

발간등록번호

『성인디지털문해능력조사』 품질개선 컨설팅
최종결과 보고서

한국통계학회

2026.06.

제 출 문

국가데이터처장 귀하

본 보고서를 「성인디지털문해능력조사」 품질개선 컨설팅의 최종 결과보고서로 제출합니다.

2026년 6월

한국통계학회

연구진

책 임 연 구 원	이상인, 충남대학교 정보통계학과 부교수
연 구 원	박승환, 강원대학교 정보통계학전공 부교수
연 구 보 조 원	임경목, 충남대학교 통계데이터사이언스학과 연구원

최종결과보고서 요약문

연구과제명	「성인디지털문해능력조사」 품질개선 컨설팅
주제어	디지털 문해능력, 표본설계, 무응답 파라데이터 분석
연구기간	2026. 04. 01. ~ 2026. 06. 30.
연구기관	한국통계학회
연구진구성	이상인, 충남대학교 정보통계학과 부교수 박승환, 강원대학교 정보통계학전공 부교수

「성인디지털문해능력조사」는 통계법 제18조에 따라 승인된 일반조사통계(승인번호 제420002호)로 우리나라 성인의 디지털 문해능력 수준을 국가 차원에서 정기적으로 파악하여 관련 정책 수립 및 디지털 문해교육 지원을 위한 기초자료를 제공하는 데 그 목적을 두고 있다. 2021년 기초연구와 2023년 조사도구 개발을 거쳐 2024년 국가승인통계로 작성승인을 받았으며 같은 해 제1차 본조사가 실시되어 2025년 8월 그 결과가 공표되었다.

본 조사의 통계정보보고서 및 품질개선 컨설팅 신청서를 검토한 결과 무응답 편향 및 표본대체 증가 가능성 진단, 무응답 보정 가중치 산출체계의 개선, 지역별 공표를 위한 적정 표본규모 산출 등이 필요한 것으로 파악되었다. 이에 본 연구에서는 현행 표본설계, 가중치 산출체계, 지역별 공표 가능성 등을 종합적으로 검토하여 성인디지털문해능력조사의 신뢰성과 정책 활용도를 제고하기 위한 실질적인 개선방안을 도출하였다.

본 연구의 주요결과는 다음과 같다.

첫째, 현행 표본설계와 무응답 조정 가중치 산출체계를 종합적으로 검토하였다. 그 결과 현행 전원조사 방식의 유지가 적절함을 확인하였으며, 무응답 조정셀 안정화 및 사후층화 가중치 개선방향을 제시하였다.

둘째, 디지털 문해능력 수준별 지역 공표 가능 범위를 상대표준오차와 로짓 신뢰구간으로 진단하고 적정 표본규모를 산출하였다. 그 결과 권역 및 시도 단위의 공표 가능 범위와 지역×연령 교차 공표를 위한 현실적 대안을 제시하였다.

셋째, 표본대체 및 무응답 실태를 진단하였다. 그 결과 예비표본이 본표본을 전반적으로 적절히 대체하고 있으며, 무응답은 접촉 이후 협조 확보 단계에서 주로 발생하는 것으로 확인되었다.

마지막으로 무응답 파라데이터 분석 등을 통해 조사운영 및 활용 체계 개선방안을 제시하였다. 이를 통해 향후 실사 관리, 표본대체 운영, 지역별 공표 확대 등에 필요한 시사점을 제시하였다.

차 례

제 1 장 서론	1
제 1 절 연구배경 및 목적	1
제 2 절 연구내용 및 방법	4
 제 2 장 표본설계 및 무응답 조정 가중치 산출 검토	6
제 1 절 표본설계 개요	6
제 2 절 현행 조사의 가구구조 및 무응답 실태	10
제 3 절 가구원 추출 방식 검토	12
제 4 절 무응답 가중치 조정 산출방안	15
제 5 절 1인 가구 등 가구구조 변화의 반영	21
제 6 절 소결	22
 제 3 장 지역별 공표를 위한 적정 표본규모 산출	24
제 1 절 개요	24
제 2 절 상대표준오차 분석결과	25
제 3 절 2범주 지표 기준 공표 가능성 검토	30
제 4 절 소결	34
 제 4 장 표본대체 및 무응답 실태 분석	36
제 1 절 분석 개요	36
제 2 절 조사구 및 가구대체 현황	36
제 3 절 본표본과 예비표본 분포 특성 비교	39
제 4 절 무응답 파라미터 분석	43
제 5 절 유사통계 사례분석	49
 제 5 장 조사운영 및 활용 체계 개선방안	51

제 1 절 무응답 대응 및 조사원 교육 체계 개선	51
제 2 절 표본대체 운영 및 현장 품질관리 고도화	52
제 3 절 지역별 공표 확대 및 통계 활용 방안	54
 제 6 장 결론	 57
 참고문헌	 59
 부록 1. 지역×연령대 디지털 문해능력 수준 정밀도 진단 상세 결과	 60
부록 2. 본표본·예비표본 비교 상세 결과	63

표 차 례

〈표 2-1〉 표본설계 요약	6
〈표 2-2〉 전체/적격/조사참여 가구원수 구분별 가구 분포	10
〈표 2-3〉 적격가구원수별 가구 분포와 일부 무응답률	11
〈표 2-4〉 구분 별 응답가구 내 개인 응답율	15
〈표 2-5〉 시도 별 응답가구 내 개인 응답율	16
〈표 2-6〉 무응답 조정셀 정의 수준별 셀 규모 비교	19
〈표 2-7〉 조사자료의 1인 가구 비율 분석	21
〈표 3-1〉 응답자 전체 추정 결과	26
〈표 3-2〉 3개 권역별 디지털 문해능력 수준 정밀도 진단 결과	26
〈표 3-3〉 6개 권역별 디지털 문해능력 수준 정밀도 진단 결과	27
〈표 3-4〉 시도별 디지털 문해능력 수준 정밀도 진단 결과	28
〈표 3-5〉 지역×6개 연령대 교차 공표 가능성 요약	30
〈표 3-6〉 응답자 전체 추정 결과(2범주 기준)	31
〈표 3-7〉 3개 권역별 디지털 문해능력 수준 정밀도 진단 결과(2범주 기준)	31
〈표 3-8〉 6개 권역별 디지털 문해능력 수준 정밀도 진단 결과(2범주 기준)	32
〈표 3-9〉 시도별 디지털 문해능력 수준 정밀도 진단 결과(2범주 기준)	32
〈표 3-10〉 지역×연령대 교차 공표 가능성 요약(2범주 기준)	34
〈표 4-1〉 조사구 대체 현황	37
〈표 4-2〉 가구 단위 대체 현황	38
〈표 4-3〉 본표본과 예비표본 응답 현황	39
〈표 4-4〉 본표본·예비표본 비교 주요 결과 요약	40

〈표 4-5〉 로지스틱 회귀분석 주요 결과 요약	41
〈표 4-6〉 응답 결과의 범주별 집계	43
〈표 4-7〉 응답률 지표	44
〈표 4-8〉 컨택횟수별 최종 조사완료율	45
〈표 4-9〉 1차 컨택 결과별 2차 조사완료 전환율	45
〈표 4-10〉 최종 조사코드별 분포	46
〈표 4-11〉 적격 가구원 중 조사 완료 현황	47
〈표 4-12〉 완료가구 수 구분별 조사구 현황	48
〈표 4-13〉 초과표본 발생 원인별 분류	49
〈표 4-14〉 유사 국가통계조사의 무응답 대응 및 표본대체 방식 비교	49
〈표 4-15〉 유사 국가통계조사의 가구원 추출방식 비교	50
〈부록 표 1〉 시도 × 3개 연령대 디지털 문해능력 수준 정밀도 진단 결과	60
〈부록 표 2〉 6개 권역 × 3개 연령대 디지털문해능력 수준 정밀도 진단 결과	62
〈부록 표 3〉 주요변수별 본표본과 예비표본간 비율 차이	64
〈부록 표 4〉 시도별 본표본 비율	66
〈부록 표 5〉 주요변수별 본표본과 예비표본간 가구원단위 평균 차이	67
〈부록 표 6〉 주요변수별 본표본과 예비표본 간 가구원단위 비율 차이	67
〈부록 표 7〉 로지스틱 회귀 모형 적합도 (개인 모형, n=10,062)	69
〈부록 표 8〉 변수별 결합 p-value	69
〈부록 표 9〉 유의한 변수별 회귀계수 추정 결과	70

그 립 차 례

〈그림 1-1〉 디지털문해능력 5개 영역 및 수준 정의	2
--------------------------------------	---

요약

성인디지털문해능력조사는 디지털 전환이 사회 전반으로 확대되는 환경에서 우리나라 성인의 디지털 문해능력 수준을 국가 차원에서 정기적·공식적으로 파악하여, 연령·지역·교육수준·경제활동 상태 등에 따른 디지털 문해능력 격차와 관련된 연구를 수행하고, 디지털 문해교육 정책 수립에 필요한 기초자료를 제공하는 데 그 목적이 있다.

성인디지털문해능력조사의 통계정보보고서 및 품질개선 컨설팅 신청서를 검토한 결과, 가구방문조사 환경 변화에 따른 무응답 편향 진단, 무응답 보정 가중치 산출체계의 개선, 지역별 공표를 위한 적정 표본규모 산출방안 마련 등이 필요한 것으로 파악되었다.

따라서 본 연구에서는 현행 가구 표본설계와 본표본 및 예비표본 운영체계, 무응답 발생 구조 및 가중치 산출절차, 지역별 공표 가능성을 종합적으로 검토하여 성인디지털문해능력조사의 품질을 향상시키기 위한 개선방안을 제시하였다.

현행 조사설계 및 가중치 산출과 그에 대한 개선안을 비교하여 요약하면 다음과 같다.

구분	현행 조사설계	조사설계 개선안
표본 규모	- 전국 약10,000명, 5,000가구, 500개 표본 조사구	- 기본 표본규모는 유지 - 지역별 공표 확대가 필요한 경우 정 밀도 기준에 따라 일부 지역 표본규모 보강 검토
표본 추출 방법	- 시도, 동·읍면, 주택유형 등을 고려 한 층화 다단계 집락추출	- 좌동

구분	현행 조사설계	조사설계 개선안
가구원 조사	- 표본조사구 내 가구를 계통추출하고, 표본가구 내 만 18세 이상 적격가구원 전원조사	- 좌동 - 전원조사 방식을 유지하되 다인가구 응답 부담 및 일부 가구원 무응답 관리 강화
표본 대체	- 본표본 조사 불가 시 동일 조사구 내 예비표본을 순차적으로 활용	- 본표본 우선 원칙을 명확히 하고, 예비표본 사용 단계, 대체 사유, 대체 시점, 완료 여부를 체계적으로 기록
설계 가중값	- 조사구 추출률의 역수와 조사구 내 가구 추출률의 역수를 곱하여 산출	- 좌동
표본 조사구내 개인 응답률 반영	- 가구 내에서 적격 가구원을 모두 조사하지만 일부 조사거부할 경우에는 표본 조사구 내에서 연령 및 성별 응답률의 역수를 가중값으로 적용함	- 조사구 단위 조정셀의 불안정성을 완화하기 위해 시도×가구유형 또는 시도 수준으로 조정셀 확대 검토
사후 층화 보정	- 시도×성×연령대 기준으로 모집단 분포에 맞게 사후층화 조정	- 성·연령·지역 보정은 유지 - 1인 가구 등 가구구조 특성을 반영할 수 있도록 레이킹 방식 적용 검토
최종 가중값	- 설계가중값 × 무응답 보정계수 × 개인 응답률 보정계수 × 사후층화 조정계수	- 기본 산출구조는 유지하되, 무응답 조정셀 안정화와 사후층화 보정방식 개선을 통해 최종가중값의 안정성 제고

제1장 서론

제1절 연구배경 및 목적

1. 성인디지털문해능력조사 개요

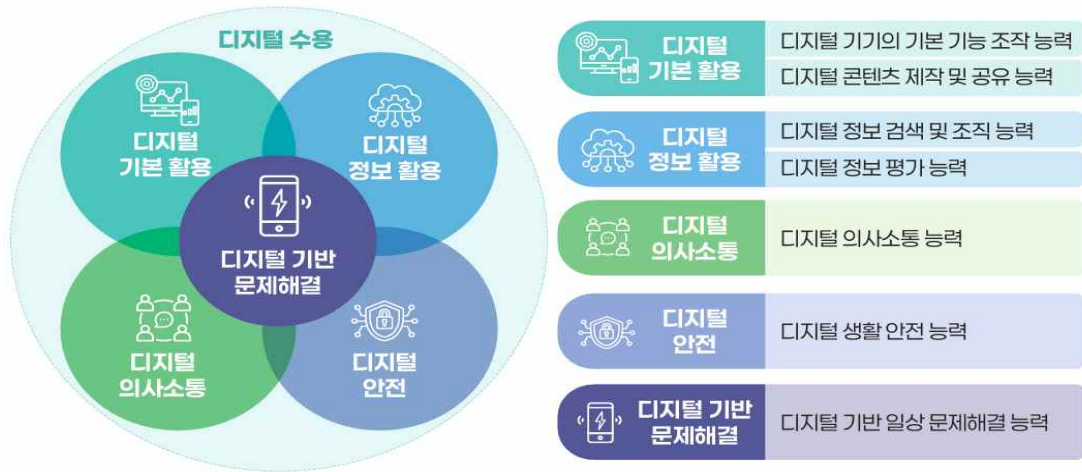
성인디지털문해능력조사는 통계법 제18조에 따라 승인된 일반조사통계(승인번호 제420002호)로 최근 디지털 문해능력 및 디지털 문해 교육의 중요성이 확대되는 환경에서 우리나라 성인의 디지털 문해능력 수준을 국가 차원에서 정기적·공식적으로 파악하고, 관련 정책 수립과 디지털 문해 교육 지원에 필요한 기초자료를 제공하는 데 그 목적을 두고 있다.

본 조사는 국가평생교육진흥원에서 작성하는 3년 주기의 조사로 조사시점 기준 대한민국에 거주하는 18세 이상 성인 남녀를 조사대상으로 한다. 조사의 주요 연혁을 살펴보면, 2021년 기초연구를 시작으로 2023년에는 조사도구 개발연구와 함께 측정 예비조사 및 시범조사가 실시되었다. 이후 2024년 국가승인통계로 작성승인을 받아 같은 해 9월부터 10월까지 제1차 본조사가 수행되었으며 그 결과는 2025년 8월에 공표되었다.

본 조사에서 디지털문해능력은 “디지털 기기와 기술을 활용하여 삶의 문제를 해결하면서 디지털 사회의 구성원으로 살아가기 위한 능력”으로 정의된다. 또한 본 조사의 표본규모는 전국 1만 명이며 95% 신뢰수준에서 예상표본오차는 $\pm 1.39\%$ 이다.

조사내용은 크게 디지털 문해능력 조사 문항, 디지털 문해교육 실태조사 문항, 그리고 응답자 정보 및 개인배경 문항으로 다음과 같이 구성되어 있다.

- 디지털 문해능력 조사: 39문항(A유형 3문항, B유형 36문항)
- 디지털 문해교육 실태조사: 38문항(C유형)
- 응답자 정보(7개) 및 개인배경(6개): 13문항



[표 1-3] 성인디지털문해능력 수준 정의

구분	정의
수준 1	디지털에 대한 기본적인 이해와 경험이 부족하거나 일상생활에서의 기본적인 디지털 기기 조작을 어려워하는 수준
수준 2	기본적인 디지털 이해와 기기 조작이 가능하지만, 일상생활에 활용하기에는 미흡한 수준
수준 3	디지털 기기나 기술을 활용하여 일상생활의 문제를 해결할 수 있으나, 비판적으로 안전하게 활용하기에는 부족한 수준
수준 4	디지털 기기나 기술을 능숙하게 활용하여 일상생활에서의 다양한 문제를 원활히 해결할 수 있는 수준

〈그림 1-1〉 디지털문해능력 5개 영역 및 수준 정의(결과보고서 일부)

디지털 문해능력 조사 문항은 디지털 기본 활용, 디지털 정보 활용, 디지털 의사소통, 디지털 안전, 디지털 기반 문제해결의 다섯 개 영역으로 구성된다. 이를 통해 디지털 기기의 기본 조작, 정보 검색 및 평가, 온라인 의사소통, 개인정보 보호와 디지털 생활안전 등 실제 활용역량을 측정하도록 설계되었으며 문항은 체크리스트형, 선다형, 상황판단형, 수행형으로 구분되어 있다.

디지털 문해교육 실태조사 문항은 디지털 기기 활용 목적, 디지털 활용경험

및 태도, 디지털 문해교육 참여경험 및 요구 등을 포함하고 있어 성인의 디지털 활용 실태와 학습 수요를 파악할 수 있도록 구성되어 있다. 또한 본 조사는 응답자의 교육수준, 경제활동 상태, 가구소득, 생활만족도, 정치관심도 등 개인배경 정보와 성별, 출생년도, 가구주와의 관계, 주사용 스마트폰 유형 등 응답자 정보를 함께 수집하고 있다. 조사방법은 구조화된 설문지를 토대로 전문 조사원이 가구를 방문하여 진행하는 가구방문 면접조사 방식으로 조사는 TAPI 시스템을 활용하여 수행된다.

2. 연구배경 및 목적

성인디지털문해능력조사는 국가 차원에서 성인의 디지털 역량 수준을 파악하고 디지털 문해교육 정책 수립을 지원하는 중요한 기초통계이다. 특히 디지털 전환이 사회 전반으로 확대되면서 연령, 지역, 교육수준, 경제활동 상태 등에 따른 디지털 문해능력 격차를 정확히 파악하는 일이 더욱 중요해지고 있다. 그러나 현행 조사는 가구 방문조사 방식에 기반하고 있어 조사환경 변화에 따른 여러 한계를 존재한다.

1인 가구 증가, 맞벌이 가구 확대, 개인정보 제공 기피, 방문조사에 대한 거부감 증가 등으로 응답자 접촉과 응답 확보가 갈수록 어려워지고 있고, 이에 따라 무응답 편향과 표본대체 증가의 가능성이 제기되고 있다. 또한 본 조사는 전국 단위의 대표성 확보를 주된 목적으로 설계되었기 때문에 시도별 또는 시도×연령별 등 세부 공표 단위에서는 충분한 표본 규모가 확보되지 않을 수 있다. 이는 지역별 디지털 문해교육 정책 수립과 취약계층 지원방안 마련에 필요한 통계 활용성을 제한하는 요인으로 작용할 수 있다.

이에 본 연구는 성인디지털문해능력조사의 표본설계, 표본대체 운영, 무응답 발생 구조, 가중치 산출체계, 지역별 공표 가능성 등을 종합적으로 검토하여 조사의 신뢰성과 정책 활용도를 높이는 것을 목표로 한다.

첫째, 표본설계 검토 및 무응답 분석 측면에서는 현행 가구 표본설계와 본 표본·예비표본 운영체계, 표본대체 방식의 적절성을 점검하고, 무응답 파라

데이터 분석을 통해 무응답 발생 패턴과 편향 가능성을 진단하여 무응답 편향을 최소화하고 표본설계를 고도화하는 방안을 마련한다. 둘째, 무응답 대응가중치 산출 측면에서는 가구유형, 적격자 수, 주거형태, 접촉 시도 특성 등의 정보를 활용해 응답확률에 기반한 무응답 보정 방안을 검토하고, 성·연령·지역 등 모집단 정보를 반영한 사후보정 체계를 포함하여 조사 목적에 적합한 가중치 산출방안을 제시한다. 셋째, 지역별 공표를 위한 적정 표본규모 산출 측면에서는 시도별·연령별 공표 수요를 반영하여 공표단위별 목표정밀도 기준을 설정하고, 1차 조사 결과를 활용한 상대표준오차 분석을 통해 시도별·연령별 공표가 가능한 적정 표본규모 산출방안을 제시한다.

이처럼 본 연구는 표본설계 점검, 무응답 파라미터 분석, 가중치 산출방법에 대한 심도 있는 분석을 통해 성인디지털문해능력조사의 신뢰성과 조사효율성을 높이고 조사 품질을 개선하고자 한다.

제2절 연구내용 및 방법

본 연구의 세부 연구내용은 다음과 같다.

(1) 가구표본 설계 개선 및 무응답 편향 최소화 방안 마련

- 무응답 파라미터(거절 사유 등)에 대한 심층 분석
 - 조사 과정에서 발생한 접촉 실패, 거절, 부재 등 무응답 사유를 검토하여 무응답 발생 특성과 편향 가능성을 파악한다.
- 표본 대체 전략 및 표본 설계 고도화
 - 본표본과 대체표본의 응답 특성을 비교하고, 표본 대체가 조사 결과에 미치는 영향을 검토한다.
- 무응답 보정 및 가중치 산출 로직 설계
 - 설계가중치, 무응답 보정, 사후층화 등 가중치 산출 절차를 검토하여 표본 대표성을 높일 수 있는 보정 방안을 마련한다.

○ 유사 국가통계 사례 비교 분석

- 유사 국가통계의 표본설계, 무응답 대응 및 가중치 산출 사례를 검토하여 본 조사에 적용 가능한 시사점을 도출한다.

(2) 지역 통계 활성화를 위한 표본 규모 산출 및 설계 고도화

○ 시도별·연령별 목표 정밀도 기준 설정

- 시도별·연령별 통계 산출에 필요한 정밀도 기준을 설정하고, 공표 가능성을 판단하기 위한 기준을 검토한다.

○ 적정 표본 규모 산출

- 설정된 목표 정밀도 기준을 바탕으로 지역별 통계 산출에 필요한 적정 표본 규모를 산출하고, 표본 배분 및 설계 개선 방향을 검토한다.

(3) 조사 운영 및 활용 체계 개선 방안 마련

○ 무응답 대응 전략을 반영한 조사 지침 및 조사원 교육 체계 개선

- 조사 현장에서 활용 가능한 무응답 대응 절차를 정비하고, 조사원 교육 내용에 반영하여 조사 수행의 일관성을 높인다.

○ 조사 품질 관리 체계 고도화 방안 제시

- 조사 진행 상황, 응답률, 표본 대체, 일부 무응답 등을 체계적으로 관리할 수 있는 품질 관리 방안을 제시한다.

○ 시도별 공표 확대를 고려한 통계 활용 방안

- 시도별 통계 공표 확대 가능성을 검토하고, 지역별 정책 수립 및 디지털문해 지원 사업에 활용할 수 있는 방안을 마련한다.

제2장 표본설계 및 무응답 조정 가중치 산출 검토

본 장에서의 조사 개요, 모집단 현황, 표본설계 및 가중치 산출 등에 대한 전반적인 내용은 결과보고서 및 표본설계 내역서를 참고하여 작성하였다.

제1절 표본설계 개요

1. 표본설계 검토

〈표 2-1〉 표본설계 요약

분 류	내 용
모집단 및 표본추출틀	① 목표 모집단: 조사시점 당시 대한민국에 거주하는 18세 이상의 성인 남녀 ② 조사 모집단: 조사시점 당시 일반조사구(섬, 기숙시설, 특수사회시설 해당 조사구 제외)에 거주하는 만 18세 이상의 성인 남녀 ② 표본추출틀: 2022년 기준 인구주택총조사 조사구 가구 명부
목표오차 또는 표본규모	전국 10,000명(5,000가구 적격가구원 전원조사) - 95% 신뢰수준, 기대표본오차 $\pm 1.39\%p$
표본추출방식 (층화, 특성, 분류지표 포함)	① 층화 : 시/도(21개 권역), 조사구 특성(동부/읍면부, 주택유형) ② 표본 추출단위별 표본추출 · 1차(psu) : 조사구 → 확률비례계통추출 · 2차(ssu) : 조사구 내 가구 → 계통추출 (분류지표: 1인가구 비율, 19-34세 인구비율) · 최종조사단위 : 표본가구 내 적격가구원 전원조사
표본배정방식	- 시도 : 가구수 기준 제곱근 비례배분 - 권역(서울, 경기)/동부/읍면부/주택유형 : 조사구 내 가구수 기준 비례배분
추정식	- 표본가중값 산출하여 복합표본설계에 적합한 모수추정법 적용
주요항목별 공표범위	- 성별, 연령별, 학력별, 가구소득별, 경찰상태, 권역, 생활만족도, 정치적 관심, 디지털 문해능력 수준 등 RSE 값을 고려하여 최종 공표단위 결정 -> 서울 및 특광역시: 서울, 부산, 인천, 대구, 대전, 광주, 울산, 세종 중소도시: 도지역 중 동 지역 농산어촌: 도지역 중 읍/면 지역 ※ 도지역: 경기도, 충청북도, 충청남도, 전라남도, 경상북도, 경상남도, 강원특별자치도, 전북특별자치도, 제주특별자치도

현행 표본설계는 전국 단위 성인 디지털 문해능력 수준을 추정하기 위한 기본 구조를 갖추고 있으며 시도별 제곱근 비례배분을 통해 소규모 지역의 표본 확보도 고려하고 있다. 다만, 표본가구 내 적격가구원 전원조사 방식은 가구 내 성인 수가 많은 가구의 응답부담을 높일 수 있으며 가구 규모 및 적격가구원 수에 따른 선택확률과 응답확률 차이를 가중치 산정 과정에서 적절히 반영할 필요가 있다. 또한 지역별 공표 또는 시도×연령별 세부 공표를 고려할 경우 현재 표본규모(전국 1만명)만으로는 일부 세부 영역에서 상대표준오차가 크게 나타날 가능성이 있으므로 공표단위별 표본규모와 추정 안정성에 대한 추가 검토가 필요하다.

따라서 향후 분석에서는 모집단 및 조사표본의 지역별·연령별 분포를 비교하고 주요 지표별 표준오차와 상대표준오차를 산출하여 현행 표본설계의 공표 가능 범위와 개선 필요성을 검토할 필요가 있다.

2. 가중치 산출과정 검토

○ 설계 가중값

- 표본조사구에 대한 추출률의 역수와 표본조사구에서 가구 추출률의 역수를 곱하여 산정

$$\text{설계가중값} = \frac{S_h}{n_h S_{hi}} \times \frac{M_{hi}}{m_{hi}} \approx \frac{S_h}{n_h \times 10}$$

- 여기서, S_h 는 h 층의 가구 수, S_{hi} 는 h 층 내 i 조사구의 가구 수이고, M_{hi} 와 m_{hi} 는 h 층 내의 i 조사구의 모집단 적격가구 수와 표본 가구 수를 나타내며 n_h 는 h 층에 할당된 표본 조사구수임.(일반적으로 모집단 가구 수와 모집단 적격가구 수가 유사하다고 가정함($S_{hi} \cong M_{hi}$ 가정))

○ 무응답 보정 계수

- 해당 지역 내 세부 층에서 조사 가구에 대한 설계가중값은 원칙적으로

동일하며, 무응답 조정은 조사구당 10가구를 조사하지만 조사성공 가구 수가 10가구 미만일 경우에 아래식으로 보정계수를 산출함

$$\text{무응답 보정 가중값} = \frac{m_{hi}}{r_{hi}}$$

- 여기서 m_{hi} 는 h 층 내의 i 조사구의 표본 가구 수, r_{hi} 는 h 층 내의 i 조사구의 응답 가구 수임

○ 표본 조사구 내 개인 응답률 반영

- 디지털 문해능력은 성 및 연령대별 차이가 뚜렷하며, 이에 따라 표본 가구 내 가구원 응답률을 반영한 무응답 보정 방법보다는 조사구 단위에서 연령대 및 성별 구분을 무응답 조정셀로 이용하는 무응답 조정 방법을 사용함
- 가구 내에서 적격 가구원을 모두 조사하지만 일부 조사거부할 경우에는 표본 조사구 내에서 연령(7)¹⁾ 및 성별(2) 응답률의 역수를 가중값으로 적용함

$$\text{표본 조사구 내 개인 응답률} = \frac{\text{표본조사구 내 성 및 연령대별 응답자 수}}{\text{표본조사구 내 적격 조사대상자 수}}$$

○ 사후층화 조정 계수

- 모집단과 가중표본의 분포가 유사하도록 최신의 인구주택총조사 결과를 기준으로 시/도별(17)에서 연령대(6) 및 성별(2) 구분을 사후층화 변수로 이용하는 사후층화 조정 방법을 적용함

$$\text{최종가중값} = \text{설계가중값} \times \text{무응답 조정 계수} \times \text{성 및 연령대별 개인 응답률 역수(조사구별)} \times \text{사후층화 조정 계수}$$

가중치 산출과정은 설계가중값, 가구 무응답 조정, 표본조사구 내 개인 응

1) 연령대 구분은 18-29세, 30-39세, 40-49세, 50-59세, 60-69세, 70세 이상으로 구분함

답률 조정, 사후층화 조정으로 구성되어 있으며 표본설계와 조사과정에서 발생하는 선택확률 차이 및 무응답을 보정하기 위한 일반적인 절차를 따르고 있다. 또한 설계가중값과 가구 무응답 조정, 사후층화 조정은 층화 다단계 집락추출 설계와 모집단 분포 보정을 반영하고 있어 전반적으로 적절한 것으로 판단된다.

다만 표본조사구 내 개인 응답률 조정의 경우 “표본조사구 내 적격 조사대상자 수”의 범위를 명확히 할 필요가 있다. 표본조사구는 원칙적으로 선정된 조사구 전체를 의미하므로 계산식의 분모가 조사구 전체 적격자 수인지, 추출 표본가구 내 적격자 수인지, 또는 조사완료 가구 내 적격자 수인지 구분되어야 한다. 현실적으로 조사미완료 가구의 성·연령대별 적격자 정보를 확보하기 어렵기 때문에 개인 응답률 조정은 조사완료 가구 내에서 파악된 성·연령대별 적격 조사대상자 수와 실제 응답자 수를 기준으로 산출하는 방식으로 이해하는 것이 적절하다.

따라서 개인 응답률 조정계수 산식은 조사완료 가구 내 성·연령대별 적격자 수와 응답자 수를 기준으로 명확히 재정의할 필요가 있으며, 조사구 단위 성·연령대 조정셀에서 표본수가 작을 경우 조정계수의 불안정성이 발생할 수 있으므로 이에 대한 추가 검토가 필요하다.

3. 표본설계 및 무응답 조정 가중치 산출 검토방향

현행 성인디지털문해능력조사는 표본조사구에서 표본가구를 추출한 뒤 표본가구 내 적격가구원(18세 이상) 전원을 조사하는 방식을 취하고 있다. 디지털 문해능력은 가구가 아니라 개인 단위의 특성이므로 가구를 추출단위로 하면서 가구원을 어떻게 조사하고 무응답을 어떻게 보정하여 가중치를 산출할지가 표본의 효율성 및 대표성에 직접 영향을 준다. 본 보고서는 이러한 관점에서 표본설계의 두 축을 검토한다.

- ① 가구원 선정 방식: 가구 내 무작위 1인 선정 / 가구당 최대 2인 제한 / 현행 전원조사 유지 방식을 비교

- ② 무응답 조정 가중치 산출: 가구 내 일부 무응답 처리, 응답확률모형·사후층화·레이킹 보정을 검토하여 최종가중치 산출방안 제시

이를 위해 먼저 현행 가구구조 및 무응답 실태를 진단한 뒤(2절), 가구원 선정 방식(3절)과 무응답 조정 가중치 산출방안(4절)을 검토하고, 1인 가구 등 가구구조 변화 반영(5절)을 다룬다.

제2절 현행 조사의 가구구조 및 무응답 실태

본 조사는 500개 표본 조사구에서 조사구당 약 10가구를 추출하여 총 5,030가구에 대한 조사를 완료하였다. 표본 5,030가구의 가구원수 분포는 다음과 같다. 가구당 적격가구원(18세 이상)은 평균 2.11명, 실제 조사참여 가구원은 평균 2.00명으로 표본설계가 가정한 ‘가구당 성인 2명’과 부합하는 것으로 대체로 확인되었다.

〈표 2-2〉 전체/적격/조사참여 가구원수 구분별 가구 분포

구분	가구원수 구분별 가구 수(가구)					평균 가구원수(명)
	1명	2명	3명	4명	5명 이상	
전체가구원 기준	732	2,765	1,126	381	26	2.25
적격가구원 기준	742	3,243	814	218	13	2.11
조사참여가구원 기준	854	3,435	627	113	1	2.00

다음으로 조사완료 가구 내 일부 무응답 실태를 검토하였다. 조사완료 5,030가구에서 파악된 전체 적격가구원은 10,607명이며 이 중 실제 조사에 참여한 가구원은 10,062명으로 나타났다. 이에 따라 조사완료 가구 내 개인 응답률은 94.9%이며 조사에 참여하지 않은 적격가구원은 545명이며 적격가구원 중 일부만 응답한 가구는 443가구로 전체 조사완료 가구의 8.8%를 차지하였다. 응답자의 대부분은 적격가구원 2명 가구에서 확보되었으며 전체 응답자의 63.5%가 적격가구원 2명 가구, 21.8%가 적격가구원 3명 가구에서 조사되었다.

한편 적격가구원 기준 3명 이상 가구는 1,045가구로 전체 조사완료 가구의

20.8%를 차지하였다. 가구 내 일부 무응답은 적격가구원 수가 많을수록 증가하는 경향을 보이며 적격가구원 2명 가구의 일부 무응답률은 2.9%에 불과하였으나 3명 가구는 28.3%, 4명 가구는 49.1%, 5명 이상 가구는 92.3%로 나타났다. 이는 적격가구원 전원조사 방식에서 가구 규모가 클수록 응답부담이 증가하고 일부 가구원의 무응답 가능성이 높아질 수 있음을 시사한다.

〈표 2-3〉 적격가구원수별 가구 분포와 일부 무응답률

적격 가구원수	가구 수	평균 조사참여 가구원	일부무응답 가구	일부무응답 가구 비율	응답자 비중
1명	742	1.00	-	0.0%	7.4%
2명	3,243	1.97	94	2.9%	63.5%
3명	814	2.70	230	28.3%	21.8%
4명	218	3.18	107	49.1%	6.9%
5명	13	2.77	12	92.3%	0.4%
전체	5,030	2.00	443	8.8%	100%

가구원 조사방식의 효율성은 동일 가구 내 구성원의 디지털 문해능력이 서로 얼마나 유사한지에 영향을 받는다. 이에 2명 이상이 응답한 4,176개 가구를 대상으로 디지털 문해능력 수준의 급내상관계수(Intraclass Correlation Coefficient, ICC)를 산출하였다. 급내상관계수는 전체 변동 중 가구 간 변동이 차지하는 비율을 의미하며 본 분석에서는 일원배치 분산분석 기반의 적률추정법을 적용하였다.

분석 결과, 디지털 문해능력 수준의 급내상관계수는 0.60으로 디지털 문해능력 수준의 전체 변동 중 약 60%가 가구 간 차이에 의해 설명되고 약 40%가 가구 내 차이에 해당함을 의미한다. 또한 2명 이상 응답가구의 58.8%는 모든 응답 가구원이 동일 연령대(예: 부부 등)에 속하는 것으로 나타났다. 디지털 문해능력이 연령과 밀접하게 관련되어 있다는 점을 고려하면 동일 가구 내 성인 구성원의 디지털 문해능력은 상당히 유사할 가능성이 있다. 이러한 결과는 적격가구원 전원조사 방식에서 동일 가구 내 추가 응답자가 제공하는 정보량이 서로 독립적인 개인 표본에 비해 상대적으로 작을 수 있음을

시사한다.

가구 내 군집으로 인한 설계효과는 다음과 같이 계산할 수 있다.

$$DEFF = 1 + (b - 1)\rho$$

여기서 b 는 가구당 평균 응답자 수, ρ 는 급내상관계수(ICC)이다. 위 식에서 $b \approx 2$, $\rho = 0.60$ 을 적용하면 설계효과는 약 1.60이 된다. 이는 현행 전원조사 방식은 한 번 접촉한 가구에서 평균 2명의 응답자를 확보할 수 있다는 점에서 조사 효율성이 높으나 동일 가구 내 응답자의 유사성으로 인해 통계적 효율은 일부 감소할 수 있다. 향후 표본규모 산출 및 공표 가능성 검토 시 이러한 가구 내 군집효과를 함께 고려할 필요가 있다.

제3절 가구원 추출 방식 검토

1. 검토개요

성인디지털문해능력조사는 가구를 표본추출 단위로 활용하되 최종 분석단위는 18세 이상 성인 개인이다. 따라서 표본가구 내 적격가구원을 어떤 방식으로 조사할 것인지는 조사 효율성, 응답부담, 표본 대표성 및 무응답 처리 측면에서 중요한 검토사항이다. 본 연구에서는 가구원 추출방식을 다음 세 가지로 구분하여 검토하였다.

방식	내용
무작위 1인 선정	표본가구에서 적격가구원 중 1명을 무작위로 선정하여 조사: 가구당 응답자 1명
가구당 최대 2인	적격가구원이 3명 이상인 가구는 2명까지만 선정하여 조사: 가구당 응답자 최대 2명
전원조사 (현행)	표본가구의 적격가구원 전원을 조사: 가구당 응답자 평균 약 2명.

2. 조사 효율성 및 응답부담 비교

무작위 1인 선정 방식은 가구당 1명만 조사하므로 가구의 응답부담이 낮으

며 가구당 최대 2인 조사 방식은 중간 수준의 부담을 갖는다. 반면 전원조사 방식은 적격가구원 전원을 조사하므로 가구원 수가 많은 대가구일수록 응답 부담이 증가할 수 있다.

조사 효율성 측면에서 전원조사는 한 번 접촉한 가구에서 평균 약 2명의 응답자를 확보할 수 있어 가구 접촉 비용 대비 응답자 확보 효율이 높다. 그에 비해 무작위 1인 선정 방식은 동일한 응답자 수를 확보하기 위해 더 많은 가구를 접촉해야 하므로 조사비용과 현장 부담이 증가할 수 있으며 가구당 최대 2인 조사 방식은 적격가구원 3명 이상 가구에서만 조사대상을 축소하므로 조사 효율성은 전원조사 방식과 비교적 유사할 것으로 판단된다.

3. 표본의 대표성 및 가중치 설계 비교

표본 대표성 측면에서 무작위 1인 선정 방식은 가구당 1명만 조사하므로 가구 내 군집효과를 줄일 수 있다는 장점이 있다. 다만 적격가구원 중 1명을 선정하는 과정에서 가구 내 선정확률이 발생하므로 이를 반영한 개인가중값 산출이 필요하다. 가구당 최대 2인 조사 방식 또한 적격가구원 수가 3명 이상인 가구에서 일부 가구원이 조사대상에서 제외되므로 가구원 선정확률 차이를 반영한 추가적인 가중치 조정이 필요하다.

반면 전원조사 방식은 표본가구 내 적격가구원 전원을 조사하므로 원칙적으로 가구가중값을 개인가중값으로 확장하여 활용할 수 있어 가중치 구조는 비교적 단순하다. 다만 조사완료 가구 내 일부 적격가구원이 응답하지 않는 경우에는 성·연령대별 개인 무응답 조정이 필요하다. 한편 현행 전원조사 방식은 동일 가구 내 응답자의 디지털 문해능력이 유사할 경우 통계적 효율이 감소할 수 있다. 앞서 분석한 바와 같이 디지털 문해능력 수준의 급내상관계수는 0.60으로 나타났으며 가구당 평균 응답자 수를 2명으로 가정할 경우 가구 내 군집에 따른 설계효과는 약 1.60 수준으로 산출된다.

4. 무응답에 대한 강건성 비교

가구원 추출방식은 가구 내 무응답이 발생할 때의 안정성 측면에서도 차이가 있다. 특히 디지털 문해능력은 연령대별 차이가 크기 때문에 가구 내 무응답이 특정 연령대에 집중될 경우 조사결과에 편향이 발생할 수 있다.

무작위 1인 선정 방식은 무작위 선정(최근 생일자 방식, Kish 격자 등)과 적절한 가중치 보정이 이루어진다면 확률표본으로서의 성격을 가질 수 있다. 그러나 선정된 1인이 응답을 거부할 경우 해당 가구가 통째로 조사에서 제외되거나 가구 내 다른 구성원으로 대체조사가 이루어질 가능성이 있다. 이 경우 최초의 무작위 선정 원칙이 훼손될 수 있으며 재택 가능성이 높은 고령층 또는 비경제활동 인구가 과대대표될 위험이 있다. 반면 전원조사 방식은 일부 가구원이 응답하지 않더라도 응답한 가구원의 자료를 보존할 수 있다. 또한 미응답 가구원에 대해서는 성·연령대별 조정셀을 활용한 개인 무응답 보정이 가능하므로 무응답에 상대적으로 강건한 구조를 갖는다.

5. 요약

세 가지 가구원 추출방식 중 어느 하나가 모든 기준에서 우월하다고 보기는 어렵다. 무작위 1인 선정 방식은 응답부담과 가구 내 군집효과 측면에서 장점이 있으나 동일 표본규모 확보를 위해 더 많은 가구 접촉이 필요하고 선정 가구원의 무응답 발생 시 가구 전체가 소실될 수 있다는 한계를 가지고 있다. 가구당 최대 2인 조사 방식은 대가구의 응답부담을 일부 완화할 수 있으나 3명 이상 가구에서 가구원 선정확률 차이를 반영한 가중치 조정이 필요하다. 현행 전원조사 방식은 대가구 응답부담과 가구 내 군집효과라는 한계가 있으나 한 번 접촉한 가구에서 평균 2명의 응답자를 확보할 수 있어 조사 효율성이 높고 일부 가구원 무응답에도 응답자료를 보존할 수 있다는 장점이 있다.

따라서 현 단계에서는 현행 전원조사 방식을 유지하되 대가구의 응답부담 완화, 가구 내 일부 무응답 보정, 성별·연령대별 조정셀의 안정성 확보 방안을 함께 검토하는 것이 적절하다고 판단된다.

제4절 무응답 가중치 조정 산출방안

1. 응답가구 내 개인 응답률 현황

성인디지털문해능력조사는 표본가구 내 18세 이상 적격가구원 전원을 조사하는 방식이므로 무응답은 가구 단위 무응답과 가구 내 개인 단위 무응답으로 구분하여 검토할 필요가 있다. 가구 단위 무응답은 예비표본 대체를 통해 대부분 보완되지만 조사구당 조사성공 가구 수가 10가구 미만인 경우에는 응답가구 수를 기준으로 가구 단위 무응답 보정계수를 산출하여 반영한다. 한편, 조사완료 가구 내 일부 적격가구원이 응답하지 않은 경우에는 개인 단위 무응답 조정이 필요하다.

다음 <표 2-4>는 응답 완료 가구를 기준으로 가구 내 적격가구원수와 조사 참여 가구원수를 집계한 결과이다. 이는 전체 표본가구 응답률이 아니라 “응답가구 내부의 개인 참여율”로 해석해야 한다.

<표 2-4> 구분 별 응답가구 내 개인 응답율

구분	범주	응답완료 가구수	적격 가구원수 합계	조사참여 가구원수 합계	응답가구 내 개인 참여율	일부응답 가구 비율
전체	전체	5,030	10,607	10,062	94.9%	8.8%
지역 유형	서울·특광역시	2,285	4,888	4,575	93.6%	10.7%
	중소도시	1,763	3,674	3,521	95.8%	7.4%
	농산어촌	982	2,045	1,966	96.1%	6.9%
권역	서울	581	1,279	1,165	91.1%	14.1%
	경기·인천	971	2,008	1,946	96.9%	5.6%
	충청권	860	1,803	1,720	95.4%	8.3%
	호남권	743	1,510	1,485	98.3%	3.1%
	영남권	1,484	3,193	2,960	92.7%	12.7%
	강원·제주	391	814	786	96.6%	6.4%
적격 가구원수	1명	742	742	742	100.0%	0.0%
	2명	3,243	6,486	6,392	98.6%	2.9%
	3명	814	2,442	2,198	90.0%	28.3%
	4명	218	872	694	79.6%	49.1%
	5명	13	65	36	55.4%	92.3%

응답완료 5,030가구에서 파악된 적격가구원은 10,607명이며 이 중 실제 조사에 참여한 가구원은 10,062명으로 나타났다. 이에 따라 응답가구 내 개인 참여율은 94.9%이며 일부 적격가구원만 응답한 가구는 443가구로 전체 응답완료 가구의 8.8%를 차지하였다.

〈표 2-5〉 시도 별 응답가구 내 개인 응답율

시도	응답완료 가구수	적격 가구원수 합계	조사참여 가구원수 합계	응답가구 내 개인 참여율	일부응답 가구 비율
전체	5,030	10,607	10,062	94.9%	8.8%
서울	581	1,279	1,165	91.1%	14.1%
부산	340	719	680	94.6%	9.7%
대구	292	642	584	91.0%	15.8%
인천	310	635	620	97.6%	4.5%
광주	232	478	466	97.5%	4.3%
대전	230	481	460	95.6%	6.5%
울산	190	427	380	89.0%	20.0%
세종	110	227	220	96.9%	5.5%
경기	661	1,373	1,326	96.6%	6.1%
강원	240	504	485	96.2%	7.1%
충북	240	517	480	92.8%	14.2%
충남	280	578	560	96.9%	5.7%
전북	255	509	508	99.8%	0.4%
전남	256	523	511	97.7%	4.7%
경북	322	677	636	93.9%	9.6%
경남	340	728	680	93.4%	11.8%
제주	151	310	301	97.1%	5.3%

권역별로는 서울·특광역시의 개인 참여율이 93.6%로 상대적으로 낮고 농산어촌은 96.1%로 상대적으로 높게 나타났다. 시도별로는 울산, 대구, 서울 등에서 응답가구 내 개인 참여율이 비교적 낮은 수준을 보임과 동시에 적격가구원 수가 많을수록 가구 내 일부 무응답은 증가하는 경향을 보였다. 적격가구원 2명 가구의 개인 참여율은 98.6%로 높았으나 3명 가구는 90.0%, 4명 가구는 79.6%, 5명 이상 가구는 55.4%로 감소하였다. 이는 전원조사 방식에서 대가구일수록 일부 가구원 무응답 가능성이 높아질 수 있음을 시사한다.

2. 현행 무응답 조정 방식 검토

현행 최종가중값은 다음과 같이 산출된다.

$$\text{최종가중값} = \text{설계가중값} \times \text{무응답 조정 계수} \times \text{성 및 연령대별 개인 응답률 역수(조사구별)} \times \text{사후층화 조정 계수}$$

무응답 보정은 크게 두 단계로 구분된다. 첫째, 표본가구 자체가 조사되지 않은 경우를 보정하는 가구 단위 무응답 조정이고, 둘째, 조사완료 가구 내 일부 적격가구원이 응답하지 않은 경우를 보정하는 개인 단위 무응답 조정이다.

가구 내 일부 무응답을 단순히 해당 가구의 가구원 응답률의 역수로 보정하는 방식은 적절하지 않을 수 있다. 예를 들어 한 가구에 20대와 50대 적격가구원이 있고 20대만 응답한 경우 가구 내 응답률의 역수를 적용하면 20대 응답자가 50대 무응답자를 대신 대표하게 된다. 디지털 문해능력은 연령대별 차이가 크므로 이러한 방식은 추정 편향을 유발할 가능성이 있다. 따라서 가구 내 일부 무응답은 동일 가구 내 응답자에게 단순 이전하기보다 성별·연령대 등 유사한 특성을 가진 응답자 집단을 기준으로 보정하는 것이 타당하다. 이러한 점에서 현행 조사에서 성별×연령대 조정셀을 활용하는 방향은 방법론적으로 적절한 것으로 판단된다.

본 연구 제안 단계에서는 각 표본단위의 응답확률 $P(\text{응답}|관련변수)$ 을 추정하고 그 역수를 무응답 가중치로 쓰는 응답확률모형 기반 무응답 가중치 산출방안을 검토 대상으로 두었다. 이 방법으로 모형을 적합하려면 응답자뿐 아니라 무응답자에 대해서도 설명변수가 관측되어야 하지만 현 분석자료에는 응답자 10,062명만 수록되어 있을 뿐 무응답 가구와 무응답 가구원에 대한 기록이 없다. 가구 내 무응답자의 연령·성별조차 파악되지 않아 개인 단위 응답확률모형을 직접 적합하기 어렵다.

따라서 현 시점에서 가용한 정보로 실현 가능한 방법은 셀(가중층) 기반 무응답 보정이다. 이는 표본단위를 유사한 특성의 셀로 나눈 뒤 셀별 응답률의 역수를 적용하는 방식으로, 현행 조사의 「조사구×연령대×성별」 보정이 여기에 해당한다. 셀 기반 보정은 각 셀을 하나의 응답확률 계급으로 보는 비모

수적 응답확률모형으로 이해할 수 있어 타당한 대안이다. 다만 향후 모형 기반 고도화를 위해서는 표본으로 추출된 전체 가구와 가구원을 파라미터 및 가구원 명부와 연결한 「표본관리 파일」의 축적이 선행되어야 한다.

3. 무응답 조정셀 정의의 적절성 검토

앞서 무응답 조정셀을 통한 가중치 보정의 적정성을 살펴보았다. 그러나 성별·연령대별 개인 응답률을 산출하기 위해서는 응답자뿐만 아니라 무응답자를 포함한 전체 적격가구원의 성별 및 연령대 정보가 필요하다.

$$(\text{조정셀 } j\text{내})\text{개인 응답률} = \frac{\text{조정셀 } j\text{의 응답자 수}}{\text{조정셀 } j\text{의 적격 조사대상자 수}}$$

현재 원자료에는 적격가구원 수가 가구 단위의 「수」로만 기록되어 있어, 무응답 가구원이 어느 연령대·성별 셀에 속하는지 파악할 수 없다. 성별×연령대 조정셀별 응답률을 산출하려면 조사완료 가구 내 미응답 가구원의 성별과 연령대가 가구원 명부에 기록되어 있어야 하는데, 적격가구원 수만 가구 단위로 집계되어 있고 미응답 가구원의 성별·연령대 정보가 없으면 개인 응답률을 직접 산출하기 어렵다.

따라서 개인 무응답 조정은 아래와 같이 무응답자를 포함한 전체 가구원 명부의 확보 여부에 따라 구분하여 검토하였다.

가. 무응답자를 포함한 전체 가구원 명부가 있는 경우

조사원이 가구 방문 시 작성하는 가구원 명부에 무응답 가구원의 연령·성별이 기록되어 있다면 연령대×성별 셀의 분모인 적격 조사대상자 수를 산출할 수 있으므로 셀 응답률을 정의할 수 있다. 이 경우 핵심 쟁점은 무응답 조정셀을 어느 지역 수준에서 정의할 것인가이다. 현행 방식은 조사구(500개) 단위에서 연령대(6)×성별(2) 조정셀을 구성하는 방식인데, 조사구당 응답자는 평균 약 20명에 불과한 반면 연령대×성별 조정셀은 최대 12개로 구성되므로 셀별 표본규모가 매우 작아질 수 있다. 실제로 원자료를 분석한 결과 응답자

가 1명 이상인 조사구×연령대×성별 셀은 4,441개였으며 셀당 평균 응답자 수는 2.3명에 그쳤다. 또한 응답자 5명 미만 셀이 전체의 93.0%에 달해 현행 조사구 단위 조정셀은 과도하게 희박한 것으로 판단된다.

〈표 2-6〉 무응답 조정셀 정의 수준별 셀 규모 비교

조정셀 정의 (× 연령대 × 성별)	셀 수	셀당 평균 응답자 (분자)	5명 미만 셀 비율	비고
조사구(500개)	4,441	2.3명	93.0%	과도하게 희박
시군구(179개)	1,844	5.5명	56.3%	여전히 희박
층 수준(시도×가구유형)	508	19.8명	9.6%	안정적
시도 (17개)	204	49.3명	0.5%	매우 안정적

이처럼 현행 조사구 단위 셀은 이미 지나치게 희박하여 추가 세분화가 불가능하다고 판단되므로 조정셀을 조사구가 아니라 층(시도×가구유형) 또는 시도 수준에서 연령대×성별로 정의하여 셀당 응답자를 20~50명 수준으로 확보할 필요가 있다. 아울러 최소 셀 크기 기준(예: 응답자 20명)과 이에 미달할 경우의 셀 병합 규칙을 명문화하는 방안도 함께 고려할 필요가 있다.

나. 무응답자를 포함한 전체 가구원 명부가 없는 경우

무응답자의 연령·성별이 기록된 명부가 없다면, 연령대×성별 셀의 분모를 알 수 없으므로 셀 응답률은 조사구든 시도든 어느 지역 단위로도 산출할 수 없다. 이는 앞서 살펴본 셀 희박성과는 차원이 다른 보다 근본적인 자료상의 제약이다. 희박성은 셀을 키워 해결할 수 있지만, 분모 자체가 없으면 셀을 어떻게 정의하더라도 응답률이 정의되지 않는다. 이 경우 「연령대×성별 셀 무응답 조정」 단계 자체를 수행할 수 없으며 그 기능을 다른 보정 단계로 이전해야 한다.

실현가능한 방안으로는 가구 단위 무응답 조정은 그대로 유지하는 것이다. 명부가 없더라도 표본가구 수와 응답가구 수 등 가구 단위 정보는 확보되어 있으므로 성·연령과 무관한 가구 단위 무응답 조정계수는 종전과 같이 적용할 수 있다. 다만 미응답 적격가구원의 성별·연령대 정보가 없어 성별×연령

대별 개인 응답률을 산출할 수 없으므로 개인 단위 무응답 조정은 직접 적용하기 어려우며 이 경우 가구 내 일부 무응답으로 발생할 수 있는 응답자의 성·연령 분포 불균형은 최종 사후층화 또는 레이킹 조정 단계에서 보완하는 방안을 검토할 수 있다.

이 방식은 개인 무응답을 발생 단계에서 직접 보정하지 않고 최종 응답표본의 성·연령 분포를 결과적으로 모집단 분포에 맞춘다는 점에서 한계가 있으며 조정셀 안에서 무응답이 무작위로 발생한다는 가정을 필요로 한다. 그러나 명부가 없는 한 이는 현실적으로 최선의 방안이며 실제로 다수의 가구 조사가 이러한 구조를 취하고 있다.

다. 요약

무응답자를 포함한 전체 가구원 명부가 확보되어 있는 경우에는 연령대×성별 조정셀을 이용하여 가구 내 개인 무응답을 직접 보정할 수 있다. 다만 조사구 단위 조정셀은 규모가 지나치게 작아 불안정할 수 있으므로 층 수준 또는 시도 수준에서 연령대×성별 조정셀을 구성하는 방안을 검토할 필요가 있다.

반면 무응답자의 연령·성별 정보가 포함된 명부가 없는 경우에는 연령대×성별 개인 응답률을 산출할 수 없어 해당 조정 단계를 직접 적용하기 어려우므로 가구 단위 무응답 조정은 유지하되 가구 내 개인 무응답으로 발생할 수 있는 성·연령 분포의 불균형은 최종 사후층화 또는 레이킹 조정 단계에서 보완하는 것이 현실적이다.

따라서 조사기관은 우선 전체 가구원 명부의 존재 여부와 보존 상태를 확인하고 향후 조사에서는 무응답 적격가구원의 연령·성별 정보를 가구원 명부에 기록·관리하여 셀 기반 개인 무응답 보정이 가능하도록 자료 기반을 마련할 필요가 있다.

4. 사후층화·레이킹 보정 방안

현행 사후층화는 시도(17)×연령대(6)×성별(2), 총 204개 셀을 인구주택총조사 분포에 일치시킨다. 이 204개 셀의 셀당 응답자는 평균 49명, 5명 미만 셀은 0.5%로 사후층화에 안정적이며, 디지털 문해능력의 주요 결정요인(연령·성별·지역)을 모두 반영하므로 현행 사후층화 변수 구성은 적절하다.

한편 가구원 수(1인/다인)를 사후층화 보정 변수로 추가한다면, 1인 가구의 과소대표를 가중 단계에서 교정하는 부수적 효과가 있을 것으로 기대할 수 있을 것이다. 가구원 수 등 추가 변수를 반영하려면 사후층화보다는 레이킹(반복비례적합)으로 완전교차 셀이 아니라 각 변수의 주변분포만 모집단에 일치시키는 보정 방안을 고려하여야 할 것이다.

제5절 1인 가구 등 가구구조 변화의 반영

최근 1인 가구 비율은 지속적으로 증가하고 있다. 1인 가구 비율은 2005년 20.0%에서 2022년 34.5%(약 750만 가구), 2024년 36.1%로 증가한 것으로 나타나 표본설계 및 가중치 산출 과정에서 이러한 가구구조 변화가 적절히 반영되고 있는지 점검할 필요가 있다.

현행 표본설계는 조사구를 층화할 때 1인 가구 비율을 분류지표로 사용하여 1인 가구가 많은 지역과 적은 지역이 표본 조사구에 고르게 포함되도록 한다. 따라서 추출틀 차원에서 1인 가구 분포는 적절히 반영되어 있다고 볼 수 있다. 그러나 표본의 1인 가구(전체가구원수 1명)는 732가구로 전체의 14.6%이다. 권역·주택유형별로는 13.7~16.1%로 비교적 고르며 권역별로는 경기·인천(17.4%)이 가장 높고 충청권(8.6%)이 가장 낮다.

〈표 2-7〉 조사자료의 1인 가구 비율 분석

구분	항목	1인 가구 수	1인 가구 비율
표본 전체	5,030 가구 전체	732	14.6%

구분	항목	1인 가구 수	1인 가구 비율
지역유형)	서울·특광역시	313	13.7%
	중소도시	284	16.1%
	농산어촌	135	13.7%
주택유형별	1. 아파트	336	14.1%
	2. 일반	261	15.7%
	3. 읍/면부	135	13.8%
권역별	서울	77	13.3%
	수도권	169	17.4%
	충청권	74	8.6%
	전라권	105	14.1%
	경상권	253	17.0%
	기타	54	13.8%
응답자 기준	단독 성인가구 거주 응답자(비가중)	742명	7.4%

또한 표본의 1인 가구 비율(14.6%)이 2024년 인구주택총조사의 1인 가구 비율(36.1%)에 비하여 크게 낮게 나타났다. 이는 조사구 추출의 문제가 아니라 실사 단계의 차별적 무응답에서 비롯된 것으로 추측된다. 1인 가구는 접촉이 어렵고 단독 거주자 1인이 거절하면 가구 전체가 집계되지 않는다. 또한 현행 사후층화 가중은 성·연령·시도만 보정하며 가구유형(1인/다인)은 보정 변수에 포함되지 않는다.

이를 해결하기 위해 조사 단계에서 1인 가구 관리 강화가 필요하다. 1인 가구 표본의 접촉 시도 횟수·방문 시간대 다양화, 예비표본 대체 시 1인 가구 대체 원칙을 명확히 해야 한다. 사후층화 또는 레이킹에 가구유형(또는 가구원 수)을 추가하여 가구구조 과소대표를 보정하는 방안도 고려할 수 있을 것이다.

제6절 소결

가구원 추출 방식에서 전원조사는 가구당 약 2명의 면접을 추가적인 가구 접촉 비용 없이 확보할 수 있어 동일한 정밀도를 더 적은 가구 접촉으로 달성한다. 또한 디지털 문해능력처럼 연령차가 큰 특성에서도 가구 내 무응답 대

체에 강건하므로 현행 전원조사 방식을 유지하는 것이 적절하다고 판단된다.

무응답 가중치 조정에서 가구원 응답률 역수 방식은 무응답자를 같은 가구의 다른 연령대 응답자로 대체하는 결과를 초래하므로 부적절하며 현행의 연령대×성별 조정셀 방식이 타당하다. 다만 조사구 단위 셀은 응답자가 평균 2.3명에 불과하여 응답률 추정이 불안정하므로 조정셀을 시도×가구유형 또는 시도 수준으로 정의하고 최소 셀 크기 기준에 따른 병합 규칙을 마련할 필요가 있다. 더 나아가 모형 기반 응답확률 보정을 적용하기 위해서는 무응답자 명부와 파라미터의 축적이 선행되어야 한다.

최종 가중치 산출에서는 현행 산출체계를 대체로 유지하되 가구원 무응답 조정셀의 정의를 안정화하고 사후층화 조정을 레이킹 방식으로 확장하는 것이 적절하다. 특히 1인 가구의 과소대표 가능성을 완화하기 위해 가구원 수 또는 1인·다인 가구 여부를 레이킹의 주변분포 보정변수로 추가하는 방안을 검토할 필요가 있을 것이다.

제3장 지역별 공표를 위한 적정 표본규모 산출

제1절 개요

본 장에서는 2024년 성인디지털문해능력조사의 핵심 결과변수인 디지털 문해능력 수준(수준 1~수준 4)을 대상으로 어느 공표단위까지 통계의 신뢰성을 확보하면서 결과를 공표할 수 있는지를 파악하고자 한다. 이를 위해 공표가능 범위를 상대표준오차(RSE)로 진단하고 공표가 어려운 영역에 대해서는 필요한 표본규모를 산출하고자 한다. 공표단위는 시도, 3개 권역, 6개 권역을 기본으로 하였으며 각 지역 단위에서 6개 연령대 구간과의 교차 공표가 가능한 시도 함께 검토하였다. 범주별 비율 추정에서 상대표준오차는 일반적으로 다음과 같이 정의된다.

$$RSE(\hat{p}) = \frac{SE(\hat{p})}{\hat{p}} \times 100 (\%)$$

비율이 작은 범주일수록 절대 정밀도가 동일하더라도 상대표준오차는 구조적으로 크게 나타난다. 디지털 문해능력 수준에서는 수준 1이 전국 비율 8.2%로 가장 작은 범주이며 일부 시도에서는 그 비율이 1% 미만으로 떨어져 상대표준오차가 100%에 육박하기도 한다. 이 경우 상대표준오차만으로 보면 해당 통계가 “공표 불가” 수준으로 판단될 수 있으나 실제로는 신뢰구간 반폭이 좁아 절대 정밀도가 양호한 경우가 많다. 따라서 본 분석에서는 상대표준오차를 기본 지표로 사용하되 다음의 두 가지 보조지표를 함께 검토하였다.

첫째는 절대 신뢰구간 반폭으로, 로짓 변환에 기반한 95% 신뢰구간의 반폭, 즉 신뢰구간 길이의 절반(단위 %p)을 의미한다. 이 지표는 비율이 0이나 1에 가까운 경계 영역에서도 안정적으로 산출되어 소수 범주의 실질적인 정밀도를 직접 보여준다. 본 분석에서는 로짓 신뢰구간의 반폭이 5%p 이하인 경우를 절대정밀도 충족 기준으로 삼았으며 이는 본 장의 모든 표에 제시한 '로짓 CI 반폭'과 동일한 지표이다.

둘째는 다범주 종합 판정 기준이다. 디지털 문해능력 수준은 4개 범주로 구성되므로 단일 RSE가 아니라 범주별 RSE를 종합하여 공표 가능 여부를 판정하였다. 구체적으로 비율이 충분히 큰 핵심 범주인 수준 2·3·4의 최대 RSE가 25% 이하인지를 공표 판정의 구속 기준으로 사용하였다. 반면 비율이 작아 RSE가 구조적으로 팽창하는 수준 1은 이 기준에서 제외하고, 앞서 설명한 절대 신뢰구간 반폭 기준(5%p 이하)으로 별도 판정하였다.

본 분석의 모든 신뢰구간(표의 95%CI 반폭)은 단순 Wald 구간($\hat{p} \pm 1.96 \times SE(\hat{p})$)이 아니라 로짓(logit) 변환 기반 신뢰구간으로 산출하였다. 복합표본 추정으로 비율 $\hat{p}$ 과 표준오차 $SE(\hat{p})$ 를 구한 뒤 다음 절차를 적용한다.

- ① 로짓 변환: $\text{logit}(\hat{p}) = \log(\hat{p} / (1 - \hat{p}))$
- ② 델타법으로 로짓 척도의 표준오차: $SE(\text{logit}) = SE(\hat{p}) / [\hat{p}(1 - \hat{p})]$
- ③ 로짓 척도에서 Wald 구간: $\text{logit}(\hat{p}) \pm 1.96 \times SE(\text{logit})$
- ④ 역로짓 변환으로 비율 척도 복원: $p = 1 / (1 + \exp(-\text{logit}(\hat{p})))$

이 방식은 신뢰구간이 항상 (0,1)범위 안에 있고 비대칭 형태를 가지므로 비율이 0 또는 1에 가까운 소수 범주의 정밀도를 안정적으로 표현할 수 있다. 예를 들어 부산의 수준1 추정비율이 0.2%인 경우, 단순 Wald 구간으로는 하한이 음수가 되어 비율로 해석하기 어렵다. 반면 로짓 변환 기반 신뢰구간은 항상 (0,1) 범위 안에 머물기 때문에 소수 범주의 신뢰구간도 해석 가능하다. 따라서 로짓 신뢰구간은 비율 추정에서 단순 Wald 구간과 RSE가 갖는 한계를 분산안정화 변환을 통해 보완하는 표준적인 방법이며 R survey 패키지의 `svyciprop(method="logit")`과 동일한 산출방식이다.

제2절 상대표준오차 분석결과

1. 지역별 상대표준오차 분석결과

다음 <표 3-1>에는 응답자 전체에 대한 디지털 문해능력 수준별 비율 추정 결과가 제시되어 있다. 전국 기준에서는 디지털 문해능력 4개 범주 모두 안정적인 정밀도를 보였다. 수준 1은 4개 범주 중 비율이 가장 작은 범주이나 RSE는 7.9%, 로짓 변환 신뢰구간 반폭은 1.2%p로 나타나 절대정밀도 기준을 충분히 충족하였다. 수준 2~4 역시 RSE가 각각 4.1%, 3.9%, 2.3%로 낮게 나타나, 전국 단위에서는 디지털 문해능력 수준별 비율을 안정적으로 공표할 수 있는 것으로 판단된다.

<표 3-1> 응답자 전체 추정 결과

디지털문해능력	표본수	추정비율	RSE	로짓 CI 반폭
수준1	897	8.2%	7.9%	1.2%p
수준2	2,242	17.7%	4.1%	1.4%p
수준3	2,409	21.4%	3.9%	1.6%p
수준4	4,514	52.8%	2.3%	2.4%p

3개 권역별 디지털 문해능력 비율 추정 결과를 <표 3-2>에 제시하였다. 3개 권역은 모든 권역에서 수준1 절대정밀도와 수준2~4 핵심범주 RSE 기준을 모두 충족한다. 따라서 3개 권역 단위에서는 디지털 문해능력 수준별 비율을 안정적으로 공표할 수 있는 것으로 판단된다.

<표 3-2> 3개 권역별 디지털 문해능력 수준 정밀도 진단 결과

3개권역	표본수	수준1 비율	수준1 RSE	수준 1 로짓 CI 반폭	수준2~4 최대 RSE
서울 및 특광역시	4,575	6.1%	10.9%	1.3%p	6.9%
중소도시	3,521	8.6%	13.9%	2.4%p	7.1%
농산어촌	1,966	12.7%	15.9%	4.0%p	9.2%

6개 권역별 디지털 문해능력 수준 정밀도 진단 결과를 <표 3-3>에 제시하였다. 6개 권역 중 서울, 경기·인천, 충청권, 호남권, 영남권은 수준 1의 절대정밀도 기준과 수준 2~4 핵심범주의 RSE 기준을 모두 충족하였다. 반면 강

원·제주 권역은 수준 2~4의 최대 RSE가 17.0%로 비교적 양호하나, 수준 1의 로짓 신뢰구간 반폭이 7.3%p로 절대정밀도 기준을 초과하였다. 이는 강원·제주 권역의 표본수가 상대적으로 작고 수준 1 비율의 불확실성이 크게 나타난 결과로 해석된다. 따라서 6개 권역 단위 공표는 전반적으로 가능하나 강원·제주 권역의 수준 1 비율은 공표 시 주석을 병기하거나 해석에 주의할 필요가 있다.

〈표 3-3〉 6개 권역별 디지털 문해능력 수준 정밀도 진단 결과

6개권역	표본수	수준1 비율	수준1 RSE	수준 1 로짓 CI 반폭	수준2~4 최대 RSE
서울	1,165	8.2%	15.5%	2.5%p	11.3%
수도권	1,946	10.5%	14.8%	3.1%p	9.4%
충청권	1,720	7.2%	23.3%	3.4%p	10.0%
전라권	1,485	11.1%	16.4%	3.6%p	12.1%
경상권	2,960	3.1%	21.9%	1.4%p	7.4%
기타	786	15.5%	23.8%	7.3%p	17.0%

시도별 디지털 문해능력 수준 정밀도 진단 결과는 〈표 3-4〉와 같다. 시도 단위에서는 수준 2~4 핵심범주가 대부분 안정적으로 추정되었으나 수준 1은 지역별 비율과 표본수 차이로 인해 정밀도 편차가 크게 나타났다. 분석 결과 서울, 대구, 인천, 광주, 경기, 경북, 경남은 수준 1 절대정밀도 기준과 수준 2~4 RSE 기준을 모두 충족하여 디지털 문해능력 수준별 진단 결과를 공표할 수 있는 것으로 판단된다.

한편 부산과 울산은 수준 1 RSE가 각각 99.9%, 68.6%로 매우 높게 나타났는데 이는 수준 1 비율이 각각 0.2%, 1.1%로 매우 작은 데서 비롯된 현상으로 추정된다. 실제로 로짓 신뢰구간 반폭은 각각 0.8%p, 1.9%p로 절대정밀도는 양호한 수준이다. 다만 두 지역 모두 수준 1의 비가중 표본수가 10명 미만이므로 공표 시 주석을 병기할 필요가 있다.

반면 강원, 충북, 충남, 전북, 전남, 제주 등은 수준 2~4의 상대정밀도는 대체로 양호하나 수준 1의 로짓 신뢰구간 반폭이 5%p를 초과하였다. 따라서 해

당 시도의 수준 1 비율은 제한 공표하거나 해석에 주의할 필요가 있다. 세종은 수준 2~4의 최대 RSE가 34.8%로 30%를 넘고 수준 1 표본수도 2명에 그쳐 4범주 전체 표를 그대로 공표하기보다는 수준 1과 수준 2, 3, 4를 통합한 2범주 지표 형태로 공표하는 방안이 적절할 것으로 판단된다.

〈표 3-4〉 시도별 디지털 문해능력 수준 정밀도 진단 결과

시도	표본수	수준1 표본수	수준1 비율	수준1 RSE	수준1 로짓 CI 반폭	수준2~4 최대 RSE
서울	1,165	81	8.2%	15.5%	2.5%p	11.3%
부산	680	1	0.2%	99.9%	0.8%p	21.0%
대구	584	37	5.3%	31.0%	3.4%p	17.3%
인천	620	54	9.6%	22.9%	4.4%p	24.3%
광주	466	37	5.7%	30.0%	3.5%p	19.5%
대전	460	19	5.2%	44.8%	5.1%p	27.4%
울산	380	4	1.1%	68.6%	1.9%p	16.6%
세종	220	2	0.6%	101.7%	2.2%p	34.8%
경기	1,326	142	10.8%	17.1%	3.7%p	10.4%
강원	485	118	17.7%	28.8%	10.1%p	22.5%
충북	480	44	8.6%	33.0%	5.8%p	18.8%
충남	560	33	8.6%	40.0%	7.1%p	19.3%
전북	508	75	7.0%	36.5%	5.3%p	24.8%
전남	511	137	19.8%	21.8%	8.5%p	22.4%
경북	636	47	5.5%	34.3%	3.9%p	12.5%
경남	680	20	3.0%	57.5%	4.0%p	20.6%
제주	301	46	10.1%	28.3%	5.8%p	19.3%

2. 시도별 공표를 위한 적정 표본 규모 산출

목표 RSE를 달성하기 위한 필요 표본규모는 RSE가 표본수의 제곱근에 반비례하는 성질을 이용해 산출하였다.

$$n_{\text{목표}} = n_{\text{현행}} \left(\frac{RSE_{\text{현행}}}{RSE_{\text{목표}}} \right)^2$$

수준 2·3·4에 대하여 $RSE \leq 25\%$ 를 목표로 하면 대부분의 시도는 현행 표본으로 충분한 것으로 나타났다. 표본 증원이 필요한 곳은 세종과 대전 정도

이며, 세종은 약 220명에서 430명으로 약 210명, 대전은 약 460명에서 555명으로 약 95명의 추가 표본이 필요한 것으로 산출되었다.

따라서 수준 2·3·4 기준으로는 시도 단독 공표가 전반적으로 가능하다고 볼 수 있으며 이 경우 공표 가능성은 표본규모 자체보다 수준 1을 어떻게 평가하고 표기할 것인지가 더 중요한 쟁점으로 판단된다. 반면 수준 1까지 $RSE \leq 25\%$ 기준을 강제하면 필요한 표본규모가 과도하게 증가한다. 전국 기준으로 약 31,560명이 필요하며, 부산의 경우 약 10,900명이 필요한 것으로 산출되어 현실적인 조사 규모로 보기 어렵다.

이는 수준 1의 RSE 팽창이 단순한 표본 부족 때문이라기보다는 수준 1 비율 자체가 작은 데서 비롯된 구조적 현상임을 보여준다. 따라서 수준 1은 RSE 목표를 일률적으로 적용하기보다는 절대 신뢰구간 반폭 기준으로 관리하는 것이 타당하다. 즉, 수준 2·3·4는 RSE 기준을 중심으로 판단하고 수준 1은 로짓 신뢰구간 반폭 등 절대정밀도 기준으로 별도 평가하는 방식이 적절하다고 판단된다.

3. 지역×연령대 공표 가능성 검토

지역과 6개 연령대의 교차 공표 가능성을 검토한 결과는 <표 3-5>에 제시하였다. 지역과 연령대를 교차하면 공표단위가 세분화되어 각 셀의 표본수가 급격히 감소하며 특히 수준 1이 관측되지 않는 셀이 다수 발생한다. 분석 결과 시도×6개 연령대에서는 전체 102개 조합 중 수준 1 절대정밀도 기준과 수준 2·3·4 RSE 기준을 모두 충족한 조합이 2개에 그쳤다. 또한 수준 1 무관측 셀은 47개로 나타났다.

따라서 시도×6개 연령대 공표단위에서는 현재 표본규모로 디지털 문해능력 진단 결과의 4범주 전체표를 공표하기 어렵다. 3개 권역×6개 연령대 및 6개 권역×6개 연령대는 시도×6개 연령대보다 상대적으로 안정적이지만 이 경우에도 모든 조합에서 4범주 전체표를 공표하기에는 한계가 있다. 따라서 연령 교차 공표가 필요한 경우에는 6개 연령대보다 3개 연령대를 사용하는

방안 혹은 결과변수를 2범주 지표로 단순화하는 방안을 검토할 필요가 있다.

표본규모 측면에서도 시도×6개 연령대 전면 공표를 위해서는 목표 표본크기가 약 43,748명으로 산출되어 현행 표본의 약 4배 규모가 필요하다. 이는 현실적인 조사 운영 측면에서 부담이 큰 수준이다. 반면 6개 권역×3개 연령대는 목표 표본크기가 약 12,256명으로 산출되어 현행과 비슷하거나 소폭 증가한 표본규모를 통해 공표 가능성이 있을 것으로 보인다.

따라서 지역×연령 교차 공표가 필요한 경우, 6개 권역×3개 연령대가 통계적 안정성과 조사 현실성을 함께 고려한 가장 효율적인 절충안으로 판단된다.

〈표 3-5〉 지역×6개 연령대 교차 공표 가능성 요약

공표단위	전체 조합	수준2~4 최대 RSE<25	수준1 로짓 반폭≤5%p	두 기준 모두 충족	수준1 무관측	목표 표본크기
시도 × 6연령	102	11	19	2	47	43,748
시도 × 3연령	51	10	15	4	16	26,951
6권역 × 6연령	36	11	13	8	10	19,787
6권역 × 3연령	18	12	8	7	3	12,256
3권역 × 6연령	18	10	10	7	3	18,401
3권역 × 3연령	9	7	6	5	1	11,931

제3절 2범주 지표 기준 공표 가능성 검토

앞 절에서는 디지털 문해능력 수준을 수준1~수준4의 4개 범주로 구분하여 공표단위별 정밀도를 진단하였다. 그 결과 수준2~4는 대체로 안정적이었지만 비율이 작은 수준1은 상대표준오차가 구조적으로 팽창하여 시도 및 교차 공표단위에서 4범주 전체표를 그대로 공표하기 어려운 경우가 확인되었다.

이에 본 절에서는 디지털 문해능력을 수준1~3(일상생활 활용에 제약이 있는 수준)과 수준4(능숙하게 활용하는 수준)로 통합한 지표를 기준으로 동일한 공표단위별 정밀도를 재진단하였다. 2범주 지표는 두 범주의 비율이 모두 상대적으로 크기 때문에 소수 범주의 비율이 0에 가까워질 때 발생하는 RSE 팽창

문제를 완화할 수 있다.

1. 지역별 상대표준오차 분석결과

다음 <표 3-6>에는 응답자 전체에 대한 2범주 지표 기준 추정 결과가 제시되어 있다. 전국 기준 수준4 비율은 52.8%, 수준1~3 비율은 47.2%로 나타나, 성인의 절반가량이 디지털 기기·기술을 능숙하게 활용하는 수준(수준4)에 해당하는 것으로 확인되었다.

<표 3-6> 응답자 전체 추정 결과(2범주 기준)

디지털문해능력	표본수	추정비율(%)
수준1~3	5,548	47.2
수준4	4,514	52.8
합계	10,062	100.0

3개 권역별 결과는 <표 3-7>과 같다. 세 권역 모두 두 범주의 RSE가 최대 7.9%(농산어촌 수준4)에 그쳐 2범주 지표 기준으로 모든 권역에서 안정적으로 공표할 수 있는 것으로 나타났다. 수준4 비율은 서울 및 특광역시 55.7%, 중소도시 53.7%, 농산어촌 42.7%로 도시화 수준이 낮을수록 낮아지는 경향을 보여, 지역 간 디지털 문해능력 격차를 확인할 수 있다.

<표 3-7> 3개 권역별 디지털 문해능력 수준 정밀도 진단 결과(2범주 기준)

3개권역	수준1~3		수준4		현행 표본수	목표 표본수
	비율(%)	RSE(%)	비율(%)	RSE(%)		
서울 및 특광역시	44.3	3.9	55.7	3.1	4,575	4,575
중소도시	46.3	4.3	53.7	3.7	3,521	3,521
농산어촌	57.3	5.9	42.7	7.9	1,966	1,966
합계	47.2	-	52.8	-	10,062	10,062

6개 권역별 결과는 <표 3-8>과 같다. 모든 권역에서 두 범주의 RSE가 최대

12.4%(강원·제주 수준4) 이하로 나타나 6개 권역 단위에서도 2범주 지표를 안정적으로 공표할 수 있는 것으로 판단된다. 권역별로는 강원·제주가 수준 1~3 비율 64.9%로 가장 높고 수준4 비율 35.1%로 가장 낮았던 반면 서울(58.9%)과 경기·인천(58.4%)은 수준4 비율이 가장 높게 나타났다.

〈표 3-8〉 6개 권역별 디지털 문해능력 수준 정밀도 진단 결과(2범주 기준)

6개권역	수준1~3		수준4		현행 표본수	목표 표본수
	비율(%)	RSE(%)	비율(%)	RSE(%)		
서울	41.1	6.8	58.9	4.7	1,165	1,165
경기·인천	41.6	5.9	58.4	4.2	1,946	1,946
충청	59.0	6.0	41.0	8.6	1,720	1,720
호남	57.2	6.1	42.8	8.2	1,485	1,485
영남	46.8	5.0	53.2	4.4	2,960	2,960
강원·제주	64.9	6.7	35.1	12.4	786	786
전국	47.2	-	52.8	-	10,062	10,062

시도별 결과는 〈표 3-9〉와 같다. 17개 시도 모두 두 범주 중 비율이 작은 쪽을 기준으로 한 최대 RSE가 25% 이내로 나타났으며, 세종이 23.1%로 가장 높았다. 따라서 2범주 지표 기준으로는 17개 시도 전체를 현행 표본규모로 공표할 수 있으며 별도의 표본 증원이 필요하지 않은 것으로 산출되었다(목표 표본수 = 현행 표본수).

이는 4범주 기준에서 수준1의 RSE 팽창으로 일부 시도가 제약을 받고 대전·세종 등에서 증원이 요구되었던 것과 대비되며 결과변수를 2범주로 단순화하면 시도 단위의 공표 가능 범위가 크게 확대됨을 보여준다. 다만 세종은 수준1~3 비율이 27.3%로 작아 해당 범주의 RSE가 상대적으로 높게 나타났으나, 절대정밀도 측면에서는 양호한 수준이라 판단된다.

〈표 3-9〉 시도별 디지털 문해능력 수준 정밀도 진단 결과(2범주 기준)

시도	수준1~3		수준4		최대 RSE(%)	현행 표본수	목표 표본수
	비율(%)	RSE(%)	비율(%)	RSE(%)			

시도	수준1~3		수준4		최대 RSE(%)	현행 표본수	목표 표본수
	비율(%)	RSE(%)	비율(%)	RSE(%)			
서울	41.1	6.8	58.9	4.7	6.8	1,165	1,165
부산	34.6	12.6	65.4	6.7	12.6	680	680
대구	48.0	10.7	52.0	9.9	10.7	584	584
인천	53.3	9.9	46.7	11.3	11.3	620	620
광주	50.6	12.7	49.4	13.1	13.1	466	466
대전	46.7	14.5	53.3	12.7	14.5	460	460
울산	64.2	9.0	35.8	16.1	16.1	380	380
세종	27.3	23.1	72.7	8.7	23.1	220	220
경기	39.0	7.1	61.0	4.6	7.1	1,326	1,326
강원	69.7	8.4	30.3	19.2	19.2	485	485
충북	63.9	10.2	36.1	18.0	18.0	480	480
충남	69.0	8.7	31.0	19.3	19.3	560	560
전북	53.2	9.7	46.8	11.0	11.0	508	508
전남	66.6	9.9	33.4	19.7	19.7	511	511
경북	55.6	7.7	44.4	9.6	9.6	636	636
경남	45.9	12.7	54.1	10.8	12.7	680	680
제주	53.4	9.3	46.6	10.7	10.7	301	301
전국	47.2	-	52.8	-	23.1	10,062	10,062

2. 지역×연령대 공표 가능성 검토

지역과 연령대의 교차 공표 가능성은 <표 3-10>과 같다. 각 셀에서 두 범주 중 큰 RSE를 기준으로 $RSE \leq 25\%$ 를 충족하는 셀의 비율을 검토하였다. 분석 결과 3개 권역×3개 연령대는 9개 셀 전체가 기준을 충족하고 최대 RSE가 19.0%로 나타나 현행 표본규모만으로 완전한 교차 공표가 가능한 것으로 확인되었다. 다음으로 6개 권역×3개 연령대는 18개 중 16개 셀이 기준을 충족하였으며 최대 RSE는 32.2%, 목표 표본규모는 약 10,372명으로 산출되어 현행 대비 소폭 증가하는 수준이었다.

반면 시도×3개 연령대는 51개 중 29개 셀만 기준을 충족하였고 목표 표본규모가 약 16,091명(현행의 약 1.6배)으로 산출되었다. 또한 연령대를 6개 구간으로 세분할 경우 셀당 표본이 급감하여 시도×6개 연령대에서는 102개 중

39개 셀만 충족하고 목표 표본규모가 약 22,301명(현행의 약 2.2배)이 필요한 것으로 산출되었다.

이처럼 2범주 지표는 동일 공표단위에서 4범주 기준 대비 필요 표본규모를 약 절반 수준으로 낮추지만, 연령 교차 공표를 위해서는 여전히 연령을 3개 구간으로 통합하고 지역을 3개 또는 6개 권역으로 두는 것이 현실적이라 생각된다. 공표단위별 상세 셀 결과는 부록 표 1과 2에 제시하였다.

〈표 3-10〉 지역×연령대 교차 공표 가능성 요약(2범주 기준)

공표단위	전체 셀	RSE≤25 충족 셀	최대 RSE(%)	현행 표본	목표 표본
시도 × 6연령	102	39	106.2	10,062	22,301
시도 × 3연령	51	29	66.3	10,062	16,091
6권역 × 6연령	36	26	80.0	10,062	12,851
6권역 × 3연령	18	16	32.2	10,062	10,372
3권역 × 6연령	18	16	54.4	10,062	12,321
3권역 × 3연령	9	9	19.0	10,062	10,062

제4절 소결

지역별 공표와 관련하여, 4범주 기준에서는 3개 권역과 6개 권역에서 디지털 문해능력 4개 범주를 모두 안정적으로 공표할 수 있다. 시도 단독 공표는 핵심 범주(수준2·3·4)를 기준으로 대부분의 시도에서 현행 표본으로 가능하며, 세종(약 210명)과 대전(약 95명)만 소폭 증원하면 목표 RSE 25%를 달성할 수 있다. 다만 비율이 작은 수준1은 RSE가 구조적으로 팽창하므로 RSE만으로 공표 여부를 판단해서는 안 되며, RSE가 25%를 초과하더라도 절대 95% 신뢰구간 반폭이 5%p 이내이면 주석과 신뢰구간을 함께 제시하여 공표하는 방식이 바람직하다.

2범주 지표 기준에서는 공표 가능 범위가 크게 넓어진다. 디지털 문해능력을 수준1~3과 수준4로 통합하면 소수 범주에서 발생하는 RSE 팽창이 완화되

어, 전국, 3개 권역, 6개 권역은 물론 17개 시도 전체를 현행 표본규모로 안정적으로 공표할 수 있다. 또한 3개 권역×3개 연령대 교차는 현행 표본만으로 9개 셀 전체가 기준($RSE \leq 25\%$)을 충족한다. 2범주 지표는 동일 공표단위에서 4범주 대비 필요 표본규모를 약 절반 수준으로 낮추므로 세부 공표 수요에 대응하는 현실적 대안이 된다.

지역×연령 교차 공표와 관련하여, 4범주 전체표 기준으로는 시도×6개 연령대 교차 공표가 102개 셀 중 2개만 가능하고 목표 표본규모가 약 4.4만 명(현행의 약 4배)으로 현실성이 매우 낮게 나타났다. 2범주 지표를 적용하면 같은 단위에서 39개 셀이 기준을 충족하고 목표 표본규모도 약 2.2만 명(현행의 약 2.2배)으로 낮아지지만 여전히 부담이 크다.

따라서 교차 공표가 필요할 경우에는 지역을 3개 또는 6개 권역으로 두거나(6개 권역×3개 연령대 기준 약 1.0만~1.2만 명), 시도 공표를 고려하려면 연령을 3개 구간으로 통합(2범주 기준 시도×3개 연령대 약 1.6만 명)하는 절충안이 적절하다고 판단된다.

향후 과제와 관련하여, 차기 조사에서 시도×연령 교차 공표를 목표로 한다면 표본규모를 대폭 확대하거나 공표 목표를 권역 단위로 재설정할 필요가 있다. 또한 비율 추정 정밀도 평가는 RSE와 절대 신뢰구간 반폭을 병행하고, 다범주 변수는 핵심 범주의 최대 RSE를 구속 기준으로 삼는 종합 판정 틀을 표준화할 필요가 있다고 판단된다.

제4장 표본대체 및 무응답 실태 분석

제1절 분석 개요

본 장에서는 2024년 성인디지털문해능력조사의 실사 과정에서 발생한 표본 대체와 무응답 구조를 종합적으로 검토한다. 본 조사는 본표본 가구가 조사되지 못하는 경우 동일 조사구 내 예비표본 가구로 대체하는 방식을 적용하고 있으므로, 표본대체가 얼마나 발생하였고 대체표본이 본표본을 적절히 보완하고 있는지, 무응답이 어느 단계에서 어떤 사유로 발생하였는지를 실사 운영의 관점에서 진단할 필요가 있다.

이를 위해 먼저 조사구 및 가구 단위의 표본대체 현황을 정리하고(제2절), 본표본과 예비표본의 분포 특성을 주요 효과크기와 로지스틱 회귀분석 결과를 중심으로 요약한다(제3절). 본표본과 예비표본 비교의 상세한 변수별 분석결과와 표는 부록 2에 제시한다. 최종 조사결과 코드, 컨택횟수, 1차 컨택결과에 따른 조사완료 가능성 등 무응답 파라미터를 분석하고(제4절), 유사 국가통계조사의 무응답 대응 및 표본대체 방식을 비교한다(제5절).

본 장의 분석은 무응답의 원인을 인과적으로 규명하기보다는 현장 조사 운영 과정에서 확인된 표본대체와 무응답 양상을 정리하고, 이를 토대로 제5장의 조사운영 및 활용 체계 개선방안을 도출하는 데 목적이 있다.

제2절 조사구 및 가구대체 현황

1. 조사구 단위 대체현황

본 조사는 표본 운영 과정에서 조사구 단위 대체와 가구 단위 대체를 병행하였다. 조사구 단위에서는 현장조사 진행이 사실상 어려운 5개 조사구에 대해 사전 대체가 이루어졌다. 대체 사유는 아파트 관리사무소의 조사 거절에 따른

출입불가, 관사 아파트로 인한 일반인 출입 제한, 별장 밀집지역의 실거주자 부족, 외국인 숙소 밀집으로 인한 내국인 적격가구 부족 등으로 나타났다.

〈표 4-1〉 조사구 대체 현황

지역	구분	조사구코드	대체 사유
대구	변경전	22*****9A	아파트 관리사무소에서 강력 조사 거절로 출입불가
	변경후	22*****2A	
	변경전	22*****8A	아파트 관리사무소에서 강력 조사 거절로 출입불가
	변경후	22*****5A	
강원	변경전	32*****4A	관사 아파트로 일반인 출입불가
	변경후	32*****4A	
제주	변경전	39*****01	별장 밀집지역으로 실거주자가 거의 없어 조사 불가
	변경후	39*****31	
	변경전	39*****41	양식장 및 외국인 거주 숙소가 대부분으로 내국인이 거의 없어 조사 불가
	변경후	39*****21	

이들 사례는 개별 가구의 응답 거절이라기보다 조사구 접근성 또는 조사구 적격성 문제에 해당한다. 따라서 동일 시도 내 대체 조사구를 활용한 것은 실사 운영상 불가피한 조치로 판단된다. 다만 조사구 대체는 가구 단위 대체보다 표본 대표성에 미치는 영향이 클 수 있으므로 향후에는 대체 사유를 표준화하여 기록하고 대체 조사구 선정 기준을 명확히 관리할 필요가 있다. 예를 들어 관리사무소 거절, 출입 제한, 실거주자 부족, 내국인 적격가구 부족 등으로 구분하면 조사구 수준의 접근성 문제와 표본틀 적격성 문제를 구분하여 점검할 수 있다.

2. 가구 단위 표본대체 현황

가구 단위에서는 본표본 가구에서 조사완료가 이루어지지 않을 경우 동일 조사구 내 사전 지정된 예비표본 가구를 순차적으로 활용하였다. 전체 접촉가구는 11,593건이며 이 중 조사완료는 5,030건(표본 수)으로 전체 조사완료율은 43.4%이다. 본표본의 접촉가구 수는 4,963건, 조사완료 건수는 1,683건으로 조

사완료율은 33.9%였다. 예비표본에서는 예비1 1,257건, 예비2 1,204건, 예비 3 637건순으로 조사완료가 이루어졌다.

〈표 4-2〉 가구 단위 대체 현황

구분	접촉가구 건수	조사완료 건수	조사완료율(%)
본표본	4,963	1,683	33.9
예비1	3,322	1,257	37.8
예비2	2,081	1,204	57.9
예비3	881	637	72.3
예비4	247	190	76.9
예비5	58	37	63.8
예비6	22	14	63.6
예비7	9	6	66.7
예비8	2	2	100
예비번호 미상	8	0	0
합계	11,593	5,030	43.4

〈표 4-2〉와 〈표 4-3〉에는 본표본과 예비표본별로 접촉가구 건수와 조사완료율이 제시되어 있다. 전체 조사완료 5,030가구 중 본표본에서 확보된 조사완료는 1,683가구로 33.5%이며 예비표본에서 확보된 조사완료는 3,347가구로 66.5%이다. 개인 기준으로 보더라도 전체 응답자 10,062명 중 본표본 응답자는 3,361명으로 33.4%, 예비표본 응답자는 6,701명으로 66.6%를 차지하였다.

이 결과는 본 조사에서 예비표본이 단순한 보조 수단이 아니라 목표 조사완료 규모를 확보하기 위한 핵심적인 운영 장치로 활용되었다. 따라서 향후 조사에서는 예비표본을 충분히 확보하는 것뿐 아니라 본표본 종결 기준, 예비표본 투입 순서, 대체 사유 기록을 명확히 관리할 필요가 있다. 또한 본표본과 예비표본의 응답자 특성이 다를 가능성이 있으므로 사후적으로 두 집단 간 차이를 지속적으로 모니터링하는 절차가 필요하다.

〈표 4-3〉 본표본과 예비표본 응답 현황

구분	전체 응답	본표본 응답	예비표본 응답	본표본 비율(%)
가구	5,030	1,683	3,347	33.5
개인	10,062	3,361	6,701	33.4

제3절 본표본과 예비표본 분포 특성 비교

1. 검토 개요

현행 조사는 본표본 가구가 무응답, 접촉불능 등으로 조사되지 못하는 경우 예비표본 가구로 대체하는 방식을 적용하고 있다. 2024년 조사에서 전체 실현 표본 10,062명 중 예비표본 응답자가 6,701명으로 66.6%를 차지하여 예비표본 활용 비중이 높은 것으로 나타났다. 예비표본 의존도가 높은 상태에서 예비표본 응답자의 특성이 본표본 응답자와 다르게 나타날 경우 전체 표본의 대표성에 영향을 미칠 수 있으므로 본 절에서는 본표본과 예비표본의 분포 특성을 비교하여 표본대체가 조사결과에 미치는 영향을 검토하였다.

본표본과 예비표본 간 차이는 조사수행 특성, 지역 및 가구 특성, 인구·사회경제적 특성, 조사결과 특성을 중심으로 검토하였으며 분석에서는 단순한 통계적 유의정보보다는 효과크기를 중심으로 해석하였다. 표본 수가 충분히 큰 경우 실제 차이가 작더라도 검정의 p값이 유의하게 나타날 수 있으므로 범주형 변수는 Cramér's V, 연속형 변수는 Cohen's d를 활용하여 두 집단 간 차이의 실질적 크기를 함께 판단하였다. 상세한 변수별 분포 비교 결과와 효과크기 산출 결과는 부록 2에 제시되어 있다.

2. 효과크기 중심 주요 결과

본표본과 예비표본의 주요 변수별 차이를 효과크기 중심으로 요약하면 〈표

4-4>와 같다. 가구 단위 분석 결과 가구원수, 가구유형, 소득, 조사요일, 조사시간대 등 대부분의 변수에서 본표본과 예비표본은 유사한 분포를 보였다. 다만 시도와 6개 권역에서는 작은 수준의 차이가 나타났으며 시도별 본표본 비율에서는 울산(14.2%)과 인천(44.8%)의 경우와 같이 비교적 큰 차이가 나타나 예비표본 활용 정도가 지역별로 다르게 나타났음을 확인할 수 있다. 이러한 차이는 본표본 가구의 응답률 또는 접촉성공률이 지역별로 다르게 나타난 데서 비롯되었을 가능성이 있다.

〈표 4-4〉 본표본 · 예비표본 비교 주요 결과 요약

검토 범주	주요 분석 변수	효과크기	값
조사수행 특성	조사 소요시간	Cohen's d	<0.05
	조사요일(주중/주말)		0.004
	조사시간대		0.040
지역 및 가구 특성	시도		0.186
	권역(6개 권역)	Cramér's V	0.120
	권역(3개 권역)		0.014
	가구유형		0.024
인구 · 사회경제적 특성	성별		0.007
	연령		0.027
	학력		0.029
	가구소득		0.038
	경제활동상태		0.019
	생활만족도		0.032
	정치적 관심		0.040
	스마트폰 유형		0.031
조사결과 특성	디지털 문해능력 수준		0.041

가구원 단위 분석에서도 주요 인구 · 사회경제적 변수 전반에서 본표본과 예비표본 간 효과크기는 매우 작은 수준으로 나타났다. 성별, 연령, 학력, 경제활동상태, 생활만족도, 정치적 관심, 스마트폰 유형 등 대부분의 변수에서

Cramér's V가 0.05 미만으로 나타나 실질적인 차이는 거의 없는 것으로 판단된다. 특히 본 조사의 핵심 결과변수인 디지털 문해능력 수준에서도 Cramér's V는 0.041로 무시 가능한 수준이었으며 최대 범주 간 비율 차이도 2.9%p에 그쳤다.

따라서 단변량 비교 결과만으로 볼 때 예비표본은 본표본과 전반적으로 유사한 특성을 보이고 있으며 표본대체가 핵심 조사결과에 미치는 영향은 제한적인 것으로 판단된다. 다만 시도별 본표본 비율의 차이가 비교적 크게 나타났으므로 향후 조사에서는 지역별 본표본 완료율과 예비표본 사용 현황을 별도의 실사관리 지표로 지속적으로 점검할 필요가 있다.

3. 로지스틱 회귀분석 결과

앞에서 확인한 단변량 비교는 변수별로 본표본과 예비표본의 차이를 각각 살펴본 결과이다. 그러나 각 변수들은 서로 관련되어 있을 수 있으므로 변수들 간의 상관을 통제한 상태에서 본표본 여부와 관련된 요인을 추가로 검토할 필요가 있다. 이에 본표본 여부를 종속변수로 하는 개인 단위 로지스틱 회귀분석을 수행하였으며 그 결과는 <표 4-5>와 같다.

<표 4-5> 로지스틱 회귀분석 주요 결과 요약

구분	항목	값
분석 개요	분석 표본	개인 단위 응답자 10,062명
모형 적합도·분리력	LR χ^2 (df=55)	452.4(p<0.001)
	pseudo-R ²	0.0353
	AUC	0.628
유의한 변수 (결합 Wald, p< 0.05)	시도	p<0.001
	조사시간대	p<0.001
	정치적 관심	p<0.001
	가구유형	p=0.025
	스마트폰 유형	p=0.045

구분	항목	값
주요 지역 차이 (시도 OR, 기준=서울)	울산	0.273
	경남	0.417
	부산	0.446
	제주	0.447
	인천	1.344

로지스틱 회귀분석 결과 모형의 AUC는 0.628, pseudo- R^2 는 0.0353으로 나타났다. 이는 여러 변수를 동시에 고려하더라도 본표본과 예비표본을 뚜렷하게 구분하기 어렵다는 뜻으로 해석된다. 통계적으로는 일부 변수가 유의하게 나타났으나 이는 표본수가 큰 상황에서 검정이 민감하게 반응한 결과로 해석할 수 있으며 모형의 설명력과 분리력 자체는 낮은 수준이었다.

변수별로는 시도가 본표본 여부를 설명하는 가장 뚜렷한 변수로 나타났으며 이는 앞선 단변량 분석에서 확인된 시도별 본표본 비율 차이와 부합한다. 울산, 부산, 경남, 제주 등 일부 지역에서는 본표본 오즈가 상대적으로 낮게 나타나 예비표본 활용이 더 많이 발생한 것으로 해석할 수 있다. 반면 조사시간대, 정치적 관심, 가구유형, 스마트폰 유형 등에서도 일부 유의한 차이가 나타났으나 효과크기는 미미하여 본표본과 예비표본을 구분하는 주요 요인으로 보기는 어렵다. 로지스틱 회귀분석의 상세한 변수별 p값과 회귀계수 추정 결과는 부록 2에 제시하였다.

4. 소결

본표본과 예비표본의 차이를 주요 검토 변수별로 살펴본 결과 시도와 6개 권역에서만 작은 수준의 차이가 나타났으며 그 외 변수에서는 효과크기가 모두 무시 가능한 수준으로 확인되었다. 특히 핵심 결과변수인 디지털 문해능력 수준의 경우 Cramér's $V=0.041$ 로 나타나 본표본과 예비표본 간 실질적인 차이는 거의 없는 것으로 판단된다. 또한 다변량 로지스틱 회귀분석 결과에서도 모형의 분리력은 AUC=0.628로 낮은 수준으로 나타났다.

따라서 예비표본은 전반적으로 본표본을 적절히 대체하고 있으며 최종 표본의 대표성에 미치는 실질적 영향은 제한적인 것으로 판단된다. 다만 시도별 본표본 비율은 14%대에서 45% 수준까지 비교적 큰 차이를 보였는데 이는 시도별 본표본 응답률 또는 접촉 성공률의 차이가 예비표본 활용 정도에 반영된 결과로 해석될 수 있다. 이에 따라 향후 본조사에서는 본표본 응답률이 상대적으로 낮은 시도에 대해 실사 자원을 우선적으로 배분하고 접촉 관리를 강화하는 방안을 검토할 필요가 있다. 또한 차기 조사에서는 시도별 예비표본 추출 및 활용 기록을 체계적으로 보존하여 본표본과 예비표본 간 차이를 지속적으로 모니터링이 권장된다.

제4절 무응답 파라데이터 분석

1. 접촉가구의 응답결과 및 응답률 현황

파라데이터에서 최종 조사결과 코드는 분석 목적에 따라 응답, 불응, 불능, 비접촉, 부적격의 5개 범주로 재분류되며 그 결과는 <표 4-6>와 같다. 전체 11,593건 중 조사 완료는 5,030건(43.4%)이고 미완료는 6,563건(56.6%)이다. 미완료 사유 중에서는 불응이 4,428건으로 가장 많았으며 이는 전체 미완료의 약 67.5%에 해당한다.

<표 4-6> 응답 결과의 범주별 집계

범주	정의	빈도(가구)	비율(%)
응답	조사 완료	5,030	43.4
불응	접촉된 가구 중 1번 이상 조사를 거부한 경우	4,428	38.2
불능	조사에 임할 수 없는 경우	534	4.6
비접촉	응답이 가능한 가구원을 만나지 못한 경우	752	6.5
부적격	표본 자체를 이용할 수 없는 경우	849	7.3
전체	—	11,593	100.0

범주별 집계를 바탕으로 응답률 관련 지표를 산출한 결과는 <표 4-7>와 같

으며 응답률(RR5)은 46.8%, 접촉률(CON3)은 93.0%, 협조율(COOP1)은 50.3%, 거부율(REF3)은 41.2%로 나타났다. 접촉률은 93.0%로 높은 반면 협조율은 50.3%에 머물렀다. 이는 본 조사의 무응답 문제가 가구 접촉 자체의 실패보다는 접촉 이후 조사 참여를 설득하는 과정에서 주로 발생했음을 의미한다.

따라서 무응답 대응 전략은 단순히 방문 횟수를 늘리는 방식보다 접촉 이후 협조를 유도할 수 있는 안내와 설득 체계를 강화하는 방향으로 설계될 필요가 있다.

〈표 4-7〉 응답률 지표

지표	산출식	값(%)
응답률(RR5)	응답 / (응답 + 불능 + 비접촉 + 불응)	46.8
접촉률(CON3)	(응답 + 불능 + 불응) / (응답 + 불능 + 비접촉 + 불응)	93.0
협조율(COOP1)	응답 / (응답 + 불능 + 불응)	50.3
거부율(REF3)	불응 / (응답 + 불능 + 비접촉 + 불응)	41.2

2. 접촉횟수 및 결과별 조사완료 현황

접촉횟수별 최종 조사완료율을 살펴본 결과는 〈표 4-8〉과 같으며 조사완료 가능성은 주로 1~2회 컨택 이내에서 결정되는 것으로 나타났다. 1회 컨택 사례의 최종 완료율은 76.28%, 2회 컨택 사례의 최종 완료율은 65.15%였다. 반면 3회 이상 컨택 사례의 완료율은 1% 미만으로 급격히 낮아졌다. 이 결과는 조사 초반 접촉의 품질관리가 중요함을 시사한다. 최초 방문 전 안내 강화, 조사원 신분 확인 절차 명확화, 첫 방문 시 조사 목적 및 개인정보 보호 설명 강화 등이 필요하다. 또한 3회 이상 반복 접촉의 효과가 제한적으로 나타났으므로 조사 결과 코드별로 재방문 지속 여부와 예비표본 전환 기준을 명확히 설정할 필요가 있다.

〈표 4-8〉 컨택횟수별 최종 조사완료율

접촉횟수	전체사례수	최종완료수	최종미완료수	최종완료율(%)
1회	4,575	3,490	1,085	76.28
2회	2,353	1,533	820	65.15
3회	1,126	3	1,123	0.27
4회	3,539	4	3,535	0.11
합계	11,593	5,030	6,563	43.38

1차 컨택에서 조사완료가 이루어지지 않은 사례를 대상으로 2차 컨택에서 조사완료로 전환된 비율을 살펴보면 1차 미완료 사유에 따라 후속 완료 가능성이 크게 달랐다. 대상자 부재의 2차 조사완료율은 51.77%, 불신은 50.07%, 바쁘거나 귀찮음은 32.66%로 나타났다. 반면 아무도 없음은 0.20%에 그쳤고 가정 사정, 사생활 노출 기피, 적격 응답자 없음, 빈집, 장기 부재 등은 2차 조사완료 사례가 확인되지 않았다.

〈표 4-9〉 1차 컨택 결과별 2차 조사완료 전환율

1차 컨택결과	1차 미완료	2차 완료	2차 실패	2차 미시도	전환율(%)
대상자 부재	1,524	789	722	13	51.77
불신	697	349	348	0	50.07
바쁘거나 귀찮음	1,191	389	773	29	32.66
아무도 없음	2,961	6	2,953	2	0.20
가정 사정(실업/이혼 등)	458	0	421	37	-
사생활 노출 기피	248	0	248	0	-
적격 응답자 없음	558	0	0	558	-
빈 집	150	0	0	150	-
장기 부재	123	0	0	123	-
강원·제주	193	0	20	173	-
합계	8,103	1,533	5,485	1,085	18.92

〈표 4-9〉은 1차 컨택에서 실패한 가구들을 대상으로 2차컨택 성공률을 나타낸 표이다. 이를 통해 1차 컨택결과를 후속 실사 전략에 적극적으로 활용할 필요가 있음을 알 수 있다. 대상자 부재는 시간대 변경 후 재방문이 효과적일 수

있고 불신은 공식 공문, 조사원 신분 확인, 개인정보 보호 안내를 통해 재설득할 필요가 있다. 바쁘거나 귀찮음의 경우 조사 소요시간 안내와 가능한 시간 예약이 중요하다. 반면 빈집, 장기 부재, 적격 응답자 없음 등은 반복 방문보다 조기 대체가 효율적일 수 있다. 따라서 향후에는 1차 컨택결과에 따라 재방문, 재설득, 조기 대체를 구분하는 운영 기준을 마련할 필요가 있다.

3. 최종 조사코드별 미완료 사유

최종 조사코드별 분포를 살펴본 결과는 <표 4-10>과 같으며 조사 완료가 5,030건(43.40%)으로 가장 많았다. 미완료 사유 중에서는 바쁘거나 귀찮음 1,690건(14.58%), 불신 1,456건(12.56%), 사생활 노출 기피 854건(7.37%), 대상자 부재 750건(6.47%) 순으로 나타났다.

<표 4-10> 최종 조사코드별 분포

최종 조사코드	사례수	비율(%)
조사 완료	5,030	43.40
바쁘거나 귀찮음	1,690	14.58
불신	1,456	12.56
사생활 노출 기피	854	7.37
대상자 부재	750	6.47
적격 응답자 없음	558	4.81
가정 사정(실업/이혼 등)	428	3.69
고령	256	2.21
장기 부재	245	2.11
빈 집	150	1.29
대상표본 아님	89	0.77
거주지 없음	52	0.45
강원·제주	35	0.30
합계	11,593	100.00

미완료 사유 중 상당 부분은 응답 부담과 조사 신뢰도 인식에서 비롯된 것

으로 볼 수 있다. 따라서 향후에는 조사 안내문에 공식 발신 주체, 개인정보 보호 절차, 평균 조사 소요시간을 명확히 제시할 필요가 있다. 특히 디지털 문해능력조사는 응답자가 ‘능력 평가’ 또는 ‘개인정보 확인’으로 오해할 가능성이 있으므로 조사 목적이 국가통계 생산에 있으며 개인별 평가 결과를 산출하거나 공개하는 조사가 아니라는 점을 명확히 설명할 필요가 있다. 한편 적격 응답자 없음, 대상표본 아님, 거주지 없음 등은 표본들의 시의성 문제와 관련될 수 있으므로, 표본추출틀의 최신성도 정기적으로 점검할 필요가 있다.

4. 가구 내 조사완료 및 초과조사 현황

〈표 4-11〉에는 접촉에 성공한 가구 내에서 개인별로 조사에 응한 비율이 제시되어 있다. 완료 가구 5,030가구의 적격가구원 합계는 10,607명이고 이 중 조사참여가구원은 10,062명으로 나타났다. 가구 내 개인 참여율은 94.86%이며, 전원 응답 가구는 4,587가구(91.19%), 일부 응답 가구는 443가구(8.81%)이다.

〈표 4-11〉 적격 가구원 중 조사 완료 현황

구분	가구원/가구 수	비율(%)
조사참여가구원수 합계	10,062명	94.86
미참여 적격가구원수	545명	5.14
적격가구원수 합계	10,607명	100.00
전원 응답 가구	4,587가구	91.19
일부 응답 가구	443가구	8.81
완료 가구	5,030가구	100.00

일단 가구 단위의 조사 협조가 이루어진 경우 가구 내 적격가구원 대부분이 조사에 참여한 것으로 나타났다. 따라서 본 조사의 주요 무응답 문제는 가구 내부의 개인 무응답보다는 가구 단위 조사 협조 확보 단계에서 발생한다고 볼 수 있다. 표본가구 내 적격가구원 전원조사는 목표 조사인원 확보에 유리하므로 현행 방식을 유지하는 것이 적절하다고 판단된다. 다만 다인가구에

서는 응답 부담이 높을 수 있으므로 조사 소요시간 사전 안내와 가구원별 조사 일정 조정 등 현장 운영 측면의 보완이 필요하다.

한편 조사구별 조사완료 현황을 점검한 결과 전체 500개 조사구 중 완료가구 수가 10가구 미만인 조사구는 확인되지 않았다. 다만 20개 조사구에서 완료가구 수가 10가구를 초과하였으며 초과조사 사례(가구)는 총 30건으로 나타났다으며, 자세한 분포는 <표 4-12>에 나타나 있다.

<표 4-12> 완료가구 수 구분별 조사구 현황

완료가구수	조사구수	가구합	초과가구
10가구	480	4,800	-
11가구	13	143	13
12가구	4	48	8
13가구	3	39	9
합계	500	5,030	30

초과표본이 발생한 조사구를 대상으로 가구별 조사완료 시간대를 파악하여 초과조사가 발생한 원인을 추정한 결과는 <표 4-13>과 같다. 초과조사 사례 중 8건은 본표본과 예비표본이 같은 일자에 함께 접촉되어 모두 조사완료가 이루어진 경우로 추정된다. 향후 조사구별 완료가구 수와 완료인원 수를 실시간으로 확인할 수 있는 관리 체계를 마련할 필요가 있다. 20명 인원 충족을 위한 추가조사 19건은 완료가구 수 10가구를 충족한 이후에도 조사구 내 조사인원이 목표치(20명)에 미달하여 추가조사를 진행한 사례로 추정된다. 다만 조사인원 부족은 사후가중 또는 보정 과정에서 처리하는 것이 원칙적으로 적절하므로 임의적 추가조사는 지양할 필요가 있다. 기타 3건은 자료상 명확한 사유 확인이 어려운 사례이다. 해당 조사구의 정원 도달 시점과 추가조사 진행 경위를 사후 점검할 필요가 있다.

〈표 4-13〉 초과표본 발생 원인별 분류

초과 사례	가구 수	비율(%)
같은 날짜 내 접촉으로 인한 초과	8	26.7
20명 인원 충족을 위한 추가조사	19	63.3
알수 없음	3	10.0
합계	30	100.0

제5절 유사통계 사례분석

본 절에서는 성인디지털문해능력조사와 유사한 국가승인통계 사례를 검토하여 본 조사의 표본설계 및 표본대체 방식의 특성을 검토한다. 비교 대상은 성인문해능력조사, 디지털정보격차실태조사, 전국민금융이해력조사, 인터넷이용실태조사 등 가구방문조사이다. 이들 조사는 모두 인구주택총조사 조사구를 표본추출틀로 활용하여 1차로 조사구를 추출한 뒤 2차로 조사구 내 가구를 추출하는 층화집락계통추출 방식을 사용한다. 이후 표본가구 내 조사대상 가구원에 대해서는 적격가구원 전원조사 또는 가구원 1인 선정 방식이 적용된다.

〈표 4-14〉 유사 국가통계조사의 무응답 대응 및 표본대체 방식 비교

조사명	주관기관	표본규모	주요 표본대체 방식
성인디지털문해능력조사	교육부· 국가평생교육진흥원	약 10,000명	동일 조사구 내 예비표본 별도 운영
성인문해능력조사			
디지털정보격차실태조사	과기정통부	약 15,000명	조사구 내 명부상 인접 가구
전국민금융이해력조사	한국은행·금감원	약 2,400명	본표본 위·아래 가구 사전 지정
인터넷이용실태조사	과기정통부	약 50,000명	명부상 인접 가구 우선 활용

〈표 4-15〉 유사 국가통계조사의 가구원 추출방식 비교

조사명	가구원 추출방식
성인디지털문해능력조사	적격가구원 전원조사
성인문해능력조사	
디지털정보격차실태조사	최근 생일자법(가구당 1인)
전국민금융이해력조사	가장 빠른 생일자(가구당 1인)
인터넷이용실태조사	적격가구원 전원조사

가구원 조사방식은 조사마다 차이가 있다. 성인디지털문해능력조사와 성인 문해능력조사, 인터넷이용실태조사는 표본가구 내 적격가구원 전원조사 방식을 적용한다. 반면 디지털정보격차실태조사와 전국민금융이해력조사는 가구 내 적격가구원 중 1인을 선정하여 조사한다.

본 조사는 성인문해능력조사와 같이 적격가구원 전원조사 방식을 채택하고 있어 기존 문해능력 관련 조사와 유사한 운영 구조를 가진다. 전원조사 방식은 목표 조사인원 확보에 유리하지만 다인가구의 경우 응답 부담이 높을 수 있으므로 조사 소요시간 안내와 가구원별 조사 일정 조정 등 현장 운영상 보완이 필요하다. 표본대체 방식은 조사별로 차이가 있으나 대체로 동일 조사구 내 예비표본 활용, 조사구 내 인접 가구 활용, 본표본 주변 가구 사전 지정 방식으로 구분된다. 본 조사는 동일 조사구 내 예비표본을 별도로 운영하는 방식으로 성인문해능력조사와 유사한 구조를 가진다.

따라서 본 조사의 표본설계와 표본대체 방식은 유사 가구방문조사의 일반적인 운영 구조와 거의 동일하다고 볼 수 있다. 다만 본 조사에서는 예비표본을 통한 조사완료 비중이 높게 나타났으므로 향후에는 본표본 종결 기준, 예비표본 투입 순서, 대체 사유 기록을 보다 체계적으로 관리할 필요가 있다.

제5장 조사운영 및 활용 체계 개선방안

제4장에서는 조사구 및 가구대체 현황, 본표본과 예비표본의 분포 특성, 무응답 파라데이터, 유사 국가통계 사례를 검토하였다. 분석 결과 성인디지털문해능력조사는 인구주택총조사 조사구를 표본추출틀로 활용한 층화집락추출구조를 기반으로 전국 단위 대표성을 확보하고 있으며, 표본가구 내 적격가구원 전원조사 방식을 통해 개인 단위 디지털 문해능력 수준을 파악하는 데 적절한 조사체계를 갖추고 있는 것으로 판단된다.

다만 가구방문조사의 환경 변화, 예비표본 활용 비중 증가, 일부 지역과 가구유형에서의 낮은 접촉 가능성, 1인 가구 과소대표 가능성 등을 고려할 때 향후 조사에서는 실사 운영, 표본대체 관리, 무응답 보정, 공표 및 활용 체계를 보다 체계적으로 개선할 필요가 있다.

본 장에서는 이러한 진단 결과를 바탕으로 조사운영 및 활용 체계 개선방안을 무응답 대응 및 조사원 교육, 표본대체 운영 및 현장 품질관리, 지역별 공표 확대 및 통계 활용의 세 측면으로 나누어 제시한다.

제1절 무응답 대응 및 조사원 교육 체계 개선

성인디지털문해능력조사는 전문 조사원에 의한 가구방문 면접조사 방식으로 수행되므로 표본가구 접촉 가능성이 조사 성공률에 직접적인 영향을 미친다. 그러나 제4장의 파라데이터 분석 결과 접촉률은 93.0%로 높은 반면 협조율은 50.3%로 본 조사의 무응답은 가구 접촉 자체의 실패보다 접촉 이후 조사 참여를 설득하는 과정에서 주로 발생하는 것으로 나타났다.

따라서 무응답 대응 전략은 단순히 방문 횟수를 늘리는 방식보다 접촉 이후 협조를 유도하는 안내·설득 체계를 강화하는 방향으로 설계될 필요가 있다. 또한 조사완료 가능성이 주로 1~2회 접촉 이내에서 결정되고 3회 이상 접촉의 효과가 제한적이었던 점을 고려할 때, 1차 접촉 결과를 단순 기록하는

데 그치지 않고 부재·거절·불신·빈집·장기부재 등 세부 결과별로 재접촉 전략을 차별화하는 것이 바람직하다. 예를 들어 부재 가구는 시간대를 달리한 재방문을 원칙화하고, 거절·불신 가구는 공식 안내문과 조사원 신분 확인, 개인정보 보호 절차 설명을 통해 재설득할 필요가 있으며, 빈집·장기부재·적격 응답자 없음 등은 반복 방문보다 조기 대체하는 등 1차 조사결과에 따라 운영을 달리하는 효율적인 차별화 실사전략이 필요하다.

아울러 조사원 교육 및 현장 지침을 정교화할 필요가 있다. 본 조사는 표본 가구 내 적격가구원 전원을 조사하는 방식이므로 조사원은 전원조사의 원칙, 가구 내 일부 무응답 발생 시 처리 방식, 예비표본 사용 기준, 가구원 명부 작성 방법을 정확히 이해해야 한다.

특히 적격가구원 일부만 응답한 경우의 가구완료 처리 기준, 미응답 가구원에 대한 정보 수집 범위, 본표본 종결 후 예비표본 전환 기준을 현장 지침으로 명확히 규정할 필요가 있다. 또한 디지털 문해능력조사는 응답자가 조사를 능력 평가나 개인정보 확인으로 오해할 가능성이 있으므로, 조사 목적이 국가 통계 생산에 있으며 개인별 평가 결과를 산출하거나 공개하는 조사가 아니라는 점을 안내문과 조사원 설명에 명확히 포함하여야 한다.

한편 1인 가구 등 가구구조 변화에 대응한 실사 운영이 필요하다. 제2장과 제4장의 분석 결과 조사자료의 1인 가구 비율은 인구주택총조사 기준에 비해 낮게 나타났는데, 이는 표본설계 단계의 배제보다는 1인 가구의 접촉 어려움과 단독 거주자의 조사거절·부재 등 실사 단계의 차별적 무응답에서 비롯되었을 가능성이 크다. 따라서 1인 가구가 많이 분포한 조사구에 대해서는 방문 시간대를 다양화하고 예비표본 대체 시 동일한 가구유형 내 대체 원칙을 검토할 필요가 있으며, 적격가구원이 많은 대가구에 대해서는 조사 소요시간 사전 안내와 가구원별 조사 일정 조정 등 응답부담을 완화하는 현장 운영이 병행되어야 한다.

제2절 표본대체 운영 및 현장 품질관리 고도화

본 조사는 본표본 가구가 조사되지 못하는 경우에는 예비표본 가구로 대체하는 방식을 적용하고 있다. 2024년 조사에서는 전체 응답자의 66.6%가 예비표본에서 확보될 만큼 예비표본 활용 비중이 높게 나타났다. 제4장의 분포 비교 결과 디지털 문해능력 수준 등 핵심 조사결과에서는 본표본과 예비표본 간 차이가 무시 가능한 수준으로 예비표본이 본표본을 전반적으로 적절히 대체하고 있는 것으로 판단된다. 다만 시도 및 권역 등 지역 변수에서는 일정한 차이가 확인되었고 시도별 본표본 비율의 편차가 컸으므로, 예비표본 대체가 반복적으로 많이 발생하는 지역에 대해서는 별도의 실사 관리가 필요하다.

향후 조사에서는 본표본 우선 조사 원칙을 명확히 유지하되 예비표본 사용 단계, 대체 사유, 대체 시점, 대체 후 완료 여부를 표준화된 방식으로 기록하고, 조사 종료 후에는 본표본과 예비표본의 지역·가구유형·성별·연령·주요 조사결과 분포를 정례적으로 비교하여 표본대체가 추정결과에 미치는 영향을 지속적으로 점검할 필요가 있다.

또한 무응답 파라미터와 표본관리파일을 체계적으로 구축할 필요가 있다. 무응답 보정 및 가중치 산출의 고도화를 위해서는 응답자뿐 아니라 무응답 가구와 미응답 가구원에 대한 정보가 필요하다. 특히 현행 전원조사 방식에서는 조사완료 가구 내 일부 적격가구원이 응답하지 않는 경우가 발생하므로, 미응답 가구원의 성별과 연령대 정보가 확보되어야 성별×연령대별 개인 무응답 보정이 가능하다.

따라서 향후 조사에서는 표본가구별 접촉이력, 방문횟수, 방문일시, 최종 접촉결과, 거절 사유, 가구원 명부, 미응답 적격가구원의 기본 인구학적 정보 등을 하나의 표본관리파일로 관리할 필요가 있다. 이러한 자료가 축적되면 단순 셀 기반 무응답 보정을 넘어 응답확률모형 기반 보정, 고위험 무응답 집단 식별, 조사원별·지역별 실사 품질관리로 활용할 수 있다. 아울러 조사구별 완료가구 수, 일부 무응답률, 예비표본 사용률, 초과조사 발생 여부를 실사 기간 중 실시간으로 모니터링하여 특정 조사구나 지역에서 문제가 집중될 경우 즉시 보완할 수 있는 관리체계를 마련하고, 표본추출틀의 최신성도 정기적으로 점검할 필요가 있다.

제3절 지역별 공표 확대 및 통계 활용 방안

지역별 공표 가능성 검토 결과, 3개 권역과 6개 권역 수준에서는 디지털 문해능력 수준별 추정이 비교적 안정적으로 가능하지만, 시도별 또는 시도×연령별 세부 공표에서는 일부 셀의 표본규모가 부족하여 상대표준오차가 크게 나타날 수 있다. 특히 디지털 문해능력 수준1과 같이 비율이 낮은 범주는 상대표준오차가 구조적으로 크게 산출되므로, 상대표준오차만을 기준으로 공표 가능성을 판단할 경우 정책적으로 중요한 취약집단 정보가 과도하게 제한될 수 있다.

따라서 비율이 낮은 범주에 대해서는 상대표준오차와 함께 95% 신뢰구간의 절대 반폭을 병기하고 공표 시 주석을 통해 해석상 유의사항을 제시하는 방안이 바람직하다. 제3장의 분석에서 확인된 바와 같이 결과변수를 디지털 문해능력 수준1~3과 수준4의 2범주로 단순화하면 17개 시도 전체를 현행 표본 규모로 공표할 수 있으며, 시도×6개 연령대와 같이 지나치게 세분화된 공표 단위는 3개 연령대 통합, 권역 단위 공표, 2범주 지표 활용 등 대안적 공표체계를 함께 검토할 필요가 있다. 지역과 연령을 교차하여 공표할 경우에는 6개 권역×3개 연령대가 통계적 안정성과 조사 현실성을 함께 고려한 효율적인 절충안이라 판단된다.

차기조사에는 모든 결과를 동일한 수준으로 공표하기보다는 추정의 신뢰성과 정책적 활용 목적에 따라 공표체계를 차별화할 필요가 있다. 우선 수준1~3과 수준4로 구분한 2범주 지표는 시도, 연령, 사회경제적 특성 등 세부집단별로 폭넓게 공표하여 지역별 디지털 문해교육 수요와 정책 우선순위를 파악하는 지표로 활용할 수 있다. 반면 수준1~4의 4범주 지표결과는 전국·권역 및 정밀도 기준을 충족하는 시도와 세부집단을 중심으로 공표하되, 수준1과 같이 표본수가 적은 범주는 세부집단 병합하여 공표, 신뢰구간 및 주석을 함께 제시하는 것이 바람직하다.

수준1은 정책적으로 우선 관리가 필요한 집단이지만, 전체 비율이 낮아 지역, 연령 및 사회경제적 특성 등으로 세분할수록 표본수가 감소하고 추정오차

가 커질 수밖에 없다. 이는 수준1의 정책적 중요성이 낮다는 의미가 아니라, 오히려 해당 집단을 안정적으로 파악하기 위해 더 큰 표본이 필요함을 의미한다. 다만 조사 이전에는 개인의 수준1 여부를 직접 알 수 없으므로 수준1만을 대상으로 표본을 별도로 확대하는 것은 불가능하다.

따라서 차기 조사에서는 연령, 지역, 도시·농산어촌 여부, 가구유형 등 표본추출틀에서 확인 가능한 특성을 활용하여 디지털 문해 취약 가능성이 상대적으로 높은 집단의 표본을 보강하는 방안을 검토할 필요가 있다. 이 경우 특정 지역의 수준1 비율만을 근거로 표본을 배분하기보다, 고령인구 비율, 가구구조, 지역특성 등 안정적인 보조정보를 함께 활용하는 것이 바람직하다. 또한 불비례 배분된 표본은 설계가중치와 사후층화 또는 레이킹 보정을 통해 전국 및 지역 단위의 대표성을 확보하여야 한다.

더 나아가 통상 조사와 별도로 약 10년 주기의 대규모 정밀조사를 실시하는 중장기 방안을 검토할 수 있다. 현행 약 1만 명 규모의 조사는 전국 및 권역 단위의 변화 추이 파악과 기본적인 시도 단위 통계 생산에는 적절하나, 수준1과 같이 비율이 낮은 정책대상 집단에 대해 시도·연령을 교차한 세부통계를 안정적으로 생산하는 데에는 한계가 있다. 제3장의 분석에서도 4범주 기준 시도×3개 연령대 공표에는 약 2만 7천 명, 2범주 기준 시도×6개 연령대 공표에는 약 2만 2천 명 수준의 표본이 필요한 것으로 나타났다.

이에 따라 일반 조사(3년 주기)에서는 현행 표본규모를 유지하여 전국 및 권역 단위의 추세를 안정적으로 파악하고, 약 10년마다 표본규모를 약 3만 명 이상으로 확대하는 정밀조사를 실시하는 방안을 고려할 수 있다. 대규모 정밀조사는 시도별·연령별·교육수준별·경제활동상태별 디지털 문해 격차와 수준1 집단의 분포를 보다 세밀하게 파악하고, 지역별 디지털 문해교육 수요와 정책자원 배분의 근거를 마련하는 데 활용할 수 있다.

또한 조사결과가 실제 정책으로 이어질 수 있도록 통계공표와 정책활용 간의 연계체계를 강화할 필요가 있다. 시도별 수준1~3 비율과 수준1 비율을 활용하여 지역별 교육수요를 진단하고, 이를 지역 평생교육기관, 디지털배움터, 지자체 평생학습관 등의 교육과정 운영 및 지원대상 발굴에 활용할 수 있다.

아울러 연령, 교육수준, 경제활동상태, 가구유형별 결과를 함께 활용하여 고령층, 저학력층, 비경제활동인구, 1인 가구 등 정책적 관심집단에 대한 맞춤형 지원전략을 마련할 필요가 있다.

향후 조사 기획 단계에서는 공표목표, 목표 정밀도, 필요 표본규모, 예산 및 실시 가능성을 함께 고려한 표본설계 시나리오를 마련하고, 조사 종료 후에는 표본대체 현황, 무응답 구조, 가중치 조정 결과 및 공표단위별 표준오차를 정례적으로 검토하여 다음 조사설계 반영하는 환류체계를 구축하는 것이 바람직하다.

종합하면, 성인디지털문해능력조사는 단기적으로는 2범주 및 4범주 지표를 차별화하여 공표함으로써 통계 활용성을 높이고, 중장기적으로는 약 10년 주기의 대규모 정밀조사를 통해 수준1 등 정책우선 집단에 대한 세부통계 생산 기반을 확충할 필요가 있다. 이를 통해 본 조사는 지역별 디지털 문해교육 정책 수립과 취약계층 지원을 위한 실질적 활용도를 더욱 높일 수 있을 것이다.

제6장 결론

성인디지털문해능력조사는 통계법 제18조에 따라 승인된 일반조사통계(승인번호 제420002호)로, 디지털 전환이 사회 전반으로 확대되는 환경에서 우리나라 성인의 디지털 문해능력 수준을 국가 차원에서 정기적·공식적으로 파악하고 디지털 문해교육 정책 수립과 지원에 필요한 기초자료를 제공하는 데 그 목적을 두고 있다.

본 조사의 통계정보보고서 및 품질개선 컨설팅 신청서를 검토한 결과 가구 방문조사 환경 변화에 따른 무응답 편향 진단, 무응답 보정 가중치 산출체계의 개선, 지역별 공표를 위한 적정 표본규모 산출 등이 필요한 것으로 파악되었다. 이에 본 연구에서는 현행 가구 표본설계와 가중치 산출체계, 본표본 및 예비표본 운영, 무응답 발생 구조, 지역별 공표 가능성을 종합적으로 검토하여 조사의 신뢰성과 정책 활용도를 제고하기 위한 실질적 개선방안을 도출하였다.

본 연구의 주요결과를 정리하면 다음과 같다.

첫째, 현행 표본설계와 무응답 조정 가중치 산출체계를 종합적으로 검토하였다. 가구원 추출방식을 무작위 1인 선정, 가구당 최대 2인, 전원조사로 구분하여 비교한 결과 한 번 접촉한 가구에서 평균 약 2명의 응답자를 확보할 수 있고 일부 가구원 무응답에도 강건한 현행 전원조사 방식을 유지하는 것이 적절함을 확인하였다. 다만 조사구 단위 무응답 조정셀은 셀당 평균 응답자가 2.3명에 불과하여 불안정하므로 조정셀을 시도×가구유형 또는 시도 수준으로 확대하고, 사후층화를 레이킹 방식으로 확장하여 1인 가구 과소대표를 완화하는 개선방향을 제시하였다.

둘째, 디지털 문해능력 수준별 지역 공표 가능 범위를 상대표준오차와 로짓 신뢰구간으로 진단하고 적정 표본규모를 산출하였다. 3개 및 6개 권역에서는 4개 수준 모두 안정적으로 공표가 가능하며 시도 단독 공표는 핵심 범주인 수준2~4를 기준으로 대부분 가능하였다. 비율이 작은 수준1은 상대표준오차가 구조적으로 팽창하므로 절대 신뢰구간 반폭 기준으로 별도 평가할 것을 제안

하였다. 결과변수를 2범주로 단순화하면, 17개 시도 전체를 현행 표본규모로 공표할 수 있고 지역×연령 교차 공표가 필요한 경우 6개 권역×3개 연령대가 현실적인 절충안임을 확인하였다.

셋째, 표본대체 및 무응답 구조를 진단하였다. 본표본과 예비표본의 분포를 효과크기와 로지스틱 회귀분석으로 비교한 결과 핵심 결과변수인 디지털 문해능력 수준에서 두 집단의 차이는 무시 가능한 수준으로 예비표본이 본표본을 전반적으로 적절히 대체하고 있음을 확인하였다. 다만 시도별 본표본 비율은 큰 편차를 보였으며, 무응답 파라데이터 분석 결과 무응답은 가구 접촉 실패보다 접촉 이후 협조 확보 단계에서 주로 발생하는 것으로 나타났다.

넷째, 이상의 진단을 바탕으로 조사운영 및 활용 체계 개선방안을 제시하였다. 접촉 이후 설득을 강화하는 무응답 대응과 조사원 교육, 본표본 우선 원칙과 표본관리파일 축적을 통한 표본대체·현장 품질관리 고도화, 정밀도 기준에 따른 지역별 공표 확대와 통계 활용 방안을 제안하였다.

종합하면, 성인디지털문해능력조사는 현행 전원조사 방식과 표본설계 구조를 기본적으로 유지하되, 본 연구에서 제시한 무응답 보정, 표본대체 관리, 파라데이터 축적, 지역별 공표체계 등을 점검 및 보완할 필요가 있다. 이러한 개선이 이루어진다면 본 조사는 국가통계로서의 신뢰성을 높이고, 지역별 디지털 문해교육 정책 수립과 취약계층 지원을 위한 기초통계로서의 활용 가치를 더욱 강화할 수 있을 것이다.

참고문헌

- 국가평생교육진흥원(2025). 「제1차 성인디지털문해능력조사」 결과보고서.
- 국가평생교육진흥원(2025). 「제1차 성인디지털문해능력조사」 조사지침서.
- 국가평생교육진흥원(2025). 「제1차 성인디지털문해능력조사」 표본설계내역서.
- 교육부(2025). 「제1차 성인디지털문해능력조사 결과 발표」. 교육부 보도자료.
- 통계청(2025). 「표본설계 및 관리지침」. 통계청 예규 제300호.
- 김영원, 류제복, 박진우, 홍기학(2019). 「표본조사의 이해와 활용」. 교우사.
- Deville, J.-C., & Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418), 376-382.
- Little, R. J. A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54(2), 139-157.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

부록 1. 지역×연령대 디지털 문해능력 수준 정밀도 진단 상세 결과

<부록 표 1>과 <부록 표 2>는 제3장 지역별 공표를 위한 적정 표본규모 산출에서 활용한 공표단위별 상세 정밀도 진단 결과이다. 각 셀에 대해 디지털 문해능력 수준별 추정비율과 상대표준오차(RSE)를 산출한 뒤 두 범주 중 큰 RSE가 25% 이하인지를 기준으로 공표 가능 여부를 판정하였다.

<부록 표 1> 시도 × 3개 연령대 디지털 문해능력 수준 정밀도 진단 결과

시도	연령	수준1~3 비율(%)	수준4 비율(%)	최대 RSE(%)	현행 표본수	목표 표본수
서울	18-39세	11.0	89.0	22.3	323	323
	40-59세	39.2	60.8	10.5	458	458
	60세+	79.5	20.5	16.4	384	384
부산	18-39세	8.9	91.1	43.1	204	606
	40-59세	20.6	79.4	22.9	224	224
	60세+	68.3	31.7	22.5	252	252
대구	18-39세	18.3	81.7	33.4	135	241
	40-59세	39.3	60.7	19.8	214	214
	60세+	84.4	15.6	19.4	235	235
인천	18-39세	28.6	71.4	19.5	218	218
	40-59세	51.5	48.5	13.5	205	205
	60세+	82.3	17.7	22.4	197	197
광주	18-39세	18.3	81.7	35.4	136	273
	40-59세	51.8	48.2	20.9	159	159
	60세+	86.1	13.9	38.0	171	395
대전	18-39세	24.9	75.1	36.7	129	278
	40-59세	38.5	61.5	17.2	211	211
	60세+	83.4	16.6	32.4	120	202
울산	18-39세	35.9	64.1	20.0	97	97
	40-59세	68.8	31.2	22.5	174	174
	60세+	85.9	14.1	30.8	109	166
세종	18-39세	12.1	87.9	57.2	57	299

시도	연령	수준1~3 비율(%)	수준4 비율(%)	최대 RSE(%)	현행 표본수	목표 표본수
	40-59세	19.5	80.5	42.9	93	274
	60세+	70.1	29.9	31.3	70	110
경기	18-39세	28.7	71.3	13.4	397	397
	40-59세	23.6	76.4	14.0	496	496
	60세+	73.2	26.8	12.2	433	433
강원	18-39세	43.1	56.9	20.2	72	72
	40-59세	63.5	36.5	26.3	141	157
	60세+	93.7	6.3	51.5	272	1,153
충북	18-39세	42.7	57.3	23.6	112	112
	40-59세	50.1	49.9	17.7	193	193
	60세+	97.0	3.0	59.4	175	987
충남	18-39세	46.2	53.8	17.6	133	133
	40-59세	62.6	37.4	22.2	217	217
	60세+	94.9	5.1	45.0	210	680
전북	18-39세	4.0	96.0	63.1	32	204
	40-59세	46.6	53.4	18.0	100	100
	60세+	93.7	6.3	31.9	376	612
전남	18-39세	31.8	68.2	33.8	78	143
	40-59세	52.5	47.5	19.5	188	188
	60세+	98.6	1.4	66.3	245	1,724
경북	18-39세	23.2	76.8	27.5	103	125
	40-59세	38.1	61.9	15.1	204	204
	60세+	92.1	7.9	30.9	329	503
경남	18-39세	31.3	68.7	20.7	184	184
	40-59세	30.0	70.0	19.0	248	248
	60세+	74.5	25.5	27.6	248	303
제주	18-39세	22.4	77.6	23.7	49	49
	40-59세	50.2	49.8	16.1	123	123
	60세+	85.9	14.1	27.4	129	156
전체		47.2	52.8	66.3	10,062	16,091

<부록 표 2> 6개 권역 × 3개 연령대 디지털문해능력 수준 정밀도 진단 결과

권역	연령	수준1~3 비율(%)	수준4 비율(%)	최대 RSE(%)	현행 표본수	목표 표본
서울	18-39세	11.0	89.0	22.3	323	323
	40-59세	39.2	60.8	10.5	458	458
	60세+	79.5	20.5	16.4	384	384
경기·인천	18-39세	28.6	71.4	11.5	615	615
	40-59세	28.6	71.4	10.3	701	701
	60세+	74.9	25.1	11.0	630	630
충청	18-39세	36.5	63.5	13.4	431	431
	40-59세	49.5	50.5	8.9	714	714
	60세+	91.8	8.2	20.6	575	575
호남	18-39세	17.5	82.5	24.5	246	246
	40-59세	50.2	49.8	10.6	447	447
	60세+	93.9	6.1	25.7	792	839
영남	18-39세	21.3	78.7	12.6	723	723
	40-59세	34.7	65.3	8.4	1,064	1,064
	60세+	79.4	20.6	13.2	1,173	1,173
강원·제주	18-39세	36.8	63.2	16.9	121	121
	40-59세	59.1	40.9	17.1	264	264
	60세+	91.7	8.3	32.2	401	664
전체		47.2	52.8	32.2	10,062	10,372

부록 2. 본표본과 예비표본 비교 상세 결과

가. 효과크기 지표(Cramér's V, Cohen's d)

표본 수가 큰 경우 실제 차이가 작더라도 카이제곱검정이나 t-검정의 p값이 유의하게 나타날 수 있다. 따라서 본 분석에서는 p값의 유의성과 함께 효과크기를 산출하여 두 집단 간 차이의 실질적 크기를 판단하였다. 범주형 변수에는 Cramér's V, 연속형 변수에는 Cohen's d를 적용하였으며 산출식과 해석 기준은 다음과 같다.

Cramér's V는 두 범주형 변수 간 관련성 또는 분포 차이의 크기를 나타내는 효과크기 지표이며 카이제곱 통계량 χ^2 을 이용하여 다음과 같이 산출한다.

$$V = \sqrt{\frac{\chi^2}{n \times \min(r-1, c-1)}}$$

여기서 n 은 표본 수, r 은 행의 수, c 는 열의 수를 의미한다. Cramér's V는 0에서 1 사이의 값을 가지며 0에 가까울수록 두 집단 간 분포 차이가 작고 1에 가까울수록 분포 차이가 크다는 것을 의미한다.

- $V < 0.1$: 무시 가능한 차이
- $0.1 \leq V < 0.2$: 작은 차이
- $0.2 \leq V < 0.3$: 중간 수준의 차이
- $V \geq 0.3$: 큰 차이

Cohen's d는 두 집단의 평균 차이를 표준화한 효과크기 지표로 두 집단 평균의 차이를 통합표준편차로 나누어 산출한다.

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_p}, \quad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

이때 $\bar{x}_1, \bar{x}_2$ 는 두 집단의 평균, s_1^2, s_2^2 는 두 집단의 분산, n_1, n_2 는 두 집단의 표본수, s_p 는 통합표준편차를 의미한다.

- $|d| < 0.2$: 무시 가능한 차이

- $0.2 \leq |d| < 0.5$: 작은 차이
- $0.5 \leq |d| < 0.8$: 중간 수준의 차이
- $|d| \geq 0.8$: 큰 차이

나. 단변량 분석(효과크기) 결과

<부록 표 3>에는 가구 단위 주요 변수의 효과크기와 분포 차이가 제시되어 있다. 가구원수, 가구유형, 가구소득, 조사요일, 조사시간대 등 대부분의 변수에서 본표본과 예비표본은 거의 동일한 분포를 보였으며 시도(Cramér's $V=0.186$)와 6개 권역(0.120)에서만 작은 수준의 차이가 나타났다.

<부록 표 3> 주요변수별 본표본과 예비표본간 비율 차이

변수	범주	본표본 빈도	본표본 비율	예비표본 빈도	예비표본 비율	비율차 (본-예비)	Cramér V
시도(17)	서울	214	12.7%	367	11.0%	+1.8%	0.186
	부산	68	4.0%	272	8.1%	-4.1%	
	대구	108	6.4%	184	5.5%	+0.9%	
	인천	139	8.3%	171	5.1%	+3.2%	
	광주	88	5.2%	144	4.3%	+0.9%	
	대전	93	5.5%	137	4.1%	+1.4%	
	울산	27	1.6%	163	4.9%	-3.3%	
	세종	28	1.7%	82	2.4%	-0.8%	
	경기	268	15.9%	393	11.7%	+4.2%	
	강원	99	5.9%	141	4.2%	+1.7%	
	충북	60	3.6%	180	5.4%	-1.8%	
	충남	113	6.7%	167	5.0%	+1.7%	
	전북	99	5.9%	156	4.7%	+1.2%	
	전남	65	3.9%	191	5.7%	-1.8%	
	경북	112	6.7%	210	6.3%	+0.4%	
	경남	69	4.1%	271	8.1%	-4.0%	
	제주	33	2.0%	118	3.5%	-1.6%	
3개 권역	서울 특광역시	765	45.5%	1520	45.4%	+0.0%	0.014

변수	범주	본표본 빈도	본표본 비율	예비표본 빈도	예비표본 비율	비율차 (본-예비)	Cramér V
	중소도시	601	35.7%	1162	34.7%	+1.0%	
	농산어촌	317	18.8%	665	19.9%	-1.0%	
6개 권역	서울	214	12.7%	367	11.0%	+1.8%	0.120
	경기·인천	407	24.2%	564	16.9%	+7.3%	
	충청권	294	17.5%	566	16.9%	+0.6%	
	호남권	252	15.0%	491	14.7%	+0.3%	
	영남권	384	22.8%	1100	32.9%	-10.0%	
	강원·제주	132	7.8%	259	7.7%	+0.1%	
가구유형	아파트	826	49.1%	1557	46.5%	+2.6%	0.024
	일반	540	32.1%	1127	33.7%	-1.6%	
	읍/면부	317	18.8%	663	19.8%	-1.0%	
주중/주말	주중	1192	70.8%	2386	71.3%	-0.5%	0.004
	주말	491	29.2%	961	28.7%	+0.5%	
시간대	오전 9~11시	336	20.0%	566	16.9%	+3.1%	0.040
	오후 12~17시	912	54.2%	1839	54.9%	-0.8%	
	저녁 18시 이후	435	25.8%	942	28.1%	-2.3%	
가구소득1	100만원 미만	82	4.9%	187	5.6%	-0.7%	0.054
	100~200만원 미만	210	12.5%	449	13.4%	-0.9%	
	200~300만원 미만	255	15.2%	570	17.0%	-1.9%	
	300~400만원 미만	300	17.8%	609	18.2%	-0.4%	
	400~500만원 미만	286	17.0%	548	16.4%	+0.6%	
	500~600만원 미만	282	16.8%	567	16.9%	-0.2%	
	600~700만원 미만	176	10.5%	282	8.4%	+2.0%	
	700만원 이상	92	5.5%	135	4.0%	+1.4%	
가구소득2	300만원 미만	546	32.4%	1201	35.9%	-3.4%	0.038
	300~500만원 미만	587	34.9%	1157	34.6%	+0.3%	
	500만원 이상	550	32.7%	989	29.5%	+3.1%	

시도별 본표본 비율은 <부록 표 4>와 같다. 울산 14.2%에서 인천 44.8%까지 약 3배의 편차를 보였는데 이는 본표본 가구의 응답률·접촉성공률이 지역별로 다른 데서 비롯된 것으로 볼 수 있다. 다만 이러한 시도 단위의 불균형은 3개 권역 수준에서는 거의 사라져(Cramér's $V=0.014$) 권역 단위 추정에는 영향이 작을 것으로 추정된다. 따라서 시도별 추정 시에는 사후층화 가중을 통해 시도×연령×성별 모집단 분포에 맞추는 보정을 권장한다.

<부록 표 4> 시도별 본표본 비율

시도	가구 수	본표본 가구수	예비표본 가구수	본표본 비율
인천	310	139	171	44.8%
강원	240	99	141	41.3%
경기	661	268	393	40.5%
대전	230	93	137	40.4%
충남	280	113	167	40.4%
전북	255	99	156	38.8%
광주	232	88	144	37.9%
대구	292	108	184	37.0%
서울	581	214	367	36.8%
경북	322	112	210	34.8%
세종	110	28	82	25.5%
전남	256	65	191	25.4%
충북	240	60	180	25.0%
제주	151	33	118	21.9%
경남	340	69	271	20.3%
부산	340	68	272	20.0%
울산	190	27	163	14.2%
전국	5030	1683	3347	33.5%

가구원 단위에서 본표본과 예비표본의 평균 및 비율 차이를 <부록 표 5>·<부록 표 6>에 각각 나타내었다. 개인 단위 주요 변수 전반에서 효과크기는 매우 작아(연속형 Cohen's $d<0.05$, 범주형 Cramér's $V<0.05$) 실질적 차이는 거의 없는 것으로 확인되었다. 또한 핵심 결과변수인 디지털 문해능력 수준도 Cramér's $V=0.041$, 최대 범주 간 비율차 2.9%p에 그쳤다.

〈부록 표 5〉 주요변수별 본표본과 예비표본간 가구원단위 평균 차이

변수	본표본 평균	본표본 SD	예비표본 평균	예비표본 SD	평균 차이	Cohen's d
전체 소요시간	18.785	11.001	19.028	11.400	-0.243	-0.022
SQ 소요시간	2.660	5.112	2.523	3.389	0.137	0.034
A문항 소요시간	1.266	3.166	1.413	3.615	-0.147	-0.042
B문항 소요시간	11.812	9.397	11.967	9.396	-0.155	-0.017
C문항 소요시간	4.406	4.703	4.410	4.718	-0.004	-0.001
DQ 소요시간	1.302	2.891	1.239	3.664	0.063	0.018
만 나이	53.083	16.060	53.350	16.267	-0.267	-0.016

〈부록 표 6〉 주요변수별 본표본과 예비표본 간 가구원단위 비율 차이

한글명	범주	본표본 빈도	본표본 비율	예비표본 빈도	예비표본 비율	비율차 (본-예비)	Cramér V
성별	남자	1,563	46.5%	3,164	47.2%	-0.7%	0.007
	여자	1,798	53.5%	3,537	52.8%	+0.7%	
연령 (6구간)	18-29세	328	9.8%	624	9.3%	+0.4%	0.027
	30-39세	490	14.6%	1,017	15.2%	-0.6%	
	40-49세	473	14.1%	895	13.4%	+0.7%	
	50-59세	759	22.6%	1,521	22.7%	-0.1%	
	60-69세	781	23.2%	1,474	22.0%	+1.2%	
	70세+	530	15.8%	1,170	17.5%	-1.7%	
스마트폰 유형	안드로이드(갤럭시)	2,717	80.8%	5,524	82.4%	-1.6%	0.031
	아이폰(애플)	435	12.9%	729	10.9%	+2.1%	
	기타 일반폰 (실버폰, 키즈폰 등)	209	6.2%	448	6.7%	-0.5%	
학력1	무학	45	1.3%	142	2.1%	-0.8%	0.029
	초등학교 졸업	249	7.4%	528	7.9%	-0.5%	
	중학교 졸업	361	10.7%	698	10.4%	+0.3%	
	고등학교 졸업	1,237	36.8%	2,464	36.8%	+0.0%	
	대학교 졸업 이상	1,469	43.7%	2,869	42.8%	+0.9%	
학력2	중학교 졸업 이하	655	19.5%	1,368	20.4%	-0.9%	0.012
	고등학교 졸업	1,237	36.8%	2,464	36.8%	+0.0%	

한글명	범주	본표본 빈도	본표본 비율	예비표본 빈도	예비표본 비율	비율차 (본-예비)	Cramér V
	대학교 졸업 이상	1,469	43.7%	2,869	42.8%	+0.9%	
경찰 상태	취업	2,411	71.7%	4,746	70.8%	+0.9%	0.019
	비취업	49	1.5%	133	2.0%	-0.5%	
	비경제활동	901	26.8%	1,822	27.2%	-0.4%	
생활 만족도	만족	1,797	53.5%	3,373	50.3%	+3.1%	0.032
	보통	1,382	41.1%	2,900	43.3%	-2.2%	
	불만족	182	5.4%	428	6.4%	-1.0%	
정치적 관심	관심있음	1,032	30.7%	1,846	27.5%	+3.2%	0.040
	보통	1,188	35.3%	2,340	34.9%	+0.4%	
	관심없음	1,141	33.9%	2,515	37.5%	-3.6%	
디지털 문해 수준	수준 1	327	9.7%	570	8.5%	+1.2%	0.041
	수준 2	683	20.3%	1,559	23.3%	-2.9%	
	수준 3	853	25.4%	1,556	23.2%	+2.2%	
	수준 4	1,498	44.6%	3,016	45.0%	-0.4%	

다. 다변량 분석 결과(로지스틱 회귀)

본표본과 예비표본을 다변량 차원에서 비교하기 위해 응답자가 본표본에 해당하는지 여부(본표본=1, 예비표본=0)를 종속변수로 하는 개인 단위 이항 로지스틱 회귀모형을 설정하였다(n=10,062). 독립변수로는 단변량 비교에 사용한 조사수행 특성(조사 소요시간·조사요일·조사시간대), 지역 및 가구 특성(시도·권역·가구유형), 인구·사회경제적 특성(성별·연령·학력·가구소득·경제활동상태·생활만족도·정치적 관심·스마트폰 유형)을 투입하였다.

모형의 적합도와 분리력을 우도비 검정통계량(LR χ^2), pseudo-R², AUC로 평가한 결과는 <부록 표 7>과 같다. AUC는 0.628, pseudo-R²는 0.0353으로 나타나 여러 변수를 동시에 고려하더라도 본표본과 예비표본을 뚜렷하게 구분하기 어려운 것으로 확인되었다. AUC는 0.5에 가까울수록 두 집단의 구분이 어렵고 1에 가까울수록 완전한 구분을 의미한다.

〈부록 표 7〉 로지스틱 회귀 모형 적합도 (개인 모형, n=10,062)

표본수	LL (모형)	LL (영가설)	LR χ^2 (df)	pseudo-R ²	AUC (비가중)
10,062	-6,183.3	-6,409.5	452.4 (55)	0.0353	0.628

개별 변수의 기여도는 더미를 함께 고려하는 결합 Wald 검정으로 판정하였으며 그 결과는 〈부록 표 8〉과 같다. 시도, 조사시간대, 정치적 관심, 가구유형, 스마트폰 유형 5개 변수만 통계적으로 유의(p<0.05)하였으며 그중 시도가 가장 강한 분리 변수로 나타났다.

〈부록 표 8〉 변수별 결합 p-value

한글명	자유도	Wald p값
시도(17)	16	0.000
시간대	2	0.000
정치적 관심	2	0.000
가구유형	2	0.025
스마트폰 유형	2	0.045
A문항 소요시간	1	0.092
연령	5	0.141
학력1	4	0.142
가구소득1	7	0.186
성별	1	0.218
DQ 소요시간	1	0.244
SQ 소요시간	1	0.252
B문항 소요시간	1	0.272
C문항 소요시간	1	0.273
전체 가구원수	1	0.325
만 나이	1	0.394
주중/주말	1	0.471
경찰 상태	2	0.636
조사참여 가구원수	1	0.757
생활만족도	2	0.940
적격 가구원수	1	0.984
가구소득2	2	1.000

유의한 변수에 대해서는 범주별 회귀계수와 오즈비(OR)를 산출한 결과는 <부록 표 9>와 같다. 서울을 기준으로 울산(OR=0.273), 부산·경남·제주 등은 본표본 오즈가 절반 이하로 낮은 반면 인천(OR=1.344)은 유의하게 높아 단변량에서 확인된 시도별 본표본 비율 차이와 일관된 결과를 보였다. 그 외 조사 시간대·정치적 관심·가구유형·스마트폰 유형의 오즈비는 1에 가까워 실질적 영향은 작은 것으로 해석된다.

<부록 표 9> 유의한 변수별 회귀계수 추정 결과

변수	범주	회귀계수 β	표준오차	p 값	OR
시도	부산	-0.808	+0.114	+0.000	+0.446
	대구	+0.020	+0.109	+0.853	+1.020
	인천	+0.296	+0.103	+0.004	+1.344
	광주	+0.043	+0.120	+0.721	+1.044
	대전	+0.121	+0.115	+0.295	+1.128
	울산	-1.298	+0.166	+0.000	+0.273
	세종	-0.598	+0.170	+0.000	+0.550
	경기	+0.044	+0.085	+0.607	+1.045
	강원	+0.147	+0.118	+0.214	+1.158
	충북	-0.550	+0.126	+0.000	+0.577
	충남	+0.026	+0.114	+0.819	+1.026
	전북	+0.090	+0.120	+0.456	+1.094
	전남	-0.573	+0.128	+0.000	+0.564
	경북	-0.115	+0.112	+0.306	+0.892
	경남	-0.874	+0.117	+0.000	+0.417
	제주	-0.805	+0.160	+0.000	+0.447
시간대	오후 12~17시	-0.183	+0.060	+0.002	+0.833
	저녁 18시 이후	-0.291	+0.067	+0.000	+0.748
정치적관심	보통	-0.095	+0.056	+0.090	+0.909
	관심없음	-0.242	+0.059	+0.000	+0.785
가구유형	일반	-0.138	+0.051	+0.007	+0.872
	읍/면부	-0.065	+0.069	+0.344	+0.937
스마트폰유형	아이폰(애플)	+0.192	+0.081	+0.018	+1.212
	기타 일반폰 (실버폰, 키즈폰 등)	+0.083	+0.111	+0.454	+1.087