2011년도 국가통계 품질개선지원 연구용역

# 『평생학습개인실태조사』
# 품질개선지원 최종결과보고서

## - 표본설계 개선방안-

**2011. 9.**

통계청 STATISTICS KOREA

제 출 문

# 제 출 문

통계청장 귀하

본 보고서를 『평생학습개인실태조사 통계』 품질 개선지원 연구용역 과제의 최종 연구결과물로 제출합니다.

[개선지원 부문]: 표본설계 개선방안

2011년 9월

**한국통계진흥원장 이 동 명 ㊞**


연구진

| | |
|---|---|
| 책임연구원 | 임 명 선 (한국통계진흥원 사무국장) |
| 연 구 원 | 임 대 철 (한국통계진흥원 조사부 과장) |
| 연 구 원 | 김 종 환 (한국통계진흥원 기획연구부 과장) |

요 약 문

# 최종결과보고서 요약문

| 연구과제명 | 「평생학습개인실태조사」 표본설계 개선방안 |
|---|---|
| 주 제 어 | 평생학습개인실태조사, 표본설계 |
| 연 구 기 간 | 2011. 06. 17 ~ 2011. 9. 30 |
| 연 구 기 관 | (재)한국통계진흥원 |
| 연 구 진 | 임명선, 임대철, 김종환 |


본 연구는 한국교육개발원에서 매년 조사하는 평생학습개인실태조사의 표본설계 개선을 위한 연구이다. 지난 2007년 실시한 통계청의 정기품질진단 결과에서 '표본설계' 부분에 대해 제기된 문제점에 대해 개선안을 제공하는데 그 목적이 있다.

○ 조사 현황을 검토한 결과, 인구통계학적 특성이 전혀 고려되지 않은 표본설계로 인하여 응답자 특성이 편향되어 있다. 따라서 인구통계학적 특성이 고려된 표본설계 방법을 제시한다.
- 먼저 2010년도 인구총조사 자료를 토대로 모집단을 정의하고 표본설계에 활용가능한 표본추출틀을 마련한다.
- 서울 및 광역시와 기타 도지역을 동지역 및 읍면지역으로 층화하여 행정구역을 25개로 층화한다.
- 표본크기는 조사예산을 고려하여 모집단 분포에 따라 비례하게 지역별로 할당한다.
- 인구통계학적 특성을 반영하기 위해 성별, 연령, 학력, 점유유형의 특성을 분류지표로 선정하여 추출함으로써 부차층의 효과를 얻는다.
- 추출된 가구에 대해서는 조사대상자로 해당되는 가구원을 모두 조사하는 것을 원칙으로 한다.

○ 본 연구를 통해 인구통계학적 특성이 반영된 조사대상자를 선정함으로써 모집단과 가장 유사한 표본을 대상으로 조사가 이루어질 것으로 예상된다.

# 목 차


제1장 서론 ····· 1

1절 연구배경 ····· 1

2절 연구목적과 필요성 ····· 1

(1) 필요성 ····· 1

(2) 목적 ····· 1

3절 연구내용 및 방법 ····· 3

(1) 연구방법 ····· 3

(2) 모집단 분석 ····· 3

(3) 표본설계 방안제시 ····· 4

4절 연구흐름도 및 협의내용 ····· 4

(1) 연구흐름도 ····· 4

(2) 연구협의 내용 ····· 6

제2장 통계 및 개선과제 개요 ····· 7

1절 통계개요 ····· 7

(1) 조사목적 ····· 7

(2) 조사설계 ····· 7

(3) 조사현황 ····· 8

2절 개선과제 ····· 9

(1) 조사현황 ····· 9

(2) 문제점 ····· 10

(3) 개선사항 ····· 11

3절 유사조사 분석 ····· 11

제3장 표본설계 ····· 14
1절 모집단 분석 ····· 14
(1) 모집단 설정 ····· 14
(2) 층화 ····· 16
2절 표본설계 ····· 17
(1) 표본크기 ····· 17
(2) 표본추출 ····· 21
3절 추정 ····· 28
(1) 가중값 ····· 28
(2) 추정 ····· 30
제4장 결론 및 제안 ····· 32

# - 표 목 차 -

<표 1> 연구내용 및 방법 ····· 3
<표 2> 2010년 시도별 조사현황 ····· 8
<표 3> 2010년 조사결과 중 연령, 학력, 고용상태 분포 ····· 9
<표 4> 연령분포 비교(2010년 인구총조사) ····· 10
<표 5> 만 25세~64세 개인 모집단 분포 ····· 14
<표 6> 지역별 동 · 읍면 비중 ····· 15
<표 7> 지역별 가구수 및 가구원수 ····· 15
<표 8> 지역별 개인 및 가구모집단 및 표본규모 ····· 17
<표 9> 추출할 가구규모 및 조사구 수 ····· 18
<표 10> 지역별 개인 및 가구모집단 및 표본규모 ····· 19
<표 11> 전국단위 공표시 표본오차별 표본규모 ····· 20
<표 12> 지역별 개인 및 가구모집단 및 표본규모 ····· 20
<표 13> 층화 및 분류지표 ····· 23
<표 14> 지역별 조사구수 ····· 24
<표 15> 분류지표 기준 ····· 25
<표 16> 조사구 정렬 기준(예시) ····· 26
<표 17> 표본조사구와 모집단의 인적특성 비교 ····· 27

# - 그 림 목 차 -

<그림 1> 연구흐름도 ····· 5

# 제1장 서론

## 1절 연구배경

국가통계 품질진단 결과로 조기 품질개선이 필요한 통계에 대하여 개선안과 구체적인 실행방안을 제시함으로써 해당 통계의 신속한 품질개선 및 정책 자료로의 활용도를 제고하기 위하여 국가통계정기품질진단의 내용을 토대로 표본설계 부분에서 이론적ㆍ실증적 연구가 필요하다.

## 2절 연구목적과 필요성

### (1) 필요성

일반적으로 통계조사는 모집단의 일부분인 표본(sample)에 의존한다. 조사대상 전체의 집합인 모집단(population)을 대상으로 하는 통계조사는 현실적으로 거의 불가능하기 때문이다. 따라서 표본은 모집단을 가장 잘 대표할 수 있는 것이어야 한다.

표본추출을 어떻게 하는가는 통계조사의 성패를 좌우할 뿐만 아니라 나아가 통계의 품질에도 직접적인 영향을 미치게 된다. 표본추출을 하기 위해서는 표본의 크기(규모)나 추출방법 등을 결정하는 일련의 과정이 선행되어야 하는데 이를 표본설계라 하며 실제 모집단과의 차이가 최소가 되도록 하는 것이 중요하다.

### (2) 목적

통계청의 2007년도 정기품질진단 결과에서 '표본설계' 부분에 대해 제기된 문제점과 개선안은 다음과 같다.

□ 문제점

○ 가구원의 인구통계학적 특성 미반영

- 행정구역별로 층화하여 조사구 추출시 단순임의추출함에 따라 가구원들의 성별, 직업, 연령 등의 인구통계학적 특성이 전혀 고려되지 않고 있다.

○ 가구원

- 최종적으로 선정된 가구내에 만25세 이상 64세 이하의 가구원 중 1~2명만 선정하여 조사함에 따라 비교적 낮시간에 집에 많이 머물러 있는 가정주부들의 응답비율이 높게 나타난다.

□ 개선안

- 지역, 성별, 연령 및 직업 등을 고려하여 모집단을 대표할 수 있는 최적의 표본설계 방법을 연구한다.
- 모수에 대한 추정이 가능하도록 추정식을 제시한다.

2007년도에 지적된 문제점들과 개선사항들이 현재까지 반영되지 않고 있기 때문에 본 연구를 통해 앞서 제기된 문제점을 보완하고 요구되는 개선안을 최대한 반영한 최적의 표본설계방법을 찾고자 한다.

## 3절 연구내용 및 방법

### (1) 연구방법

- 「평생학습개인실태조사」의 연구내용 및 방법은 아래와 같다.

<표 1> 연구내용 및 방법

| 연구 내용 | 연구 방법 |
|---|---|
| - 평생학습개인실태조사의 조사설계, 조사방법 등 검토 | - 통계작성기관 방문 면담<br>- 통계청 '국가통계품질진단'보고서 검토<br>- 유사 조사 수집 및 검토 |
| - 조사 모집단 검토 및 문제점 파악 | - 통계작성기관으로부터 제공받은 조사모집단에 대해 분류 기준 및 표본설계 활용시 문제점 파악 |
| - 표본추출틀 구축 | - 표본설계시 사용가능한 표본추출틀 정의 |
| - 적합한 표본설계 제시 | - 통계작성기관과의 협의와 활용방안을 고려한 적합한 표본설계 방법을 제시 |
| - 최종 표본설계 방안 정립 | - 통계작성기관과의 협의회와 자문회의를 통한 최종 실행가능한 표본설계방안 제시 |

### (2) 모집단 분석

○ 조사대상

- 만25세 이상 64세 이하의 일반가구에 거주하는 개인

○ 조사지역

- 전국 16개 시ㆍ도

○ 검토 사항

- 현 상황에서 가장 적절한 표본추출틀을 결정하고 활용가능한 분류 기준 변수를 선택한다.

### (3) 표본설계 방안제시

- 모집단 현황을 분석하여 사용가능한 표본추출틀을 정의한다.
- 목표모집단에 가장 적합한 표본 추출방법을 제시하여 현실가능한 표본설계 방안을 제시한다.
- 또한 조사모집단의 변화에 따른 표본추출틀의 변화를 반영할 방법을 제시한다.

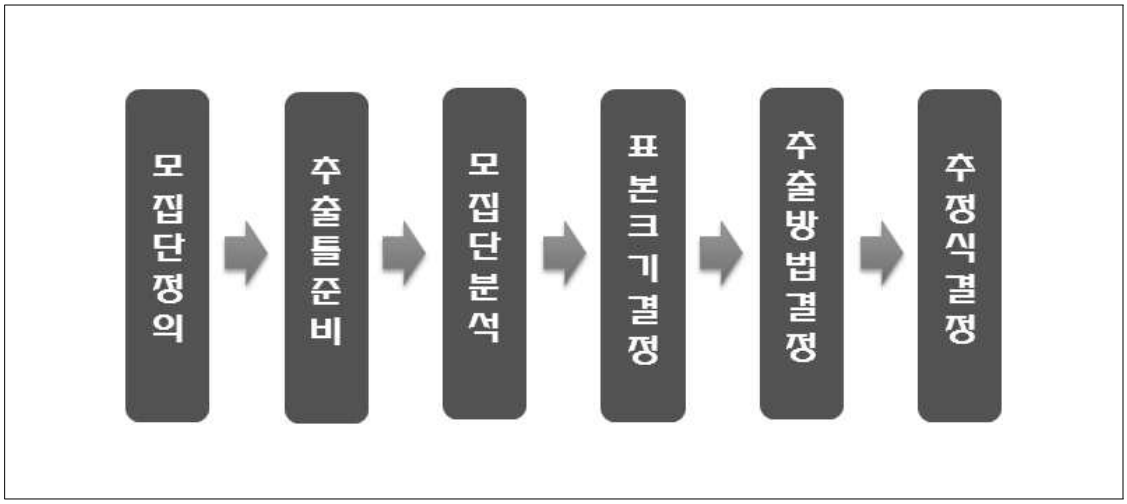
## 4절 연구흐름도 및 협의내용

### (1) 연구흐름도

- 본 연구의 전체 흐름도는 다음과 같은 과정을 통해 이루어진다.
- 먼저 조사대상이 되는 모집단인 조사모집단을 명확하게 정의하고 정확한 표본조사를 위해서는 개념적으로 정의한 목표모집단과 표본추출을 위한 실제 조사대상인 조사모집단의 차이가 최소가 되도록 하는 것이 중요하다.
- 조사모집단이 정해지면 조사모집단을 최대한 포함하는 표본추출틀을 준비하고 조사목적에 맞으면서 이용 가능한 표본추출틀을 확보하여 표본설계시 사용가능한 항목들을 포함하여 결정한다.
- 또한  조사모집단의 조사사업체가 빠지거나 중복되는 경우가 없도록 해야 하며, 조사모집단은 가능한 한 동일한 표본추출틀을 사용해서 일관성을 유지하도록 한다.
- 표본추출틀이 마련되면, 조사일정을 수립하고 일정에 따른 비용을 산정하여 예산을 배정한다.
- 그런 다음, 실제 조사에 필요한 표본크기와 가장 효율적인 추출법을 정하고, 그에 따라 층화와 표본배분, 표본추출단계, 조사방법 등을

결정하는 등 구체적인 계획들을 마련한다.

- 마지막으로 모수 추정식을 제시한다.




<그림 1> 연구흐름도

### (2) 연구협의 내용

#### □ 협의 일시 및 장소

○ 일 시 : 2011. 8. 12(금)

○ 장 소 : 한국교육개발원

○ 참석자 : 이경미(한국교육개발원 평생교육통계팀 팀장), 이진아(한국교육개발원 평생교육통계팀 연구원), 김용환(한국통계진흥원 조사부 부장), 임대철(한국통계진흥원 조사부 과장)

#### □ 협의내용

○ 응답대상

- 인구통계학적 특성이 반영되지 않은 표본설계로 인하여 응답대상 중 화이트칼라들의 응답비중이 낮게 나타난다.

○ 모집단 층화

- 행정구역별 층화는 서울 및 광역시 7개 지역과 기타 도지역을 동지역과 읍면지역으로 나누어 18개 지역으로 층화하였다.

○ 조사의 문제점

- 표본수가 작아 분석단위를 지역으로 나누어 분석할 수 없으며 향후 점차적으로 표본수를 늘려갈 계획이다.
- 2011년 조사시 표본가구 2,700가구에서 5,000명을 조사할 계획이다.

# 제2장 통계 및 개선과제 개요

## 1절 통계개요

### (1) 조사목적

평생학습개인실태조사의 목적으로는 첫째, 학습사회 성숙도 가늠을 위한 국민의 평생학습 참여 양상 및 학습역량 진단, 학습의 활동패턴을 이해한다. 둘째, 평생학습과 노동시장과의 이행과정에 대한 심층적 이해를 통해 모든 연령대에 걸친 효과적인 인적자원개발 계획을 수립한다. 셋째, 평생학습자 공급과 수요 사이의 불일치(mismatch)문제 개선을 위하여 '공급관점'의 평생교육 기본통계조사와 더불어 '수요관점'의 개인표집조사를 실시한다. 넷째, 평생학습 영역에 있어서 우리나라의 현황 진단 및 국가 간 비교를 위하여 OECD 등 국제기구 표준과 호환가능한 조사를 실시한다. 본 조사결과는 OECD 등 국제기구에 제출되어 우리나라의 평생학습 실태에 관한 국제적 비교, 분석 자료로 활용하게 된다.

### (2) 조사설계

○ 조사 내용

- 일반사항, 형식교육, 비형식교육, 무형식 학습, 장애요인 및 접근성, 학습 역량, 사회적 성과, 응답자의 일반적 배경

○ 조사 대상

- 만 25세 이상 64세 이하의 일반가구에 거주하는 개인

○ 조사 지역

- 전국 16개 시ㆍ도

○ 조사 기준

- 조사 기간 : 7월 1일부터 8월 31일(2개월)
- 조사 주기 : 1년

○ 조사 방법

- 해당 가구를 직접 방문함으로써 조사 대상자를 직접 대면하여 면접하는 개별방문조사(face-to-face interview)를 원칙으로 한다. 단, 부득이한 경우 유치조사를 병행 실시한다.

### (3) 조사현황

2010년 평생학습개인실태조사는 전국 16개 시·도의 전체 표본조사구에서 추출된 270개 조사구별로 할당된 가구를 대상으로 하였다. 만 25세 이상 64세 이하 가구원 4,500명을 대상으로 조사하였다.

<표 2> 2010년 시도별 조사현황

(단위 : 명, %)

| | 빈도 | 구성비 |
|---|---|---|
| 서울 | 994 | 22.1 |
| 부산 | 336 | 7.5 |
| 대구 | 218 | 4.8 |
| 인천 | 254 | 5.6 |
| 광주 | 132 | 2.9 |
| 대전 | 142 | 3.2 |
| 울산 | 103 | 2.3 |
| 경기 | 987 | 21.9 |
| 강원 | 135 | 3.0 |
| 충북 | 135 | 3.0 |
| 충남 | 169 | 3.8 |
| 전북 | 159 | 3.5 |
| 전남 | 160 | 3.6 |
| 경북 | 233 | 5.2 |
| 경남 | 288 | 6.4 |
| 제주 | 55 | 1.2 |
| 합계 | 4,500 | 100.0 |

## 2절 개선과제

### (1) 조사현황

2010년 평생학습개인실태조사결과 중 연령, 학력, 고용상태의 분포를 살펴본 결과는 다음과 같다. 특히, 학력은 고등학교 졸업이 43.0%로 가장 높았고 고용상태에서는 가정주부가 전체 응답자 중 27.6%로 높게 나타났다.

<표 3> 2010년 조사결과 중 연령, 학력, 고용상태 분포

(단위 : 명, %)

| 구분 | | 빈도 | 구성비 |
|---|---|---|---|
| 연령 | 25~29세 | 261 | 5.8 |
| | 30~39세 | 1,099 | 24.4 |
| | 40~49세 | 1,536 | 34.1 |
| | 50~59세 | 1,130 | 25.1 |
| | 60~64세 | 474 | 10.5 |
| 학력 | 무학 | 27 | 0.6 |
| | 초등학교 | 284 | 6.3 |
| | 중학교 | 315 | 7.0 |
| | 고등학교 | 1,935 | 43.0 |
| | 대학(2·3년제 대학) | 538 | 12.0 |
| | 대학교(4년제 대학교) | 1,272 | 28.3 |
| | 대학원 이상 | 129 | 2.9 |
| 고용상태 | 취업 및 자영업 | 3,072 | 68.3 |
| | 실업 및 구직 | 115 | 2.6 |
| | 학생 | 37 | 0.8 |
| | 은퇴, 조기퇴직 | 35 | 0.8 |
| | 가정주부 | 1,241 | 27.6 |
| 합계 | | 4,500 | 100.0 |

## (2) 문제점

### □ 조사결과 분석

- 2010년 인구총조사 결과 중 25세 이상 64세 이하의 인구 분포를 살펴보면 평생학습개인실태조사 결과와 다소 차이가 있는 것으로 나타난다.

<표 4> 연령분포 비교(2010년 인구총조사)

(단위 : 명, %)

| | 2010년 평생학습개인실태조사 | | 2010년 인구총조사 결과 | |
|---|---|---|---|---|
| | 빈도 | 구성비 | 빈도 | 구성비 |
| 25~29세 | 261 | 5.8 | 3,538,949 | 12.5 |
| 30~39세 | 1,099 | 24.4 | 7,794,495 | 27.6 |
| 40~49세 | 1,536 | 34.1 | 8,204,781 | 29.0 |
| 50~59세 | 1,130 | 25.1 | 6,564,826 | 23.2 |
| 60~64세 | 474 | 10.5 | 2,182,236 | 7.7 |
| 합계 | 4,500 | 100.0 | 28,285,287 | 100.0 |

- 학력 또한 인구총조사 결과와 다른 분포를 가지는 것으로 나타났으며 조사결과 중 고용상태를 살펴보면 취업 및 자영업이 68.3%로 현재 우리나라의 고용상태와 크게 다른 것으로 나타났다. 특히, 주간 면접이 비교적 쉬운 가정주부의 응답비율이 크게 높은 것으로 나타났다.
- 행정구역별로 층화하여 조사구 추출시 단순임의추출함에 따라 가구원들의 성별, 직업, 연령 등의 인구통계학적 특성이 고려되지 않은 조사구로 조사를 실시하므로 인구통계학적 특성을 통제하지 않고 있다.
- 또한, 가구원 중 주민등록상의 생일이 빠른 1~2명을 조사대상자로만 선정함에 따라 응답자 특성이 편향될 수 있는 가능성을 가지고 있다.

### (3) 개선사항

현 표본설계 방법에 대한 문제점을 체계적인 표본설계를 통하여 인구통계학적 특성이 잘 반영될 수 있고 현실가능한 표본설계 방안을 제시하는 것이 필요하다.

## 3절 유사조사 분석

본 조사에 대해 2007년 정기품질진단 결과에서 제시된 개선사항 중 「가족실태조사」와 「국민독서실태조사」, 「국민생활체육참여실태조사」 등의 조사들과 성격이 유사하여 정보교류의 필요성과 협조체제 구축을 강조하였다. 본 조사와 가장 유사하고 표본조사로 시행되고 있는 조사로 「가족실태조사」를 들수 있다. 따라서 여성가족부의 「가족실태조사」 표본설계 방법을 중심으로 살펴본다.

**□ 조사명칭 : 가족실태조사**

**□ 조사목적**

- 한국 가족의 특성과 변화 추이를 통합적으로 포착할 수 있는 자료 확보
- 맞벌이 가정의 보편화 현상에 주목하여, 가족원 돌봄 및 일상적인 생활양식을 파악하여 일-가정 양립을 위한 정책 마련에 필요한 기초자료 제공
- 제2차 건강가정기본계획 수립을 위한 기초자료 및 모델의 근거 마련
- 조사결과를 토대로 맞춤형·생활밀착형 가족정책을 위한 방향 제시

**□ 조사개요**

○ 조사대상

- 전국 일반가구의 만 15세 이상 가구원 전체

○ 조사방법

- 컴퓨터를 이용한 면접조사(CAPI)

## □ 표본설계

○ 모집단

- 2005년 인구주택총조사의 전국 16개 시도의 일반 가구 + 2005년 이후 2010년 6월까지 신축된 아파트 가구

○ 표본추출틀

- 2005년 인구주택총조사의 조사구와 가구수 자료 외에 2005년 12월 이후～2010년 6월까지 신축하여 입주한 아파트에 관한 정보를 보조적인 자료로 활용해 모집단 보완

○ 표본규모

- 표본규모 : 2,500가구
- 추정치의 허용오차 : 95%신뢰수준에서 ±2.0%p (모비율 추정 기준)

○ 표본할당

- 시도별 지역별 주택유형별 층화집락계통추출

○ 표본추출

- 광역시도별, 주택유형별, 동/읍면부별 조사구를 행정지역번호 및 조사구의 가구수 크기 순으로 정렬하여 표본을 추출
- 1차 추출 : 조사구를 조사구의 가구수 크기에 비례하는 추출확률을

부여하여 계통추출로 조사구를 추출

- 2차 추출 : 표본조사구내에서 가구명부를 작성하여 조사구마다 추출간격을 결정하여 5가구를 계통추출로 가구를 추출

## □ 본 연구에 반영사항

○ 모집단 보완

- 가족실태조사는 조사모집단을 2005년 인구주택총조사의 일반가구 뿐만 아니라 2005년 이후 신축한 아파트 가구를 조사모집단으로 포함함으로써 현재와 가장 비슷한 조사모집단을 구축하였다. 평생학습개인실태조사에서도 2011년 이후 조사모집단 선정에 있어서 신규 가구에 대해 반영이 되지 않으므로 조사모집단을 보완하도록 한다.

○ 층화

- 가족실태조사에서는 주택유형을 고려하여 층화하였으나 평생학습개인실태조사의 표본설계에서는 인구통계학적 특성이 고려된 층화가 이루어져야 한다.

# 제3장 표본설계

## 1절 모집단 분석

### (1) 모집단 설정

### □ 모집단

본 연구에서의 목표모집단은 만25세 이상 64세 이하의 개인이며 조사 모집단은 2010년 인구주택총조사 중 일반가구에 거주하는 만 25세~64세 개인이다. 2010년 인구주택총조사 결과에 따르면 만 25세~64세 인구는 다음과 같다.

<표 5> 만 25세~64세 개인 모집단 분포

(단위 : 명)

| | 25~29세 | 30~34세 | 35~39세 | 40~44세 | 45~49세 | 50~54세 | 55~59세 | 60~64세 | 합계 |
|---|---|---|---|---|---|---|---|---|---|
| 전 국 | 3,538,949 | 3,695,348 | 4,099,147 | 4,131,423 | 4,073,358 | 3,798,131 | 2,766,695 | 2,182,236 | 28,285,287 |
| 서 울 | 898,900 | 857,764 | 850,730 | 797,360 | 784,467 | 777,432 | 592,471 | 468,836 | 6,027,960 |
| 부 산 | 247,229 | 239,936 | 256,860 | 271,955 | 293,334 | 312,412 | 248,722 | 192,906 | 2,063,354 |
| 대 구 | 169,158 | 174,275 | 201,620 | 217,539 | 219,070 | 200,207 | 148,331 | 113,154 | 1,443,354 |
| 인 천 | 200,048 | 207,697 | 232,540 | 240,289 | 248,709 | 217,853 | 140,003 | 103,653 | 1,590,792 |
| 광 주 | 113,039 | 114,706 | 126,152 | 127,527 | 118,774 | 105,584 | 73,008 | 60,344 | 839,134 |
| 대 전 | 114,615 | 115,328 | 129,017 | 129,921 | 125,691 | 114,889 | 79,161 | 58,205 | 866,827 |
| 울 산 | 75,807 | 85,342 | 95,126 | 105,585 | 105,437 | 91,127 | 61,284 | 40,023 | 659,731 |
| 경기도 | 821,115 | 924,453 | 1,074,259 | 1,059,502 | 992,903 | 837,407 | 544,701 | 409,903 | 6,664,243 |
| 강원도 | 85,501 | 91,411 | 110,150 | 117,177 | 121,477 | 124,021 | 93,016 | 69,649 | 812,402 |
| 충청북도 | 99,156 | 102,444 | 116,506 | 123,479 | 123,065 | 117,176 | 87,616 | 64,533 | 833,975 |
| 충청남도 | 132,577 | 141,500 | 155,044 | 153,553 | 153,480 | 143,493 | 110,994 | 96,025 | 1,086,666 |
| 전라북도 | 102,874 | 109,492 | 131,138 | 135,282 | 136,916 | 133,760 | 104,671 | 98,041 | 952,174 |
| 전라남도 | 85,461 | 98,140 | 119,167 | 129,597 | 133,735 | 132,614 | 107,133 | 103,877 | 909,724 |
| 경상북도 | 159,212 | 167,811 | 190,804 | 200,609 | 206,740 | 203,823 | 165,679 | 136,430 | 1,431,108 |
| 경상남도 | 202,144 | 229,358 | 264,987 | 275,367 | 266,282 | 247,848 | 182,499 | 142,494 | 1,810,979 |
| 제주도 | 32,113 | 35,691 | 45,047 | 46,681 | 43,278 | 38,485 | 27,406 | 24,163 | 292,864 |

기타 도지역 중 충남을 제외하고는 동지역 비중이 읍면지역에 비해 높았다.

<표 6> 지역별 동 · 읍면 비중

(단위 : %)

| 구분 | 경기 | 강원 | 충북 | 충남 | 전북 | 전남 | 경북 | 경남 | 제주 |
|---|---|---|---|---|---|---|---|---|---|
| 동지역 | 83.6 | 63.6 | 63.8 | 41.7 | 74.7 | 50.6 | 56.8 | 66.8 | 76.5 |
| 읍면 | 16.4 | 36.4 | 36.2 | 58.3 | 25.3 | 49.4 | 43.2 | 33.2 | 23.5 |
| 합계 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

## □ 표본추출틀(Sampling Frame) 검토

표본추출틀로는 2010년 인구주택총조사의 전국 16개 시 · 도의 일반가구로서 지역별 가구수와 가구원수는 아래와 같다.

<표 7> 지역별 가구수 및 가구원수

(단위 : 가구, 명, %)

| | 가구수 | | 가구원수 | |
|---|---|---|---|---|
| | | 구성비 | | 구성비 |
| 전 국 | 17,339,422 | 100.0 | 46,650,668 | 100.0 |
| 서 울 | 3,504,297 | 20.2 | 9,458,577 | 20.3 |
| 부 산 | 1,243,880 | 7.2 | 3,306,198 | 7.1 |
| 대 구 | 868,327 | 5.0 | 2,376,725 | 5.1 |
| 인 천 | 918,850 | 5.3 | 2,584,917 | 5.5 |
| 광 주 | 515,855 | 3.0 | 1,422,908 | 3.1 |
| 대 전 | 532,643 | 3.1 | 1,441,102 | 3.1 |
| 울 산 | 373,633 | 2.2 | 1,042,247 | 2.2 |
| 경기도 | 3,831,134 | 22.1 | 10,925,576 | 23.4 |
| 강원도 | 557,751 | 3.2 | 1,392,032 | 3.0 |
| 충청북도 | 558,796 | 3.2 | 1,439,229 | 3.1 |
| 충청남도 | 749,035 | 4.3 | 1,902,317 | 4.1 |
| 전라북도 | 659,946 | 3.8 | 1,698,825 | 3.6 |
| 전라남도 | 681,431 | 3.9 | 1,668,499 | 3.6 |
| 경상북도 | 1,005,349 | 5.8 | 2,457,756 | 5.3 |
| 경상남도 | 1,151,172 | 6.6 | 3,019,068 | 6.5 |
| 제주도 | 187,323 | 1.1 | 514,692 | 1.1 |

### (2) 층화

표본설계에서 조사목적에 부합하도록 효과적인 층화를 하는 것은 무엇보다도 중요한 일이다. 층화란 모집단을 특성에 따라 서로 동질적인 몇 개의 그룹(층)으로 나누는 과정이다. 모집단을 몇 개의 층으로 나눈 후에 각 층별로 독립적인 표본추출이 이루어진다.

전국 16개 시·도를 서울 및 광역시와 도지역으로 층화한 후, 도지역을 동지역 및 읍면지역으로 층화하여 행정구역별로 층화를 수행한다.

#### □ 층화구조

○ 1차 층화 : 16개 시·도별 1차 층화

○ 2차 층화 : 동지역 및 읍면지역으로 2차 층화

- 서울 및 광역시를 제외한 기타 도지역 9개에 대해 행정구역에 따라서 동지역과 읍면지역으로 구분한다.

## 2절 표본설계

### (1) 표본크기

#### □ 표본크기

표본크기는 조사예산을 고려하여 2012년 조사에서는 허용오차 1.4% 수준으로 5,000명을 조사할 계획으므로 모집단 분포에 따라 지역별로 비례할당하여 개인 표본규모를 산출하였다.

#### □ 표본배분

① 1단계 : 만 25세~64세 모집단 분포에 따라서 지역별로 비례할당하여 개인표본규모를 산출한다.

<표 8> 지역별 개인 및 가구모집단 및 표본규모

(단위 : 명)

| | 개인모집단 | 가구모집단 | 가구당 조사대상수 | 개인표본규모 |
|---|---|---|---|---|
| 전 국 | 28,285,287 | 17,339,422 | 1.63 | 5,000 |
| 서 울 | 6,027,960 | 3,504,297 | 1.72 | 1,066 |
| 부 산 | 2,063,354 | 1,243,880 | 1.66 | 365 |
| 대 구 | 1,443,354 | 868,327 | 1.66 | 255 |
| 인 천 | 1,590,792 | 918,850 | 1.73 | 281 |
| 광 주 | 839,134 | 515,855 | 1.63 | 148 |
| 대 전 | 866,827 | 532,643 | 1.63 | 153 |
| 울 산 | 659,731 | 373,633 | 1.77 | 117 |
| 경기도 | 6,664,243 | 3,831,134 | 1.74 | 1,178 |
| 강원도 | 812,402 | 557,751 | 1.46 | 144 |
| 충청북도 | 833,975 | 558,796 | 1.49 | 147 |
| 충청남도 | 1,086,666 | 749,035 | 1.45 | 192 |
| 전라북도 | 952,174 | 659,946 | 1.44 | 168 |
| 전라남도 | 909,724 | 681,431 | 1.34 | 161 |
| 경상북도 | 1,431,108 | 1,005,349 | 1.42 | 253 |
| 경상남도 | 1,810,979 | 1,151,172 | 1.57 | 320 |
| 제주도 | 292,864 | 187,323 | 1.56 | 52 |

② 2단계 : 개인 모집단을 가구모집단으로 나누어 가구당 조사대상수를 산출한다.

③ 3단계 : 지역별 개인표본 규모를 가구당 조사대상 인원수로 나누어 표본가구 규모를 산출한다.

④ 4단계 : 지역별 가구 표본규모를 유효 표본 가구수인 10가구로 나누어 지역별 표본 조사구 규모를 결정한다.

<표 9> 추출할 가구규모 및 조사구 수

(단위 : 명, 개)

| | 개인표본규모 | 가구당 조사대상수 | 표본 가구규모 | 추출할 조사구수 |
|---|---|---|---|---|
| 전 국 | 5,000 | 1.63 | 3,065 | 313 |
| 서 울 | 1,066 | 1.72 | 619 | 62 |
| 부 산 | 365 | 1.66 | 220 | 22 |
| 대 구 | 255 | 1.66 | 153 | 16 |
| 인 천 | 281 | 1.73 | 162 | 17 |
| 광 주 | 148 | 1.63 | 91 | 10 |
| 대 전 | 153 | 1.63 | 94 | 10 |
| 울 산 | 117 | 1.77 | 66 | 7 |
| 경기도 | 1,178 | 1.74 | 677 | 68 |
| 강원도 | 144 | 1.46 | 99 | 10 |
| 충청북도 | 147 | 1.49 | 99 | 10 |
| 충청남도 | 192 | 1.45 | 132 | 14 |
| 전라북도 | 168 | 1.44 | 117 | 12 |
| 전라남도 | 161 | 1.34 | 120 | 12 |
| 경상북도 | 253 | 1.42 | 178 | 18 |
| 경상남도 | 320 | 1.57 | 203 | 21 |
| 제주도 | 52 | 1.56 | 33 | 4 |

⑤ 5단계 : 기타 도지역에 대해서는 동지역 및 읍면지역 비중에 따라서 조사구를 할당한다.

<표 10> 지역별 개인 및 가구모집단 및 표본규모

(단위 : 개, 명)

| 기타 도지역 | 동/읍면 구분 | 동, 읍면 비율(%) | 할당 조사구수 | 가구 규모 | 개인 표본규모 |
|---|---|---|---|---|---|
| 경기도 | 동 | 83.6 | 57 | 566 | 979 |
| | 읍면 | 16.4 | 11 | 111 | 192 |
| 강원도 | 동 | 63.6 | 6 | 63 | 95 |
| | 읍면 | 36.4 | 4 | 36 | 54 |
| 충청북도 | 동 | 63.8 | 6 | 63 | 98 |
| | 읍면 | 36.2 | 4 | 36 | 56 |
| 충청남도 | 동 | 41.7 | 6 | 55 | 85 |
| | 읍면 | 58.3 | 8 | 77 | 119 |
| 전라북도 | 동 | 74.7 | 9 | 87 | 136 |
| | 읍면 | 25.3 | 3 | 30 | 46 |
| 전라남도 | 동 | 50.6 | 6 | 61 | 91 |
| | 읍면 | 49.4 | 6 | 59 | 88 |
| 경상북도 | 동 | 56.8 | 10 | 101 | 149 |
| | 읍면 | 43.2 | 8 | 77 | 114 |
| 경상남도 | 동 | 66.8 | 14 | 136 | 216 |
| | 읍면 | 33.2 | 7 | 67 | 108 |
| 제주도 | 동 | 76.5 | 3 | 25 | 42 |
| | 읍면 | 23.5 | 1 | 8 | 13 |

## □ 표본오차 및 공포범위

- 표본크기가 5,000명일 때의 모수추정에 대한 95% 신뢰수준의 최대 허용오차는 약 1.4% 수준으로 예상된다.
- 현재 표본설계에서의 공표단위는 전국단위이며 분석단위별로 일정 크기 이상의 표본이 배분되어야 한다.

## □ 표본오차별 유효표본수

○ 전국 단위 표본규모

- 현재 표본배분은 전국단위 공표에 맞게 표본이 배분되어 있다. 표본오차별 전국 공표단위 표본규모는 아래와 같다.

<표 11> 전국단위 공표시 표본오차별 표본규모

(단위 : 명)

| 표본오차 | 0.5% | 1.0% | 1.5% | 2.0% | 2.5% | 3.0% |
|---|---|---|---|---|---|---|
| 표본규모 | 38,311 | 9,597 | 4,267 | 2,401 | 1,536 | 1,067 |

○ 16개 시·도 단위 표본규모

- 16개 시·도별로 공표를 하기 위한 표본규모를 표본오차 1.0%~5.0%에 따라 살펴본 결과는 아래와 같다. 시·도별 공표를 위해 표본오차 3.0%에서는 17,037명 이상의 표본이 필요한 것으로 보인다. 표본규모는 지역별로 표본을 배분하는 방법과 지역별로 설정된 표본오차에 따라 달라질 수 있으며 조사비용과 여건에 맞게 표본규모를 결정하고 그에 맞는 표본설계가 필요로 한다.

<표 12> 지역별 개인 및 가구모집단 및 표본규모

(단위 : 명)

| 표본오차 | 1.0% | 2.0% | 2.5% | 3.0% | 3.5% | 4.0% | 5.0% |
|---|---|---|---|---|---|---|---|
| 표본규모 | 150,784 | 38,230 | 24,627 | 17,037 | 12,554 | 9,592 | 6,142 |

## (2) 표본추출

### □ 분류지표 결정

본 표본설계에는 성별, 연령, 직종 등의 인구통계학적 특성을 고려하여 표본을 추출함으로써 부차층을 구성하여 층화효과를 얻고자 한다. 평생학습개인실태조사에서 표본추출시 가장 고려되어야 할 특성으로는 대상자의 지역, 성별, 연령, 직종 등이 고려되어야 할 사항이다. 지역은 층화를 통해 통제하였으며 성별, 연령, 직종 등의 특성은 표본추출틀에서 제공가능한 항목을 살펴보고 적절한 분류지표를 선정하여 인구통계학적 특성을 반영한 표본설계를 실시한다.

○ 분류지표 항목 검토

- 표본추출틀로 사용하는 통계청의 인구총조사 조사항목 중 분류지표로 사용가능한 조사항목을 살펴보면 아래와 같다.

| 조사항목 | 코드명 |
|---|---|
| 성별 | ①남자, ②여자 |
| 연령 | 세 |
| 교육정도 | ①미취학, ②초등학교, ③중학교, ④고등학교, ⑤4년제 대학 미만, ⑥4년제 대학교 이상, ⑦대학원 석사, ⑧대학원 박사 |
| 점유형태 | ①자기집, ②전세, ③보증금 있는 월세, ④보증금 없는 월세, ⑤사글세, ⑥무상 |
| 거처종류 | ①단독주택, ②아파트, ③연립주택, ④다세대주택, ⑤비거주용 건물내 주택, ⑥오피스텔, ⑦호텔, 여관, ⑧기숙사, ⑨판작집, ⑩기타 |
| 총 방수 | 방+거실+식당 |
| 연건평 | 평, ㎡ |
| 대지면적 | 평, ㎡ |

○ 분류변수 선정

- 성별, 연령에 대한 특성은 표본추출틀에서의 조사항목 중 성별, 연령에 대한 항목이 있으므로 분류변수로 선택하도록 한다.
- 직종은 교육정도와 가장 관련이 높은 항목으로 분류변수로 교육정도로 대체한다.
- 소득은 주택의 점유형태와 가장 관련이 높은 항목으로 분류변수를 주택의 점유형태로 대체한다.
- 따라서 성별, 연령, 교육정도, 주택의 점유형태을 분류변수로 선정하여 표본추출시 반영하도록 한다.

○ 분류지표 생성

- 인구총조사 자료에서 조사구 단위별로 분류지표를 생성하여 표본조사구 추출시 분류지표로 사용한다.
- 조사구를 성별, 연령, 교육정도, 주택의 점유형태에 따라 분류하기 위해 조사구별 여성구성비, 50세 이상 구성비, 대졸 구성비, 자가 구성비를 파악하여 조사구를 분류한다.
- 새로 생성된 분류지표의 평균, 중위값 또는 사분위값으로 조사구를 부차층으로 분류한다.

## □ 표본추출

표본추출틀에서 지역별 층화를 한 후 각 층에 조사구수를 할당하고, 할당된 표본 조사구의 크기대로 각 층에서 조사구를 추출하기 위해 적절한 분류변수를 선정하여 분류변수에 따라 조사구 명부를 정렬 후 층별로 가구 수(MOS)를 기준으로 확률비례계통추출방법(PPS: Probability proportional to size)을 이용하여 추출한다.

<표 13> 층화 및 분류지표

| | | |
|---|---|---|
| 층화 | ▪ 지역 층화 : 25개 층<br>- 서울, 부산, 대구, 인천, 광주, 대전, 울산(7개)<br>- 경기, 강원, 충북, 충남, 전북, 전남, 경북, 경남, 제주 지역은 동부 및 읍면부로 구분 | |
| 분류 지표 | 1차 분류 | ▪ 여성 구성비(여성 비율)<br>예시) ① 50% 미만, ② 50% 이상 |
| | 2차 분류 | ▪ 50세 이상 구성비<br>예시) ① 20% 미만, ② 20% 이상 |
| | 3차 분류 | ▪ 교육정도(대졸 구성비)<br>예시) ① 20% 미만, ② 20% 이상 |
| | 4차 분류 | ▪ 점유형태(자가 구성비)<br>예시) ① 50% 미만, ② 50% 이상 |
| | 5차 분류 | ▪ 행정구역 및 조사구번호 |

추출된 표본조사구의 가구를 일련번호를 부여한 후 조사구번호의 끝번호와 가구 일련번호가 같은 가구를 최초가구로 결정하여 조사구의 명부순으로 10가구를 표본가구로 확정한다. 조사거부 또는 불능가구는 실제 조사의 진행 순서에 따라 10번째 이후 가구로 대체한다.

표본가구의 조사대상자인 만 25세 이상 만 64세 이하인 가구원을 모두 조사한다.

*표본대체 예시) 5번째 가구가 조사거부, 9번째 가구가 불능인 경우 대체방법
: 조사거부는 최소한 3번 이상 방문하여야 하므로, 9번째 불능인 가구를 먼저 11번째 가구로 대체하고 5번째 조사거부 가구는 12번째 가구로 대체함

## □ 표본추출절차(예시)

본 표본설계에서는 표본추출틀로 사용된 2010년 인구총조사 자료가 아직 미공표상태임으로 2005년 인구총조사 자료를 사용하여 분류지표를 사용한 표본추출방법을 제시하였다.

○ 표본추출틀

- 추출단위 : 2005년 인구총조사의 일반가구 조사구
- 조사항목 : 시도, 시군구, 읍면동, 조사구, 조사구특성, 성별, 연령, 교육정도, 점유형태

○ 층화

- 서울 및 광역시, 기타 도지역을 읍면지역과 동지역으로 구분하여 25개 층으로 구분

<표 14> 지역별 조사구수

(단위 : 명)

| 층 | | 조사구 | 가구 | 가구원 |
|---|---|---|---|---|
| 전 국 | | 265,341 | 15,887,067 | 26,607,534 |
| 서 울 | | 54,558 | 3,309,889 | 5,883,964 |
| 부 산 | | 20,065 | 1,186,378 | 2,069,327 |
| 대 구 | | 13,383 | 814,585 | 1,406,151 |
| 인 천 | | 14,186 | 823,021 | 1,450,789 |
| 광 주 | | 7,510 | 460,090 | 765,679 |
| 대 전 | | 8,018 | 478,865 | 800,834 |
| 울 산 | | 5,560 | 339,095 | 604,977 |
| 경기도 | 동 | 44,944 | 2,757,407 | 4,923,817 |
| | 읍면 | 10,010 | 571,769 | 970,347 |
| 강원도 | 동 | 5,297 | 315,406 | 488,640 |
| | 읍면 | 3,576 | 205,222 | 298,482 |
| 충청북도 | 동 | 5,005 | 307,284 | 492,668 |
| | 읍면 | 3,426 | 197,919 | 277,502 |
| 충청남도 | 동 | 3,752 | 231,786 | 374,383 |
| | 읍면 | 7,352 | 428,030 | 598,941 |
| 전라북도 | 동 | 6,854 | 407,983 | 665,576 |
| | 읍면 | 3,745 | 211,974 | 260,145 |
| 전라남도 | 동 | 4,456 | 260,478 | 431,264 |
| | 읍면 | 6,886 | 405,840 | 497,290 |
| 경상북도 | 동 | 7,715 | 462,200 | 751,174 |
| | 읍면 | 8,186 | 476,640 | 628,439 |
| 경상남도 | 동 | 10,094 | 607,248 | 1,039,067 |
| | 읍면 | 7,760 | 448,759 | 642,648 |
| 제주도 | 동 | 2,141 | 129,071 | 212,980 |
| | 읍면 | 862 | 50,128 | 72,450 |

○ 분류지표 생성

- 전국 265,341개 조사구별로 여성 가구원수, 50세 이상 가구원수, 대졸이상 가구원수, 자가 가구수를 집계하여 조사구별 여성구성비, 50세 이상 구성비, 대졸이상 구성비, 자가 구성비를 생성한다.
- 생성된 분류지표의 중위값을 기준으로 상·하로 2개 층으로 분류하여 조사구를 구분한다.

<표 15> 분류지표 기준

(단위 : 명)

| 분류지표 | 중위값 | 부차층 |
|---|---|---|
| 성별 구성비 | 50.5% | ① 50.5% 이상, ② 50.5% 미만 |
| 50세 이상 구성비 | 26.3% | ① 26.3% 이상, ② 26.3% 미만 |
| 대졸 이상 구성비 | 17.1% | ① 17.1% 이상, ② 17.1% 미만 |
| 자가비율 구성비 | 59.2% | ① 59.2% 이상, ② 59.2% 미만 |

○ 조사구 추출

- 분류지표별로 집계된 조사구를 여성구성비, 50세 이상 구성비, 대졸 구성비, 자가 구성비에 따라 분류된 부차층에 따라 조사구를 정렬하여 층별 가구수를 기준으로 확률비례계통추출방법을 이용하여 층별로 할당된 조사구를 추출한다.
- 분류지표에 의해 정렬된 조사구 집계된 조사구를 여성구성비, 50세 이상 구성비, 대졸 구성비, 자가구성비에 따라 아래와 같이 조사구를 정렬하여 추출한다.

<표 16> 조사구 정렬 기준(예시)

| 조사구번호 | 여성구성비 | 50세이상구성비 | 대졸 구성비 | 자가 구성비 | 표본 |
|---|---|---|---|---|---|
| 11100002354 | 50.5% 이상 | 26.3% 이상 | 17.1% 이상 | 59.2% 이상 | 선택 |
| 11100002355 | 50.5% 이상 | 26.3% 이상 | 17.1% 이상 | 59.2% 미만 | |
| 11100002356 | 50.5% 이상 | 26.3% 이상 | 17.1% 미만 | 59.2% 이상 | |
| 11100002357 | 50.5% 이상 | 26.3% 이상 | 17.1% 미만 | 59.2% 미만 | 선택 |
| 11100002358 | 50.5% 이상 | 26.3% 미만 | 17.1% 이상 | 59.2% 이상 | |
| 11100002359 | 50.5% 이상 | 26.3% 미만 | 17.1% 이상 | 59.2% 미만 | |
| 11100002360 | 50.5% 이상 | 26.3% 미만 | 17.1% 미만 | 59.2% 이상 | 선택 |
| 11100002361 | 50.5% 이상 | 26.3% 미만 | 17.1% 미만 | 59.2% 미만 | |
| 11100002362 | 50.5% 미만 | 26.3% 이상 | 17.1% 이상 | 59.2% 이상 | |
| 11100002363 | 50.5% 미만 | 26.3% 이상 | 17.1% 이상 | 59.2% 미만 | 선택 |
| 11100002364 | 50.5% 미만 | 26.3% 이상 | 17.1% 미만 | 59.2% 이상 | |
| 11100002365 | 50.5% 미만 | 26.3% 이상 | 17.1% 미만 | 59.2% 미만 | |
| 11100002366 | 50.5% 미만 | 26.3% 미만 | 17.1% 이상 | 59.2% 이상 | 선택 |
| 11100002367 | 50.5% 미만 | 26.3% 미만 | 17.1% 이상 | 59.2% 미만 | |
| 11100002368 | 50.5% 미만 | 26.3% 미만 | 17.1% 미만 | 59.2% 이상 | |
| 11100002369 | 50.5% 미만 | 26.3% 미만 | 17.1% 미만 | 59.2% 미만 | 선택 |

○ 추출결과

- 목표모집단으로 사용한 인구주택총조사의 일반가구에 거주하는 만25세 이상 64세 이하 개인들의 인적특성과 추출된 조사구의 개인인적특성을 비교한 결과는 아래와 같다. 분류지표에 따라 추출된 조사구의 인구통계학적 특성의 구성비가 모집단의 인구통계학적 특성의 구성비와 유사하게 추출된 것을 확인할 수 있다.

<표 17> 표본조사구와 모집단의 인적특성 비교

(단위 : 명, %)

| 구분 | | 추출 조사구 | | 인구총조사(2005년) | |
|---|---|---|---|---|---|
| | | 명 | 구성비 | 명 | 구성비 |
| 성별 | 남자 | 16,020 | 49.7 | 13,227,492 | 49.7 |
| | 여자 | 16,243 | 50.4 | 13,380,042 | 50.3 |
| 연령 | 25~29세 | 4,344 | 13.5 | 3,546,030 | 13.3 |
| | 30~39세 | 10,098 | 31.3 | 8,123,316 | 30.5 |
| | 40~49세 | 9,586 | 29.7 | 7,960,330 | 29.9 |
| | 50~59세 | 5,952 | 18.5 | 5,099,409 | 19.2 |
| | 60~64세 | 2,283 | 7.1 | 1,878,449 | 7.1 |
| 학력 | 무학 | 450 | 1.4 | 379,274 | 1.4 |
| | 초등학교 | 2,962 | 9.2 | 2,329,194 | 8.8 |
| | 중학교 | 3,815 | 11.8 | 2,989,629 | 11.2 |
| | 고등학교 | 13,286 | 41.2 | 10,797,106 | 40.6 |
| | 대학(2·3년제 대학) | 3,553 | 11.0 | 2,943,839 | 11.1 |
| | 대학교(4년제 대학교) | 6,981 | 21.6 | 6,124,280 | 23.0 |
| | 대학원 이상 | 1,216 | 3.8 | 1,044,212 | 3.9 |
| 주택 점유 형태 | 자기집 | 10,576 | 55.3 | 8,828,100 | 55.6 |
| | 전세 | 4,493 | 23.5 | 3,556,760 | 22.4 |
| | 보증금 있는 월세 | 2,676 | 14.0 | 2,394,557 | 15.1 |
| | 보증금 없는 월세 | 374 | 2.0 | 333,206 | 2.1 |
| | 사글세 | 387 | 2.0 | 284,092 | 1.8 |
| | 무상 | 619 | 3.2 | 490,413 | 3.1 |

## 3절 추정

### (1) 가중값

각 표본단위들이 모집단 단위를 얼마나 대표하는지를 조정하기 위해 가중값을 계산한다.

통계조사에서 가중값을 적용하는 이유는 첫째, 표본설계특성에 따라 존재하는 불균등한 표본단위의 추출확률을 조정하고 둘째, 표본단위의 무응답을 보정하고 셋째, 모집단의 알려진 특성을 일치시키기 위해 보정하기 위함이다.

일반적으로 표본조사의 가중치는 ①설계가중치 산정, ②무응답에 대한 조정, ③모집단 정보를 이용한 조정의 세 가지 과정을 통해서 산정된다. 본 연구에서 가중치는 설계가중치와 무응답 조정의 단계를 거쳐서 산출한다.

본 연구에서는 인구총조사 추출틀을 이용하여 서울 및 광역시(7개 지역)와 기타 도지역을 동지역과 읍면지역(18개 지역)으로 구분하여 총 25개 층으로 구분한다.

표본추출은 층화 2단추출법을 적용하는데, 1차추출단위는 조사구이고 2차추출단위는 가구 및 개인이다. 따라서 조사구 내의 가구별 추출확률이 차이가 있음으로 이를 반영할 수 있도록 가중치를 산정한다.

가중치 작성 과정에서 사용될 기호들을 정리하면 다름과 같다.

- $L$ : 층의 수
- $N_h$ : 층 $h$의 모집단 조사구 수
- $n_h$ : 층 $h$의 표본 조사구 수
- $S_{hi}$ : 층 $h$의 $i$번째 조사구에 대한 크기의 측도

- $S_h = \sum_{i=1}^{N_h} S_{hi}$ : 층 $h$에서 크기의 측도에 대한 총합

- $M_{hi}$ : 층 $h$의 $i$번째 조사구 내 조사착수 가구 수

- $r_{hi}$ : 층 $h$의 $i$번째 표본조사구 내 조사완료 가구 수

① 설계가중치

이 조사의 설계가중치는 각 표본조사구에 대한 표본추출률의 역수와 표본조사구에서 가구조사 착수율의 역수를 곱하여 다음과 같이 산출한다.

$$\text{설계가중치} = \frac{S_h}{n_h S_{hi}} \times \frac{M_{hi}}{m_{hi}}$$

원칙적으로 각 표본 조사구에서는 10가구씩을 표본으로 조사하였기 때문에 $m_{hi}$=10 이다. $S_{hi} \approx M_{hi}$를 가정할 수 있는 경우 각 층에서 설계가중치는 해당 층 내에서 일정한 값이 되어 설계가중치=$\frac{S_h}{n_h \times 10}$으로 표현할 수 있다. 이때 각 지역 내 층에서 표본 가구들은 모두 동일한 설계가중치를 갖게 된다.

② 무응답 조정

해당 지역 내 세부 층에서 조사 가구에 대한 설계가중치는 원칙적으로 같다. 본 조사에서 무응답 조정은 표본 조사구 단위로 진행된다. 무응답 조정계수는 다음 식에 따라 구한다.

$$\text{무응답 조정계수} = \frac{m_{hi}}{r_{hi}}$$

### (2) 추정

#### □ 추정량

본 조사에서 주요 추정대상은 모평균이나 모비율이다. 모비율 추정 및 추정량의 분단은 모평균 추정방법과 동일하다. 다만, 모비율 추정의 경우 조사변수 값이 어떤 특성을 갖고 있는가에 따라 1 또는 0의 값을 갖는다는 점이 차이이다.

본 조사에서 각종 모평균 추정을 위해서 사용된 가중치를 이용한 추정량은 다음과 같이 정의된다.

$$\bar{y} = \frac{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} w_{hij} y_{hij}}{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} w_{hij}} = \frac{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} w_{hij} y_{hij}}{w_{...}}$$

여기서, $w_{hij}$는 각 응답자에 부여된 가중치이고, $y_{hij}$는 각 응답결과로 모비율 추정의 경우는 특성 속성을 갖고 잇는 경우는 1 아니면 0의 값을 갖는다.

$L$은 층의 수, $n_h$는 층 $h$에서의 1차 표본추출단위인 표본조사구의 수, $m_{hi}$는 층 $h$내 $i$번째 표본조사구의 응답자 수이다. $w_{...} = \sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} w_{hij}$은 전체 응답자에 대한 가중치의 합계이다.

#### □ 추정량의 분산

앞서 제시한 모평균 추정량에 대해서 층화와 2단 집락추출 등의 표본설계를 반영한 추정분산은 다음과 같이 계산된다.

$$var(\bar{y}) = \sum_{h=1}^{L} \frac{n_h(1-f_n)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi.} - \overline{e_{h..}})^2$$

여기서, $L$은 층의 수, $n_h$는 층 $h$에서의 1차 표본추출단위인 조사구 수, $m_{hi}$는 층 $h$내 $i$번째 표본조사구의 응답자 수이다.

여기서, $f_h = n_h / N_h$, $e_{hi.} = \left( \sum_{j=1}^{m_{hi}} w_{hij}(y_{hij} - \bar{y}) \right) / w_{...}$, $\bar{e}_{h..} = \left( \sum_{i=1}^{n_h} e_{hi.} \right) / n_h$이다.

모평균 및 모비율 추정에 대한 표준오차, 상대표본오차, 95% 신뢰수준 오차의 한계는 다음과 같다.

$$s.e(\bar{y}) = \sqrt{var(\bar{y})}$$

$$rse(\bar{y}) = \frac{s.e(\bar{y})}{\bar{y}} \times 100(\%)$$

오차의 한계=$1.96 \times \sqrt{var(\bar{y})}$

# 제4장 결론 및 제안

본 연구는 한국교육개발원에서 매년 조사하는 평생학습개인실태조사의 표본설계 개선을 위한 연구이다. 지난 2007년 실시한 통계청의 정기품질진단 결과에서 '표본설계' 부분에 대해 제기된 문제점에 대해 개선안을 제공하는데 그 목적이 있다.

## □ 결론

○ 조사현황을 분석하여 현 표본설계의 문제점을 파악하고 체계적이고 현 상항을 잘 대표할 수 있는 현실가능한 표본설계방안을 제시한다.

○ 조사 현황을 검토한 결과, 인구통계학적 특성이 전혀 고려되지 않은 표본설계로 인하여 응답자 특성이 편향되어 있다. 따라서 인구통계학적 특성이 고려된 표본설계 방법을 제시한다.

- 먼저 2010년도 인구총조사 자료를 토대로 모집단을 정의하고 표본설계에 활용가능한 표본추출틀을 마련한다.
- 서울 및 광역시와 기타 도지역을 동지역 및 읍면지역으로 층화하여 행정구역을 25개로 층화한다.
- 표본크기는 조사예산을 고려하여 모집단 분포에 따라 비례하게 지역별로 할당한다.
- 인구통계학적 특성을 반영하기 위해 성별, 연령, 학력, 점유형태의 특성을 분류지표로 선정하여 추출함으로써 부차층의 효과를 얻는다.
- 추출된 가구에 대해서는 조사대상자로 해당되는 가구원을 모두 조사하는 것을 원칙으로 한다.

○ 본 연구를 통해 인구통계학적 특성이 반영된 조사대상자를 선정함으로써 모집단과 가장 유사한 표본을 대상으로 조사가 이루어질 것으로 예상된다.

□ 제안

○ 본 연구를 수행하면서 향후 조사모집단의 관리 및 공표단위의 변화에 따른 표본설계 방법의 변경 등이 필요할 것으로 보인다.

- 향후 조사에서는 조사모집단인 2010년 인구총조사의 모집단 분포가 변화함으로서 이에 맞게 적절한 조사모집단 보완이 필요로 한다. 신규 아파트 가구명부 등을 보완하여 모집단을 변화에 맞게 적절한 조사모집단을 구축하여야 한다.

- 표본규모를 증가하여 공표단위를 전국에서 시·도 단위에서 공표할 경우 시·도 단위 공표에 맞게 표본설계 변경이 불가피할 것으로 사료된다.

- 그리고 조사결과의 편향을 방지하기 위해 한 가구에서의 응답대상자를 주민등록상 생일이 가장 빠른 1명만 작성하도록 하는 것이 표본가구 수는 늘어나지만 조사결과의 편향은 감소할 것으로 판단된다.

[참고문헌]


- 2010 평생학습개인실태조사(한국교육개발원, 2010)
- 표본조사의 이해와 활용(김영원·류제복·박진우·홍기학, 2009)
- 축산물생산비조사를 위한 복합표본설계(김수택, 2008)
- 표본조사를 위한 설계과정(류제복, 2008)
- 노동통계조사를 위한 표본설계(매월노동통계조사, 노동력수요동향조사를 중심으로-, 응용통계연구 제10권 2호, p215-226.)
- Sampling Techniques, Wiley(Cochran, W. G., 1977)
- Survey Sampling, Wiley(Kish, L., 1965)
- Sampling: Design and Analysis, Duxbury Press(Lohr, S., 1999)
- SAS/STAT User's Guide, Version 8(SAS, 1999)