

장애인삶 패널조사 가중치 산정 방안 연구

김현지 · 김태용

Korea Disabled people's Development Institute



본 보고서의 연구내용은 연구진의 의견으로 (재)한국장애인개발원의 공식적인
의견과는 다를 수 있습니다.

연구진

연구책임 : 김현지(한국장애인개발원 정책연구실 조사패널팀 부연구위원)

공동연구 : 김태용(한국장애인개발원 정책연구실 조사패널팀 연구원)

자문위원

변 종 석 (한신대학교 응용통계학과 교수)

박 민 규 (고려대학교 통계학과 교수)

임 찬 수 (통계청 통계개발원 사무관)

발간사

한국장애인개발원은 장애인 관련 조사·연구, 정책개발 등을 위해 다양한 연구를 수행하고 있습니다. 이러한 연구 활동의 일환으로 장애인삶 패널조사를 실시하고 있으며, 이는 2018년 통계청의 국가승인통계 지위를 확보하였습니다. 2018년 1차년도 본조사를 실시하였습니다.

장애인삶 패널조사는 장애등록 이후 삶의 변화(장애수용, 사회복귀, 가족구조 등)를 소득수준, 건강, 일상생활, 사회참여 등에서 파악하고, 향후 관련 정책 수립·지원에 필요한 기초자료 마련에 목적을 두고 있습니다. 이에 본 조사는 국가승인통계로서 통계적 방법에 따른 과학적 조사가 실시될 수 있도록 정밀한 표본 설계방법에 따라, 2015년부터 2017년 사이에 장애등록을 마친자 중 본 조사에 참여할 대상자를 선정하였습니다. 이때 특정 장애유형이나 특정 지역에 치중하지 않고 조사될 수 있도록 하였습니다.

이 같은 배경 하에서 장애인삶 패널조사의 1차년도 조사 결과 분석을 위한 가중치 연구를 수행하게 되었습니다. 본 연구를 통해 조사 자료를 활용하여 생산하는 통계의 신뢰성을 제고하고, 국가통계가 필요로 하는 요건을 갖추고자 하였습니다. 아무쪼록 본 조사가 장애등록 이후 장애수용 과정과 사회관계 속에서 겪는 다양한 변화를 개인·가족·사회적 관점에서 이해하는데 귀중한 자료로 활용되었으면 하는 바램입니다.

마지막으로 본 연구를 위해 노력해주신 한국장애인개발원 정책연구실 연구진과 연구 진행 과정에서 아낌없는 조언을 주신 학계 및 실무계 인사들께도 깊은 감사를 드립니다.

2019년 10월
한국장애인개발원 원장 **최 경 속**

➤ 목 차

연구요약	ix
I. 서 론	1
1. 개요	3
2. 장애인삶 패널조사	4
1) 법적근거 및 통계작성승인번호	4
2) 조사목적	4
3) 목표모집단	4
4) 조사대상 및 조사방법	5
5) 조사내용	5
6) 통계 자료 산출 범위	6
3. 연구필요성 및 목적	7
1) 연구배경 및 필요성	7
2) 연구의 목적	8
4. 연구내용 및 방법	8
1) 연구내용	8
2) 연구 수행 방법	9
II. 가중치 작성 이론 고찰	13
1. 개요	15
2. 기본 이론	16
1) 사후층화	16
2) Calibration	18



3) Raking ratio	20
4) 무응답 조정	21
Ⅲ. 타 조사 가중치 설계 및 참고자료 분석	23
1. 개요	25
2. 타조사 가중치 산출 방법 분석	29
1) 장애인경제활동실태조사	29
2) 특수교육실태조사	32
3) 장애인고용패널조사	35
4) 산재보험패널조사	42
5) 한국·아동청소년패널조사	46
6) 여성가족패널조사	50
3. 관련 자료 분석	55
4. 시사점 및 주요고려 사항	59
1) 가중치 작성 방안 검토를 위한 주요 시사점	59
2) 관련 자료에서의 주요 시사점	61
Ⅳ. 가중치 부여 방안	63
1. 개요	65
2. 장애인삶 패널조사 구축 현황	66
1) 개요	66
2) 표본추출	67
3) 패널구축 및 1차년도 조사 현황	69
4) 고려사항	71
3. 가중치 부여 방향	72
1) 고려사항	72
2) 공표범위 검토	73
3) 기본방향	73

4. 가중치 부여를 위한 검토 방안	74
1) 개요	74
2) 가중치 부여 방안	74
3) 검토방안별 비교 검토	75
5. 주요항목 분석	80
1) 개요	80
2) 주요항목 비교·분석	81
 V. 결론 및 유의사항	 83
 참고문헌	 86



표목차

〈표 I-1〉 조사영역 및 세부영역	5
〈표 I-2〉 조사표 구성	6
〈표 I-3〉 연구 수행 과정	10
〈표 I-4〉 연구 추진 절차	11
〈표 II-1〉 거리함수 종류와 캘리브레이션 식	20
〈표 III-1〉 조사별 최종가중치 요약	26
〈표 III-2〉 층별 표본크기 (2018년 장애인경제활동실태조사)	30
〈표 III-3〉 장애인고용패널 표본배분 및 조사완료 표본 수	37
〈표 III-4〉 장애인고용패널조사 (2008) : 표본배분 및 조사 완료 표본 수	40
〈표 III-5〉 산재보험패널조사 층별 표본할당	43
〈표 III-6〉 산재보험패널조사 : 권역별 층 통합	44
〈표 III-7〉 산재보험패널조사 : 표본할당 및 조사 완료 표본 수	45
〈표 III-8〉 한국·아동청소년패널조사 : 층별 표본 크기	47
〈표 III-9〉 여성가족패널 1차 조사 목표 및 조사완료 표본 수	52
〈표 III-10〉 여성가족패널 : 표본추출 결과 및 1차년도 조사완료 표본 수	52
〈표 III-11〉 여성가족패널 가중치 분포	53
〈표 III-12〉 검토 내용 및 기준 (통계조정업무 매뉴얼)	56
〈표 IV-1〉 표본추출결과	68
〈표 IV-2〉 지역별 조사완료 현황	69
〈표 IV-3〉 장애유형, 장애정도, 성별 조사완료 현황	70
〈표 IV-4〉 최종가중치 산출을 위한 검토 방안	76
〈표 IV-5〉 검토방안별 $\sqrt{1+cv^2}$	77
〈표 IV-6〉 초기가중치 대비 최종가중치 비율 분포	78
〈표 IV-7〉 최종가중치의 기술통계량	79
〈표 IV-8〉 주요변수별 모집단 분포	79
〈표 IV-9〉 변수유형별 주요 문항	81
〈표 IV-10〉 주요문항별 상대표준오차(%) - 전체	82

▶ 그림목차

[그림 Ⅲ-1] 여성가족패널 가중치	51
[그림 Ⅲ-2] 표본조사 정확도 기준	57
[그림 Ⅲ-3] 표본설계 및 관리지침 제9조 및 제10조	58
[그림 IV-1] 검토방안별 가중치 분포	78

연구 요약



연구요약

I. 서론

1. 개요

- 본 연구는 장애인삶 패널조사의 1차년도 조사 결과 분석에 필요한 가중치 작성을 위한 연구를 수행함. 이를 통해 본 조사의 조사자료로 생산하는 통계 신뢰성을 제고하고, 국가통계가 필요로 하는 요건을 갖추 수 있도록 함
- 이에 문헌과 선행자료 분석을 토대로 이론적 배경을 검토, 타 조사의 가중치 산출 과정을 분석하여 시사점을 도출하고, 장애인삶 패널조사의 표본설계를 기초로 가중치 산정 방향을 설정하여, 최종가중치 제시함

2. 연구필요성 및 목적

1) 연구필요성

- 장애인삶 패널조사의 모집단 분포를 살펴보면, 모집단은 전국에 분산되어 거주하고 있으며, 모집단 규모는 장애유형별, 장애정도별, 성별, 연령별 큰 차이가 있음
 - 이에 적정 수의 표본 확보, 모집단 대표성, 추정의 정도를 높일 수 있는 표본추출을 위해 표본설계에 기반을 둔 다양한 설계요소를 반영하여 특정 연령대, 특정 장애유형의 누락이 없으면서, 통계적으로는 분석 가능할 수 있는 크기의 표본을 확보하고자 하였음
- 본 조사는 조사 예산 및 조사 일정을 고려하여 조사 효율성을 제고할 수 있도록 이중추출(double sampling)을 적용하고 있음
- 따라서 가중치를 무시하고 통계분석을 수행할 경우 추정치의 편향이 발생할 수 있음. 이에 모수에 대한 비편향 추정량을 얻을 수 있도록 추정단계에서 적절한 가중치 조정(adjustment)이 필요함



2) 연구목적

- 본 연구는 장애인삶 패널조사의 1차년도 조사 결과 분석에 필요한 가중치 작성을 주목적으로 하며, 본 연구를 통해 다음의 목적을 달성하고자함
 - 첫째, 장애인삶 패널조사의 가중치 조정에 대한 이론적 배경 제공
 - 둘째, 1차년도 조사 결과 분석과 본 조사의 조사자료를 활용한 결과분석시 추정치의 정도를 높일 수 있는 적절한 가중치 작성 방안 마련
 - 셋째, 조사통계가 국가통계로서 갖춰야할 요건을 갖추고, 향후 관련 자료로의 활용 기대

3. 연구의 내용 및 방법

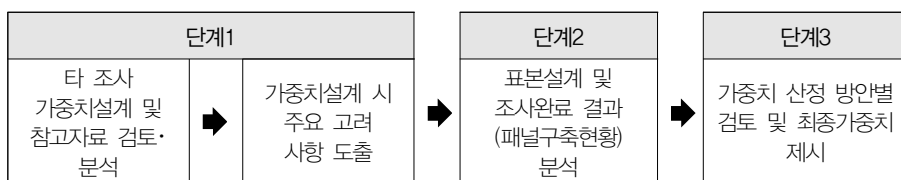
1) 연구의 내용

- 가중치 이론 및 타조사 가중치 설계 및 참고자료 분석
 - 선행 참고자료와 문헌을 토대로 가중치 설계를 위한 이론적 측면을 검토하고, 장애인삶 패널조사 가중치 산출을 위해 벤치마킹이 필요한 사항과 설계상 시사점을 도출함
- 표본설계 및 패널 구축 결과 분석
 - 표본설계와 패널 구축 결과를 분석하여 가중치 설계의 기본 방향을 설정함. 이때 통계의 공표단위를 고려하여 가중치 설계시 반영 가능성을 검토함
- 1차년도 가중치 산정 방안 검토 및 최종가중치 제시
 - 문헌 및 선행자료 분석과 표본설계 및 패널 구축 결과를 토대로 1차년도 조사자료의 결과분석을 위한 1차년도 가중치 산정 방안을 마련함. 연구 전반에 포함되는 내용에 관해서는 통계승인업무 처리지침과 통계청의 품질진단 매뉴얼(조사통계)을 참고, 제시되어 있는 기준을 전반적인 연구 진행의 기본 틀로 활용함

2) 연구방법

- 문헌과 선행자료 분석을 토대로 이론적 배경을 검토하고, 타 조사의 가중치 산출 과정을 분석하여 시사점 도출
- 장애인삶 패널조사의 표본설계를 고려한 가중치 산정 방향 설정
- 가중치 산정 방안 검토 과정에서 표본이론 및 실무적으로 발생한 이슈들은 학계 및 통계전문가와의 지속적 논의를 통해 검토·결정

[연구 수행 과정]



II. 가중치 작성 이론 고찰

- 장애인삶 패널조사의 1차년도 조사결과 분석에 필요한 적절한 가중치 작성 방안을 마련할 수 있도록 관련된 기본 이론 검토
 - 사후층화, 캘리브레이션, 레이킹비, 무응답조정

III. 타 조사 가중치 설계 및 참고자료 분석

1. 타 조사 가중치 설계 분석

- 장애인삶 패널조사 가중치 작성시 고려할 필요가 있는 사항을 사전에 점검하고, 이를 토대로 가중치 작성시 참고할 주요 시사점 도출



- 가중치 적용의 기본 아이디어는 비편향 추정이지만, 표본조사 수행시 발생하는 무응답, 표본추출틀의 결함 등을 극복하기 위해 추정식을 보정하는 수단으로도 가중치 조정이 실시되며, 이러한 가중치 조정을 통해 비편향 추정량을 얻을 수 있음 (김영원 외, 2012).
- 표본조사의 경우 다양한 사유로 발생한 무응답으로 사전 계획된 표본 크기와 조사 완료 후 표본크기간의 불일치 발생 가능성이 있음. 이렇게 발생하는 규모 차이는 추출확률에 응답률을 보정한 무응답 조정 가중치로 보정함
- 사후조정은 표본설계 단계에서 다양한 이유(주요 문항에 영향을 미칠 것으로 예상되나 표본크기의 한계, 모집단 목록에의 정보 부족 등)로 고려할 수 없었던 변수에 대해서 조사가 완료된 이 후, 부분적으로 확보된 모집단 정보를 활용하여 사후층화, 벤치마킹조정, 캘리브레이션 등의 방법을 적용하여 실시함. 이때 활용 변수의 범위는 통계의 실제 공표와 통계를 공표하진 않더라도 향후 자료 사용 측면을 함께 고려할 필요가 있음
- 층화변수 이외의 주요 결과에 영향을 줄 수 있는 변수들은 사후층화 등을 으로 가중치 보정 단계에서 고려할 수 있음
- 사후층은 표본설계 단계에서 고려되지 않은바, 사후층의 표본은 조사되지 않았을 가능성이 높음. 이러한 경우 층 통합을 고려할 수 있음. 가중치를 적용한 분석 수행시 층 통합된 분석단위 주의 필요
- 표본설계에서 고려하지 않은 변수를 가중치 작성 단계에서 고려하거나, 층 통합 후 캘리브레이션 가중치 부여시 최종가중치의 합계는 모집단 합계와 일치 하지 않을 수 있음

2. 참고자료 분석

- 관련 규칙 및 매뉴얼을 토대로 가중치 산정시의 주요 고려 사항 파악
 - 통계조정업무 매뉴얼 및 자체통계품질진단(통계법 제9조)에서 가중치와 관련된 검토 사항을 제시하고 있음. 표본추출과정을 고려한 가중치, 무응답 및 사후층화 가중치의 고려, 복합표본설계의 경우에는 선형화나 반복분산 추정방법에 대한 고려, 표본오차, 분산, 상대표준오차 등을 검토요건으로 삼고 있음

- 장애인삶 패널조사 역시 최종가중치 산출 단계에서 관련 규칙 및 매뉴얼에서 주요 검토 사항으로 삼고 있는 항목을 검토하여 연구를 진행함

IV. 가중치 부여 방안

1. 개요

- 일반적으로 표본조사에서는 표본의 대표성을 확보할 수 있도록 불균등 추출 확률, 응답률을 보정하여 가중치를 부여한 후 추정치를 산출함 (Kish, 1992; 류제복, 1993; 김규성, 2006)
 - 표본가중치를 무시하고 통계분석을 수행할 경우 추정치는 심각한 편향이 발생할 수 있어, 대규모 조사에서는 모수추정과정에서 반드시 표본가중치를 적용한 분석방법을 활용하여 추정량의 비편향성을 확보함
- 장애인삶 패널조사의 패널구축 및 1차년도 조사완료에 따른 결과의 분석에 필요한 가중치를 산출하고자, 다양한 방안을 검토하여 최종가중치 작성을 위한 적절한 방안을 마련하고자 함



2. 장애인삶 패널조사 구축 현황

- 패널 구축 현황을 토대로 가중치 산출과정에서 고려 필요한 사항 파악
 - 장애인삶 패널조사는 사전 계획된 목표 표본 크기보다 약 300여명 많은 표본이 조사완료됨
 - 패널 구축 이후 차년도 부터 발생할 수 있는 패널 마모로 인한 표본 규모의 감소를 고려하여, 조사 완료된 모든 장애인이 최종 표본에 포함될 수 있도록 조사완료된 모든 표본에 가중치를 부여하는 방안을 마련할 필요가 있음

3. 가중치 부여 방향

1) 고려 사항

- 장애인삶 패널조사의 패널 구축 이후 차년도부터 발생할 수 있는 패널마모로 인해 표본 규모의 감소가 예상됨을 감안하여 조사 완료된 모든 장애인이 최종 표본에 포함될 수 있도록 하는 방안을 마련하기로 함
- 표본설계 과정에서 표본크기의 한계 및 공표단위 등을 고려하여 장애유형은 7개 범주로 설계됨
 - 가중치 작성시 향후 자료의 활용성 측면을 고려하여 추가로 활용 가능한 변수에 대한 검토를 수행하기로 결정함
- 표본설계 과정에서 고려하지 않은 사후층의 표본은 조사되지 않았을 가능성이 높으니, 이러 경우 해결하는 방안으로 층 통합을 검토하기로 함. 연령대 검토시의 층은 만18세 이하(학령기), 만19-49세(청·중년), 만50세(장·노년)로 사후층을 구성하기로 결정함

2) 공표범위 검토

- 주요항목별 상대표준오차 검토 후 공표항목 조정
 - 상대표준오차는 양적변수에 대해서만 검토하는 것이 적절함. 그러나 질적 변수에 대한 검토가 병행 되는 경우에는 항목간 통합, 리커트 척도의 경우 평균 점수, 비율 등을 이용하고 향후 자료제공 시 명시 필요
- 공표단위별 상대표준오차 검토 및 공표단위 조정

3) 기본방향¹⁾

- 기본가중치는 설계가중치에 응답률을 조정하여 산출하고, 가중값이 부여되는 인원은 조사가 완료된 인원²⁾에 대하여 산출
 - 목표 장애인 수보다 최종 표본이 많은 것으로 인한 표본의 영향력을 줄이고, 집단별 분포의 반영으로 편향을 줄일 수 있는 방안으로 검토
- 현실적으로 공표 및 데이터 사용의 용이성을 담보할 수 있는 방안 마련
 - 통계를 공표하지 않더라도 통계 작성을 위한 집단별 분포를 추정에 반영할 수 있으므로 편향을 줄이는 방안을 검토함. 실제 공표와 향후 자료 사용을 고려하여, 최종적인 캘리브레이션 범위를 결정함
- 도출된 가중치 작성방안별로 주요문항에 대한 상대표준오차(또는 표준오차)를 함께 검토하여 최종가중치 작성방안 마련
 - 통계의 공표단위, 가중치 작성시 반영하고자하는 변수, 표본크기, 사후층수 등을 종합적으로 고려하여 가중값을 산출함

1) 표본설계 내역과 패널구축 현황을 토대로 1차년도 조사결과 분석을 위한 최종가중치 산출을 위한 자문회의 실시. 공표범위, 모집단 및 표본원료 현황에 따른 가중치 산출시 활용 변수와 조사원료 인원내 대한 가중치 부여 및 이론적 타당성을 검토(자문회의)하여 최종가중치 산출 방안을 마련하여 적용 결과를 검토한 후, 최종가중치를 산출(자문회의)함

2) 목표 표본보다 조사 완료 표본크기가 큰 상황임



4. 가중치 부여를 위한 검토 방안

1) 가중치 부여 방안

- 최종가중치를 부여할 표본은 총 6,121명임
 - 패널 구축 이후 다음연도 부터 발생할 수 있는 패널마모로 인한 표본 규모의 감소를 고려하여, 조사완료된 모든 장애인은 최종 표본에 포함하여 가중치 작성 방안을 검토함
- 가중치 산출을 위하여 활용할 변수들로는 표본추출틀내 정보가 존재하는 변수인 장애유형, 장애정도, 성별, 연령, 지역변수임
 - 기본가중치는 무응답을 고려, 설계가중치에 무응답 조정계수를 적용하여 산출함. 기본가중치 계산한 후, 사후조정을 적용하여 최종가중치를 작성함
 - 사후조정 방안으로는 사후층화와 레이킹을 적용하는 방안을 검토하여 적절한 가중치를 산출함. 이때 표본추출틀내 정보가 존재하는 변수를 조정변수로 활용함
- 최종가중치 산정을 위하여 사후층화와 레이킹 적용 검토
 - 캘리브레이션 범위는 실제 공표와 향후 자료 활용도 측면을 고려하여 결정함. 통계 공표를 하지 않더라도 통계 작성을 위한 집단별 분포를 추정에 반영, 편향을 줄일 수 있도록 연령대를 고려하고, 층을 재구성하는 방안으로 검토³⁾
- 최종가중치 작성방안은 가중값으로 인한 효과와 검토방안별 가중치를 적용하여 변수유형별 주요문항에 대한 상대표준오차를 비교·검토하여 자문회의를 거쳐 최종 결정

3) 이에 관해서는 자문회의를 통해 통계적 타당성을 검토하였으며, 이는 공표 단위가 아니더라도 통계 작성을 위한 집단별 분포를 추정에 반영함으로써 편향을 줄일 수 있도록 하는 방안이 된다. 연구진 논의 결과를 토대로 사후층의 연령대 범주를 결정함

2) 검토방안별 비교 검토⁴⁾

- 장애유형의 경우 향후 자료 사용 측면을 고려하여 15개 범주로 가중값을 사후보정함. 목표 장애인 수 보다 최종 표본이 많은 것으로 인한 표본의 영향력을 줄이고, 집단별 분포를 반영함으로써 편향을 줄일 수 있도록 최종가중치 작성시 연령대, 그리고 지역변수까지 고려함
- 연령대의 경우 통계 공표를 하지 않더라도 통계 작성을 위한 집단별 분포를 추정에 반영하고, 편향을 줄이는 효과를 얻을 수 있도록 사후층으로 구성하여 적용함
- 사후보정은 사후층화, 레이킹비를 적용한 결과를 비교·검토하여 적절한 최종가중치 산출 방안을 마련함
- 가중치의 선택 문제는 통계적 추정과 더불어 현실적인 공표 및 데이터 사용의 용이성을 담보하는 차원에서 이루어져함. 이에 다음 사항을 검토함
 - 가중값으로 인한 효과를 살펴보면, 검토 방안별로 차이를 보이지 않음
 - 초기가중치 대비 최종가중치의 변화를 검토방안별로 살펴보면, 초기가중치 대비 최종가중치 분포간에 유의한 차이를 보이지 않음
 - 단순히 최종가중치의 크기만으로 판단하기 보다는 초기가중치 대비 최종가중치 변동량을 검토하는 것이 최종가중치를 통해 작성되는 추정량의 성질에 미칠 영향력을 판단하는데 있어서 좋은 지표가 됨⁵⁾

5. 주요항목 분석

- 상대표준오차 분석은 변수의 유형에 따라 실시함
 - 장애인삶 패널조사의 조사문항을 토대로 산출되는 통계표는 월평균 가구소득 등과 같은 연속형 변수, 장애수용과 같이 세부분항으로 구성되어 문항별 응답값의 평균을 활용하는 척도 변수, 비율추정을 하는 범주형 변수임

4) 본 절은 자문회의를 통해 마련한 방안을 토대로 작성함

5) Deville, Sarndal, 그리고 Sautory(1993)은 프랑스 가구 조사 분석을 위한 가중치 산출 연구에서 이 비율의 범위를 정의하고 여러 다양한 캘리브레이션 방안을 연구함. (출처) 변중석 외 (2018). 국민건강영양조사 제8기(2019-2021) 표본설계 및 가중치 계산. 보건복지부 질병관리본부 연구용역.



- 검토방안에 따른 상대표준오차를 주요문항별로 살펴보면, 모집단 전체 및 관심 도메인 추정량의 통계적 정확성에 대해서 유의한 차이를 나타내지 않았음
- 전체적으로 전국 단위 상대표준오차는 5% 미만으로 나타나고 있어 전반적으로 만족할 만한 수준인 것으로 보임⁶⁾
- 장애인삶 패널조사는 기본가중치를 레이킹 조정하여 최종가중치 산출함
 - 가중치를 적용한 패널조사 전체 분석(1차년도) 및 상대표준오차 분석 결과를 토대로, 마이크로데이터 제공시의 자료제공의 범위를 추가적으로 검토할 필요가 있음

[주요문항별 상대표준오차(%) - 전체]

구분	A1	B1	B2	B3
노후생활에 필요한 1달 평균 생활비	1,171	1,120	1,128	1,109
월평균 가구소득	2,519	2,537	2,547	2,530
장애수용(평균)*	0,368	0,353	0,357	0,351
일상생활존중받는 정도(존중받고 있는 비율)**	1,789	1,699	1,720	1,693

* 주 1) 12개 세부 항목의 전체 평균을 사용하여 산출(일부 문항 역코딩)

2) 4점 리커트 척도 : 전혀그렇지 않다, 거의 그렇지 않은 편이다, 그런 편이다, 매우 그렇다

** 주 1) 일상생활존중받는 정도 : 전혀 존중받지 못한다, 거의 존중받지 못하는 편이다, 존중받고 있는 편이다, 매우 존중받고 있다.

2) 존중받고 있는 비율은 '존중받고 있는 편이다'와 '매우 존중받고 있다'을 의미한다.

6) 상대표준오차가 25%가 넘을 경우는 해석에 유의해야함. (출처) 표본설계 및 관리지침. 통계청 예규 제198호, 2017. 3.10.

I. 서론

I 서론

1. 개요

표본조사에서 가중치(sampling weights)란 하나의 표본이 대표하는 모집단 단위의 수를 의미 한다 (2018년 자체통계품질진단 매뉴얼).⁷⁾ 대부분의 표본조사는 복합표본조사⁸⁾를 사용하고 있는데, 이러한 조사 자료를 활용하여 추정할 때에 편향 없는 결과를 얻기 위해서는 가중치 사용을 필요로 한다. 이에 자료 분석시 일반적인 통계교과서에 나오는 일반적인 추정 및 검정 방법(estimation and testing procedures)을 적용하여 통계분석을 수행하는 것은 적절하지 않다.⁹⁾

복합표본조사 자료로 올바른 추론을 위한 첫 번째 고려사항은 표본추출에서 불균등 선택확률을 사용한다는 것이다. 만약 표본설계에서 모집단의 단위들을 선택할 확률들이 서로 다르다면, 비편향추정량(unbiased estimator)이 되기 위해 각 표본 단위별로 반드시 적절한 가중치를 적용해 분석해야 한다. 만약 표본으로 추출된 모든 단위들이 응답을 하고 이 관측값을 추정에 모두 사용한다면, 선택확률의 역수를 가중치로 사용함으로써 어떤 경우든지 비편향추정량을 얻을 수 있다. 추론에 영향을 주는 표본추출의 두 번째 측면은 표본설계와 연관성 정도(precision)의 수준인 표본추출오차(sampling error)이다.¹⁰⁾

따라서 가중치 작업은 표본추출을 통해 얻어지는 표본에 적절한 가중치를 부여해 줌으로써 추정량의 효율성과 신뢰성을 제고하는 방법으로 표본조사에서의 필수적인 요소이다 (김재광, 2017).¹¹⁾

7) 2018년 자체통계품질진단 매뉴얼(조사통계). 통계청.

8) 복합표본설계는 다음과 같은 특성들을 갖게 된다. (1) 층화(stratification), (2) 집락(clustering), (3) 불균등 선택 확률(unequal probabilities of selection), (4) 다단계 추출(multiple stages of selection), 그리고 (5) 계통 추출(systematic selection). 이 같은 특성으로 자료 분석시 일반적인 통계교과서에서 나오는 추정 및 검정 방법을 적용하는 것은 적절하지 않다.

9) Richard L. Scheaffer et. al. (2006). Elementary Survey Sampling. DUXBURY. 6th Edition. 김영원 외(2012). 표본조사의 이해와 활용. CENGAGE Learning.

10) Richard L. Scheaffer et. al. (2006). Elementary Survey Sampling. DUXBURY. 6th Edition. 김영원 외(2012). 표본조사의 이해와 활용. CENGAGE Learning.

11) 김재광 (2017). 표본조사론. 제2판, 자유아카데미.



2. 장애인삶 패널조사¹²⁾

1) 법적근거 및 통계작성승인번호

장애인삶 패널조사는 장애인복지법 시행령 제19조 제2항 및 통계법 제18조 및 같은 법 시행령 제24조에 의해 실시되는 국가승인통계(승인번호 : 제438001호)¹³⁾¹⁴⁾이다.

2) 조사목적

장애인삶 패널조사는 장애 발생에 따른 장애수용 과정과 사회관계 속에서 겪는 변화에 대해서 개인·가족·사회적 요인의 동태를 파악 하여, 향후 관련 정책의 수립·지원에 필요한 기초자료를 확보하는데 그 목적이 있다.¹⁵⁾ 이에 장애 발생에 따른 변화를 탐색할 수 있는 과학적이고, 체계적 종단자료의 구축을 목표로 2018년 패널구축 및 1차년도 본조사를 실시하였다. 장애인삶 패널조사는 통계법 제18조 및 동법시행령 제24조에 의해 실시되는 국가승인통계이다.

3) 목표모집단

장애인삶 패널조사의 모집단은 장애인복지법 제2조에서 정의하는 장애인으로 같은 법 제32조에 의거 보건복지부에 장애등록을 2015년 1월1일부터 2017년 12월 31일 사이에 마친 등록장애인(단, 장애인거주시설 거주 장애인 제외)과 그들의 가구원이다.

12) 본 절은 법적근거 및 통계작성승인번호를 제외하고, 그 외 내용은「김현지, 김태웅 (2018). 장애인패널조사 표본 설계, 한국장애인개발원.」,「강정배 외(2018). 장애인패널 예비조사, 한국장애인개발원.」에서 일부 발췌하여 요약·정리 하였다.

13) 통계청 고시 제2018-373호

14) 장애인삶 패널조사는 2018년 패널구축과 1차 년도 조사가 함께 이루어졌다. 이에 패널조사 실시 후 조사결과 분석, 공표항목 등의 검토를 필요로 한다. 따라서 통계청으로부터 통계작성승인번호(제438001호, 작성승인일 : 2018.09.06., 조건부승인)을 받았다.

15) 장애인삶 패널조사는 장애등록 이후 삶의 변화(장애수용, 사회복귀, 가족구조 등)를 파악하고, 삶과 관련된 일상 생활, 소득수준, 건강상태, 복지욕구, 사회참여 등의 탐색을 통해 향후 관련 정책의 수립·지원에 필요한 자료를 마련하고자 한다. 이에 개인·가족·사회적 관점에서 장애 발생에 따른 장애수용과정과 사회관계의 변화 등을 탐색하고, 장애수용 및 변화, 건강·의료, 일상생활 등의 이해를 통해 관련 정책의 수립·지원 과정에서 활용 할 수 있는 기초자료를 확보하고자 한다.

4) 조사대상 및 조사방법

조사대상은 장애인복지법 제2조에서 정의하는 장애인으로 같은 법 제 32조에 의거 2015년 1월 1일부터 2017년 12월 31일 사이에 보건복지부에 장애등록을 한 등록장애인(단, 장애인거주시설 거주 장애인 제외)과 그들의 가구원으로 한다.

조사대상은 모집단 대표성과 추정치의 정도(precision)를 높일 수 있는 표본이 선정될 수 있도록 표본설계를 바탕으로 다양한 설계요소를 반영, 통계적 방법에 기반을 두었다.¹⁶⁾ 표본설계 연구결과에 따라 장애유형, 장애정도, 성별 등을 고려하여 표본배분 및 표본추출 하였다. 조사도구는 TAPI¹⁷⁾을 활용하며, 조사방법은 전문면접원에 의한 대면면접조사를 실시한다.

5) 조사내용

장애인삶 패널조사는 장애인과 그들의 가구원이 장애수용 과정과 사회관계 속에서 겪는 변화를 '장애수용 및 가계변화, 건강·의료, 자립, 사회참여'라는 관점에서 접근하고자 하였으며, 영역별 세부 포함 내용¹⁸⁾은 다음과 같다.

〈표 1-1〉 조사영역 및 세부영역

주요 조사 영역	세부 포함 내용
장애수용 및 변화	장애수용, 심리상태, 사회관계 경제상황, 주거, 일반현황 등
건강·의료	건강, 의료서비스, 안전
자립	일상생활, 여가활동, 자립생활, 교육, 고용(취업)
사회참여	장애관련서비스, 인권, 결혼생활 및 성, 노후, 삶의 만족도

주) 조사영역은 패널용(장애인용) 조사표 기준임

패널조사의 특성을 고려하여 추적조사를 위해 마지막에 패널 관리 정보(향후 이사의향 등)를 추가로 조사하였다.

16) 이에 관해서는「김현지, 김태용. (2018). 장애인패널조사 표본설계. 한국장애인개발원.」을 참고한다.

17) TAPI : Tablet Assisted Personal Interviewing

18) 조사대상의 범위에 대한 자문회의(2017.11.28) 결과를 바탕으로 보건복지부와 협의를 통해 주요 조사 영역을 결정하였다. 이에 관한 자세한 내용은 장애인패널 예비조사 (2018)을 참고한다.



본 패널조사의 조사표는 총 3종 (패널용, 가구주용, 가구원용¹⁹⁾)으로 구성되었으며, 각각의 조사내용은 다음과 같다.

〈표 1-2〉 조사표 구성

장애인용 (패널)	가구주용	가구원용
표본으로 선정된 장애인(패널)에 관한 문항	표본으로 선정된 장애인의 가계변화에 관한 문항	표본으로 선정된 ① 패널에 관한 가구원의 경험·생각 (인식)관련 문항 ② 가구원 본인과 관련한 문항
조사표 1	조사표 2	조사표 3

- 패 널 용 : 장애수용, 심리상태, 사회관계 및 건강, 의료서비스, 안전, 일상 및 여가생활, 자립생활, 교육, 고용(취업), 장애관련 서비스, 인권, 결혼생활 및 성, 노후, 삶의 만족도 등에 대하여 조사
- 가구주용 : 가구원 현황, 경제상황, 주거에 대하여 조사
- 가구원용 : 장애수용, 심리상태, 사회관계 및 건강, 의료서비스, 안전, 일상 및 여가생활, 자립생활, 교육, 고용(취업), 인권, 노후, 삶의 만족도 등에 대하여 조사

6) 통계 자료 산출 범위

장애인삶 패널조사를 통한 자료 산출은 전국 및 장애유형별, 장애정도별, 성별 주요 항목의 통계 생산을 목표로 한다.

19) 가구원 설문 경우 조사 참여 의향을 묻고, 참여 의향이 있는 경우만 응답 받도록 하였다. 가구원 조사표는 가구원 수에 따라 반복될 수 있다.

3. 연구필요성 및 목적

1) 연구배경 및 필요성

장애인삶 패널조사 모집단은 2015년 1월 1일부터 2017년 12월 31일 사이에 보건복지부에 장애등록을 마친 장애인(장애인거주시설에 거주하는 장애인 제외)과 그들의 가구원으로, 매년 추적 조사를 실시하는 전국 단위 조사이다.

모집단 분포를 살펴보면, 모집단은 전국에 분산되어 거주하고 있으며, 모집단 규모는 장애유형별, 장애정도별, 성별, 연령별 큰 차이가 있다. 장애정도별 분포 역시 장애유형별로 차이가 있으며, 연령별 모집단 규모도 큰 차이를 보였다. 특히, 심장, 호흡기, 간, 안면, 장루·요루, 뇌전증, 언어, 자폐성 장애를 지니고 있는 장애인은 모집단의 2% 미만인 것으로 나타났다.²⁰⁾

이에 장애인삶 패널조사는 적정 수 표본 확보와 모집단 대표성, 추정의 정도를 높일 수 있는 표본추출을 위해 표본설계에 기반을 두어 다양한 설계요소를 반영하여 통계적 방법에 따른 과학적 조사가 실시되도록 하였다. 이를 통해 특정 연령대, 특정 장애유형의 누락이 없으면서, 통계적으로는 분석 가능할 수 있는 표본크기의 표본을 확보할 수 있도록 하였다. 조사결과와 활용성 등을 고려하여 통계적으로 유의미한 수준의 분석이 가능할 수 있도록, 모집단 특성 변수별 적정 수의 표본이 확보될 수 있도록 하였다. 표본추출은 조사 예산 및 조사 일정을 고려하여 조사 효율성을 제고할 수 있도록 이중추출(double sampling)을 적용하였다.²¹⁾

이렇듯 장애인삶 패널조사는 조사목적상 일부 층은 과대 추출하여 불균등 추출 확률을 가지고 있으며, 표본추출에 있어서는 이중추출을 적용하고 있다. 따라서 가중치를 무시하고 통계분석을 수행할 경우에는 추정치의 편향이 발생할 수 있다. 이러한 이유로 모수에 대한 비편향 추정량을 얻을 수 있도록 추정단계에서 적절한 가중치 조정(adjustment)이 필요하다.

20) 본 문단은「김현지, 김태용. (2018). 장애인패널조사 표본설계. 한국장애인개발원.」에서 발췌하였다. 장애인삶 패널 조사는 조사목적과 조사대상에 있어서 특정 연령대나, 특정 장애유형에 초점을 둔 조사가 아니다. 위와 같은 모집단 특성이 고려되지 않고 일반적인 방식으로 표본이 추출될 경우, 본 조사의 목적을 충족하는 표본 수의 확보가 어려울 수 있다. 이에 모집단 비율이 낮은 장애유형(언어, 자폐성, 심장, 심장, 호흡기, 간, 안면, 장루·요루, 뇌전증)을 가진 장애인, 만18세 이하의 영유아기·학령기에 있는 장애인이 최종 표본에서 누락되지 않고 일정 수 이상 추출될 수 있는 방안을 마련하였다. (출처) 김현지, 김태용. (2018). 장애인패널조사 표본설계. 한국장애인개발원.

21) 본 문단을 포함한 본 절은 전체적으로 「김현지, 김태용. (2018). 장애인패널조사 표본설계. 한국장애인개발원.」의 내용을 일부 발췌하여 요약·재정리 하였다.



2) 연구의 목적

본 연구는 장애인삶 패널조사의 1차년도 조사 결과 분석에 필요한 가중치 작성을 주목적으로 하며, 이를 통해 다음의 목적을 달성하고자 한다.

첫째, 장애인삶 패널조사의 가중치 조정에 대한 이론적 배경을 제공한다. 둘째, 1차년도 조사 결과 분석과 조사자료를 활용한 추정치의 정도를 높여 추정량의 효율성과 신뢰성을 제고할 수 있는 적절한 가중치 작성 방안을 마련한다. 일반적으로 표본가중치는 표본설계시 사용한 자료의 시의성, 무응답, 조사여건으로 인한 현실적 제약, 표본추출틀과 모집단간의 불일치 등으로 인해 적절한 조정을 거쳐 최종가중치로 산출된다. 추가적으로 조사통계가 국가통계²²⁾로 갖춰야할 요건을 검토하여, 향후 관련 자료로 활용 할 수 있도록 한다.

4. 연구내용 및 방법

1) 연구내용

(1) 개요

장애인삶 패널조사 가중치 산정 방안 연구의 주요 내용은 첫째, 가중치 작성에 대한 통계 이론을 검토하고, 둘째 타 조사의 가중치 산출 과정을 검토·분석하여 장애인삶 패널조사의 1차년도 조사결과 분석을 위한 가중치 산정에 대한 시사점을 도출하고, 관련 규칙 및 매뉴얼 분석한다. 이를 토대로 1차년도 조사결과 분석을 위한 최종가중치 산정의 검토방안을 마련 한다.

(2) 연구의 주요 내용

본 연구는 서론을 제외하고, 총 네 개의 장으로 구성된다. 보고서는 가중치 산출

22) 국가통계에는 보고통계, 가공통계, 조사통계가 있다. 조사통계란 통계작성을 목적으로 통계작성기법을 사용하여 조사한 자료를 통해 작성한 통계를 말한다. 보고통계란 신고, 보고, 신청, 인·허가 등과 같이 다른 행정업무에 수반하여 수집된 자료를 이용하여 작성한 통계를 말한다. 가공통계란 한 종류 이상의 통계와 추가로 수집한 통계자료 또는 행정자료를 이용하여 작성한 통계를 말한다(통계승인업무 처리지침(통계청 예규 제213호) 제2조). 참고로 장애인삶 패널조사는 표본조사로 실시되는 조사통계이다. (출처) 김현지, 김태웅 (2018). 장애인삶 패널조사 표본설계. 한국장애인개발원.에서 재인용.

에 대한 통계 이론을 검토하는 것으로 시작한다. 이는 제2장에서 살펴보고자 한다.

제3장에서는 타 조사의 가중치 산출 과정을 검토·분석하여 가중치 산정에 대한 주요 쟁점 사항과 시사점을 도출한다. 이를 토대로 최종가중치 작성시에 발생 가능한 이론적 또는 실무적 애로사항을 사전에 검토한다. 이를 위해 전국 단위 대규모 조사를 중심으로 검토하고 가중치 작성에 필요한 사항을 벤치마킹 한다. 더불어 가중치 작성과 관련된 규칙과 매뉴얼을 분석하여 가중치 작성시 함께 검토할 사항을 살펴보고자 한다. 관련 자료로는 통계조정업무매뉴얼(통계청), 표본설계 및 관리지침(통계청 예규 제198호, 2017.3.10.), 자체통계품질진단 매뉴얼(조사통계), 국가통계품질진단 매뉴얼(조사통계) 등이 있다.

제4장에서는 패널구축 현황을 정리하고, 가중치 산출시 고려해야할 요건을 점검한다. 그리고 장애인삶 패널조사 1차년도 조사결과 분석에 필요한 적절한 가중치 산정을 위한 방안들을 검토하여, 최종가중치 산출 방안을 제시한다. 마지막 장에서는 결론을 제시하였다.

2) 연구 수행 방법

(1) 기본프레임

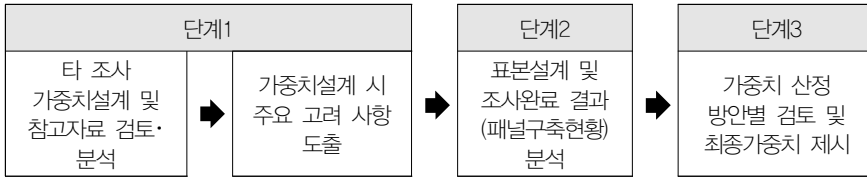
장애인삶 패널조사를 위한 적정 가중치 산출 방안 연구를 위해 문헌과 선행자료 분석을 토대로 이론적 배경을 검토하고, 타 조사의 가중치 산출을 분석하여 시사점을 도출하여, 장애인삶 패널조사의 표본설계를 기초로 가중치 산정 방향을 설정한다. 가중치 산정 방안 검토 과정에서 표본이론 및 실무적으로 발생한 이슈에 관해서는 학계 및 통계전문가와의 지속적 논의²³⁾를 토대로 검토·결정하였다.

연구절차는 다음 세 단계로 요약할 수 있다. ① 가중치 이론 검토, 타 조사 가중치설계 및 참고자료를 검토, 분석 하여 시사점을 도출하고, ⇨ ② 표본설계 및 패널 구축 결과를 분석한 후, ⇨ ③ 장애인삶 패널조사를 위한 적절한 최종가중치를 산출한다.

23) 가중치에 대한 자문은 장애인삶 패널조사 표본설계 내역에 대한 이해와 통계분야의 표본이론에 대한 자문을 필요로 하는바, 통계조사 분야의 학계 및 외부전문가로 장애인삶패널조사 표본설계 자문을 맡았던 전문가로 구성하였다. 가중치 산출 방안 검토를 위한 이론 및 실무적 이슈를 자문회의를 통해 검토하였다.



〈표 1-3〉 연구 수행 과정



(2) 단계별 주요 내용

- (단계1) 가중치 이론 및 타 조사 가중치 설계 및 참고자료 분석
 - 선행 참고자료와 문헌을 토대로 가중치 설계를 위한 이론적 측면을 검토하고, 장애인삶 패널조사 가중치 산출을 위해 벤치마킹이 필요한 사항과 설계상 시사점을 도출한다. 추출 단위가 개인인 전국 단위 조사들을 중심으로 가중치 산출시 고려사항 및 주요 특징을 검토하였다.
- (단계2) 표본설계 및 패널 구축 결과 분석
 - 표본설계와 패널 구축 결과를 분석하여 가중치 설계의 기본 방향을 설정한다. 이때 통계의 공표단위를 고려하여 가중치 설계시 추가적으로 반영 가능한 사항들을 검토한다.
- (단계3) 1차년도 가중치 산정 방안 검토 및 최종가중치 제시
 - 문헌 및 선행자료 분석과 표본설계 및 패널 구축 결과를 토대로 장애인삶 패널조사 1차년도 가중치 산정 방안을 마련한다. 연구의 전반에 포함하는 내용은 「통계승인업무 처리지침」과 통계청의 품질진단 매뉴얼(조사통계)²⁴⁾을 참고하여, 제시되어 있는 기준을 검토의 기본 틀로 삼았다. 그리고 전문가 자문회의를 거쳐 최종가중치를 제시하였다.

24) 자체통계품질진단 및 정기통계품질진단을 모두 포함한다.

〈표 1-4〉 연구 추진 절차



Ⅱ. 가중치 작성 이론 고찰

II 가중치 작성 이론 고찰

1. 개요

표본조사에 가중치 적용에 대한 기본 발상은 비편향 추정을 염두 한 것이지만, 이는 표본조사 수행 과정에서 나타나는 다양한 문제를 보정할 수 있다는 측면에서도 중요한 역할을 한다. 가중치는 기본적인 추정량 공식(기본 가중치 적용)을 설명해줄 뿐만 아니라 무응답, 추출틀 결함을 포함하여 다양한 문제를 해결하는데 있어서 추정식을 보정하는 수단이다. 이러한 이유로 표본조사에서 가중치를 부여한 후 추정치를 산출하는 것이 일반화되고 있다 (김규성, 2006).

일반적으로 조사결과에 대한 가중치 작업은 설계가중치, 무응답조정, 사후보정(김재광, 2006)의 과정 등을 거치게 된다. 장애인삶 패널조사 역시 큰 틀에서는 일반적인 과정을 거치게 될 것이다. 따라서 본 장에서는 장애인삶 패널조사의 1차년도 조사결과 분석에 필요한 적절한 가중치 작성 방안을 마련할 수 있도록 관련된 기본적 이론을 살펴보고자 한다. 이때 본 장에서 사용하는 기호(notation)들은 참고문헌의 기호를 그대로 활용하여 혼란을 최소화 하고자 하였다.

가중치를 부여하는 첫 번째 이유는 불균등한 추출확률이나 응답률 보정을 통한 표본의 대표성 확보이다 (Kish 1992; 류제복 1993; 김규성, 2006에서 재인용). 이는 작은 확률로 얻어진 데이터에 상대적으로 큰 가중치를 부여함으로써 표본데이터의 대표성을 보정해 주는 것이다 (김규성, 2006)²⁵⁾. 무응답에 의한 결측에 대해서는 결측으로 인한 부족분을 보충하기 위해 응답한 단위에 대한 가중치를 일부 부풀림으로써 해결한다. 또한 추출틀이 가지고 있는 결함을 극복하기 위해 추정식을 보정하는 수단으로도 가중치를 조정한다. 이런 경우 조사에서 추정되는 결과가 센서스에서 얻은 추정값 또는 매우 정확도가 높은 다른 조사에서 얻은 추정값을 벤치마킹하도록 가중치를 조정하게 된다.²⁶⁾ 가중치를 부여하는 두 번째 이유는 보조변수의 일치성을 토대로 추론의 효율을 높이기 위함이다.²⁷⁾

본 연구에서는 장애인삶 패널조사 가중치 사후보정을 위해 레이킹비, 사후층화

25) 김규성. (2006). 유한모집단에서 가중평균에 포함된 가중치의 효과. 조사연구, 7(2), 53-69.

26) 김영원 외. (2012). 표본조사의 이해와 활용, 7th edition. CENGAGE Learning.

27) 본 문단은「김규성. (2006). 유한모집단에서 가중평균에 포함된 가중치의 효과. 조사연구, 7(2), 53-69.」을 참고하여 요약·정리하였다.



를 적용해 보고 그 결과를 비교·검토하여 적절한 최종가중치 산출 방안을 마련하고자 한다.

2. 기본 이론

1) 사후층화²⁸⁾

일반적으로 표본설계 단계에서 다양한 보조정보(auxiliary information)를 활용하기 위해서는 해당 보조정보가 표본추출틀에 직접 존재해야하고, 이는 표본설계 시 직접 사용되어야 한다 (권순필 외, 2017). 이에 층화추출법은 표본추출틀에 층화(stratification)에 필요한 모든 변수들의 정보가 모든 추출단위에 대해 완벽하게 주어진 경우에만 사용할 수 있는 방법이다 (박민규와 강현철, 2016).

그러나 국가기관의 공식 통계 산출 목적의 표본설계에서 사전에 층화를 위한 정보를 모든 추출단위에 대해 확보할 수 없는 경우가 왕왕 발생한다. 예를 들어 개별 조사대상 응답자들의 나이, 성별, 교육 등의 정보에 대해서는 조사 완료 전에 층화 시 사용할 수 없다. 하지만 총조사(census)에서는 전체 모집단을 구성하는 각 조사 단위들에 대해서 이 변수들의 정보를 얻을 수 있는데, 이렇게 총조사를 통해 알려진 특성별 총수들을 표본조사 결과 분석시 모집단 총계들의 추정치를 얻기 위한 모집단 크기의 척도로 활용할 수 있다. 이러한 방법을 사후층화(post-stratification)라고 한다 (양미애, 2005).

이와 같이 보조정보가 표본추출틀에 존재하진 않지만 외부로부터 모집단에 대한 신뢰할 수 있는 보조정보가 주어진 경우에는 이를 이용하여 비편향추정량(일반적으로 Horvitz-Thompson; HT 추정량)보다 효율적인 추정량을 정의할 수 있는데, 이러한 대표적인 추정량이 사후층화 추정량(poststratified estimator)이다 (박민규, 강현철, 2016).

28) 본 절은 “권순필, 김경미, 최재혁, 임찬수 (2017). 2017년 가계지출조사 가중치 사후보정방법 연구, 통계개발원.”, “양미애 (2005). 사후층화 추정법의 효율성에 관한 연구, 숙명여자대학교 석사학위 논문.”, “박민규, 강현철 (2016). SAS를 활용한 표본설계와 통계적 추론, 자유아카데미.”, “Li-Chun Zhang (1998). Post-Stratification and Calibration, Statistics Norway Research Department, Discussion Papers, No. 2016”, “D. Holt and T. M. F. Smith (1979). Post Stratification, Journal of the Royal Statistical Society. Series A (General), Vol. 142, No. 1, pp. 33-46”을 참고하여 요약·정리 하였다.

전체 모집단이 크기가 $\sum_{h=1}^H N_h = N$ ($h = 1, \dots, H$)인 H 개의 사후층으로 구성되고, 전체 조사단위수는 N 개라고 가정하자. y_{hi} 는 h 층의 i 번째 조사단위에 대한 관심변수 값이라 하자($h = 1, \dots, H; i = 1, \dots, N_h$). 각 층에서 추출된 조사단위 (n_1, n_2, \dots, n_H) 개가 사후에 확인되었다고 하자($\sum n_h = n$). 이때 유의할 점은 각 사후층에서 관측된 표본크기 (n_1, n_2, \dots, n_H)는 조사가 진행된 후에 관측된 값들이라는 점이다.²⁹⁾ 여기서 전체 표본크기 n 은 사전에 고정된 상수이다. 이때 SRS 하에서 모평균의 사후층화 추정량은 다음과 같이 정의된다.

$$\bar{y}_{post} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} \frac{y_{hi}}{n_h} = \sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} y_{hi} / w_{..}$$

, 여기서 $w_{hi} = \frac{N_h}{n_h}$, $w_{..} = \sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi}$. 이때 개별 관측치 y_{hi} 에 부여되는 w_{hi}

를 사후층화 가중치(post-stratified weight)라고 한다. 이러한 사후층화 가중치를 계산하는 일반적인 절차는 다음과 같다. d_{hi} 를 사용된 표본설계에 근거하여 h 번째 사후 층에 속한 i 번째 개체에 부여된 설계가중치라고 하자. h 번째 사후층의 모집단 크기를 N_h ($h = 1, \dots, H$)라고 할 때, 사후층화 가중치는 다음과 같이 된다 (박민규와 강현철, 2016).³⁰⁾

$$w_{hi} = d_{hi} \frac{N_h}{\sum_j d_{hj}}$$

이러한 사후층화 추정량은 널리 사용되지만, 사후층이 어떻게 구성되는지에 따라서 사후층에 할당된 표본이 없는 극단적인 상황이 발생할 수 있다(즉, 몇몇의 n_h 가 0 (또는 1 또는 2)인 경우 이다³¹⁾). 사후층화의 경우는 이런 경우에 대해서는 통제할 수 없다는 것이다.

29) 양미애. (2005). 사후층화 추정법의 효율성에 관한 연구. 숙명여자대학교 석사학위 논문.

30) 박민규, 강현철. (2016). SAS를 활용한 표본설계와 통계적 추론. 자유아카데미.

31) 양미애. (2005). 사후층화 추정법의 효율성에 관한 연구. 숙명여자대학교 석사학위 논문.

$W_j = N_j/N$ for $j = 1, \dots, J$ and $w_i = N_{j(i)}/(N n_{j(i)})$ for $i = 1, \dots, n$. Problems arise if $n_j = 0$ (or 1 or 2) for some cell j . (노르웨이 보고서)



2) Calibration³²⁾³³⁾

보조정보가 표본추출틀에 존재하지 않지만 외부로부터 모집단에 대한 신뢰할 수 있는 보조정보를 얻을 수 있는 경우 이를 이용하여 비편향추정량(일반적으로 HT 추정량)보다 효율적인 추정량을 정의할 수 있다(박민규와 강현철, 2016; 권순필 외, 2017에서 재인용). 이에 추정의 효율을 높이고자 여러 보조변수를 활용하게 된다.

Calibration 추정량은 사후층화를 일반화 한 것으로, 보조정보를 적절히 이용하는 추정방법으로 Deville과 Sarndal (1992)은 보조정보의 활용 측면에서 매우 효율적인 것으로 알려진 회귀추정량을 포함한 여러 추정량들의 집합을 Calibration 추정량으로 정의하였다 (박서연과 박민규, 2017). 이러한 Calibration은 대표적인 가중치 사후보정 방법이다 (양미애, 2005).³⁴⁾

Calibration 추정량에 대한 통계 이론적 논의의 시작은 Deville과 Sarndal (1992)이었으며, 현재까지도 활발하게 이루어지고 있다. Deville과 Sarndal (1993)은 교차표에서 모집단의 주변합(marginal count)은 알지만 각 셀의 빈도(cell count)는 알려져 있지 않은 경우, 또는 소지역 추정의 경우에는 셀가중이 아닌 레이킹 비 가중법과 같은 주변합에 대한 캘리브레이션 보정이 더 효율적이라고 제시하였다(이은, 2019).³⁵⁾

Deming과 Stephan(1940)은 보조정보가 표본설계시 사용되지 않은 보조정보의 주변분포만 알고 있는 경우 최소제곱방법과 반복비례적합 방법으로 표본조사 결과를 모집단 총계와 일치시켰고, 이 방법들이 동일한 결과를 갖음을 발견하였다. 이후 Deville과 Sarndal(1992)은 보조정보가 주어진 경우 유한모집단의 총합 추정량을 구하는 방법으로 Calibration 방법을 정의하고 제안하였다. 이들이 제안한 Calibration 추정량은 거리함수 $G_k(w_k, d_k)$ 의 형태에 따라 회귀추정량, 레이킹비 추정량 등 다양한 형태로 나타나지만, 표본의 크기가 충분히 커지면 모든 Calibration

32) 본 절은 “권순필, 김경미, 최재혁, 임찬수 (2017). 2017년 가계지출조사 가중치 사후보정방법 연구, 통계개발원.”, “통계개발원 (2017). 국가통계연구 소식 제12호.”, 양미애 (2005). 사후층화 추정법의 효율성에 관한 연구, 숙명여자대학교 석사학위 논문.”, “박민규, 강현철 (2016). SAS를 활용한 표본설계와 통계적 추론, 자유아카데미.”, “Li-Chun Zhang (1998). Post-Stratification and Calibration, Statistics Norway Research Department, Discussion Papers, No. 2016”, “D. Holt and T. M. F. Smith (1979). Post Stratification, Journal of the Royal Statistical Society. Series A (General), 142(1), 33-46”을 참고하여 요약·정리 하였다.

33) 김재광 (2006). 표본조사의 가중치 작업에 대한 고찰 : Composite calibration estimator의 소개, 한국조사연구학회 춘계학술대회 발표논문집, pp. 125-130.

34) 양미애 (2005). 사후층화 추정법의 효율성에 관한 연구, 숙명여자대학교 석사학위 논문.

35) 이은. (2019). 국민건강영양조사 가중치 산출 방안 연구. 고려대학교 석사학위 논문.

추정량들이 회귀추정량과 근사적으로 같고 Calibration 추정량들의 분산도 회귀 추정량의 분산과 근사적으로 같게 된다 (Deville과 Sarndal, 1992; 권순필 외 (2017)에서 재인용)

유한모집단 $U = \{1, \dots, k, \dots, N\}$ 에서 확률표본 s ($s \subset U$)가 표본선택 $p(\cdot)$ 으로 추출되었을 때, $p(s)$ 는 표본 s 가 뽑힐 확률이 된다. 개체포함확률은 $\pi_k = \Pr(k \in s) > 0$, $\pi_{kl} = \Pr(k, l \in s) > 0$ 라고 하자. y_k 를 k 번째 모집단 개체의 관심값, $\mathbf{x}_k = (x_{k1}, \dots, x_{kj}, \dots, x_{kJ})^T$ 을 보조변수 x 들로 이루어진 벡터라고 하면, y_k 는 보조변수 벡터 $\mathbf{x}_k = (x_{k1}, \dots, x_{kj}, \dots, x_{kJ})^T$ 와 연관되어 있으므로 표본 s 에 포함된 개체 k 에 대해 (y_k, \mathbf{x}_k) 가 관측된다. \mathbf{x} 의 모집단 총계 $\mathbf{t}_x = \sum_U \mathbf{x}_k$ 이며, 이는 정확히 알려진(known) 값이다. 이 보조정보는 총조사, 행정자료 등 다양한 출처로부터 얻을 수 있다. 추정 목표는 모집단 총계 $t_y = \sum_U y_k$ 로, 설계 가중치 $d_k = 1/\pi_k$ 는 비편향추정량인 HT 추정량을 정의하기 위해 사용된다($\hat{t}_{y\pi} = \sum_s y_k/\pi_k = \sum_s d_k y_k$).

유한모집단 총합에 대한 Calibration 추정량은 다음과 같다 (Deville과 Sarndal, 1992).

$$\hat{t}_{yw} = \sum_{k \in s} w_k y_k$$

여기서 w_k 를 Calibration 가중치라고 하며, w_k 는 Calibration식 $\sum_{k \in s} w_k \mathbf{x}_k = \mathbf{t}_x$ 의 제약하에서 주어진 거리함수를 $\sum_s G_k(w_k, d_k)$ 를 최소화 해야 한다. $\min\left(\sum_s G_k(w_k, d_k)\right)$ 을 만족하는 가중치 w_k , $k \in s$ 는 $w_k = d_k F_k(\mathbf{x}_k' \boldsymbol{\lambda})$ 와 같은 형식으로 표현될 수 있고, $\boldsymbol{\lambda}$ 는 라그랑지 승수 이다. 이러한 Calibration 추정량은 사용되는 거리함수에 따라 다르게 정의된다.³⁶⁾ Calibration 추정량이 근사적으로라도 설계비편향(Asymptotically Design Unbiased) 이기 위해서는 각 추정량에 대하여 $T_s = \sum_s \mathbf{x}_k \mathbf{x}_k'$ 의 역행렬이 존재해야만 한다(권순필 외, 2017).

36) 이에 대해서는 하단의 표를 참고하고, 그 외 자세한 사항은 Deville and Sarndal (1992)와 박서연, 박민규 (2017)를 참고하기 바란다.



〈표 II-1〉 거리함수 종류와 캘리브레이션 식³⁷⁾

Case	$q_k G_k(w_k, d_k)$	$F_k(q_k \mathbf{x}_k' \lambda)$
1	$(w_k - d_k)^2 / 2d_k$	$1 + q_k \mathbf{x}_k' \lambda$
2	$w_k \log(w_k / d_k) - w_k + d_k$	$\exp(q_k \mathbf{x}_k' \lambda)$
3	$2(\sqrt{w_k} - \sqrt{d_k})^2$	$(1 - q_k \mathbf{x}_k' \lambda / 2)^{-2}$
4	$-d_k \log(w_k / d_k) + w_k - d_k$	$(1 - q_k \mathbf{x}_k' \lambda)^{-1}$
5	$(w_k - d_k)^2 / 2w_k$	$(1 - 2q_k \mathbf{x}_k' \lambda)^{-1/2}$

3) Raking ratio³⁸⁾

레이킹 비(raking ratio) 추정량은 앞 절에서 살펴본 Calibration 추정량의 한 형태이며 case2이다. 레이킹비추정량(\hat{t}_r)은 다음과 같이 정의 된다.

$$\sum_{k \in s} G_k(w_k, d_k) = \sum_{k \in s} w_k \log(w_k / d_k) - w_k + d_k$$

$$w_{k, rak} = d_k \exp(\mathbf{x}_k' \lambda)$$

$$G_k(w_k, d_k) = w_k \log(w_k / d_k) - w_k + d_k, \quad F_k(\mathbf{x}_k' \lambda) = \exp(\mathbf{x}_k' \lambda)$$

여기서 d_k 는 설계가중치이며, λ 는 $w_k = d_k F_k(\mathbf{x}_k' \lambda)$ 의 라그랑지 승수이다. 레이킹 비 가중치는 주변합에 Calibration 보정을 하는 것으로, 이때 λ 의 해는 Newton-Raphson방법으로 산출된다.

$$\sum_{i \in s_{i+}} w_{k, rak} = N_{i+}, \quad \sum_{i \in s_{+j}} w_{k, rak} = N_{+j}$$

$$F(\mathbf{x}_k' \lambda) = F(\lambda_{ij}) = N_{ij} / \hat{N}_{ij}$$

레이킹 비 추정은 반복비례법(iterative proportional fitting), 반복비례가중법 이라도 부른다.

37) (출처) 이은 (2019). 국민건강영양조사 가중치 산출 방안 연구. 고려대학교 석사학위 논문.

38) 본 절은 “이은. (2019). 국민건강영양조사 가중치 산출 방안 연구. 고려대학교 석사학위 논문.”에서 발췌하였다

4) 무응답 조정³⁹⁾

단위무응답의 경우 응답률에 따른 가중치 조정으로 편향을 줄인다. 표본설계 시점에 갖고 있던 제한적 정보 이외에 조사로 추가적인 정보를 얻는 것이 불가능하기 때문이다. 이에 관한 일반적인 처리방법으로 응답가능성(확률)이 유사한 단위로 조정층을 구성하는 방법⁴⁰⁾, 관심변수가 비슷한 값을 가질 것으로 예측되는 단위로 조정층을 구성하는 방법⁴¹⁾이 있다.

y_i 는 관심변수, Y 를 모집단 관심변수의 총계라 하자. 무응답이 발생하지 않은 경우 HT 추정량 $\hat{Y}_{HT} = \sum \pi_i^{-1} y_i$ 는 Y 에 대한 비편향 추정량이 된다. 여기서 π_i 는 i 번째 단위의 포함확률이다.

단위무응답이 발생하는 경우에 추출된 표본단위가 응답할 확률 ϕ_i 를 알 수 있다고 하면, HT 추정량은 $\hat{Y}_{HT} = \sum \pi_i^{-1} \phi_i^{-1} y_i$ 이 모집단 총계 Y 에 대한 비편향추정량이 된다. 여기서 ϕ_i^{-1} 을 무응답 조정 상수라고 한다.

무응답 조정상수를 추정하는 방법 중 하나는 무응답 층을 구성한 후 각 조정층내에서 응답률을 계산하는 것이다. 만약 단위 i 가 조정층 c 에 속하는 경우 가장 간단한 방법은 ϕ_i 를 해당 조정층 내의 전체 단위 중 응답한 단위의 비율로 추정하는 것이다. 그러나 이는 응답률에 대한 비편향 추정량이 아니게 된다.

$$\hat{\phi}_c = \frac{r_{c^+}}{n_{c^+}}$$

최소한 근사적으로 응답률에 대한 비편향성을 만족하는 결과를 얻을 수 있는 가중치는 다음과 같다.

$$\hat{\phi}_c = \frac{\sum_{i \in R_c} \pi_i^{-1}}{\sum_{i \in S_c} \pi_i^{-1}}$$

, 여기서 S_c 는 조정층 c 에 포함된 전체 단위들의 집합을 나타내고, R_c 는 조정층 c 에서 응답단위들의 집합을 의미한다.

39) 본 절은 「김영원, 남시주. (2009). 표본조사에서 무응답 가중치 조정층 구성방법에 따른 효과. 한국통계학회논문집, 16(1), 103-113.」, 「김재광. (2017). 표본조사론 (2판). 자유아카데미.」를 참고하여 작성하였다.

40) 편향을 줄이는 것에 초점을 맞추고 있다.

41) 이는 추정량의 분산을 줄이는 것에 초점을 맞추고 있다. (출처) 김영원, 남시주. (2009). 표본조사에서 무응답 가중치 조정층 구성방법에 따른 효과. 한국통계학회논문집, 16(1), 103-113.

Ⅲ. 타 조사 가중치 설계 및 참고자료 분석

III 타 조사 가중치 설계 및 참고자료 분석

1. 개요

본 장에서는 타조사의 가중치 작성 방법을 살펴보고자 한다. 검토를 진행한 조사는 전국 단위 대규모 조사, 장애인삶 패널조사와 마찬가지로 추출단위가 개인인 조사, 등록장애인 명부를 표본추출틀로 활용한 전국 단위 조사⁴²⁾등을 중심으로 살펴보았다.

가중치 작성을 위해서는 먼저 기존의 표본설계 및 표본추출과정을 충분히 검토하는 것이 필요하므로, 각 조사별 표본설계를 기초로 표본추출과정을 살펴보았다. 이를 토대로 가중치의 특성을 살펴보고, 가중치 작성 단계에서 주로 고려해야할 사항을 사전에 점검하여, 가중치 작성에 있어서 주요 시사점을 도출하고자 한다. 기존 조사들의 특징을 파악·검토하는 것은 가중치 작성의 통계 일반적 측면과 모집단의 특수성을 장애인삶 패널조사의 가중치를 작성함에 있어서 고려할 수 있어 가중치 설계 단계에서 많은 도움이 될 것으로 생각된다.

주로 살펴본 조사는 장애인경제활동조사, 특수교육실태조사(특수교육대상자 부문), 장애인고용패널조사, 산재보험패널조사, 한국아동·청소년패널조사, 여성가족패널조사이다. 장애인경제활동조사와 특수교육실태조사는 횡단면 조사이며, 후술된 조사는 패널조사로 종단면 조사이다. 장애인삶 패널조사의 1차년도 가중치는 횡단면 조사 가중치와 동일하다. 본 장에서 다른 조사들을 검토·분석하는 것은 앞서 언급한 바와 장애인삶 패널조사에 적절한 가중치 설계를 수행하고자 하는 목적으로 가중치 작성시 필요한 사항에 대한 벤치마킹사항 및 시사점을 도출하는 것이다. 따라서 이러한 내용이 다른 조사의 단점이나 개선필요 사항 등에 대한 논의가 아님을 사전에 밝혀둔다. 이러한 이유로 각 조사에 대한 조사목적 및 조사배경, 조사내용 등에 관련한 내용은 본 장의 목적을 벗어나기 때문에 이에 관해서는 다루지 않는다. 따라서 조사배경 및 조사내용, 조사영역 등과 관련한 사항에 대해서는 해당 조사의 연구보고서를 참고하기 바란다.

본 장에서는 각 조사들의 표본추출과정, 가중치 산출 과정 측면에서 주요 특징을 분석하고, 가중치 산정시의 시사점을 도출한다. 이와 관련한 사항은 해당 조사의

42) 보건복지부의 등록장애인 DB 명부를 활용한 조사를 의미한다.



기초분석 보고서, 통계청 통계설명자료 홈페이지⁴³⁾에서 제공하는 메타자료와 통계청의 정기통계품질진단(또는 자체통계품질진단) 보고서를 주로 참고하여 정리하였다. 따라서 표본추출과정과 가중치 산출 과정 및 검토사항에 대한 설명이 부족한 부분이 있어서, 구체적인 내용을 파악하기에는 어려운 부분들이 존재할 수밖에 없었다. 이러한 부분은 각 조사와 관련된 문헌과 보고서를 참고하여, 일부 추론하여 가중치 작성 절차에 대해 정리하였다. 하단의 표는 본 장에서 살펴볼 조사별 최종 가중치에 대해 간략하게 요약한 것이다. 주어진 정보를 이용하여 파악 또는 추론한 내용은 각 절에서 자세히 논하고자 한다.

〈표 III-1〉 조사별 최종가중치 요약

구분	장애인고용패널조사(2차웨이브)	장애인고용패널조사(1차웨이브)
표본추출방법	층화추출	층화계통추출
공표범위	성별, 연령별, 장애유형별, 장애정도별	—
층화변수	장애유형(6), 장애정도(2), 연령(4)	지역(15), 장애유형, 연령대
최종가중치	<ul style="list-style-type: none"> 설계가중치 → 레이킹 비 조정 → 이상가중치 조정 레이킹 변수 : 성, 연령, 지역, 장애 유형, 장애정도, 고용 여부 장애유형 수 8개 이하 읍면동 제외로 표본추출틀과 모집단 불일치 발생하였으며, 레이킹 비 기법 적용 	<ul style="list-style-type: none"> 설계가중치 → 벤치마킹보정 설계가중치 : 지역별, 연령대별, 성별 벤치마킹 : 장애유형, 성별에 대한 벤치마킹보정 가중치 계산 후 경제활동상태에 대해 벤치마킹보정
구분	특수교육실태조사(2014) : 특수교육대상자	장애인경제활동실태조사
표본추출방법	층화2단집락추출	이상추출
공표범위	장애유형, 학교과정, 성, 배치형태	장애유형별(15), 장애등급별(2)
층화변수	장애유형(10개), 학교과정(4개)	장애유형(15), 장애등급(2), 지역
최종가중치	<ul style="list-style-type: none"> 설계가중치 → 무응답조정 → 사후층화 조정 조정변수 : 장애유형, 학교과정, 성, 배치형태 (레이킹비 조정) 	<ul style="list-style-type: none"> 설계가중치 → 레이킹 비 조정 레이킹 변수 : 성, 연령, 지역, 장애 유형, 장애등급 장애유형 수 8개 이하 읍면동 제외로 표본추출틀과 모집단 불일치가 발생하였으며, 레이킹 비 기법 적용

43) 통계청 통계설명자료 : <http://meta.narastat.kr/>

구분	산재보험패널	한국아동·청소년패널조사	여성가족패널조사
표본추출 방법	층화추출	2단 집락 추출	2단 집락 추출
공표범위	성별, 연령별, 장애등급별, 경제활동상태별	전국	전국
층화변수	<ul style="list-style-type: none"> • 6개 장애등급, 전국 9개 권역, 2개 재할 서비스 이용여부 	<ul style="list-style-type: none"> • 16개 시도, 도시규모별 	<ul style="list-style-type: none"> • 조사구 추출 : 시도별 (도시화 수준, 산업별 종사자 수 비중, 주거형태별 가구 비중, 가구원수별 가구분포, 가구의주의 연령, 가구의주의 성별 변수 고려) • 조사구내 가구 수 규모를 크기변수로 확률비례계통 추출 적용
최종 가중치	<ul style="list-style-type: none"> • 설계가중치 → 사후 층화 → 레이킹 비 조정 	<ul style="list-style-type: none"> • 설계가중치 → 무응답 조정 → 사후층화 	<ul style="list-style-type: none"> ◦ 기본가중치* → 이상가중치 조정 → 사후층화 *일반조사구내 거주 가구 응답률에 체계적 차이가 없다는 가정하에서 응답률 보정하여 산출
	<ul style="list-style-type: none"> • 설계가중치 : 권역, 장애등급, 재할서비스 여부 • 사후 층 : 성, 연령대, 요양기간 	<ul style="list-style-type: none"> • 설계가중치 : 학교, 학급 • 사후 층 : 16개 광역시도 × 성별 	<ul style="list-style-type: none"> • 벤치마킹조정 변수 : 시도

추가적으로 가중치 작성과 관련한 참고자료로서 관련 규정 및 매뉴얼을 살펴보고자 한다. 이를 분석함으로써 현행 국가통계에서 요구하는 조사통계의 기본특성을 파악하고, 조사결과 공표 및 품질진단 등에서 이슈가 되는 일반적인 사항을 전반적으로 파악할 수 있을 것이다. 이러한 목적으로 본 장에서 살펴본 관련 자료는 다음과 같다.



□ 타조사의 표본설계 분석

- └ 장애인경제활동실태조사
- └ 특수교육실태조사
- └ 장애인고용패널조사
- └ 산재보험패널조사
- └ 한국아동·청소년패널조사
- └ 여성가족패널조사

□ 규칙·매뉴얼

- └ 통계 승인업무 처리지침 (통계청 예규 제213호, 2017.12.13.)
- └ 통계 조정 업무 매뉴얼 (2015)
- └ 국가통계품질관리 정기통계품질진단 매뉴얼 (2017)

2. 타조사 가중치 산출 방법 분석

1) 장애인경제활동실태조사⁴⁴⁾

장애인경제활동 실태조사는 1년 주기의 횡단면 조사이다. 본 절의 내용은 가장 최근 보고서인 2018년 장애인경제활동실태조사 보고서를 토대로 필요한 내용을 발췌하여 재구성 하였다. 이때 부족한 정보는 공개되어 있는 장애인경제활동 실태조사 표본설계 보고서 중 최신 버전인 2014년 표본설계 보고서를 참조하였다.⁴⁵⁾

(1) 표본추출

장애인경제활동실태조사는 전체 표본크기 11,000명을 목표로 표본추출은 이상 추출법(two-phase sampling)을 적용하고 있다.⁴⁶⁾ 표본추출은 1상에서 집락으로 읍·면·동을 추출하고, 2상에서 장애인 표본을 추출하여 최종표본을 구성하였다. 읍·면·동⁴⁷⁾ 표본은 1단계 집락추출법을 통해 추출하였다. 이때 읍·면·동을 집락으로 고려하였으며, 추출시 17개 특별광역시·시도별로 층화하여 계통추출 하였다. 이때 내재적 층화 효과를 얻고자 지역 층 내에서 장애유형 수와 장애인 수, 그리고 집단거주 장애인 수로 정렬하여 읍·면·동을 추출하였다. 추출된 읍·면·동 내 거주하는 모든 장애인을 2상의 장애인 표본추출을 위한 1상의 표본으로 구성하였다.

2상에서는 1상의 표본을 장애유형, 장애등급, 지역별로 층화⁴⁸⁾하여 최종 표본을 추출하였다. 이때 각 층⁴⁹⁾에서 장애유형, 장애등급, 읍·면·동, 성별, 연령, 그리고 집단거주, 시설명으로 정렬하여 계통추출 하였다.

44) 본 절에 이어지는 내용은 2014년 장애인 경제활동 실태조사 표본설계 연구 보고서 및 통계청의 통계설명자료, 2018년 장애인경제활동실태조사 보고서에 공개된 내용을 바탕으로 재구성 하였다. (출처) 박만규. (2014). 2014년 장애인 경제활동 실태조사 표본설계 연구. 한국장애인고용공단 고용개발원 연구용역: 통계설명자료: <https://meta.narastat.kr/>; 박혜원, 김호진. (2017). 2017년 장애인경제활동실태조사. 한국장애인고용공단 고용개발원.

45) 장애인경제활동조사는 표본설계 방법의 큰 틀을 매년 유지하고 있다. 이는 표본설계에 사용되는 변수(장애유형, 장애등급, 지역별)에서 뚜렷한 모집단 구성 변화가 나타나지 않았기 때문이다.

46) 장애유형과 장애등급만을 이용하여 장애인 표본을 직접 추출할 경우 지나치게 많은 읍·면·동이 추출되어 조사비용의 증가와 조사기간의 장기화로 인한 비표본 오차가 커질 수 있다는 것을 고려한 것이다.

47) 각 지역별 표본 읍·면·동 추출시 장애유형의 비율이 1% 미만인 8개 장애유형(간, 뇌전증, 심장, 안면, 언어, 지폐성, 장루·요루, 호흡기 장애)이 표본에서 누락되지 않도록 표본설계 단계에서 고려 하였다.

48) 장애유형(15개) × 장애등급(2개)을 기준으로 총 27개 층이며, 장애등급은 중증과 경증으로 통합하여 사용하였다. 이때 지적장애, 지폐성장애, 정신장애는 중증에만 해당 된다.

49) 각 층별 표본크기 및 배분에 대한 사항은 해당 보고서의 표본설계 내역 참고. 층별 표본크기는 제곱근비례배분을 기준으로 하였다. 각 층별(장애유형(15) × 장애등급(2)) 결정된 표본을 각 17개 특별·광역시·도 모집단 장애인 분포에 따라 비례배분 하였다.



〈표 Ⅲ-2〉 증별 표본크기 (2018년 장애인경제활동실태조사)

(단위:명)

장애등급 장애유형	중증	경증	총합계	(구성비)
지체장애	534	1,472	2,006	18.2
뇌병변장애	808	561	1,369	12.4
시각장애	410	826	1,236	11.2
청각장애	419	1029	1,448	13.2
언어장애	65	325	390	3.5
지적장애	805	0	805	7.3
자폐성장애	511	0	511	4.6
정신장애	644	0	644	5.9
신장장애	605	346	951	8.6
심장장애	140	57	197	1.8
호흡기장애	410	22	432	3.9
간장애	45	209	254	2.3
안면장애	39	86	125	1.1
장루, 요루장애	11	391	402	3.7
간질장애	80	150	230	2.1
총합계 (구성비 %)	5,526 (50.2)	5,474 (49.8)	11,000 (100)	(100)

(2) 가중치 산출 과정(및 추론)

장애인경제활동실태조사는 개인단위 조사로 가중치는 개인단위로 부여하고 있다. 최종가중치는 설계가중치, 레이킹 비 조정을 통해 작성되며, 가중치 산출은 다음의 절차로 작성된다.

먼저 표본가중치는 읍면동 추출과 장애인 추출률 곱의 역수로 정의된다. 이에 i 읍면동에 거주하는 j 번째 장애인에 대한 표본가중치는 다음과 같다.

$$w_{ij} = w_{aij}w_{ijs_a} = \left[\sum_h \pi_{aij}^{-1} I_{aij}(h) \right] \left[\sum_{a_h} \pi_{ij|s_a}^{-1} I_{ij|s_{aj}}(a_h) \right]$$

$$, \quad \text{여기서} \quad I_{aij} = \begin{cases} 1, & (ij) \in U_h \\ 0, & (ij) \notin U_h \end{cases}, \quad I_{ij|s_a} = \begin{cases} 1, & (ij) \in s_{a_h} \\ 0, & (ij) \notin s_{a_h} \end{cases}, \quad \pi_{aij} = \frac{n_h}{N_h},$$

$\pi_{ij|s_a} = \frac{m_{a_h}}{M_{a_h}}$ 을 나타낸다. h 는 1상 표본 추출시에 모집단에서 정의된 h 번째 층을 의미하며, a_h 는 2상 표본 추출을 위해 1상 표본에서 정의된 a_h 번째 층을 나타낸다. U_h 와 s_{a_h} 는 각각 h 번째 장애인 모집단 층과 a_h 번째 장애인 1상 표본 층을 의미한다. N_h 는 h 번째 모집단 층의 읍면동 수이다. n_h 는 h 번째 모집단 층으로부터 1상 표본추출을 통해 추출된 표본 읍면동 수이다. M_{a_h} 는 a_h 번째 1상 표본 층의 장애인 수를 의미한다. m_{a_h} 는 a_h 번째 1상 표본 층으로부터 2상 표본추출시 추출된 표본 장애인 수를 나타낸다. 장애인경제활동실태조사는 효율적인 표본대체를 통해 조사를 완료⁵⁰⁾한 것으로 보이며, 따라서 가중치 조정 단계에서 별도의 무응답 조정은 이뤄지지 않을 것으로 파악된다.

표본추출시 장애유형 수 8개 이하 읍면동을 표본추출틀에서 제외하여 발생한 모집단과의 불일치는 산출된 표본가중치를 레이킹 비 조정하여 보정하였다. 이때 레이킹 비 조정 변수는 성, 연령, 지역, 장애유형, 장애등급 5개 변수를 활용하였으며 보정된 가중치는 다음과 같이 정의된다. 조정된 가중치를 적용하여 추정한 성, 연령대, 지역, 장애유형, 그리고 장애등급⁵¹⁾ 그룹의 분포는 모집단 분포와 일치하게 된다.

$$w_{ij}^{rak} = w_{ij} \exp(\lambda' x_{ij})$$

, 여기서 x_{ij} 는 레이킹에 적용된 변수들을 나타내는 지시변수 벡터이다. λ 는 $\sum w_{ij}^{rak} x_{ij} = t_x$ 의 해이고, t_x 는 각 레이킹에 사용된 벤치마킹 변수들의 범주에 해당하는 모집단 분포를 나타낸다.

50) 층별 조사원료 인원 수에 대한 정보는 보고서상에 제시되어 있지 않지만, 전체 조사인원수는 확인할 수 있었다.

표본	원표본	대체표본	계
조사원료	8,584명	2,416명	11,000명

주) 2018년 장애인경제활동조사 보고서의 결과를 재구성 하였다. (출처) 2018년 장애인경제활동실태조사, 한국장애인고용공단 고용개발원.

51) 장애유형별 장애등급 범주 병합 기준은 해당 보고서를 참고한다.



(3) 시사점

장애인경제활동실태조사는 향후 지역별 통계의 공표 가능성을 염두에 두어 표본 설계시 15개 장애유형을 모두 고려하였다. 통계의 공표시 층을 통합하여 6개로 범주로 공표하고 있다. 그리고 주요 문항에 영향을 줄 것으로 예상되는 변수를 모두 활용하여 레이킹 비 방법을 토대로 가중치 조정하여, 최종가중치를 작성하였다.

장애인삶 패널조사 역시 가중치 작성 단계에서 공표범위를 고려하여 층 통합(collapse)의 필요여부를 검토할 필요가 있다. 층화변수로 장애유형(7개 범주), 장애정도(2개 범주), 성별(2개 범주)을 활용하고 있는바, 레이킹 비 조정시 활용 가능한 변수를 검토하여 조정변수를 결정하는 것이 필요함을 알 수 있다.

2) 특수교육실태조사⁵²⁾53)

(1) 표본추출

특수교육대상자 부문 조사의 표본 추출단위는 특수교육대상자이다. 이는 이중추출법을 적용하고 있다. 표본추출은 시군구를 추출하는 단계와 대상자를 추출하는 단계로 이루어져 있다. 시군구 추출은 시군구를 집락으로 고려하여 1단계 집락 추출법을 적용하였다. 이때 시군구를 6개 군집으로 층화하고, 시군구에 따른 장애유형 비율 등 총 11개 변수로 장애유형별 대상자 분포를 반영하여 표본 시군구를 계통 추출하였다.⁵⁴⁾

특수교육대상자 추출은 층내에서⁵⁵⁾ 성별, 배치형태⁵⁶⁾ 순으로 정렬하여 앞서 추출한 시군구의 모든 특수교육대상자에 대하여 계통추출 하였다. 이때 표본크기는

52) 장애인복지법상의 장애와 일부 차이가 있지만 장애특성에 대해서는 유사성을 가졌다고 볼 수 있어, 통계의 공표 범위나 모집단의 일부 특성 측면에서 참고할 점이 있다고 판단하였다. 특수교육실태조사에 관해서는 공개되어 접근 가능한 최신 자료인 2014년 특수교육대상자 부문 조사를 검토한다.

53) 이어지는 내용은 「2014 특수교육실태조사 표본설계 (2014.6) 통계청 통계정책국 표본과」를 참고하여 재구성하였다.

54) 군집분석 결과의 6개 군집으로 층화 하였다. 1번 군집 : 정신지체 비중은 상대적으로 낮고 의사소통장애, 학습장애, 건강장애의 비중은 상대적으로 높음, 2번 군집 : 자폐성장애의 비중이 상대적으로 높음, 3번 군집 : 각 장애 유형들이 특별한 유형 없이 평균적으로 분포, 4번 군집 : 정서행동장애와 발달지체 비중이 상대적으로 높음, 5번 군집 : 정신지체 비중이 가장 높은 나머지 장애유형들은 상대적으로 가장 낮음, 6번 군집 : 시각, 청각, 지체 비중이 상대적으로 가장 높음

55) 장애유형(10개) × 학교과정(4개) = 38개 층. 발달지체는 만9세까지만 한정하기 때문에 중·고등학교는 없음. 공표단위인 장애유형, 학교과정, 성별, 배치형태 중에서 가장 중요하고 분포의 쓸림이 심한 장애유형과 학교과정을 층화변수로 선정하였다.

56) 성별, 배치형태 변수는 정렬 변수로 활용하여 내재적 층화 효과를 얻을 수 있도록 하였다.

장애유형별 배분 후, 각 장애유형별로 학교과정별 배분한 결과를 활용하였다. 그리고 약 20%의 표본을 과대추출하여, 최종 유효표본 8,400명을 목표로 하였다.

(2) 가중치 산출 과정(및 추론)

특수교육실태조사(특수교육대상자 부문)의 가중치는 설계가중치 산출, 무응답 조정, 극단가중치 조정, 그리고 레이킹비 조정을 거쳐 이루어진다. 1단계에서 시군구를 추출하고, 2단계에서는 층별 대상자를 추출하여, 표본설계 기반 설계가중치는 다음과 같이 부여되었다.

$$w_h^0 = \frac{N_h}{n_h}$$

, 여기서 h 는 층 ($h = 1, 2, \dots, 38$), N_h 는 h 층 모집단 대상자수, n_h 는 h 층 표본 대상자 수를 나타낸다.

이 후 무응답에 대해서는 적절한 조정을 거쳐 무응답조정 가중치를 산출하였는데, 응답자가 조사 자체에 불응한(단위무응답) 경우를 가중셀 조정으로 무응답조정 가중치를 산출하였다. 무응답 조정 후 발생한 극단가중치는 절사(trimming)하여 조정하였다.

특수교육실태조사(조사대상자부문)는 설계가중치와 무응답조정, 그리고 극단가중치 절사를 통해 작성된 가중치를 레이킹 비 조정을 걸쳐서 최종가중치를 산출한다. 레이킹 비 조정시 조정변수로는 활용한 변수는 장애유형, 학교과정, 성, 배치형태 이다. 이때 벤치마킹 모집단은 2014 특수교육통계를 활용하였다. 비 조정시 중요한 변수를 뒤쪽에 배치하여, 조정변수별 주변 모집단 분포에 따른 조정을 실시하였다. 레이킹 비 조정은 다음 절차에 따라 수렴조건을 만족할 때까지 수행한다.

$$1) w_{h,a}^2 = w_{h,a}^1 \times \frac{X_{h,a}}{x_{h,a}}, a = 1, 2, 3$$

, 여기서 $w_{h,a}^1$ 은 직전가중치, $x_{h,a}$ 는 변수 a (배치형태)에 대해 가중된 표본의 주변 분포를 의미한다. $X_{h,a}$ 는 변수 a 에 대한 벤치마킹 모집단의 주변분포를 나타낸다.

$$2) w_{h,a}^2 \text{를 성별 분포로 조정 } w_{h,b}^2 = w_{h,a}^2 \times \frac{X_{h,b}}{x_{h,b}}, b = 1, 2$$



3) $w_{h,b}^2$ 를 학교과정 분포로 조정 $w_{h,c}^2 = w_{h,b}^2 \times \frac{X_{h,c}}{x_{h,c}}, c = 1, 2, 3, 4$

4) $w_{h,c}^2$ 을 장애유형 분포 이용 $w_{h,d}^2 = w_{h,c}^2 \times \frac{X_{h,d}}{x_{h,d}}, d = 1, 2, \dots, 10$

(3) 시사점

조사거절, 장기적 부재 등으로 응답자가 조사 자체에 불응하여 발생한 무응답으로 인해 사전 계획된 표본 크기와의 불일치가 발생할 수 있다. 이러한 단위무응답에 대해서는 적절한 조정을 거쳐 무응답 조정 가중치를 산정하게 된다.

설계가중치 이후 가중치 작성 단계를 거치면서 지나치게 큰 가중치가 발생할 수 있는데, 이를 막기 위해 지나치게 큰 부분을 절사하는 극단가중치 조정 등을 고려하여 최종가중치를 산출할 수 있다. 또한 비 조정시 조정변수에 대한 분포를 알 수 없을 경우, 조정변수별 주변 모집단 분포에 따른 조정을 실시한다. 이때 조정변수의 적용 순서의 배치도 검토할 필요가 있다.

장애인삶 패널조사 역시 조사 과정 중 조사거절, 장기적 부재 등의 다양한 이유로 인해 발생한 무응답으로 계획한 표본보다 많은 인원이 조사완료 되었다.⁵⁷⁾ 이러한 이유로 목표표본과 조사완료 표본 규모간의 불일치가 발생하였다. 이에 레이킹 비 조정을 적용하는 것을 검토해볼 수 있다. 조정변수는 공표단위를 고려하여 적용하는 것이 필요할 것이다.

57) 장애인삶 패널조사의 패널 구축 현황에 관해서는「IV. 가중치 부여 방안. 2. 장애인삶 패널조사 구축 현황」을 참고한다.

3) 장애인고용패널조사

① 2016 장애인고용패널조사

(1) 표본추출

장애인고용패널조사는 이상추출법(two-phase sampling)을 적용하여 4,400명을 조사목표로 하였다. 표본추출은 장애인 표본을 추출하는 단계와, 읍면동을 추출하는 단계로 이뤄진다. 장애인 표본 추출에 앞서 읍면동 추출이 선행되는데, 이때 읍면동을 집락으로 고려하여 추출하였다. 그리고 추출된 전체 읍면동내에서 장애인 표본을 추출하였다.

먼저 읍면동 추출은 확률비례계통추출을 적용하였으며 이때 장애인 수와 장애유형 수를 크기변수로 고려하였다. 읍면동 추출시⁵⁸⁾에서는 지역변수를 이용하여 읍면동을 정렬한 후 추출하여 내재적 층화(implicit stratification) 효과를 얻도록 하였다. 최종적으로 총 492개⁵⁹⁾의 읍면동이 구성되었다.

그리고 추출된 읍면동내 모든 장애인을 장애유형, 장애등급, 연령 기준으로 층화⁶⁰⁾하여, 최종적으로 장애인 표본을 배분된 장애인 표본크기에 따라 계통추출하였다.⁶¹⁾

(2) 가중치 산출 과정(및 추론) : 1차년도

전체 표본크기 4,400명 조사완료율 목표를 하였으며, 1차 년도는 최종 4,577명이 조사 완료되었다. 조사완료 인원은 조사 목표 인원보다 많은 인원이다. 이는 조사가 진행되는 동안 부재 등으로 인해 발생한 무응답에 대하여 표본대체가 이루어지면서, 복수 접촉으로 인해 조사가 이루어진 장애인 모두를 포함시켰기 때문이다. 장애인고용패널조사는 조사완료한 전체 표본에 대해 가중치를 부여하였다. 이

58) 이때 시간 및 비용 등 주어진 현실적인 제약을 감안 조사 효율성을 높이기 위해 장애유형 수가 8개 이하인 읍면동은 표본추출에서 제외하였다.

59) 장애인고용패널조사는 희귀장애유형이 표본에서 누락되지 않도록 하고자 읍면동 표본 추출시 희귀장애유형을 가진 장애인이 거주하는 읍면동이 뽑힐 수 있도록 하였다. 여기서 희귀장애유형이란 모집단 비율이 낮은 간, 뇌전증, 심장, 안면, 언어, 자폐성, 장루·요루, 호흡기 장애를 희귀 장애유형으로 범주화 하고 있다.

60) 장애유형, 장애정도, 연령을 기준으로 총44개 층을 구성하고, 각 층별 표본은 비례배분법을 사용하였다. 그리고 각 부모집단별 결정된 표본을 경제활동 여부(경제활동을 참여하거나 참여를 희망하는 장애인, 희망하지 않는 장애인)에 따라 장애인의 비율이 평균적으로 3:1이 되도록 배분하였다.

61) 2016 장애인고용패널조사(2차웨이브 1차조사)보고서에서 제공하고 있는 정보만으로는 1상의 표본크기 결정에 대한 과정을 파악하기에는 정보가 부족하다.



는 패널구축 이후 부터 발생이 예상되는⁶²⁾ 패널 마모 등으로 인해 표본 규모가 감소될 수 있음을 감안한 것이다.

장애인고용패널조사는 개인 패널조사로, 개인 횡단면 가중치는 1차 년도 조사에 대해 산출된다. 최종가중치는 설계가중치, 벤치마킹 조정, 레이킹 비 조정, 이상가중치 조정을 통해 작성되며, 가중치가 부여된 1차 년도장애인 수는 4,577명이다. 파악된 절차는 다음과 같다.

먼저 설계가중치인 i 읍면동에 j 번째 장애인을 위한 표본가중치는 다음과 같다.⁶³⁾

$$w_{ij} = w_{aij}w_{ij|s_a} = \left[\sum_h \pi_{aij}^{-1} I_{aij}(h) \right] \times \left[\sum_{a_h} \pi_{ij|s_a}^{-1} I_{ij|s_a}(a_h) \right]$$

$$, \text{여기서 } I_{aij} = \begin{cases} 1, & (ij) \in U_h \\ 0, & (ij) \notin U_h \end{cases}, I_{ij|s_a} = \begin{cases} 1, & (ij) \in s_{a_h} \\ 0, & (ij) \notin s_{a_h} \end{cases}, \pi_{aij} = \frac{n_h x_{hi}}{\sum_j x_{hj}}, \pi_{ij|s_a} = \frac{m_{a_h}}{M_{a_h}} \text{ 이}$$

다. h 는 1상 표본 추출을 위해 모집단에서 정의된 h 번째 층, a_h 는 2상 표본을 추출하기 위해 1상 표본에서 정의된 a_h 번째 층을 나타낸다. U_h 와 s_{a_h} 는 h 번째 장애인 모집단 층과 a_h 번째 장애인 1상 표본 층을 의미한다. N_h 와 n_h 는 h 번째 모집단 층의 읍면동 수와 h 번째 모집단 층으로부터 1상 표본추출을 통해 추출된 표본 읍면동 수를 의미한다. M_{a_h} 와 m_{a_h} 는 a_h 번째 1상 표본 층의 장애인 수와 a_h 번째 1상 표본 층으로부터 2상 표본추출을 통해 추출된 표본 장애인 수 이다.

장애인고용패널조사는 무응답 처리를 위해 조사 단계에서 표본 대체방안을 적용하였다. 이에 1차 조사에서는 별도의 무응답 조정은 이뤄지지 않은 것으로 파악된다.

앞의 단계를 거쳐 산출된 가중치는 레이킹 비 기법을 적용하여 보정하였다. 이를 통해 표본추출틀 구성시 시간 및 비용으로 인한 현실적 제약으로 장애유형의 종류가 8개 이하인 읍면동을 제외함으로써 발생한 표본추출틀과 모집단간의 불일치를 보정하였다. 레이킹 비 보정시 성, 연령, 지역, 장애유형, 장애정도, 경제활동상태 변수 등을 벤치마킹 변수로 활용하였다. 경제활동상태 변수의 경우에는 같은 시기에 실시된 장애인경제활동실태조사의 결과를 활용하였다. 레이킹 비 보정 가중치는 다음과 같다.

62) 1차년도 패널조사는 패널구축 조사의 성격을 함께 가지고 있다.

63) 임예직 외. (2018). 2017년 2차웨이브 2차 조사. 장애인고용공단 고용개발원.

$$w_{ij}^{rak} = w_{ij} \exp(\lambda' x_{ij})$$

,여기서 x_{ij} 는 레이킹 비에 적용된 변수들을 나타내는 지시변수 벡터, $\lambda = \sum w_{ij}^{rak} x_{ij} = t_x$ 의 해 이다. t_x 는 각 레이킹 비에 적용된 벤치마킹 변수들의 범주에 해당되는 모집단 장애인 분포 이다.

보정된 가중치가 5,000을 넘는 경우에 대해서는 극단값 조정을 실시하고 있다.⁶⁴⁾ 극단 가중치 조정 후에는 특성 변수(성, 연령, 장애유형, 장애정도 등)에 대한 레이킹 비 보정을 실시하여, 1차 조사 최종가중치가 산출된다.

〈표 III-3〉 장애인고용패널 표본배분 및 조사완료 표본 수

(단위 : 명)

구분		표본배분	조사 완료 표본 수	성공률(%)
성별	남성	3,060	3,009	98.3
	여성	1,340	1,568	117
연령별	15-29세	802	831	103.6
	30-39세	1,125	1,157	102.8
	40-49세	1,306	1,341	102.7
	50-64세	1,167	1,248	106.9
권역별	수도권	2,278	2,183	95.8
	경남권	677	700	103.4
	경북권	384	385	100.3
	광주권	415	512	123.4
	충청권	437	528	120.8
	강원/제주	209	269	128.7
장애정도별	중증	1,287	1,443	112.1
	경증	3,113	3,134	100.7
계		4,400	4,577	104

출처) 장애인고용패널조사 통계정보보고서, 장애인고용공단.

64) 2018년 2차웨이브 2차 조사



(3) 시사점

장애인고용패널조사(2016)은 표본배분 수에 비해 많은 인원이 조사완료 되었다. 조사 과정 중 다양한 이유로 발생한 무응답으로 인한 표본대체 과정 중 조사완료된 표본이 포함되면서 발생하였는데, 이를 다음 년도부터 발생 가능한 패널 마모 등을 감안하여 표본에 포함하여 가중치를 작성하였다. 모집단과의 불일치는 레이킹 비 보정하였다.

장애인삶 패널조사 역시 조사완료된 인원이 표본배분 수에 비해 많은 상황인데, 다음 년도부터 발생할 수 있는 패널 마모 등을 고려하여 최종 표본에 포함시키기 위한 방안을 검토할 필요가 있다. 또한 표본추출시 표본추출틀에서 제외한 읍면동으로 인해 발생한 표본추출틀과 모집단간의 불일치는 가중치 작성 단계에서 레이킹 비 보정 등을 적용하는 방안이 있다. 레이킹 보정 변수는 공표단위 등을 고려하여, 검토할 필요가 있다.

② 2008 장애인고용패널조사⁶⁵⁾

(1) 표본추출

장애인고용패널조사(2008)는 표본추출방법으로 층화집락추출법을 적용하고 있으며, 전체 표본크기는 5,000명이다. 그리고 통계청의 2005년 인구주택총조사 10% 표본조사의 육체적·정신적 활동제한자 가구 리스트와 읍면동 등록장애인 명부를 표본추출틀로 사용하여 구축하였다.⁶⁶⁾

표본 읍면동은 인구 비례확률추출을 적용하였다. 이는 활동제한자 현황을 고려한 것이다. 표본 읍면동 추출시에는 인구주택총조사의 활동제한자 자료를 이용하였다. 장애인 표본은 추출된 표본 읍면동내에서 장애유형별 할당 후 5,000명을 계통추출 하였다. 장애인 표본의 층별 표본크기는 연령대별로 9:1 비율로 배분하였다.⁶⁷⁾ 그리고 장애유형, 지역(15개)이 고려되었다.⁶⁸⁾ 연령과 장애유형별 배분은 모

65) <https://www.narastat.kr/metascv/svc/SvcMetaDcDtaPopup.do?confmNo=383003&inputYear=2008>, (출처) 통계설명자료

66) 표본구축 당시 장애인 전체 집단의 표본추출틀의 확보가 불가능하여 다양한 자료를 활용하였다. 활동제한자 명부에는 등록장애여부에 대한 정보가 없고, 등록장애인 명부에는 미등록장애인이 포함되지 않을 수 있어 두 명부를 추출틀로 활용하여 조사모집단을 구축하였다. (출처) 장애인고용패널조사 통계정보보고서, 장애인고용공단.

67) 연령대는 15세~60세 (4,500명)와 61세~75세 연령대 (500명)로 배분하였다. 장애인의 경제활동 특성을 고려하기 위해 연령대를 구분하여 표본배분 한 것이다. 연령 비율을 그대로 적용할 경우 고연령층의 비율이 상대적으로 높게 될 것이기 때문에 연령대 구분을 고려하였다.

집단의 시도, 연령대, 장애유형 분포 기준으로 비례배분 하였다.

(2) 가중치 산출 과정(및 추론) : 1차년도

장애인고용패널(2008)의 1차 년도 조사는 5,092명을 조사 완료 하였다. 가중치는 설계가중치와 벤치마킹보정을 통해 작성되었으며, 층화변수로 활용한 지역(j), 연령대(k), 장애유형(g)과 성별(s), 경제활동상태가 가중치 산출에 반영되었다.

설계가중치는 표본추출률의 역수와 응답률의 역수로 산정되었다. j 지역 k 연령대의 s 성별을 가지는 표본의 가중치는 다음과 같이 정의되며, 시도별⁶⁹⁾, 연령대별⁷⁰⁾, 성별 조사모집단 및 표본할당 분포로부터 계산된다.

$$W_{1jks} = \frac{N_{jks}}{n_{jks}}, \quad j = 1, \dots, 15, \quad k = 1, 2, \quad s = 1, 2.$$

, 여기서 N_{jks} 는 j 지역의 k 연령대의 s 성별의 모집단 빈도 수를 의미한다. n_{jks} 는 j 지역의 k 연령대의 s 성별의 유효표본 수를 의미한다.

설계가중치는 벤치마킹 보정 하였으며, 벤치마킹 변수는 장애유형, 성별, 경제활동상태변수 이다. 먼저, 장애유형별, 성별에 대해 벤치마킹보정가중치를 계산하고, 경제활동상태에 대해서 벤치마킹보정가중치를 계산하였다.

장애유형, 성별 벤치마킹을 위한 준거집단은 2008년 3월 31일 기준 등록장애인 현황자료이며, 벤치마킹보정가중치는 다음과 같다.

$$W_{2jks} = W_{1jks} \times \frac{T_g}{S_g}, \quad j = 1, \dots, 15, \quad k = 1, 2, \quad s = 1, 2, \quad g = 1, \dots, 15.$$

, 여기서 T_g 는 g 유형별 모집단 빈도 수, $S_g = \sum_{j=1}^{15} \sum_{k=1}^2 \sum_{s=1}^2 W_{1jks}$ 이다.

경제활동상태에 대한 벤치마킹 보정은 장애인고용패널 1차년도 본조사 실시 직전에 실시한 장애인고용패널 경제활동조사 자료와 본 조사의 조사자료를 활용하였

68) 등록장애인의 특성을 대표할 수 있고 시도별로도 정책수립에 필요한 통계를 생산할 수 있도록 시도별, 장애유형별, 연령별로 층화하였다.

69) 제주도를 제외한 15개 특광역시·도

70) 15-60세, 61-75세



다.71) 장애인고용패널 경제활동조사와 장애인고용패널조사의 응답자 층, 무응답자 층 비율과 연령층을 고려하여 벤치마킹 보정 가중치72)를 계산하였다. 경제활동상태를 보정한 가중치로 분석한 결과는 모집단의 연령별 현황과 차이가 발생할 수 있다. 이에 연령층을 세분화하여 모집단과의 차이를 줄이는 방향으로 가중치를 반복적으로 보정하였다. 가중치 보정은 일차적으로 경제활동상태, 그리고 연령층을 4개 층(15-34세, 35-49세, 50-64세, 65세 이상)으로 구분하여 최종가중치를 산출하였다.

$$W_3 = W_2 \times \text{벤치마킹 보정 가중치 73)}$$

〈표 III-4〉 장애인고용패널조사 (2008) : 표본배분 및 조사 완료 표본 수

(단위:명)

장애유형	표본배분			조사 완료 표본 수		
	15세~60세	61세~75세	계	15세~60세	61세~75세	계
지체	2,312	238	2,550	2,252	304	2,556
뇌병변	402	72	474	422	85	507
시각	509	73	582	499	90	589
청각	530	63	593	502	81	583
언어	44	6	50	39	5	44
지적	211	9	220	214	8	222
자폐성	14	—	14	30	0	30
정신	143	2	145	151	4	155
신장	100	8	108	112	7	119
심장	65	9	74	72	15	87

71) 장애인고용패널조사는 1차년도 본조사를 실시하기 직전 장애인고용패널 대상자에게 통계청 '경제활동인구조사'와 최대한 동일한 조건과 기준으로 장애인고용패널 경제활동조사를 실시하였다. 여기서 파악된 횡단면적 경제활동 상태 정보를 장애인고용패널에서 추정시 활용하였다. 또한 본 조사에서 발생한 취업자의 단위무응답을 보완하는 측면에서도 활용하였다.

장애인고용패널 경제활동조사는 2008년 4월 14일 기준으로 실시되었고, 본 조사는 2008년 6월에 실시되었고 두 조사 사이의 경제활동 상태변동을 반영하여 모집단의 경제활동상태별 구조를 추정해야 한다. 이를 위해 먼저 응답자(두 조사에 모두 응답)와 무응답자(본 조사에만 응답)간의 특성이 상이할 것이라는 가정 하에 모집단을 응답자층과 무응답자층으로 구분하고 경제활동상태가 연령층별로 차이가 있을 것이므로, 표본설계시 반영된 연령 구분으로 고려하여 각 층별로 본 조사 시점에서 경제활동상태별 구성비를 추정하였다.

72) 이때 벤치마킹 보정 가중치는 경제활동상태별 추정 인구수와 경제활동상태별 가중치 합계의 비로 계산하였다. (출처) 민준영, 김호진, 강혜승, 김성천. (2015). 제8차 장애인고용패널조사. 장애인고용공단 고용개발원.

73) 취업/실업/비경제활동인구로 구분하여 연령층별로 계산

장애유형	표본배분			조사 완료 표본 수		
	15세~60세	61세~75세	계	15세~60세	61세~75세	계
호흡기	51	12	63	54	12	66
간	29	—	29	29	1	30
안면	28	1	29	27	4	31
장루.요루	31	6	37	30	9	39
간질	31	1	32	33	1	34
전체	4,500	500	5,000	4,466	626	5,092

지역	표본배분			조사 완료 표본 수		
	15세~60세	61세~75세	계	15세~60세	61세~75세	계
서울	877	85	962	866	101	967
부산	332	39	371	342	50	392
대구	263	29	292	246	32	278
인천	238	23	261	231	37	268
광주	82	10	92	93	10	103
대전	108	12	120	103	16	119
울산	63	6	69	64	7	71
경기	808	86	894	801	96	897
강원	218	24	242	216	32	248
충북	132	12	144	141	13	154
충남	209	33	242	206	39	245
전북	275	31	306	282	30	312
전남	272	37	309	276	46	322
경북	278	36	314	276	51	327
경남	345	37	382	323	66	389
전체	4,500	500	5,000	4,466	626	5,092

자료) 「제1차 장애인고용패널조사(2008년, 한국장애인고용공단)」의 내용을 재구성함



(3) 시사점

장애인고용패널조사(2008)는 층화변수로 활용한 지역, 연령대, 장애유형을 고려하였으며, 가중치 산출시 지역, 연령대, 장애유형, 성별, 경제활동상태 변수를 적용하였다. 최종가중치는 설계가중치와 벤치마킹조정을 통해 작성되었으며, 준거집단은 2008년 3월 31일 기준 등록장애인 현황자료를 활용하였다. 그리고 경제활동상태는 장애인고용패널 경제활동상태조사 자료를 활용하였다.

표본추출틀에서 정보가 확보되지 않았으나 반영이 필요한 경우에는 조사완료 후 벤치마킹조정하여 가중치작성 단계에서 반영할 수 있다. 장애인삶 패널조사 역시 설계변수가 아니더라도, 반영할 필요가 있는 경우 가중치 조정 단계에서 반영하는 방안을 검토해 볼 수 있다.

4) 산재보험패널조사⁷⁴⁾

(1) 표본추출

산재보험패널조사는 층화추출법을 적용하고 있으며, 표본규모는 2012년 산재요양 종결 근로자 2,000명⁷⁵⁾이다. 층화변수는 장해등급(6개), 권역(9개), 재활서비스 이용여부(2개)를 활용하고 있다.⁷⁶⁾ 층화변수 이외에도 패널조사결과에 영향을 미칠 것으로 보이는 성, 연령, 요양기간 변수를 표본추출 단계에서 고려하였다.

표본추출은 각 층의 표본할당 결과를 기준으로 층내에서 성, 연령대, 요양기간 등⁷⁷⁾으로 정렬한 후, 계통추출 하였다. 성, 연령대, 요양기간⁷⁸⁾에 대해 표본 추출시, 비례배분효과를 얻기 위한 추출방안이 적용되어 모집단 층의 과도한 세분화로 발생할 수 있는 불편함을 해결할 수 있도록 하였다. 산재보험패널 1차년도 조사는 본 표본 이외 예비표본을 함께 구성하였다.⁷⁹⁾

74) 통계청, 통계설명자료 : <https://www.narastat.kr/metasvc/index.do>.

이정화, 신술비 (2018). 5차 산재보험패널조사 이용자안내서. 근로복지공단.

75) 2012년 1월부터 12월까지 발생한 산재요양종결근로자(무장해자 포함)를 목표모집단으로, 모집단 규모는 89,921명이다. 여기서 주소불명과 외국인 제주도 거주자를 제외한 82,498명을 조사모집단으로 정의하였다.

76) 장해등급 : 1-3급, 4-7급, 8-9급, 10-12급, 13-14급, 무장해자. 권역 : 서울, 부산, 대구/경북, 강원, 울산/경남, 경기, 인천, 전라권, 충청권, 재활서비스 이용여부 : 이용, 미이용/

77) 모집단의 층화와 표본배분에서 고려한 층화변수는 권역, 장해등급, 재활서비스 이용여부 등이었으나 이외에도 성, 연령대, 요양기간이 패널조사결과에 영향을 미친 것으로 생각되어 표본 추출과정에서 반영하였다.

78) 성은 남자, 여자, 연령대 범주는 5개로 20대 이하, 30대, 40대, 50대, 60대 이상이며, 요양기간 범주는 4개로 3개월 이하, 3개월 초과~6개월 이하, 6개월 초과~9개월 이하, 9개월 초과 이다.

79) 산재보험패널 1차년도 조사에서는 무응답 조정을 적용하지 않은 것으로 파악된다.

〈표 III-5〉 산재보험패널조사 층별 표본할당

(단위 : 명)

권역 별	1-3급		4-7급		8-9급		10-12급		13-14급		무장해자		계
	미이용	이용	미이용	이용	미이용	이용	미이용	이용	미이용	이용	미이용	이용	
1	0	4	0	10	2	21	18	102	52	14	48	7	278
2	0	2	1	7	2	9	26	45	47	10	22	6	177
3	0	4	1	9	2	14	20	62	47	20	28	8	215
4	0	2	0	3	0	6	3	27	16	7	8	5	77
5	0	3	1	8	2	13	20	77	66	15	26	6	237
6	0	6	2	16	9	28	65	121	98	23	73	8	449
7	0	3	1	5	2	10	21	45	36	9	18	6	156
8	0	2	0	8	1	13	8	72	47	15	36	7	209
9	0	4	1	10	2	16	18	65	36	12	31	7	202
계	0	30	7	76	22	130	199	616	445	125	290	60	2,000

주) 제4차 산재보험패널조사 기초분석 보고서

(2) 가중치 산출 과정(및 추론) : 1차년도

산재보험패널조사는 개인단위 패널조사로 2012년 1월부터 12월까지 발생한 산재요양종결근조자(무장해자 포함)가 조사대상 이다. 1차년도 조사의 최종가중치는 설계가중치, 사후층화, 레이킹 비 보정을 통해 작성되며, 개인단위 가중치가 부여된다.

먼저 설계가중치는 권역(k), 장애등급(g), 재활서비스 이용여부(j)를 기준으로 표본추출률의 역수이다. 따라서 (k, g, j) 층의 l 번째 표본에 대한 설계가중치는 다음과 같다.

$$w_{(k,g,j,l)} = \frac{N_{(k,g,j)}}{n_{(k,g,j)}}, \quad k = 1, \dots, 9, \quad g = 1, \dots, 6, \quad j = 1, 2.$$

여기서 $N_{(k,g,j)}$ 는 모집단 빈도 수, $n_{(k,g,j)}$ 는 표본 빈도 수를 나타낸다. k 는 권역, g 는 장애등급, j 는 재활서비스 이용여부를 의미한다. 권역은 서울, 부산, 대구/경북, 강원, 울산/경남, 경기, 인천, 전라권, 충청권 9개 범주이다. 장애등급은 6개 범주로 1-3급, 4-7급, 8-9급, 10-12급, 13-14급, 무장해자 이다.



설계가중치 산출 단계에서 장애등급별 재활서비스 미이용층은 권역별로 적절히 통합하였다. 층화변수인 장애등급과 재활서비스 이용 여부의 교차표를 살펴보면, 4-7급, 8-9급에 대한 권역별 재활서비스 미이용자 층의 경우 층내 표본 수가 없거나 2개 미만임을 알 수 있다. 이런 경우 층내 분산을 구할 수 없다. 이에 장애등급 4-7급의 재활서비스 미이용자 층과 장애등급 8-9급의 재활서비스 이용자 층에 대해서는 유사한 성격을 지니는 권역을 하나로 묶어 권역 범주를 조정, 층을 통합 한 후 가중값을 산출하였다. 장애등급 4-7급 층은 설계가중치 산출시 전체 권역을 기준으로 산출하고, 8-9급의 재활서비스 미이용 층은 권역별 표본 수가 2이상 되도록 권역 범주를 확대하여 산출하였다.

〈표 III-6〉 산재보험패널조사 : 권역별 층 통합

구분	4-7급					8-9급				
	미이용	⇒	미이용	이용	계	미이용	⇒	미이용	이용	계
서울	0	통합	7	10	10	2	통합	2	21	23
강원	0			3	3	0		6	6	6
부산	1			7	8	2		9	11	11
대구/경북	1			9	10	2		6	14	16
울산/경남	1			8	9	2		13	15	15
경기	2			16	18	9		9	28	37
인천	1			5	6	2		2	10	12
전라권	0			8	8	1		3	13	14
충청권	1			10	11	2		16	18	18
계	7		7	76	82	22		22	130	152

주) 제4차 산재보험패널조사 기초분석 보고서(2017)에서 발췌하여 재구성함

설계가중치의 합계는 모집단 구성비와 다를 수 있는데, 이는 사후층화하여 보정하였다.⁸⁰⁾ 사후층화보정은 표본추출 단계에서 반영한 내재적층화변수인 성(s), 연령대(a), 요양기간(i) 변수를 활용하였다. 사후층화 보정계수는 다음과 같이 성, 연령대, 요양기간의 모집단 빈도수와 이들의 설계가중 빈도수 비로 산출한다.

80) 표본추출단계에서 계통추출이 반영되어 설계가중치의 합계는 모집단 구성비와 다를 수 있다.

$$w_{k,g,j,s,a,i,l} = w_{k,g,j,l} \times BF_{s,a,i}$$

, 여기서 $BF_{s,a,i} = \frac{N_{s,a,i}}{\sum w_{s,a,i}}$ 이다. 사후층화 보정을 실시하면서 기존의 설계가

중값은 변하게 된다. 따라서 사후층화보정으로 산출된 층화변수들의 가중값 합계와 모집단 빈도는 불일치 할 수 있다. 레이킹 비 조정을 통해 가중값을 재보정하여 최종가중치를 산출함으로써 가중치가 적용된 표본 빈도와 모집단 규모 차이가 줄어들 수 있도록 하였다. 이때 레이킹 비 보정은 권역×장애유형×재활서비스 이용여부별 가중치 합과 성×연령별×요양기간별 가중값 합이 각 모집단 합으로 수렴할 때까지 반복한다.

〈표 Ⅲ-7〉 산재보험패널조사 : 표본할당 및 조사 완료 표본 수

(단위 : 명)

구분		표본 할당	조사 완료 표본 수	구분		표본 할당	조사 완료 표본 수
권역별	서울	278	278		13~14급	570	570
	부산	177	177		무장애자	350	350
	대구/경북	215	215	재활서비스 이용여부별	이용	1,037	1,037
	강원	77	77		이용안함	963	963
	울산/경남	237	237	성별	남자	1,685	1,686
	경기	449	449		여자	315	314
	인천	156	156	연령대	20대 이하	178	118
	전라권	209	209		30대	325	295
	충청권	202	202		40대	510	522
장애등급별	1~3급	30	30		50대	649	705
	4~7급	83	83		60대 이상	338	360
	8~9급	152	152	계		2,000	2,000
	10~12급	815	815				

주 1) 표본할당결과와 조사완료 표본 수는 제4차 산재보험패널조사 기초분석 보고서(2017) 에서 제시하고 있는 표를 발췌하여 재구성함

2) 조사완료표본은 1차 조사(2013) 결과 기준임

3) 자료 : 제4차 산재보험패널조사 기초분석 보고서(2017)



(3) 시사점

산재보험패널조사는 표본설계 단계에서 층화변수로 활용한 장애등급, 권역, 재활서비스 이용여부를 적용하고, 주요 결과에 영향을 미칠 수 있는 변수(성별, 연령대, 요양기간)를 표본추출단계에서 활용하였다. 그리고 가중치 산출 단계에서 장애등급, 권역, 재활서비스 이용여부, 성별, 연령대, 요양기간 변수를 활용하여 최종가중치를 산출하였다. 가중치 산출 과정에서 장애등급별 재활서비스 미이용자 층을 권역별로 통합하여 설계가중치를 산출하였는데, 이는 해당 층의 표본규모가 2미만으로 분산을 산출할 수 없었기 때문이다. 또한 이렇게 산출된 가중치의 합계와 모집단 구성비가 다를 수 있어, 성별, 연령대, 요양기간 변수를 사후층화보정 단계에서 활용하였다. 그리고 모집단 규모와 층화변수와 사후층화변수에 대한 표본가중값 합계와의 차이를 줄이고자 레이킹 추정법을 이용하여 가중값을 재보정 하였다.

장애인살 패널조사 역시 표본규모가 2미만인 층(0인 경우 포함) 이 나타나는 경우 분산을 구할 수 없기 때문에, 가중치 산출시 활용하고자 하는 변수별 세부 층분포를 파악할 필요가 있다. 그리고 해당 층은 유사한 층과 통합을 검토해야 한다. 또한 층화변수 이외 주요 결과에 영향을 줄 수 있는 변수들은 사후보정 등 가중치 보정 단계에서 고려할 수 있다. 가중치 보정을 거치며 변화하는 설계가중값으로 인해 모집단 구성비와 표본가중치와의 차이는 레이킹 비 조정을 통해 보정하는 것을 검토할 수 있을 것이다.

5) 한국·아동청소년패널조사⁸¹⁾

(1) 표본추출

연령코호트로 조사설계 되어 있는 한국아동·청소년패널조사의 코호트별 표본 크기는 각각 2,691명, 2,791명, 2,716명이다. 표본추출은 학교를 추출한 후 추출된 학교 내에서 학급을 추출하는 층화 2단 집락 추출법을 적용하고 있다. 따라서 학교가 1차 표본추출단위이며, 2차 표본추출단위가 학급이 된다.

81) 다음의 이어지는 내용은 한국아동·청소년패널조사 2010과 관련한 보고서를 참조하여 내용을 재구성 하였다. (출처) 김지경 외(2010). 한국아동·청소년패널조사2010I, 한국청소년정책연구원.; 이경상 외 (2011). 한국아동·청소년패널조사 2010 II 사업보고서, 한국청소년정책연구원.; 이경상 외 (2012). 한국아동·청소년패널조사 2010 III 사업보고서, 한국아동·청소년패널조사(KCYPS) 사업개요(2014), 제4회 한국아동·청소년패널학술대회. 한국청소년정책연구원. 한국아동청소년패널조사 통계정보보고서 (2017), 한국청소년정책연구원.

표본학교는 16개 시도별⁸²⁾, 도시규모별⁸³⁾ 층화하여 추출하였다. 이때 각 시도에 따른 층별 배분은 학급수를 기준으로 비례배분 하였다.⁸⁴⁾ 각 층내에서는 학교 주소를 기준으로 학교 목록을 정렬한 후, 층별 할당된 표본 크기만큼의 학교⁸⁵⁾를 확률비례계통 추출(크기변수 : 학급 수)하였다. 추출된 표본학교의 학생 조사는 추출된 표본학교 내에서 학급을 랜덤하게 추출하여 조사하였다.⁸⁶⁾ 층별 표본크기는 다음과 같다.

〈표 III-8〉 한국·아동청소년패널조사 : 층별 표본 크기

(단위 : 명)

시도	할당된 표본학교 수(목표)			조사완료 학교 수					
	초1	초4	중1	초1	학생 수	초4	학생 수	중1	학생 수
서울	12	11	8	12	327	11	321	8	280
부산	6	6	5	6	165	6	164	5	159
대구	5	5	4	5	149	5	152	4	161
인천	6	5	5	6	163	5	134	5	187
광주	5	5	4	5	148	5	141	4	135
대전	5	5	4	5	149	5	137	4	130
울산	4	4	4	4	126	4	117	4	138
경기	15	15	10	15	453	15	477	10	385
강원	4	4	4	4	113	4	127	4	129
충북	5	5	4	5	163	5	154	4	147
충남	5	5	4	5	145	5	142	4	112
전북	5	5	4	5	135	5	154	4	125
전남	5	5	4	5	138	5	126	4	135
경북	6	5	5	6	181	5	138	5	172
경남	6	6	5	6	18	6	182	5	166
제주	4	4	4	4	118	4	125	4	155
계	98	95	78	98	2,691	95	2,791	78	2,716

주) 한국아동청소년패널조사 2010 I, 한국아동청소년패널조사 2010 II, 한국아동청소년패널조사 통계정보보고서(2017).

82) 각 시도로 통계를 생산하지는 않지만 시도별 특성이 뚜렷하고 조사된 데이터를 이용하여 권역별로 특성분석이나 통계를 생산할 수 있도록 최소 표본으로 표본학교를 3개씩 우선할당한 후에 학생 수를 기준으로 비례배분 하였다. (출처) 한국아동·청소년패널조사 2010 I. 한국청소년정책연구원.

83) 도시규모는 대도시, 중소도시, 군지역으로 층화하였다.

84) 특히 중학교 1학년 코호트의 경우에는 남녀공학과 여학교 및 남자학교로 구분하여 할당 한 후 표본학교를 추출 함으로서 표본학생들의 성별 분포가 모집단의 분포와 유사할 수 있도록 하였다.

85) 추출한 표본학교의 학급 수와 학생 수가 조건을 충족(학급 수 2개 이상, 학생수 50명 이상)하지 않을 경우에는 바로 인접한 학교를 표본학교로 대체 하였다.

86) 표본학교는 표본학교의 특성, 소재지를 확인하고 학급수와 학급별 학생 수에 대한 정보를 확인한 뒤 랜덤으로 표본학급을 선정하였다.



(2) 가중치 산출 과정(및 추론)⁸⁷⁾ : 1차년도

한국아동·청소년패널조사의 최종가중치는 설계가중치, 무응답 조정, 사후층화조정을 거쳐 작성되고 있다. 종단면 가중치는 본 연구에 포함되지 않는 바, 본 절에서는 한국아동·청소년패널조사의 횡단면 가중치 부여 방법에 대해서 중점적으로 살펴보고자 한다.

한국아동·청소년패널조사는 1차 추출단위인 학교를 추출 하고, 추출된 표본학교에서 학급을 랜덤하게 선정한 후 학생을 조사하고 있다. 따라서 표본추출률은 $f_{hij} = n_h \times (l_h / M_h)$ 으로, 여기서 l_h 는 h 층에서 추출된 학교의 수, M_{hi} 는 h 층의 i 번째 학교의 학생 수, n_{hi} 는 추출된 표본학교에서 조사한 학생 수 이다. 이로부터 표본추출률의 역수인 설계가중치는 다음과 같이 정의된다.

$$W_{1hij} = \frac{1}{f_{hij}} = \frac{M_h}{n_h \times l_h}$$

, 여기서 $f_{hij} = n_h \times \frac{l_h}{M_h}$, $M_h = \sum_{i=1}^{K_h} M_{hi}$ 를 의미한다. 이때 K_h 는 h 층에 속한 학교의 수, M_{hi} 는 h 층의 i 번째 학교의 학생 수를 나타낸다.

무응답 조정 가중치는 표본으로 선정된 조사 단위 중 응답거부 등으로 인해 조사 완료 되지 않은 경우나 조사내용이 부실하여 무응답 처리한 경우의 조사단위를 조정해준 가중치이다. 무응답 조정 가중치는 다음과 같이 계산된다. 여기서 r_{hij} 은 완전 응답한 표본크기, n_{hij} 은 추출된 표본크기를 의미한다.

$$W_{2hij} = W_{1hij} \times \frac{n_{hij}}{r_{hij}}$$

최종가중치는 무응답 조정 후 16개 광역시도×성별에 대한 사후층화를 실시하여 작성되었으며, 다음과 같이 정의된다.

$$W_{fij} = W_{2hij} \times BF_{hij}$$

87) 다음의 이어지는 내용은 한국아동·청소년패널조사 2010과 관련한 보고서를 참조하여 내용을 재구성 하였다(이 부분은 강현철 교수(호서대학교 정보통계학과)에 의해 수행됨)

,여기서 BF_{hij} 는 사후층화 보정계수로 다음과 같이 정의된다. PS_{hij} 는 외부보조 정보에 대한 셀 빈도수 이다.

$$BF_{hij} = \frac{PS_{hij}}{\sum W_{1hij} W_{2hij}}.$$

(3) 시사점

한국아동·청소년패널조사 2010은 표본추출방법으로 층화2단집락추출 방법을 적용하고 있으며, 최종가중치는 성별 정보를 16개 광역시도×성별로 사후층화하여 작성하고 있다. 이는 한국아동·청소년패널조사 2010에서 표본추출틀로 활용한 2009년 전국 학교일람표에 학교에 관한 사항만이 포함되어 있고, 학교내 개별 학생의 성별과 같은 학생 관련 정보가 포함되어 있지 않았기 때문인 것으로 추측된다.⁸⁸⁾

사후층화 추정량은 모든 추출단위가 층화에 필요한 모든 변수 정보에 대하여 표본추출틀내 완벽히 가지고 있지 않은 경우, 주변 합과 같은 부분적인 보조정보만을 활용하는 대표적인 추정량 이다. 표본설계단계에서 반영하지 못한 보조정보를 추정 단계에서 활용하는 경우 비편향 HT 추정량보다 효율적인 추정량을 정의할 수 있다. 따라서 장애인삶 패널조사 가중치 작성시에도 연구변수에 영향을 미치는 변수이지만 표본설계 단계에서 여러 이유로 층화변수로 고려되지 않은 변수를 사후층화할 필요가 있는지 검토할 필요가 있다.

88) 한국아동·청소년패널조사 2010의 모집단은 2010년 현재 초등학교 1학년과 4학년, 그리고 중학교 1학년에 재학 중인 모든 학생과 그들의 보호자이다. 표본추출틀은 교육과학기술부에서 제공하는 2009년 전국 학교일람표에 수록된 초등학교와 중학교의 학교리스트이다.



6) 여성가족패널조사⁸⁹⁾⁹⁰⁾

(1) 표본추출

여성가족패널조사의 전체 표본크기는 8,500가구이다.⁹¹⁾ 표본추출틀로는 2005년 인구주택총조사의 260,000개 조사구 중 10%인 26,000개 일반조사구를 사용하고 있다. 따라서 표본추출은 1단계로 일반조사구를 추출한 후, 2단계로 적격가구와 적격가구를 추출하는 방법을 택하고 있다. 따라서 1차 표본추출단위(Primary Sampling Unit: PSU)는 2005년 인구주택총조사의 10% 일반조사구⁹²⁾이다. 조사구 추출을 위해 일반조사구를 도시화 수준, 산업별 종사자 수 비중, 주거형태별 가구 비중, 가구원수별 가구분포, 가구주 연령, 가구 주 성별 변수에 따라 시도별로 층화하였으며, 지역별 가구수에 따라 제곱근 배분하여 최종 1,700개⁹³⁾ 조사구를 추출하였다. 이때 조사구내 가구 수 규모를 크기변수로 확률비례계통추출 하였다.

2차 표본추출단위(Secondary Sampling Unit: SSU)는 PSU로 선정된 조사구에 거주하는 가구로, 조사가구는 계통추출을 적용하여 선정하였다. 최종 목표 표본인 여성 가구원 조사를 위해 1개 표본 조사구에서 5가구를 계통 추출하여 최종 8,500가구를 선정⁹⁴⁾ 하였다.

(2) 가중치 산출 과정(및 추론) : 1차년도

2007년 첫 조사가 실시된 여성가족패널조사는 횡단면 가중치와 종단면 가중치를 부여하고 있다.⁹⁵⁾ 그러나 본 연구는 장애인살 패널조사 1차년도 조사 결과 분석

89) 다음 이어지는 내용은 한국여성정책연구원의 여성가족패널조사 보고서와 통계청의 통계설명자료에서 제공하는 내용을 바탕으로 내용을 재구성 하였다. (출처) 「2018년 여성가족패널조사」, 「2007년 여성가족패널조사」, 「2008년 여성가족패널조사」, 「여성가족패널조사의 이용자용 통계정보보고서」

90) 여성가족패널은 여성에 대한 장기적인 추적조사를 통하여 급격한 변화의 중심에 있는 여성과 가족, 그리고 그 속에서의 여성노동과 일상의 변화실태를 가시화할 수 있는 학문적, 정책적 기초자료를 제공하고, 여성을 둘러싼 가족 및 사회적 배경에 대한 동태적 정보와 양성평등정책의 기초자료를 제공하는데 목적을 두고 있다. 이를 통해 변화하는 사회구조 속에서 여성의 의식이나 행위양식에 영향을 미치는 다양한 요인을 광범위하게 수집, 분석할 수 있다. (출처) 여성가족패널조사의 이용자용 통계정보보고서.

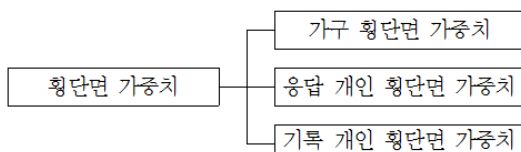
91) 여성가족패널은 2007년 조사 시작 이래로 2018년 7차 웨이브 조사까지 완료 되었다.

92) 일반조사구 표본추출틀에는 시설단위 조사구 및 제주도 이외의 섬지역 조사구는 제외된다.

93) 여성가족패널은 300개의 예비 조사구를 동일한 방법으로 별도 추출하였는데, 이는 조사구의 소멸 또는 개발 등의 사유로 조사 불가능한 표본 조사구 발생시 활용하기 위함이었다.

94) 선정된 조사대상 가구에 19세 이상 64세 이하의 여성가구원이 한명도 없는 경우에는 차순위에 있는 가구를 조사대상 가구로 선정하고, 차순위 가구에도 19세 이상 64세 이하의 여성가구원이 한 명도 없는 경우에는 차차순위에 있는 가구를 조사대상 가구로 선정하였다. 최종응답자는 표본조사구 내에 거주하는 표본가구의 가원 중 19세 이상 64세 이하의 여성을 모두 조사대상으로 하였다.

에 필요한 가중치 작성을 목적으로 하며, 종단면 가중치는 연구의 범위를 벗어난다. 이에 대해서는 추가적인 연구를 필요로 한다. 1차년도 조사의 경우 횡단면과 종단면의 구분이 없어 일반적으로 횡단면 조사에서의 가중치 부여 방법을 사용하게 된다. 여성가족패널조사의 횡단면 가중치로는 개인가중치와 가구단위 분석을 실시할 때 사용하는 가구가중치가 있다. 개인가중치의 경우에는 적격여성가구원 외에도 가구에 존재하는 모든 가구원에게 부여한다.



[그림 III-1] 여성가족패널 가중치

- 주 1) 여성가족패널은 가구가중치, 개인가중치를 부여하고 있으며, 개인가중치는 응답가구원가치, 기록가구원가중치로 나뉜다.
2) 응답개인가중치는 가구에서 응답한 적격여성가구원을 단위로 분석할 때 사용하는 가중치이다.
3) 기록가구원가중치는 가구에 존재하는 적격여성가구원을 단위로 분석할 때 사용할 수 있는 가중치이다.
4) 여성가족패널 3차 보고서(한국여성정책연구원)에서 발췌하여 재구성하였다.

추출단위는 가구이며, 개인과 가구를 조사단위로 하여 횡단면분석을 위해 개인 및 가구단위 가중치를 필요로 한다. 여성가족패널조사는 개인조사임에도 불구하고 개인을 추출할 수 있는 표본추출틀의 부재로 가구단위 표본을 뽑을 수 밖에 없다.⁹⁶⁾ 따라서 개인이나 가구가중치 산출시 표본 추출단위인 가구의 추출을 기초로 적격대상가구원의 응답확률을 고려하였다. 따라서 본 절에서는 여성가족패널조사의 횡단면 가구 가중치 부여 방법에 대해서 중점적으로 살펴보고자 한다.

여성가족패널조사의 가중치는 표본추출확률, 응답률보정, 사후보정 등의 과정을 거쳐 작성되고 있다. <표 II-7>과 같이 전체 표본크기 8,500가구 목표표 실제 9,084가구에 대해 조사완료 되었다. 따라서 조사구 추출확률에 응답률을 보정한 가중치를 기본가중치⁹⁷⁾로 부여하였다. 다시 말해 기본가중치는 실제 조사완료된 9,084가구에 부여되었다.

95) 여성가족패널조사의 최종분석 단위는 가구단위와 개인단위 모두이며, 다시 개인단위는 기록개인인 응답개인인
로 구분한다. 기록개인이란 개인용 설문지의 응답대상은 아니지만 가구원기록지에 포함되어 있는 개인들을 의
미한다. 기록개인인 응답 개인 모두에 대해 종단면과 횡단면 가중치를 부여한다. 패널가구에 대해서는 시간의
흐름에 따라 가구원의 전입과 전출에 따라 가구의 개념이 약화됨으로 인해 종단면 가중치보다는 횡단면 가중치
만으로 고려한다.

96) 「여성가족패널조사」통계정보보고서. 한국여성정책연구원. (출처) 통계청 통계설명자료.

97) 일반조사구내 거주 가구 응답률에 체계적인 차이가 없다는 가정하에서 기본가중치를 구하였다.



〈표 III-9〉 여성가족패널 1차 조사 목표 및 조사완료 표본 수

(단위: 개, 명)

구분	조사구	가구	여성개인
목표표본	1,700	8,500	10,000
성공표본	1,771	9,084	10,013
목표대비	104.2%	106.9%	100.1%

출처) 2008년 여성가족패널조사, 한국여성정책연구원.

〈표 III-10〉 여성가족패널 : 표본추출 결과 및 1차년도 조사완료 표본 수

(단위 : 가구)

지역	모집단 가구 수②	일반조사구수②	추출가구수 비율(%)③	추출가구 수	조사완료 수②
서울	3,312,858	210	12.32	1048	1128
부산	1,186,890	125	7.38	627	676
대구	814,886	104	6.11	520	476
인천	823,579	104	6.14	522	529
광주	460,300	78	4.59	390	433
대전	479,318	80	4.69	398	428
울산	338,991	67	3.94	335	371
경기	3,331,792	210	12.36	1051	1040
강원	520,039	83	4.88	415	447
충북	505,722	82	4.82	409	445
충남	660,715	94	5.5	468	494
전북	620,104	91	5.33	453	493
전남	666,736	94	5.53	470	566
경북	941,871	112	6.57	559	617
경남	1,056,589	118	6.96	592	676
제주	179,197	49	2.87	244	265
계	15,899,587	1,700	100	8,500	9,084

주 1) 「2007년 여성가족패널조사」, 「2008년 여성가족패널조사」에서 관련내용을 발췌하여 재구성함

2) 모집단 가구 수, 일반조사구수, 조사완료 수 : 「2008년 여성가족패널조사」

3) 추출가구수 비율 : 「2007년 여성가족패널조사」

기본가중치는 다음과 같이 정의되며, 여기서 일반조사구 추출확률은 시도별 모집단 가구 수에 따라 제공된 배분한 결과에 따라 계산된다.

$$\text{기본가중치} = \text{일반조사구 추출 확률} \times \frac{\text{적격응답가구}}{\text{적격가구}}$$

최종가중치는 기본가중치를 이상가중치와 벤치마킹조정 후 작성된 것으로 판단된다. 이상가중치 조정은 가중치의 지나친 변동을 제어하기 위한 목적으로 기본가중치 분석 결과(평균가중치는 150.25, 표준편차 73.30, 최소값 16.15, 최대값 627.98)를 토대로 평균값의 3배 이상인 가중치에 대해서 실시되었다. 이상가중치 조정된 가중치는 벤치마킹조정을 실시하여, 시도별 가중치 총합은 시도별 적격가구 수와 일치할 수 있도록 하였다. 이 때 벤치마킹자료로는 표본추출에 사용된 2,000개⁹⁸⁾ 일반조사구에 나타난 자료를 사용하였다.

〈표 III-11〉 여성가족패널 가중치 분포

구분	기본가중치	가구가중치
가구 수	9,084	
평균	150.25	1,385.70
표준편차	73.30	788.83
최소값	16.15	210.34
최대값	627.98	4,137.03

주) 보고서 내용을 토대로 내용을 재구성함 (출처) 2008년 여성가족패널조사, 한국여성정책연구원.

(3) 시사점

목표 표본 수와 조사완료 표본 수와의 표본크기에 차이가 발생하는 경우 추출확률에 응답률을 보정하여 가중치를 산정한다. 이상가중치 조정으로 층별 가중치 총합이 일치하지 않을 수 있다. 조정된 가중치는 벤치마킹조정 등 적절한 방법을 통해 가중치를 조정하여 층별 총합이 일치할 수 있도록 해준다.

장애인실태 패널조사 역시 목표 표본수와 조사완료 표본 수간의 규모 차이가 발생

98) 여성가족패널조사는 1,700개 조사구와 조사구의 소멸 또는 개발 등의 사유로 조사 불가능한 표본조사구를 대체하기 위해 300개의 예비표본 조사구를 동일한 방식으로 별도 추출하였다.



하였다.⁹⁹⁾ 이에 설계가중치에 응답률을 보정하여 기본가중치를 산정하는 것을 검토해볼 수 있다. 조정된 가중치의 층별 총합과 모집단 합계가 일치하지 않을 수 있다. 이는 적절한 방법을 검토하여 가중치 합계와 모집단 총계가 일치하도록 해주도록 한다.

99) 이에 관해서는「IV.. 가중치 부여 방안. 2. 장애인살 패널조사 구축 현황」에서 자세히 기술하였다.

3. 관련 자료 분석

본 절에서는 가중치 설계시 검토가 필요한 주요 사항을 파악하고자 관련 규칙과 매뉴얼 분석을 실시하고자 한다. 이를 통해 고려가 필요한 사항을 사전에 파악하여 가중치 부여 방향을 설정하는데 도움을 얻고자 한다.

관련 규칙·매뉴얼로는 통계조정업무매뉴얼(통계청), 표본설계 및 관리지침(통계청 예규 제198호 2017.3.10.)¹⁰⁰⁾, 자체통계품질진단 매뉴얼(조사통계), 국가통계품질진단 매뉴얼(조사통계)이 있다. 국가통계는 통계법 제11조(자체통계품질진단) 및 제9조(정기통계품질진단)에 의거 통계품질진단을 수행해야하기 때문에 관련 자료를 함께 살펴보는 것이 필요할 것으로 판단된다.

100) 표본설계 및 관리지침(통계청 예규 제198호 2017.3.10.)

- 제9조(표본오차의 작성) ① 표본조사 결과를 공표할 때는 표본조사의 정확성을 평가하고 이용자에게 정확한 통계 정보를 제공하기 위하여 표본오차를 작성하여야 한다.
- ② 표본오차는 조사결과 공표 1개월 전까지 표본설계 소관부서에 요청하여야 한다. 다만, 월별조사의 경우는 공표 1주일 전까지 요청할 수 있다.
- ③ 제2항에 요청한 내용에 대하여 표본설계 소관부서는 공표하기 전 최대한 빠른 시간 내에 조사 담당부서에 제공하여야 한다.
- 제10조(표본오차의 기준) ① 표본조사 결과를 공표할 때에는 표본조사의 정확성을 측정하기 위해 상대표본오차(RSE)를 보고서 및 보도자료 등에 수록하는 것을 원칙으로 한다. 다만, 보도자료는 주요 결과표에 대하여서만 작성할 수 있다.
- ② 정확성 측정은 상대표본오차(RSE)와 최소 표본규모를 기준으로 측정하며 통계 공표에 필요한 평가 기준을 따르는 것을 원칙으로 한다. 다만, 모집단수가 10개 미만이거나 발견률이 낮은 특수한 경우 등에는 최소 표본규모의 기준을 적용하지 않을 수 있다. 표본조사 정확도 평가 기준은 다음과 같다.
- ③ 조사결과를 공표할 때에는 정확성 평가 기준에 맞게 유의사항을 표기하여 자료 이용자로부터 신뢰도가 낮은 통계결과를 이용하는데 주의하도록 알리는 것을 원칙으로 한다. 표본조사 유형별 유의사항 표기 및 해석기준은 다음과 같다.

[표1] 가구 통계 유의사항 표기 및 해석

	RSE < 25%		25% ≤ RSE < 50%	50% ≤ RSE
	n ≥ 10	n < 10		
표기	표기 없음	*	*	**
해석	신뢰도 높음	신뢰도 다소 낮음	신뢰도 다소 낮음	신뢰도 낮음

[표2] 농어업사업체 통계 유의사항 표기 및 해석

	RSE < 30%		30% ≤ RSE < 60%	60% ≤ RSE
	n ≥ 10	n < 10		
표기	표기 없음	*	*	**
해석	신뢰도 높음	신뢰도 다소 낮음	신뢰도 다소 낮음	신뢰도 낮음

- ④ 보존용 자료(MD)를 이용하여 생산된 가구 통계는 공표수준 보다 세분된 경우가 많고, 자료 활용성 및 이용자 혼란을 방지하기 위해 농어업사업체 통계 기준을 사용할 수 있다.
- ⑤ 조사의 특성에 따라 별도의 기준 적용이 필요한 경우에는 다른 기준을 적용할 수 있다. <개정 2017.3.10>
- ⑥ 통계작성 승인 및 통계품질진단 관련 부서에서는 통계법에 의한 통계작성기관에서 표본조사 결과를 공표할 때에 위의 기준을 활용하도록 권고하여야 한다. <개정 2017.3.10>



표본조사에서는 추정시 편향 없는 결과를 얻을 수 있도록 가중치 조정을 실시하는데, 통계조정업무 매뉴얼에서는 가중치 조정 및 통계 결과 분석시 검토가 필요한 주요 요건을 제시하고 있다. 이를 기초로 가중치 설계시 필요한 요소를 점검하고자 한다.

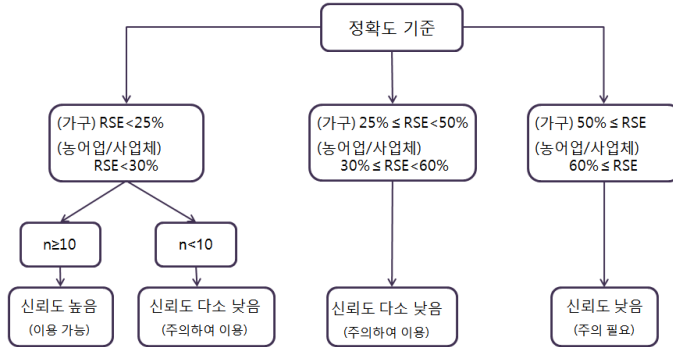
〈표 III-12〉 검토 내용 및 기준 (통계조정업무 매뉴얼)

구분	검토 내용	주요 검토 기준
추정식 (평균, 총합, 비율, 분산)	<ul style="list-style-type: none"> 표본추출방식에 의한 가중치를 적용하였는가? 무응답 및 사후층화 가중치를 고려하였는가? 가중치는 표본추출방식을 고려한 평균, 총합, 비율, 분산식을 채택하였는가? 복합설계의 경우 선형화 또는 반복분산 추정방법을 고려하였는가? 	<ul style="list-style-type: none"> 층별 표본배정한 경우, 가중치를 고려한 추정식을 적용해야 함 인구, 연령, 규모별 사업체수 등 모집단 정보를 이용하여 사후층화 가중치를 적용하여야 함 평균, 비율, 총합에 대한 추정식 명시 표준오차, 분산, 상대표준오차 등에 대한 추정식 명시
주요 항목별 공표 범위	<ul style="list-style-type: none"> 주요항목별 공표범위를 제시하였는가? 공표범위별 표본규모가 적절한가? 	<ul style="list-style-type: none"> 세분화된 통계표 작성 시 각 셀별 응답표 분수가 최소한 10~15 이내가 되어야 함 응답표분수가 적은 셀은 유사한 셀간 병합하여 통계표 작성해야 함 원하는 셀대로 통계표를 작성하기 위해서는 표본규모를 확대하여 조사하여야 함 표본설계 시 고려하지 않은 “분류/항목”에 대한 통계표 작성 시 대표성 검토 반드시 필요

출처)「통계조정업무 매뉴얼 (2015). 통계청」에서 발췌하여 내용을 재구성함

품질진단 시 에서는 통계추정 및 분석 파트가 가중치 조정¹⁰¹⁾에 관한 사항을 포함하고 있다. 주요 검토 사항은 가중치 조정 및 조정 방법, 주요항목의 상대표준오차, 통계결과의 정확성을 위해 제시하는 값에 대한 것이다.¹⁰²⁾ 특히 주요항목의 추정값에 대한 정확성을 평가하는 수치로 상대표준오차를 제시하는 경우에 대한 기준치를 함께 제시하고 있다. 상대표준오차를 활용한 추정치 평가 기준은 다음과 같다.

101) 2019년 국가통계 품질관리 매뉴얼 : 정기통계품질진단, 통계청.



주) n은 표본규모
출처) 표본설계 및 관리지침(통계청 예규 제198호 2017.3.10.)

[그림 III-2] 표본조사 정확도 기준

표본조사의 조사결과를 통계표를 산출할 때는 표본오차 또는 상대표준오차의 제시를 필요로 함을 알 수 있다(「표본설계 및 관리지침」, 통계청 예규 제198호 2017.3.10.). 이는 향후 장애인삶 패널조사의 조사결과를 활용한 보고서 작성시 참고가 필요하다.

VII. 통계추정 및 분석

작성지침

1. 통계추정

1-1. 가중치 조정

- 무응답 가중치 조정, 사후가중치 조정(Calibration) 등을 하는 경우 각각에 대하여 필요성, 구체적인 조정과정 및 방법, 주요 가중치 분포 등을 기술한다.

1-2. 이상치 식별 및 처리

- 이상치 처리를 하는 경우 이상치의 기준, 식별 및 처리방법, 처리결과 등에 대하여 기술한다.

1-3. 통계추정 산식 및 내용

- 통계추정 산식과 추정내용(예: 총계, 평균, 중위수, 비율 등)에 대하여 기술한다.

작성지침

2. 표집오차 추정 방법 및 결과(표본조사)

- 분산, 표준오차 등의 추정 방법에 대하여 기술한다.
- 주요 항목들에 대한 변동계수, 표준오차, 신뢰구간 등을 정리해 제시하고, 이들 오차의 특성과 이용 시 고려사항을 기술한다.
- 마이크로데이터 이용자가 스스로 표집오차를 계산할 수 있도록 관련 방법을 제공하는 경우 이에 대하여도 사용방법과 함께 기술한다.

102) 통계청. (2016). 자체통계품질진단 매뉴얼 : 조사통계 부문.



[표본설계 및 관리지침] 통계청 예규 개정 2017.3.10.

제9조(표본오차의 작성) ① 표본조사 결과를 공표할 때는 표본조사의 정확성을 평가하고 이용자에게 정확한 통계 정보를 제공하기 위하여 표본오차를 작성하여야 한다.

제10조(표본오차의 기준) ① 표본조사 결과를 공표할 때에는 표본조사의 정확성을 측정하기 위해 상대 표준오차(RSE)를 보고서 및 보도자료 등에 수록하는 것을 원칙으로 한다. 다만, 보도자료는 주요 결과표에 대하여서만 작성할 수 있다.

③ 조사결과를 공표할 때에는 정확성 평가 기준에 맞게 유의사항을 표기하여 자료 이용자로 하여금 신뢰도가 낮은 통계결과를 이용하는데 주의하도록 알리는 것을 원칙으로 한다.

[그림 III-3] 표본설계 및 관리지침 제9조 및 제10조

출처) 「표본설계 및 관리지침」, 통계청 예규 제198호 「2018년 자체통계품질진단 매뉴얼(조사통계)」, 통계청,에서 발췌함.

앞서 검토한 사항을 종합해보면, 가중치 조정이 표본조사 결과를 통해 편향 없는 추정치를 얻기 위한 주요 사항임을 알 수 있다. 따라서 장애인삶 패널조사의 가중치 작성시에도 관련 규칙·매뉴얼상의 검토 요건을 토대로 진행하는 것이 필요할 것으로 판단된다.

4. 시사점 및 주요고려 사항

본 절에서는 논의 내용을 토대로 장애인삶 패널조사 가중치 설계를 위한 주요 시사점을 도출한다. 이를 통해 가중치 작성을 위한 검토 필요 사항을 사전에 발굴함으로써 설계에 도움을 얻고자 한다.

복합표본설계는 층화, 집락, 불균등 선택 확률, 다단계 추출, 계통 추출 등이 적용되기 때문에 일반적인 추정 및 검정 방법으로는 적절한 통계분석을 할 수 없다. 따라서 일반적으로 표본조사에서는 불균등 추출확률, 응답률 보정으로 가중치를 부여한 뒤 추정치를 산출한다 (Kish, 1992; 류제복, 1993; 김규성, 2006). 가중치를 무시하고 통계분석을 할 경우, 추정치의 편향이 발생할 수 있으며, 이에 추정시에 반드시 가중치를 이용할 필요가 있다.

앞서 살펴본 조사들의 가중치 작성시 고려사항을 종합해 보면 일반적으로 복합표본조사의 최종가중치는 설계가중치, 무응답 조정, 사후조정 등의 과정을 거쳐 산출됨을 확인할 수 있으며, 통계의 공표단위, 향후 자료 활용성 등을 종합적으로 검토하여 최종가중치를 작성하고 있다.

1) 가중치 작성 방안 검토를 위한 주요 시사점

가중치 적용의 기본 아이디어는 비편향 추정이지만, 표본조사 수행시 발생하는 무응답, 표본추출틀의 결함 등을 극복하기 위해 추정식을 보정하는 수단으로도 가중치 조정이 실시된다. 우리는 이러한 가중치 조정을 통해 비편향 추정량을 얻을 수 있다 (김영원 외, 2012).

표본조사에서는 조사거절, 장기적 부재 등으로 응답대상자의 조사 자체에 대한 불응으로 발생하는 무응답으로 인하여 사전 계획된 표본 크기와 조사 완료 후 표본 크기간의 불일치가 발생할 수 있다. 이렇게 목표 표본 수와 조사완료 표본 수간에 차이가 발생하는 경우, 무응답으로 인한 결측 문제를 해결하기 위해 추출확률에 응답률을 보정하여 무응답 조정 가중치를 산출한다.

층화변수 이외 주요 결과에 영향을 줄 수 있는 변수들은 사후층화 등 가중치 보정 단계에서 고려할 수 있다. 표본설계단계에서 반영하지 못한 보조정보를 추정단계에서 활용하는 경우 비편향 HT 추정량보다 효율적인 추정량을 정의할 수 있다. 사후조정은 표본설계 단계에서 다양한 이유(주요 문항에 영향을 미칠 것으로 예상



되나 표본크기의 한계, 모집단 목록에의 정보 부족 등)로 고려할 수 없었던 변수에 대해서 조사 완료 후 이에 대한 모집단 정보가 부분적으로 확보되었을 경우 사후층화, 벤치마킹조정, 캘리브레이션 등을 통해 실시된다. 이때 활용 변수의 범위 결정은 통계의 실제 공표와 통계를 공표하진 않더라도 향후 자료 사용 측면을 함께 고려하여 가중치 조정 변수를 결정할 필요가 있다.

사후층은 표본설계 단계에서 고려되지 않았기에 표본설계시 고려하지 않은 사후층에 대해서는 표본이 조사되지 않았을 가능성이 높다. 또는 층의 표본규모가 2미만으로 분산을 산출할 수 없을 수 있다. 이러한 경우 층 통합(collapse)을 통해 해결할 수 있다. 층 통합을 검토 하는 경우에는 가중치 산출시 활용하고자 하는 변수별 세부 층 분포를 파악할 필요가 있으며, 이때 유사한 층과의 통합을 검토해야 한다. 즉, 모집단과 조사완료 결과를 사후층별로 분석하여 공표범위를 고려하여 층 통합 여부를 결정해야 한다.

표본설계에서 고려하지 않은 변수를 가중치 작성 단계에서 고려하거나, 층 통합 후 캘리브레이션(calibration) 가중치를 부여하게 되면 최종가중치 합계는 모집단 합계와 일치 하지 않을 수 있다. 층화변수와 사후층화변수의 가중합과 모집단 규모간의 차이를 줄이기 위해 레이킹 조정을 검토할 수 있다. 그리고 가중치를 이용하여 분석을 수행할 때에는 층 통합이 적용된 분석단위에 대해서는 주의해야 한다.

설계가중치 이후 가중치 작성 단계를 거치면서 지나치게 큰 가중치가 발생할 수 있는데, 이를 막기 위해서 지나치게 큰 부분을 절사하는 극단가중치 조정 등을 통해 최종가중치를 산출할 수 있다. 이때 고려할 사항은 극단가중치의 평가 및 결정은 값 자체보다는 초기가중치 대비 최종가중치의 비율로 검토하는 것이 바람직하다는 것이다. 이는 표본설계 및 무응답 모형에 기반해 산출된 초기가중치는 추정량의 근사 비편향성 및 일치성을 보장하며 따라서 단순히 최종 가중치의 크기 보다는 초기가중치 대비 최종가중치의 변동량이 최종가중치를 통해 작성되는 추정량의 성질에 미치는 영향을 판단하는 좋은 지표가 되기 때문이다 (Deville, Sarndal, and Sautory, 1993; 변종석 외, 2018에서 재인용).

장애인삶 패널조사의 1차년도 조사 자료에 대한 가중치는 횡단면가중치와 동일하다. 종단면가중치(longitudinal weights)의 작성은 향후 조사가 진행됨에 따라 필요하다.

2) 관련 자료에서의 주요 시사점

앞서 살펴본 자료들을 종합해보면, 통계조정업무 매뉴얼 및 자체통계품질진단(통계법 제9조)에서 가중치 조정시 주요 검토 사항을 확인할 수 있다. 검토 요건으로는 표본추출과정을 고려한 가중치, 무응답 및 사후층화 가중치의 고려, 복합표본 설계의 경우에는 선형화나 반복분산 추정방법에 대한 고려, 표본오차, 분산, 상대표준오차 등에 대한 추정식 등이다. 즉 통계 자료 분석을 위한 가중치 조정 및 조정 방법 등에 대한 사항을 주요하게 여감을 확인할 수 있다.

타(他)조사 가중치 조정 방법을 살펴보면, 일반적인 가중치 작성 단계의 큰 틀은 설계가중치, 무응답조정, 사후조정의 단계를 거치고 있으며, 조사 특성(공표단위, 벤치마킹 여부 등)을 고려하여 최종가중치 작성이 이루어지고 있다. 따라서 장애인 삶 패널조사에서도 최종가중치 산출을 위해서 사후조정 방안을 검토해볼 필요가 있다.

Ⅳ. 가중치 부여 방안

IV 가중치 부여 방안

1. 개요

앞서 살펴본 바와 같이 일반적으로 표본조사에서는 표본의 대표성을 확보할 수 있도록 불균등 추출확률, 응답률 보정하여 가중치를 부여한 후 추정치를 산출한다(Kish, 1992; 류제복, 1993; 김규성, 2006). 추정량의 효율성과 신뢰성 제고를 위해 표본에 적절한 가중치를 부여하는 것은 표본조사의 필수 요소이다. 일반적으로 추정단계에서 가중치를 이용하게 되면 모집단 특성치인 모수에 대한 비편향 추정량(unbiased estimator)을 얻을 수 있다(김재광, 2006). 표본가중치를 무시하고 통계분석을 수행할 경우 추정치는 심각한 편향이 발생할 수 있다. 이에 대규모 조사에서는 모수추정과정에서 반드시 표본가중치를 적용한 분석방법을 활용하여 추정량의 비편향성을 확보하고 있다.

복합표본조사의 최종 가중치는 일반적으로 표본추출확률을 이용한 설계 가중치(design weight) 외에도 단위 무응답 자료의 처리를 위한 무응답 가중치 조정, 사후가중치(외부 보조정보¹⁰³)와의 일치성을 위해 조정해 주는 calibration 등, 지나치게 큰 가중치를 판별하게 이를 처리하는 가중치 이상값 처리와 같은 요소들로 구성되어 있다(김재광, 2006).¹⁰⁴

따라서 장애인삶 패널조사의 가중치 산정 방안 검토 시에는 다음 사항을 고려하여 최종가중치를 제시하고자 한다. 첫째, 표본추출 과정에서 모집단으로부터 표본을 선정한 표본추출률이다. 둘째, 조사과정에서 표본들의 응답 정도이다. 이를 통해 본 장에서는 장애인삶 패널조사의 패널구축 및 1차년도 조사완료에 따른 결과의 분석에 필요한 가중치를 산출하고자 다양한 방안을 검토하여 최종가중치 작성을 위한 적절한 방안을 마련하고자 한다. 우선적으로 장애인삶 패널조사의 패널 구축 현황을 살펴보고, 가중치 산출과정에서 고려가 필요한 사항을 파악한다.

103) 일반적으로 보조정보를 이용함으로써 변수들의 표본분산을 줄일수 있고, 무응답(nonresponse)과 비 포괄성(non-coverage)에 의한 편향을 감소시킬 수 있다는 장점으로 인해 조사의 질을 개선할 수 있다.(출처) 손창균(2002). 완전응답과 단위무응답하에서의 사후중화와 보정의 연관성에 대한 연구, Journal of the Korean Data Analysis Society, 4(3), 297-305.

104) 김재광(2006), 표본조사의 가중치 작업에 대한 고찰, 한국조사연구학회 춘계학술대회 발표논문집.



2. 장애인삶 패널조사 구축 현황

1) 개요¹⁰⁵⁾

장애인삶 패널조사는 2015년 1월 1일부터 2017년 12월 31일 사이에 장애등록을 마친 전국의 장애인(장애인 거주시설에 거주하는 장애인 제외) 및 함께 거주하는 가구원을 조사대상으로 하고 있다. 이에 장애인삶 패널조사는 생산 통계의 정도(precision)를 높일 수 있는 방안을 마련하면서, 주어진 시간과 조사예산 등 현실적 제약을 고려한 표본추출을 실시하는 것을 목표로 표본설계를 실시하였으며, 표본설계 내역을 토대로 추출된 표본을 조사하여 패널 구축 및 1차년도 조사를 완료하였다.¹⁰⁶⁾ 조사완료된 장애인 표본은 6,121명이다.

장애인삶 패널조사의 목표모집단은 장애인복지법 제2조에서 정의하는 장애인으로 같은 법 제32조에 의거 보건복지부에 장애등록을 2015년 1월 1일부터 2017년 12월 31일 사이에 마친 등록장애인(단, 장애인거주시설 거주 장애인 제외)과 그들의 가구원이다. 조사모집단은 사망, 주소불명 등의 사유로 조사 불가능한 경우를 제외한 장애인 251,277명¹⁰⁷⁾과 그들의 가구원으로 정의하였다. 표본추출틀은 보건복지부의 등록장애인 DB 명부(장애인거주시설장애인, 사망 및 주소불명 등 조사

105) 본 절은「김현지, 김태용. (2018). 장애인삶패널조사 표본설계 연구. 한국장애인개발원.」에서 일부 발췌하여 요약·정리하였다.

106) 모집단 분석은 모집단의 특성 변수이면서 조사항목에 영향을 줄 수 있는 장애유형과 장애정도, 그리고 인구사회학적 특징인 성별, 연령, 지역을 중심으로 실시하여 표본설계를 위한 기초자료로 확보하였다. 모집단 분석 결과, 장애유형별 장애정도(중증, 경증) 및 각 장애유형별 성별, 연령별 분포가 상이한 것으로 나타났다. 이와 같이 장애유형별 모집단 수의 차이가 큰 경우 장애유형별 통계 작성에 필요한 충분한 표본 수가 확보되지 않을 수 있을 뿐만 아니라 추정 정도에도 영향을 미칠 수 있다. 따라서 표본설계에서는 이러한 특성을 고려하여 충분한 표본 수 확보를 위한 방안을 마련하고자 하였다. 이와 더불어 최종 표본에 특정 장애유형이나 연령대의 누락 없이 이들이 표본에 포함될 수 있도록 하고자 하였다. 장애인삶 패널조사를 위한 표본설계는 다음의 사항을 고려하였다. 첫째로 조사결과와 공표단위 이다. 장애유형별, 장애정도별, 그리고 성별에 대한 공표를 계획하고 있는바, 공표단위에 따른 모집단 대표성이 확보될 수 있도록 하였다. 둘째로 추정량의 정도를 높일 수 있는 적절한 표본배분을 고려하였다. 또한, 특정 장애유형의 누락 없이 모든 법정장애유형이 최종 표본에 충분히 포함될 수 있는 방안을 검토하였다. 마지막으로 조사예산과 조사일정을 고려하여 조사의 효율성을 높일 수 있는 표본추출방법 방안을 고려하였다.

이러한 표본설계를 통해 모든 법정장애유형 및 다양한 특성의 표본이 최종표본에 충분히 포함될 수 있으면서, 공표단위에 대한 모집단 대표성, 통계적으로 분석 가능한 수준의 표본크기를 확보하고, 조사의 효율성(조사예산 및 일정)을 높일 수 있는 방안을 마련하고자 하였다. (출처) 김현지, 김태용 (2018).「장애인삶패널조사 표본설계 연구」, 한국장애인개발원.

107) 본 조사의 모집단은 2015년 1월 1일부터 2017년 12월 31일 사이에 보건복지부에 장애등록을 마친 장애인은 총 258,016명이다. 따라서 2017년 12월 31일 기준 등록장애인은 2,545,637명(보건복지부 홈페이지 장애인 등록 현황 및 http://kosis.kr/statHtml/statHtml.do?orgId=117&tblId=DT_11761_N004, 접속: 2019.6.24.)과는 차이가 있다.

불가능자 제외)를 이용하였다.¹⁰⁸⁾ 여기에는 장애유형¹⁰⁹⁾, 장애정도¹¹⁰⁾, 지역, 성별, 나이 정보가 포함되어있다. 장애인삶 패널조사의 조사대상은 장애인 및 함께 거주하는 가구원으로, 장애인 표본크기는 전체 예산과 조사비용 등을 고려하여 결정되었다. 이는 전국 단위 통계의 95% 신뢰수준에서 단순임의추출의 모비율 추정을 가정할 경우 목표허용오차 1.27% 수준이다.

가중치 작성을 위해서는 먼저 기존의 표본설계 및 표본추출과정을 충분히 검토하고, 조사완료 현황을 파악하는 것이 필요하다. 따라서 본 장에서는 가중치 산정 방안에 필요한 자료를 마련하고자 장애인삶 패널조사의 표본추출결과 및 구축 현황을 살펴보고자 한다. 이를 통해 조사완료 되어 패널로 구축된 표본의 특성을 파악하여 가중치 산정을 위한 방안을 검토시 주요한 요소들을 탐색하고 기본 방향을 설정하고자 한다.

2) 표본추출¹¹¹⁾

장애인삶 패널조사는 이중추출(double sampling)을 적용하였으며, 이에 장애인 표본 추출에 앞서 읍면동 추출이 선행되었다. 일상의 읍면동은 희소장애인과 만 18세 이하 장애인이 최종 표본에서 누락되지 않도록 고려하였으며, 일상의 읍면동 내 거주하는 전체 장애인을 장애유형, 장애정도, 성별로 층화한 후 최종 표본인 장애인을 계통추출 하였다. 지역은 표본추출시 고려함으로서 비례배분하여 할당한 것과 동일한 효과를 가질 수 있도록 하였다. 최종 추출된 표본의 주요변수별 모집단 및 표본추출결과는 하단 표와 같다.

108) 본 패널조사의 조사목적 및 모집단 정의에 따라 2017년 12월 31일 기준 보건복지부의 등록장애인 중 2015년 1월 1일부터 2017년 12월 31일 사이에 장애등록을 마친 장애인 명부를 표본추출틀로 확보하였다.

109) 장애인복지법 제2조에서 정의한 15개 법정장애유형을 의미한다.

110) 본 패널조사를 위해 확보한 표본추출틀에는 장애정도(장애등급 1급~3급은 중증, 4급~6급은 경증)에 대한 정보가 포함되어 있다.

111) 본 절은「김현지, 김태용. (2018). 장애인패널조사 표본설계 연구. 한국장애인개발원.」에서 일부 발췌하여 요약·정리하였다.



〈표 IV-1〉 표본추출결과

(단위 : 명, %)

장애유형(15)	장애유형(7)	모집단	표본	
			빈도	비율(%)
지체	지체	49,532	935	16.12
시각	시각	21,557	732	12.62
청각	청각,언어	73,818	1,090	18.79
언어				
뇌병변	뇌병변	42,177	951	16.4
지적	지적,자폐성	23,852	538	9.28
자폐성				
정신	정신	8,894	321	5.53
신장	내부,안면	31,447	1,233	21.26
심장				
호흡기				
간				
안면				
장루,요루				
뇌전증				
계		251,277	5,800	100

(단위 : 명, %)

장애 정도	모집단	표본		성별	모집단	표본	
		빈도	비율(%)			빈도	비율(%)
중증	92,616	3,018	52.03	남자	142,540	3,189	56.73
경증	158,661	2,782	47.97	여자	108,737	2,611	43.27
계	251,277	5,800	100	계	251,277	5,800	100

장애인삶 패널조사는 표본대체를 허용하지 않는 것을 기본 원칙으로 삼았다. 다만 조사과정에서 다양한 이유로 부득이하게 발생할 수 있는 조사 불가능 상황을 고려하여, 초기 패널 구축 단계에서는 예외적으로 예비표본을 활용한 표본대체를

허용하였다.¹¹²⁾

3) 패널구축 및 1차년도 조사 현황

장애인삶 패널조사는 최종 6,121명¹¹³⁾의 장애인 표본을 패널로 구축하였다. 지역별로 조사완료 현황을 살펴보면 서울 1,117명(18.25%), 부산 503명(8.22%), 대구 329명(5.38%), 인천 358명(5.85%), 광주 213명(3.48%), 대전 192명(3.14%), 울산 152명(2.48%), 세종 35명(0.57%), 경기 1,492명(24.38%), 강원 167명(2.73%), 충북 161명(2.63%), 충남 220명(3.59%), 전북 191명(3.12%), 전남 209명(3.41%), 경북 310명(5.07%), 경남 379명(6.19%), 제주 93명(1.52%)이다.

〈표 IV-2〉 지역별 조사완료 현황

지역	조사완료		지역	조사완료	
	빈도	비율(%)		빈도	비율(%)
서울	1,117	18.249	강원	167	2.728
부산	503	8.218	충북	161	2.63
대구	329	5.375	충남	220	3.594
인천	358	5.849	전북	191	3.12
광주	213	3.48	전남	209	3.414
대전	192	3.137	경북	310	5.065
울산	152	2.483	경남	379	6.192
세종	35	0.572	제주	93	1.519
경기	1,492	24.375	계	6,121	100

장애유형별로는 지체 974명(15.91%), 시각 757명(12.37%), 청각·언어 1,141

112) 본 패널조사는 원칙적으로 대체표본을 허용하지 않는다. 하지만 조사과정에서 패널의 이사 및 장기부재, 또는 지속적이고 강력한 조사거절로 인한 무응답 및 기타 다양한 이유로 부득이하게 조사가 불가능한 경우가 발생하게 된다. 이러한 경우를 대비하여 예비표본을 추출하여, 표본대체를 할 수 있도록 하였다. 단 이러한 표본대체는 초기 패널 구축 단계에서만 적용하고, 이후 조사는 표본대체 하지 않는 것을 원칙으로 한다. 표본대체로 인한 비표본 오차를 줄이기 위하여 표본설계에 따라 각각의 원표본과 동일 속성을 가진 표본을 추가로 추출하였다.

113) 초기에 조사 대상으로 선정한 패널(원표본)이 다양한 이유로 조사에 참여를 하지 못한 경우, 동일한 속성을 가진 예비표본을 사전 원칙에 따라 차례대로 접촉 및 조사를 실시하였다. 다만, 실사 과정에서 강정거절, 부재 등의 이유로 무응답이 발생하여 대체가 이뤄지는 과정에서 복수 접촉으로 조사가 완료된 장애인이 패널 구축 시 포함되어 조사목표 인원보다 더 많은 인원이 구축되었다.



명(18.64%), 뇌병변 980명(16.01%), 지적·자폐성 567명(9.26%), 정신 328명(5.36%), 내부·안면 1,374명(22.45%)이다. 성별로 살펴보면 남자 3,389명(55.37%), 여자 2,732명(44.63%)이며, 장애정도별로는 중증 3,242명(52.97%), 경증 2,879명(47.03%)이다. 최종 구축 패넬은 다음과 같다.

〈표 IV-3〉 장애유형, 장애정도, 성별 조사완료 현황

(단위 : 명, %)

장애유형(15)	장애유형(7)	모집단	표본배분		조사완료	
			빈도	비율(%)	빈도	비율(%)
지체	지체	49,532	935	16.12	974	15.91
시각	시각	21,557	732	12.62	757	12.37
청각	청각,언어	73,818	1,090	18.79	1,141	18.64
언어						
뇌병변	뇌병변	42,177	951	16.4	980	16.01
지적	지적, 자폐성	23,852	538	9.28	567	9.26
자폐성						
정신	정신	8,894	321	5.53	328	5.36
신장	내부,안면	31,447	1,233	21.26	1,374	22.45
심장						
호흡기						
간						
안면						
장루,요루						
뇌전증						
계		251,277	5,800	100	6,121	100

(단위 : 명, %)

장애정도	모집단	표본배분		조사완료	
		빈도	비율(%)	빈도	비율(%)
중증	92,616	3,018	52.03	3,242	53
경증	158,661	2,782	47.97	2,879	47
계	251,277	5,800	100	6,121	100

(단위 : 명, %)

성별	모집단	표본배분		조사완료	
		빈도	비율(%)	빈도	비율(%)
남자	142,540	3,189	56.73	3,389	55.37
여자	108,737	2,611	43.27	2,732	44.63
계	251,277	5,800	100	6,121	100

4) 고려사항

장애인삶 패널조사는 사전에 계획된 목표 표본 크기보다 약 300여명 많은 표본이 조사완료 되었다. 패널 구축 이후 차년도 부터 발생할 수 있는 패널마모로 인해 표본 규모의 감소가 예상됨을 감안하여 조사 완료된 모든 장애인이 최종 표본에 포함될 수 있도록 조사완료된 모든 표본에 대한 가중치를 산출하는 방안을 마련할 필요가 있다.



3. 가중치 부여 방향

1) 고려사항¹¹⁴⁾

장애인삶 패널조사는 장애수용 과정과 사회관계 속에서 겪는 변화를 개인·가족·사회적 관점에서 탐색하고자 2018년 패널 구축 조사 및 1차년도 조사를 통해 장애 등록 초기 장애인을 대상으로 과학적이고 체계적인 종단면 패널을 구축하였다.

패널 구축 이후 차년도부터 발생할 수 있는 패널마모로 인해 표본 규모의 감소가 예상됨을 감안하여 조사 완료된 모든 장애인이 최종 표본에 포함될 수 있도록 하는 방안을 마련하기로 하였다.

표본설계 과정에서 표본크기의 한계 및 공표단위 등을 고려하여 장애유형은 7개 범주로 설계되었다. 통계의 공표단위는 전국 및 장애유형(7), 장애정도(2), 성별(2)이다. 가중치 작성에서 활용되는 변수는 기본적으로 설계변수 및 통계의 공표단위가 고려된다. 가중치 작성시 향후 자료의 활용성 측면을 고려하여 추가로 활용 가능한 변수(장애유형, 연령대, 지역 검토)에 대한 검토를 수행하기로 하였다. 따라서 장애유형은 데이터 활용성 및 용이성을 감안하여 15개 범주를 기준으로 가중값을 산출할 수 있는 방안 검토한다. 또한(표본크기의 제약으로 인해) 설계단계에서 반영하지 못한 주요변수를 반영할 수 있는 방안을 마련하기로 한다.

충분치 않은 표본크기로 인해 통계표 산출의 한계가 있는바, 공표단위 및 현실적으로 산출 가능한 통계표를 고려하여 가중치 작성시 활용할 변수를 결정하기로 하였다. 모집단 크기가 작은 경우 표본설계 과정에서 고려하지 않은 변수를 사후층으로 구성할 경우 층이 지나치게 세분화되어 비현실적인 결과를 야기하거나, 표본설계 과정에서 고려하지 않은 사후층의 표본은 조사되지 않았을 가능성이 있기 때문이다. 이를 해결하는 방안으로 층 통합을 검토한다. 이때 연령대 검토시의 층은 만 18세 이하(학령기), 만19-49세(청·중년), 만50세(장·노년)로 사후층을 구성하기로 결정하였다. 이는 만6세 이하의 자연 마모, 연령대 모집단 분포, 정책적 수요 등을 고려하여 구성한 것이다.

114) 장애인삶 패널조사는 한국장애인개발원 정책연구실 조사패널팀에서 담당하고 있다. 본 절은 장애인삶 패널조사의 전반적인 사항과 관련한 사항으로 연구진간 논의를 통해 결정된 사항이다.

2) 공표범위 검토

주요항목별 상대표준오차와 공표단위별 상대표준오차 검토 후 공표항목 및 공표단위 조정을 검토할 필요가 있다. 상대표준오차의 경우에는 양적변수에 대해서만 검토하는 것이 적절하다. 그러나 질적변수에 대한 검토가 병행 되는 경우에는 항목 간 통합, 리커트 척도의 경우 평균 점수, 비율 등을 이용하고 향후 자료제공 시 명시하도록 한다.

3) 기본방향¹¹⁵⁾

기본가중치를 설계가중치에 응답률을 조정하여 산출하고, 이때 가중값이 부여되는 인원은 조사가 완료된 인원¹¹⁶⁾으로 한다. 목표 장애인 수보다 최종 표본이 많음으로 인한 표본의 영향력을 줄이고, 집단별 분포를 반영하여 편향을 줄일 수 있는 방안을 검토한다. 가중치는 장애유형 15개 범주를 기준으로 산출할 수 있도록 적절한 사후조정 방안을 검토 한다.

통계를 공표하지 않더라도 통계 작성을 위한 집단별 분포를 추정에 반영하여 편향을 줄이는 방안을 검토한다. 이를 통해 통계의 공표 및 데이터 사용의 용이성을 높일 수 있도록 한다. 가중치 작성시 장애유형, 장애정도, 성별, 연령대, 지역변수를 고려할 수 있는 방안을 검토하고, 실제 공표와 향후 자료 사용을 고려하여 최종적인 캘리브레이션 범위를 결정하고자 한다.

가중치 작성방안별 검토 후 주요문항에 대한 상대표준오차 또는 표준오차를 함께 검토하여 최종가중치 작성방안을 마련 한다. 따라서 통계의 공표단위, 가중치 작성시 반영하고자 하는 변수, 표본크기, 사후층 수 등을 종합적으로 고려한 가중값을 산출할 수 있도록 한다.

115) 표본설계 내역과 패널구축 현황을 토대로 가중치 산출을 위한 자문회의를 실시하였다. 공표범위, 모집단 및 표본완료 현황에 따른 가중치 산출시 활용 변수와 조사완료 인원내 대한 가중치 부여 및 이론적 타당성을 검토(자문회의)하여 최종가중치 산출 방안을 마련, 적용 결과 검토 후, 최종가중치를 산출(자문회의)하였다.

116) 목표 표본보다 조사 완료 표본크기가 큰 상황이다.



4. 가중치 부여를 위한 검토 방안¹¹⁷⁾

1) 개요

장애인삶 패널조사는 전체 표본크기의 제약으로 설계 단계에서 장애유형은 7개 범주, 연령대는 내재적 층으로 적절히 표본을 할당하였으며, 지역은 표본배분이 아닌 계통추출로 표본이 추출되었다. 따라서 해당변수들의 최종 계산된 가중합은 모집단 구성비와 다를 수 있다. 장애인삶 패널조사의 1차년도 조사결과 분석에 필요한 적절한 가중치 산출을 위해서는 표본설계 및 실제 공표와 향후 자료 사용 측면을 함께 고려하여 최종 calibration 범위가 결정되어야 한다.

2) 가중치 부여 방안

최종가중치를 부여할 표본은 총 6,121명으로 목표 장애인 수보다 많은 인원이다. 이는 조사 진행 중 발생한 지속적이고 강력한 거절, 장기부재 등의 이유로 발생한 무응답으로 인해 표본 대체가 이뤄지는 과정에서 복수 접촉이 발생하여 조사가 완료된 장애인을 모두 최종 표본으로 포함하였기 때문이다. 이들을 최종 표본으로 포함한 것은 패널 구축 이후 다음연도 부터 발생할 수 있는 패널마모로 인한 표본 규모의 감소를 고려한 것이다. 따라서 조사완료된 모든 장애인은 최종 표본에 포함하여 가중치 작성 방안을 검토하였다.

가중치 산출을 위하여 활용할 변수들로 표본추출틀내 정보가 존재하는 변수는 장애유형, 장애정도, 성별, 연령, 지역변수이다. 먼저 기본가중치를 산출하였는데, 이는 무응답을 고려하여 설계가중치에 무응답 조정계수를 적용하였다. 즉 기본가중치를 먼저 계산한 후, 사후조정을 적용하여 최종가중치를 산출하였다. 사후조정 방안으로는 사후층화와 레이킹을 적용하는 방안을 검토하여 적절한 가중치를 산출하였다. 이때 활용한 변수들은 표본추출틀내 정보가 존재하는 변수들이다.

본 연구에서는 최종가중치 산정을 위하여 다음과 같이 사후층화와 레이킹 적용을 검토한 후 적절한 방안을 적용하였다. 이때 캘리브레이션 범위는 실제 공표와 향후 자료 활용도 측면을 고려하여 결정하였다. 통계 공표를 하지 않더라도 통계

117) 본 절은 최종가중치 부여 인원, 가중치 부여 방안 마련, 결과 검토 등 최종가중치 산출을 위해 실시한 자문회의 결과를 바탕으로 작성하였다.

작성을 위한 집단별 분포를 추정에 반영, 편향을 줄일 수 있도록 연령대를 고려하고, 층을 재구성하는 방안으로 검토하였다.¹¹⁸⁾

최종가중치 작성방안은 가중값으로 인한 효과¹¹⁹⁾와 검토방안별 가중치를 적용하여 변수유형별 주요 문항에 대한¹²⁰⁾상대표준오차를 비교·검토하고, 자문회의를 거쳐 최종 결정하였다.

3) 검토방안별 비교 검토¹²¹⁾

검토한 방안은 <표 IV-4>과 같다. 전체 표본크기의 제약으로 설계 단계에서 장애유형은 7개 범주, 연령대¹²²⁾는 내재적 층으로 표본이 적절히 할당되었으며, 지역은 추출단계에서 계통추출로 표본이 추출되었다. 따라서 가중값을 산출시 최종 계산된 가중값 합계가 모집단 구성비와 다를 수 있다.

장애유형은 공표단위는 아니지만 향후 자료 사용 측면을 고려하여 15개 범주로 가중값을 사후조정 하였다. 또한 목표 장애인 수 보다 최종 표본이 많은 것으로 인한 표본의 영향력을 줄이고, 집단별 분포의 반영을 통해 편향을 줄일 수 있도록 최종가중치 작성시 연령대, 그리고 지역변수까지 고려하였다.

연령대 변수에 대해서는 통계 공표를 하지 않더라도 통계 작성을 위한 집단별 분포를 추정에 반영하고, 편향을 줄이는 효과를 얻을 수 있도록 사후층으로 구성하여 적용하였다. 사후층 구성시 층이 지나치게 세분화 되어 사후층별 표본이 존재하지 않을 수 있으며, 표본이 없는 층으로 인해 가중치 산출시 발생하는 문제는 층을 통합하는 방안으로 해결할 수 있도록 검토하였다.¹²³⁾ 가중치 산출시 최종가중치의

118) 이에 관해서는 자문회의를 통해 통계적 타당성을 검토하였으며, 이는 공표 단위가 아니더라도 통계 작성을 위한 집단별 분포를 추정에 반영함으로써 편향을 줄일 수 있도록 하는 방안이 된다. 연구진 논의 결과를 토대로 사후층의 연령대 범주를 결정하였다.

119) Kish는 불균등 가중값으로 인한 효과 UWE(Unequal Weighting Effect)를 $UWE = 1 + L$ 로 정의하였는데, 여기서 $L = L = Var(W_i) / \bar{W}^2 = CV(W_i)^2$ 이다. 따라서 $UWE = 1 + CV(W_i)^2$ 으로 표현된다. 이는 총합, 비율에 대한 UWE에도 적용되며, 부차그룹과 조사설계시 고려되었던 층화에도 적용 가능하다. 이에 대한 해석은 $UWE = 1 + CV^2 = 3.72$ 라고 하면, 이는 모집단 평균, 총합, 비율 등에 대한 표본 오차가 가중값으로 인해 약 1.9배($= \sqrt{3.72}$) 증가한다고 판단할 수 있다. (출처) 임찬수 (2012). 다목적 표본조사에서 극단 가중값 탐색과 조정 방법. 성균관대학교 박사학위 논문.; Paul S. Levy, and Stanley Lemeshow. (2008). Sampling of Populations: Methods and Applications(4th) - Chapter16, John Wiley & Sons.

120) 변수유형별 주요문항은 연구진 논의를 토대로 결정되었다.

121) 본 절은 자문회의를 통해 마련한 방안을 토대로 작성하였다.

122) 연령대는 모집단 분포의 치우침이 심해 설계단계에서 반영하지 않을 경우, 최종표본이 장애유형별 전 연령대를 골고루 포함하지 못한 가능성이 높다. 따라서 전 연령대가 모두 최종 표본에 포함되면서, 연령대별 표본이 골고루 포함될 수 있도록 하고, 노년층 패널의 자연마모로 인한 표본이탈을 사전에 고려하여 표본을 구성하였다.



최소값이 1이 될 수 있도록 하였다.

〈표 IV-4〉 최종가중치 산출을 위한 검토 방안

방안	A1	B1	B2	B3
상세 내용	<ul style="list-style-type: none"> 사후증화 : 장애유형×장애정도×성별×연령대 	<ul style="list-style-type: none"> 레이킹 <ul style="list-style-type: none"> -장애유형 × 장애정도×성별 -연령대×지역¹²⁴⁾ 	<ul style="list-style-type: none"> 레이킹 <ul style="list-style-type: none"> -장애유형 -장애정도 -성별 -연령대 -지역 	<ul style="list-style-type: none"> 레이킹 <ul style="list-style-type: none"> -장애유형 -장애정도 -성별 -연령대 -권역(서울, 부산, 대구/경북, 강원, 경기, 울산/경남, 인천, 전라, 충청, 제주)

A1의 경우 사후증에 해당하는 ‘안면·여자·연령대별’ 표본은 조사되지 않아, 해당층은 통합하여 가중치를 산출하였다¹²⁵⁾. 이러한 경우 층 통합과 관련된 분석단위는 가중치 적용시 주의해야한다.

B1, B2, B3는 레이킹 보정을 적용한 방안이다. 연령대는 내재적 층으로 표본이 구성되었으며, 지역은 계통추출을 통해 표본이 추출되어, 가중합은 모집단 구성비와 다를 수 있다. 이를 레이킹 조정하여 가중값을 보정하였다.

레이킹을 적용한 가중값 보정 방안으로 B1에서는 설계변수의 분포와 내재적 층 변수들로 구성된 이차원 이상의 분포를 레이킹을 위한 분포로 활용하였다. 지나치게 세분화된 층으로 가중값의 변동을 크게 할 가능성이 있으며, 가중치 조정단계에

123) 자문회의 결과, 기본적으로 모집단 크기가 작은 경우에는 표본설계과정에서 고려하지 않은 사후층 표본은 조사되지 않았을 가능성이 높기 때문에, 이러한 점은 층 통합(collapse)을 통해 해결하는 방안을 검토해볼 수 있다. 다만, 층 통합으로 인해 모집단 구성비와 가중치 합계가 일치하지 않을 수 있음에 유의해야 하고, 가중치를 이용한 분석에서는 층 통합과 관련된 분석단위는 가중치 적용상의 주의가 필요하다.

124) 지역은 17개 특별광역시·도를 의미한다.

125) 표본 수가 2개 미만인 층의 경우 층내 분산을 계산할 수 없다. 따라서 유사한 성격을 지니는 성별로 범주를 확대조정하여 가중값을 산출하였다.

장애 유형	장애 정도	모집단			조사원료		
		여자			여자		
		만0-18세	만19-49세	만50세 이상	만0-18세	만19-49세	만50세 이상
안면	경증	2	13	13	0	4	5
		↓			↓		
안면	경증	28			9		

서 고려하지 않은 분포에 대한 가중합은 모집단 구성비와 다를 수 있다.

B2에서는 통계의 공표단위, 가중치 작성시 반영하고자하는 변수, 표본크기, 사후 층 수 등을 고려하여 주변분포를 맞추었다. B3은 지역을 권역 10개로 통합하는 방안이다. 장애유형별 지역분포를 검토하여 지리적으로 인접한 지역으로 통합하였다.¹²⁶⁾ 가중값으로 인한 효과를 살펴보면, 검토 방안별로 차이를 보이지 않았다.

〈표 IV-5〉 검토방안별 $\sqrt{1+cv^2}$

구분	A1	B1	B2	B3
평균	41.05	41.05	41.05	41.05
표준편차	50.43	47.74	47.97	46.67
$\sqrt{1+cv^2}$	1.58	1.534	1.535	1.51

가중치 비교를 위해서는 산출된 가중치 값을 직접 비교하는 것과 더불어 초기가중치 대비 최종가중치의 변화를 살펴보는 것이 필요하다(표 IV-6)¹²⁷⁾. 왜냐하면 극단가중치에 대한 평가 및 이에 대한 결정은 초기가중치 대비 최종가중치의 비율로 평가하는 것이 가중값 자체만을 평가하는 것 보다 바람직하기 때문이다. 표본설계 및 무응답 모형을 기반해 산출된 초기가중치는 추정량의 근사 비편향성과 일치성을 보장한다. 이러한 이유로 단순히 최종가중치의 크기만으로 판단하기 보다는 초기가중치 대비 최종가중치 변동량을 검토하는 것이 최종가중치를 통해 작성되는 추정량의 성질에 미칠 영향력을 판단하는데 있어서 좋은 지표가 되기 때문이다.¹²⁸⁾ 이러한 측면에서 살펴 볼 때, 검토방안별로 초기가중치 대비 최종가중치 분포간에 유의한 차이를 보이지 않았다. 즉 고려한 방안을 통해 산출된 가중치를 비교한 결과 거의 유사하게 나타나고 있으며, 이를 이용한 모집단 전체 및 관심 도메인 추정량의 통계적 정확성 역시 유의한 차이가 나타나지는 않는다(이는 다음의 이어지는 절을 참고한다.).

126) 장애범주 15개의 지역분포 중 지역별 셀빈도가 2이하인 지역인 경우, 조사인원이 2 이하인 경우에 대해 표본수가 20 이상이 되도록 지역을 10개 권역으로 통합하였다. 10개 권역은 대구/경북, 울산/경남, 전라권, 충청권, 서울, 부산, 강원, 경기, 인천, 제주 이다.

127) 기본가중치를 초기가중치로, 초기가중치 대비 최종가중치 비율을 분포를 살펴보았다.

128) Deville, Sarndal, 그리고 Sautory(1993)은 프랑스 가구 조사 분석을 위한 가중치 산출 연구에서 이 비율의 범위를 최대 1.85 수준으로 정의하여 여러 다양한 캘리브레이션 방안을 연구하였다. (출처) 변종석 외 (2018). 국민건강영양조사 제8기(2019-2021) 표본설계 및 가중치 계산. 보건복지부 질병관리본부 연구용역.

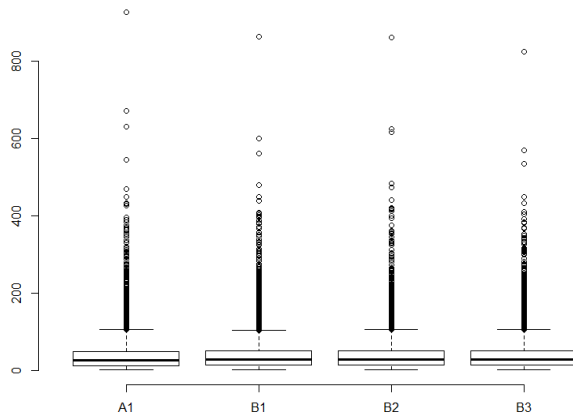


가중치 선택 문제는 통계적 추정과 더불어 현실적으로 공표 및 데이터 사용의 용이성을 담보하는 차원에서 이루어져야 한다. 따라서 향후 자료 분석의 활용도 측면 등을 종합적으로 고려한 B2를 최종가중치로 적용하는 것이 적절하다고 판단된다. 이러한 이유로 장애인삶 패널조사의 1차년도 조사 결과 분석을 위한 최종가중치는 B2를 고려하고자 한다.

〈표 IV-6〉 초기가중치 대비 최종가중치 비율 분포

통계량	A1	B1	B2	B3
max	1,377	1,284	1,280	1,227
95%	1,114	1,028	1,031	1,032
Q_3	0,985	0,990	0,999	1,008
중위수	0,895	0,947	0,933	0,938
Q_1	0,784	0,925	0,920	0,923
5%	0,819	0,840	0,929	0,927
min	1,000	1,000	1,000	1,000

추가적으로 검토방안별 상대표준오차를 살펴볼 필요가 있다. 분석이나 공표 수준이 다양해짐에 따라 해당 수준별 작성된 통계의 정확성 향상을 위해, 적절한 수준의 상대표준오차를 정의하고 이에 부합하는 표본수에 따라 공표 범위를 결정할 필요가 있기 때문이다. 다음절에서 상대표준오차를 살펴보고자 한다.



[그림 IV-1] 검토방안별 가중치 분포

〈표 IV-7〉 최종가중치의 기술통계량

관측수	6121	가중합	6,121
평균	41.05	관측치 합	251,277
표준 편차	47.97	분산	2300.73
왜도	4.31	첨도	34.54
제곱합	24395809.1	수정 제곱합	14080479.8
변동계수	116.84	평균의 표준 오차	0.61
중위수	27.55	표준 편차	47.97
사분위 범위	36.72	범위	859.90

〈표 IV-8〉 주요변수별 모집단 분포

(단위 : 명, %)

장애유형		장애정도		성별		연령대		지역	
지체	49,532	중증	92,616	남자	142,540	만0-18세	23,935	서울	39,850
시각	21,557	경증	158,661	여자	108,737	만19-49세	40,067	부산	17,791
청각	69,732	계	251,277	계	251,277	만50세 이상	187,275	대구	13,393
언어	4,086					계	251,277	인천	13,921
지적	19,086							광주	6,635
자폐성	4,766							대전	7,121
뇌병변	42,177							울산	5,026
정신	8,894							세종	1,022
신장	21,421							경기	52,921
심장	726							강원	9,077
호흡기	2,161							충북	9,267
간	3,027							충남	12,457
안면	102							전북	11,602
장루,요루	3,076							전남	12,075
뇌전증	934							경북	17,918
계	251,277							경남	17,414
								제주	3,787
								계	251,277



5. 주요항목 분석

1) 개요

표본조사의 통계 품질을 가늠하는 척도로 추정량의 상대표준오차(Relative Standard Error; RSE)가 일반적으로 사용되고 있다. 표준오차는 변수마다 단위와 절대적 크기가 다르기 때문에, 이를 표준화한 값이 상대표준오차를 보편적으로 활용하고 있다.¹²⁹⁾ 모든 통계표에 대해 오차분석을 실시하는 것은 현실적으로 불가능할 것이다. 본 절에서는 주요문항을 중심으로 추정값의 정확성에 대한 기준인 상대표준오차를 산출하여 분석을 실시하였다.

상대표준오차에 대한 국제적 기준이 있는 것은 아니지만, 각 국가는 통계품질 유지를 위한 기준을 가지고 있다. 많은 국가에서 캐나다 통계청의 기준을 준용하고 있다. 상대표준오차 5%미만은 매우 우수(excellent), 5~10%미만은 우수(very good), 10~15%미만은 좋음(good), 15~25%미만은 허용가능(acceptable), 25~35% 미만은 주의사항과 함께 사용 가능(use with caution), 35% 이상은 공표 시 신뢰불가(too unreliable to publish)의 기준을 가지고 있다. 우리나라의 경우 캐나다 통계청의 기준을 준용하고 있다.¹³⁰⁾

조사규모나 목적에 따라 다르게 나타날 수 있지만 일반적으로 상대표준오차의 값이 30%미만이면 신뢰할 수 있는 추정값으로 사용할 수 있으며, 30%이상~60%미만의 값은 주의와 함께 이용 가능하지만, 60%이상 값은 이용에 주의가 필요하다.¹³¹⁾ 전국 추정값의 상대표준오차 5% 내외가 되는 경우에 대해서 추정 결과의 정도가 높은 비교적 만족스러운 수준으로 볼 수 있다.¹³²⁾

본 절에서 상대표준오차 분석은 변수의 유형에 따라 실시하였으며, 분석을 위한 주요 문항 선정은 장애인삶 패널조사 연구진 논의를 토대로 결정하였다. 장애인삶 패널조사의 조사문항을 토대로 산출되는 통계표는 월평균 가구소득, 노후생활에 필요한 1달 생활비 평균과 같은 연속형 변수 형태, 장애수용과 같이 세부문항으로 구성되어 문항별 응답값을 평균 내어 활용하는 척도 변수, 비율추정을 하는 일상생활

129) 변종석 외. (2018). 국민건강영양조사 제8기 표본설계 및 가중치 계산. 보건복지부 질병관리본부 연구용역.

130) 이인선 외. (2016). 2016년 가정폭력 실태조사 연구. 한국여성정책연구원. 여성가족부 연구용역.

131) 통계청. (2018). 2018년 자체통계품질진단 매뉴얼 : 조사통계 부문.

132) 통계청. (2015). 자체통계품질진단 매뉴얼 : 조사통계.

존중정도¹³³⁾를 묻는 범주형 변수가 있다. 본 절에서는 이들 변수(문항)에 대한 상대표준오차를 검토하였다.

〈표 IV-9〉 변수유형별 주요 문항

변수유형	문항	조사영역(세부영역)
연속형 변수	• 월평균 가구소득	장애수용 및 변화(장애수용, 심리상태, 사회관계, 경제상황, 주거)
	• 노후생활에 필요한 1달 생활비	사회참여(장애관련 서비스, 인권, 결혼생활 및 성, 노후, 삶의 만족도)
척도 변수(평균) : 세부문항으로 구성	• 장애수용	장애수용 및 변화(장애수용, 심리상태, 사회관계, 경제상황, 주거)
단일 범주형 변수(비율)	• 일상생활존중정도	사회참여(장애관련 서비스, 인권, 결혼생활 및 성, 노후, 삶의 만족도)

2) 주요항목 비교·분석

앞 절에서 고려한 방안들을 적용하여 산출한 가중치를 적용하여, 검토한 주요 변수 유형들의 주요 문항들에 대한 상대표준오차를 비교하였다. 이들의 전국 단위 상대표준오차는 5% 미만으로 나타났으며, 전반적으로 만족할 만한 수준인 것으로 나타났다. 노후생활에 필요한 1달 평균 생활비에 대해 청각·언어장애의 상대표준오차가 상대적으로 큰 값을 가지는 것으로 나타났는데, 일반적으로 허용되는 상대표준오차의 범위로 알려져 있는 20%를 넘는 경우가 나타나지 않았다. 상대표준오차가 25%가 넘는 경우는 해석에 유의해야 한다.¹³⁴⁾

검토방안에 따른 상대표준오차는 모집단 전체 및 관심 도메인 추정량의 통계적 정확성 역시 유의한 차이를 나타내지 않았다. 따라서 앞 절의 가중치 부여 방안의 결과와 종합해 볼 때 B2를 적용하는데 무리가 없을 것으로 보인다.

133) 일상생활존중받는 정도 : 전혀 존중받지 못한다, 거의 존중받지 못하는 편이다, 존중받고 있는 편이다, 매우 존중받고 있다.

134) 표본설계 및 관리지침. 통계청 예규 제198호, 2017. 3.10.



〈표 IV-10〉 주요문항별 상대표준오차(%) - 전체

구분	A1	B1	B2	B3
노후생활에 필요한 1달 평균 생활비	1,171	1,120	1,128	1,109
월평균 가구소득	2,519	2,537	2,547	2,530
장애수용(평균)*	0.368	0.353	0.357	0.351
일상생활존중받는 정도(존중받고 있는 비율)**	1,789	1,699	1,720	1,693

* 주 1) 12개 세부 항목의 전체 평균을 사용하여 산출(일부 문항 역코딩)

2) 4점 리커트 척도 : 전혀 그렇지 않다, 거의 그렇지 않은 편이다, 그런 편이다, 매우 그렇다

**주 1) 일상생활존중받는 정도 : 전혀 존중받지 못한다, 거의 존중받지 못하는 편이다, 존중받고 있는 편이다, 매우 존중받고 있다.

2) 존중받고 있는 비율은 '존중받고 있는 편이다'와 '매우 존중받고 있다'를 합친 비율을 의미한다.

V. 결론 및 유의사항

V 결론 및 유의사항

본 연구에서는 문헌연구와 선행자료 분석을 바탕으로 장애인삶 패널조사의 1차년도 조사결과 분석을 위한 가중값 산출 방안에 대한 연구를 진행하였다. 장애인삶 패널조사의 1차년도 가중치 산출의 큰 틀은 일반적인 횡단조사의 가중값 산출단계와 기본적으로 동일한 맥락을 가진다.¹³⁵⁾

장애인삶 패널조사 가중치 작성 방안 마련을 위해 검토방안별 가중치를 비교한 결과, 방안별 유의한 차이가 있는 것으로 보이지 않았다. 최종적으로 기본가중치를¹³⁶⁾ 레이킹 조정하여 최종가중치를 산출하였다. 이에 향후 자료 분석의 활용도 측면 등을 종합적으로 고려하여 최종가중치를 적용하였다. 가중치 선택의 문제는 통계적 추정과 더불어 현실적으로 공표 및 데이터 사용의 용이성을 담보하는 차원에서 이루어져야 하며 고려한 방안 역시 그러한 차원에서 결정하였다.

공표단위의 검토시에는 가중치를 적용하여 1차년도 패널조사의 전체 분석 및 상대표준오차 분석 결과를 기초로 하고, 마이크로데이터 제공시에는 표본설계시 고려한 7개 장애유형으로 공개하는 등 자료제공 범위에 대해서는 추가적인 면밀한 검토가 필요할 것이다.

135) 백지호, 심규호. (2012). 가구패널조사에서 횡단면 가중값 작성방법. 통계개발원.

136) 설계가중치를 무응답 조정하여 기본가중치를 산출하였다.

설계가중치 : (출처) 김현지, 김태웅. (2018). 장애인삶패널조사 표본설계. 한국장애인개발원.

• 설계가중치 : 읍면동 i 에 거주하는 j 번째 장애인을 위한 표본가중치

$$w_{ij} = \left[\sum_{1h} \pi_{aij}^{-1} I_{aij}(1h) \right] \times \left[\sum_{2h} \pi_{ijs_{2h}}^{-1} I_{ijs_{2h}}(2h) \right]$$

여기서 $I_{aij} = \begin{cases} 1, & (i, j) \in U_{1h} \\ 0, & (i, j) \notin U_{1h} \end{cases}$, $I_{ijs_{2h}} = \begin{cases} 1, & (i, j) \in s_{2h} \\ 0, & (i, j) \notin s_{2h} \end{cases}$, $\pi_{aij} = \frac{n_{1h} x_{hi}}{\sum_j x_{hj}}$, $\pi_{ijs_{2h}} = \frac{m_{2h}}{M_{2h}}$

$1h$: 일상의 층, $1h = 1, \dots, 17$.

$2h$: 이상의 층, $2h = 1, \dots, 24$. (장애유형 \times 장애정도 \times 성별)

U_{1h} : 일상 표본 추출을 위한 모집단 $1h$ 번째 층

s_{2h} : 이상 표본 추출을 위한 일상 표본의 $2h$ 번째 층

x_{hj} : 이상 표본 추출을 위해 정의된 $1h$ 번째 모집단 층의 읍면동내 장애인 수

n_{1h} : $1h$ 번째 모집단 층에서 추출된 읍면동 수

M_{2h} : $2h$ 번째 1상 표본 층 내 전체 장애인 수

m_{2h} : $2h$ 번째 층에서 추출된 표본 장애인 수

• 무응답조정 : $w_{2hj} = w_{1hj} \times \frac{n_h}{r_h}$, $h = 1, \dots, 24$. 여기서 w_{1hj} 은 설계가중치, r_h 은 완전 응답한 표본 수, n_h 은 추출된 표본 수이다.



참고문헌

- 강석훈. (2003). KLIPS의 가중치 부여방안 연구. 한국노동패널연구 2003-04.
- 권순필, 김경미, 최재혁, 임찬수. (2017). 2017년 가계지출조사 가중치 사후보정방법 연구. 통계개발원.
- 김규성. (2006). 유한모집단에서 가중평균에 포함된 가중치 효과, 조사연구, 7(2), 53-69.
- 김규성, 황영은, 박진우. (2006). 패널조사에서 가중치 부여 방법 및 효과에 관한 연구.
- 김기현. (2003). 한국노동패널조사의 2003년 가중치, 노동통계개선시리즈5, 한국노동연구원
- 김영원. (2009). 표본조사에서 무응답 가중치 조정층 구성방법에 따른 효과. 한국통계학회논문집
- 김영원. (2011). 표본조사 가중치 작성 및 적용, 제1회 국가통계 방법론 심포지엄 : 튜토리얼
- 김영원 외. (2012). 표본조사의 이해와 활용, 7th edition, CENGAGE Learning.
- 김영원, 김재광, 이기재, 조유미. 한국노동패널 표본의 대표성과 가중치 보정 방법
- 김영원, 박민규, 변종석. (2013). 한국노동패널 가중치 연구
- 김영원, 이용희. (2008). 조사통계의 정확성지표 품질관리 매뉴얼. 통계청 연구용역.
- 김재광. (2006). 표본조사의 가중치 작업에 대한 고찰 : Composite calibration estimator의 소개, 한국조사연구학회 춘계학술대회 발표논문집, 125-130.
- 김재광. (2017). 표본조사론. 제2판, 자유아카데미.
- 김지경 외. (2010). 한국아동·청소년패널조사 2010 I, 한국청소년정책연구원.
- 김지경, 백혜정, 임희진, 이계오. (2010). 한국아동·청소년패널조사 2010 I, 한국청소년정책연구원.
- 김현지, 김태용. (2018). 장애인패널조사 표본설계. 한국장애인개발원.
- 박민규. (2014). 2014년 장애인 경제활동 실태조사 표본설계 연구. 한국장애인고용공단 고용개발원 연구용역.
- 박민규, 강현철. (2016). SAS를 활용한 표본설계와 통계적 추론. 자유아카데미.
- 박민규, 김사라. (2015). 한국노동패널조사 자료의 분석을 위한 패널 가중치 산출 및 사용방안 사례 연구. 조사연구, 16(1), 49-71.

- 박민규, 이정상, 박현수, 강현철. (2011). 한국청소년패널조사(KYPS) 가중치 부여 방법 연구 : 중학교 2학년 패널의 경우, 조사연구, 12(3), 173-186.
- 박서연, 박민규. (2017). 캘리브레이션 추정량의 소표본 성질에 대한 연구. Journal of the Korean Data Analysis Society, 19(1), 119-128.
- 박수미, 이택면, 김승연, 우원규, 강석훈. (2008). 2008년 여성가족패널조사 사업보고서 및 제 1차 기초분석보고서, 한국여성정책연구원.
- 백지선, 심규호. (2012). 가구패널조사에서 횡단면 가중값 작성방법, 통계개발원.
- 변종석. (2007). 통계조사에서 발생하는 오차와 보정방법. IT산업통계 품질관리 세미나.
- 변종석 외. (2018). 국민건강영양조사 제8기(2019-2021) 표본설계 및 가중치 계산. 보건복지부 질병관리본부 연구용역.
- 손창균. (2002). 완전응답과 단위무응답하에서의 사후층화와 보정의 연관성에 대한 연구. Journal of the Korean Data Analysis Society, 4(3), 297-305.
- 손창균. (2008). 한국복지패널의 가중치 조정과 향후 과제. 한국보건사회연구원, 6-23.
- 양미애. (2005). 사후층화 추정법의 효율성에 관한 연구, 숙명여자대학교 석사학위 논문.
- 이정상 외. (2011). 한국아동·청소년패널조사 2010 II 사업보고서. 한국청소년정책연구원.
- 이정상 외. (2012). 한국아동·청소년패널조사 2010 III 사업보고서. 한국청소년정책연구원.
- 이은. (2019). 국민건강영양조사 가중치 산출 방안 연구, 고려대학교 석사학위 논문.
- 이정화, 신슬비. (2018). 5차 산재보험패널조사 이용자안내서. 근로복지공단.
- 이택면, 김승연, 우원규, 이주영, 강석훈. (2009). 2009년 여성가족패널조사 사업보고서 및 제 2차 기초분석보고서, 한국여성정책연구원.
- 임정은. (2009). 경제활동인구조사의 가구가중치 산정 방안. 통계개발원.
- 임찬수. (2012). 다목적 표본조사에서 극단 가중값 탐색과 조정방법. 성균관대학교 박사학위 논문.
- 주재선, 김영란, 정성미, 송치선, 한진영, 이진숙, 손창균, 윤자영. (2018). 2018년 여성가족패널조사. 한국여성정책연구원.
- 최재혁, 김경미. (2014). 순환표본조사 표본설계 및 가중치 작성방법 연구. 통계개발원.



- 통계청. (2018). 2018년 자체통계품질진단 매뉴얼(조사통계). 통계청.
- 한국여성정책연구원. (2017). 여성가족패널조사 이용자용 통계정보보고서. 한국여성정책연구원.
- 한국조세연구원. (2012). 재정패널 가중치 보고서.
- 한국청소년정책연구원. (2014). 한국아동·청소년패널조사(KCYPS) 사업개요-제4회 한국아동·청소년패널학술대회용. 한국청소년정책연구원.
- 한국청소년정책연구원. (2017). 한국아동청소년패널조사 통계정보보고서. 한국청소년정책연구원.
- Battaglia, M., D. Izrael, D. C. Hoaglin, and M. Frankel. (2004). Tips and Tricks for Raking Survey Data (Aka Sample Balancing). American Association of Public Opinion Research.
- DeBell, Matthew and Jon A. Krosnick. (2009). "Computing Weights for American National Election Study Survey Data." Stanford University.
- Deville, J. C., and Sarndal, C. E. (1992). Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 87(376), 376-382.
- Deville, Sarndal, and Sautory. (1993). Generalized raking procedures in survey sampling. Journal of the American Statistical Association, 88(423), 1013-1020.
- Graham Kalton, and Ismael Flores-Cervantes. (2003). Weighting Methods. Journal of Official Statistics, 19(2), 81-97.
- Groves, R., D. Dillman, J.L. Eltinge, and R.J.A. Little. (2002). Survey Nonresponse. New York: Wiley.
- Holt, D., and Smith, T. M. F.(1979). Post Stratification, Journal of the Royal Statistical Society. Series A (General), 142(1), 33-46.
- Li-Chun Zhang. (1998). Post-Stratification and Calibration, Statistics Norway Research Department. Discussion Papers, No. 2016.
- Paul S. Levy, and Stanley Lemeshow. (2008). Sampling of Populations: Methods and Applications(4th). John Wiley & Sons.
- Poter, F. A Survey of procedures to identify and trim extreme sampling weights, Potter, F., and Zheng, Y. (1988). Survey of procedures to control extreme sampling weights

- Richard L. Scheaffer et al. (2006). Elementary Survey Sampling. DUXBURY 6th Edition.
- Sarndal, C.E., and S. Lundstorm. (2006). Estimation in Survey with Nonresponse. New York: Wiley.
- Smith, T. M. F. (1990). Post-Stratification. The Statistician, 40, 315-23.
- U. Rendtel., and T.Harms. (2006). Weighting and Calibration for Household Panels.
- Valliant, Richard, Jill A. Dever, and Frauke Kreuter (2013). Practical Tools for Designing and Weighting Survey Samples. New York, NY: Springer New York.

〈법률 및 웹사이트〉

- 통계승인업무 처리지침(통계청 예규 제213호)
- 통계청 고시 제2018-373호
- 통계청 통계설명자료 : <http://meta.narastat.kr/>

장애인삶 패널조사 가중치 산정 방안 연구

발 행 일 : 2019년 10월

발 행 인 : 최경숙

발 행 처 : (재)한국장애인개발원

서울시 영등포구 의사당대로22 이룸센터 5층

Tel. 02-3433-0600

Fax. 02-3433-9567

<http://www.koddi.or.kr>

ISBN 978-89-6921-337-2 93330

※ 본 연구보고서의 내용을 무단 복사하는 것을 금합니다.