{hi} : h 층 내의 i 번째 조사구의 표본 가구 수(도시지역은 10가구이고 농촌지역은 5가구임)
- y_{hijk} : h 층 내의 i 번째 조사구 내 j 가구의 15세 이상 인구 중 k 번째 사람의 취업 여부를 조사한 값($y_{hijk}=1$ 취업자, $y_{hijk}=0$ 실업자)

h 층 내 i 번째 조사구 내 j 가구의 15세 이상 인구 중 k 번째 사람의 가중치를 W_{hijk} 라고 하면 아래 식으로 나타낼 수 있다.

$$W_{hijk} = (\text{조사구 추출률의 역수}) * (\text{조사구 내 가구 추출률의 역수})$$

$$= \left(\frac{N_h}{n_h} \right) \cdot \left(\frac{M_{hi}}{m_{hi}} \right) \dots\dots\dots (1)$$

다음으로 무응답 보정 가중치의 산출을 살펴보면 앞의 사례에서 조사 대상 가구는 모두 응답하였으나 조사 대상 가구 내의 15세 이상 사람 중에서 일부는 장기 출타 또는 다른 사유로 응답을 하지 않은 경우가 있을 것이다. 식(1)에 주어진 가중치 산출식은 표본 가구 내의 15세 이상 가구

원들은 모두 응답했다는 가정하에 개인별 가중치를 계산하는 데 유효하지만, 만일 가구원 중 일부가 무응답한 경우에는 무응답을 보정하는 가중치를 적용해야 할 것이다. 무응답 보정 가중치 산출의 전체조건은 무응답자와 응답자의 특성이 유사하다는 것이며 이 경우에 응답자의 정보를 무응답자의 정보로 대체하는 효과가 있다. 무응답 보정 가중치를 계산하기 위해서 G_{hij} 는 h 층 내의 i 번째 조사구 내 j 가구의 15세 이상 인구수라 하고 이 중에서 응답한 사람 수를 g_{hij} 라 한다면 설계 가중치에 무응답 보정 가중치를 곱해서 아래 식(2)로 가중치를 나타낼 수 있다.

$$W_{hijk}^* = W_{hijk} \cdot \frac{G_{hij}}{g_{hij}} = \left(\frac{N_h}{n_h} \right) \cdot \left(\frac{M_{hi}}{m_{hi}} \right) \cdot \left(\frac{G_{hij}}{g_{hij}} \right) \dots\dots\dots (2)$$

취업률이 성별, 연령대별로 차이가 있다면 모집단의 성별, 연령대별 인구의 분포가 주어졌을 경우 표본의 성별, 연령대별 분포를 모집단의 분포에 맞추는 조정을 해야 한다. 이때 조정인자를 사후층화 보정인자라 한다. 사후층화 보정인자의 계산은 식(2)에서 주어진 가중치를 성별, 연령대에 대해서 합산하고 계산 결과와 모집단 성별, 연령대의 인구수 간의 비(ratio)로 계산한다.

$$BF_{sa} = \frac{P_{sa}}{\sum_{h,i,j,k} I_{sahijk} W_{hijk}} = \frac{\text{모집단}_{sa}}{\text{표본가중치 } sa \text{에 대한합계}}$$

여기서 I_{sahijk} 는 일종의 지시함수이며 동일한 성별과 연령대이면 1이고 그 외의 경우에는 0이므로 지수함수 I_{sahijk} 와 가중치 W_{hijk} 를 곱해서 더하면 해당되는 성별과 연령대에 대해서만 가중치들의 합계를 계산하면

된다.

최종 가중치는 식(2)에 BF_{sa} 를 곱해서 산출하며 아래 식으로 주어진다.

$$W_{hijk}^f = BF_{sa} \cdot W_{hijk}^* \dots\dots\dots (3)$$

제2절 표본조사에서 자체 가중

표본조사에서 가중치를 산출하는 과정이 단순하지 않을 뿐 아니라 표본 가중치를 적용하여 모수추정을 할 경우에는 추정량의 분산 계산이 복잡하므로 가중치를 적용하지 않아도 제대로 분석할 수 있는 표본설계를 설명하고자 한다.

먼저 자체 가중 표본설계라는 것은 표본으로 선정된 모든 조사 단위들의 가중치가 동일한 값을 갖는 표본설계를 의미한다. 예를 들어 직원이 100명인 회사에서 10월 한 달간 초과근무시간의 합계를 추정하기 위해 20명을 단순확률추출법으로 선정하여 그들의 초과근무시간을 조사한 후에 전체 직원의 초과근무시간을 산출할 경우 아래 식을 사용할 수 있다.

$$\hat{Y} = \sum_{i=1}^{20} W_i y_i = \sum_{i=1}^{20} \frac{100}{20} \cdot y_i = \frac{100}{20} \sum_{i=1}^{20} y_i \dots\dots\dots (4)$$

식(4)에 주어진 가중치 W_i 는 100명 중에서 20명을 단순확률비복원추출법으로 선정하므로 $5(=100/20)$ 가 되고 표본으로 선정된 20명이 동일한 가중치를 갖게 되므로 자체 가중설계가 되며 식(4)의 마지막 우측 항과 같이 표본 관찰값들만의 함수로 표기할 수 있으므로 표본 데이터 분석이 용이하다.

다음에는 몇 가지 자체 가중 표본설계의 사례를 살펴보자.

(예제 1) 모집단이 H 개의 층으로 구성되었으며 각 층의 크기가 N_h 이고 전체 표본 크기가 n 이다. 또한 층별 표본 할당을 비례배분법으로 하고 각 층에서 할당된 표본을 단순확률추출법으로 선정한다. 여기서 모총계에 대한 추정량은 아래 식으로 표현할 수 있다.

$$\hat{Y}_{st} = \sum_{h=1}^H \sum_{i=1}^{n_h} W_{hi} y_{hi} = \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi} = \frac{N}{n} \sum_{h=1}^H \sum_{i=1}^{n_h} y_{hi} \quad \dots\dots\dots (5)$$

여기서 층별 표본 할당을 비례배분법으로 하면 $n_h = n \cdot \frac{N_h}{N}$ 가 성립하므로 $\frac{N_h}{n_h} = \frac{N}{n}$ 이 된다.

식(5)의 5번째 항에서 보는 바와 같이 모든 표본 관찰값이 동일한 값인 N/n 을 가중치로 갖게 되므로 자체 가중 표본설계이다.

(예제 2) A 개의 조사구로 구성된 도시지역에서 a 개 조사구를 조사구 내의 가구 수 기준으로 확률비례추출법으로 추출하고, 선정된 조사구에서는 b 가구를 랜덤추출법으로 선정한 후에 가구 내의 모든 가구원을 조사한다면 2단 집락추출법이 되며 또한 자체 가중 표본설계가 된다.

i 번째 조사구 내의 가구 수를 B_i 라고 한다면 확률비례추출법으로 1단계 추출에서 a 개 조사구를 선정할 확률은 $a \cdot \frac{B_i}{\sum_{i=1}^A B_i}$ 가 되고 2단계 추출

에서 선정된 조사구 내에서 b 가구를 추출할 확률은 $\frac{b}{B_i}$ 이므로 전체적인 표본추출 확률은 아래 식으로 표현할 수 있다.

$$f = f_1 \cdot f_{2i} = a \cdot \frac{B_i}{\sum_{i=1}^A B_i} \cdot \frac{b}{B_i} = \frac{ab}{\sum_{i=1}^A B_i} = \frac{n}{N} \quad \dots\dots\dots (6)$$

따라서 도시지역에서 모든 가구들이 표본으로 선정될 확률이 n/N 으로 동일하므로 자체 가중 표본설계가 된다.

위에서 설명한 자체 가중 표본설계의 장점은 조사된 관찰값만으로 모수추정량을 나타낼 수 있으므로 계산과 분석이 용이하고, 특히 분산추정이 용이할 뿐만 아니라 산출한 통계를 설명하기도 쉽다는 점이다. 그러나 다단계 층화집락추출법과 같은 복합적인 표본설계에서는 표본추출 단계별로 세부적인 정보가 필요하지만 실제로는 이런 정보를 얻을 수 없기 때문에 불균등확률추출법을 적용하게 된다.

제 3 장

패널 조사의 가중치

제1절 한국노동패널의 가중치

제2절 한국복지패널의 가중치

제3절 소결

제1절 한국노동패널의 가중치

한국복지패널은 2006년에 조사를 시작하였으며 빈곤층과 차상위계층 가구들의 소득 수준과 경제활동 상태 등의 변화와 변동 규모를 동적으로 분석할 수 있는 자료를 수집하는 데 목적을 두었다. 그러나 패널 조사의 특성상 조사 차수를 거듭할수록 패널 유지율이 낮아질 수밖에 없어 표본의 대표성을 유지하기 위해서는 추가 패널 구성이 필요하였다. 이에 따라 7년차에는 1800가구를 신규로 추가하였는데, 비슷한 예로 한국노동패널에서도 12년차(2009년)에 추가 패널을 구축한 적이 있다. 이에 따라 한국노동패널에서 신규 추가 패널을 구축한 후에 원패널과 신규 추가 패널을 통합하고 가중치를 산출하는 과정 등에 관한 부분을 살펴보고자 한다.

1. 한국노동패널의 표본 규모

가. 패널 원표본 가구 구축 과정¹⁾

1차 연도인 1998년에 노동패널 조사를 위해 진행한 표본추출 과정에 대해서 살펴보면 다음과 같다. 1차 연도의 노동패널 조사만을 위한 표본 설계는 연구되지 않았고 1997년 고용구조특별조사에 사용되었던 조사구를 추출 틀로 사용하였다.

1) “노동패널의 가중치 산출방안 : 원패널과 추가패널”(이계오·임찬수, 2012) 보고서의 패널 구축과정 및 가중치 산출과정의 부분을 인용(포괄적 인용)함.

노동패널의 조사 대상 지역을 7개 특·광역시와 8개 도에서 시 지역으로 제한하였다. 그러나 1997년 고용구조특별조사의 추출 틀은 1995년 인구주택총조사에서 표본조사로 사용했던 전체의 10% 조사구를 사용하였고 이에 1997년 고용구조특별조사의 전체 표본조사구 수는 2382개였다. 이 중에서 시 지역 조사구는 2025개였다.

1차 연도 한국노동패널 조사 표본 크기는 1000개로 하였고, 15개 시·도별 배분은 표본조사의 조사구 수에 비례하도록 하였으며 추출 틀의 규모는 전국도시조사구인 1만 9025개 조사구이다. 최종적으로는 시·도별로 할당된 결과와 1997년 고용구조특별조사의 도시지역 조사구 규모를 비교·검토하여 15개 시·도별 동부와 읍·면별로 표본 할당을 결정하였다. 예를 들면 서울특별시에는 1995년 10% 표본조사에 사용한 조사구 수가 5186개로 1만 9025개의 27.3%에 해당되므로 273개의 할당이 계산되었으나, 1997년 고용구조특별조사에서 서울시의 조사된 조사구 수는 227개이므로 227개를 서울시의 표본 규모로 결정하였다. 위와 같은 방법으로 시·도별로 조사구 수를 산출하였고 이를 요약하면 전체 표본 규모는 951개 조사구이다. 이렇게 시·도별 동부/읍·면부별로 할당된 조사구는 계통추출법으로 선정하였다. 15개 시·도별로 조사 모집단과 표본 할당 결과는 아래 <표 3-1-1>과 같다.

<표 3-1-1> 1998년 노동패널 조사의 시·도별 표본 할당 결과

(단위: 조사구, %)

시·도	서울	부산	대구	인천	광주	대전	울산	경기
고용구조특별조사구	227	176	155	150	106	101	70	280
할당 조사구	227	97	64	59	32	31	24	167
조사구 추출률	4.38	5.27	5.28	5.29	5.27	5.24	5.24	5.15
시·도	강원	충북	충남	전북	전남	경북	경남	합계
고용구조특별조사구	165	159	122	170	152	129	220	2382
할당 조사구	26	22	25	40	25	54	58	951
조사구 추출률	5.16	5.21	5.12	5.28	5.13	5.29	5.30	5.00

자료: 강석훈, (2003), KLIPS의 가중치 부여방안 연구. p4. <표 I-1> 재구성, 한국노동패널연구 2003-04

나. 패널 신규 표본 가구 구축 과정²⁾

1998년에 시작된 한국노동패널 조사가 2008년 11차 연도까지 이어져 오는 동안 중간에 이탈하지 않고 계속 응답한 가구 수는 2607가구로 원 표본 5000가구의 52.1%에 해당한다. 특히 1998년 1차 연도 조사 대상을 시지역으로 제한하여 패널을 구축하였으나 패널 가구가 시지역에서 군지역으로 이동하거나 군이 시로 승격하는 등의 모집단 변화를 반영하지 못함으로써 전국을 대표하는 데 한계가 있음이 지적되어 왔다.

지역 통계 생산의 요구를 반영하고 세부 영역에 대한 분석을 시도하기 위해서는 조사 대상을 군지역으로 확대하고 표본 규모도 늘릴 필요가 있다. 따라서 2008년 11차 연도에 응답한 4837가구의 특성을 동부/읍·면부별, 16개 시·도별 및 가구원 수별로 분석하여 현재 시점에서 조사 모집단을 대표할 수 있는지를 검토한 후 미흡한 부분에 대해 추가 패널을 구축하기 위한 연구를 하였다. 조사 모집단을 전국 가구로 확장하는 데 있어 가구원 수의 분포 변화도 동시에 반영할 수 있는 연구를 하였다.

추가 패널 구축 설계에 관한 내용은 2009년 연구 결과(김재광, 2009)를 요약 정리하였으며 이는 가중치 산출 방안 연구에 필요한 기초 정보이다. 추가 패널 구축의 기준이 되는 조사 모집단은 2005년 인구주택총조사 가구 데이터이다.

추가 패널 구축의 규모는 조사 비용의 수용성을 감안하여 1998년의 패널 조사 시작 당시와 같은 규모인 5000가구로 하였으며 이를 기준으로 11차 연도의 응답 가구와 2005년 인구주택총조사 가구의 16개 시·도별 동부/읍·면부별 분포를 비교한 후에 추가해야 할 가구 수를 계산하였다. 추가 표본은 조사구당 5가구를 추출한다는 가정하에 추출해야 할 조사구 수를 산출하였으며 16개 시·도별 동부/읍·면부별로 추가 표본 규모와 조

2) “노동패널의 가중치 산출방안 : 원패널과 추가패널”(이계오·임찬수, 2012) 보고서의 패널 구축과정 및 가중치 산출과정의 부분을 인용(포괄적 인용)함.

사구 수를 <표 3-1-2>에 정리하였다.

<표 3-1-2> 시·도별 동부/읍·면부별 추가 패널 구축 현황

(단위: 조사구)

시·도	동부			읍·면부		
	현행	추가	조사구	현행	추가	조사구
서울	1,303	185	37	0	0	0
부산	461	65	13	9	0	0
대구	330	20	4	18	5	1
인천	319	40	8	7	0	0
광주	180	25	5	0	0	0
대전	168	30	6	0	0	0
울산	107	25	5	20	5	1
경기	1,004	240	48	194	50	10
강원	120	25	5	29	60	12
충북	108	30	6	17	70	14
충남	80	30	6	95	85	17
전북	161	25	5	60	30	6
전남	102	15	3	51	125	25
경북	158	50	10	139	65	13
경남	234	40	8	108	85	17
제주	2	55	11	1	20	4
합계	4,837	900	180	748	600	120

자료: 김재광, (2009), 한국노동패널조사 표본추가 연구. p4~p5 <표4>, <표5> 재구성. 한국노동연구원 미발간보고서

표본조사구의 추출은 동부의 경우 가구원 1~2인 가구 수와 단독주택 수를 기준으로 확률비례추출법을 적용하여 계통추출법으로 선정하였고, 읍·면부는 조사구 내의 가구 수를 기준으로 확률비례추출법으로 선정하였다. 조사 대상 가구의 추출은 선정된 표본조사구 내에서 5가구를 계통추출법으로 선정하였다.

위 표에서 읍·면부가 없는 시·도인 서울, 광주와 대전에서는 읍·면부의 추가 표본이 없으나 군지역이 있는 부산, 대구, 인천, 울산에서 부산과 인천은 1998년에 구축한 원패널 가구들이 잘 유지되어 신규 패널 가구를 추가하지 않았다.

2. 한국노동패널의 가중치 산출 방법³⁾

한국노동패널 조사의 1차 연도 가중치는 종단 가중치와 횡단 가중치가 동일한 값이다. 가중치 계산은 1단계인 표본설계 가중치와 2단계인 무응답 보정 가중치 및 3단계인 사후층화 보정 가중치로 나누어 계산된다.

12차 연도 패널은 원패널과 추가 신규 패널로 구성되며 원패널은 1998년에 구축한 원패널과 분가 가구 패널로 구성된다. 12차 연도 패널의 가중치는 원패널 가중치와 추가 신규 패널 가중치로 구분하여 산출하였다.

12차 연도 조사에서 응답한 가구는 원패널 가구와 추가 패널 가구로 구분되지만 이를 통합한 통합 패널의 가중치는 표본 크기를 가중값으로 원패널과 추가 패널의 가중평균에서 계산할 수 있으며 아래 식으로 나타낼 수 있다.

$$\frac{n_{h0}}{n_h} \sum_{(ij) \in s_0} w_{0hij} + \frac{n_{h1}}{n_h} \sum_{(ij) \in s_1} w_{1hij} = M_h$$

여기서 n_{h0} 와 n_{h1} 은 각각 h 층의 원표본 크기와 추가 표본 크기를 나타내며 M_h 는 h 층의 2009년도 추계가구 수를 나타낸다.

통합 표본의 가중치는 추가 표본 가중치와 원표본 가중치에 가중치 조정인자와 추가 표본 크기와 원표본 크기의 구성비를 곱한 값을 다시 곱해서 산출한다.

통합 표본에서 개인 종단 가중치와 개인 횡단 가중치도 가구 횡단 가중치와 동일한 방법으로 산출하게 된다.

한국노동패널의 13차 조사 개인 가중치에 대한 설명은 다음과 같다.

3) “한국노동패널조사(KLIPS) 13차년도 기초분석보고서”(천영민, 최형아, 권순범, 이상희, 2012)의 한국노동패널의 가중치 부분을 인용(포괄적 인용)함.

13차 연도의 원패널 가중치 산출은 총 3단계(1단계 : 종단면 가구원 가중치 산출 → 2단계 : 가구 가중치 산출 → 3단계 : 횡단면 가구원 가중치 산출)의 과정으로 수행되었다.

가. 개인 종단면 가중치 산출 과정

종단면 가구원 가중치 산출은 기본적으로 무응답 보정의 과정을 통해 이루어진다. 무응답 보정을 위해 13차 연도 가구원의 응답 여부를 종속변수로, 12차 연도의 변수(성, 연령, 지역, 교육 수준, 주된 활동, 가구 소득)를 설명변수로 하는 로지스틱 회귀분석을 실시하고 이를 통하여 무응답 보정 가중치를 산출하였다. 이 단계에서 작성되는 가중치는 다음과 같다.

$$\hat{p}_i^{-1} w_i^{12}, \hat{p}_i = [1 + \exp(x_i' \hat{\beta})]$$

여기서 w_i^{12} 는 12차 연도에 개인(가구원) i 에 부여된 최종 종단면 가중치를 의미한다. 또한 x 는 로지스틱 회귀분석을 위해 사용된 설명변수 벡터를 나타낸다. 로지스틱 회귀분석을 위해 사용된 가구(가구원)은 원패널 가구의 가구원이면서 동시에 12차 연도에 응답한 가구원의 집합이다. 즉, 추정하는 응답 확률은 $P(R_i=1 : \text{원패널 가구원, 12차 응답 가구원})$ 로 표기할 수 있다. 여기서 R_i 는 13차 연도 응답 여부를 나타내는 지시변수이다. Duncan(1995)의 경우, 로지스틱 회귀분석을 위해서 원패널 가구원(원패널 가구의 가구원)의 해당 시점 응답 여부를 종속변수로, 그리고 최초 시점의 관측된 원패널 가구원의 특성변수를 설명변수로 사용하였다.

원패널의 가구원이면서 12차 연도 이전 1회 이상 응답하였고 12차 연도에 응답하지 않았다가 13차 연도에 다시 응답한 가구원에게는 개인이 응답한 최근 시기의 가구원이 속한 가구 가중치를 종단 가구원 가중치로

부여하거나 혹은 0을 부여할 수 있다. 원패널 종단 가구원 가중치는 이전 가중치와 0을 부여하는 두 가지를 모두 적용한다.

13차 연도에 유입된 비원가구 가구원의 경우, 종단면 가중치로 0을 부여한다.

원패널 가구에 속하며 13차 연도에 만 15세가 된 가구원의 경우, 단계 2에서 산출된 가구원이 소속되어 있는 가구의 가중치를 종단면 가중치로 정의하고 이후 지속적으로 종단면 가중치를 부여해 산출한다.

나. 가구 가중치 산출 과정

가구 가중치는 기본적으로 1단계에서 산출된 개인(가구원) 가중치들의 가구별 평균으로 정의되며 가구별 평균 산출 시 13차 연도 가구 아이디가 사용되었다. 가구 가중치를 부여할 때, 한 명 이상의 원 가구 가구원이 13차 연도에 조사된 경우, 해당 가구는 0이 아닌 가구 가중치를 부여받게 된다. 그러나 13차 연도에 조사된 가구 내의 가구원이 모두 원가구의 가구원이 아닌 경우, 즉 원년 이후 새로이 유입된 가구원들만이 조사된 경우 각 가구원들의 종단면 가중치는 0이 되므로 해당 가구의 가구 가중치 역시 0으로 부여된다. 가구별 가구원의 종단면 가중치 평균 산출 시 주의할 점들은 다음과 같다.

원가구에 포함되어 있지 않고 13차 연도에 새롭게 패널에 유입되어 조사된 가구원의 경우(예: 결혼, 입양 등), 평균 계산을 위한 신규 유입 가구원(개인)의 종단면 가중치로는 0을 사용한다.

원패널 가구 가구원 중 만 15세가 되어 패널에 새로이 유입된 경우, 평균 계산에서 제외되나 해당 가구원의 종단면 가중치로는 이 단계에서 계산된 가구 가중치가 부여되고, 이 시점 이후의 종단면 가중치는 지속적으로 산출한다.

다. 개인 횡단면 가중치 산출 과정

13차 연도에 조사된 모든 가구원에게 부여되는 횡단면 가중치는 각 가구원이 속한 가구마다 2단계에서 산출된 가구 가중치로 정의된다. 앞서 언급한 바와 같이 13차 연도에 조사된 가구 내 가구원 중 원패널 가구 가구원이 한 명도 존재하지 않는 경우를 제외하고 모든 가구의 가중치는 양의 값이 부여되기 때문에 원가구원이 조사되지 않은 가구에 속한 비원가구원을 제외하고 13차 연도에 조사된 모든 가구원은 횡단면 가중치를 부여받게 된다.

제2절 한국복지패널의 가중치

1. 한국복지패널의 표본 규모

가. 패널 원표본 가구 구축 과정⁴⁾

한국복지패널의 원표본인 7,000가구를 선정하기 위한 자료는 ‘2006년 국민생활실태조사’를 이용하였으며, 국민생활실태조사는 2005년도 인구주택총조사의 90% 조사구인 23만여 개 조사구 중 517개 조사구를 지역별 조사구 규모에 따라 층화 확률 비례 추출하였다.

전체 517개 표본조사구 중에서 수해와 같은 천재지변으로 조사가 불가능한 지역을 제외하고 487개 조사구에 대한 조사를 완료하였다.

조사 대상 지역은 섬을 제외한 전국의 각 시·도이며, 제주도를 포함한다. ‘2006년 국민생활실태조사’에서 최종적으로 조사가 완료된 2만 4711가

4) “2006 한국복지패널 기초분석 보고서”(김미곤 외, 2006)의 표본추출 및 추출방법을 인용(포괄적 이용)함.

구의 소득을 기준으로 7000가구를 2단계 표본추출하여 최종적으로 7072 가구를 패널로 구축하였다.

1단계 표집 자료인 ‘2006년 국민생활실태조사’ 자료를 바탕으로 패널 가구의 대표성을 확보하기 위해 중위소득 60% 이하인 저소득층 가구 3500가구와 중위소득 60% 이상인 일반 가구 3500가구를 각각 표본으로 추출하여 조사를 수행하였다. 이때 저소득 가구와 일반 가구를 구분하는 기준으로는 다음과 같은 3가지 대안을 고려하였고, 최종적으로 ‘공공부조 전 경상소득’의 중위소득 60%를 기준으로 구분하였다.

〈표 3-2-1〉 일반 가구와 저소득 가구의 분류 기준

(단위: 가구, %)

중위소득	경상소득		가처분소득		공공부조 전 경상소득	
	가구 수	백분율	가구 수	백분율	가구 수	백분율
< 40%	2,481	10.00	2,489	10.09	3,477	13.96
< 50%	4,016	16.12	3,880	15.62	4,757	19.04
< 60%	5,227	22.56	5,473	22.25	6,128	24.76

자료: 김미곤 외(2006). 2006 한국복지패널 기초분석 보고서. p80. 〈표 2-2-3〉 인용.

소득 규모별로 구분된 2개의 층에서 지역별, 조사구별로 확률 비례 계통 추출에 의해 일반 가구와 저소득 가구를 표본으로 추출하였다. 상대적으로 작은 규모를 가진 저소득층 가구에 대해서는 추출률을 상향 조정하여 일반 가구와 동일한 수준인 3500가구를 표본 가구로 선정하였다.

KOWEPS의 1차 조사에서 구축된 패널 가구를 소득 기준별로 살펴보면 〈표 3-2-1〉과 같다.

패널로 구축된 표본 가구를 소득 기준별로 살펴보면 당초 각 층별로 3500가구씩 배분하였으나 조사 거절, 패널 참여 거부 등의 사유로 저소득층에서는 표본설계 당시의 3500가구를 약간 밑도는 규모인 3283가구가 패널로 구축되었다.

일반 가구의 경우 평균적으로 100% 이상의 완료율을 보인 반면, 저소득 가구의 경우 약 94%의 완료율을 보여 저소득 가구의 조사 거부가 많았던 것으로 나타났다. 완료 가구 수를 기준으로 지역별 목표 오차는 약 2.7~11.1%로 지역별 변동이 큰 것으로 나타났다. 패널 가구의 탈락률을 최소화하는 것은 종단면 조사인 패널 조사에서 중요하게 고려해야 하는 사항으로, 패널 자료의 대표성을 유지하는 데 가장 중요한 부분이기도 하다. 패널 조사의 가장 큰 문제점은 시간의 흐름에 따라 패널 대상 가구의 탈락률이 지속적으로 증가한다는 점이다. 이러한 점을 고려할 때 1차 조사 당시 패널 가구를 구축하는 과정에서 거부율과 비협조성을 최소화하는 방향으로 표본 가구를 선정하도록 하였다. 이때 조사구별 표본 가구는 계통 추출을 하여 선정하도록 하였으며, 만일 조사 대상 표본 가구가 조사를 거절할 경우 순서상 다음 가구를 조사 대상 가구로 선정하도록 하였다.

〈표 3-2-2〉 KOWEPS 1차 조사의 지역별 조사구와 가구 분포 현황

(단위: 가구)

지 역	조사구 수		일반 가구		저소득 가구		합 계	
	표본 조사구 수	완료 조사구	표본 배분 가구 수	완료 가구 수	표본 배분 가구 수	완료 가구 수	표본 배분 가구 수	완료 가구 수
서울	93	93	811	886	506	449	1,317	1,335
부산	33	33	254	277	272	251	526	528
대구	25	25	187	204	227	218	414	422
인천	28	28	228	247	193	190	421	437
광주	15	15	114	123	130	121	244	244
대전	14	14	118	132	89	76	207	208
울산	14	14	120	124	82	79	202	203
경기	76	76	644	706	471	426	1,115	1,130
강원	14	14	102	118	131	118	233	236
충북	14	14	108	113	113	107	221	221
충남	20	20	153	161	168	160	321	321
전북	20	20	138	140	209	207	347	347
전남	19	19	104	109	273	268	377	377
경북	26	26	152	164	339	329	491	494
경남	30	30	229	240	254	248	483	488
제주	5	5	38	45	43	36	81	81
합 계	446	446	3,500	3,789	3,500	3,283	7,000	7,072

자료: 김미곤 외(2006). 2006 한국복지패널 기초분석 보고서. p80. 〈표 2-2-4〉 인용.

나. 패널 신규 표본 가구 구축 과정⁵⁾

한국복지패널의 6차 연도 조사 이후에 원표본 가구 유지율이 감소하는 상황에서 신규 표본 가구의 추가 필요성이 제기되었다. 저소득 가구 및 가구원의 분포가 치우치는 현상이 발생하였고, 지역별 표본 규모는 잦은 이주와 탈락 등의 사유로 변동이 발생하였다. 따라서 7차 조사에서는 1차 조사 패널 표본 규모를 유지하고자 약 1800가구를 추가하여 신규 패널을 구축하도록 하였다.

한국복지패널의 경우 1차 조사 당시 인구주택총조사에서 표본을 직접 추출하지 않고, ‘2006년 국민생활실태조사’ 표본으로부터 이중 추출(double sampling) 방식으로 7072가구를 일반 가구와 저소득 가구로 구분하여 추출하였다. 신규 표본의 추가를 위한 표본추출 방법도 1차 조사와 동일한 방식으로 고려하였다.

1차 조사 당시 저소득 가구와 일반 가구를 각각 3500가구씩 패널로 구축하여 6차 조사까지 조사를 수행하였고, 1차 조사의 1단계 표본은 ‘2006년 국민생활실태조사’에서 선정된 약 2만 4000여 가구의 대규모 표본으로부터 2단계로 추출한 표본이기 때문에 패널의 특성을 유지하기 위해 가능한 한 저소득 및 일반 가구 표본을 현재 6차 조사에서 탈락한 비율에 따라 선정하는 방안을 고려하였다. 이 경우 활용 가능한 표본은 ‘2011년 복지욕구실태조사(1만 5000여 가구 자료)’를 통한 소득 자료를 이용하여 저소득 및 일반 가구를 1단계 표본으로 파악하였기에 동일한 속성의 패널 자료 구축이 용이하며, 현재 패널 가구의 가중치 조정에도 맞추어 조정 가능하다는 장점이 있다. 일반 가구와 저소득 가구의 비율도 1

5) “2013 한국복지패널 기초분석 보고서”(이현주 외, 2013)의 표본특성 및 가중치 조정 부분을 인용(포괄적 인용)함.

차 조사와 동일하게 저소득 가구를 과대 표집하였다. 지역별 표본 배분 또한 1차 조사 당시의 지역별 가구 비율과 유사하게 표본 가구를 배분하여 표본의 대표성을 확보하고자 하였다. <표 3-2-3>은 7차 조사 당시 신규 패널의 지역별 표본 배분 결과를 나타낸다. ‘2011년 복지욕구실태조사’에서 일반 가구와 저소득 가구로 구분하여 지역별 가구 분포에 따라 1800가구를 비례 배분하였다. 서울과 경기 지역은 표본 가구가 가장 많이 배분된 지역이며, 일반 가구가 저소득 가구에 비해 상대적으로 많이 배분되었다. 대구, 강원, 충북, 충남, 전북, 전남, 경북, 경남 지역은 저소득 가구가 일반 가구에 비해 상대적으로 많이 배분되었다.

<표 3-2-3> KOWEPS 7차 조사 신규패널 지역별 표본 배분 결과

(단위: 가구)

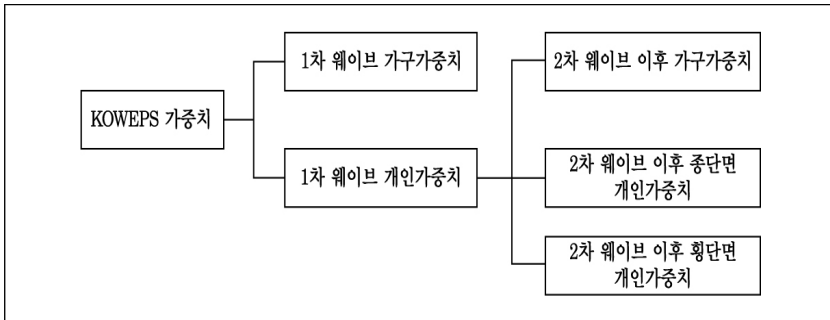
지역	2011년 복지욕구실태조사 표본			표본 배분		
	일반	저소득	계	일반	저소득	계
전 국	10,745	4,298	15,043	900	900	1,800
서울	1,494	343	1,837	125	72	197
부산	821	330	1,151	69	69	138
대구	616	313	929	52	66	118
인천	706	243	949	59	51	110
광주	504	200	704	42	42	84
대전	535	172	707	45	36	81
울산	457	142	599	38	30	68
경기	1,510	342	1,852	126	72	198
강원	443	286	729	37	60	97
충북	490	235	725	41	49	90
충남	482	344	826	40	72	112
전북	524	278	802	44	58	102
전남	501	300	801	42	63	105
경북	600	351	951	50	73	123
경남	751	303	1,054	63	63	126
제주	311	116	427	27	24	51

자료: 이현주 외(2013). 2013 한국복지패널 기초분석 보고서. p39. <표 2-1-4> 인용.

2. 한국복지패널의 가중치 산출 방법⁶⁾

한국복지패널 가중치 산출 방법은 13차 조사 가중치를 기준으로 설명하고자 한다. KOWEPS의 13차 조사 가중치 산정은 기본적으로 12차 조사 당시 부여된 개인 가중치를 기준으로 가구 가중치와 개인 가중치를 산정하였다. 기존의 KOWEPS 가중치 부여 체계는 [그림 3-2-1]과 같다. 가구 가중치는 조사 차수가 증가함에 따라 가구원의 변동으로 인해 가구의 개념이 1차 조사 정의와 상이해질 수 있으므로 2차 조사 이후의 가중치는 개인 가중치를 중심으로 산정하도록 한다. 또한 가구는 개인과 달리 차수가 증가함에 따라 생성과 소멸이 반복적으로 이루어지기 때문에 종단면과 횡단면의 구분이 모호해져 이를 구분하는 것은 별 의미가 없다. 따라서 가구 가중치는 개인 가중치를 기준으로 종단면과 횡단면 가중치를 구분하지 않고 단일 가중치를 부여한다.

[그림 3-2-1] KOWEPS의 가중치 부여 체계



자료: 김미곤 외(2007). 2007 한국복지패널 기초분석 보고서. p102. [그림 2-3-1] 인용.

6) “2018년 한국복지패널 기초분석 보고서”(김태완 외, 2018)의 표본특성 및 가중치 조정 부분을 인용(포괄적 인용)함.

한국복지패널 7차 조사의 경우 기존의 1~6차 조사에서 구축된 패널 표본과 7차 조사에서 새롭게 추가된 신규 표본 패널을 병합하여 새롭게 가중치를 조정할 필요가 있었다. 먼저 7차 조사에서의 가중치 조정 과정을 간략히 살펴보면 다음과 같다.

1차 조사 이후 7차 조사까지 구축된 기존의 패널 표본은 6차 조사의 응답률을 로지스틱 회귀모형에 적용해 응답 확률을 추정하고, 6차 조사의 종단면 가중치에 기초하여 7차 조사의 개인 종단면 가중치를 조정하였다. 7차 조사에 새롭게 추가된 신규 패널은 '2011년 복지욕구실태조사'의 최종 가중치를 기본 가중치로 고려하여 1800가구의 추출 확률의 역수를 곱하여 신규 패널의 설계 가중치를 조정하였다.

이와 같이 각각 독립적으로 조정된 가중치에서 신규 패널의 7차 조사 종단면 가중치는 2012년 새롭게 진입한 신규 가구이므로 모두 '0'으로 부여하였고, 7차 조사 횡단면 가중치의 경우 기존 패널과 신규 패널을 병합하여 새롭게 가중치를 부여하였다.

8차 조사 개인 가중치는 7차 조사의 개인 종단면 가중치를 기준으로 종단면과 횡단면 가중치를 부여하였다. 단 7차 조사에 새롭게 추가된 신규패널 1800가구의 가구원은 개인 종단면 가중치가 0이므로 7차 조사 개인 횡단면 가중치를 기준으로 원표본 패널 가중치와 병합하여 8차 조사의 종단면 가중치와 횡단면 가중치를 산정하였다.

9차 조사 개인 가중치는 8차 조사의 개인 종단면 가중치를 기준으로 종단면과 횡단면 가중치를 부여하고, 10차 조사 개인 가중치는 9차 조사의 개인 종단면 가중치를 기준으로 종단면과 횡단면 가중치를 부여하였다. 13차 조사 개인 가중치 산출 역시, 12차 조사의 개인 종단면 가중치를 기준으로 종단면과 횡단면 가중치를 부여하였다. 13차 조사 개인 가중치에 대한 설명은 다음과 같다.

가. 개인 종단면 가중치 산출 과정

13차 조사 개인 종단면 가중치는 13차 조사 종단면 기본 가중치에 무응답 보정과 개인별 변동 상황을 고려하여 조정한 뒤, 사후 조정으로 계산된다.

13차 조사 종단면 기본 가중치는 12차 조사의 개인 종단면 가중치를 기본 가중치로 한다. 13차 조사의 무응답으로 인한 실제 표본 크기 감소를 반영하기 위한 무응답 보정은 다음과 같다.

13차 조사에서 조사가 완료된 개인의 가중치 조정은 패널 탈락으로 인한 무응답 보정을 위해서 13차 조사 종단면 응답 여부 변수와 12차 조사 변수들의 관계를 로지스틱 회귀모형을 이용하여 응답 확률을 추정하였다. 로지스틱 모형의 설명변수로는 응답자의 성별, 연령, 지역, 경제활동 상태 변수를 사용하였고, 13차 조사 종단면 응답 여부 변수는 13차와 12차 조사 모두 응답하면 1, 그렇지 않으면 0을 갖는다.

다음으로 13차 조사의 개인 종단면 기본 가중치에 로지스틱 모형으로 추정된 응답 확률의 역수를 곱하여 13차 조사의 개인 종단면 가중치를 조정한다. 13차 조사의 기본 가중치를 $w_{\text{기본 13차}}$, 로지스틱 회귀분석을 통해 예측된 예측 응답 확률을 $\hat{p}_{13\text{차}}$ 라고 하면 무응답 보정을 통해 얻어지는 가중치는

$$w_{13\text{차}} = w_{\text{기본 13차}} \times \hat{p}_{13\text{차}}^{-1}$$

로 표현된다.

그리고 개인별 변동 상황에 따라 13차 종단면 가중치를 조정한 뒤, 지역 및 응답자의 인구학적 특성에 따라 2017년 인구 추계값을 이용하여 사후 조정을 실시하였다.

개인별 변동 상황의 예는 다음과 같다.

1차 조사 패널 가구원으로 군 입대 및 해외여행 등의 사유로 조사 당시 응답하지 않았던 가구는 가구의 평균 가중치를 적용받고, 결혼 등의 사유로 원패널 가구에 진입한 가구는 0의 가중치를 부여받는다. 또한 13차 조사에서 원패널 가구로부터 분가한 가구의 경우 원래의 가구원 가중치를 부여받지만, 분가한 후 결혼 등의 사유로 새로 진입한 신규 가구의 가구원은 0의 가중치를 부여받는다. 12차 조사에서 탈락했던 가구원이 13차 조사에 재진입한 경우에는 지역별 평균 가중치를 부여받는다.

지나치게 큰 가중치값들은 추정량의 분산을 크게 만들어 추정의 정확도에 영향을 미칠 수 있는데 이러한 가중치의 과도한 변동을 줄이고자 13차 조사에서는 사후 조정 후 극단 가중치 조정을 하였다.

레이킹을 통한 사후 조정 후 지나치게 큰 가중치들(상위 1%)을 찾아서 절단(trimming)한 뒤 절단값을 제외한 나머지 값을 극단 관측치가 속한 지역*성별*연령 내의 관측치에 동일하게 배분하였다.

나. 개인 횡단면 가중치 산출 과정

13차 조사 개인 횡단면 가중치는 13차 조사 종단면 기본 가중치에 무응답 보정과 개인별 변동 상황을 고려하여 조정한 뒤, 13차 조사 종단면 가중치 값이 0인 가중치에 대해 가구별 평균 가중치를 적용하고 사후 조정으로 계산된다.

패널 탈락으로 인한 무응답 보정을 위해서 13차 조사 종단면 응답 여부 변수와 12차 조사 변수들의 관계를 로지스틱 회귀모형을 이용하여 응답 확률을 추정하였다.

다음으로 13차 조사의 개인 종단면 기본 가중치에 로지스틱 모형으로

추정된 응답 확률의 역수를 곱하여 13차 조사의 개인 종단면 가중치를 조정한다.

그리고 개인별 변동 상황에 따라 13차 조사 종단면 가중치를 조정한 뒤, 13차 조사 종단면 가중치값이 0인 가중치에 대해 가구별 평균 가중치를 적용하고, 지역 및 응답자의 인구학적 특성에 따라 2017년 인구 추계 값을 이용하여 사후 조정을 실시하였다.

13차 조사 횡단면 가중치에서는 레이킹을 통한 사후 조정 후 13차 조사 종단면 가중치와 마찬가지로 지나치게 큰 가중치들(상위 1%)을 찾아서 절단한 뒤 절단값을 제외한 나머지의 값을 극단 관측치가 속한 지역*성별*연령 내의 관측치에 동일하게 배분하였다.

다. 가구 가중치 산출 과정

13차 조사 가구 가중치의 산출 과정은 먼저 가구의 패널 진입 차수에 따라 13차 조사 개인 종단면 가중치의 가구 내 평균값을 계산하여 산출하였다. 13차 조사에서 새롭게 진입한 신규 가구의 경우 신규 가구 내에 원표본(또는 신규 표본) 가구원들의 가중치 평균을 적용하였고, 원표본(또는 신규 표본) 가구원들만으로 구성된 가구는 이들의 평균으로 가구 가중치를 산출하였다. 다음으로 이와 같이 계산된 가구 가중치를 2017년 추계 가구 수에 따라 지역별로 사후 조정하고, 극단값 조정을 통해 최종 가중치를 산출하였다.

극단 값 조정은 레이킹을 통한 사후 조정 후 지나치게 큰 가중치들(상위 1%)을 찾아서 절단한 뒤 절단값을 제외한 나머지 값을 극단 관측치가 속한 지역*가구원 수 내의 관측치에 동일하게 배분하였다.

제3절 소결

3장에서는 한국노동패널과 한국복지패널의 가중치 산출 방식을 설명하였다. 가중치는 연구 목적과 연구 설계 방법에 따라 산출 과정에 차이가 있을 수 있다. 한국노동패널과 한국복지패널은 우리나라에서 10년 이상 된 패널이기 때문에 패널 노후화 및 마모에 대한 이슈 및 연구를 지속적으로 해오고 있다. 두 패널 모두 신규 표본 추가의 경험을 가지고 있기 때문에 여기에서는 두 패널의 가중치 부여 방법에 대한 비교를 통해 한국복지패널의 가중치 산정 과정을 전반적으로 짚어보고자 하였다.

한국노동패널의 원표본 가구 구축은 고용구조특별조사에서 사용된 조사구를 추출 틀로 하였고, 한국복지패널의 원표본 가구 구축은 국민생활실태조사의 조사구를 활용하였다. 한국노동패널의 경우 신규 표본 가구를 생성하는 목적이 기존의 조사 대상이 시지역에 한정됨에 따른 한계점을 극복하기 위해 새롭게 지역 통계를 생산하고 세부 영역에 대해 분석을 시도하기 위한 표본 규모 확대였다. 한국복지패널은 원표본 가구 유지율이 감소하는 상황에서 지역별 표본 규모의 변동, 저소득 가구 및 가구원의 분포가 치우치는 현상이 발생하였고 이를 위해 신규 표본 가구 추가 필요성이 논의되었다. 한국노동패널은 신규 표본의 추출 틀이 인구주택총조사였음에 반해, 한국복지패널은 1차 표본설계 구조를 유지하고자, 복지육구실태조사의 조사구를 활용하여 설계하였다.

한국노동패널은 원패널 가중치 산출 단계가 종단면 가중치 산출 후 가구 가중치 산출, 횡단면 가구원 가중치 산출의 과정으로 이루어지는데, 한국복지패널은 종단면 가중치 산출, 횡단면 가구원 가중치 산출 후 가구 가중치 산출 과정으로 이루어진다. 두 패널의 가중치 산출 단계에서 가구 가중치 산출과 횡단면 가중치 산출 과정 단계가 바뀌어 있는데, 가중치가

0인 대상자에 대해 가구 평균값을 적용하는 것은 동일하기 때문에 크게 다르지 않다고 판단된다.

개인별 변동 상황에 따른 가중치 조정 과정은 한국복지패널이 한국노동패널에 비해 다양한 상황을 고려하고 있다.

한국복지패널의 가중치 산출 방법을 다른 패널과 비교하는 것은 한국복지패널의 가중치 산출 과정에 대해 전반적으로 검토 및 검증할 수 있고 향후 가중치 산출 과정에 있어서 보완하고 고려해야 할 점이 무엇인지 고민할 수 있다는 점에서 필요한 과정이다. 이번 연구에서는 한국노동패널의 가중치 산출 과정과 비교하였는데, 가중치 산출 과정에 대한 점검은 지속적으로 이루어져야 할 것이다.

제 4 장

마모패턴별 특성 비교 분석

제1절 패널 마모패턴 분석

제2절 패널 마모패턴별 패널 가구 특성 분석

제3절 차수별 횡단 가중치 분포 분석

제4절 패널 마모가 가구 소득(가처분소득/경상소득)에
미치는 영향 분석

제5절 소결

4

마모패턴별 특성 비교 분석 <<

제1절 패널 마모패턴 분석

한국복지패널의 연간 데이터를 이용하여 표본 가구의 마모패턴을 살펴 본다. 먼저 마모패턴을 나타내는 변수를 생성하였다. 기존 패널의 경우 총 13차에 걸쳐 조사가 이루어졌으므로 자릿수는 13자리로 하고, 각 자리는 조사년을 나타내는 것으로 한다. 각 연도에서 패널은 응답을 할 경우 패널로 유지되고 응답을 거절할 경우 패널에서 제외된다. 패널 유지인 경우에 '1'의 값을 지정하고, 패널 마모인 경우는 '0'의 값을 지정하였다. 또한 신규 패널 가구는 7차부터 조사에 포함되고 분가 가구는 분가 이후에 패널에 포함되기 때문에 신규 패널 가구와 분가 가구는 패널에 포함되지 않는 연도가 존재한다. 이러한 경우에는 '.'의 값을 지정한다. 예를 들어 마모패턴 '1111100000000'은 원패널에서 5차까지 응답을 하고 6차부터는 응답 거절로 데이터가 없는 경우이며, 마모패턴 '..11110000000'은 3차에서 생성된 분가 가구가 6차까지 응답하고 그 이후부터 응답 거절한 경우이다.

1. 기존 패널 마모패턴

기존 패널 가구를 원가구, 분가 가구, 기타 가구로 구분하였는데, 원가구는 13차 조사까지 계속 응답을 한 가구와 계속 응답을 하다가 일정 시점부터 응답을 하지 않아 패널 마모가 일어난 가구이다. 분가 가구는 1차

이후에 패널에 진입한 가구이며, 기타 가구는 1차 혹은 그 후 차수까지 응답을 연속으로 하고 그 후 응답을 거절하였다가 다시 응답한 마모패턴이 불규칙한 가구이다. <표 4-1-1>은 기존 패널의 원가구에 대한 마모패턴을 정리한 것이다.

<표 4-1-1> 기존 패널 마모패턴별 원가구 범주 구분

가구 구분	범주	마모패턴
원가구	p1	"11111111111111"
	p2	"11111111111110"
	p3	"11111111111100"
	p4	"11111111111000"
	p5	"11111111110000"
	p6	"11111111000000"
	p7	"11111110000000"
	p8	"11111100000000"
	p9	"11111000000000"
	p10	"11110000000000"
	p11	"11100000000000"
	p12	"11000000000000"
	p13	"10000000000000"

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

<표 4-1-2>는 <표 4-1-1>의 원가구 마모패턴에 따른 원가구 가구 수를 차수별로 정리한 것이다. 각 차수별 전체 가구 수와 원가구 수의 비율도 함께 나타나 있다. 13차 조사에는 1차부터 13차까지 모두 응답한 가구 4170가구와 분가 가구, 기타 가구 827가구(분가 가구 731가구, 기타 가구 96가구)를 합쳐 총 4997가구가 조사 완료되었음을 알 수 있다. 1차에서 13차까지 전체 가구 중 원가구의 비율은 83~98%로 상당히 높은 편임을 알 수 있다. 따라서 추후 마모패턴별 가구 특성 분석은 각 차수의 원가구를 중심으로 진행하려 한다.

〈표 4-1-2〉 기존 패널 차수별 마모패턴별 원가구 가구 수

마모패턴	가구 수	(단위: 가구)										
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차
p1	4,170	O	O	O	O	O	O	O	O	O	O	O
p2	139	O	O	O	O	O	O	O	O	O	O	O
p3	165	O	O	O	O	O	O	O	O	O	O	O
p4	206	O	O	O	O	O	O	O	O	O	O	O
p5	124	O	O	O	O	O	O	O	O	O	O	O
p6	191	O	O	O	O	O	O	O	O	O	O	O
p7	146	O	O	O	O	O	O	O	O	O	O	O
p8	131	O	O	O	O	O	O	O	O	O	O	O
p9	289	O	O	O	O	O	O	O	O	O	O	O
p10	251	O	O	O	O	O	O	O	O	O	O	O
p11	205	O	O	O	O	O	O	O	O	O	O	O
p12	384	O	O	O	O	O	O	O	O	O	O	O
p13	527	O	X	X	X	X	X	X	X	X	X	X
원가구 총합		6,928	6,401	6,017	5,812	5,561	5,272	5,141	4,995	4,804	4,680	4,474
원가구 비율		98%	97%	95%	94%	92%	92%	90%	89%	88%	88%	86%
전체 가구		7,072	6,580	6,314	6,207	6,034	5,735	5,732	5,619	5,438	5,343	5,189

주: 'O': 패널 유지, 'X': 패널 마모
자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

〈부록 1〉에 1차부터 13차까지 조사에서 기존 패널 가구에 대한 마모패턴을 상세히 기록한 표를 수록하였다. 〈부표 1-1-1〉을 참조하면 1차부터 13차까지 모두 응답한 가구는 4170가구이고 도중에 이탈한 가구는 2758가구이다. 또한 1차 조사 이후 생성된 분가 가구는 998가구이며 기타 가구는 144가구이다.

2. 신규 패널 마모패턴

〈표 4-1-3〉은 7차부터 시행된 조사에서 신규 패널 가구에 대한 마모패턴을 정리한 것이다. 기존 패널과 마찬가지로 신규 패널에서도 패널 가구를 원가구, 분가 가구, 기타 가구로 구분하였다. 7차부터 13차까지 모두 응답한 가구는 1389가구이고 도중에 이탈한 가구는 408가구이다. 또한 7차 조사 이후 생성된 분가 가구는 93가구이며 기타 가구는 3가구다. 〈표 4-1-4〉는 신규 패널의 원가구 마모패턴을 범주별로 정리한 것이다.

〈표 4-1-5〉는 〈표 4-1-4〉의 원가구 마모패턴에 따른 원가구 가구 수를 차수별로 정리한 것이다. 각 차수별 전체 가구 수와 원가구 수의 비율도 함께 나타나 있다. 13차 조사에서는 7차부터 13차까지 모두 응답한 1389가구와 분가 가구 85가구, 기타 가구 3가구를 합쳐 총 1477가구가 조사 완료되었음을 알 수 있다. 13차까지 전체 가구 중 원가구의 비율은 94% 이상으로 상당히 높은 편임을 알 수 있다. 따라서 추후 마모패턴별 가구 특성 분석은 기존 패널과 마찬가지로 각 차수의 원가구를 중심으로 진행하려 한다.

〈표 4-1-3〉 신규 패널 마모패턴별 가구 수

(단위: 가구)			
가구 구분	패턴	가구 수	합계
원가구	"1111111"	1,389	1,389
	"1111110"	36	408
	"1111100"	52	
	"1111000"	55	
	"1110000"	59	
	"1100000"	96	
	"1000000"	110	
분가 가구	"0111111"	3	93
	"0011111"	11	
	"0011000"	2	
	"0001111"	18	
	"0001110"	2	
	"0001000"	1	
	"0000111"	20	
	"0000110"	1	
	"0000100"	1	
	"0000011"	18	
	"0000010"	1	
	"0000001"	15	
기타 가구	"111011"	1	3
	"1111101"	1	
	"1110001"	1	

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

〈표 4-1-4〉 신규 패널 마모패턴별 원가구 범주 구분

가구 구분	범주	마모패턴
원가구	np1	"1111111"
	np2	"1111110"
	np3	"1111100"
	np4	"1111000"
	np5	"1110000"
	np6	"1100000"
	np7	"1000000"

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

〈표 4-1-5〉 신규 패널 차수별 마모패턴별 원가구 가구 수

(단위: 가구)

마모패턴	가구 수	7차	8차	9차	10차	11차	12차	13차
np1	1,389	O	O	O	O	O	O	O
np2	36	O	O	O	O	O	O	X
np3	52	O	O	O	O	O	X	X
np4	55	O	O	O	O	X	X	X
np5	59	O	O	O	X	X	X	X
np6	96	O	O	X	X	X	X	X
np7	110	O	X	X	X	X	X	X
원가구 총합		1,797	1,687	1,591	1,532	1,477	1,425	1,389
원가구 비율		100%	100%	99%	98%	96%	95%	94%
전체 가구		1,800	1,693	1,610	1,571	1,534	1,500	1,477

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

제2절 패널 마모패턴별 패널 가구 특성 분석

본 절에서는 패널 마모패턴별 가구 특성을 살펴보기로 한다. 〈표4-2-1〉은 기존 패널과 신규 패널의 차수별 전체 가구 수와 마모패턴에 따라 정의된 원가구의 수가 나타나 있다. 패널 가구의 특성 비교를 위하여 차수별 전체 가구보다는 제1절에서 정의된 원가구를 추후 분석에 사용하기로 한다. 제1절에서 마모패턴이 기타 가구(불규칙)로 분류된 가구는 대상 가구 수에 비하여 분석의 종류가 매우 많아지고 복잡하여 특성 분석에서는 제외하였다.

〈표 4-2-1〉 연간 자료의 마모패턴별 패널 가구 수

(단위: 가구)

패널 구분	기존 패널		신규 패널		전체 패널	
가구 구분	원가구	전체 가구	원가구	전체 가구	원가구	전체 가구
1차	6,928	7,072	-	-	6,928	7,072
2차	6,401	6,580	-	-	6,401	6,580
3차	6,017	6,314	-	-	6,017	6,314
4차	5,812	6,207	-	-	5,812	6,207
5차	5,561	6,034	-	-	5,561	6,034
6차	5,272	5,735	-	-	5,272	5,735
7차	5,141	5,732	-	-	5,141	5,732
8차	4,995	5,619	1,687	1,693	6,682	7,312
9차	4,804	5,438	1,591	1,610	6,395	7,048
10차	4,680	5,343	1,532	1,571	6,212	6,914
11차	4,474	5,189	1,477	1,534	5,951	6,723
12차	4,309	5,081	1,425	1,500	5,734	6,581
13차	4,170	4,997	1,389	1,477	5,559	6,474

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

1. 패널 가구 특성 분석

패널 가구 특성으로 균등화 소득에 따른 가구 구분, 총가구원 수, 가구 형태, 주택 유형의 4개 변수에 대해 살펴보았다. 가구 총소득에 대한 패널 가구의 특성은 다음 장에 자세히 기술하였다. 균등화 소득에 따른 가구 구분은 이진변수로 마모패턴별 관심 범주에 대한 비율을 구하여 비교하였다. 총가구원 수, 가구 형태, 주택 유형은 범주형 변수이므로 차수별로 빈도수를 구하여 경향을 살펴보았다.

가. 균등화 소득에 따른 가구 구분

일반 가구(1)와 저소득 가구(2)로 구분되어 있는 균등화 소득에 따른 가구 구분 변수를 이용해 저소득 가구의 비율을 마모패턴별로 비교하였다. <표 4-2-2>와 <표 4-2-3>은 기존 패널 및 신규 패널에 대해 마모패턴별 저소득 가구 비율을 나타낸 것이다. 신규 패널의 마모패턴 np1은 7차부터 13차까지 모두 응답한 가구를 뜻하며 np7은 7차만 응답하고 그 이후 차수에서 탈락한 가구를 뜻한다.

〈표 4-2-2〉 기준 패널 마모패턴별 저소득 가구 비율

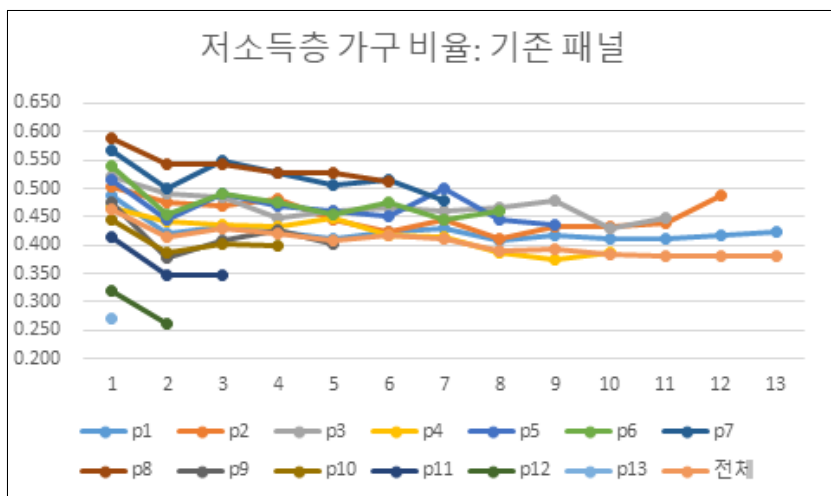
패널 구분	마모패턴	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기준 패널	p1	0.487	0.421	0.433	0.423	0.412	0.424	0.43	0.409	0.418	0.411	0.410	0.418	0.423
	p2	0.504	0.475	0.468	0.482	0.446	0.424	0.446	0.410	0.432	0.432	0.439	0.489	
	p3	0.521	0.491	0.485	0.448	0.461	0.467	0.461	0.467	0.479	0.430	0.448		
	p4	0.466	0.442	0.437	0.432	0.447	0.417	0.413	0.388	0.374	0.388			
	p5	0.516	0.444	0.492	0.468	0.460	0.452	0.500	0.444	0.435				
	p6	0.539	0.455	0.492	0.476	0.455	0.476	0.445	0.461					
	p7	0.568	0.500	0.548	0.527	0.507	0.514	0.479						
	p8	0.588	0.542	0.542	0.527	0.527	0.511							
	p9	0.474	0.377	0.408	0.426	0.401								
	p10	0.446	0.386	0.402	0.398									
	p11	0.415	0.346	0.346										
	p12	0.318	0.260											
	p13	0.271												
전체		0.464	0.414	0.431	0.420	0.408	0.416	0.41	0.389	0.392	0.384	0.381	0.382	0.381

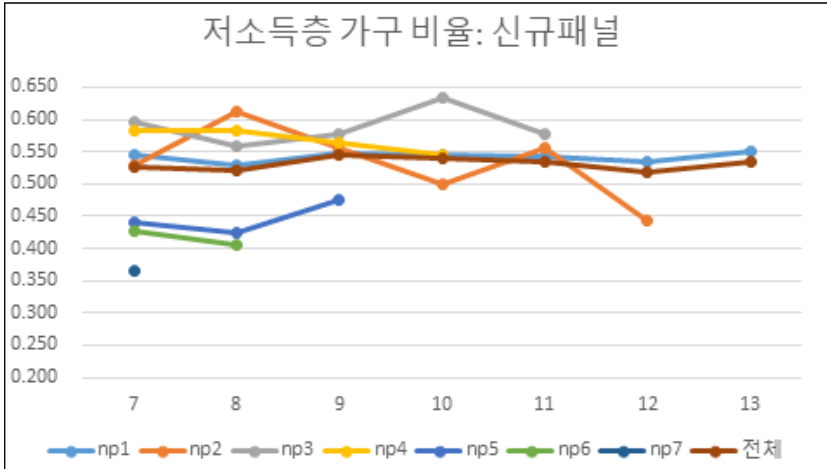
자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

패널 구분	마모패턴	7차	8차	9차	10차	11차	12차	13차
신규 패널	np1	0.544	0.528	0.549	0.544	0.542	0.535	0.550
	np2	0.528	0.611	0.556	0.500	0.556	0.444	
	np3	0.596	0.558	0.577	0.635	0.577		
	np4	0.582	0.582	0.564	0.545			
	np5	0.441	0.424	0.475				
	np6	0.427	0.406					
	np7	0.364						
	전체	0.526	0.522	0.546	0.539	0.535	0.517	0.534

〈표 4-2-2〉, 〈표 4-2-3〉과 [그림 4-2-1]의 결과를 보면 13차까지 모두 응답한 패널 가구(마모패턴 p1)의 저소득 가구 비율이 상대적으로 일찍 탈락한 패널에 비하여 높게 나타남을 알 수 있다. 이는 기존 패널과 신규 패널 모두에서 같은 경향을 확인할 수 있다. 신규 패널의 저소득 가구 비율이 전체적으로 기존 패널의 저소득 가구 비율보다 높게 나타남을 알 수 있다.

[그림 4-2-1] 마모패턴별 저소득 가구 비율





자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

나. 총가구원 수

〈표 4-2-4〉와 〈표 4-2-5〉는 전체 가구를 대상으로 기존 패널 및 신규 패널에 대한 차수별 총가구원 수 및 비율을 보여 준다. 기존 패널 및 신규 패널에서 시간이 흐름에 따라 1인, 2인 가구의 비율은 증가하고, 3인 이상 가구의 비율은 감소하는 경향을 보였다. 일반 가구와 저소득 가구에서도 비슷한 경향을 나타낸다. 〈부표 2-1-1〉 ~ 〈부표 2-1-4〉에 일반 가구와 저소득 가구에 대한 분석 결과를 각각 기술하였다. 분석 결과 1인, 2인 저소득 가구의 비율은 일반 가구 대비 높게 나타났다. 마모패턴별 분석 결과 13차까지 모두 응답한 1인, 2인 가구의 비율이 조기 탈락한 1인, 2인 가구에 비하여 상대적으로 높게 나타났다.

〈표 4-2-4〉 차수별 총가구원 수

패널 구분	총가구원 수	(단위: 가구)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기존 패널	1인	1,521	1,452	1,437	1,449	1,474	1,457	1,474	1,468	1,454	1,471	1,475	1,464	1,505
	2인	2,095	1,936	1,888	1,855	1,798	1,698	1,728	1,682	1,634	1,609	1,572	1,568	1,539
	3인	1,357	1,255	1,174	1,131	1,092	1,029	1,009	1,000	970	911	882	878	849
	4인	1,555	1,415	1,316	1,275	1,202	1,114	1,085	1,061	997	974	920	854	810
	5인 이상	544	522	499	497	468	437	436	408	383	378	340	317	294
	합계	7,072	6,580	6,314	6,207	6,034	5,735	5,732	5,619	5,438	5,343	5,189	5,081	4,997
신규 패널	1인							536	532	536	546	555	569	589
	2인							694	637	603	584	551	528	500
	3인							266	244	217	203	202	193	178
	4인							238	214	195	183	179	169	168
	5인 이상							66	66	59	55	47	41	42
	합계							1,800	1,693	1,610	1,571	1,534	1,500	1,477

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

〈표 4-2-5〉 차수별 총가규원 수 비율

패널 구분	총가규원 수	(단위: %)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기존 패널	1인	21.5	22.1	22.8	23.3	24.4	25.4	25.7	26.1	26.7	27.5	28.4	28.8	30.1
	2인	29.6	29.4	29.9	29.9	29.8	29.6	30.2	29.9	30.1	30.1	30.3	30.9	30.8
	3인	19.2	19.1	18.6	18.2	18.1	17.9	17.6	17.8	17.8	17.1	17.0	17.3	17.0
	4인	22.0	21.5	20.8	20.5	19.9	19.4	18.9	18.9	18.3	18.2	17.7	16.8	16.2
	5인 이상	7.7	7.9	7.9	8.0	7.8	7.6	7.6	7.3	7.0	7.1	6.6	6.2	5.9
	합계	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
신규 패널	1인							29.8	31.4	33.3	34.8	36.2	37.9	39.9
	2인							38.6	37.6	37.5	37.2	35.9	35.2	33.9
	3인							14.8	14.4	13.5	12.9	13.2	12.9	12.1
	4인							13.2	12.6	12.1	11.7	11.7	11.3	11.4
	5인 이상							3.7	3.9	3.7	3.5	3.1	2.7	2.8
	합계							100.0	100.0	100.0	100.0	100.0	100.0	100.0

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

다. 가구 형태

〈표 4-2-6〉과 〈표 4-2-7〉는 전체 가구를 대상으로 기존 패널 및 신규 패널에 대한 차수별 가구 형태 범주의 가구 수 및 그 비율을 보여 준다. 가구 형태는 1. 단독, 2. 모자, 3. 부자, 4. 조손가구 또는 소년소녀가장, 5. 기타로 구성되어 있다.

분석 결과, 기존 패널 및 신규 패널에서 시간이 흐름에 따라 1. 단독 가구의 비율이 증가하는 경향을 보였고, 이와 반대로 5. 기타 가구의 비율은 감소하는 경향을 보였다. 일반 가구와 저소득 가구에서도 비슷한 경향을 나타낸다.

〈표 4-2-6〉 차수별 가구 형태 범주 가구 수

패널 구분		범주	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기존 패널	1.단독		1,520	1,451	1,437	1,448	1,479	1,457	1,474	1,468	1,453	1,471	1,475	1,464	1,505
	2.모자		112	102	90	77	72	62	66	65	66	70	65	64	48
	3.부자		39	36	34	32	28	26	24	32	29	27	28	33	25
	4.조손가구 또는 소년소녀가장		39	37	55	45	39	39	40	38	40	34	31	24	23
	5.기타		5,362	4,954	4,698	4,605	4,416	4,151	4,128	4,016	3,850	3,741	3,590	3,496	3,396
합 계			7,072	6,580	6,314	6,207	6,034	5,735	5,732	5,619	5,438	5,343	5,189	5,081	4,997
신규 패널	1.단독								536	532	536	546	555	569	589
	2.모자								41	27	24	25	24	17	12
	3.부자								11	8	7	7	6	6	3
	4.조손가구 또는 소년소녀가장								9	13	12	9	8	7	3
	5.기타								1,203	1,113	1,031	984	941	901	870
합 계									1,800	1,693	1,610	1,571	1,534	1,500	1,477

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

〈표 4-2-7〉 차수별 가구 형태 범주 가구 수 비율

패널 구분	범주	(단위: %)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기초 패널	1.단독	21.5	22.1	22.8	23.3	24.5	25.4	25.7	26.1	26.7	27.5	28.4	28.8	30.1
	2.모자	1.6	1.6	1.4	1.2	1.2	1.1	1.2	1.2	1.2	1.3	1.3	1.3	1.0
	3.부자	0.6	0.6	0.5	0.5	0.5	0.5	0.4	0.6	0.5	0.5	0.5	0.7	0.5
	4.조손가구 또는 소년소녀가장	0.6	0.6	0.9	0.7	0.7	0.7	0.7	0.7	0.7	0.6	0.6	0.5	0.5
	5.기타	75.8	75.3	74.4	74.2	73.2	72.4	72.0	71.5	70.8	70.0	69.2	68.8	68.0
합 계		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
신규 패널	1.단독							29.8	31.4	33.3	34.8	36.2	37.9	39.9
	2.모자							2.3	1.6	1.5	1.6	1.6	1.1	0.8
	3.부자							0.6	0.5	0.4	0.5	0.4	0.4	0.2
	4.조손가구 또는 소년소녀가장							0.5	0.8	0.8	0.6	0.5	0.5	0.2
	5.기타							66.8	65.7	64.0	62.6	61.3	60.1	58.9
합 계								100.0	100.0	100.0	100.0	100.0	100.0	100.0

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

라. 주택 유형

〈표 4-2-8〉과 〈표 4-2-9〉는 전체 가구를 대상으로 기존 패널 및 신규 패널에 대한 차수별 주택 유형 범주의 가구 수 및 그 비율을 보여 준다. 주택 유형은 1. 단독주택, 2. 다세대주택, 3. 연립주택, 4. 아파트(일반, 임대), 5. 기타로 구성되어 있다.

분석 결과, 기존 패널 및 신규 패널에서 시간이 흐름에 따라 1. 단독주택의 비율은 감소하는 경향을, 4. 아파트(일반, 임대)의 비율은 증가하는 경향을 보였다.

〈표 4-2-8〉 처수별 주택 유형 범주 가구 수

패널 구분	범주	(단위: 가구)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기존 패널	1. 단독주택	3,716	3,515	3,359	3,204	3,140	2,910	2,839	2,776	2,620	2,498	2,351	2,215	2,156
	2. 다세대주택	209	188	491	472	393	491	473	523	446	487	507	509	498
	3. 연립주택	721	649	256	256	328	190	218	151	194	148	137	129	128
	4. 아파트(일반, 임대)	2,167	1,983	1,972	1,878	1,909	1,825	1,886	1,908	1,848	1,844	1,825	1,833	1,822
	5. 기타	259	245	236	397	264	319	316	261	330	366	369	395	393
	합계	7,072	6,580	6,314	6,207	6,034	5,735	5,732	5,619	5,438	5,343	5,189	5,081	4,997
신규 패널	1. 단독주택							968	904	854	822	777	751	726
	2. 다세대주택							118	157	125	109	112	126	128
	3. 연립주택							77	42	61	58	63	54	45
	4. 아파트(일반, 임대)							567	530	520	503	500	477	486
	5. 기타							70	60	50	79	82	92	92
	합계							1,800	1,693	1,610	1,571	1,534	1,500	1,477

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

〈표 4-2-9〉 차수별 주택 유형 범주 가구 수 비율

패널 구분	범주	(단위: %)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기존 패널	1.단독주택	52.6	53.4	53.2	51.6	52.0	50.7	49.5	49.4	48.2	46.8	45.3	43.6	43.2
	2.다세대주택	3.0	2.9	7.8	7.6	6.5	8.6	8.3	9.3	8.2	9.1	9.8	10.0	10.0
	3.연립주택	10.2	9.9	4.1	4.1	5.4	3.3	3.8	2.7	3.6	2.8	2.6	2.5	2.6
	4.아파트(일반, 임대)	30.6	30.1	31.2	30.3	31.6	31.8	32.9	34.0	34.0	34.5	35.2	36.1	36.5
	5.기타	3.7	3.7	3.7	6.4	4.4	5.6	5.5	4.6	6.1	6.9	7.1	7.8	7.9
	합계	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
신규 패널	1.단독주택							53.8	53.4	53.0	52.3	50.7	50.1	49.2
	2.다세대주택							6.6	9.3	7.8	6.9	7.3	8.4	8.7
	3.연립주택							4.3	2.5	3.8	3.7	4.1	3.6	3.1
	4.아파트(일반, 임대)							31.5	31.3	32.3	32.0	32.6	31.8	32.9
	5.기타							3.9	3.5	3.1	5.0	5.4	6.1	6.2
	합계							100.0	100.0	100.0	100.0	100.0	100.0	100.0

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

제3절 차수별 횡단 가중치 분포 분석

본 절에서는 차수 변화에 따른 가구 가중치 분포의 변화를 살펴보고자 한다. 전체 패널에 대한 가중치는 연구진에서 생성한 가구 통합가중치를 사용하였다. 차수별로 전체 가구, 일반 가구, 저소득 가구를 구분하여 가중치의 분포를 비교하도록 한다.

〈표 4-3-1〉 ~ 〈표 4-3-3〉에는 기존 패널의 전체 가구, 일반 가구, 저소득 가구의 가중치에 대하여 최댓값, 99%, 95%, 90%, 75%, 50%, 25%, 10%, 5%, 1% 백분위수, 최솟값, 평균, CV, 4분위수 범위(3분위수-1분위수)를 구하여 가중치 분포를 분석하였다. 〈표 4-3-4〉 ~ 〈표 4-3-6〉에는 신규 패널에 대하여 결과를 기술하였다. [그림 4-3-1] ~ [그림 4-3-6]에는 기존 패널과 신규 패널의 가중치 분포를 상자 그림을 통하여 나타내었다.

기존 패널 전체 가구의 가중치 분포를 보면 차수가 지남에 따라 표본수가 감소함으로써 전체적으로 가중치 값이 증가하는 것을 알 수 있다. 전반적인 값이 증가했을 뿐만 아니라, 4분위수 범위를 보면 차수가 지남에 따라 가중치의 산포도 증가했음을 알 수 있다. CV는 1차에 150% 정도였으나 3차 이후로는 120% 정도를 유지하고 있다.

기존 패널에서는 일반 가구의 가중치 값 평균이 전체 가구에 비하여 크고 저소득 가구의 가중치 값 평균은 전체 가구에 비하여 작음을 알 수 있다. 일반 가구와 전체 가구 모두 차수가 지남에 따라 전체적인 값이 증가하고 가중치의 산포가 증가함을 알 수 있다.

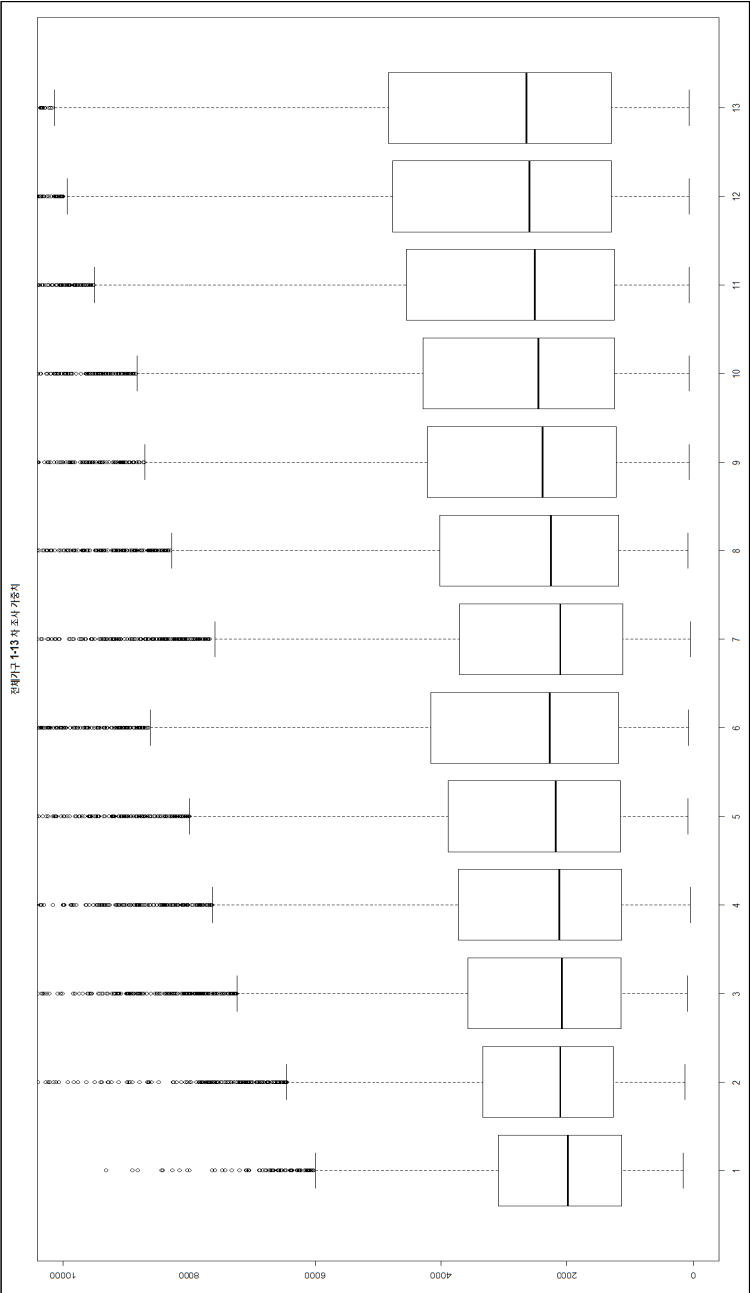
신규 패널의 가중치 값은 같은 차수의 기존 패널 가중치 값에 비하여 약 절반 정도이다. 신규 패널의 가중치는 전체 가구, 일반 가구, 저소득 가구에서 서로 비슷한 값을 지닌다. 차수가 지남에 따라 일반, 저소득 모두 평균 가중치 값이 증가함을 알 수 있고 사분위수 범위로 나타나는 산포 역시 증가함을 알 수 있다. 또한, 저소득 가구 가중치 증가량은 일반 가구에 비하여 적게 나타남을 알 수 있다.

〈표 4-3-1〉 차수별 기존 패널 전체 가구 가중치 분포

	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
최대	9321	15053	28136	20714	21593	21633	34994	10399	11282	11795	12171	12643	13300
99%	6666	7674	9166	9862	10074	10948	11293	10250	10940	11620	11830	12360	13300
95%	4939	5483	6454	6614	7062	7639	6966	7486	7866	8297	8634	8895	9295
90%	4231	4572	5118	5357	5731	6239	5594	5993	6304	6609	6978	7162	7503
75%	3090	3340	3578	3730	3885	4164	3709	4022	4216	4283	4548	4766	4829
50%	1986	2111	2078	2136	2193	2281	2116	2266	2400	2464	2524	2604	2649
25%	1138	1264	1140	1133	1150	1178	1116	1177	1221	1245	1250	1293	1294
10%	554	676	624	628	611	635	583	607	608	624	647	655	659
5%	386	515	450	451	440	445	419	429	438	445	455	468	452
1%	257	316	298	296	279	271	234	236	249	235	219	225	219
최소	160	131	95	48	83	75	49	87	66	71	65	69	70
평균	2247	2456	2600	2686	2804	2991	2741	2884	3021	3125	3260	3368	3464
CV	154%	151%	131%	128%	126%	122%	118%	129%	128%	125%	125%	125%	122%
4범위	1953	2075	2438	2598	2736	2986	2592	2845	2994	3038	3298	3473	3535

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

[그림 4-3-1] 기존 패널 전체 가구 1~13차 조사 가중치(10,000까지)



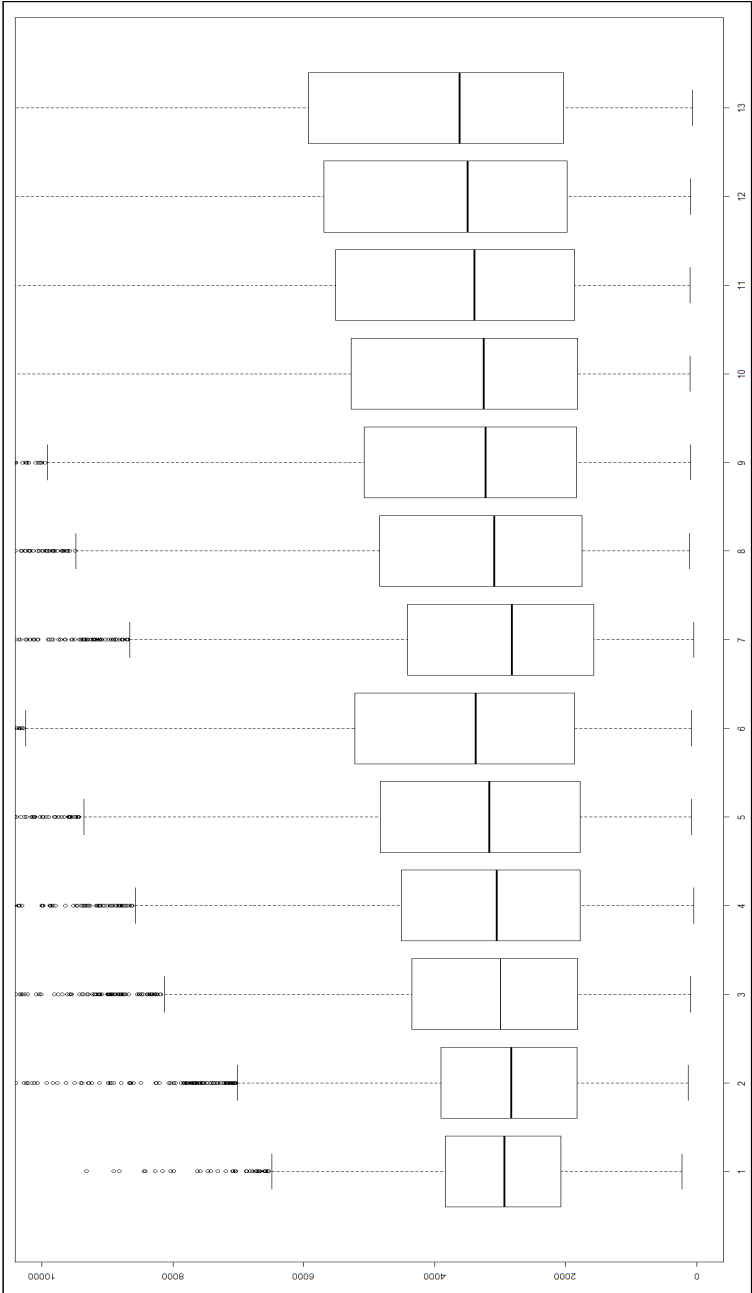
자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

〈표 4-3-2〉 차수별 기존 패널 일반 가구 가중치 분포

	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
최대	9321	15053	28136	20714	21593	21633	34994	10399	11282	11795	12111	12643	13300
99%	7053	8259	10255	10956	11074	12612	12299	10250	10940	11620	11830	12360	13300
95%	5691	6220	7369	7459	7908	8783	8013	8373	8813	9403	9576	10024	10501
90%	4828	5197	5964	6193	6546	7221	6255	6701	7210	7613	7795	8123	8512
75%	3837	3903	4346	4510	4832	5214	4411	4840	5079	5269	5504	5683	5920
50%	2924	2820	2990	3061	3165	3379	2815	3098	3228	3253	3392	3501	3626
25%	2066	1819	1810	1775	1774	1857	1571	1745	1835	1815	1862	1974	2027
10%	1284	1027	988	969	917	957	810	890	993	963	990	1009	1044
5%	893	691	682	651	603	671	551	602	676	635	643	658	662
1%	462	338	346	357	327	358	283	328	333	303	285	270	255
최소	229	136	95	48	83	91	49	116	94	104	101	97	70
평균	3039	3035	3334	3404	3548	3852	3334	3546	3736	3847	3987	4130	4289
CV	212%	177%	157%	153%	150%	145%	133%	152%	151%	145%	147%	147%	144%
4범위	1772	2084	2536	2735	3058	3357	2840	3095	3244	3454	3641	3710	3892

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일: 내부자료]

[그림 4-3-2] 기존 패널 일반 가구 1~13차 조사 가중치(10,000까지)



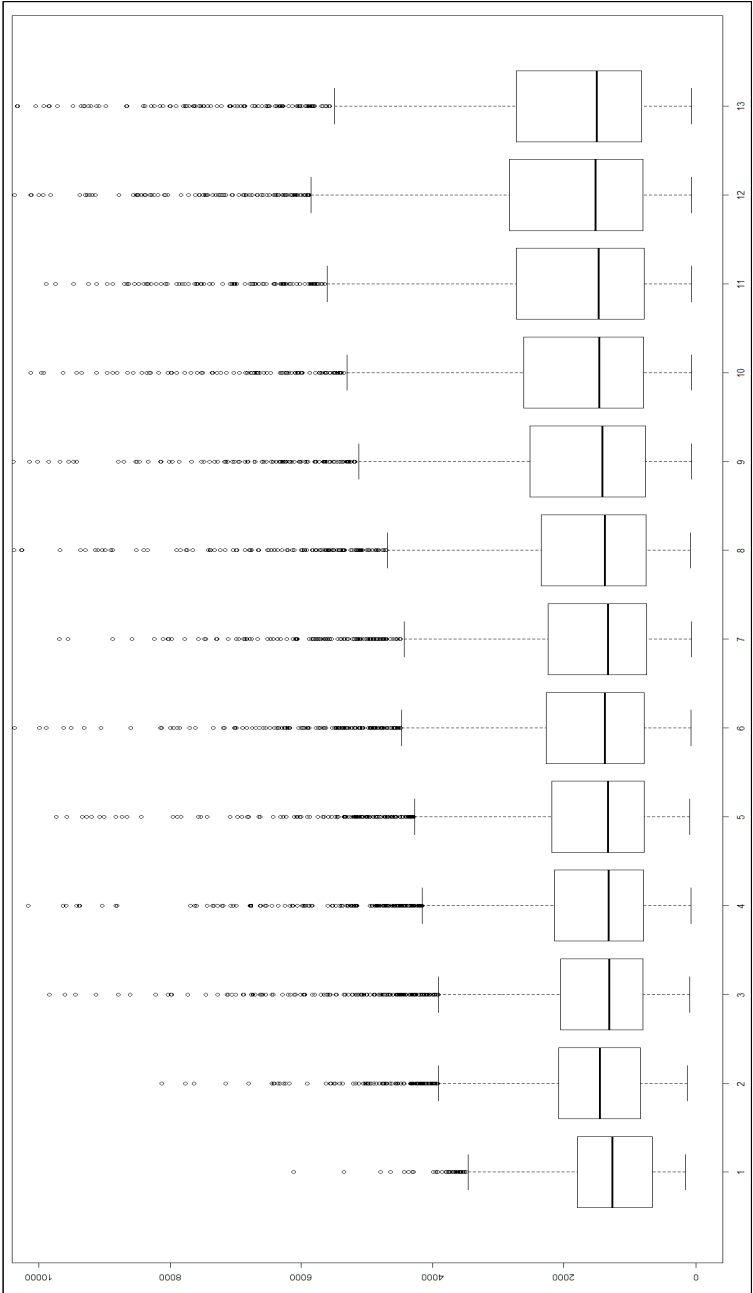
자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

〈표 4-3-3〉 차수별 기존 패널 저소득 가구 가중치 분포

	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
최대	6109	8127	13275	17782	14988	15350	16096	10372	10940	11620	12171	12444	13300
99%	3598	5007	6576	6988	6862	7042	7205	8448	8487	8466	9506	10051	9809
95%	2743	3711	3979	4287	4300	4690	4759	4982	5192	5363	5806	6024	6066
90%	2371	3035	3198	3368	3440	3658	3492	3734	4068	4191	4504	4726	4577
75%	1792	2072	2046	2138	2180	2265	2237	2335	2513	2604	2719	2824	2716
50%	1265	1456	1308	1325	1334	1376	1330	1379	1418	1467	1475	1519	1501
25%	663	842	801	790	786	787	749	756	762	794	788	807	821
10%	385	535	473	470	463	455	439	448	439	461	471	488	462
5%	307	448	370	369	379	370	356	343	351	342	357	351	331
1%	241	263	251	239	246	213	192	180	218	201	188	186	188
최소	160	131	97	78	94	75	71	87	66	71	65	69	72
평균	1329	1635	1632	1693	1721	1784	1753	1824	1924	1989	2089	2149	2145
CV	169%	156%	129%	119%	123%	121%	116%	115%	115%	115%	110%	109%	108%
4범위	1128	1230	1246	1348	1394	1477	1488	1579	1751	1810	1931	2017	1895

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

[그림 4-3-3] 기존 패널 저소득 가구 1~13차 조사 기중치(10,000까지)



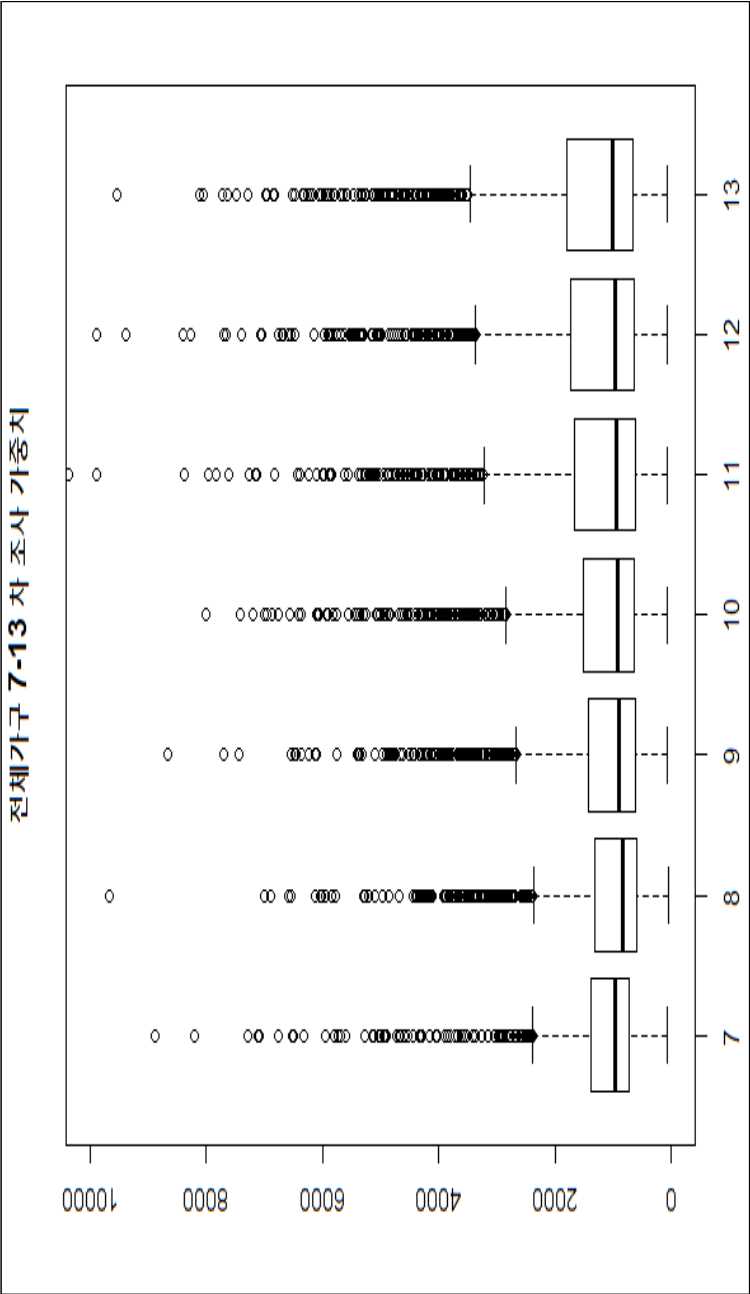
자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

〈표 4-3-4〉 차수별 신규 패널 전체 가구 기중치 분포

	7차	8차	9차	10차	11차	12차	13차
최대	8874	9668	8665	8001	10358	12360	13300
99%	5789	4700	5077	5837	5943	6660	6826
95%	2755	3047	3454	3580	4013	4237	4392
90%	2113	2126	2349	2541	2930	3100	3159
75%	1392	1307	1435	1519	1654	1726	1793
50%	959	840	902	932	942	972	1007
25%	730	592	607	633	622	631	662
10%	589	434	452	438	441	464	467
5%	500	365	377	355	363	386	395
1%	340	242	264	231	241	257	285
최소	80	60	66	64	70	79	82
평균	1246	1132	1219	1277	1368	1447	1499
CV	128%	123%	121%	118%	112%	108%	106%
4범위	662	715	828	885	1033	1095	1131

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일: 내부자료]

[그림 4-3-4] 신규 패널 전체 가구 1~13차 조사 가중치(10,000까지)



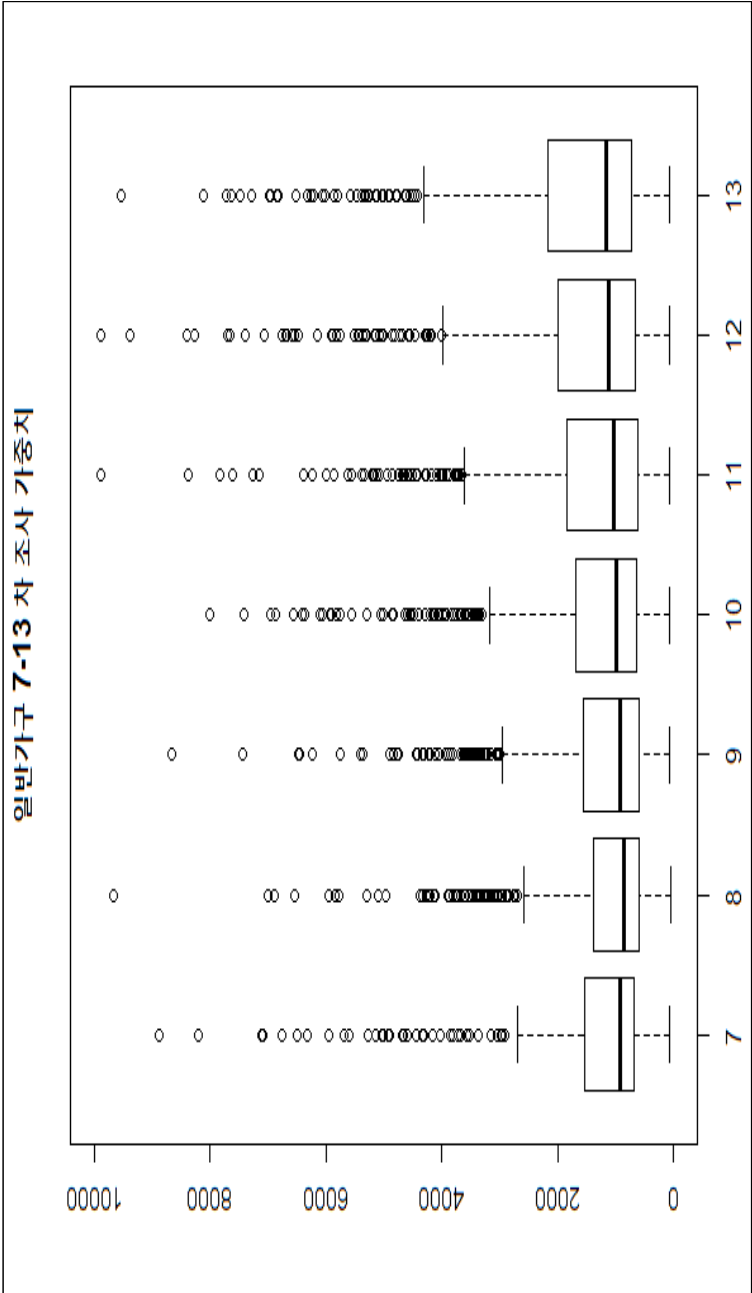
자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

〈표 4-3-5〉 차수별 신규 패널 일반 가구 가중치 분포

	7차	8차	9차	10차	11차	12차	13차
최대	8874	9668	8665	8001	9883	12360	13300
99%	6120	5080	5226	6086	6202	7589	7309
95%	3037	3131	3535	3823	4123	4796	4985
90%	2225	2215	2581	2800	3032	3524	3821
75%	1530	1384	1547	1692	1834	1994	2175
50%	930	862	920	983	1038	1118	1170
25%	694	587	590	639	620	665	717
10%	555	424	451	449	445	473	469
5%	480	353	361	360	372	395	386
1%	327	229	226	234	255	256	261
최소	80	60	66	64	70	79	82
평균	1283	1171	1268	1373	1469	1625	1714
CV	123%	119%	120%	116%	115%	107%	110%
4범위	836	798	958	1053	1214	1329	1458

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일: 내부자료]

[그림 4-3-5] 신규 패널 일반 가구 1~13차 조사 가중치(10,000까지)



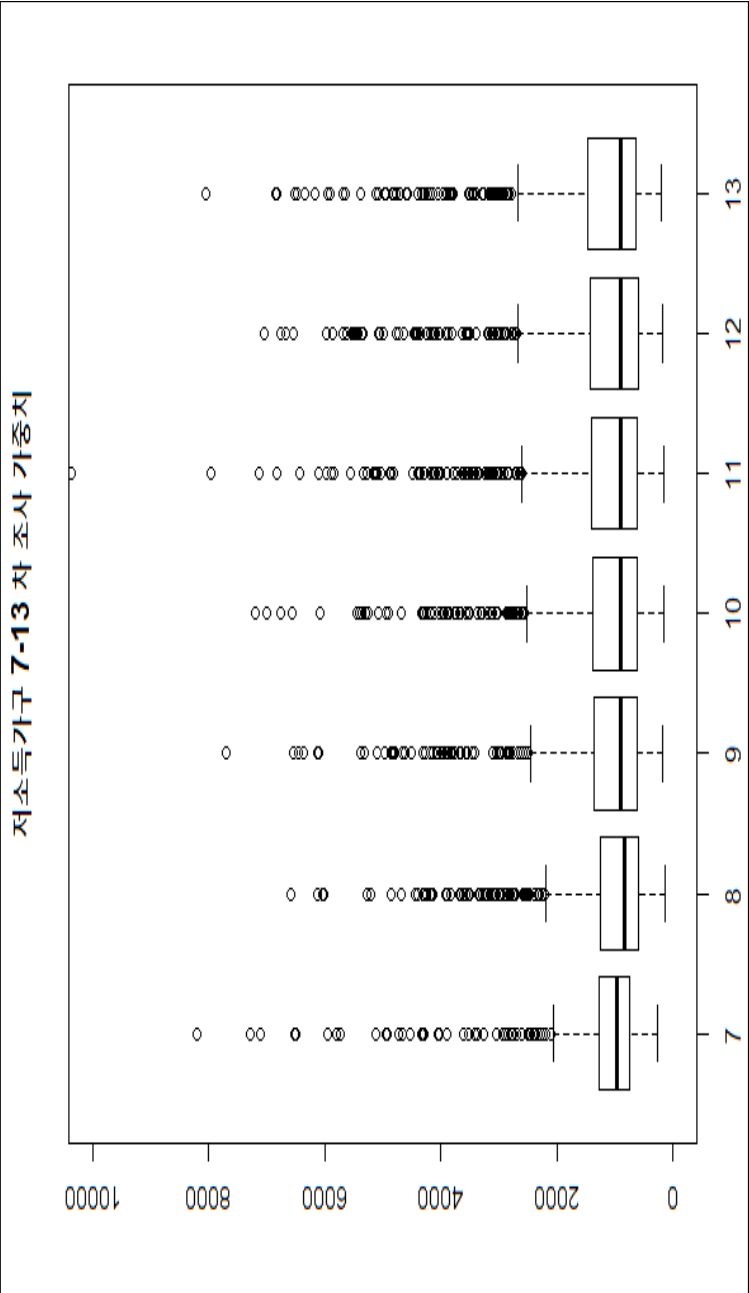
자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일: 내부자료]

〈표 4-3-6〉 차수별 신규 패널 저소득 가구 가중치 분포

	7차	8차	9차	10차	11차	12차	13차
최대	8196	6586	7685	7193	10358	7043	13300
99%	5461	4477	4995	5290	5788	5553	6184
95%	2486	2907	3055	3133	3885	3884	3794
90%	2000	2055	2132	2270	2645	2657	2524
75%	1281	1244	1365	1391	1412	1433	1477
50%	970	827	896	900	895	901	903
25%	759	599	621	625	622	604	632
10%	610	444	457	429	439	459	466
5%	540	382	392	354	361	384	404
1%	371	253	283	232	240	277	294
최소	264	129	171	169	158	176	210
평균	1212	1096	1178	1195	1281	1281	1311
CV	135%	128%	123%	122%	111%	115%	106%
4범위	522	645	745	766	791	829	845

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일: 내부자료]

[그림 4-3-6] 신규 패널 저소득 가구 1~13차 조사 가중치(10,000까지)



자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일: 내부자료]

제4절 패널 마모가 가구 소득(가처분소득/경상소득)에 미치는 영향 분석

본 절에서는 패널 마모가 가구 소득에 미치는 영향을 살펴보고자 한다. 분석 대상 변수로 가처분소득과 경상소득을 사용한다. 두 변수에 대하여 가구 소득은 가중 평균과 비가중 평균으로 추정 가능하다. 한국복지패널은 차수별 횡단면 가중치를 제공하므로 횡단면 가중치로 가구 소득의 평균을 추정할 수 있다. 전체 패널에 대한 가중치는 연구진에서 생성한 가구 통합 가중치를 사용하였다.

마모패턴별로 가중/비가중 평균을 비교하여 그 차이를 살펴본다. 차수별 자료에 대해서는 일반 가구와 저소득 가구를 구분하여 가중/비가중 평균을 비교 분석한다.

1. 가처분소득

가. 비가중 평균

기존 패널 및 신규 패널에 대해 차수별로 전체 가구, 일반 가구, 저소득 가구에 대하여 가구 가처분소득의 비가중 평균을 <표 4-4-1>에 기술하였다. <표 4-4-2>와 <표 4-4-3>은 기존 패널 및 신규 패널에 대해 마모패턴별 가처분소득의 비가중 평균을 계산한 결과이다.

〈표 4-4-1〉 가구별, 패널별, 차수별 가처분소득 비가중 평균

가구 구분	패널 구분	(단위: 만 원)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
전체	기준	2,068	2,455	2,691	2,871	3,071	3,177	3,285	3,471	3,572	3,612	3,802	3,900	4,000
	신규	-	-	-	-	-	-	2,435	2,556	2,538	2,733	2,811	3,014	3,120
	통합	2,068	2,455	2,691	2,871	3,071	3,177	3,082	3,259	3,336	3,412	3,576	3,698	3,800
일반	기준	3,311	3,654	4,038	4,234	4,461	4,661	4,610	4,927	5,194	5,204	5,429	5,552	5,750
	신규	-	-	-	-	-	-	3,989	4,143	4,326	4,618	4,703	4,941	5,238
	통합	3,311	3,654	4,038	4,234	4,461	4,661	4,490	4,778	5,036	5,097	5,297	5,437	5,657
저소득	기준	629	755	913	984	1,049	1,098	1,080	1,142	1,086	1,104	1,183	1,257	1,200
	신규	-	-	-	-	-	-	1,033	1,100	1,055	1,117	1,169	1,213	1,274
	통합	629	755	913	984	1,049	1,098	1,065	1,130	1,077	1,108	1,179	1,245	1,222

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

〈표 4-4-2〉 기준 패널 마모패턴별 가치분소득 비가중 평균

패널 구분	마모 패턴	(단위: 만 원)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기준 패널	p1	1,946	2,392	2,658	2,848	2,985	3,171	3,229	3,397	3,362	3,470	3,717	3,775	3,855
	p2	1,989	2,298	2,436	2,577	2,837	3,081	3,090	3,310	3,476	5,288	3,532	3,686	
	p3	1,133	2,147	2,414	2,593	2,844	3,199	3,170	3,556	3,613	3,599	3,878		
	p4	2,099	2,447	2,836	3,192	3,527	3,420	3,606	3,867	3,920	3,862			
	p5	2,019	2,430	2,690	2,791	3,025	3,482	3,319	3,524	3,761				
	p6	1,927	2,187	2,467	2,526	2,607	2,892	3,100	3,166					
	p7	1,889	2,227	2,700	2,576	2,663	3,044	2,989						
	p8	1,741	2,808	2,388	2,417	2,525	2,574							
	p9	2,110	2,550	2,698	2,985	3,026								
	p10	2,333	2,645	2,833	3,138									
	p11	2,398	2,731	3,314										
	p12	2,570	2,948											
	p13	2,809												

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일, 내부자료]

〈표 4-4-3〉 신규 패널 마모패턴별 가처분소득 비가중 평균

(단위: 만 원)

패널 구분	마모패턴	7차	8차	9차	10차	11차	12차	13차
신규 패널	np1	2,298	2,487	2,506	2,659	2,756	2,960	2,916
	np2	2,587	2,678	2,625	3,071	3,362	3,275	
	np3	2,522	2,940	2,920	2,970	3,208		
	np4	2,194	2,208	2,164	2,283			
	np5	3,248	2,777	2,709				
	np6	3,170	3,423					
	np7	3,145						

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

〈표 4-4-1〉 ~ 〈표 4-4-3〉의 결과, 기존 패널과 신규 패널 모두에서 가처분소득의 평균은 시간이 흐름에 따라 증가하는 경향을 보였다. 기존 패널의 경우 일반 가구의 가처분소득 증가율이 저소득 가구의 가처분소득 증가율에 비하여 크게 나타난다. 또한 마모패턴별로 비교하였을 때, 패널에서 초기에 이탈한 가구의 가처분소득이 높은 경향을 보였다. 이는 총소득이 낮은 가구, 즉 저소득 가구가 패널에 많이 유지됨을 의미한다. 가처분소득의 평균은 전체 가구와 일반 가구에서는 기존 패널이 신규 패널보다 더 큰 값을 갖는 반면 저소득 가구에서는 두 패널의 추정값이 비슷하게 나타난다.

나. 가중 평균

기존 패널 및 신규 패널에 대해 차수별로 전체 가구, 일반 가구, 저소득 가구의 가처분소득 가중 평균을 〈표 4-4-4〉에 기술하였다. 〈표 4-4-5〉와 〈표 4-4-6〉은 기존 패널 및 신규 패널의 마모패턴별 가처분소득 가중 평균을 계산한 결과이다.

그림 [4-4-1]에는 기존 패널과 신규 패널을 합한 통합 패널을 대상으로 전체 가구, 일반 가구, 저소득 가구 각각의 가처분 소득에 대한 가중/비가중 평균이 나타나 있다. 차수가 지남에 따라 전체 가구, 일반 가구, 저소득 가구의 가처분소득은 증가하는 경향을 보인다. 전체 가구와 일반 가구의 가중 가처분소득의 평균은 비가중 가처분소득의 평균보다 큼을 알 수 있다. 저소득 가구의 가처분소득은 가중/비가중 평균의 값이 서로 비슷하게 나타난다.

[그림 4-4-2]와 [그림 4-4-3]에는 기존/신규 패널 가구별 가처분소득의 비가중, 가중 평균이 나타나 있다. 전체 가구와 일반 가구에서는 가중 평균, 비가중 평균 모두 기존 패널의 값이 신규 패널보다 더 크게 나타난다. 하지만 저소득 가구에서는 기존 패널과 신규 패널의 가처분소득 평균값이 비슷하게 나타났으며 13차 조사에서는 신규 패널의 값이 더 크게 나타나기도 하였다.

〈표 4-4-4〉 가구별, 패널별, 차수별 가처분소득 기준 평균

가구 구분	패널 구분	(단위: 만 원)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
전체	기준	2,825	3,069	3,483	3,659	3,903	4,145	4,009	4,282	4,473	4,406	4,773	4,718	4,835
	신규	-	-	-	-	-	-	2,438	2,552	2,576	2,769	2,921	3,218	3,342
	통합	2,825	3,069	3,483	3,659	3,903	4,145	3,812	4,099	4,271	4,231	4,568	4,549	4,666
일반	기준	3,615	3,965	4,419	4,626	4,836	5,133	4,928	5,286	5,603	5,479	5,947	5,821	6,021
	신규	-	-	-	-	-	-	3,940	4,093	4,312	4,493	4,729	4,932	5,138
	통합	3,615	3,965	4,419	4,626	4,836	5,133	4,845	5,200	5,513	5,407	5,855	5,747	5,949
저소득	기준	734	712	958	971	1,106	1,159	1,100	1,163	1,110	1,138	1,164	1,323	1,041
	신규	-	-	-	-	-	-	1,003	1,042	1,025	1,071	1,121	1,184	1,295
	통합	734	712	958	971	1,106	1,159	1,077	1,139	1,093	1,125	1,156	1,297	1,092

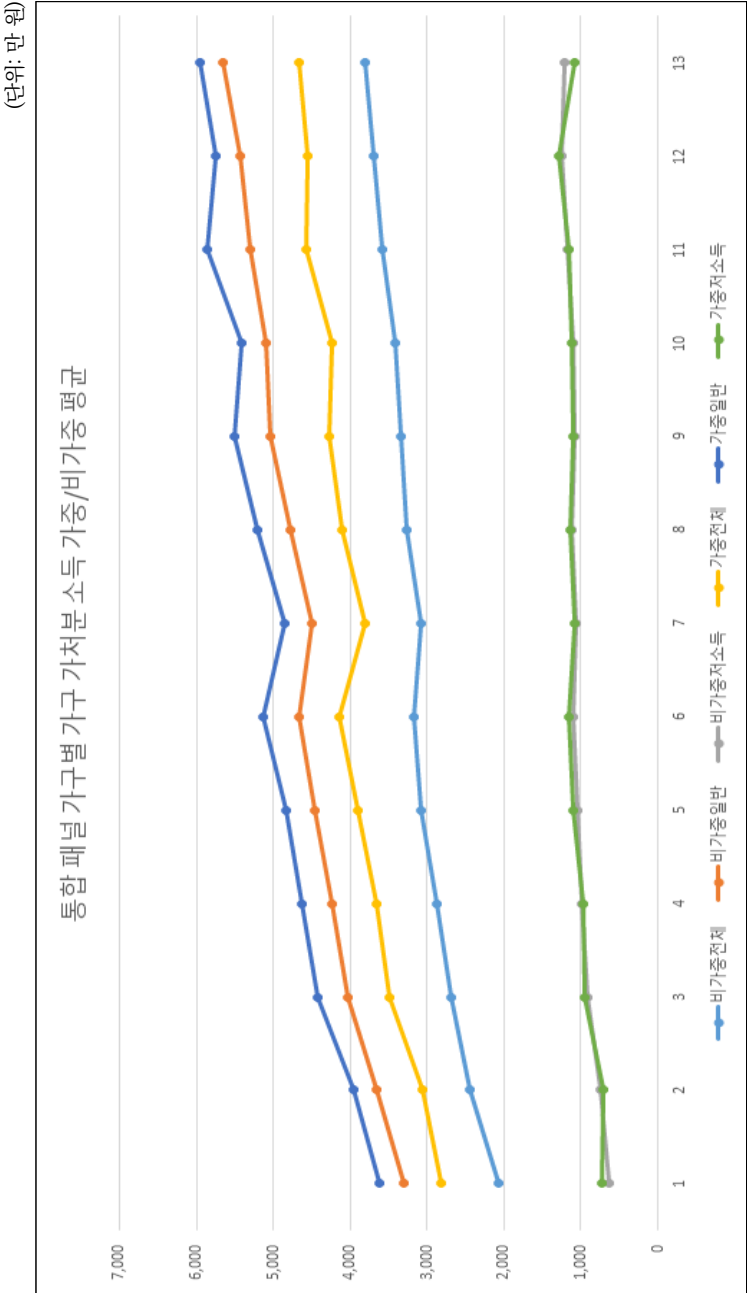
자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

〈표 4-4-5〉 기준 패널 마모패턴별 가치분소득 기중 평균

패널 구분	마모 패턴	(단위: 만 원)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기준 패널	p1	2,660	2,971	3,414	3,634	3,829	4,193	4,033	4,317	4,293	4,367	4,889	4,779	4,886
	p2	2,765	2,849	3,243	3,303	3,804	4,114	3,950	4,419	4,643	6,980	4,520	4,959	
	p3	2,456	2,839	3,350	3,639	3,844	4,503	4,309	4,840	4,796	4,765	5,117		
	p4	2,759	3,083	3,700	4,531	4,961	4,708	4,500	4,909	4,943	4,543			
	p5	2,938	3,273	4,226	3,748	4,431	5,117	4,653	4,722	5,403				
	p6	2,667	2,806	3,222	3,362	3,563	3,934	3,874	3,897					
	p7	2,880	3,300	4,577	4,059	4,240	4,725	4,173						
	p8	2,464	3,449	3,229	3,369	3,526	3,565							
	p9	2,819	3,200	3,478	3,828	3,841								
	p10	3,187	3,391	3,840	4,190									
	p11	3,073	3,267	4,284										
	p12	3,189	3,460											
	p13	3,432												

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일: 내부자료]

[그림 4-4-1] 통합 패널 가구별 가치분소득 가중/비가중 평균



자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

〈표 4-4-6〉 신규 패널 마모패턴별 가처분소득 기중 평균

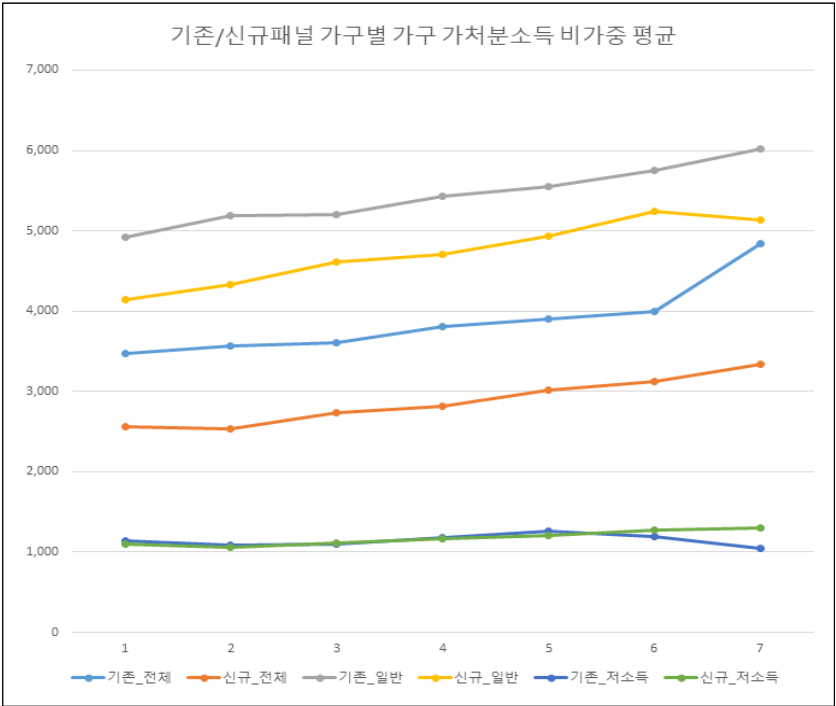
(단위: 만 원)

패널 구분	마모패턴	7차	8차	9차	10차	11차	12차	13차
신규 패널	np1	2,301	2,478	2,536	2,707	2,876	3,197	3,241
	np2	2,581	2,681	2,761	3,085	3,569	3,527	
	np3	2,314	2,692	2,651	2,818	3,078		
	np4	2,512	2,589	2,392	2,587			
	np5	3,629	2,969	3,145				
	np6	3,078	3,309					
	np7	2,821						

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

[그림 4-4-2] 기존/신규 패널 가구별 가처분소득 비가중 평균

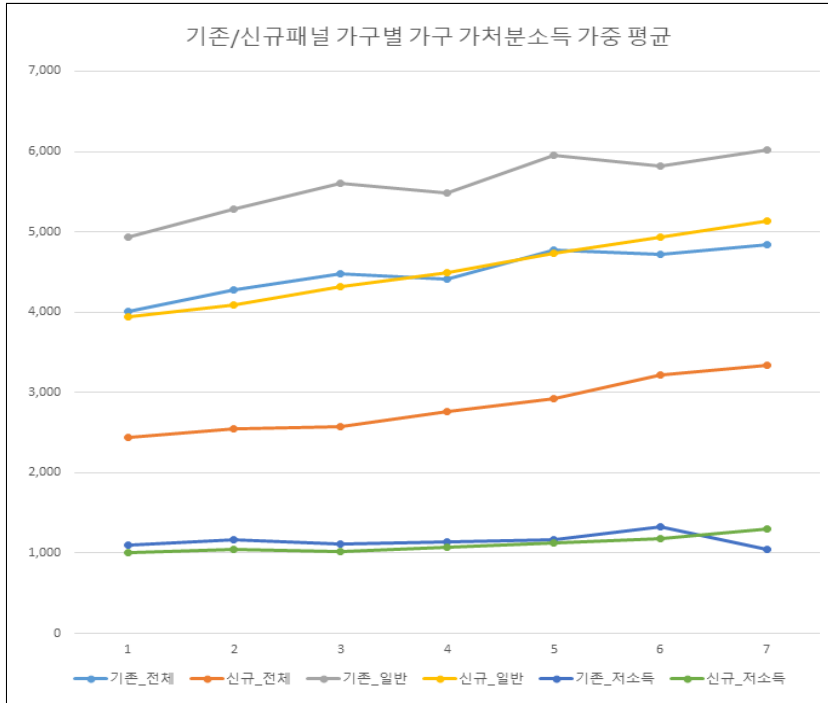
(단위: 만 원)



자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

[그림 4-4-3] 기존/신규 패널 가구별 가처분소득 가중 평균

(단위: 만 원)



자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

2. 경상소득

가. 비가중 평균

기존 패널 및 신규 패널에 대해 차수별로 전체 가구, 일반 가구, 저소득 가구의 경상소득 비가중 평균을 <표 4-4-7>에 기술하였다. <표 4-4-8>과 <표 4-4-9>는 기존 패널 및 신규 패널의 마모패턴별 경상소득 비가중 평균을 계산한 결과이다.

〈표 4-4-7〉 ~ 〈표 4-4-9〉의 결과, 기존 패널과 신규 패널 모두에서 경상소득의 평균은 시간이 흐름에 따라 증가하는 경향을 보였다. 기존 패널에서는 일반 가구의 경상소득 증가율이 저소득 가구의 경상소득 증가율에 비하여 크게 나타난다.

또한 마모패턴별로 비교하였을 때, 패널에서 초기에 이탈한 가구의 경상소득이 높은 경향을 보였다. 이는 총소득이 낮은 가구, 즉 저소득 가구가 패널에 많이 유지됨을 의미한다. 전체 가구와 일반 가구에서 경상소득의 평균은 기존 패널이 신규 패널보다 더 큰 값을 갖는 반면 저소득 가구에서는 두 패널의 추정값이 비슷하게 나타난다.

〈표 4-4-7〉 가구별, 패널별, 차수별 경상소득 비기준 평균

(단위: 만 원)

가구 구분	패널 구분	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
전체	기준	2,324	2,661	2,931	3,096	3,292	3,423	3,547	3,753	3,864	3,923	4,143	4,267	4,395
	신규	-	-	-	-	-	-	2,606	2,727	2,710	2,928	3,025	3,252	3,378
	통합	2,324	2,661	2,931	3,096	3,292	3,423	3,322	3,516	3,600	3,697	3,888	4,036	4,163
일반	기준	3,659	3,982	4,433	4,595	4,813	5,053	5,009	5,363	5,652	5,691	5,961	6,122	6,365
	신규	-	-	-	-	-	-	4,319	4,471	4,669	5,009	5,129	5,396	5,749
	통합	3,659	3,982	4,433	4,595	4,813	5,053	4,876	5,194	5,474	5,567	5,809	5,986	6,252
저소득	기준	780	787	948	1,022	1,081	1,138	1,115	1,179	1,122	1,140	1,220	1,299	1,244
	신규	-	-	-	-	-	-	1,060	1,128	1,084	1,145	1,197	1,246	1,311
	통합	780	787	948	1,022	1,081	1,138	1,098	1,164	1,111	1,141	1,213	1,284	1,263

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

〈표 4-4-8〉 기준 패널 마모패턴별 경상소득 비기준 평균

패널 구분	마모 패턴	(단위: 만 원)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기준 패널	p1	2,201	2,585	2,880	3,064	3,197	3,408	3,474	3,659	3,636	3,753	4,037	4,118	4,216
	p2	2,151	2,481	2,751	2,804	3,071	3,369	3,394	3,633	3,802	5,648	3,891	4,055	
	p3	1,994	2,316	2,613	2,792	3,064	3,448	3,463	3,870	3,917	3,957	4,276		
	p4	2,287	2,663	3,130	3,451	3,790	3,708	3,939	4,248	4,267	4,298			
	p5	2,226	2,657	2,975	3,041	3,293	3,798	3,660	3,843	4,078				
	p6	2,086	2,365	2,679	2,707	2,815	3,113	3,331	3,433					
	p7	2,089	2,403	2,921	2,783	2,858	3,271	3,245						
	p8	1,956	3,056	2,764	2,702	2,681	2,739							
	p9	2,312	2,773	2,979	3,248	3,298								
	p10	2,530	2,913	3,122	3,426									
	p11	2,628	2,978	3,612										
	p12	2,823	3,218											
	p13	3,096												

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일, 내부자료]

〈표 4-4-9〉 신규 패널 마모패턴별 경상소득 비가중 평균

(단위: 만 원)

패널 구분	마모패턴	7차	8차	9차	10차	11차	12차	13차
신규 패널	np1	2,451	2,646	2,672	2,846	2,961	3,187	3,166
	np2	2,830	2,874	2,825	3,308	3,599	3,699	
	np3	2,755	3,180	3,180	3,277	3,493		
	np4	2,319	2,342	2,293	2,434			
	np5	3,459	3,007	2,944				
	np6	3,442	3,720					
	np7	3,413						

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

나. 가중 평균

기존 패널 및 신규 패널에 대해 차수별로 전체, 일반, 저소득 가구의 경상소득 가중 평균을 〈표 4-4-10〉에 기술하였다. 〈표 4-4-11〉과 〈표 4-4-12〉는 기존 패널 및 신규 패널의 마모패턴별 경상소득 가중 평균을 계산한 결과이다.

[그림 4-4-4]에는 기존 패널과 신규 패널을 합한 통합 패널을 대상으로 전체 가구, 일반 가구, 저소득 가구 각각의 경상소득에 대한 가중/비가중 평균을 계산한 결과가 나타나 있다. 차수가 지남에 따라 전체 가구, 일반 가구, 저소득 가구의 경상소득은 증가하는 경향을 보인다. 전체 가구와 일반 가구의 가중 경상소득의 평균은 비가중 경상소득의 평균보다 큼을 알 수 있다. 저소득 가구의 경상소득은 가중/비가중 평균의 값이 서로 비슷하게 나타난다.

[그림 4-4-5]와 [그림 4-4-6]에는 기존/신규 패널 가구별 경상소득의 비가중, 가중 평균이 나타나 있다. 전체 가구와 일반 가구에서는 가중 평균, 비가중 평균 모두 기존 패널의 값이 신규 패널보다 더 크게 나타난다. 하지만 저소득 가구에서는 기존 패널과 신규 패널의 경상소득 평균값이 비슷하게 나타났으며 13차 조사에서는 신규 패널의 값이 더 크게 나타나기도 하였다.

〈표 4-4-10〉 기구별, 패널별, 차수별 경상소득 기준 평균

기구 구분	패널 구분	(단위: 만 원)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
전체	기준	3,153	3,367	3,846	4,001	4,235	4,516	4,387	4,681	4,887	4,846	5,257	5,227	5,390
	신규	-	-	-	-	-	-	2,613	2,730	2,761	3,003	3,161	3,492	3,654
	통합	3,153	3,367	3,846	4,001	4,235	4,516	4,165	4,475	4,660	4,648	5,025	5,032	5,194
일반	기준	4,023	4,361	4,899	5,069	5,265	5,605	5,410	5,797	6,139	6,046	6,572	6,478	6,732
	신규	-	-	-	-	-	-	4,271	4,425	4,673	4,937	5,179	5,411	5,688
	통합	4,023	4,361	4,899	5,069	5,265	5,605	5,314	5,699	6,036	5,965	6,467	6,389	6,646
저소득	기준	850	754	1,008	1,029	1,149	1,223	1,150	1,213	1,162	1,189	1,217	1,380	1,100
	신규	-	-	-	-	-	-	1,029	1,068	1,053	1,099	1,152	1,216	1,334
	통합	850	754	1,008	1,029	1,149	1,223	1,122	1,184	1,140	1,171	1,203	1,348	1,147

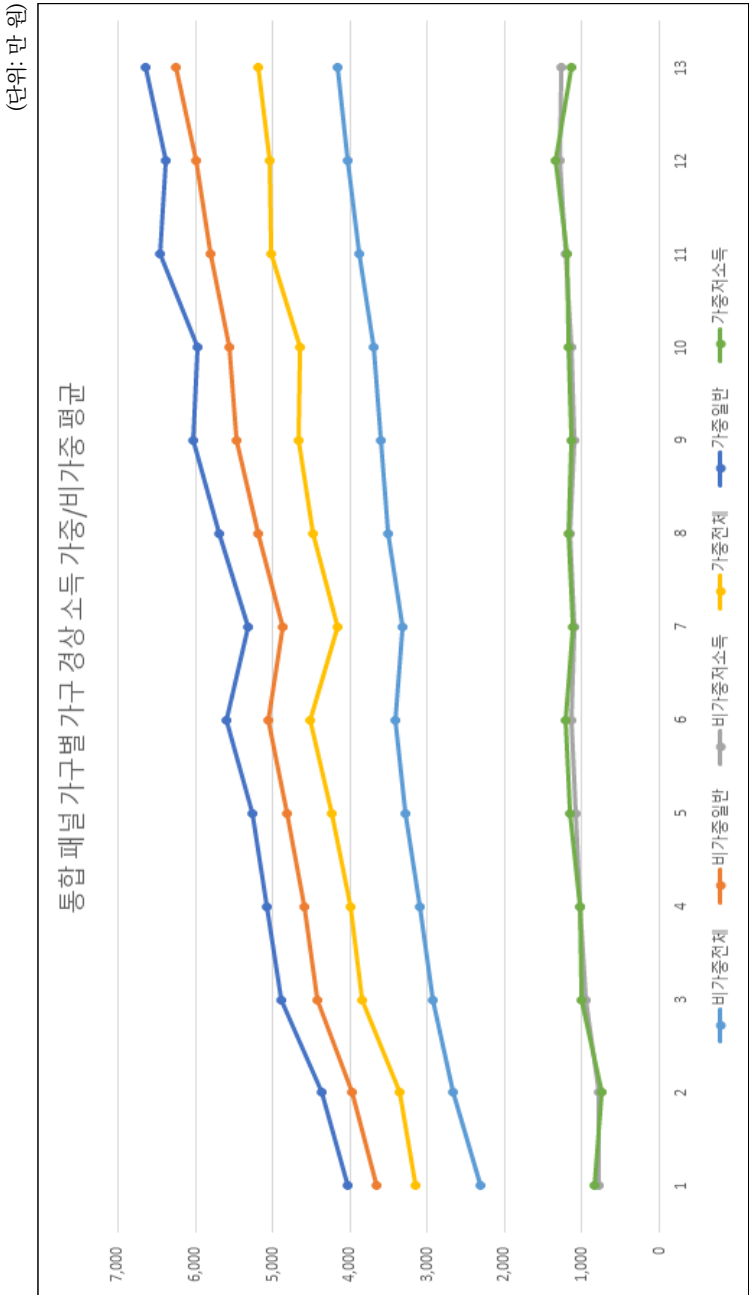
자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

〈표 4-4-11〉 기준 패널 마모패턴별 경상소득 기중 평균

패널 구분	마모 패턴	(단위: 만 원)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기준 패널	p1	2,983	3,252	3,751	3,963	4,154	4,561	4,404	4,707	4,702	4,783	5,369	5,288	5,422
	p2	3,025	3,120	3,790	3,665	4,179	4,570	4,397	4,901	5,138	7,505	5,043	5,505	
	p3	2,890	3,127	3,685	3,971	4,209	4,930	4,782	5,381	5,235	5,334	5,756		
	p4	3,039	3,383	4,142	4,922	5,370	5,160	4,988	5,393	5,389	5,099			
	p5	3,312	3,624	4,714	4,143	4,878	5,632	5,185	5,219	5,880				
	p6	2,928	3,079	3,542	3,621	3,879	4,277	4,190	4,263					
	p7	3,206	3,583	4,969	4,415	4,574	5,131	4,609						
	p8	2,815	3,826	3,871	3,910	3,758	3,846							
	p9	3,130	3,536	3,901	4,223	4,259								
	p10	3,493	3,776	4,290	4,641									
	p11	3,391	3,595	4,703										
	p12	3,521	3,803											
	p13	3,813												

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일, 내부자료]

[그림 4-4-4] 통합 패널 가구별 경상소득 가중/비가중 평균



자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

〈표 4-4-12〉 신규 패널 마모패턴별 경상소득 가중 평균

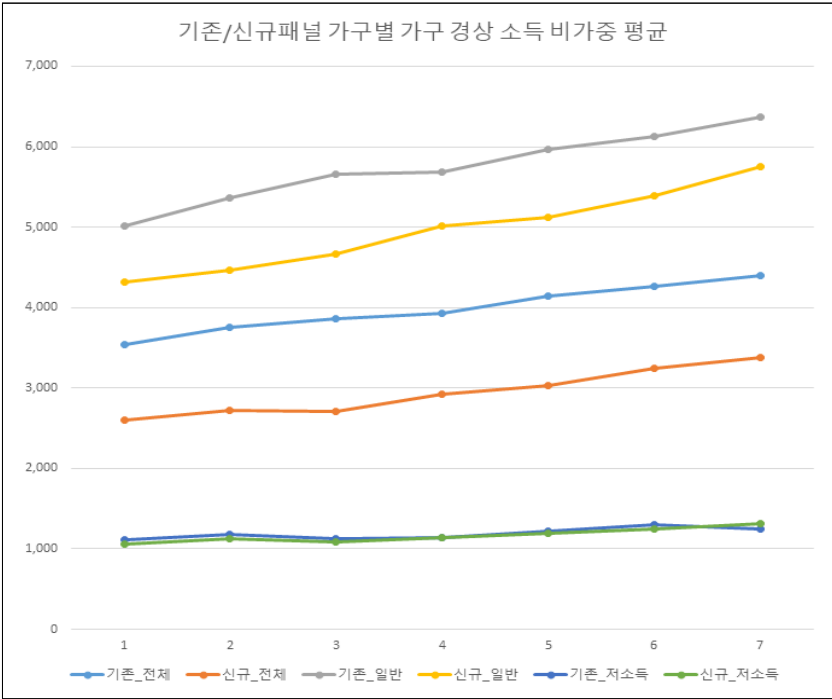
(단위: 만 원)

패널 구분	마모패턴	7차	8차	9차	10차	11차	12차	13차
신규 패널	np1	2,461	2,642	2,713	2,934	3,110	3,456	3,553
	np2	2,837	2,891	3,007	3,354	3,828	4,200	
	np3	2,487	2,880	2,867	3,100	3,358		
	np4	2,681	2,759	2,561	2,778			
	np5	3,867	3,236	3,442				
	np6	3,348	3,604					
	np7	3,049						

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

[그림 4-4-5] 기존/신규 패널 가구별 경상소득 비가중 평균 비교

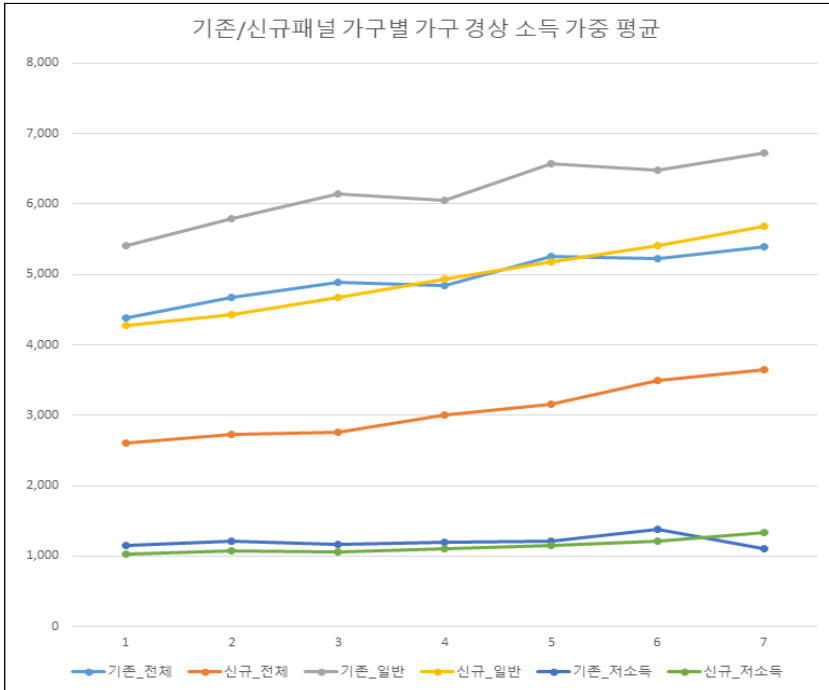
(단위: 만 원)



자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

[그림 4-4-6] 기존/신규 패널 가구별 경상소득 가중 평균 비교

(단위: 만 원)



자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

제5절 소결

이 장에서는 한국복지패널의 패널 마모패턴에 따른 가구 특성의 차이에 대한 분석, 차수별 가구별 가중치 분포 분석, 가중/비가중 가구 소득 추정 결과 분석을 진행하였다.

한국복지패널에서는 패널 마모로 인한 원표본 유지율 저하로 신규 표본이 추가되었다. 기존 패널과 신규 패널이 가구 특성, 가중치 분포, 가구 소득 추정에 있어 비슷한 모습을 보이는지 비교 분석하였다.

한국복지패널에서의 패널 마모는 다양한 형태로 발생한다. 단조패널 마모패턴 이외의 패턴을 보이는 가구가 있으나 원패널의 99.7%이상은 단조패널 마모패턴이다. 대부분의 가구는 차수별로 응답을 계속하다가 한번 패널에서 이탈을 하면 다시 패널로 돌아오지 않는 것이다. 단조 마모패턴을 이용하여 패널로 유지되는 가구와 패널에서 이탈하는 가구를 비교 분석하여 패널 마모가 조사에 미치는 영향을 파악하였다. 패널 마모 패턴으로 살펴본 패널 가구의 주요 특성 차이는 다음과 같다.

- (1-1) 기존 패널과 신규 패널 모두에서 소득 기준 구분 일반 가구 패널이 일찍 탈락하는 경향이 있다.
- (1-2) 전체적으로 신규 패널의 저소득 가구 비율이 기존 패널의 저소득 가구 비율보다 높게 나타났다.
- (2) 기존 패널 및 신규 패널 모두에서 1인, 2인 가구 패널이 유지되는 경향이 있다.
- (3) 기존 패널 및 신규 패널 모두에서 차수가 지남에 따라 단독 가구 형태가 증가하는 경향이 있다.
- (4) 기존 패널 및 신규 패널 모두에서 차수가 지남에 따라 1. 단독주택의 비율은 감소하는 경향을, 4. 아파트(일반, 임대)의 비율은 증가하는 경향을 보였다.

다음은 기존 패널과 신규 패널에 대한 마모패턴별 가중치 분포 분석과 가중/비가중 가구 소득 추정에서 나타난 주요 결과이다.

- (1) 차수가 지남에 따라 기존 패널 및 신규 패널 모두 가중치의 평균과 산포가 커진다. 대체로 기존 패널 일반 가구는 저소득 가구보다 큰 가중치 값을 갖는다.
- (2) 신규 패널의 가중치는 같은 차수 기존 패널의 가중치에 비하여 약 절반 정도이다. 신규 패널에서는 일반 가구와 저소득 가구의 가중

치가 크게 차이 나지 않는다.

- (3) 패널에서 초기에 이탈한 가구의 가처분소득과 경상소득이 높은 경향을 보였다.
- (4) 전체 가구와 일반 가구의 가중 가처분소득의 평균은 비가중 가처분소득의 평균보다 큰 값을 갖는다. 반면, 저소득 가구의 가처분소득은 가중/비가중 평균의 값이 서로 비슷하게 나타난다.
- (5) 전체 가구와 일반 가구의 가중 경상소득의 평균은 비가중 경상소득의 평균보다 큰 값을 갖는다. 반면, 저소득 가구의 경상소득은 가중/비가중 평균의 값이 서로 비슷하게 나타난다.

제 5 장

종단면 결측 패턴에 대한 통계적 가설 검정

제1절 종단적 결측자료 이슈

제2절 결측자료 모형

제3절 완전랜덤결측 검정

제4절 한국복지패널 사례 분석

제5절 결측 패턴에 대한 통계적 검증

제6절 소결

5

종단면 결측 패턴에 대한 통계적 가설 검정

제1절 종단적 결측자료 이슈

종단적 연구는 횡단면 연구에 비해 많은 장점을 가진다. 첫째, 동일한 개체로부터 반복적으로 측정된 자료는 완전하게 상관관계가 없는 경우 고정된 개체에 대해 종단적 연구가 횡단면 연구에 비해 효과적이다. 이러한 이유는 동일한 개체로부터 반복적으로 관찰함으로써 단일 개체에 대한 독립적인 정보를 순수하게 도출할 수 있기 때문이다.

둘째, 종단적 연구에서는 각 개체들은 자신을 통제할 수 있다. 즉, 개체 내 변동성이 개체 간 변동성보다 사실상 덜하기 때문에 통계적으로 강력한 검증이 가능하다. 시간의 흐름에 따라 관찰되는 자연현상이나 또는 연구 관찰에서 어떠한 외부적 충격에 의한 변동은 결과적으로 개체 내 변동의 결과와 관련될 수 있기 때문이다. 즉, 순수한 변동성은 측정오차로부터 개체 간 변동성을 배제함으로써 보다 효과적으로 처리와 관련된 효과만을 도출하여 분석할 수 있기 때문에 횡단면 연구에 비해 효율적인 추정이 가능하다.

셋째, 종단적 연구에서는 코호트 효과(cohort effect)로부터 aging 효과를 분리하여 분석할 수 있다. 이러한 코호트 효과는 개체 내에서 발생하는 변동을 잘못 처리할 수 있다.

마지막으로 종단적 자료는 개체 변동에 관한 정보를 제공하는 반면, 횡단면 자료는 그렇지 못하다. 개체들의 추세에 관한 통계적 추정은 모집단의 이질성과 개별적인 수준에서 성장과 변동의 결정을 보다 잘 이해할 수

있도록 한다.

이와 같은 종단적 연구 또는 자료의 장점에도 불구하고, 종단적 자료의 분석에서는 반드시 고려해야 할 사항들이 있다.

우선, 관찰값들은 서로 독립적이지 않기 때문에 통계분석 과정에서 종속성(dependency)을 반드시 고려해야 한다. 따라서 보다 복잡한 통계분석 모형을 필요로 하며, 이를 처리할 수 있는 프로그램 또한 복잡할 수밖에 없다. 또한 종단적 자료가 피할 수 없는 것이 결측치의 발생이다. 즉 종단적 횡수가 증가함에 따라 관측 대상의 탈락 등으로 인해 결측이 발생하게 된다. 이 경우 가장 간단히 고려할 수 있는 방법은 결측된 자료를 데이터로부터 제거하고, 완전히 관찰된 자료만을 이용하는 것이다. 또 하나의 방법은 마지막에 관찰된 관측값으로 결측치를 대체하여 이용하는 방법인데, 이 또한 결측의 유형에 따라 달라질 수 있기 때문에 현실성이 떨어진다.

한편 종단적 자료의 분석에서 고려해야 할 또 다른 측면은 종속변수뿐만 아니라, 설명변수들도 시간의 흐름에 따라 변화한다는 것이다. 따라서 종단적 분석에서는 시간의 변동에 따른 설명변수의 처리를 통해 횡단면 자료와는 달리 개체들의 역동성(dynamic)을 분석할 수 있는 통계적 방법론들을 고려해야 한다.

결과적으로 종단적 자료는 횡단면 자료와는 다르게 분석 과정에서 여러 측면이 동시에 고려되어야 하는 제한점이 있으나, 그럼에도 불구하고 개체들의 다양한 정보를 내포하고 있기 때문에 통계분석의 다양성을 꾀할 수 있다는 장점을 가진다.

앞에서 언급한 바와 같이 종단 연구에서 결측자료는 필연적으로 발생하게 된다. 개체가 어떤 특정 관측 시점에서 결측될 수 있으며, 연구 시점에서 일부 개체는 자료를 제공하지만, 다른 개체들은 그렇지 못할 수 있

다. 특정한 시점에서 접근된 개체들은 단지 일부 연구변수에 반응하여 결측인 자료를 만들어 내게 된다.

혼합효과회귀모형(MRM: Mixed-effect Regression Model)과 일반화 추정방정식(GEE: Generalized Estimating Equation) 모형에 대해 개체들은 동일한 시점 수에서 관측됨을 가정하지 않으며, 실제로 서로 다른 시점에서 관측될 수 있다. 이러한 개체별 관측 수에 대한 제한이 없기 때문에 주어진 조사 시점에서 결측이 발생한 개체들은 분석에서 제외되지 않는다. 그러나 이러한 모형들은 결측자료에 대한 서로 다른 가정을 만들어 내며, 따라서 특정한 분석을 수행할 때 이러한 결측자료에 대한 가정은 매우 중요하다.

최근까지 보건분야와 사회학 분야에서 종단적 연구에서 결측치 문제에 대한 다양한 논문들이 발표되었다(Demirtas, 2004b; Molenberghs et al., 2004; Hogan & Larid, 1997; Gornbein, Lazaro & Little. 1992). 이러한 연구들 중에서 Little(1995)은 결측치 연구에 대한 중요한 통계적인 체계를 수립하였고, 최근에 Hogan, Roy, and Korkontzelou(2004)는 결측 연구에 대한 다양한 실모형의 적용을 소개하였다. 한편, Little and Rubin(2002), Diggle et al.(2002), Verbeke and Molenberghs(2000) 등은 결측치 처리에 대한 연구의 기틀을 마련하였다.

본 연구에서는 연속형 자료에 대해 혼합효과회귀모형(MRM)을 이용한 종단적 모형에서 결측자료⁷⁾를 다루는 방법을 제시한다. 그러나 이 방법들의 적용은 시간의 흐름에 대해(GEE와 공분산 모형) 결측인 자료 또는 비정규적 결과(범주형 자료)를 허용한 그 외 다른 방법으로 보다 일반화시킬 수 있다.

7) “패널자료의 종단적 결측패턴에 관한 실증분석 연구”(손창균, 2011)에서 데이터를 확장(1~13차)하여 연구한 내용으로 분석방법은 재인용(포괄적 인용) 함.

제2절 결측자료 모형⁸⁾

결측자료 모형(Missing data mechanism)은 결측자료의 원인을 특성화하는 것이다. 다시 말해서 모형은 “왜 결측자료가 존재하는 가?”에 대한 기본적인 문제를 다루는 것이다. 나중에 다루겠지만, 종단적 자료 분석은 결측자료 모형에 결정적으로 좌우되기 때문에 적절한 분석을 선택함에 있어 모형을 고려해야 한다. 이러한 결측자료 모형에 대한 용어는 Rubin(1976)에 의해 소개되었다. 결측자료 방법을 접근 가능한 제고를 제시한 연구자들로서 Schafer and Graham(2002)에서 언급한 것처럼, 이 용어는 광범위하게 인용된 반면에 이해가 덜된 측면이 있다. 여기서는 모형에 대한 정의와 부수적으로 이들을 이해할 수 있도록 모의실험을 실시하였다.

이러한 모형들을 설명하기 위해 결측자료에 대한 어떤 기호를 도입할 필요가 있다. R_{ij} 는 만일 개체 i 가 시점 j 에서 응답하였다면 1의 값을 갖고, 만일 개체 i 가 시점 j 에서 응답하지 않았다면 0의 값을 가지는 지시변수이다. 여기서 종속변수 y 가 관측되는지 그렇지 않은지를 고려한다. 만일 어떤 연구가 n 개의 시점에서 측정을 요구한다면, $n \times 1$ 차원의 완전한 종속변수벡터는 다음과 같다.

$$\mathbf{y}' = (y_{i1}, y_{i2}, \dots, y_{in}) \quad (2.1)$$

임의의 개체에 대한 $n \times 1$ 차원의 결측자료 지시벡터는 다음과 같다.

$$\mathbf{R}_i' = (R_{i1}, R_{i2}, \dots, R_{in}) \quad (2.2)$$

8) 본절은 “패널자료의 종단적 결측패턴에 관한 실증분석 연구”, 손창균(2011)의 내용을 재인용(포괄적 인용)함.

여기서 R_{ij} 는 y_{ij} 가 관측되는지에 따라 1 또는 0의 값을 갖는다(즉, 개체 i 가 시점 j 에서 관측되면 $R_{ij} = 1$, 개체 i 가 시점 j 에서 관측되지 않으면 $R_{ij} = 0$ 을 갖는다).

R_i 에 기초하여 주어진 개체 i 에 대해 완전 종속변수벡터 y_i 를 관측된 자료 y_i^O 와 비관측된 자료 y_i^M 로 분해할 수 있다. 즉, $y_i = (y_i^O, y_i^M)'$. 여기서 y_i 를 개체 i 에 대해 계획되거나 또는 잠재적인 종속변수벡터로 고려하며, y_i^O 는 개체 i 에 대해 실제로 관측된 종속변수벡터로, y_i^M 은 개체 i 에 대해 계획되거나 관측되지 않은 종속변수벡터 원소로 볼 수 있다.

종단적 연구에서 하나의 개체가 탈락하여 결코 패널로 되돌아오지 않는 경우는 매우 흔하게 발생한다. 만일 결측자료가 단지 그러한 탈락만 있다면, 결측자료 지시자는 단순히 탈락 시점 변수 D_i 로 요약할 수 있다. 여기서 $D_i = j'$ 는 개체 i 가 $(j' - 1)$ 번째와 j' 사이 시점에서 탈락되었음을 나타낸다. 즉, $y_{i1}, y_{i2}, \dots, y_{i, j' - 1}$ 은 관측을 나타내고, $y_{i, j'}, \dots, y_{in}$ 은 결측을 나타낸다. 따라서 만일 하나의 개체가 맨 마지막 차수에서 결측되었다면, 그 개체에 대해서는 $D_i = n$ 으로 나타낸다. 시간에 따라 완전한 자료를 가진 개체에 대해 D_i 는 0으로 정의될 수 있다. 어떤 목적을 위해서 D_i 를 $n + 1$ 로 정의할 수 있는데 이는 가장 마지막 측정 시점을 초과하는 탈락 시점을 나타낸다.

1. 완전랜덤결측(MCAR: Missing Completely at Random)

결측자료를 처리하기 위한 가장 기본적인 방법은 결측자료들이 완전랜덤결측(missing completely at random)이라고 가정하는 것이다. 즉, 어떤 개체가 완전하게 확률적인 원인으로 특정 시점에 결측된 것으로 본

다. 이러한 모형을 ‘완전랜덤결측’ 또는 MCAR이라 한다. 이 경우 결측자료 지시변수 벡터 R_i 는 y_i^O 와 y_i^M 에 모두 독립적이다. 즉 이들은 관측된 종속변수와 비관측된 종속변수에 좌우되지 않음을 의미한다.

완전랜덤결측의 덜 엄격한 경우는 Little(1995)가 공변량-종속결측으로 언급한 바 있다. 여기서 결측자료 지시자는 완전히 관측된 공변량 X_i 와 연계될 수 있다. 이 경우는 시간이 경과함에 따라 결측의 증가를 허용하기 때문에 매우 중요하게 특별한 경우이다(변동하는 시간이 X_i 에 포함되어 있다). 거의 모든 종단적 연구에서 시간이 흐름에 따라 결측자료는 증가하며 따라서 완전랜덤결측에서 이러한 가정을 허용하는 것은 매우 유용하다. 공변량-종속결측하에서 완전랜덤결측은 조건부 독립성의 가정을 고려할 수 있다. 즉, 공변량 X_i 의 조건하에서 결측 R_i 은 관측치 y_i^O 와 비관측 종속변수 y_i^M 와 독립적이다. Fitzmaurice, Larid, and Ware(2004)가 지적했듯이, 이러한 사실은 드물지만 중요한 점이다. 결측자료 모형이 다음을 만족하면 완전랜덤결측이라 한다.

$$f(R_i|y_i, X_i, \beta) = f(R_i|X_i, \beta) \quad (2.3)$$

완전랜덤결측과 일관된 분석을 위해 X_i 에 있는 공변량으로서 결측과 연관된 변량을 고려하고, 결측 R_i 과 y_i^O 또는 y_i^M 간의 가능한 관계를 조정할 수 있다.

예를 들어 결측치와 종속변수들 둘 다 시간에 따라 증가한다고 가정하면, 결측치와 종속변수들은 연관되지 않는다(즉, 특정한 시점에서 결측치를 가진 개체와 과거에 y 값을 가진 개체 또는 결측이 없이 y 값을 관측한 개체 간에는 아무런 관련이 없다). 이 경우 만일 분석에서 공변량으로 시

간을 포함하지 않으면 자료 분석은 더 이상 완전랜덤결측에 일치하지 않는다. 주어진 분석에서 결측치들의 예측값을 포함하는 것은 매우 중요한데, Schafer and Graham(2002), Demirtas and Schafer(2003)은 종단적 설문지에서 “당신은 다음 연도 조사 기간 동안 이 조사에 계속 남아있겠습니까?”와 같은 질문을 포함하도록 권고하였다. 이러한 질문이 다음의 결측과 관련된다면 이 변수를 분석에 공변량으로 포함하여 비-완전랜덤결측 상황을 완전랜덤결측 상황으로 변환할 수 있을 것이다.

2. 랜덤결측(MAR: Missing at Random)

랜덤결측(MAR)은 완전히 관찰된 공변량 \mathbf{X}_i 와 관찰된 종속변수 벡터 \mathbf{y}_i^O 둘 다에 결측치들의 종속성을 허용함으로써 보다 중요하게 한 단계 더 확장된 모형이다. 랜덤결측은 이와 같은 조건하에서 결측치들이 관측되지 않은 종속변수 벡터 \mathbf{y}_i^M 과 관련되지 않는다고 가정한다. 다시 말하면 조건부 독립성 가정이라고 생각할 수 있다. 즉, 공변량 \mathbf{X}_i 와 관찰된 종속변수 벡터 \mathbf{y}_i^O 의 조건하에서 결측 \mathbf{R}_i 는 종속변수 벡터 \mathbf{y}_i^M 에 독립적이다. 즉, 다음과 같은 관계를 가지면 결측자료 모형은 랜덤결측이라 한다.

$$f(R_i | \mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\beta}) = f(R_i | \mathbf{y}_i^O, \mathbf{X}_i, \boldsymbol{\beta}) \quad (2.4)$$

예를 들어 랜덤결측하에서 종속변수들의 값이 어떤 기각치의 상한(또는 하한) 위쪽(또는 아래쪽)으로 떨어졌기 때문에 개체들이 종단연구에서 이탈된 것을 나타낸다. 예를 들어 복지패널 연구에서 월소득이 1000만원 이상인 개체들은 해당 연구에서 제외하게 된다(특정 시점에서 1000만원 이상의 월소득이 발생한 것이며, 향후 미래 시점에서 얻은 월소득은

아니다). 결측치에 관측 종속변수 벡터 y_i^O 의 종속성을 추가하면 완전랜덤결측과 랜덤결측 간의 중요한 차이점이 발생하며, 결과적으로 결측자료에 대한 제한적 가정을 줄이는 것이다. 이것이 완전랜덤결측 가정을 지지하는 강한 이유가 됨에도 불구하고, 결측자료 처리에 관한 많은 전문가들은 기본적인 접근 방법으로 랜덤결측을 적극적으로 추천한다. 나중에 살펴보겠지만, 관측치 벡터 y_i^O 의 포함 여부가 차이점이기 때문에 랜덤결측과 비교하여 완전랜덤결측이 적절한지에 대해 검증을 실시하였다.

랜덤결측은 결측자료들이 관측자료들과(X_i 와 y_i^O) 연관된다고 가정하지만, 결측은 비관측자료 y_i^M 과는 추가적으로 관련되지 않음을 가정한다. 결과적으로 종단적 분석모형에 대한 랜덤결측은 적절한 공변량들이 X_i 에 포함되고, y_i 의 분산-공분산 구조가 적절히 규정되어야 한다. 만일 이들 중 하나의 조건이 성립되지 않으면, 주어진 분석은 기본적인 랜덤결측 모형과 일치하지 않는다.

랜덤결측 모형하에서 y_i 의 분산-공분산 구조에 대해 오분류가 얼마나 편향을 야기하는지를 모의실험을 통해 보여 주고 있다. Schafer(1997), Demirtas(2004b)에서 다루었듯이, 랜덤결측과 매우 연관된 주제는 ‘무시 가능성’ 개념이다. (1) 만일 결측자료가 랜덤결측이고, (2) 이 자료 모형의 모수(즉, 종속변수에 대한 종단적 모형)와 결측 모형의 모수가 서로 다르면 결측자료 모형은 무시할 수 있다(Little and Rubin, 2002). 이러한 후자의 조건을 ‘모수 상이성’ 조건이라고 부르며, Shih(1992)에 의해 자세히 설명되었다. 필연적으로 이 조건은 두 과정에서의 모수들이 독립적임을 의미한다. 별도의 결측 모형을 고려하지 않아도 된다는 것은 결측자료 모형이 y_i 의 종단적 모형에서 무시될 수 있음을 의미한다. 만일 모수 상이성 조건이 성립하고 결측자료들이 랜덤결측이면, 이는 완전 우도모

형의 경우이며, 그렇지 않으면 GEE1과 같은 강한 완전랜덤결측 가정이 필요하게 된다.

3. 비랜덤결측(MNAR: Missing Not at Random)

비랜덤결측(MNAR)은 결측이 관측자료(X_i 와 y_i^O)를 고려한 후 비관측된 종속변수 벡터 y_i^M 과 관련된 상황을 의미한다. 여기서는 관측된 값(즉, y_i^M 의 값)과 결측 R_{ij} 간의 관계를 의미한다. 만일 개체들이 특정한 시점에서 그들의 종속변수 값들이 어떤 기각치의 아래쪽(또는 위쪽)에 떨어져 관측되지 않을 때 비랜덤결측이 발생할 수 있다. 예를 들어 랜덤결측과는 대조적으로 비랜덤결측은 관측치가 해당 시점에서 소득이 1000만 원 이상인 가구가 관측되지 않은 경우 발생한다.

때때로 연구가들은 비랜덤결측 가정으로 금연 연구에서 발견할 수 있음을 지적한다. 이러한 연구에서 종속변수는 특정한 시점에서 한 개인의 흡연 여부가 되며, 연구자는 만일 어떤 개체가 특정한 시점에서 결측이면, 그들은 흡연 때문이라고 가정한다. 물론 결측 R 과 비관측된 y 간의 연관성이 1임을 의미하기 때문에 이것은 비랜덤결측보다 강한 형식이다.

랜덤결측과 비랜덤결측 간의 차이점이 비관측된 자료 y_i^M 을 포함하는지의 여부이기 때문에 비랜덤결측 대비 랜덤결측을 확신하거나 기각할 수 있는 방법은 없다. 나중에 살펴보겠지만, 특정한 비랜덤결측 모형에 비교하여 특정한 랜덤결측 모형을 확신하거나 기각할 수 있지만, 이것이 보다 일반적인 모형을 고려하는 것은 아니다. 일반적으로 비랜덤결측의 이용은 자료가 심각하게 랜덤결측에 위배된다고 의심되는 경우로 한정하며, 이 경우 비랜덤결측의 가정하에 다양한 민감성 분석을 실시하는 것이 유용하다.

제3절 완전랜덤결측 검정

여기서 고려한 불완전 종단적 자료분석을 위한 근본적인 방법의 측면에서 Larid(1998)이 지적한 바와 같이 최대 우도추정과 종속변수의 평균과 분산-공분산 구조의 적절한 모형인 혼합모형은 무시할 수 있는 무응답이 존재하는 경우 적합한 추론을 제공한다. 완전우도과정하에서 추정될 때 공분산패턴모형(CPM: Covariance Pattern Model)에 대해 성립한다.

결측모형을 고려하지 않는 무응답(ignorable nonresponse)은 분석 모형에 따라 완전랜덤결측과 랜덤결측 모형을 포함할 수 있지만 비랜덤결측은 성립하지 않는다. 대체적으로 GEE1 모형은 공변량-종속 완전랜덤결측을 허용한다. 언급한 바와 같이 랜덤결측과 공변량-종속 완전랜덤결측 간의 필수적인 차이는 결측자료와 모형공변량 간의 종속성을 허용하는지 여부이며, 랜덤결측은 종속변수의 관측치와 관련된 결측자료를 허용한다. 따라서 보통의 GEE1 모형은 비랜덤결측과 공분산패턴모형의 가정보다 결측자료에 대해 강한 가정을 적용한다. 결과적으로 GEE 모형들 내에서 대다수 학자들이 제안한 가중-추정방정식(WEE: Weighted-estimating equations)은 비-완전랜덤 결측을 허용한다(Robins, Rotnizky, & Zhao, 1995; Rotnizky & Robins, 1999). Hogan et al.(2004)는 표준적인 소프트웨어를 이용하여 그러한 모형을 추정할 수 있음을 보였다.

이론에 따라 만일 결측자료가 완전랜덤결측이면 공변량 행렬 X_i 가 임의의 결측의 예측치를 포함할 경우 GEE1이나 비랜덤결측 모형을 이용한 분석은 적절하다. 그러나 만일 결측이 랜덤결측이면 GEE1은 적절하지 않은 반면 비랜덤결측 분석은 평균과 분산-공분산 구조가 올바르게 모형화 되는 한 허용될 수 있다. 따라서 주어진 분석에 대해 하나의 방법을 선택하는 것은 완전랜덤결측이 채택 가능한지 여부를 결정하는 가장 큰 요

인이 될 수 있다. 다시 말해서 완전랜덤결측과 랜덤결측 간의 결정적인 차이는, 완전랜덤결측은 결측이 종속변수의 관측값 y_i^O 에 의존하지 않는 반면에 랜덤결측은 의존할 수 있다는 것이다. 완전랜덤결측이 적절한지 여부를 검정하는 것은 y_i^O 을 포함한 분석에 근거할 수 있다. 그러한 검정은 Little(1988)과 Diggle and Hutchinson(1989)에 의해 설명되었으며, 여기서는 그것을 활용하여 설명한다.

예로서 모든 개체들이 시점 1에서 관측되었고, 어떤 개체들은 시점 2에서 결측이라고 가정하자. 변수 $D_i = 0$ 은 개체들이 두 시점 모두에서 관찰된 경우를 나타내며, $D_i = 1$ 은 단지 첫 번째 시점에서 관측된 경우를 나타낸다. 그러면 만일 y_1 자료를 두 그룹 $D_i = 0$ 과 $D_i = 1$ 에서 비교한다면, 완전랜덤결측의 경우 y_1 자료는 평균적으로 다르지 않아야 한다. 그러므로 두 그룹 간의 y_1 평균을 비교한 단순 t-검정은 이러한 차이를 규명하는데 이용할 수 있다. 보다 일반적으로 완전랜덤결측은 결측이 공변량을 허용하기 때문에 다음과 같은 회귀모형을 수행할 수 있다.

$$y_{i1} = \beta_0 + \beta_1 D_i + \beta_2 \mathbf{x}_i + \epsilon_i \quad (3.1)$$

여기서 β_2 는 \mathbf{x}_i 에 포함된 공변량에 대한 회귀계수 벡터이다.

또한 공변량과 탈락 변량 간의 상호작용을 고려하면

$$y_{i1} = \beta_0 + \beta_1 D_i + \beta_2 \mathbf{x}_i + \beta_3 (D_i \times \mathbf{x}_i) + \epsilon_i \quad (3.2)$$

이고, β_3 는 공변량과 이탈변수 간의 상호작용에 대한 회귀계수 벡터이다.

이 모형에서 완전랜덤결측은 $\beta_1 = \beta_3 = 0$ 이다. Ridout(1991)에서 언

급한 바와 같이 이러한 의문에 대해 다음과 같이 로지스틱 회귀로 탈락 여부를 모형화하는 것이 바람직하다.

$$\log \left[\frac{P(D_i = 1)}{1 - P(D_i = 1)} \right] = \alpha_0 + \alpha_1 y_{i1} + \alpha_2 x_i + \alpha_3 (y_{i1} \times x_i) \quad (3.3)$$

여기서 α_2 와 α_3 는 공변량 x_i 와 y_{i1} 간의 상호작용에 대한 회귀계수 벡터를 나타낸다. 완전랜덤결측은 결과적으로 $\alpha_1 = \alpha_3 = 0$ 가 된다.

만일 연구에서 두 시점 이상인 경우라면 이 모형은 방정식의 양변을 일반화시킬 수 있다. 먼저 방정식의 오른쪽 항에 대한 일반화를 고려하자. 예측치의 측면에서 y_{i1} 의 위치에서 한 개체의 y_{ij} 의 최종 관측치를 이용하여 위의 모형을 수정할 수 있다. 이 경우 관측된 종속변수 값의 함수 $h(y_i^O)$ 를 이용할 수 있다. 이에 대해 간단하게 평균으로 하면 다음과 같다.

$$\bar{y}_i = \frac{1}{n_i} \sum_j^{n_i} y_{ij} \quad (3.4)$$

보다 일반적으로 관측치들의 가중평균으로 h 를 고려할 수 있으며 다음과 같다.

$$h(y_i^O) = \sum_{j=1}^{n_i} w_j y_{ij} \quad (3.5)$$

w_j 에 대한 정의에 따라 이 식은 최초 관측 시점과 마지막 관측 시점 간의 차이, 최근 두 번째 시점과 마지막 시점 간의 차이 등과 같은 관측된 시

점에 대해 선형 추세를 허용한다. 주어진 연구에 따라 이용하는 관측된 종속변수 값의 함수를 선택할 수 있지만, 이러한 검정을 실시하는 데는 다양한 함수를 고려할 수 있다.

방정식의 좌측 항에 대해 2개 이상의 시점으로 이탈한 것에 대한 모형화는 이탈된 시점에 대한 모형화로 재구성할 수 있다. 이 경우 만일 개체 i 가 $j'-1$ 과 j' 시점 사이에 이탈되었다면 $D_i = j'$ 이다. 즉 $y_{i,j'-1}$ 가 관측되고, $y_{i,j'}, \dots, y_{i,n}$ 은 모두 결측이다. 여기서 중간에 결측된 것은 무시하며, 이탈 시점이 완전랜덤결측인지에 초점이 맞춰진다.

결측에 대한 단순화는 중도 결측이 완전랜덤결측이 된다는 범위에 합당하지만, 이탈은 잠재적으로 완전랜덤결측은 아니다. 이탈이 중도 결측보다는 완전랜덤결측이 되는 경향이 덜하다는 가정이 적절하기 때문에 이러한 단순화는 우선적으로 고려할 수 있는 자연스러운 방법이다. 더욱이 검정은 지시변수 R_{ij} 를 이용하기 때문에 여기서는 고려되지 않는다. 이탈이 있기까지 시간에 대한 모형의 측면에서 이산적 시간모형이나 집단화된 시간생존모형 분석을 이용할 수 있다(Allison, 1994; D'agostino et al., 1990; Singer and Willett, 1993). 로짓 링크함수를 이용하여 이 모형은 다음과 같이 표현된다.

$$\log \left[\frac{P(D_i = j | D_i \geq j)}{1 - P(D_i = j | D_i \geq j)} \right] = \alpha_{0j} + \alpha_1 h(\mathbf{y}_i^O) + \alpha_2 \mathbf{x}_i + \alpha_3 (h(\mathbf{y}_i^O) \times \mathbf{x}_i) \quad (3.6)$$

여기서 시점들은 $j = 1, \dots, n$ 이며, 최종 시점은 $D_i = n + 1$ 이다(이때 n 은 마지막 시점이다). 여기서 로짓은 주어진 시점에서 이탈할 확률이며, 이는 이미 발생한 것은 아니다. 앞에서 언급한 바와 같이 만일 $\alpha_1 = \alpha_3 = 0$ 을

기각하면, 결과적으로 완전랜덤결측을 기각하며 랜덤결측을 받아들인다.

생존분석 모형에서와 같이 이 모형에서 시간 변동 공변량을 고려할 수 있다. 즉, 보다 일반적으로 다음과 같이 모형을 설정할 수 있다.

$$\log \left[\frac{P(D_i = j | D_i \geq j)}{1 - P(D_i = j | D_i \geq j)} \right] = \alpha_{0j} + \alpha_1 h(\mathbf{y}_{ij}^O) + \alpha_2 \mathbf{x}_{ij} + \alpha_3 (h(\mathbf{y}_{ij}^O) \times \mathbf{x}_{ij}) \quad (3.7)$$

예를 들어 함수 $h(\mathbf{y}_{ij}^O)$ 는 시간에 따라 종속변수의 평균 변화를 나타낼 수 있다.

즉, $h(\mathbf{y}_{i1}^O) = y_{i1}$, $h(\mathbf{y}_{i2}^O) = (y_{i1} + y_{i2})/2$, $h(\mathbf{y}_{i3}^O) = (y_{i1} + y_{i2} + y_{i3})/3$ 이다.

이와 유사하게, x_{ij} 는 시간의 흐름에 따른 스트레스나 기분의 수준을 나타내거나, Demirtas and Schafer(2003)이 언급한 바와 같이 “이 연구에 얼마나 남아 있으시겠습니까?”와 같은 변수가 될 수 있다.

이 모형을 추정하기 위해서 표준적인 로지스틱 회귀프로그램을 이용할 수 있지만, 데이터 세트를 ‘개인-시점 데이터’로 생성하는 것이 필요하다 (Singer and Willett, 2003). 필연적으로 이러한 자료 세트에서 각 개인들은 이탈할 위험에 있는 시점 수만큼 많은 레코드에 포함된다.

예를 들어 4개의 시점을 가진 연구가 있다고 하고 모든 개체들이 첫 번째 시점(기준시점)에서 관측되었으나, 다음 연구에서 어떤 개체들은 이탈되었다고 하자. 그러면 $D_i = 1, 2, 3$ 은 첫 번째 관측을 제외한 나머지 시점에서의 참여 횟수를 나타내며, 만일 $D_i = 4$ 이면 완전하게 4개의 시점에 대해 관측된 것을 의미한다. 다음의 <표 5-3-1>은 4개의 개체에 대해 D_i 는 개인별로 시점별 연구 참여 횟수를 나타낸 것으로서 y_{ij} 가 1인 경우는

개체 i 가 시점 j 에서 이탈된 경우이며, 이 개체들은 더 이상 데이터에 레코드를 가지지 않는다. 0인 경우는 개체 i 가 시점 j 에서 관측된 경우이거나 아직 관측에서 이탈되지 않은 경우를 나타낸다. 여기서는 이탈이 있을 때까지의 시점을 고려하며, 하나의 개체가 종속변수에 더 이상 자료를 제공하지 않는 시점을 말한다.

〈표 5-3-1〉 개인-시점 자료의 예

ID	D_i	시점	y_{ij}
101	1	1	1
-	-	-	-
102	2	1	0
102	2	2	1
-	-	-	-
103	3	1	0
103	3	2	0
103	3	3	1
-	-	-	-
104	4	1	0
104	4	2	0
104	4	3	0

주어진 분석에 대해 위의 레코드는 시간 변동 또는 시간-불변이며, 종속변수의 관찰값의 함수로서 공변량에 정보를 포함할 수 있다. 예를 들어 성별(시간 불변)과 스트레스 수준(시간 변동)이라는 공변량이 이용 가능하다고 하자. 또한 종속변수의 관측값에 대해 시간 변동 평균을 고려하자. 그리고 $j = 0$ 을 기준시점이라 하자. 그러면 위와 같은 데이터 세트는 〈표 5-3-2〉와 같이 표현된다.

〈표 5-3-2〉 공변량을 가진 개인-시점 자료의 예

ID	D_i	시점	y_{ij}	Sex	Stress	y -평균
101	1	1	1	Sex101	Stress101,0	$y_{101,0}$
-	-	-	-	-	-	-
102	2	1	0	Sex102	Stress102,0	$y_{102,0}$
102	2	2	1	Sex102	Stress102,1	$(y_{101,0} + y_{102,1})/2$
-	-	-	-	-	-	-
103	3	1	0	Sex103	Stress103,0	$y_{103,0}$
103	3	2	0	Sex103	Stress103,1	$(y_{103,0} + y_{103,1})/2$
103	3	3	1	Sex103	Stress103,2	$(y_{103,0} + y_{103,1} + y_{103,2})/3$
-	-	-	-	-	-	-
104	4	1	0	Sex104	Stress104,0	$y_{104,0}$
104	4	2	0	Sex104	Stress104,1	$(y_{104,0} + y_{104,1})/2$
104	4	3	0	Sex104	Stress104,2	$(y_{104,0} + y_{104,1} + y_{104,2})/3$

이러한 방법으로 자료를 구성하면, 로지스틱 회귀분석은 시점, 성별, y 평균 그리고 상호작용들에 관한 이탈지시변수 y_{ij} 의 회귀를 수행할 수 있다. 비록 로짓 회귀를 이용하는 것이 보통이지만, 보충적 log-log(clog-log) 연 계함수를 고려할 수 있다.

$$\log(-\log(1 - P(D_i = j | D_i \geq j))) = \alpha_{0j} + \alpha_1 h(\mathbf{y}_{ij}^O) + \alpha_2 \mathbf{x}_{ij} + \alpha_3 (h(\mathbf{y}_{ij}^O) \times \mathbf{x}_{ij}) \quad (3.8)$$

단순히 이 모형에 대한 변환은 다른 연계함수로 정의된 로짓 대신에 clog-log 모형이다. 실제로 로짓이나 clog-log를 이러한 의도로 선택하는 것은 큰 문제가 되지 않는다.

제4절 한국복지패널 사례 분석

1. 한국복지패널 가구의 횡단적 변화

2006년 시작된 한국복지패널(KOWEPS)은 1차 웨이브(조사 기준 시점은 2005년임)에서 7072가구를 조사하여 2019년 현재 13차 웨이브까지 조사된 종단적 가구 조사이다. 시간이 경과함에 따라 원표본 가구가 점점 소멸함으로 인하여 7차 웨이브 이후 추가 표본 2000가구를 투입하여 13차 웨이브 현재 약 6581가구가 패널로 유지되고 있다.

조사표는 가구 및 개인 조사표, 부가조사표 등으로 구성되며, 가구원에 대한 추적 규칙은 원표본 가구원의 2세대까지(1차 분가 가구원) 추적하는 것으로 한다.

한국복지패널의 종단적 결측 패턴에 대해 언급하기에 앞서 패널의 횡단적 분포를 먼저 파악하는 것이 결측 패턴을 이해하는 데 도움이 될 수 있기 때문에 여기서는 횡단면적 특성을 파악하고자 한다. 횡단적 패턴 분석을 위해 패널 가구를 기준으로 가구원 수, 가구주 성별, 연령, 가구 소득, 가구주 혼인 상태 등에 대해 살펴보려고 한다. 이때의 데이터는 1~12차 웨이브 자료를 연계하여 사용하였다. 이와 더불어 횡단면적 분석에서는 보다 객관적인 비교를 위해 통계청의 KOSIS 자료를 함께 인용하였다.

가. 가구원 수 분포

〈표 5-4-1〉은 한국복지패널의 1~12차 웨이브까지 패널에 참여한 가구의 가구원 수 분포를 나타낸 것으로서 1인 가구의 비율은 1차 웨이브에서 22%였다가 12차 웨이브에서는 31%로 약 9%포인트 상승한 것으로 나타났으며, 2인 가구의 비율은 1차 웨이브 29.7%에서 32%로 약 3%포

인트 증가한 반면, 4인 가구는 1차 웨이브 22%에서 16%로 약 6%포인트 감소한 것으로 나타났다.

〈표 5-4-1〉 KOWEPS의 횡단면 가구원 수

(단위: 가구)

웨이브	1인	2인	3인	4인	5인	6인 이상	계
1	1,521	2,095	1,357	1,555	430	114	7,072
2	1,452	1,936	1,255	1,415	420	102	6,580
3	1,437	1,888	1,174	1,316	399	100	6,314
4	1,449	1,855	1,131	1,275	391	106	6,207
5	1,474	1,798	1,092	1,202	356	112	6,034
6	1,457	1,698	1,029	1,114	342	95	5,735
7	2,010	2,422	1,275	1,323	388	114	7,532
8	2,000	2,319	1,244	1,275	362	112	7,312
9	1,990	2,237	1,187	1,192	334	108	7,048
10	2,017	2,193	1,114	1,157	331	102	6,914
11	2,030	2,123	1,084	1,099	286	101	6,723
12	2,033	2,096	1,071	1,023	276	82	6,581

주: 비가중(unweighted) 실측치임

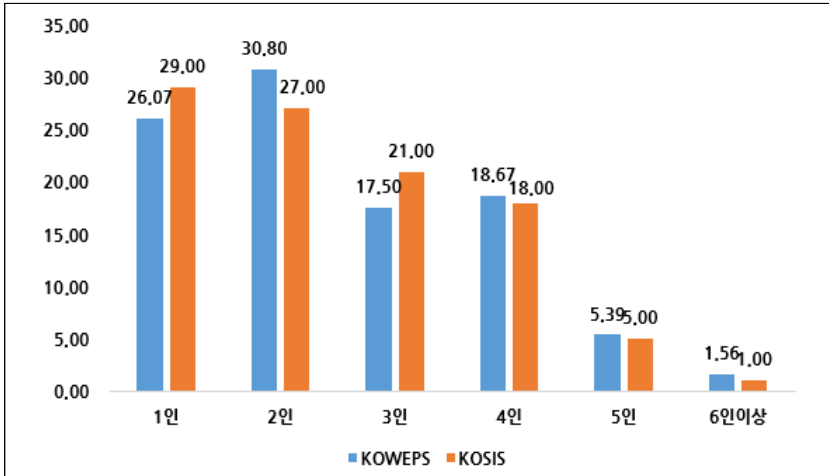
자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

한편 1~12차 웨이브를 pooling한 전체 패널을 하나의 표본으로 간주하여 가구원 수 분포를 분석하면⁹⁾ 1인 가구는 약 26%, 2인 가구는 31%, 3인 가구는 18%, 4인 가구는 19%, 5인 가구는 5.4%, 6인 이상은 1.6%로 2016년 기준으로 KOSIS와 비교하면 [그림 5-4-1]과 같이 나타난다. 가구원 수 분포의 경우 1인, 2인 가구의 비율에서 KOSIS와 차이를 보이고 있으며, 이러한 현상은 추후 논의될 결측 패턴과 관련성이 있을 것으로 추측된다.

9) 12차 웨이브에 응답한 가구만을 대상으로 비교분석하기보다는 12차 웨이브의 기준 연도인 2016년 기준으로 1~12차 웨이브를 하나의 전체 표본으로 간주하여 분석함으로써 가중의 효과를 부여할 수 있기 때문에 횡단적 비교의 경우 현재 시점 기준의 전체 표본으로 간주하여 분석하는 경우도 있음

[그림 5-4-1] KOWEPS와 KOSIS의 가구원 수 분포 비교

(단위: %)



자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료
 통계청, 인구총조사-“2016년 가구원수별 가구” 재가공, KOSIS(<http://kosis.kr>), 2019.
 08.01.

나. 가구주 성별 분포

가구주 성별 분포를 웨이브별로 비교하면, 1차 웨이브인 2005년 KOSIS는 남녀의 비율이 78:22, KOWEPS는 76:24로 KOWEPS가 KOSIS에 비해 여성 가구주의 비율이 약간 높았으며, 최종 차수인 12차 웨이브(2016년)의 경우 KOSIS는 70:30인 반면 KOWEPS는 67:33으로 KOWEPS의 여성 가구주 비중이 3%포인트 높게 나타나고 있다(〈표 5-4-2〉 참조). 앞서 분석한 방법에 따라 1~12차 웨이브를 pooling한 전체 표본과 KOSIS의 2016년도 자료를 비교하면 KOWEPS는 남녀 가구주의 비율이 71:29로 나타났고, KOSIS는 〈표 5-4-2〉의 70:30으로 나타나 KOWEPS의 여성 가구주 비율이 1%포인트 낮았다.

〈표 5-4-2〉 KOWEPS와 KOSIS의 가구주 성별 분포

(단위: %)

웨이브	KOWEPS		KOSIS	
	남	여	남	여
1	76	24	78	22
2	75	25	77	23
3	74	26	76	24
4	73	27	75	25
5	72	28	75	25
6	71	29	74	26
7	70	30	73	27
8	70	30	72	28
9	69	31	72	28
10	68	32	71	29
11	68	32	71	29
12	67	33	70	30

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료
 통계청, 인구총조사-“2005~2016년 가구주의 성별 가구” 재가공, KOSIS(<http://kosis.kr>),
 2019.08.01.

다. 가구주 연령 분포

가구주의 연령 분포는 다른 배경변수의 분포와 달리 KOWEPS와 KOSIS 간에 많은 차이를 보이는 것으로 나타났다.

30~40대 저연령대의 분포는 KOWEPS가 상대적으로 낮은 반면, 50대 이상 고연령층의 분포는 KOWEPS가 상대적으로 높게 나타나, 가구주 연령 분포에서는 왜곡 현상이 뚜렷하게 나타나고 있다(〈표 5-4-3〉 참조).

전체 표본을 pooling한 결과 KOWEPS의 경우 30대 이하는 17.6%, 40대는 18.2%, 50대는 16.7%, 60대는 19.5%, 70대 이상은 28.1%였으며, KOSIS는 2016년 기준 30대 미만은 24%, 40대는 23.3%, 50대는 24%, 60대는 15.5%, 70대 이상은 13.2%였다. 즉 가구주 연령 분포에서 50대 이하에서는 KOWEPS의 분포가 낮게 나타났고, 60대 이상에서는 KOWEPS가 높게 나타났으며, 이는 KOWEPS의 가구 및 가구주 분포에

서 나타난 특성이다.

〈표 5-4-3〉 KOWEPS와 KOSIS의 가구주 연령 분포 비교(2016년 기준)

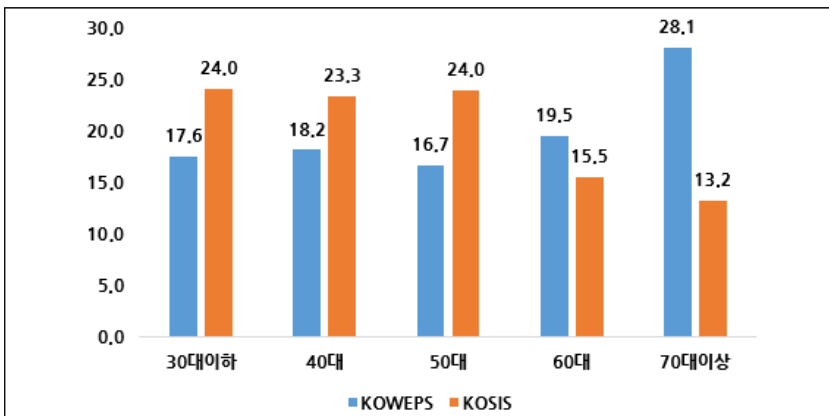
(단위: %)

차수	30대이하		40대		50대		60대		70대이상	
	KOWEPS	KOSIS	KOWEPS	KOSIS	KOWEPS	KOSIS	KOWEPS	KOSIS	KOWEPS	KOSIS
1	24	32	21	28	16	19	21	13	19	9
2	22	31	20	27	16	19	21	13	21	9
3	20	31	19	27	15	20	21	13	24	9
4	20	30	19	27	15	20	20	13	26	10
5	18	29	19	27	16	21	19	13	28	10
6	17	29	18	26	17	22	18	13	30	11
7	14	28	18	26	17	23	19	13	32	11
8	13	27	17	25	18	23	18	13	34	12
9	13	26	17	25	18	24	17	14	36	12
10	12	25	17	25	18	24	17	14	37	13
11	11	25	16	24	17	24	17	15	38	13
12	28	24	17	23	18	24	26	15	11	13

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일], 내부자료
통계청, 인구총조사-“2016년 가구주의 연령별 가구” 재가공, KOSIS(<http://kosis.kr>), 2019.08.01.

〔그림 5-4-2〕 KOWEPS와 KOSIS의 가구주 연령 분포 비교(2016년 기준)

(단위: %)



자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일], 내부자료
통계청, 인구총조사-“2016년 가구주의 연령별 가구” 재가공, KOSIS(<http://kosis.kr>), 2019.08.01.

라. 가구주 혼인 상태 분포

가구주 혼인 상태는 미혼, 유배우, 사별, 이혼 등 4개 범주에 대해 웨이브별로 분포를 비교하였다. KOWEPS와 KOSIS의 분포를 비교한 결과 미혼과 사별에서 분포의 차이가 크게 나타났다. 즉, KOWEPS가 KOSIS에 비해 미혼 비율은 낮게, 사별 비율은 높게 나타났으며, 유배우 비율은 웨이브별로 약간의 변동을 보였다.

〈표 5-4-4〉 KOSIS와 KOWEPS의 가구주 혼인 상태 비교

(단위: %)

차수	KOWEPS				KOSIS			
	미혼	유배우	사별	이혼	미혼	유배우	사별	이혼
1	7.1	68.2	18.1	6.6	12.8	70.4	11.1	5.7
2	6.7	67.2	19.3	6.8	13.4	69.5	11.1	6.0
3	6.3	66.3	20.4	7.0	13.8	68.7	11.1	6.4
4	6.1	65.6	20.9	7.3	14.3	67.9	11.1	6.7
5	6.5	64.3	21.5	7.8	14.7	67.1	11.1	7.1
6	6.1	63.8	22.2	7.9	15.1	66.3	11.1	7.4
7	6.0	62.4	23.4	8.2	15.3	65.8	11.1	7.7
8	6.0	61.8	23.8	8.5	15.5	65.3	11.1	8.1
9	5.8	61.4	24.4	8.5	15.7	64.8	11.2	8.4
10	5.8	60.2	25.1	8.8	15.9	64.3	11.2	8.7
11	6.0	59.3	25.7	9.0	16.0	63.8	11.2	9.0
12	6.5	58.7	25.8	9.0	16.3	63.1	11.2	9.3

주: 혼인 상태에 대해 무응답 및 비해당은 제외하고 분석하였음.

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

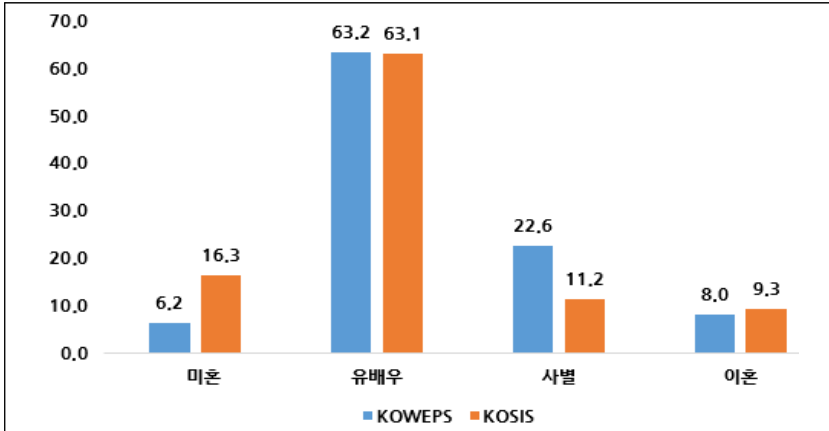
통계청, 장래가구추계-“2016년 가구주의 혼인상태별 가구” 재가공, KOSIS(<http://kosis.kr>), 2019.08.01.

패널 가구 전체를 pooling 데이터를 이용하여 2016년 KOSIS와 비교한 결과 미혼은 KOWEPS가 6.2%, KOSIS는 16.3%로 10.1%포인트 차이가 발생하고, 유배우의 경우 KOWEPS와 KOSIS의 분포가 63.2%와 63.1%로 거의 일치하며, 사별의 경우 KOWEPS는 22.6%, KOSIS는 11.2%로 11.4%포인트 차이가 발생하였고, 미혼의 경우 KOWEPS가

8.0%, KOSIS는 9.3%로 1.3% 포인트 차이가 나타났다.

[그림 5-4-3] KOWEPS와 KOSIS의 가구주 혼인 상태 분포 비교

(단위: %)



자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료
통계청, 장래가구추계-“2016년 가구의 혼인상태별 가구” 재가공, KOSIS(<http://kosis.kr>),
2019.08.01.

마. 가구 소득 분포

가구 소득의 경우 통계청의 가계동향 조사자료를 이용하여 KOWEPS의 가구 소득과 분포를 비교하고자 한다. 가계동향조사(2006~2016)의 경우 신분류 기준으로 가구원 2인 이상의 경상소득을 기준으로 각 연도의 분기별 평균을 이용하여 KOWEPS의 연간 경상소득과 비교하였다.

전체 패널 표본을 pooling하여 가구원 2인 이상에 대한 평균 가구 소득을 추정한 결과 501만 2000원으로 산출되었으며, KOSIS의 2016년도 평균 경상소득이 423만 4000원으로 추정된 것으로 볼 때 KOWEPS 소득 분포가 상대적으로 높았으나, 각 차수별로 비교해 보면 KOWEPS의 가구 소득 증가 속도가 KOSIS에 비해 빠르게 진행되고 있는 것으로 나타

났다([그림 5-4-4] 참조).

〈표 5-4-5〉 KOWEPS와 KOSIS의 경상소득 비교

(단위: 천원)

조사 차수	KOWEPS	KOSIS
1	3,490	3,542
2	3,858	3,626
3	4,417	3,721
4	4,588	3,781
5	4,805	3,730
6	5,199	3,842
7	4,971	3,908
8	5,345	4,063
9	5,501	4,106
10	5,582	4,182
11	6,119	4,223
12	6,132	4,234

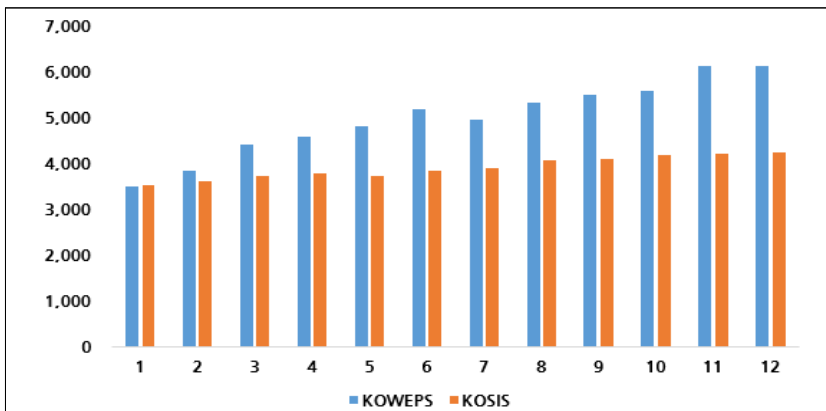
주: 2인 이상 전체 가구를 대상으로 횡단면 가중치를 부여하여 추제한 소득임.

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

통계청, 가계동향조사(신분류)-“2006~2016년 가구당 월평균 가계수지” 재가공, KOSIS (<http://kosis.kr>), 2019.08.01.

[그림 5-4-4] KOWEPS와 KOSIS의 가구 평균 소득 분포 비교

(단위: 천원)



자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

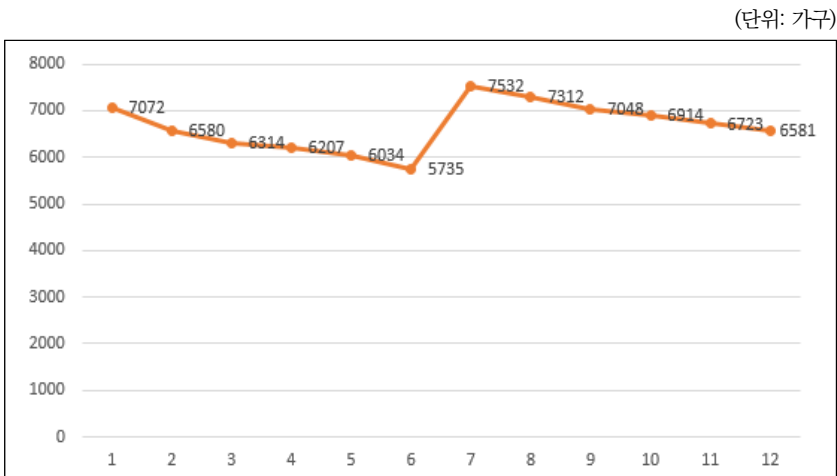
통계청, 가계동향조사(신분류)-“2006~2016년 가구당 월평균 가계수지” 재가공, KOSIS (<http://kosis.kr>), 2019.08.01.

2. 한국복지패널 가구의 종단적 변화

패널 가구의 종단적 변화를 파악하고, 본 연구에서 다루고자 하는 가구 탈락의 패턴 유형을 통계적 모형으로 가설 검정하기 위해 먼저 1~12차 패널 가구에 대해 종단적으로 가구 형태의 변화와 의미를 파악하고자 한다.

2006년 1차 웨이브 당시 패널로 참여한 가구는 7072가구이며, 2011년 6차 웨이브에서 5735가구까지 줄어들었다가, 2012년 7차 웨이브에 약 2000가구를 추가하여 새롭게 패널을 구성하였고, 그 이후 12차 웨이브인 2016년 6581가구가 유지되고 있다([그림 5-4-5] 참조).

[그림 5-4-5] KOWEPS 차수별 유지 현황



자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

1차 웨이브 이후 12차까지 계속 참여한 가구는 4309가구로 1차 웨이브 7072가구에서 출발하여 약 61%가 지속적으로 참여한 것으로 파악되고 있다(<표 5-4-6> 참조).

또한 패널에 단속적으로 참여한 가구와 12차까지 지속적으로 참여한 가구 간의 특성을 파악하기 위해 가구 소득을 비교해 보면 지속적으로 참여한 가구의 소득이 약간 높게 나타나고 있어 가구의 소득 수준이 패널 지속 참여 여부에 영향을 미치는 것으로 파악된다(〈표 5-4-7〉 참조).

〈표 5-4-6〉 1회 이상 패널에 참여한 가구 수

(단위: 가구, %)

패널 참여 횟수	빈도	백분율	누적빈도	누적백분율
1	819	8.29	819	8.29
2	619	6.27	1,438	14.56
3	380	3.85	1,818	18.40
4	413	4.18	2,231	22.58
5	430	4.35	2,661	26.94
6	1,638	16.58	4,299	43.52
7	217	2.2	4,516	45.71
8	279	2.82	4,795	48.54
9	236	2.39	5,031	50.93
10	287	2.91	5,318	53.83
11	252	2.55	5,570	56.38
12	4,309	43.62	9,879	100.00

주: 비가중(unweighted) 실측치임.

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

한편 두 집단의 소득 편차를 살펴보면 단속적으로 패널에 참여한 가구들의 소득 편차가 상대적으로 크게 나타나고 있어 지속적인 패널 참여 여부에 소득이 일정 정도 영향을 준 것으로 사료된다.

〈표 5-4-7〉 지속적 참여 여부에 대한 패널 가구의 소득 비교

(단위: 가구, 만 원)

구분	가구 수	소득	평균	표준편차	신뢰하한	신뢰상한
비지속참여 패널 가구	5,570	경상소득	3,273.3	3,194.3	3,189.4	3,357.2
		가처분소득	3,012.7	2,924.1	2,935.9	3,089.5
지속참여 패널 가구	4,309	경상소득	3,337.0	2,867.9	3,251.4	3,422.7
		가처분소득	3,080.9	2,578.0	3,003.9	3,157.9

주: 가구 수는 비가중(unweighted) 실측치임.

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

1차 웨이브 이후 가구주 변동과 패널 지속 참여 여부를 분석해 보면 12차까지 지속적으로 참여한 가구 중에 가구주 변동이 있는 가구는 약 15%인 641가구이고, 가구주 변동이 없는 가구는 85%인 3668가구로 파악되었다. 또 패널에 1회 이상 참여한 가구 중 가구주 변동이 있는 가구는 전체의 약 10%인 981가구이고, 나머지 가구는 가구주 변동이 없었던 것으로 나타났다(〈표 5-4-8〉 참조).

〈표 5-4-8〉 가구주 변동 유무와 패널 참여 횟수

(단위: 가구)

패널 참여 횟수	가구주 변동		
	변동 없음	변동 있음	합계
1	819	0	819
2	610	9	619
3	370	10	380
4	396	17	413
5	396	34	430
6	1,540	98	1,638
7	199	18	217
8	248	31	279
9	201	35	236
10	247	40	287
11	204	48	252
12	3,668	641	4,309
합계	8,898	981	9,879

주: 비가중(unweighted) 실측치임.

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일], 내부자료

다음으로 가구주 변동과 소득 간의 관계를 파악해 보면 가구주 변동이 없는 가구의 가구 소득이 가구주 변동이 있는 가구의 가구 소득에 비해 높게 나타나고 있으며, 이러한 차이는 통계적으로 유의하게 나타났다(〈표 5-4-9〉 참조).

〈표 5-4-9〉 가구주 변동 유무와 가구 소득

(단위: 가구, 만 원)

가구주 변동	가구 수	소득	평균	표준편차	95% 신뢰하한	95% 신뢰상한
변동 없음	8,898	경상소득	3,392.3	3,094.6	3,328	3,456.6
		가처분소득	3,124	2,811.1	3,065.6	3,182.5
변동 있음	981	경상소득	2,474.1	2,539	2,315.1	2,633.2
		가처분소득	2,302.3	2,337.8	2,155.8	2,448.8

주: 가구 수는 비가중(unweighted) 실측치임.

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

패널 참여 지속 여부 및 가구주 변동에 대한 가구의 소득을 살펴보면 전술한 바와 같이 가구주 변동이 있으면서 소득이 낮은 그룹이 패널에 단속적으로 참여하는 경향성을 보이는 것으로 나타났다.

12회차까지 지속적으로 참여한 가구와 그렇지 않은 가구에 대해 가구주 변동이 없는 가구와 변동이 있는 가구 간의 소득을 비교한 결과에서는 특별한 경향성을 파악하기 어려우나, 횡단면적으로는 앞서 분석한 바와 같이 소득이 높은 그룹의 패널 지속 경향성이 높았으며, 이 중에서도 가구주 변동이 없는 가구의 소득이 높게 나타나고 있다(〈표 5-4-10〉 참조).

〈표 5-4-10〉 가구주 변동 유무와 패널 참여 횟수별 가구 소득 비교

(단위: 가구, 만 원)

참여 횟수	가구주 변동	가구 수	평균 소득	표준편차	95% 신뢰하한	95% 신뢰상한
1	변동무	819	3,484	4,651	3,165	3,803
2	변동무	610	3,314	2,661	3,102	3,526
	변동유	9	2,168	1,142	1,291	3,046
3	변동무	370	3,428	3,135	3,107	3,748
	변동유	10	2,311	1,514	1,227	3,394
4	변동무	396	3,431	3,446	3,090	3,771
	변동유	17	3,166	2,503	1,879	4,453
5	변동무	396	3,250	2,745	2,979	3,521
	변동유	34	2,660	1,626	2,093	3,228
6	변동무	1,540	2,907	2,670	2,773	3,040
	변동유	98	2,389	2,996	1,788	2,990
7	변동무	199	3,283	3,042	2,858	3,708
	변동유	18	3,095	2,018	2,091	4,099

참여 횟수	가구주 변동	가구 수	평균 소득	표준편차	95% 신뢰하한	95% 신뢰상한
8	변동무	248	3,746	3,033	3,367	4,126
	변동유	31	2,716	1,689	2,097	3,336
9	변동무	201	3,737	2,854	3,340	4,134
	변동유	35	2,921	1,881	2,275	3,567
10	변동무	247	3,848	3,109	3,458	4,237
	변동유	40	2,832	1,984	2,198	3,467
11	변동무	204	3,843	3,179	3,404	4,282
	변동유	48	2,570	2,106	1,959	3,182
12	변동무	3,668	3,504	2,869	3,411	3,597
	변동유	641	2,383	2,667	2,176	2,590

주: 가구 수는 비가중(unweighted) 실측치임.

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

한편 패널 참여 가구의 가구원 수 변동을 파악하기 위해 먼저 패널 참여 횟수별 가구원 수의 분포를 파악한 결과 1인 가구의 비율이 1회 참여 가구의 경우 27%, 2회 참여 가구의 경우 30.2% 등으로 나타났으며, 12회 모두 참여한 가구는 22.7%로 파악되고 있다(〈표 5-4-11〉 참조).

〈표 5-4-11〉 패널 참여 횟수별 가구원 수

(단위: 케이스)

참여 횟수	가구원 수						합계
	1인	2인	3인	4인	5인	6인 이상	
1	220	208	151	190	44	6	819
2	374	292	216	270	74	12	1,238
3	354	312	233	187	39	15	1,140
4	583	353	316	305	69	26	1,652
5	703	466	422	441	109	9	2,150
6	3,300	3,575	1,375	1,235	255	88	9,828
7	573	342	272	254	65	13	1,519
8	679	537	376	516	99	25	2,232
9	683	461	436	417	104	23	2,124
10	915	563	547	667	123	55	2,870
11	769	683	498	587	167	68	2,772
12	11,717	16,868	9,171	9,877	3,167	908	51,708
합계	20,870	24,660	14,013	14,946	4,315	1,248	80,052

주: 비가중(unweighted) 실측치임.

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

패널에 참여한 가구들에 대해 가구원 수 변동(분가, 사망, 이혼, 출생, 합가 등의 사유)이 있는 가구와 그렇지 않은 가구의 가구 소득을 분석한 결과 가구원 수 변동(증/감)이 있는 가구의 소득이 높게 나타났으며, 통계적으로 유의미한 차이를 보이고 있다(〈표 5-4-12〉 참조).

〈표 5-4-12〉 가구원 수 변동 유무와 가구 소득

(단위: 가구, 만 원)

가구원 수 변동	가구 수	소득	평균	표준편차	95% 신뢰하한	95% 신뢰상한
변동 없음	5,777	경상소득	2,899.1	3,101.6	2,819.1	2,979.1
		가처분소득	2,679.4	2,852.8	2,605.8	2,753
변동 있음	4,102	경상소득	3,867.3	2,898.5	3,778.5	3,956
		가처분소득	3,553.8	2,585.5	3,474.6	3,632.9

주 : 가구 수는 비가중(unweighted) 실측치임.

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

패널에 참여한 가구의 가구원 변동 유무에 따른 가구 소득을 분석한 결과 가구원 변동이 있는 가구의 소득이 상대적으로 높았으며, 이러한 가구 소득의 차이는 통계적으로 유의미하게 나타났다. 가구원 변동에 따른 가구 소득의 변화는 패널 초기 시점에서는 가구원 변동 유무에 따른 소득 변화가 미미하지만, 패널이 지속될수록 가구원 변동이 있는 가구의 소득이 변동이 없는 가구의 소득보다 높게 나타나고 있다(〈표 5-4-13〉 참조).

〈표 5-4-13〉 가구주 변동 유무와 패널 참여 횟수별 가구 소득 비교

(단위: 가구, 만 원)

참여 횟수	가구원 변동	가구 수	평균 소득	표준편차	95% 신뢰하한	95% 신뢰상한
1	변동무	819	3,484	4,651	3,165	3,803
2	변동무	557	3,223	2,661	3,002	3,445
	변동유	62	3,964	2,449	3,342	4,586
3	변동무	300	3,045	2,831	2,724	3,367
	변동유	80	4,721	3,708	3,896	5,546
4	변동무	305	3,027	3,265	2,659	3,394
	변동유	108	4,531	3,581	3,847	5,214
5	변동무	266	2,844	2,747	2,512	3,175
	변동유	164	3,786	2,458	3,407	4,165
6	변동무	1,169	2,499	2,440	2,359	2,639
	변동유	469	3,815	3,041	3,539	4,091
7	변동무	126	2,645	2,889	2,135	3,154
	변동유	91	4,129	2,873	3,531	4,727
8	변동무	138	2,970	2,785	2,501	3,439
	변동유	141	4,279	2,934	3,791	4,768
9	변동무	109	3,062	2,792	2,532	3,593
	변동유	127	4,092	2,621	3,631	4,552
10	변동무	111	2,865	2,921	2,316	3,414
	변동유	176	4,237	2,928	3,801	4,672
11	변동무	91	2,837	2,919	2,229	3,444
	변동유	161	4,033	3,034	3,560	4,505
12	변동무	1,786	2,762	2,811	2,631	2,892
	변동유	2,523	3,744	2,838	3,634	3,855

주: 가구 수는 비가중(unweighted) 실측치임.

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

가구원 변동의 방향성(증/감)에 따른 가구 소득의 변화를 파악하기 위해 가구원이 감소한 경우와 가구원이 증가한 경우, 가구원 변동이 없는 경우로 구분하여 그룹별로 소득의 변화를 살펴본 결과 가구원이 증가한 경우 소득이 상대적으로 높았으며, 변화가 없는 경우가 가장 낮게 나타나고 있다(〈표 5-4-14〉 참조).

〈표 5-4-14〉 가구원 수 변화에 따른 경상소득 비교

(단위: 가구원 수, 만 원)

패널 참여 횟수	가구원 수 증감 여부	케이스	평균	표준편차	95% 신뢰하한	95% 신뢰상한
1	변화 없음	819	3,484	4,651	3,165	3,803
2	변화 없음	1,104	3,220	2,842	3,052	3,387
	평균 이하	67	3,549	2,331	2,980	4,117
	평균 이상	67	4,327	3,048	3,583	5,070
3	변화 없음	880	3,028	3,447	2,800	3,256
	평균 이하	119	4,440	6,412	3,276	5,604
	평균 이상	141	4,830	3,211	4,295	5,365
4	변화 없음	1,163	2,947	4,456	2,691	3,203
	평균 이하	247	4,520	7,466	3,584	5,455
	평균 이상	242	4,571	3,306	4,152	4,989
5	변화 없음	1,306	2,777	3,032	2,612	2,941
	평균 이하	408	3,334	2,644	3,077	3,592
	평균 이상	436	4,358	3,072	4,069	4,647
6	변화 없음	6,873	2,471	3,235	2,395	2,548
	평균 이하	1,413	3,272	2,936	3,119	3,425
	평균 이상	1,542	4,315	3,796	4,125	4,504
7	변화 없음	834	2,571	3,616	2,325	2,817
	평균 이하	322	3,197	2,510	2,922	3,473
	평균 이상	363	4,929	3,662	4,551	5,307
8	변화 없음	1,062	2,928	3,586	2,712	3,144
	평균 이하	546	3,684	5,552	3,217	4,151
	평균 이상	624	4,784	3,342	4,521	5,046
9	변화 없음	941	2,973	3,040	2,778	3,167
	평균 이하	543	3,165	2,737	2,934	3,396
	평균 이상	640	4,945	3,354	4,685	5,206
10	변화 없음	1,077	2,870	3,666	2,651	3,090
	평균 이하	814	3,160	3,133	2,944	3,375
	평균 이상	979	5,080	3,949	4,832	5,328
11	변화 없음	959	2,757	3,134	2,558	2,955
	평균 이하	825	3,313	3,083	3,103	3,524
	평균 이상	988	4,660	4,067	4,406	4,914
12	변화 없음	20,764	2,709	4,133	2,653	2,765
	평균 이하	14,608	3,109	2,927	3,061	3,156
	평균 이상	16,336	4,339	5,099	4,261	4,418

주: 가구원 수는 비가중(unweighted) 실측치임.

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

이러한 소득의 변화 양태는 패널 초기에는 가구원 변동과 소득 간에 유의미한 관계가 나타나지 않다가 웨이브가 지속될수록 이러한 관계가 밀접하게 나타나고 있는 것으로 판단된다.

3. 한국복지패널 가구원 수 변동의 요인 분석¹⁰⁾

시간이 경과함에 따라 가구 변동이 발생하게 되는데 이러한 변동의 효과가 여러 가지 다양한 가구 속성의 변화에 따른 것인지를 종단적으로 분석하고자 한다. 이를 위해 먼저 전체 패널 표본을 pooling한 자료를 분석하여 각 변동 요인을 파악하고자 한다.

가구를 집락(cluster)으로 고려할 때 가구 내 변동은 집락 내 변동(MSW)으로, 가구들 간의 변동은 집락 간 변동(MSB)으로 고려할 수 있다. 그러면 전체 패널의 변동에 가구원 수 변동이 얼마나 효과를 끼쳤는가를 파악할 수 있다. 즉, 패널 가구원 수 변동이 패널 가구원 전체 변동에서 차지하는 비율이 어느 정도인지를 파악함으로써 향후 다룰 결측 가구의 무응답 모형에 대한 가설 검정의 요인으로 활용할 수 있기 때문이다.

가. 가구원 변동의 효과 분석

가구원 변동 효과를 알아보기 위해 집락 간 변동(MSB)과 집락 내 변동(MSW)을 살펴보면 집락 간 변동이 상대적으로 크게 나타나고 있으며(12.02), 패널 가구 자체적인 가구원 변동(0.395)에 의한 효과보다는 전체적인 가구 변동에 기인한 가구원 변동이 크게 나타나고 있다(그림 5-4-6) 참조). 즉, 가구원 수의 변동은 패널 가구의 변동-어떠한 원인인

10) 한국복지패널 가구원 수 변동의 요인 분석은 가중치를 적용하지 않고 분석한 결과임

지는 아직 모르지만-에 기인한 것이며, 이러한 변동은 가구 내 변동에 따른 것으로 분석되고 있다.

[그림 5-4-6] 가구 변동 효과에 대한 분석

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8871	106606.8310	12.0175	30.42	<.0001
Error	71180	28123.7585	0.3951		
Corrected Total	80051	134730.5895			

R-Square	Coeff Var	Root MSE	h01_1 Mean
0.791259	24.99143	0.628576	2.515165

Source	DF	Anova SS	Mean Square	F Value	Pr > F
h_id	8871	106606.8310	12.0175	30.42	<.0001

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

나. 웨이브 효과 분석: 시간 효과

패널 가구의 가구원 변동의 원인을 웨이브 변동(시간 효과)으로 모형화하여 분석한 결과 전체 변동 중에서 시간 효과가 146.5로 크게 나타났다 ([그림 5-4-7] 참조).

[그림 5-4-7] 웨이브 효과에 대한 분석

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	1611.4340	146.4940	88.08	<.0001
Error	80040	133119.1555	1.6632		
Corrected Total	80051	134730.5895			

R-Square	Coeff Var	Root MSE	h01_1 Mean
0.011960	51.27436	1.289635	2.515165

Source	DF	Anova SS	Mean Square	F Value	Pr > F
wv_num	11	1611.434041	146.494004	88.08	<.0001

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

그러나 분석 모형은 통계적으로 유의하지만 실제로 패널의 웨이브 효과가 가구원 수 변동을 설명하는 것은 1.2%밖에 되지 않기 때문에 웨이브 효과에 의한 가구원 수 변동 모형은 검토가 필요한 것으로 보인다.

다. 혼인 상태 변동 효과

가구원 변동을 혼인 상태 변동에 대한 효과로 모형화하여 분석한 결과 가구주의 혼인 상태 변동에 따라 가구원 수 변동을 설명할 수 있는 것으로 나타났다. 즉 가구 간 혼인 상태 변동이 1만 2560.6으로 크게 나타났으며, 가구원 변동은 1.056으로 전체 가구원 변동 중 가구주 혼인 상태 변동 효과가 크게 나타나고 있다([그림 5-4-8] 참조).

[그림 5-4-8] 가구주 혼인 상태 변동 효과 분석

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	50242.4886	12560.6221	11900.4	<.0001
Error	80047	84488.1009	1.0555		
Corrected Total	80051	134730.5895			

R-Square	Coeff Var	Root MSE	h01_1 Mean
0.372911	40.84687	1.027366	2.515165

Source	DF	Anova SS	Mean Square	F Value	Pr > F
MAR	4	50242.48857	12560.62214	11900.4	<.0001

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일], 내부자료

라. 가구주 연령 변동 효과

가구원 변동을 가구주의 연령 변화에 대한 효과로 모형화하면 가구주 연령 변화에 따라 가구원 수 변동이 크게 나타나고 있다. 즉, 전체 변동

중에서 패널 가구의 가구주 연령대 변동에 따라 가구원 수 변동을 28.9% 정도 설명하는 것으로 나타났다([그림 5-4-9] 참조).

[그림 5-4-9] 가구주 연령 변동 효과 분석

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	39034.0838	9758.5209	8162.68	<.0001
Error	80047	95696.5057	1.1955		
Corrected Total	80051	134730.5895			

R-Square	Coeff Var	Root MSE	h01_1 Mean
0.289720	43.47194	1.093391	2.515165

Source	DF	Anova SS	Mean Square	F Value	Pr > F
AGE1	4	39034.08378	9758.52094	8162.68	<.0001

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

마. 전체 변동 효과

지금까지 각각의 개별 효과에 대한 가구원 수 변동 효과를 분석하였는데, 여기서는 전체 변수를 모두 고려하여 가구원 수 변동에 가장 큰 효과를 끼치는 변인이 무엇인지를 살펴보고자 한다. 이때 가구 변동(가구 id)을 제외하고 나머지 요인들에 대해 가구원 수에 어떠한 영향을 끼치는지를 분석한 결과 모형은 통계적으로 유의하며, 각 요인별로 가구원 수 변동에 영향을 끼치는지를 파악해 보면 모든 변인들이 영향을 끼치는 것으로 나타났다([그림 5-4-10] 참조).

[그림 5-4-10] 가구원 수 변동 효과 분석

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	63	75601.8543	1200.0294	1623.37	<.0001
Error	79988	59128.7352	0.7392		
Corrected Total	80051	134730.5895			

R-Square	Coeff Var	Root MSE	h01_1 Mean
0.561134	34.18380	0.859779	2.515165

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ww_num	11	1611.43404	146.49400	198.17	<.0001
MAR	4	49313.88518	12328.47130	16677.7	<.0001
AGE1	4	24444.84656	6111.21164	8267.11	<.0001
ww_num*MAR	44	231.68853	5.26565	7.12	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
ww_num	11	112.28663	10.20788	13.81	<.0001
MAR	4	12182.24977	3045.56244	4119.97	<.0001
AGE1	4	24186.00785	6046.50196	8179.57	<.0001
ww_num*MAR	44	231.68853	5.26565	7.12	<.0001

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일], 내부자료

4. 한국복지패널 가구 소득 변동의 요인 분석¹¹⁾

앞 절에서는 가구원 수 변동에 영향을 끼치는 가구주 특성을 기반으로 분석하였다면, 여기서는 가구 소득 변동에 영향을 끼치는 요인을 분석하고자 한다. 이때 분석 요인들은 앞 절에서 사용한 변인들을 사용하였다.

11) 한국복지패널 가구 소득 변동의 요인 분석은 가중치를 적용하지 않고 분석한 결과임

가. 가구 변동 분석

가구 변동 효과 분석은 가구의 소득 변동이 가구의 변동에 얼마나 영향을 받는지를 분석한 것이다. [그림 5-4-11]을 보면 가구 변동에 따라 가구 소득이 변동한 것으로 나타나고 있다. 모형의 설명력은 66.8%로 가구 변동과 소득 변동의 관계는 매우 높게 나타났다.

[그림 5-4-11] 가구 변동 효과 분석

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8871	54622.08670	6.15738	16.15	<.0001
Error	71180	27139.93238	0.38129		
Corrected Total	80051	81762.01908			

R-Square	Coeff Var	Root MSE	In_cin Mean
0.668062	8.029095	0.617484	7.690574

Source	DF	Anova SS	Mean Square	F Value	Pr > F
h_id	8871	54622.08670	6.15738	16.15	<.0001

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

나. 웨이브 효과 분석: 시간 효과

시간 효과 분석은 가구의 소득 변동이 시간의 변동에 얼마나 영향을 받는지를 분석한 것이다. [그림 5-4-12]를 보면 시간의 변동에 따라 가구 소득이 변동한 것으로 나타나고 있다.

[그림 5-4-12] 웨이브 효과에 대한 분석

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	213.63280	19.42116	19.06	<.0001
Error	80040	81548.38628	1.01885		
Corrected Total	80051	81762.01908			

R-Square	Coeff Var	Root MSE	In_cin Mean
0.002613	13.12488	1.009379	7.690574

Source	DF	Anova SS	Mean Square	F Value	Pr > F
ww_num	11	213.6328007	19.4211637	19.06	<.0001

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일], 내부자료

다. 혼인 상태 변동 효과

가구 소득 변동을 가구주 혼인 상태 변동에 대한 효과로 모형화하여 분석한 결과 가구주의 혼인 상태 변동에 따라 가구 소득 변동을 설명할 수 있는 것으로 나타났다([그림 5-4-13] 참조).

[그림 5-4-13] 혼인 상태 변화 효과에 대한 분석

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	18943.90374	4735.97594	6034.89	<.0001
Error	80047	62818.11533	0.78477		
Corrected Total	80051	81762.01908			

R-Square	Coeff Var	Root MSE	In_cin Mean
0.231696	11.51890	0.885870	7.690574

Source	DF	Anova SS	Mean Square	F Value	Pr > F
MAR	4	18943.90374	4735.97594	6034.89	<.0001

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일], 내부자료

라. 가구주 연령 변동 효과

가구 소득 변동을 가구주의 연령 변화에 대한 효과로 모형화하면 가구주 연령 변화에 따라 가구원 수 변동이 크게 나타나고 있다. 즉, 전체 변동 중에서 패널 가구의 가구주 연령대 변동에 따라 가구 소득 변동을 25.7% 정도 설명하는 것으로 나타났다([그림 5-4-14] 참조).

[그림 5-4-14] 가구주 연령 변동 효과 분석

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	21063.61306	5265.90327	6944.49	<.0001
Error	80047	60698.40602	0.75828		
Corrected Total	80051	81762.01908			

R-Square	Coeff Var	Root MSE	In_cin Mean
0.257621	11.32289	0.870795	7.690574

Source	DF	Anova SS	Mean Square	F Value	Pr > F
AGE1	4	21063.61306	5265.90327	6944.49	<.0001

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

마. 전체 변동 효과

각각의 개별 효과에 대한 가구 소득 변동 효과를 분석하였는데, 여기서는 전체 변수를 모두 고려하여 가구 소득 변동에 가장 큰 효과를 끼치는 변인이 무엇인지를 살펴보고자 한다.

이때 가구 변동(가구 id)을 제외하고 나머지 요인들에 대해 가구 소득에 어떠한 영향을 끼치는지를 분석한 결과 모형은 통계적으로 유의하며, 각 요인별로 가구원 수 변동에 영향을 끼치는지를 파악해 보면 모든 변인들이 영향을 끼치는 것으로 나타났다([그림 5-4-15] 참조).

[그림 5-4-15] 전체 변동 효과 분석

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8898	56693.99979	6.37154	18.08	<.0001
Error	71153	25068.01928	0.35231		
Corrected Total	80051	81762.01908			

R-Square	Coeff Var	Root MSE	In_cin Mean
0.693403	7.717997	0.593558	7.690574

Source	DF	Type I SS	Mean Square	F Value	Pr > F
h_id	8871	54622.08670	6.15738	17.48	<.0001
wv_num	11	99.90528	9.08230	25.78	<.0001
h01_1	8	1906.35327	238.29416	676.37	<.0001
AGE1	4	53.26005	13.31501	37.79	<.0001
MAR	4	12.39449	3.09862	8.80	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
wv_num	11	152.499901	13.863627	39.35	<.0001
h01_1	8	1286.427576	160.803447	456.42	<.0001
AGE1	4	49.152545	12.288136	34.88	<.0001
MAR	4	12.394491	3.098623	8.80	<.0001

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일], 내부자료

제5절 결측 패턴에 대한 통계적 검증¹²⁾

1. 종단적 결측에 대한 기초분석

본 절에서는 한국복지패널 2006~2018년의 12년도 자료를 이용하여 종단적인 결측 패턴을 통계적 방법으로 검증하고자 한다.

먼저 전술한 바와 같이 한국복지패널은 2006년에 1차 웨이브 조사를

12) 결측 패턴의 통계적 검증은 가중치를 적용하지 않고 분석한 결과임

통해 7072가구를 패널로 구축하였고, 2018년 12차 웨이브에서는 1차 웨이브 가구 중 4309가구에 대해 조사를 완료하였고, 12차 웨이브까지 조사하는 과정에서 완전히 탈락된 가구는 819가구로 나타났다. <표 5-5-1>에서는 결측 패턴에 따른 결측 횟수별로 가구의 빈도를 나타낸 것이다. <표 5-5-1>의 결측 횟수는 웨이브별 패널 참여 횟수의 반대 개념으로 보면 된다. <표 5-5-1>에서 제시한 결측 횟수는 결측 패턴 모형에서 이탈변수(drop variable)로 활용된다.

종단적 결측(또는 무응답)에 대한 패널 가구의 패턴을 검정하기 위해 결측에 투입할 변수들은 이미 가구의 변동 및 가구 소득의 변동에 영향을 끼칠 수 있는 가구 및 가구주의 특성변수를 정의하였기 때문에 이를 바탕으로 논리를 전개하고자 한다.

<표 5-5-1> 웨이브 간 결측 패턴에 따른 가구 수 분포

(단위: 가구 수, %)

결측 횟수	가구 수	백분율(%)
0	4,309	43.72
1	252	2.56
2	281	2.85
3	236	2.39
4	278	2.82
5	217	2.2
6	1,634	16.58
7	423	4.29
8	411	4.17
9	380	3.86
10	616	6.25
11	819	8.31
계	9,856	100.00

주: 비가중(unweighted) 실측치임.

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

가. 가구원 수에 대한 결측 횟수 차이 분석

결측 횟수는 0~11회까지 존재하며 가구원 수는 1~6인 이상으로 재조정하여 분석한 결과 가구원 수의 변동에 따라 결측 횟수에 차이가 나타나는 것으로 분석되었으며, 집단별 차이를 분석하기 위해 Duncan의 사후검정을 실시한 결과 1인 가구, 2·3·4인 가구, 5인 가구, 6인 이상 가구로 결측 횟수에 차이가 나타났다.

〈표 5-5-2〉 가구원 수에 따른 결측 횟수에 대한 ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	7201.838	1440.368	165.16	<.0001
Error	79853	696388.4	8.7209		
Corrected Total	79858	703590.3			

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

사후 분석 결과 1인 가구는 평균 2.38회의 결측이 발생한 반면, 5인 가구는 1.36회, 6인 이상 가구는 평균 1.23회로 가구원 수가 많을수록 패널 가구의 결측이 덜 발생하는 것으로 나타났다.

〈표 5-5-3〉 가구원 수에 따른 결측에 대한 사후 분석

Duncan Grouping	평균 결측 횟수	N	가구원 수
A	2.34871	20,774	1
B	1.81427	13,999	3
B	1.73359	14,928	4
B	1.72829	24,600	2
C	1.36071	4,311	5
D	1.22855	1,247	6

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

나. 지역에 따른 결측 횟수 차이 분석

지역은 복지패널에서 5개 권역으로 정의한 변수를 사용하였고, 지역에 따라 결측 횟수에 차이가 있는지를 분석한 결과, 거주 지역별로 결측 횟수에 차이가 있는 것으로 나타났으며, 사후 분석 결과 서울(=1), 광역시(=2), 시(=3)지역이 하나의 그룹으로 나타나며, 군지역(=4), 도농복합(=5)이 각각 별도의 그룹으로 나타났다.

〈표 5-5-4〉 지역에 따른 결측 횟수에 대한 ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	10111.77	2527.943	291.09	<.0001
Error	79854	693478.5	8.6843		
Corrected Total	79858	703590.3			

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

〈표 5-5-5〉 지역에 대한 사후 분석

Duncan Grouping	평균 결측 횟수	N	5개 권역
A	2.10821	27,678	3
A			
A	2.0524	20,991	2
A			
A	2.0222	12,660	1
B	1.48	2,450	5
C	1.20168	16,080	4

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

시지역(=3)이 결측 횟수가 2.11로 가장 빈번한 것으로 나타났으며, 상대적으로 군지역(=4)이 평균 1.20으로 가장 결측 횟수가 적게 나타났다.

다. 가구주 혼인 상태에 따른 결측 횟수 차이 분석

가구주 혼인 상태에 따른 결측 횟수의 차이에 대해 분석한 결과 <표 5-5-6>과 같이 혼인 상태별로 결측 횟수에 차이가 있는 것으로 나타났다. 미혼(=1)인 경우가 결측 횟수 평균 3.45회로 가장 많았으며, 유배우(=2)인 경우는 평균 1.69회로 가장 낮게 나타났다.

<표 5-5-6> 가구주 혼인 상태에 따른 결측 횟수에 대한 ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	13958.34	3489.585	404.07	<.0001
Error	79854	689631.9	8.6362		
Corrected Total	79858	703590.3			

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

<표 5-5-7> 가구주 혼인 상태에 대한 사후 분석

Duncan Grouping	평균 결측 횟수	N	혼인 상태
A	3.44653	4,891	1
B	2.1875	1,088	5
C	1.95198	6,268	4
C	1.92638	17,781	3
D	1.69088	49,831	2

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

라. 가구 소득에 따른 결측 횟수 차이 분석

가구 소득에 따른 결측 횟수의 차이에 대해 분석한 결과 <표 5-5-8>과 같이 소득 상태별로 결측 횟수에 차이가 있는 것으로 나타났다. 경상소득

의 5분위(20, 40, 60, 80)로 구분하여 결측 횟수에 차이가 있는지를 분석한 결과, 20분위, 60분위 소득 가구가 같은 그룹으로, 그리고 40분위 및 80분위 소득 가구가 같은 그룹으로 나타났다. 이들 각 그룹의 결측 횟수는 각각 평균 1.98번과 1.81번으로 소득계층별로 차이가 나타났다.

〈표 5-5-8〉 가구 소득 따른 결측 횟수에 대한 ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	570.7238	190.2413	21.61	<.0001
Error	79855	703019.6	8.8037		
Corrected Total	79858	703590.3			

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

〈표 5-5-9〉 가구 소득에 대한 사후 분석

Duncan Grouping	평균 결측 횟수	N	소득계층
A	2.00898	15,932	1
A			
A	1.94932	15,983	3
B	1.81122	31,948	4
B			
B	1.81014	15,996	2

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

마. 가구주 연령에 따른 결측 횟수 차이 분석

가구주 연령에 따른 결측 횟수의 차이에 대해 분석한 결과 〈표 5-5-10〉과 같이 연령대별로 결측 횟수에 차이가 있는 것으로 나타났다. 가구주 연령대는 30대 이하(=1), 40대(=2), 50대(=3), 60대(=4), 70대 이상(=5)으로 구분하여 분석하였다. 분석 결과 30대 이하가 평균 2.97번으로 가장 빈번하게 결측한 것으로 나타났으며, 상대적으로 60대는 평균 1.36번으로 결측 횟수가 낮게 나타났다.

〈표 5-5-10〉 가구주 연령에 따른 결측 횟수에 대한 ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	23181.86	5795.466	680.17	<.0001
Error	79854	680408.4	8.5207		
Corrected Total	79858	703590.3			

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

〈표 5-5-11〉 가구주 연령에 대한 사후 분석

Duncan Grouping	평균 결측 횟수	N	가구주연령대
A	2.97365	14,040	1
B	1.82666	22,412	5
B			
B	1.81279	14,508	2
C	1.48383	13,354	3
D	1.36243	15,545	4

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

바. 가구주 교육 수준에 따른 결측 횟수 차이 분석

가구주 교육 수준에 따른 결측 횟수의 차이에 대해 분석한 결과 〈표 5-5-12〉와 같이 교육 수준별로 결측 횟수에 차이가 있는 것으로 나타났다. 즉, 가구주의 교육 수준이 높을수록 결측 횟수가 낮았고, 상대적으로 교육 수준이 낮을수록 결측 횟수가 높았다. 구체적으로 중졸 이하(=1), 고졸 이하(=2), 대졸 이하(=3), 대학원 이상(=4)으로 재코딩하여 분석한 결과, 중졸 이하는 평균 결측 횟수가 1.69회로 가장 적고, 대졸 이하는 평균 2.35회로 가장 많았으며, 전반적으로는 가구주 교육 수준이 높을수록 결측 횟수가 많은 것으로 나타났다.

〈표 5-5-12〉 가구주 교육 수준에 따른 결측 횟수에 대한 ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	5525.689	1841.896	210.7	<.0001
Error	79860	698129.6	8.7419		
Corrected Total	79863	703655.3			

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

〈표 5-5-13〉 가구주 교육 수준에 대한 사후 분석

Duncan Grouping	평균 결측 횟수	N	가구주 교육 수준
A	2.34588	17,000	3
A			
A	2.27786	2,109	4
B	1.80723	22,643	2
C	1.68955	38,112	1

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

사. 가구주 경제활동 상태에 따른 결측 횟수 차이 분석

가구주의 경제활동 상태에 따른 결측 횟수의 차이에 대해 분석한 결과 〈표 5-5-14〉와 같이 경제활동 상태에 따라 결측 횟수에 차이가 있는 것으로 나타났다.

즉, 가구주가 상용직 임금근로자인 경우 평균 결측 횟수가 2.217회로 가장 많았고, 다음으로 실업 상태인 경우 2.143회, 임시일용직 임금근로자는 2.133회 등의 순으로 나타났으며, 무급가족 종사자(=7)는 1.297회로 가장 적었고, 다음으로 자영업(=6)인 경우 1.351회로 나타났다.

〈표 5-5-14〉 가구주 경제활동 상태에 따른 결측 횟수에 대한 ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	7467.046	933.3808	107.07	<.0001
Error	79876	696318.7	8.7175		
Corrected Total	79884	703785.7			

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

〈표 5-5-15〉 가구주 경제활동 상태에 대한 사후 분석

Duncan Grouping		평균 결측 횟수	N	가구주 경활 상태
	A	2.2177	18659	1
	A			
B	A	2.1428	1345	8
B	A			
B	A	2.1332	7087	2
B	A			
B	C	1.9806	1084	4
B	C			
B	C	1.9245	5496	3
	C			
	C	1.8821	27248	9
	C			
	C	1.8641	2023	5
	D	1.3505	16613	6
	D			
	D	1.297	330	7

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

아. 가구주 건강 상태에 따른 결측 횟수 차이 분석

가구주 건강 상태에 따른 결측 횟수의 차이에 대해 분석한 결과 〈표 5-5-16〉과 같이 건강 상태별로 결측 횟수에 차이가 있는 것으로 나타났

다. 즉, 가구주의 건강 상태가 좋은 경우(=1) 평균 결측 횟수가 1.98회로 가장 많았고, 건강 상태가 보통인 경우(=2) 1.67회로 가장 적었으며, 건강 상태가 나쁜 경우(=3)에는 1.85회로 보통인 경우보다 많게 나타났다.

〈표 5-5-16〉 가구주 건강 상태에 따른 결측 횟수에 대한 ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1214.76	607.3798	69.08	<.0001
Error	79949	702955.5	8.7925		
Corrected Total	79951	704170.3			

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

〈표 5-5-17〉 가구주 건강 상태에 대한 사후 분석

Duncan Grouping	평균 결측 횟수	N	가구주건강상태
A	1.98482	39983	1
B	1.846	22305	3
C	1.67391	17664	2

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

전체적으로 가구 관련 변수별로 결측 횟수의 변동에 대해 분석한 결과를 요약해 보면, 가구원 수가 적을수록 종단적 결측 횟수가 많았고, 지역 별로는 군지역에 거주하는 가구의 결측이 다른 지역 거주 가구에 비해 결측 횟수가 적었으며, 가구주 혼인 상태의 경우 사별 가구주의 결측 횟수가 가장 적었다. 가구 소득 수준별로는 저소득 가구일수록 결측 횟수가 적었고, 가구주 연령대별로는 30대 가구주의 결측 횟수가 가장 많았다. 가구주의 교육 수준에 따라서는 교육 수준이 높을수록 결측 횟수가 증가하는 것으로 나타났고, 가구주의 경제활동 상태에 따라서는 상용임시직

의 결측 횟수가 가장 많은 반면 무급가족 종사자인 경우 가장 적었다. 마지막으로 가구주의 건강 상태에 따른 결측 횟수의 차이는 건강한 경우의 결측 횟수가 가장 많았고, 상대적으로 보통인 경우가 가장 적었다.

이처럼 개별 변수에 대한 중단적 결측 패턴의 차이를 파악하여 얻은 결과는 최종적으로 복지패널 가구의 중단적 결측 패턴을 결정하는 중요한 요소로 고려될 수 있을 것이다.

2. 완전랜덤결측에 대한 통계적 가설 검정

앞 절에서 살펴본 바와 같이 패널 가구들의 결측 패턴에 대한 가설검증을 위해, 가구 및 가구주의 인구·사회학적 변수를 토대로 다음과 같이 변수들을 재정의하였다. 즉, 결측 확률이 관심변수와 연관이 있는지 또는 그 외 설명변수와 관심변수 간에 연관성이 있는지에 대한 검증을 통해 결측자료의 패턴을 식별할 수 있기 때문이다. 이를 위해 가설검증 모형에 사용하기 위한 변수를 다음의 <표 5-5-18>과 같이 정의하였다.

‘탈락(drop)’은 탈락 여부에 대한 지시변수로서 1차부터 12차까지 1회 이상의 결측이 있는 경우에 ‘1’, 그렇지 않은 경우에는 ‘0’의 값으로 재코딩하였다(오현민, 2016).

우선 결측 여부와 가구 특성에 대한 연관성을 분석하기 위해 앞서 제시한 모형(3.3)을 이용하여 분석하였다. 이 모형은 가구의 중단적 결측이 완전랜덤결측인지를 검정하는 것으로, 완전랜덤결측 가설이 위배될 경우 랜덤결측 또는 비랜덤결측의 가설을 도입하여 재검정할 것이다.

〈표 5-5-18〉 분석변수 목록

변수명	내용
drop	결측 지시변수 (결측=1, 비결측=0)
ln_cin	로그경상소득
fnum	가구원 수 (1인=1, 2인=2, 3인=3, 4인=4, 5인=5, 6인 이상=6)
age	가구주 연령대 (1=30대이하, 2=40대, 3=50대, 4=60대, 5=70대이상)
sex	가구주 성별 (1=남, 2=여)
edu	가구주 교육 수준 (1=중졸 이하, 2=고졸 이하, 3=대졸 이하, 4=대학원 이상)
mar	가구주 혼인 상태 (1=미혼, 2=유배우, 3=사별, 4=이혼, 5=기타)
reg5	권역 구분 (1=서울, 2=광역시, 3=시, 4=군, 5=도농복합)
eco	가구주 경제활동 상태 (1=상용직, 2=임시직, 3=일용직, 4=자활공공, 5=고용주, 6=자영업, 7=무급종사, 8=실업, 9=비경활)
health	가구주 건강 상태 (1=좋음, 2=보통, 3=나쁨)

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

(모형 1)

모형 1은 가구주의 기본적인 인구학적인 특성인 성별(sex), 연령(age), 혼인 상태(mar), 교육 수준(edu)과 가구 소득(ln_cin)이 가구의 종단 결측에 영향을 주는지를 검정하였다.

$$(\text{모형 1}) \quad \log \left[\frac{P(D_i = 1)}{1 - P(D_i = 1)} \right] = \alpha_0 + \alpha_1 y_{i1} + \alpha_2 \mathbf{x}_i + \alpha_3 (y_{i1} \times \mathbf{x}_i)$$

y_{i1} : 소득(ln_cin)

\mathbf{x}_i : 가구 기본 특성(성별, 연령, 혼인 상태, 교육 수준)

$y_{i1} \times \mathbf{x}_i$: 상호작용

〈표 5-5-19〉 (모형 1)에 대한 모수에 대한 검정

Effect	DF	Wald Chi-Square	Pr > ChiSq
ln_cin	1	11.4537	0.0007
SEX	1	45.1769	<.0001
AGE	4	134.7729	<.0001
MAR	4	108.7203	<.0001
EDU	3	7.8085	0.0501
ln_cin*SEX	1	51.3048	<.0001
ln_cin*AGE	4	91.4254	<.0001
ln_cin*MAR	4	97.3734	<.0001
ln_cin*EDU	3	15.4857	0.0014

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

〈표 5-5-19〉에서 살펴본 결과 모형에 포함된 모든 설명변수들의 회귀 계수들 특히 각 설명변수들의 계수는 0이 아니며, 소득과 설명변수들 간의 상호작용의 계수들 또한 0이 아닌 것으로 나타났다.

즉, 가구의 결측 여부가 가구의 기본적인 특성변수들에 영향을 받는 것으로 나타났으며, 소득이 높을수록 결측 가능성이 높고, 남성 가구주가 여성 가구주에 비해 결측 가능성이 높게 나타났다.

〈표 5-5-20〉에서 살펴본 결과 가구주의 연령은 70대 이상에 비해 50, 60대 연령층의 결측 확률이 낮게 나타났다. 혼인 상태별로는 미혼과 유배우에서 결측 확률이 낮은 반면, 사별과 이혼에서는 결측 확률이 높게 나타났다. 교육 수준은 중졸 이하가 결측 확률이 높고, 가구 소득과의 상호작용에 대해서는 모든 변수가 가구의 종단 결측에 영향을 주는 것으로 나타났다.

〈표 5-5-20〉 (모형 1)에 대한 로지스틱 분석

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.1303	0.1936	0.453	0.5009
ln_cin		1	-0.0822	0.0243	11.4537	0.0007
SEX	1	1	0.788	0.1172	45.1769	<.0001
AGE	1	1	0.2949	0.1827	2.6044	0.1066
	2	1	0.0779	0.1513	0.2654	0.6064
	3	1	-0.577	0.1344	18.4248	<.0001
	4	1	-1.0362	0.1474	49.4334	<.0001
MAR	1	1	-1.171	0.2506	21.8437	<.0001
	2	1	-1.4358	0.1771	65.7594	<.0001
	3	1	1.139	0.2157	27.8901	<.0001
	4	1	0.707	0.2185	10.4687	0.0012
EDU	1	1	0.441	0.1906	5.3523	0.0207
	2	1	0.2677	0.1775	2.2753	0.1314
	3	1	-0.1008	0.1916	0.2769	0.5987
ln_cin*SEX	1	1	-0.1137	0.0159	51.3048	<.0001
ln_cin*AGE	1	1	0.055	0.0224	6.0507	0.0139
	2	1	-0.00262	0.0187	0.0196	0.8886
	3	1	0.0437	0.0169	6.7151	0.0096
	4	1	0.0826	0.0193	18.3025	<.0001
ln_cin*MAR	1	1	0.2003	0.033	36.8067	<.0001
	2	1	0.162	0.0238	46.3854	<.0001
	3	1	-0.1551	0.0297	27.2126	<.0001
	4	1	-0.0916	0.0295	9.6206	0.0019
ln_cin*EDU	1	1	-0.0682	0.0233	8.5482	0.0035
	2	1	-0.0465	0.0211	4.8544	0.0276
	3	1	0.0266	0.0227	1.3746	0.241

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

(모형 2)

모형 2는 가구주의 기본적인 인구학적인 특성인 성별(sex), 연령(age), 혼인 상태(mar), 교육 수준(edu) 및 가구 소득(ln_cin)과 더불어 가구원 수(fnum), 가구주 건강 상태(health), 가구주 경찰 상태(eco), 지역(reg5)을 모두 모형에 포함하여 가구의 종단 결측에 영향을 주는지를 검정하였다.

$$(\text{모형 2}) \log \left[\frac{P(D_i = 1)}{1 - P(D_i = 1)} \right] = \alpha_0 + \alpha_1 y_{i1} + \alpha_2 \mathbf{x}_i + \alpha_3 (y_{i1} \times \mathbf{x}_i)$$

y_{i1} : 소득(ln_cin)

\mathbf{x}_i : 가구 변수(성별, 연령, 혼인 상태, 교육 수준, 가구원 수, 건강 상태, 경찰 상태, 지역)

$y_{i1} \times \mathbf{x}_i$: 상호작용

〈표 5-5-21〉로부터 가구주 건강 상태와 소득을 제외하고 나머지 설명 변수들이 가구의 결측에 영향을 주는 것으로 분석되었다.

〈표 5-5-21〉 (모형 2)에 대한 모수에 대한 검정

Effect	DF	Wald Chi-Square	Pr > ChiSq
ln_cin	1	4.3692	0.0366
SEX	1	60.6501	<.0001
AGE	4	102.2731	<.0001
MAR	4	38.3683	<.0001
EDU	3	13.3182	0.004
FNUM	5	32.2232	<.0001
Reg5	4	70.1172	<.0001
Eco	8	86.2175	<.0001
Health	2	3.2805	0.1939
ln_cin*FNUM	5	41.7907	<.0001
ln_cin*SEX	1	63.3102	<.0001
ln_cin*AGE	4	61.6403	<.0001
ln_cin*MAR	4	44.7439	<.0001
ln_cin*EDU	3	18.5365	0.0003
ln_cin*REG5	4	56.001	<.0001
ln_cin*ECO	8	68.8272	<.0001
ln_cin*Health	2	0.9386	0.6255

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일], 내부자료

다음으로 가구 변수와 소득이 결측 확률에 어떠한 영향을 끼치는지를 분석해 보면 〈표 5-5-22〉와 같다.

〈표 5-5-22〉 〈모형2〉에 대한 로지스틱 분석

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.5586	0.3368	21.4205	<.0001
ln_cin		1	0.0878	0.042	4.3692	0.0366
SEX	1	1	0.9353	0.1201	60.6501	<.0001
AGE	1	1	0.408	0.2021	4.076	0.0435
	2	1	0.2135	0.172	1.541	0.2145
	3	1	-0.5426	0.1433	14.3346	0.0002
	4	1	-1.0568	0.1642	41.4157	<.0001
MAR	1	1	-0.9106	0.275	10.9614	0.0009
	2	1	-0.9061	0.2353	14.8352	0.0001
	3	1	0.9371	0.2277	16.9306	<.0001
	4	1	0.2973	0.2213	1.8047	0.1791
EDU	1	1	0.5983	0.2065	8.3948	0.0038
	2	1	0.2706	0.1873	2.0882	0.1484
	3	1	-0.2532	0.2006	1.5928	0.2069
FNUM	1	1	0.4849	0.3018	2.5805	0.1082
	2	1	0.1377	0.2221	0.3842	0.5354
	3	1	0.1563	0.2366	0.4366	0.5088
	4	1	1.1928	0.2513	22.5265	<.0001
	5	1	0.7994	0.3678	4.7229	0.0298
REG5	1	1	-0.6241	0.1514	16.9975	<.0001
	2	1	0.4389	0.1313	11.1803	0.0008
	3	1	0.7804	0.1186	43.3262	<.0001
	4	1	0.0866	0.1442	0.3604	0.5483
ECO	1	1	0.6126	0.2903	4.4516	0.0349
	2	1	0.1784	0.3329	0.2872	0.592
	3	1	0.3572	0.3198	1.2482	0.2639
	4	1	-1.2289	0.7165	2.9414	0.0863
	5	1	-0.2715	0.359	0.5719	0.4495
	6	1	-0.8494	0.2229	14.5188	0.0001
	7	1	-0.7492	1.0292	0.5299	0.4667
	8	1	1.0052	0.377	7.1098	0.0077
HEALTH	1	1	-0.0175	0.1052	0.0276	0.8682
	2	1	-0.1637	0.1054	2.4134	0.1203
ln_cin*FNUM	1	1	0.0108	0.0383	0.0796	0.7779
	2	1	0.00741	0.0269	0.0759	0.7829
	3	1	-0.0114	0.0283	0.1616	0.6877
	4	1	-0.1607	0.0297	29.1813	<.0001
	5	1	-0.1428	0.0435	10.7938	0.001
ln_cin*SEX	1	1	-0.1294	0.0163	63.3102	<.0001
ln_cin*AGE	1	1	0.0429	0.0247	3.0146	0.0825
	2	1	-0.00625	0.0213	0.086	0.7694

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
ln_cin*MAR	3	1	0.0406	0.018	5.0726	0.0243
	4	1	0.0761	0.0214	12.6686	0.0004
	1	1	0.1424	0.0361	15.5343	<.0001
	2	1	0.1277	0.0308	17.1431	<.0001
	3	1	-0.1284	0.0313	16.8221	<.0001
ln_cin*EDU	4	1	-0.0444	0.0299	2.2139	0.1368
	1	1	-0.0815	0.0254	10.311	0.0013
	2	1	-0.0444	0.0223	3.9673	0.0464
	3	1	0.0414	0.0238	3.0295	0.0818
ln_cin*REG5	1	1	0.0895	0.0193	21.488	<.0001
	2	1	-0.0401	0.017	5.5591	0.0184
	3	1	-0.0744	0.0154	23.5047	<.0001
	4	1	-0.0515	0.0194	7.0623	0.0079
ln_cin*ECO	1	1	-0.0717	0.0357	4.041	0.0444
	2	1	-0.0191	0.042	0.2066	0.6495
	3	1	-0.0394	0.0417	0.8903	0.3454
	4	1	0.1857	0.0998	3.4611	0.0628
	5	1	0.033	0.0424	0.6046	0.4368
	6	1	0.0851	0.0288	8.7143	0.0032
	7	1	0.0917	0.1305	0.494	0.4821
	8	1	-0.1414	0.0507	7.7781	0.0053
ln_cin*HEALTH	1	1	0.000929	0.0136	0.0046	0.9457
	2	1	0.0121	0.014	0.7537	0.3853

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일], 내부자료

가구 기본 특성과 나머지 설명변수를 투입한 결과 가구 소득이 높을수록 결측 확률이 높게 나타나고 있다.

한편 남성 가구주의 결측 가능성이 높으며, 연령대는 70대 대비 50, 60대 가구주의 결측 확률이 낮게 나타났고, 혼인 상태는 미혼과 유배우자의 결측 확률이 높으며, 가구원 수는 4, 5인 가구의 결측 확률이 높았다. 지역은 도농지역 대비 서울지역 거주 가구의 결측 확률이 낮았으며, 상대적으로 광역시 및 시지역의 결측 확률이 높았다. 가구주 경제활동 상태별로는 자영자의 결측 확률이 낮았고, 상대적으로 실업자의 결측 확률이 높았다. 모형에서 상호작용 계수 α_3 에 대해 건강 상태를 제외하고 나머지는 0이 아닌 것으로 나타났다.

결과적으로 모형 1과 모형 2의 분석 내용으로부터 가구의 결측 패턴은 $\alpha_1 = \alpha_3 \neq 0$ 이므로 “귀무가설: 결측 패턴은 MACR이다”라는 가설은 받아들이지 못한다.

3. 랜덤결측에 대한 통계적 가설 검정

앞에서 실시한 완전랜덤결측 결측 모형에 대한 가설 검정 과정에서는 1~12차 웨이브에 관찰된 모든 가구를 대상으로 분석에 포함하였으나, 여기서는 결측자료 모형이 랜덤결측인지를 검정하기 때문에 앞서 언급한 바와 같이 1차 웨이브에 진입한 원가구만을 대상으로 하며, 종속변수의 결측이 설명변수의 영향에 기인한 것인지를 파악해야 한다. 따라서 보다 복잡한 모형을 가정하며, 이 경우 종속변수(여기서는 소득)의 평균을 산출하여 이 값을 사용한다.

완전랜덤결측 검정과 마찬가지로 가구 기본 특성을 나타내는 변수와 그 외 변수로 구분하여 모형 3과 모형 4로 구분한다.

(모형 3)

모형 3은 모형 1과 같이 가구주의 기본적인 인구학적인 특성인 성별(sex), 연령(age), 혼인 상태(mar), 교육 수준(edu)과 가구 소득(ln_cin)이 가구의 종단 결측에 영향을 주는지를 검정하였다.

$$\log(-\log(1 - P(D_i = j | D_i \geq j))) =$$

$$\alpha_{0j} + \alpha_1 h(y_{ij}^O) + \alpha_2 x_{ij} + \alpha_3 (h(y_{ij}^O) \times x_{ij})$$

$h(y_{i1}^0)$: 소득평균(ln_cin)

x_i : 가구 기본 특성(성별, 연령, 혼인 상태, 교육 수준)

$h(y_{i1}^0) \times x_i$: 상호작용

가설 검증을 위해 생존함수를 사용함으로써 가구의 종단적 결측이 관측자료에 의존하는지를 검정하는 것으로, 이는 랜덤결측 가설을 받아들일 수 있는 근거가 된다.

〈표 5-5-23〉 (모형 3)에 대한 모수에 대한 검정

Effect	DF	Wald Chi-Square	Pr > ChiSq
ymean	1	1137.5973	<.0001
SEX	1	120.1917	<.0001
AGE	4	71.5138	<.0001
MAR	4	97.3744	<.0001
EDU	3	14.9769	0.0018
ymean*SEX	1	136.5672	<.0001
ymean*AGE	4	36.7546	<.0001
ymean*MAR	4	100.5982	<.0001
ymean*EDU	3	22.6508	<.0001

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

〈표 5-5-23〉에서와 같이 모형 3에서 소득 평균과 공변량의 계수들 α_1 과 α_3 이 0이 아닌 것으로 분석되었다. 이때 분석 대상은 원패널을 대상으로 중간에 결측이 소득 또는 공변량에 의존하는지를 검증한 것이다.

〈표 5-5-24〉를 보면 평균 소득이 낮을수록 결측 확률이 높게 나타나는데, 여기서 평균 소득이 낮은 경우는 결과적으로 소득이 낮은 가구의 생존 기간이 길다는 것을 의미한다.

따라서 전체적으로 관심변수인 소득과 공변량값에 따라 가구의 종단적 결측에 영향을 주는 것으로 분석되었다.

즉, 가구주가 남성일수록 결측 확률이 높고, 연령이 높을수록 결측 확률이 낮게 나타나고 있다. 혼인 상태에서는 미혼 및 유배우일수록 결측 확률이 낮으며, 교육 수준은 대졸 이하가 결측 확률이 낮게 나타났다.

〈표 5-5-24〉 (모형 3)에 대한 로지스틱 분석

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
ymean		1	-0.1948	0.00578	1137.5973	<.0001
SEX	1	1	2.5333	0.2311	120.1917	<.0001
AGE	1	1	2.1873	0.3864	32.0529	<.0001
	2	1	1.6331	0.3707	19.4026	<.0001
	3	1	-0.7518	0.3574	4.4248	0.0354
	4	1	-2.8537	0.4169	46.8606	<.0001
MAR	1	1	-4.4665	0.538	68.9195	<.0001
	2	1	-2.331	0.3582	42.3441	<.0001
	3	1	1.1589	0.4661	6.1813	0.0129
	4	1	0.4789	0.4848	0.9756	0.3233
EDU	1	1	0.6612	0.3873	2.9141	0.0878
	2	1	-0.3613	0.35	1.0659	0.3019
	3	1	-1.2591	0.3501	12.9362	0.0003
ymean*SEX	1	1	-0.3556	0.0304	136.5672	<.0001
ymean*AGE	1	1	-0.1102	0.0476	5.3544	0.0207
	2	1	-0.1913	0.0461	17.2339	<.0001
	3	1	0.0624	0.0449	1.9329	0.1644
	4	1	0.2676	0.054	24.5776	<.0001
ymean*MAR	1	1	0.6005	0.0691	75.4437	<.0001
	2	1	0.3099	0.0472	43.1519	<.0001
	3	1	-0.1174	0.0632	3.453	0.0631
	4	1	-0.0536	0.0647	0.6867	0.4073
ymean*EDU	1	1	-0.1246	0.0501	6.1989	0.0128
	2	1	0.0206	0.0433	0.2271	0.6337
	3	1	0.1887	0.0426	19.5825	<.0001

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

(모형 4)

모형 4는 가구주의 기본적인 인구학적인 특성인 성별(sex), 연령(age), 혼인 상태(mar), 교육 수준(edu) 및 가구 소득(ln_cin)과 더불어 가구원 수(fnum), 가구주 건강 상태(health), 가구주 경찰 상태(eco), 지역(reg5)을 모두 모형에 포함하여 가구의 종단 결측에 영향을 주는지를 검정하였다.

$$\log(-\log(1 - P(D_i = j | D_i \geq j))) =$$

$$\alpha_{0j} + \alpha_1 h(\mathbf{y}_{ij}^O) + \alpha_2 \mathbf{x}_{ij} + \alpha_3 (h(\mathbf{y}_{ij}^O) \times \mathbf{x}_{ij})$$

$h(y_{it}^0)$: 소득평균(ln_cin)

x_i : 가구 변수 (성별, 연령, 혼인 상태, 교육 수준, 가구원 수, 건강 상태, 경찰 상태, 지역)

$h(y_{it}^0) \times x_i$: 상호작용

〈표 5-5-25〉 (모형 4)에 대한 모수에 대한 검정

Effect	DF	Wald Chi-Square	Pr > ChiSq
ymean	1	675.0951	<.0001
SEX	1	111.7923	<.0001
AGE	4	33.3748	<.0001
MAR	4	41.0659	<.0001
EDU	3	19.8066	0.0002
FNUM	5	100.875	<.0001
Reg5	4	25.2688	<.0001
Eco	8	78.4947	<.0001
Health	2	4.5169	0.1045
ymean*FNUM	5	122.3424	<.0001
ymean*SEX	1	124.4339	<.0001
ymean*AGE	4	17.3974	0.0016
ymean*MAR	4	46.8726	<.0001
ymean*EDU	3	26.3965	<.0001
ymean*REG5	4	26.3322	<.0001
ymean*ECO	8	66.6013	<.0001
ymean*Health	2	2.8071	0.2457

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일], 내부자료

〈표 5-5-26〉 (모형 4)에 대한 로지스틱 분석

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.2205	0.00849	675.0951	<.0001
ln_cin		1	2.5622	0.2423	111.7923	<.0001
SEX	1	1	1.2191	0.4442	7.5332	0.0061
AGE	1	1	0.6294	0.4192	2.255	0.1332
	2	1	-0.8411	0.3764	4.9948	0.0254
	3	1	-1.8563	0.4472	17.2278	<.0001
	4	1	-1.3203	0.5947	4.9299	0.0264

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
MAR	1	1	-2.9577	0.4978	35.3068	<.0001
	2	1	0.7094	0.5105	1.9308	0.1647
	3	1	0.4998	0.5128	0.95	0.3297
	4	1	1.712	0.451	14.4078	0.0001
EDU	1	1	0.1156	0.3832	0.0911	0.7628
	2	1	-0.9881	0.3901	6.415	0.0113
	3	1	-4.3817	0.5388	66.1447	<.0001
FNUM	1	1	-3.2724	0.4406	55.1531	<.0001
	2	1	-0.0626	0.4884	0.0164	0.8981
	3	1	1.2351	0.5779	4.5682	0.0326
	4	1	2.1902	0.9133	5.751	0.0165
	5	1	-1.4067	0.349	16.2436	<.0001
REG5	1	1	0.0893	0.2896	0.0951	0.7578
	2	1	0.6327	0.2558	6.1159	0.0134
	3	1	0.7321	0.3208	5.2092	0.0225
	4	1	2.8455	0.5649	25.373	<.0001
ECO	1	1	-0.5473	0.6944	0.6211	0.4307
	2	1	0.9164	0.8011	1.3088	0.2526
	3	1	-7.2472	1.7027	18.1168	<.0001
	4	1	1.1363	1.0538	1.1627	0.2809
	5	1	-1.9383	0.5913	10.7453	0.001
	6	1	4.6806	2.5063	3.4877	0.0618
	7	1	-0.6419	1.4993	0.1833	0.6685
	8	1	-0.2165	0.2466	0.7711	0.3799
HEALTH	1	1	-0.3423	0.2478	1.9077	0.1672
	2	1	0.6369	0.069	85.1574	<.0001
ymean*FNUM	1	1	0.41	0.0544	56.7163	<.0001
	2	1	0.0162	0.0598	0.0731	0.7869
	3	1	-0.1866	0.0698	7.1482	0.0075
	4	1	-0.3407	0.1103	9.5359	0.002
	5	1	-0.357	0.032	124.4339	<.0001
ymean*SEX	1	1	0.00382	0.0547	0.0049	0.9444
ymean*AGE	1	1	-0.0559	0.0522	1.1475	0.2841
	2	1	0.0764	0.0472	2.6134	0.106
	3	1	0.1388	0.0576	5.8041	0.016
	4	1	0.1574	0.0766	4.2191	0.04
ymean*MAR	1	1	0.4215	0.0648	42.3179	<.0001
	2	1	-0.0736	0.0685	1.1532	0.2829
	3	1	-0.0553	0.0679	0.6628	0.4156
	4	1	-0.2437	0.0573	18.0834	<.0001
ymean*EDU	1	1	-0.0286	0.0471	0.3684	0.5439
	2	1	0.154	0.0474	10.5747	0.0011
	3	1	0.1785	0.0436	16.7773	<.0001

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
ymean*REG5	1	1	-0.0165	0.0368	0.2018	0.6532
	2	1	-0.0751	0.0325	5.3402	0.0208
	3	1	-0.1049	0.0424	6.1123	0.0134
	4	1	-0.332	0.0709	21.9476	<.0001
ymean*ECO	1	1	0.0652	0.0878	0.5525	0.4573
	2	1	-0.1175	0.1039	1.2785	0.2582
	3	1	0.956	0.2288	17.4559	<.0001
	4	1	-0.1335	0.1266	1.1111	0.2918
	5	1	0.1954	0.0756	6.6808	0.0097
	6	1	-0.6109	0.3283	3.4637	0.0627
	7	1	0.0793	0.1886	0.1769	0.674
	8	1	0.0355	0.0317	1.2511	0.2633
ymean*HEALTH	1	1	0.0242	0.0326	0.5526	0.4573
	2	1	0.0121	0.014	0.7537	0.3853

자료: 한국보건사회연구원, 한국복지패널 1~12차 원자료[데이터파일]. 내부자료

〈표 5-5-25〉로부터 가구주의 건강 상태 변수를 제외하고 모든 공변량에 대해 0이 아닌 계수를 가지며, 특히 평균 소득과 공변량 간의 상호작용항의 계수 또한 0이 아니기 때문에 랜덤결측 가설을 받아들일 수 있다. 단, 건강 상태에서는 두 계수가 0이 되어 완전랜덤결측이 된다.

〈표 5-5-26〉으로부터 로지스틱 분석 결과를 살펴보면 남성 가구주의 결측 확률이 높고, 연령이 높아질수록 결측 확률이 낮아진다.

혼인 상태는 미혼인 경우 결측 확률이 낮고, 이혼자일수록 결측 확률이 높게 나타났다. 1인 가구의 결측 확률이 낮은 반면 3, 4인인 경우 결측 확률이 높으며, 상대적으로 5인 가구는 결측 확률이 낮았다. 지역의 경우 서울을 제외한 나머지 권역에 거주하는 가구의 결측 확률이 높게 나타났다.

상호작용항에 대해서는 먼저 가구원 수와 평균 소득 간에서는 가구원 수가 증가할수록 결측 확률이 낮아지며, 평균 소득과 성별의 상호 영향은 없으며, 평균 소득과 연령대의 상호작용은 연령이 높을수록 결측 확률이 높아지는 것으로 나타났다. 이는 평균 소득의 영향으로, 연령만의 효과에서는 반대 방향으로 나타났다.

이러한 현상은 다른 공변량에서도 나타나고 있는데, 결과적으로 패널의 웨이브가 지속됨에 따라 종단적 가구 결측은 소득 및 공변량의 효과에 영향을 받는다는 의미로서 랜덤결측 가설을 뒷받침하는 결과를 얻을 수 있다.

제6절 소결

이 장에서는 한국 복지패널 가구의 종단적 결측 패턴 특성을 살펴보았다.

이를 위해 먼저 패널의 횡단면적 특성을 파악하고자 통계청의 가구추계자료와 비교하였다. 가중치를 부여하지 않고 표본자료만으로 비교하여 표본의 변동성을 살펴보았으며, 그 결과 특별히 가구주 연령 분포에서 70대 이상의 분포가 높게 나타났다. 다른 변수들 중 특히 소득 분포의 경우 통계청 소득에 비해 높게 나타났다.

전반적으로 한국복지패널의 표본 자료는 통계청의 KOSIS 자료와 비교할 때 연령대별 차이와 소득 분포의 차이가 있으나, 이러한 차이는 가중치 조정으로 보정할 수 있을 것이다.

다음으로 한국복지패널의 종단적 변화를 파악하기 위해 먼저 패널 가구의 참여 횟수를 기준으로 다양한 변화를 살펴보았다. 이러한 분석은 궁극적으로 패널에 지속적으로 남아 있는 가구의 특성을 파악하는 데 도움이 되었다. 즉, 패널 참여 횟수가 많은 가구와 그렇지 않은 가구들 간의 차이를 파악하였으며, 이는 패널 가구의 결측 패턴과도 연계하여 특성을 파악하고자 하였다.

마지막으로 한국복지패널의 종단적 결측 패턴에 대해 보다 분석적 측면에서 결측 패턴에 대한 가설 검증 과정을 제시하였으며, 결과적으로 패

널 가구의 종단적 결측은 임의적으로 발생하기보다는 결측에 영향을 주는 공변량들이 다양하게 존재하여 이들에 의해 발생하는 것으로 설명할 수 있다.

이러한 연구는 패널 유지 방안 및 항목 결측에 대한 관리 방안 모색에 도움이 되며 단위 결측에 대한 적절한 통계적 방법 적용의 가이드라인이 될 수 있을 것이다.

제 6 장

결론

한국복지패널은 약 13년간의 시간이 지남에 따라 사회복지 및 경제학적인 측면에서의 패널 연구와 더불어 패널 데이터에 대한 근본적인 고민의 하나로 통계학적인 측면에서 패널 가구의 진입과 소멸과 같은 패널 자체의 변동성에 대한 연구가 필요한 시점이 되었다. 즉, 조사 연차가 6년 이상 지나면서 패널 탈락률이 커짐에 따라 표본 대표성을 보완하기 위해 1800가구를 신규 패널로 추가하였다. 이에 따라 처음부터 응답을 계속해 온 원패널과 신규 추가 패널을 합한 통합 패널을 대상으로 가중치 산출과 패널의 변동성에 대한 심층 연구를 실시하였다. 이를 위해 2장에서는 가중치의 개념을 설명하고 표본조사에서 가중치 산출 단계를 자세히 살펴 보았다.

3장에서는 한국노동패널과 한국복지패널의 가중치 산출 방식을 설명하였다. 한국노동패널과 한국복지패널은 우리나라에서 10년 이상 된 패널이기 때문에 패널 노후화 및 마모에 대한 이슈 및 연구를 지속적으로 해 오고 있다. 두 패널 모두 신규 표본 추가의 경험을 가지고 있기 때문에 여기에서는 두 패널의 가중치 부여 방법에 대한 비교를 통해 한국복지패널의 가중치 산정 과정을 전반적으로 짚어 보고자 하였다. 가중치는 연구 목적과 연구 설계 방법에 따라 산출 과정에 차이가 있을 수 있다. 한국노동패널과 한국복지패널의 가중치 산출 과정은 전반적으로 크게 다르지 않았다. 한국복지 패널의 가중치 산정 방법을 다른 패널과 비교하는 것은 한국복지패널의 가중치 산출 과정에 대해 전반적으로 검토 및 검증할 수 있고 향후 가중치 산출 과정에 있어서 보완하고 고려해야 할 점이 무엇인

지 고민을 할 수 있다는 점에서 필요한 과정이라고 할 수 있다.

4장에서는 한국복지패널의 패널 마모패턴에 따른 가구특성의 차이에 대한 분석, 차수별 가구별 가중치 분포 분석, 가중/비가중 가구 소득 추정 결과 분석을 진행하였다. 분석 결과, 저소득 가구는 패널에 잔류할 가능성이 큰 반면, 소득이 높은 가구 혹은 3인 이상 가구는 패널에서 이탈할 가능성이 컸다. 이러한 시간에 따른 패널 마모로 인하여 가중치 값이 증가함을 알 수 있고, 저소득 가구 가중치 증가량은 일반 가구에 비하여 적게 나타남을 알 수 있다.

5장에서는 패널 가구의 종단적 결측이 어떠한 원인에 의해 발생하는지를 통계적 관점에서 살펴봄으로써 단순하게 종단 결측을 다룰 경우 자료의 왜곡이 발생할 수 있다는 사실을 확인할 수 있었다. 즉, 패널 데이터의 분석에서 패널 가구의 결측 패턴에 부합되는 분석이 실시되어야 한다는 것이다. 단순히 종단적으로 결측된 모든 가구를 제거하고 완전하게 응답된 가구만을 대상으로 분석하는 것은 완전랜덤결측 결측 가정에서만 가능한 것이다. 따라서, 이 연구에서와 같이 결측 모형에 대한 검증을 실시한 후 그에 적합한 분석이 이루어질 경우 보다 신뢰성 있는 분석 결과를 도출할 수 있을 것이다.

이번 연구에서는 가중치 개념 설명, 가중치 산출 과정 설명, 표본관리 및 모집단 추정의 정도를 기하기 위한 마모패턴별 가구 특성 비교 분석, 패널 가구의 종단적 결측 패턴에 대한 통계적 가설 검정 분석을 다루었다. 이 연구는 패널자료의 품질 개선에 근거자료로 활용될 수 있다. 패널 연구에 있어서 대표성을 확보하기 위해 마모패턴 분석 및 가중치 산출 과정에 대한 검토는 지속적으로 이루어져야 할 것이다. 이러한 분석이 단발성 검토로 끝난다면 앞으로 가중치 산출 과정의 방향성과 마모에 대한 대처 방안을 논의하는 데 한계가 있을 수 있다. 한국복지패널의 연속적인

결과물로 분석이 이루어지고 통계 전문가뿐만 아니라 복지 전문가도 함께 논의하는 구조가 만들어진다면 한국복지패널의 품질은 지속적으로 관리될 수 있다.

한국노동패널의 경우 매해 패널자료 품질개선 연구가 발간되고 있다. 패널자료 품질개선 연구는 실험 연구, 표본 이탈, 항목 무응답 처리 방법, 표본 추가 등과 관련된 데이터 자체에 대한 내용을 다룬 보고서이다.

한국복지패널도 이번 심층분석 보고서를 기점으로 하여, 품질 개선을 포함한 다양한 통계적 이슈를 다룰 수 있는 연구가 지속되어야 하겠다.

한국복지패널 심층분석은 다양한 이슈를 중심으로 이루어질 수 있다. 앞으로 한국복지패널의 심층연구가 필요한 주제를 몇 가지 제안해 보면 다음과 같다. 한국복지패널은 13년 이상 된 중요한 패널로 세대 간 이전 연구가 가능하다. 이를 위해서는 한국복지패널 분가 가구가 원가구와 비교했을 때 어떠한 특성이 있는지에 대한 연구도 심층적으로 이루어져야 한다(부록 3 참고). 또한, 향후 신규 표본 추가를 위한 준비도 철저하게 되어 있어야 한다. 신규 표본 추가는 이전의 방법처럼 double sampling이 가능하지 않기 때문에, 어떻게 보정해 주어야 하는지에 대한 고민도 필요하다. 그리고, 다른 패널들에서 패널을 운영하면서 겪어온 경험 및 이슈를 공유할 수 있는 네트워크 구성도 필수적일 것이다. 한국복지패널의 품질 개선을 위해서 paradata 관리 틀도 재정립할 필요가 있다. 조사 방법과 관련해서도 시대의 흐름을 따라 웹(Web)을 활용한 조사(Computer Assisted Web Interviewing: CAWI)도 고려해볼 필요가 있다. 한국노동패널은 2018년 조사에서 이메일 주소를 가지고 있는 강력 거절 응답자만을 한정하여 CAWI 조사를 실시하였으며, 점진적으로 CAWI 병행도 고려하고 있다. 한국노동패널이 기준은 아니지만 한국복지패널도 조사의 어려움을 극복하기 위해 조사 방법의 다양화 문제를 검토해 볼 필요가 있

다. 마지막으로, 데이터 연계에 대한 관심이 높아지고 있는 시점에, 한국복지패널과 다른 자료와의 연계로 새로운 가치 창출을 할 수 있는 연구도 필요하다.

이러한 다양한 이슈들에 대한 한국복지패널 연구가 활발히 이루어진다면, 기존 데이터 생산 및 관리 사업에 더해 향후 한국복지패널 조사 연구의 안정적인 수행에 기여할 수 있으며, 데이터 품질을 개선하고 향상시키는 데 중요한 역할을 할 수 있을 것이다.

참고문헌 <<

- 강석훈. (2003). KLIPS의 가중치 부여방안 연구. 한국노동패널연구 2003-04.
- 김미곤, 여유진, 손창균, 이봉주, 강상경, 안상훈, . . . 구인회. (2006). 2006 한국복지패널 기초분석 보고서. 서울: 한국보건사회연구원.
- 김미곤, 여유진, 손창균, 이봉주, 강상경, 안상훈, . . . 구인회. (2007). 2007 한국복지패널 기초분석 보고서. 서울: 한국보건사회연구원.
- 김재광 (2009). 한국노동패널조사 표본추가 연구. 미발간 보고서, 서울: 한국노동연구원.
- 김태완, 오미애, 이봉주, 강상경, 박정민, 정원오, . . . 유조안. (2018). 2018년 한국복지패널 기초분석 보고서. 세종: 한국보건사회연구원.
- 이계오·임찬수. (2012). 한국노동패널 가중치 산출방안-원패널과 추가패널. 한국노동패널 학술대회 논문집, 서울: 한국노동패널 학술대회
- 이현주, 노대명, 오미애, 이봉주, 김태성, 구인회, . . . 안상훈. (2013). 2013년 한국복지패널 기초분석 보고서. 서울: 한국보건사회연구원.
- 손창균. (2011). 패널자료의 종단적 결측패턴에 관한 실증분석 연구. 한국조사연구학회 발표논문집, 서울: 한국조사연구학회, 273-275.
- 오현민. (2016). 웨이브 무응답 결측처리방법연구, (석사학위논문, 동국대학교 대학원, 서울).
- 천영민, 최형아, 권순범, 이상희. (2012). 한국노동패널조사(KILPS) 13차년도 기초분석보고서, 서울: 한국고용정보원.
- Allison, P. D. (1994) Using Panel Data to Estimate the Effects of Events. Sociological Methods & Research, 23(2), 174-199.
- Demirtas, H. (2004a), Modeling Incomplete Longitudinal Data. Journal of Modern Applied Statistical Methods, 3(2), 5.
- Demirtas, H. (2004b). Simulation driven inferences for multiply imputed longitudinal datasets. Statistica Neerlandica, 58(4), 466-482.

- Demirtas, H., & Schafer, J. L. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in medicine*, 22(16), 2553-2575.
- Diggle, P. J., & Hutchinson, M. F. (1989). On spline smoothing with autocorrelated errors. *Australian Journal of Statistics*, 31(1), 166-182.
- Diggle, P., Diggle, P. J., Heagerty, P., Liang, K. Y., Heagerty, P. J., and Zeger, S. (2002). *Analysis of longitudinal data*. London: Oxford University Press.
- D'agostino, R. B., Belanger, A., and D'agostino Jr, R. B. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4), 316-321.
- Duncan, G. J. (1995). *A simple method for weighting in household panel surveys*. Working Paper, Northwestern University.
- Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012). *Applied longitudinal analysis* (Vol. 998). New Kork: John Wiley & Sons.
- Gornbein, J. A., Lazaro, C. G., and Little, R. J. (1992). Incomplete data in repeated measures analysis. *Statistical methods in medical research*, 1(3), 275-295.
- Hogan, J. W., & Laird, N. M. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in medicine*, 16(3), 259-272.
- Hogan, J. W., Roy, J., and Korkontzelou, C. (2004). Handling drop-out in longitudinal studies. *Statistics in medicine*, 23(9), 1455-1497.
- Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in medicine*, 7(1-2), 305-315.
- Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical*

- Association, 83(404), 1198-1202.
- Little, R. J. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the american statistical association*, 90(431), 1112-1121.
- Little, R. J. A. and Rubin, D. B. (2002), *Statistical Analysis with Missing Data*, New York: John Wiley & Sons.
- Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M. G., Mallinckrodt, C., and Carroll, R. J. (2004). Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, 5(3), 445-464.
- Ridout, M. S., and Diggle, P. J. (1991). Testing for random dropouts in repeated measurement data. *Biometrics*, 1617-1621.
- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the american statistical association*, 90(429), 106-121.
- Rotnitzky, A., Scharfstein, D. O., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448), 1096-1120.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: CRC press.
- Schafer, J. L., and Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7(2), 147.
- Shih, W. J. (1992). On informative and random dropouts in longitudinal studies. *Biometrics*, 48(3), 970.
- Singer, J. D., & Willett, J. B. (1993). *It's about time: Using discrete-time*

survival analysis to study duration and the timing of events.
Journal of educational statistics, 18(2), 155-195.

Singer, J. D., Willett, J. B., & Willett, J. B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. London: Oxford university press.

Verbeke, G. and Molenberghs, G. (2000), Linear Mixture Models for Longitudinal Data. New York: Springer.

통계청. 2005년 인구주택총조사 가구 데이터

통계청, 국가통계포털(KOSIS), (<http://kosis.kr>), (2019.08.01. 인출)

한국보건사회연구원, 2006년 국민생활실태조사

한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료.

부록 1. 기존 패널 마모패턴 분석

〈부표 1-1-1〉 기존 패널 마모패턴별 가구 수

구분	패턴	가구 수	합계
원가구	"11111111111111"	4,170	2,758
	"11111111111110"	139	
	"11111111111100"	165	
	"11111111111000"	206	
	"11111111110000"	124	
	"11111111100000"	191	
	"11111111000000"	146	
	"11111100000000"	131	
	"11111000000000"	289	
	"11110000000000"	251	
	"11100000000000"	205	
	"11000000000000"	384	
	"10000000000000"	527	
분가 가구	"01111111111111"	30	998
	"01111111111110"	4	
	"01111111111000"	5	
	"01111111110000"	2	
	"01111111000000"	4	
	"01111110000000"	1	
	"01111000000000"	6	
	"01110000000000"	1	
	"01100000000000"	7	
	"01000000000000"	9	
	"00111111111111"	69	
	"00111111111110"	3	
	"00111011111110"	1	
	"00111111111100"	4	
	"00111111111000"	4	
	"00111111110000"	2	
	"00111111100000"	4	
	"00111111000000"	7	
	"00111110000000"	5	
	"00111000000000"	7	
	"00110000000000"	8	
	"00100000000000"	12	
	"00011111111111"	66	

구분	패턴	가구 수	합계
	"0001111111110"	1	
	"0001111111100"	2	
	"0001111111000"	1	
	"0001111110000"	5	
	"0001111100000"	4	
	"0001111000000"	4	
	"0001110000000"	5	
	"0001100000000"	3	
	"0001000000000"	14	
	"0000111111111"	68	
	"0000101111111"	1	
	"0000111111110"	1	
	"0000111111100"	4	
	"0000111111000"	2	
	"0000111110000"	3	
	"0000111100000"	5	
	"0000111000000"	7	
	"0000110000000"	5	
	"0000100000000"	14	
	"0000011111111"	52	
	"0000011111110"	2	
	"0000011111100"	2	
	"0000011111000"	2	
	"0000011110000"	7	
	"0000011100000"	4	
	"0000011000000"	1	
	"0000010000000"	2	
	"0000001111111"	56	
	"0000001111100"	4	
	"0000001111000"	3	
	"0000001110000"	4	
	"0000001100000"	6	
	"0000001000000"	5	
	"0000000111111"	62	
	"0000000111110"	1	
	"0000000111100"	3	
	"0000000111000"	3	
	"0000000110000"	2	
	"0000000100000"	11	
	"0000000011111"	56	
	"0000000011110"	1	
	"0000000011100"	2	
	"0000000011000"	2	
	"0000000001111"	53	
	"0000000001110"	2	

구분	패턴	가구 수	합계
	"0000000001100"	6	
	"0000000001000"	5	
	"0000000000111"	68	
	"0000000000110"	3	
	"0000000000100"	2	
	"0000000000011"	81	
	"0000000000010"	6	
	"0000000000001"	69	
기타 가구	"1011111111111"	13	144
	"1101111111111"	13	
	"1110111111111"	2	
	"1111101111111"	27	
	"1111001111111"	1	
	"1111101111110"	2	
	"1110001111111"	1	
	"1111101111100"	4	
	"1111111000111"	7	
	"1111111100011"	10	
	"1111111110001"	10	
	"1011111100011"	1	
	"1011111111000"	2	
	"1100001111111"	5	
	"1101111111000"	1	
	"1111101000111"	2	
	"1111101111000"	3	
	"1111111000110"	1	
	"1000001111111"	4	
	"1111001111000"	1	
	"1111111000100"	2	
	"1100001111100"	1	
	"1111101100000"	4	
	"1011111000000"	1	
	"1101111000000"	1	
	"1110001110000"	1	
	"1111001100000"	2	
	"1111101000000"	4	
	"1000001110000"	1	
	"1011100000000"	2	
	"1100001100000"	1	
	"1101100000000"	2	
	"1011000000000"	5	
	"1101000000000"	2	
	"1000001000000"	2	
	"1010000000000"	3	

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

부록 2. 차수별 일반/저소득 가구 총가구원 수 분석 결과

〈부표 2-1-1〉 차수별 일반 가구 총가구원 수

구분	총가구원 수	(단위: 가구)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기존 패널	1인	427	407	365	372	426	399	415	437	403	428	428	457	489
	2인	857	928	876	893	889	827	872	888	857	878	890	865	856
	3인	903	912	847	849	815	786	770	800	796	738	726	733	711
	4인	1,242	1,197	1,125	1,087	1,055	976	965	953	909	903	859	799	763
	5인 이상	365	414	379	402	390	359	358	353	344	344	310	286	273
신규 패널	1인							102	105	100	116	105	133	142
	2인							298	285	260	260	259	244	220
	3인							192	185	155	140	151	159	143
	4인							204	181	170	160	157	152	146
	5인 이상							58	54	46	48	41	37	37

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

〈부표 2-1-2〉 차수별 일반 가구 총가구원 수 비율

구분	총가구원 수	(단위: %)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기존 패널	1인	11.3	10.6	10.2	10.3	11.9	11.9	12.3	12.7	12.2	13.0	13.3	14.6	15.8
	2인	22.6	24.1	24.4	24.8	24.9	24.7	25.8	25.9	25.9	26.7	27.7	27.6	27.7
	3인	23.8	23.6	23.6	23.6	22.8	23.5	22.8	23.3	24.1	22.4	22.6	23.3	23.0
	4인	32.7	31.0	31.3	30.2	29.5	29.2	28.6	27.8	27.5	27.4	26.7	25.5	24.7
	5인 이상	9.6	10.7	10.6	11.2	10.9	10.7	10.6	10.3	10.4	10.5	9.7	9.1	8.8
신규 패널	1인							11.9	13.0	13.7	16.0	14.7	18.3	20.6
	2인							34.9	35.2	35.6	35.9	36.3	33.7	32.0
	3인							22.5	22.8	21.2	19.3	21.2	21.9	20.8
	4인							23.9	22.4	23.3	22.1	22.0	21.0	21.2
	5인 이상							6.8	6.7	6.3	6.6	5.8	5.1	5.4

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

〈부표 2-1-3〉 차수별 저소득 가구 총가구원 수

구분	총가구원 수	(단위: 가구)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기존 패널	1인	1,094	1,045	1,072	1,077	1,048	1,058	1,059	1,031	1,051	1,043	1,047	1,007	1,016
	2인	1,238	1,008	1,012	962	909	871	856	794	777	731	682	703	683
	3인	454	343	327	282	277	243	239	200	174	173	156	145	138
	4인	313	218	191	188	147	138	120	108	88	71	61	55	47
	5인 이상	179	108	120	95	78	78	78	55	39	34	30	31	21
신규 패널	1인							434	427	436	430	450	436	447
	2인							396	352	343	324	292	284	280
	3인							74	59	62	63	51	34	35
	4인							34	33	25	23	22	17	22
	5인 이상							8	12	13	7	6	4	5

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일]. 내부자료

〈부표 2-1~4〉 차수별 저소득 가구 총가구원 수 비율

구분	총가구원 수	(단위: %)												
		1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기존 패널	1인	33.4	38.4	39.4	41.4	42.6	44.3	45.0	47.1	49.4	50.8	53.0	51.9	53.3
	2인	37.8	37.0	37.2	36.9	37.0	36.5	36.4	36.3	36.5	35.6	34.5	36.2	35.9
	3인	13.9	12.6	12.0	10.8	11.3	10.2	10.2	9.1	8.2	8.4	7.9	7.5	7.2
	4인	9.6	8.0	7.0	7.2	6.0	5.8	5.1	4.9	4.1	3.5	3.1	2.8	2.5
	5인 이상	5.5	4.0	4.4	3.7	3.2	3.3	3.3	2.5	1.8	1.7	1.5	1.6	1.1
신규 패널	1인							45.9	48.4	49.6	50.8	54.8	56.3	56.7
	2인							41.9	39.9	39.0	38.3	35.6	36.7	35.5
	3인							7.8	6.7	7.1	7.4	6.2	4.4	4.4
	4인							3.6	3.7	2.8	2.7	2.7	2.2	2.8
	5인 이상							0.9	1.4	1.5	0.8	0.7	0.5	0.6

자료: 한국보건사회연구원, 한국복지패널 1~13차 원자료[데이터파일], 내부자료

부록 3. 분가 가구의 특성 분석

〈부표 3-1-1〉 분가 가구 가구 구분

구 분	일반		저소득		합계	
	가구 수	비율	가구 수	비율	가구 수	비율
기존 패널	630	86.18	101	13.82	731	100.00
추가 패널	62	72.94	23	27.06	85	100.00
전체 패널	692	84.80	124	15.20	816	100.00

자료: 한국보건사회연구원, 한국복지패널 13차 원자료[데이터파일]. 내부자료

〈부표 3-1-2〉 분가 가구 가구원 수

구 분	가구원 수	일반		저소득		합계	
		가구 수	비율	가구 수	비율	가구 수	비율
기존 패널	1인	174	27.62	61	60.40	235	32.15
	2인	123	19.52	20	19.80	143	19.56
	3인	151	23.97	14	13.86	165	22.57
	4인	152	24.13	3	2.97	155	21.20
	5인	26	4.13	2	1.98	28	3.83
	6인 이상	4	0.63	1	0.99	5	0.68
	합계	630	100.00	101	100.00	731	99.99
추가 패널	1인	26	41.94	17	73.91	43	50.59
	2인	12	19.35	4	17.39	16	18.82
	3인	13	20.97	1	4.35	14	16.47
	4인	10	16.13	1	4.35	11	12.94
	5인	-	-	-	-	-	-
	6인 이상	1	1.61	0	0.00	1	1.18
	합계	62	100.00	23	100.00	85	100.00
전체 패널	1인	200	28.90	78	62.90	278	34.07
	2인	135	19.51	24	19.35	159	19.49
	3인	164	23.70	15	12.10	179	21.94
	4인	162	23.41	4	3.23	166	20.34
	5인	26	3.76	2	1.61	28	3.43
	6인 이상	5	0.72	1	0.81	6	0.74
	합계	692	100.00	124	100.00	816	100.00

자료: 한국보건사회연구원, 한국복지패널 13차 원자료[데이터파일]. 내부자료

〈부표 3-1-3〉 분가 가구 가구 형태

구 분	가구 형태	일반		저소득		합계	
		가구 수	비율	가구 수	비율	가구 수	비율
기존 패널	단독	174	27.62	61	60.40	235	32.15
	모자	13	2.06	5	4.95	18	2.46
	부자	1	0.16	1	0.99	2	0.27
	조손가구 (소년소녀가장)	3	0.48	1	0.99	4	0.55
	기타	439	69.68	33	32.67	472	64.57
	합계	630	100.00	101	100.00	731	100.00
추가 패널	단독	26	41.94	17	73.91	43	50.59
	모자	1	1.61	2	8.70	3	3.53
	부자	-	-	-	-	-	-
	조손가구 (소년소녀가장)	-	-	-	-	-	-
	기타	35	56.45	4	17.39	39	45.88
	합계	62	100.00	23	100.00	85	100.00
전체 패널	단독	200	28.90	78	62.90	278	34.07
	모자	14	2.02	7	5.65	21	2.57
	부자	1	0.14	1	0.81	2	0.25
	조손가구 (소년소녀가장)	3	0.43	1	0.81	4	0.49
	기타	474	68.50	37	29.84	511	62.62
	합계	692	100.00	124	100.00	816	100.00

자료: 한국보건사회연구원, 한국복지패널 13차 원자료[데이터파일], 내부자료

간행물 회원제 안내

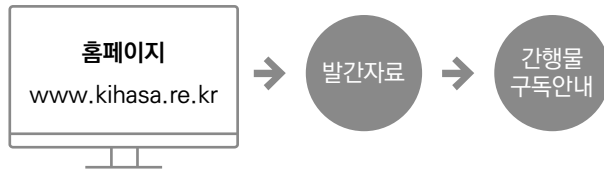
회원제에 대한 특전

- 본 연구원이 발행하는 판매용 보고서는 물론 「보건복지포럼」, 「보건사회연구」도 무료로 받아보실 수 있으며 일반 서점에서 구입할 수 없는 비매용 간행물은 실비로 제공합니다.
- 가입기간 중 회비가 인상되는 경우라도 추가 부담이 없습니다.

회원 종류

전체 간행물 회원	보건 분야 간행물 회원
120,000원	75,000원
사회 분야 간행물 회원	정기 간행물 회원
75,000원	35,000원

가입방법



문의처

- (30147) 세종특별자치시 시청대로 370 세종국책연구단지 사회정책동 1~5F
간행물 담당자 (Tel: 044-287-8157)

KIHASA 도서 판매처

- 한국경제서적(총판) 02-737-7498
- 교보문고(광화문점) 1544-1900
- 영풍문고(종로점) 02-399-5600
- 알라딘 <http://www.aladdin.co.kr>
- Yes24 <http://www.yes24.com>