

「2023년 한국수어활용조사」 표본설계

2023. 7.

I. 조사 개요	2
II. 표본설계 방향	3
III. 표본 설계	4
1. 모집단	4
2. 표본추출틀	5
3. 층화 및 분류지표	6
4. 표본규모	9
5. 표본추출	11
IV. 가중값 작성	12
1. 설계 가중값	12
2. 무응답 보정 가중값	12
3. 사후조정(층화) 가중값	12
V. 추정	13
1. 평균 추정	13
2. 평균 추정량의 분산 추정	13
3. 표준오차 및 상대표준오차	13
VI. 표본 관리	14

I 조사 개요

1. 조사명 및 목적

- (조사명) 2023년 한국수어활용조사
- (목적) 농인의 한국수어 활용 실태 등을 파악하고, 농인의 언어권리적 보장을 향한 수어 활용의 저변을 확대해 나가는 실질적인 정책 수립 기초자료 마련

2. 법적근거

- 「한국수화언어법」 제9조(실태조사) 및 동법 시행령 제3조

3. 위탁기관

- 문화체육관광부 국립국어원(특수언어진흥과)

4. 조사연혁

- (2017년) ‘2017년 한국수어 사용 실태조사’ 최초 실시
 - 국가승인통계 제113022호(2017.11.13.)
- (2020년) ‘2020년 한국수어 활용 조사’ 실시(조사명칭 변경)

5. 조사대상

- 대한민국에 거주하는 만 20세 이상 청각장애인(장애정도: 심한 장애)

6. 조사방법 및 조사기간

- (조사방법) 조사원에 의한 면접조사(필요시 자기기입식 병행)
- (조사기간) 2023. 8. 29.~9. 11. (14일간) * 준비조사: 8.28.(1일간)

II 표본설계 방향

1. 표본규모: 500명

- 원표본에 대하여 조사 불가사유 발생에 따른 예비표본(표본의 10배수) 제공 필요
- 조사대상자 불응, 불능 시 예비표본으로 대상자 대체

2. 표본설계 시 고려사항

- 성, 연령, 권역 등을 감안하여 층화 및 분류지표 설계
 - 성별: 남성, 여성
 - 연령별: 20대, 30대, 40대, 50대, 60대 이상
 - 도시규모별: 대도시¹⁾, 중소도시²⁾, 농어촌³⁾
 - 1) (대도시) 7개 특별광역시(서울, 부산, 대구, 인천, 광주, 대전, 울산)
 - 2) (중소도시) 9개 도(경기, 강원, 충북, 충남, 전북, 전남, 경북, 경남, 제주) 및 세종특별자치시의 동부
 - 3) (농어촌) 9개 도 및 세종특별자치시의 읍·면부
- 주소불분명, 도서지역은 표본 선정 시 제외 검토
- 청년 및 중년층 유효표본 확보 검토 요청

3. 기타사항

- 특이사항에 대해서는 부서 협의 후 표본설계 진행
- 표본추출틀 이용은 행정자료 개별파일의 보안관리 지침에 따라 행정자료 포털시스템 개별파일 이용신청·승인 후 사용

III 표본설계

1. 모집단

○ 모집단 정의

- 목표모집단: 대한민국에 거주 중인 재가 청각장애인
- 조사모집단: 2022년 12월 31일 기준 보건복지부 등록장애인 DB 중 만 20세 이상의 장애정도가 심한 청각장애인

※ 주 장애유형 및 정도가 청각, 심한 장애 또는 부 장애유형 및 정도가 청각, 심한 장애

< 표1 > 조사모집단의 분포

시도	총계	비율 (%)	남자	비율 (%)	여자	비율 (%)	60세이상 비율(%)
전국	83,276	100.0	43,972	100.0	39,304	100.0	71.2
서울특별시	11,853	14.2	6,427	14.6	5,426	13.8	67.2
부산광역시	4,955	6.0	2,691	6.1	2,264	5.8	71.8
대구광역시	3,944	4.7	2,038	4.6	1,906	4.8	69.4
인천광역시	4,625	5.6	2,475	5.6	2,150	5.5	69.1
광주광역시	2,360	2.8	1,248	2.8	1,112	2.8	65.2
대전광역시	1,951	2.3	1,001	2.3	950	2.4	67.4
울산광역시	1,906	2.3	1,161	2.6	745	1.9	73.5
세종특별자치시	395	0.5	194	0.4	201	0.5	70.9
경기도	17,530	21.1	9,280	21.1	8,250	21.0	66.5
강원도	3,324	4.0	1,773	4.0	1,551	3.9	78.3
충청북도	3,009	3.6	1,578	3.6	1,431	3.6	71.1
충청남도	4,747	5.7	2,461	5.6	2,286	5.8	72.0
전라북도	3,817	4.6	1,963	4.5	1,854	4.7	77.5
전라남도	5,263	6.3	2,772	6.3	2,491	6.3	78.7
경상북도	6,344	7.6	3,245	7.4	3,099	7.9	78.5
경상남도	5,997	7.2	3,059	7.0	2,938	7.5	73.3
제주특별자치도	1,256	1.5	606	1.4	650	1.7	76.7

2. 표본추출틀

○ 표본추출틀 작성

- 표본추출틀은 조사모집단 중 주소불분명¹⁾, 소규모지역²⁾을 제외한 79,076명이며, 조사모집단 대비 포함률은 95.0%
- 제외사유는 주소불분명 48명, 소규모지역 4,152명

- 1) (주소불분명) 등록장애인 명부에서 세부도로명주소가 표기되어있지 않은 사례
 2) (소규모지역) 최소 표본규모 확보 등을 위해 읍면동 내 규모가 충분하지 않은 지역은 추출틀에서 제외(제외기준: 읍면동 내 장애인수 9명 미만)

< 표2 > 표본추출틀의 포함률

(단위: 명, %)

급여유형		모집단(A)	제외 대상			표본 추출틀(B)	포함률 (B/A)
			소계	주소 불분명	소규모 지역		
전국		83,276	4,200	48	4,152	79,076	95.0
성별	남성	43,972	2,202	27	2,175	41,770	95.0
	여성	39,304	1,998	21	1,977	37,306	94.9
도시 규모별	대도시	31,594	770	15	755	30,824	97.6
	중소도시	27,794	720	20	700	27,074	97.4
	농어촌	23,888	2,710	13	2,697	21,178	88.7
연령별	20대	3,261	86	0	86	3,175	97.4
	30대	3,833	98	1	97	3,735	97.4
	40대	6,190	158	9	149	6,032	97.5
	50대	10,736	424	10	414	10,312	96.1
	60대 이상	59,256	3,434	28	3,406	55,822	94.2

○ 표본추출틀 분포

- 연령별 장애인 수는 60대 이상이 전체의 70%를 차지하고 있음
- 도시규모별 장애인 수는 대도시에서 가장 많이 나타남
- 도시규모별 농어촌 포함률은 90%를 약간 하회하고 있으나 연령별로는 94% 이상을 포함하고 있으며 전체 포함률은 95.0%로 적정하다고 판단, 위의 추출틀을 최종적으로 적용함

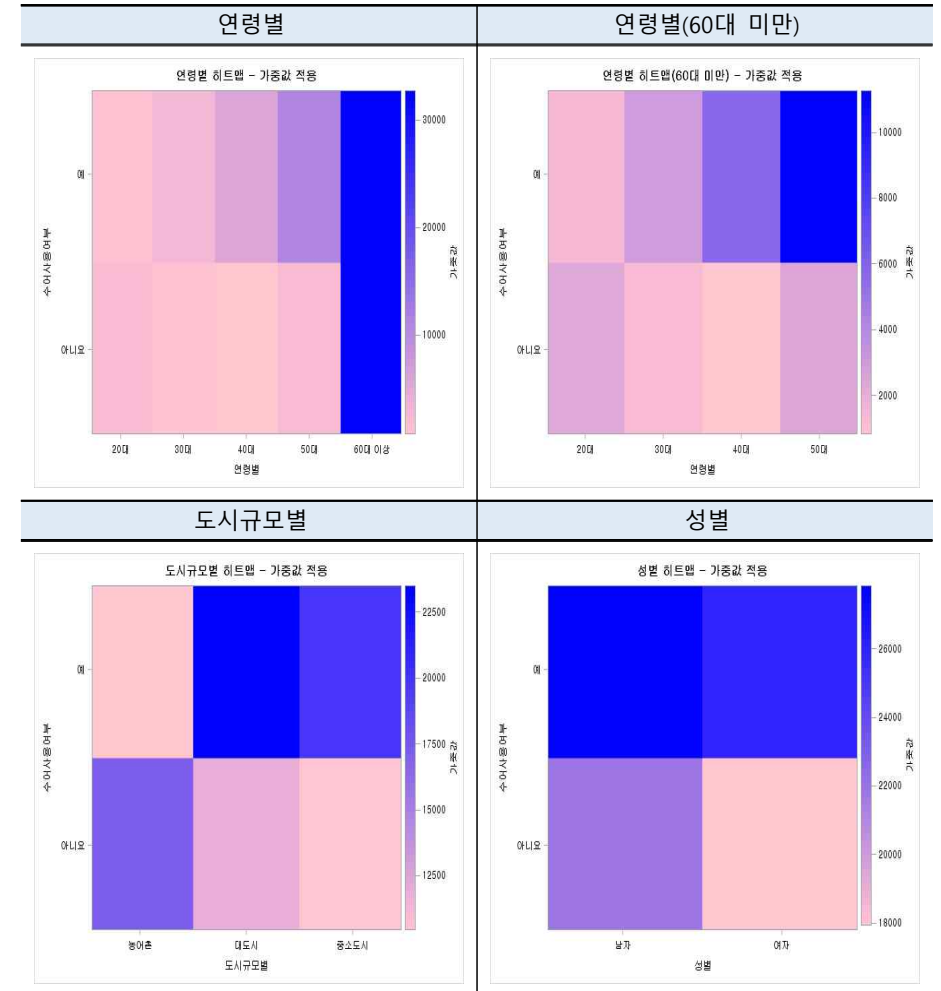
3. 층화 및 분류지표

- 전체 수어활용여부 추정치 정도향상 및 공표를 위한 유효표본 확보 등을 위해 층화 검토
 - 표본추출률 상 활용가능한 성, 연령, 도시규모별 수어활용여부 이질성 확인 후 층화변수 검토
- 층화변수 및 분류지표
 - 층의 세분화는 표본 규모의 한계 및 추정량 분산의 과소 추정 등의 사유로 층화변수는 1개로 결정
 - 2020년 한국수어활용조사 및 수요부서 요청사항을 반영하여 분류지표를 고려
 - 전주기 분류지표인 연령별, 도시규모별, 성별의 분석분석* 실시
 - * 가중값 적용 분산분석
 - 통계적으로 유의한 2개의 변수 중 도시규모의 검정통계량이 연령에 비해 약간 크지만 수요부처 요청사항(저연령층 확보) 등을 감안하여 연령층을 층화변수로 선정하고 나머지 두 변수를 분류지표로 활용

< 표3 > 분산분석 결과

종속변수	독립변수	Levels	F Value	Pr > F
수어사용 여부	연령별	4	10.3162	<.0001
	도시규모별	2	14.0898	<.0001
	성별	1	0.5688	0.4511

< 표4 > 주요지표별 수어사용여부 분포 - 히트맵(가중값 적용)



- 로지스틱모형*을 이용하여, Type III의 F Value의 크기에 따라 순서를 적용

* 가중값 적용 로지스틱분석

< 표5 > 분류지표 선정 및 순서 결과

순서	변수 분류	F Value	Pr > F
1차	도시규모별	26.3922	<.0001
2차	성별	0.5678	0.4511

* P-value < 0.05

< 표6 > 최종 분류지표

정렬 순서	지표	분류지표		
①	도시규모별	7대 광역시 및 동부 읍면부 고려		
		1구간	2구간	3구간
		대도시	중소도시	농어촌
②	성별	성별		
		1구간	2구간	
		남성	여성	

4. 표본 규모

○ 표본규모: 500명

- 전국 추정 오차한계, 상대표준오차와 함께 예산 및 조사비용 등을 고려하여 표본규모는 500명으로 설정

- 표본 불능 사유 발생 시, 대체표본(10배수)을 추가 제공

○ 예상 오차

- 층화설계에 따른 분산추정을 통한 전국 추정의 예상표본오차 산출
- 예상되는 표준오차(SE)는 5.0%, 오차한계(MOE)는 $\pm 9.8\%$ p, 상대 표준오차(RSE)는 10.0%

< 표7 > 표본 규모에 따른 예상 오차

	2020년	2023년 예상 오차			
표본 규모(명)	485	400	450	500	550
SE	4.66*	5.6	5.3	5.0	4.8
MOE(%p)	$\pm 9.13^*$	± 11.1	± 10.4	± 9.8	± 9.3
RSE(%)	9.31*	11.3	10.6	10.0	9.5

* 2020년 한국수어활용조사 주요 항목(주된 의사소통방법으로 수어 사용 여부)의 오차 평균

< 표8 > 연령별에 따른 예상 오차

구분	2020년				2023년		
	표본 규모	SE	MOE*	RSE	예상 SE	예상 MOE*	예상 RSE
전체	485	4.5	± 8.9	9.3	5.0	± 9.8	10.0
20대	26	9.5	± 18.6	25.0	6.2	± 12.2	16.4
30대	36	8.1	± 15.8	11.7	6.1	± 11.9	8.8
40대	62	6.1	± 12.0	7.0	5.5	± 10.8	6.3
50대	122	4.4	± 8.6	5.3	4.9	± 9.7	6.0
60대이상	239	3.1	± 6.1	6.3	3.4	± 6.6	6.8

* $z_{\alpha/2} = 1.96$, 95% 신뢰구간

○ 표본 배분

- 수요부처의 한국수어활용조사 조사대상자 선정시 청년 및 중년층 일정수준 확보 요청
- 60대 이상 고연령층 분포가 전체의 70% 이상을 차지하고 있어 비례배분시 청년 및 중년층의 유효표본 확보가 어려움
- 층별 유효표본 확보를 위해 최소배분 후 장애인수 기준으로 제공근 비례배분 방법 적용
- 적용 결과 청년 및 중년층 표본이 확보됨에 따라 해당연령층 예상오차는 축소된 반면, 60대 이상의 경우 예상오차한계는 $\pm 6.6\%p$ 로 수용가능한 수준

* 15명 최소 배분

< 표9 > 표본 규모

급여유형(설계층)		모집단	비례배분	제공근	제공근 배분	최소배분* (최종 표본)	설계 가중값
총계		83,276	500	544	500	500	158.2
연령 별	20대	3,261	20	57	52	60	52.9
	30대	3,833	23	62	57	63	59.3
	40대	6,190	37	79	72	76	79.4
	50대	10,736	64	104	95	96	107.4
	60대 이상	59,256	356	243	223	205	272.3

* 15명 최소배분 후 제공근 배분

5. 표본 추출

- 추출방법: 층화계통추출(Stratified Systematic Sampling)
- 조사모집단과 표본 간 주요 지표별 차이는 설계가중값 적용 시 0.4%p 이내

< 표10 > 주요 지표별 표본조사구 구성

(단위: %, %p)

주요 지표		조사모집단 구성비	표본추출률 구성비	표본구성비 (설계가중값 적용)	차이(B-A)
성별	남성	52.8	52.8	52.7	-0.1
	여성	47.2	47.2	47.3	0.1
도시규모별	대도시	37.9	39.0	38.6	-0.4
	중소도시	33.4	34.2	34.4	0.2
	농어촌	28.7	26.8	27.0	-0.2
연령층	20대	3.9	4.0	4.0	0
	30대	4.6	4.7	4.7	0
	40대	7.4	7.6	7.6	0
	50대	12.9	13.0	13.0	0
	60세 이상	71.2	70.6	70.6	0

< 표11 > 층별 표본수 및 표본추출률

층	표본추출률	표본	표본추출률 (%)
총계	79,076	500	0.63
20대	3,175	60	1.89
30대	3,735	63	1.69
40대	6,032	76	1.26
50대	10,312	96	0.93
60세 이상	55,822	205	0.37

IV 가중 값 작성

1. 설계 가중값(Design Weight)

- 설계가중값은 표본설계로부터 직접 구해지며, 층화계통추출 (Stratified Systematic Sampling)로 조사대상자 추출

설계
가중값

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

- $h = 1, 2, \dots, H$: 설계층을 나타내는 첨자
- w_h^D : h 층의 설계 가중값
- N_h : h 층의 전체 청각장애인 수
- n_h : h 층의 표본 청각장애인 수

2. 무응답 보정 가중값

- 응답자가 조사 자체에 불응한 단위무응답 등 표본의 변동 사유에 해당하는 경우 무응답 보정 실시

무응답 보정
가중값

$$w_h^R = w_h^D \times \frac{\sum w_i^D}{\sum w_i^D \times I_R}$$

- w_h^R : 무응답 보정된 가중값
- w_h^D : 설계 가중값
- $I_R = \begin{cases} 1, & \text{응답} \\ 0, & \text{무응답} \end{cases}$

3. 사후조정(층화) 가중값

- 추정량의 정도를 개선하기 위해 조사모집단 혹은 벤치마크 모집단*의 분포에 맞게 사후조정(층화)

* 조사시점 등록장애인수가 확보되는 경우 벤치마크 모집단 적용

사후층화
가중값

$$w_{hi} = w_h^R \times \frac{N}{\sum w_h^R}$$

- w_{hi} : h 층 i 번째 표본의 사후층화 가중값
- w_h^R : 무응답 보정 가중값
- N : 조사모집단 수(벤치마크 모집단 수)

V 추정

1. 평균 추정

평균 추정량 $\hat{\mu} = \hat{\bar{Y}} = \hat{Y} / w_{..}$

- y_{hi} : h 층 i 표본의 관측값(비율추정의 경우 0 또는 1)
- w_{hi} : h 층 i 표본의 최종 가중값(설계, 무응답, 사후 가중값)
- $w_{..}$: 최종 가중값의 합(모집단 개체 수($= \sum_h \sum_i w_{hi}$))

2. 평균 추정량의 분산 추정

층별 평균
분산추정량

$$\hat{V}_h(\hat{\mu}) = \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi} - \bar{e}_h)^2$$

전체 평균
분산추정량

$$\hat{V}_h(\hat{\mu}) = \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} \left[\frac{w_{hi}(y_{hi} - \hat{\mu}_h)}{\sum_{i=1}^{n_h} w_{hi}} - \frac{1}{n_h} \sum_{i=1}^{n_h} \left(\frac{w_{hi}(y_{hi} - \hat{\mu}_h)}{\sum_{i=1}^{n_h} w_{hi}} \right) \right]^2$$

- $f_h = n_h / N_h$
- $e_{hi} = \frac{w_{hi}(y_{hi} - \hat{\mu}_h)}{\sum_{i=1}^{n_h} w_{hi}} = w_{hi}(y_{hi} - \hat{\mu}_h) / w_h$
- $\bar{e}_h = \left(\sum_{i=1}^{n_h} e_{hi} \right) / n_h$

3. 표준오차 및 상대표준오차

평균 추정량의
표준오차(SE)

$$\widehat{SE}(\hat{\mu}) = \sqrt{\widehat{V}(\hat{\mu})}$$

평균 추정량의
상대표준오차(RSE)

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

1. 표본대체

- 부적격이나 불응이 발생하는 경우 대체하여 조사 진행
 - 조사 불가사유에 따른 예비표본(표본의 10배수) 제공
 - 예비표본은 본표본과 연령, 도시규모, 성별이 동일한 대상자 중에서 랜덤하게 추출

2. 표본관리 정보 입수

- 표본의 가중값 조정을 위해 조사대상자 접촉 이력정보 등 표본관리 정보(부재, 불응, 부적격, 미접촉 등) 입수 필요
 - 표본의 유고사유가 표본의 변동사유인지 모집단의 변동사유인지 구분할 필요가 있음
- ※ 유고사유가 전출, 부재, 불응, 불능 등 표본의 변동사유인 경우에는 가중값 조정을 통하여 보완하고, 조사대상 외 모집단의 변동사유인 경우는 벤치마크 단계에서 보완 예정

< 표12 > 표본의 유고사유(예시)

전출	대상자가 관할 사무소 이외 지역으로 이전한 경우
조사 불응	사생활침해, 건강불화, 가족구성원의 특이함, 정부불신 등으로 조사를 거부하는 경우
조사 불능	타지역으로 잦은 출타, 주응답자의 장기투병, 신체적, 정신적 장애 등으로 조사가 불가능한 경우
사망 등 조사대상 외	조사대상자 사망, 명부 오류 등 모집단의 변동사유