Sampling Quality Management Manual

표본 품질관리 매뉴얼

Sampling

통계청
Korea National Statistical Office

국가통계 표본 품질관리 매뉴얼 발간에 부쳐

국가통계의 작성단계가 다 중요하지만, 통계의 정확성을 좌우하는 가장 중요한 요소 하나를 들라면 표본설계과정이라고 할 수 있을 것입니다. 조사의 기획과 현장조사가 아무리 잘 되어도 표본의 대표성이 낮은 조사라면 그 조사의 결과는 신뢰하기 어렵습니다.

그 동안의 우리나라 국가통계의 관리 과정을 살펴보면 표본관리 문제는 그 중요성에 비해 매우 가볍게 다루어져 온 것이 사실입니다. 표본의 추출과 관리는 전문가의 영역으로 간주되어 일선의 통계 실무자들이 접근할 수 없는 영역으로 다루어져왔기 때문입니다. 많은 경우 표본 문제는 외부 전문가에게 용역 의뢰하는 것으로 처리하는 형편이었으므로, 통계 실무자들이 소관 통계의 표본에 문제가 있어도 잘 인식하지 못하는 경우가 많았습니다.

이러한 문제를 해결하고자 이번에 통계청에서 표본관리 매뉴얼을 발간하게 되었습니다. 이번에 발간하는 매뉴얼에는 표본추출의 과정별 자세한 설명과 함께 표본의 사후관리 문제까지 포함되어 있어, 대부분의 표본관리 과제를 망라하고 있습

니다. 또 일선의 통계조사 실무자들이 표본관리를 잘 수행하고 있는지의 여부를 손쉽게 파악할 수 있도록 표본관리 단계별 체크리스트도 첨가하였습니다.

이 매뉴얼은 어려운 수식은 가능한 한 줄이고 현장 실무자 중심으로 최대한 이해하기 쉽게 서술하였기 때문에 비전문가라도 큰 어려움 없이 활용 가능하리라 생각합니다.

아무쪼록 이 책이 국가통계조사 현장의 실무자들뿐 아니라 민간 조사기관의 표본 실무자들에게도 널리 활용되어 국가통계의 품질을 향상하기 위한 유용한 지침서가 되기를 기대합니다.

통계청에서는 앞으로도 통계조사의 각 분야별 실무 매뉴얼을 계속 발간할 것을 약속드립니다.

2007. 10.

통계청장 이 창 호

표본 품질관리 매뉴얼

CONTENTS

Sampling Quality & Management Manual

Sampling Quality Management Manual

표본 품질관리 매뉴얼

1

표본설계란 무엇인가?

표본설계는 통계조사의 정확성에 크게 영향을 미치는 매우 중요한 작업이다. 이 장에서는 표본설계가 구체적으로 무엇이며, 표본설계를 할 때 유의할 점은 무엇인지, 표본설계는 어떤 순서로 이루어지는지를 알아본다.


표본설계란
무엇일까요?
표본
설계


표본설계는 통계조사에서
가장 기본이 되는 것이죠.

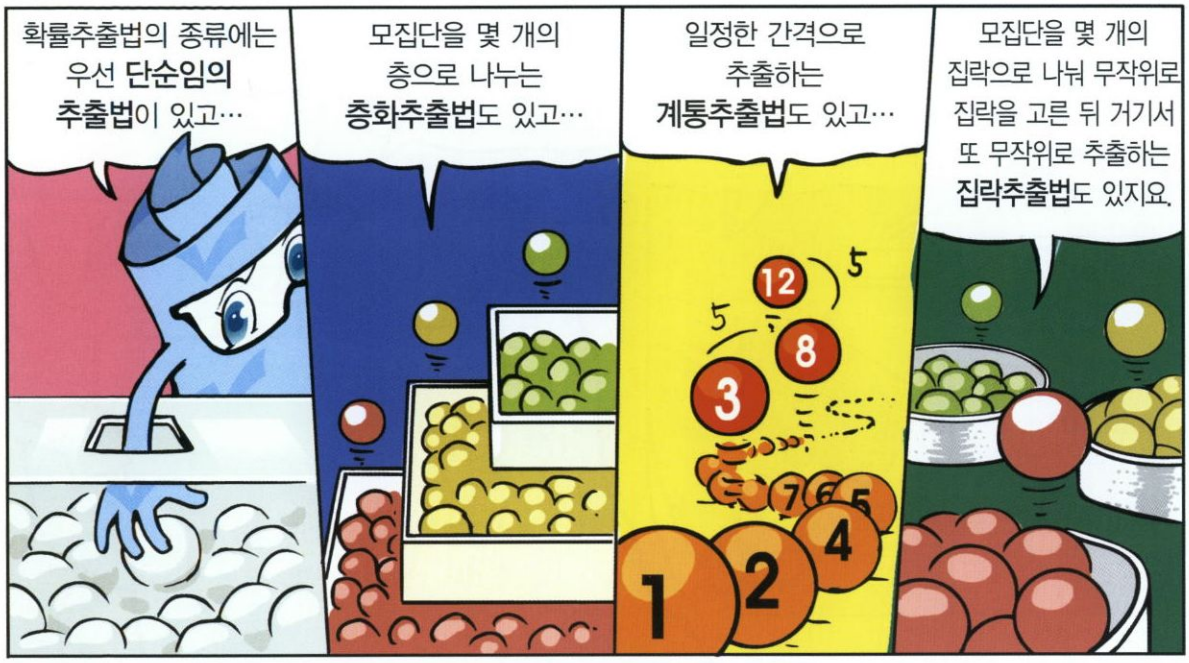

자동차로 말하면 엔진에
해당된다고 할까…

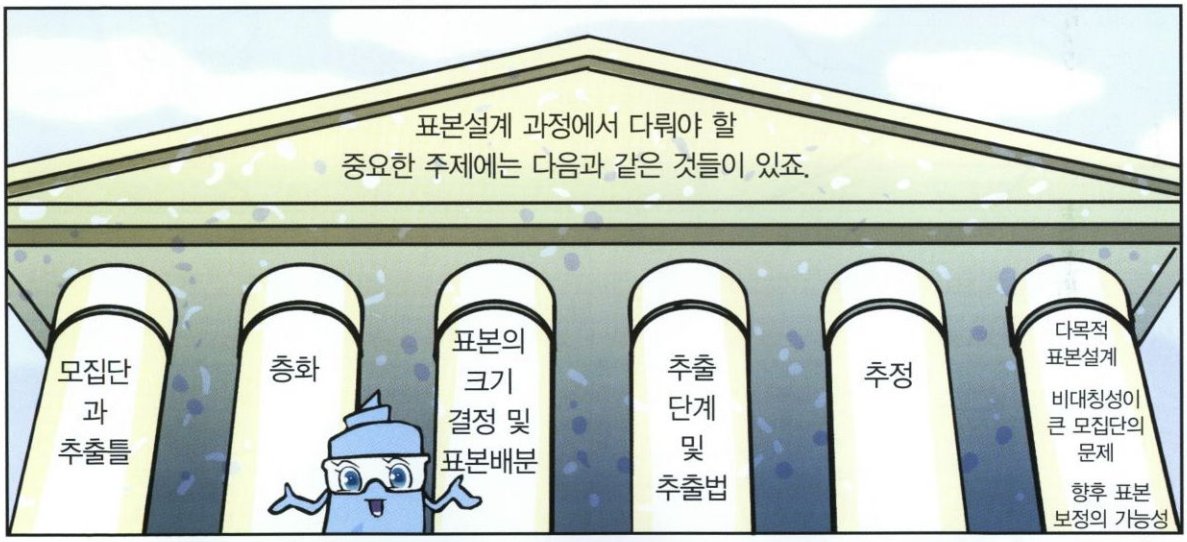

이 많은 모집단을 언제 다 조사하겠어요.
그래서 이들을 대표할 수 있는 일부의 단위
집합을 추출하는 것을 표본설계라고 하지요.
와글 바글 와글 바글

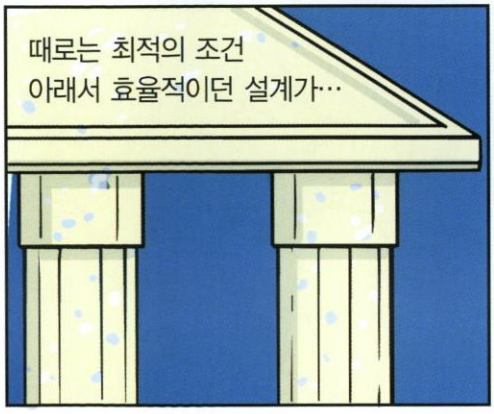

그러면 어떻게 표본을 추출하느냐…


모집단을 대표하는 표본설계는
우리야~


모집단에 속해 있는
각 구성원이 표본으로
선택될 가능성이 일정하게
되도록 하는 확률추출법을
적용하는 것이
바람직합니다.


확률추출법의 종류에는 우선 **단순임의 추출법**이 있고…
모집단을 몇 개의 층으로 나누는 **층화추출법**도 있고…
일정한 간격으로 추출하는 **계통추출법**도 있고…
모집단을 몇 개의 집락으로 나눠 무작위로 집락을 고른 뒤 거기서 또 무작위로 추출하는 **집락추출법**도 있지요.
12
5
5
8
3
7
6
5
4
1
2


표본설계 과정에서 다뤄야 할 중요한 주제에는 다음과 같은 것들이 있죠.
모집단과 추출틀
층화
표본의 크기 결정 및 표본배분
추출 단계 및 추출법
추정
다목적 표본설계
비대칭성이 큰 모집단의 문제
향후 표본 보정의 가능성


때로는 최적의 조건 아래서 효율적이던 설계가…


예상치 못하게 조건의 일부가 변할 때는 문제를 일으키기도 해요.
앗! 위험해!

따라서 예상치 못한 상황에도 충분히 대처 가능한 표본설계가 이루어져야 하죠. 표본설계는 효율성뿐만 아니라 강건성까지 고려해야 하는 겁니다.

이런 요소들이 표본설계에 영향을 미치니 항상 신경을 써야 합니다.

① 정확도 – 틀리면 안돼!

② 예산상의 제약

조사 비용이 너무 많이 드는 방법을 써선 곤란하겠지.

③ 표본추출방법의 이용 가능성

방법을 정해 놨는데, 그 방법을 쓸 수 없다고 하면 곤란해.

④ 활용 가능한 보조정보

조사 단위에 관한 정보를 많이 얻을 수 있는 추출틀 (선거인 명부, 전화번호부, 가구명부)을 마련할 것!

⑤ 이용될 조사기법

우편조사를 할까, 전화조사를 할까, 아니면 인터넷조사를 할까…

표본설계의 순서는 어떻게 될까?

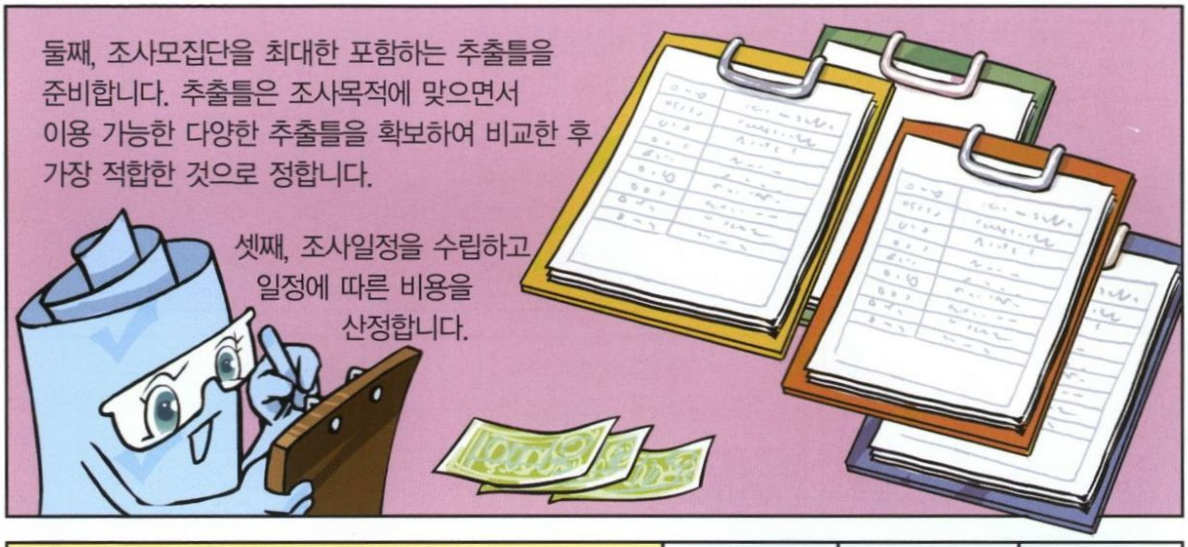
첫째, 조사대상이 되는 모집단인 조사모집단을 정의합니다.

가령 사업체 조사를 한다고 하면…

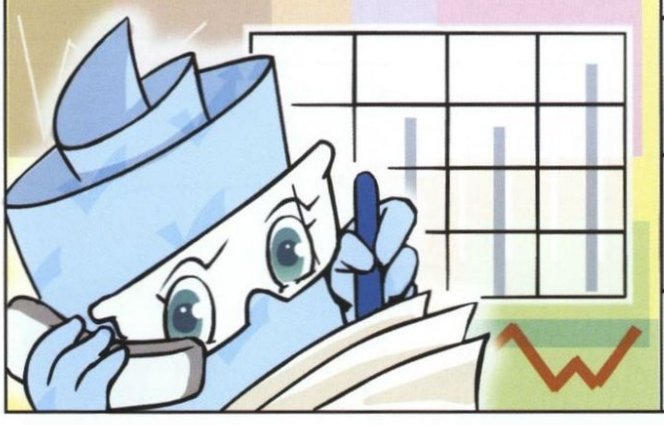
종업원 수나 매출 규모 등에 대해서 정의를 하고…

둘째, 조사모집단을 최대한 포함하는 추출틀을 준비합니다. 추출틀은 조사목적에 맞으면서 이용 가능한 다양한 추출틀을 확보하여 비교한 후 가장 적합한 것으로 정합니다.

셋째, 조사일정을 수립하고 일정에 따른 비용을 산정합니다.



넷째, 실제 조사에 필요한 표본크기와 표본추출법, 조사방법을 결정하는 등 구체적인 계획들을 세웁니다.



다섯째, 모수 추정식을 만듭니다.

$$20 - 1.75 \times \sqrt{15.875/17}$$

$$20 + 1.75 \times \sqrt{15.875/17}$$

표본크기가 클수록 조사의 정밀도는 높아지죠.

각계각층의 다양한 조사로 정확하다고 자부!

하지만 너무 커지면 조사비용이나 노력이 많이 들기 때문에 조사의 질을 떨어뜨리는 요인이 될 수도 있죠.

돈은 돈대로 들고… 힘은 힘대로 들고… 대충 하자!

모르겠다~

그러므로 조사목적에 맞는 목표 정도를 정한 후 그것을 만족시키는 범위 내에서 표본의 크기를 가능한 한 작게 하는 것이 바람직하죠.

표본설계를 할 때는 반드시 검토해야 할 사항들이 있어요.

가장 먼저 조사목적을 정해야 합니다.

조사목적

다음으로 조사범위와 조사단위를 결정해야 합니다.

잠깐! 조사단위는 조사의 대상을 뜻하는데 통계집단을 구성하는 단위와 반드시 일치하는 것은 아니에요.

예를 들면 인구주택총조사에서 조사단위는 개인이지만…

실제 조사는 가구를 단위로 해서 이루어지죠.

이 가구의 구성원은 어떻게…

다음으론 조사사항을 결정해야 하는데 이 부분이 통계조사기획 중 가장 중요한 과정의 하나예요.

조사사항을 결정할 땐 이 점을 반드시 고려해야 합니다.

① 응답자가 사실 그대로 답할 수 있는 사항인가?

② 쉽게 이해할 수 있는 사항인가?

③ 객관적 파악이 가능한 사항인가?

④ 수량에 관한 사항에서 응답자가 장부나 기록을 갖고 정확히 응답할 수 있는가?

다음엔 조사방법을 선택해야 하는데…

전부를 조사하는
전수조사를 선택할 것이냐

일부를 조사하는
표본조사를 선택할 것이냐

조사방법의 특성과 조사목적, 경비, 시간 등을
고려해서 가장 적합한 조사방법을
선택해야 합니다.

마지막으로
조사기준시점,
대상기간, 조사기간
등을 결정해야 하죠.

예를 들어
'2005년 11월 1일
0시 현재의 인구'

일정표

조사기준시점

예를 들어 '2007년 1월
1일에서 12월 31일까지의
통계 생산량'

대상기간

실제 조사기간으로
예를 들어 '2007년 3월
1일부터 3월 31일까지'

조사기간

조사방법에는
아래와 같은 것들이
있지요.

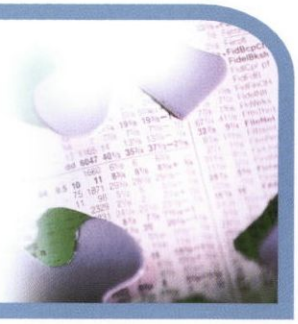
① 면접조사

② 전화조사

③ 우편조사

④ 인터넷조사

1 표본설계란?



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 한다.

표본설계는 다음과 같은 과정을 통해 이루어진다.

먼저 조사대상이 되는 모집단인 조사모집단을 명확하게 정의한다. 정확한 표본조사를 위해서는 개념적으로 정의한 목표모집단과 표본추출을 위한 실제 조사대상인 조사모집단의 차이가 최소가 되도록 하는 것이 중요하다.

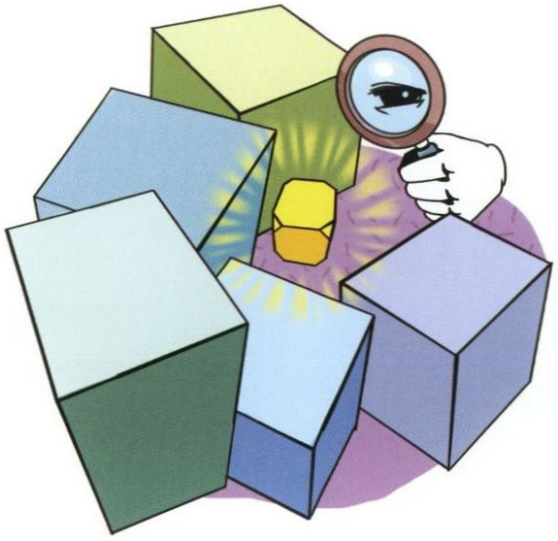
모집단이 정해지면 조사모집단을 최대한 포함하는 추출틀을 준비한다. 추출틀

은 조사목적에 맞으면서 이용 가능한 다양한 추출틀을 확보하여 비교한 후 가장 적합한 것으로 정한다. 이때 모집단의 조사단위를 빠뜨리거나 중복하는 일이 없도록 해야 하며, 동일한 모집단에 대한 조사에는 가능한 한 동일한 추출틀을 사용해서 일관성을 유지하도록 한다.

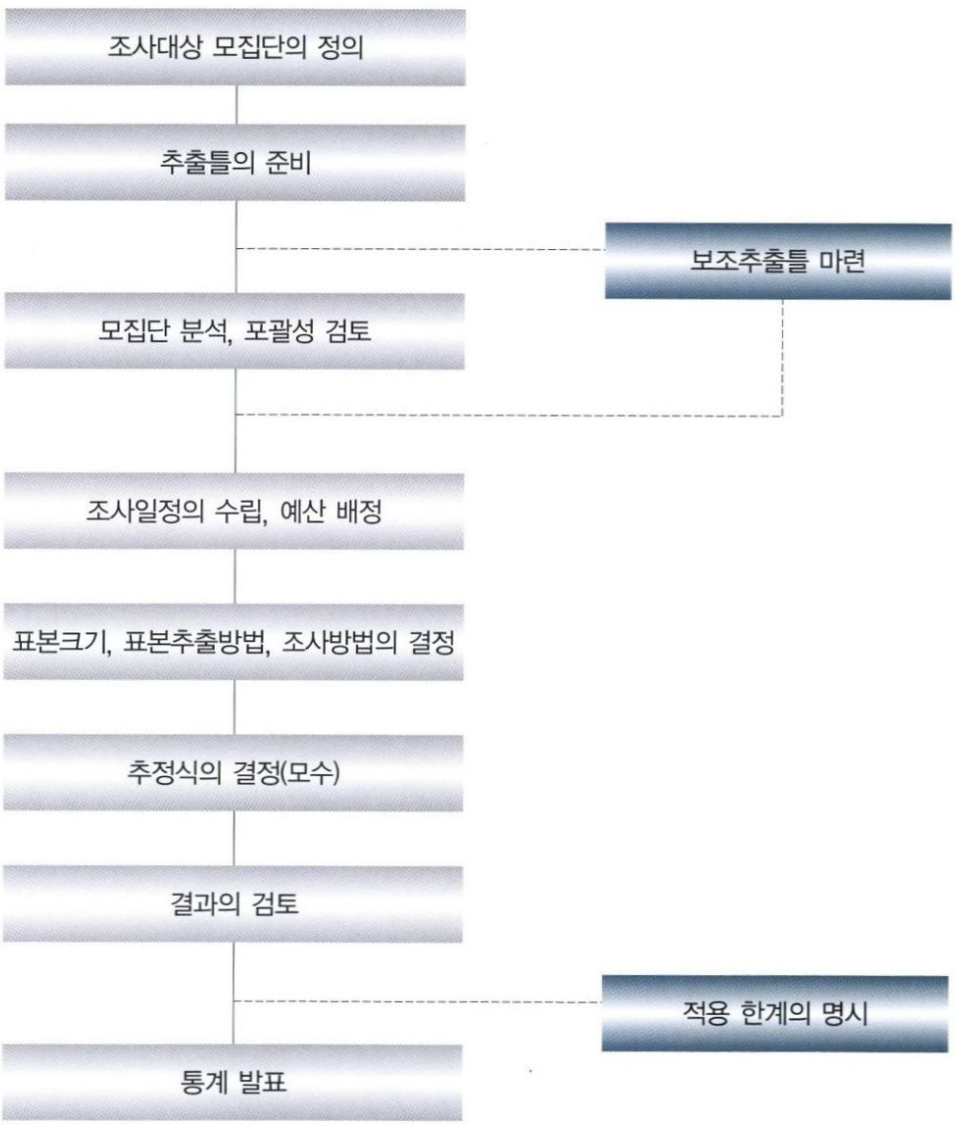
추출틀이 마련되면, 조사일정을 수립하고 일정에 따른 비용을 산정하여 예산을 배정한다.

그런 다음, 실제 조사에 필요한 표본크기와 가장 효율적인 추출법을 정하고, 그에 따라 층화와 표본배분, 표본추출단계, 조사방법 등을 결정하는 등 구체적인 계획들을 마련한다.

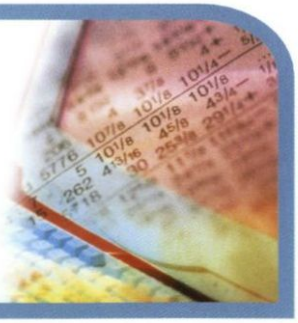
다음으로 모수 추정식을 만든다. (모수는 모집단의 특성 – 모평균, 모비율, 모분산 – 을 나타내는 값이며 모수 추정식은 이를 추정하기 위한 식이다.)



표본설계의 흐름도


조사대상 모집단의 정의
추출틀의 준비
보조추출틀 마련
모집단 분석, 포괄성 검토
조사일정의 수립, 예산 배정
표본크기, 표본추출방법, 조사방법의 결정
추정식의 결정(모수)
결과의 검토
적용 한계의 명시
통계 발표

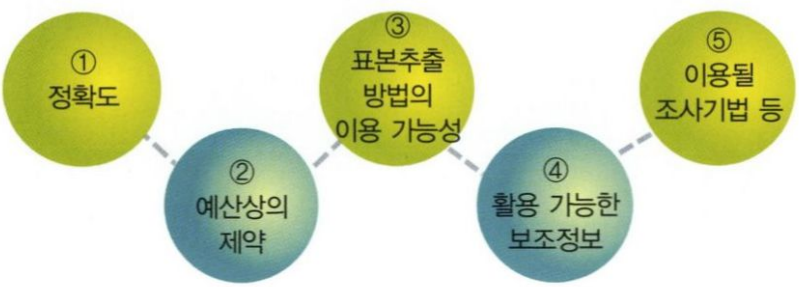
2 표본설계를 할 때 유의할 점



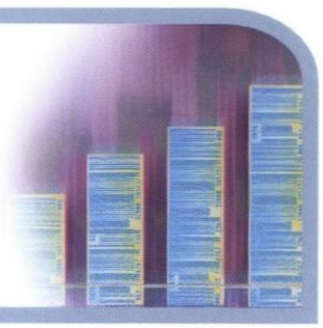
표본설계 과정에서 중요하게 다루어야 할 것으로는 모집단과 추출틀, 층화, 표본의 크기 및 표본배분, 추출단계 및 추출법, 추정식 등이 있으며, 그 밖에 다목적 표본설계, 비대칭성이 큰 모집단의 문제, 향후 표본보정의 가능성 등도 고려해야 한다.

어떤 설계는 최적의 조건에서는 매우 효율적이지만 예상치 못하게 일부 조건이 변할 때는 문제를 일으키기도 한다. 따라서 어떤 상황에도 충분히 대처 가능한 표본설계가 이루어져야 한다. 이처럼 표본설계는 효율성뿐만 아니라 강건성(robustness)까지 고려해야 한다.

표본설계에 영향을 미치는 요소들



3 표본설계 시 검토할 사항



조사목적의 설정

조사주제를 정하고 주제와 관련된 개념들을 정리하여 조사목적을 정해야 한다. 새로운 조사를 개발하거나 기존의 조사를 다시 설계할 때는 조사의 목적을 명확하고도 구체적으로 규정하는 것이 매우 중요하다. 같은 조사라 하더라도 조사목적에 따라 조사의 방향이나 규모, 방법이 달라질 수 있기 때문이다.

조사범위 및 조사단위의 결정

조사범위는 조사목적에 의해서 기본적으로 규정되지만, 인원, 비용, 시간 등의 제반 여건을 감안하여 결정해야 한다.

조사단위는 조사대상의 단위를 의미하며, 통계집단을 구성하는 단위와 반드시 일치하지는 않는다. 예를 들면 인구주택총조사에서 조사단위는 개인이지만 실제 조사는 가구를 단위로 하여 이루어진다.

*조사단위는 집계단위, 표준분류 적용단위, 표본추출단위 등과도 구별되어야 한다.

조사사항의 결정

조사사항을 결정하는 것은 통계조사기획 중 가장 중요한 과정의 하나이다. 조사사항은 조사할 내용으로서 흔히 질문 문항으로 표현되는데, 조사목적에 따라 결정하되 조사할 조직의 역량, 조사원 및 응답자의 부담 등을 고려하여 신중하게 결정해야 한다.

조사사항을 결정할 때 유의해야 할 사항들

① 응답자가 사실 그대로 답할 수 있는 사항인가?
② 응답자가 쉽게 이해할 수 있는 사항인가?
③ 객관적 파악이 가능한 사항인가?
④ 수량에 관한 사항에서 응답자가 장부나 기록을 갖고 정확히 응답할 수 있는가?

조사실시 방법의 선택

통계조사는 조사대상의 전부를 조사하느냐, 일부를 조사하느냐에 따라 전수조사와 표본조사로 구분할 수 있으며, 조사표의 배부, 기입 및 수집 방법에 따라 타계식조사와 자계식조사로 구분할 수 있다. 조사실시 방법에 따라서는 면접조사, 전화조사, 우편조사, 인터넷조사 등으로 나눌 수 있는데, 이 중 어떤 방법을 선택할 것인가는 각 조사방법의 특성과 조사목적, 경비, 시간 등을 고려하여 가장 적합한 방법으로 정한다.

① 면접조사

조사자가 응답자를 직접 만나 조사하는 개별면접(personal interview) 혹은 대면면접(face-to-face interview)을 의미한다. 넓은 의미로는 조사자가 응답자와 접촉하여 조사하는 모든 방법을 말한다. 개인주의가 만연되어가는 현대 사회에서는 점점 이 방법을 사용하기가 어려워지고 있는 실정이다.

② 전화조사

조사원이 응답자에게 전화를 걸어 질문하고 응답자의 대답을 기록하여 자료를 수집하는 방법이다. 선거 여론조사나 시장조사 등 신속히 진행해야 하는 조사에 많이 사용한다.

③ 우편조사

조사대상자가 우편으로 보내진 설문지에 응답을 기입하여 다시 우편으로 반송하게 함으로써 자료를 수집하는 방법이다.

④ 인터넷조사

인터넷을 활용하여 조사자료를 수집하는 방법, 인터넷 사용자들을 대상으로 웹 또는 전자메일로 설문을 진행하고 응답하는 방법 등을 가리킨다.

조사기준시점, 대상기간, 조사기간의 결정

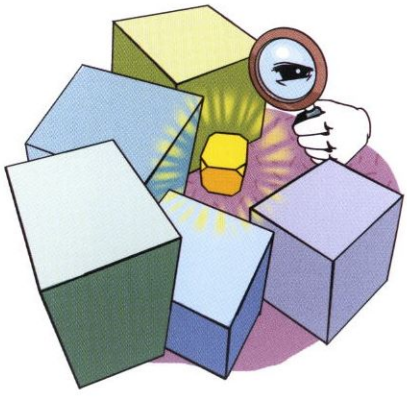
조사기준시점은 '2005년 11월 1일 0시 현재의 인구' 등과 같이 파악하고자 하는 정보의 시간 기준을 말하는 것으로 조사의 목적에 따라 결정한다.

대상기간은 '2007년 1월 1일에서 12월 31일까지의 부가가치 생산액' 등과 같이 정해진 조사대상기간을 말하며, 조사내용을 명확하게 파악할 수 있고 조사결과의 비교가 가능하도록 정해야 한다.

조사기간은 실제 조사를 실시하는 기간으로, 조사대상의 수와 조사내용 및 조사원 수에 따라 결정되며 허용되는 범위 안에서 최대한 줄이는 것이 바람직하다.

참고 조사의 종류

구분	조사명	내용
조사대상에 따라	전수조사	조사대상이 전부일 경우
	표본조사	조사대상이 일부일 경우
조사지역에 따라	전국조사	조사대상이 전국일 경우
	지역조사	조사대상이 지역일 경우
기입방식에 따라	타계식조사	조사원이 기입하는 방식
	자계식조사	응답자가 직접 기입하는 방식
조사목적에 따라	실태조사	실제 현상을 대상으로 하는 조사
	의식조사	가족의식, 정치의식 등 의식을 대상으로 하는 조사
계속성여부에 따라	연속조사	주기적으로 계속하는 조사
	일회성조사	한 번으로 끝나는 조사
조사내용에 따라	복지조사, 노동조사, 교육조사, 경제조사 등	
조사결과의 이용목적에 따라	여론조사, 학술조사, 시장조사 등	
조사실시 방법에 따라	면접조사, 전화조사, 우편조사, 인터넷조사	

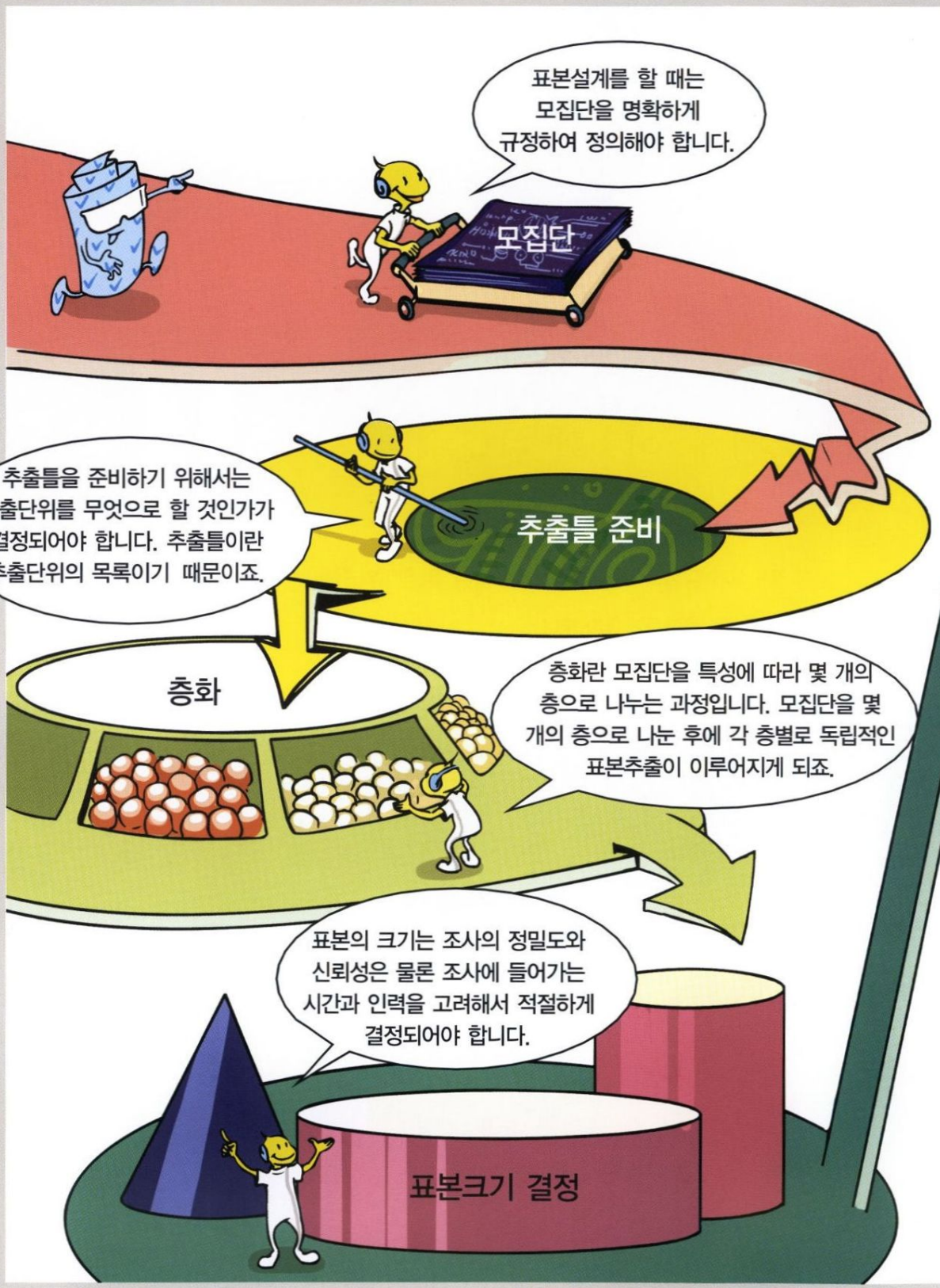


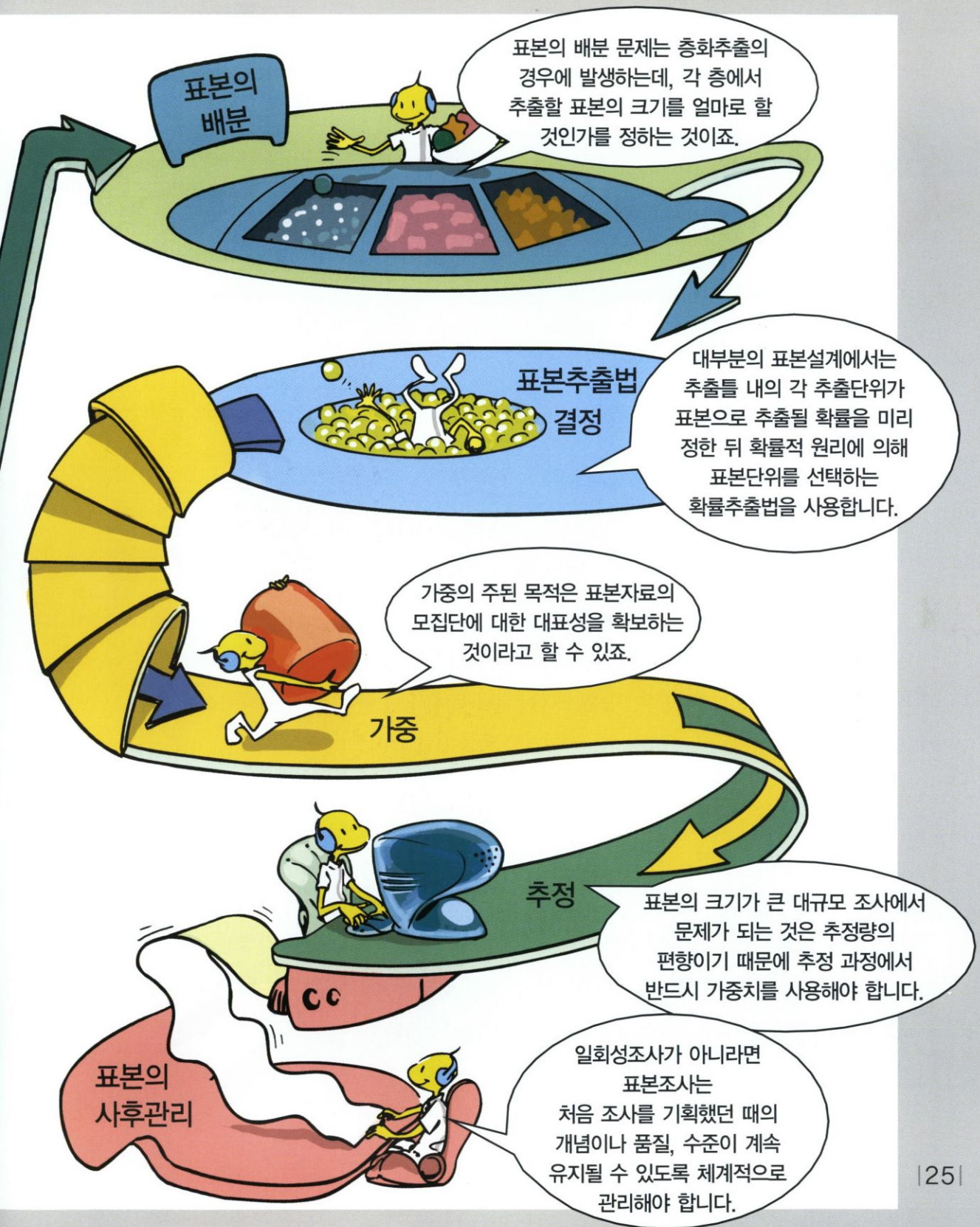
표본추출을 어떻게 하는가는 통계조사의 성패를
좌우할 뿐만 아니라 나아가 통계의 품질에도 직접적인 영향을
미치게 된다. 표본추출을 하기 위해서는 표본의 크기(규모)나
추출방법 등을 결정하는 일련의 과정이 선행되어야 하는데
이를 표본설계라 한다.

표본설계, 이렇게 한다

II

표본설계 과정을 하나하나 살펴보면서 그 의미와 유의할 점, 그리고 구체적인 실행 사항들을 꼼꼼하게 짚어본다.


표본설계를 할 때는
모집단을 명확하게
규정하여 정의해야 합니다.
모집단
추출틀을 준비하기 위해서는
추출단위를 무엇으로 할 것인가가
결정되어야 합니다. 추출틀이란
추출단위의 목록이기 때문이죠.
추출틀 준비
층화
층화란 모집단을 특성에 따라 몇 개의
층으로 나누는 과정입니다. 모집단을 몇
개의 층으로 나눈 후에 각 층별로 독립적인
표본추출이 이루어지게 되죠.
표본의 크기는 조사의 정밀도와
신뢰성은 물론 조사에 들어가는
시간과 인력을 고려해서 적절하게
결정되어야 합니다.
표본크기 결정


표본의 배분
표본의 배분 문제는 층화추출의 경우에 발생하는데, 각 층에서 추출할 표본의 크기를 얼마로 할 것인가를 정하는 것이죠.
표본추출법 결정
대부분의 표본설계에서는 추출틀 내의 각 추출단위가 표본으로 추출될 확률을 미리 정한 뒤 확률적 원리에 의해 표본단위를 선택하는 확률추출법을 사용합니다.
가중
가중의 주된 목적은 표본자료의 모집단에 대한 대표성을 확보하는 것이라고 할 수 있죠.
추정
표본의 크기가 큰 대규모 조사에서 문제가 되는 것은 추정량의 편향이기 때문에 추정 과정에서 반드시 가중치를 사용해야 합니다.
표본의 사후관리
일회성조사가 아니라면 표본조사는 처음 조사를 기획했던 때의 개념이나 품질, 수준이 계속 유지될 수 있도록 체계적으로 관리해야 합니다.

1 모집단의 정의

표본조사결과 작성되는 통계는 모집단을 설명하는 통계가 되기 때문에 표본설계를 할 때는 모집단을 명확하게 규정하여 정의해야 한다.

목표모집단

조사목적에 따라 개념적으로 규정하는 모집단을 말한다. 예를 들어 농가경제조사의 경우 목표모집단은 '농업소득이 총소득에서 중요한 부분을 차지하는 모든 가구들의 집합' 으로 정의할 수 있다.

조사모집단

실제 표본설계를 위해 규정하는 모집단이다. 위에서 예를 든 농가경제조사의 경우 실제로 각 가구의 소득을 미리 알 수 없기 때문에 조사모집단은 '경지면적 300평 이상을 경작하는 가구의 집합' 으로 정의할 수 있다.

목표모집단과 조사모집단의 차이 검정

목표모집단과 조사모집단은 가능한 한 일치하는 것이 바람직하지만, 실제로는 그렇지 못한 경우가 많다. 위의 예에서처럼 일반적으로 조사모집단은 목표모집단보다 제한되어 있기 때문이다. 따라서 두 모집단의 차이를 분석하는 작업이 반드시 필요하며, 만약 두 집단 간의 차이가 클 경우에는 다른 자료로 이 차이를 보충하는 방안을 강구해야 한다.

모집단을 정의할 때 명확히 규정해야 할 사항

① 모집단을 구성하는 조사단위들의 특성과 유형

② 모집단의 지리적 · 공간적 · 시간적 범위. 특히 연속 조사인 경우에는 시간의 흐름에 따라 모집단이 변동하는 것도 적절히 반영할 수 있도록 해야 한다.

2 추출틀의 준비

조사모집단이 정의되고 나면 그 모집단을 묘사할 수 있는 틀이 필요한데, 이를 추출틀이라고 한다. 추출틀을 준비하기 위해서는 추출단위(sampling unit)를 먼저 결정해야 한다. 추출틀이란 추출단위의 목록이기 때문이다.

추출틀은 조사의 특성에 적합하면서도 모집단에 포함된 조사단위들의 중복이나 누락을 최소화할 수 있어야 한다. 또한 추출틀의 설정, 이용, 유지 및 보완 등은 실제 적용이 가능하고 예산에 무리가 따르지 않는 범위 내에서 이루어져야 한다.

추출틀 작성 단계

1단계 : 추출단위의 선택 – 비용, 정보 형태, 단위의 안정성, 시간을 고려하여 선택

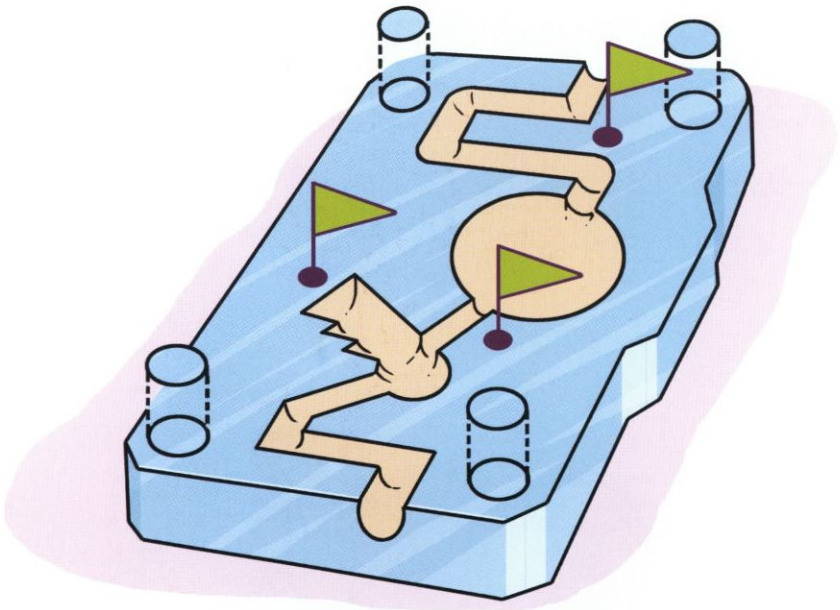
2단계 : 추출틀의 전용 – 정보의 수집 및 조직화 관리

3단계 : 추출틀의 타당도 검토 – 검토 범위(과소 포함, 과대 포함)와 정보의 품질 검토

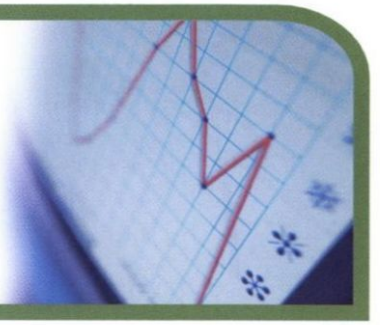
4단계 : 행정조직 – 실사조직 검토

5단계 : 유지관리 – 수정 및 보완 작업

추출틀은 표본설계뿐만 아니라 자료수집, 추적조사, 추정, 품질평가, 분석 등의 과정에서 고루 이용된다. 추출틀에 포함오차가 생기고, 추출틀 내 조사단위들의 특성을 나타내는 자료들이 낡았으면 이로 인한 편향이나 오차가 생길 가능성이 커진다. 그러므로 추출틀에는 높은 수준의 품질이 요구된다.

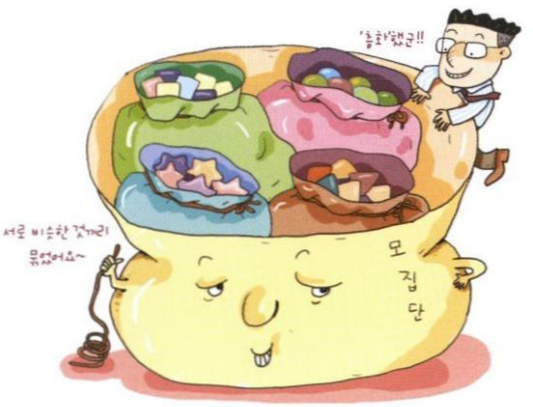


3 효과적인 층화

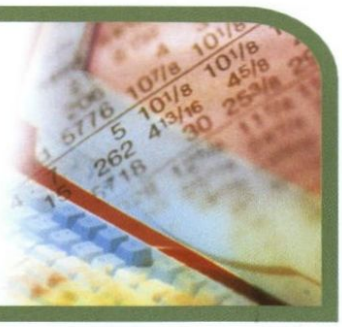


표본설계에서 조사목적에 부합하도록 효과적인 층화를 하는 것은 무엇보다도 중요한 일이다. 층화란 모집단을 특성에 따라 서로 동질적인 몇 개의 그룹(층)으로 나누는 과정이다. 모집단을 몇 개의 층으로 나눈 후에 각 층별로 독립적인 표본추출이 이루어지는데, 효과적인 층화가 이루어질수록 추정의 효율을 높일 수 있으며 부분 통계의 생산이 가능하고 경비를 줄일 수 있다.

기업체나 사업체를 대상으로 층화하는 경우 층화변수로 이용되는 것은 산업의 종류, 매출액, 종사자 수, 지역 등이다.



4 표본의 크기 결정



표본설계를 할 때는 조사목적과 여건에 맞도록 표본의 크기를 결정해야 한다. 표본의 크기는 조사의 정밀도와 신뢰도는 물론 조사에 들어가는 시간과 인력에도 영향을 끼치므로 적절하게 결정되어야 한다.

표본의 크기와 조사의 정확도

표본의 크기가 커질수록 조사의 정확도는 높아진다. 그러나 표본이 커지면 조사비용이나 조사를 위한 노력이 많이 들기 때문에 경우에 따라서는 오히려 조사의 질을 떨어뜨리는 요인이 될 수도 있다. 그러므로 표본의 크기는 조사목적에 맞는 목표정도(target precision)를 정한 후 그것을 만족시키는 범위 내에서 정하는 것이 바람직하다.

표본의 크기가 증가하면 표본분산이 감소하기 때문에 추정량의 정도는 좋아진다. 따라서 추정량의 정도는 표본의 크기에 의존하며, 추정량의 표본오차, 오차의 한계, 또는 변동계수 등으로 나타낼 수 있다. 보통 어떤 표본조사에서 표본의 크기를 구한다는 것은 주어진 추정량의 정도를 만족하도록 구하는 것을 의미한다.

표본조사에서 발생하는 표본오차

표본조사에서는 표본을 조사해서 얻은 결과를 모집단 전체에 대한 것으로 일반화하기 때문에 필연적으로 표본조사로 인한 오차가 발생하는데, 이를 표본오차라 한다. 모수를 추정하는 추정량(추정식)에 추출된 표본들의 값을 대입하면 추정값을 구할 수 있다. 표본추출과정에서 어떤 조사단위들이 표본에 포함되는가에 따라 추정값이 변하는데, 표본에 따른 추정값의 변동 정도를 수치로 나타낸 것이 추정량의 표본오차이다. 일반적으로 추정량의 표본오차는 추정량의 표준편차로 설명되는데, 추정량의 표준편차를 표준오차(SE:Standard Error)라 한다. 확률추출법에서는 추출방법에 따라 모수를 추정하는 방법 및 표준오차를 계산하는 방법이 달라진다.

표본의 크기 결정하기

표본설계를 할 때는 정해진 목표정도를 달성하기 위해서 어떤 확률추출법을 사용하는 것이 효율적인지 검토하고 이를 바탕으로 구체적인 표본의 크기를 정한다. 추정량의 표준오차는 사용하는 확률추출법에 따라 달라지기 때문에 원하는 목표정도를 달성하기 위한 표본의 크기도 확률추출법에 따라 달라진다.

일반적으로 표본의 크기는 복합표본설계에서 이용 가능한 정보(층별 총합 또는 집락총합)가 없을 경우 먼저 단순임의추출법(simple random sampling)에 근거해서 구한다. 따라서 층화추출법(stratified sampling)이나 집락추출법(cluster sampling) 등에 적합한 표본의 크기를 구하기 위해서는 설계효과(deff:design effect)를 고려해야 한다.

5 표본의 배분

층화를 할 때는 각 층에 표본을 배분(allocation)하는 방법을 고려해야 한다. 즉, 표본을 구성할 때 각 층에서 추출할 표본의 크기를 얼마로 해야 가장 효율적인가를 결정해야 한다. 일정한 비용하에서 추정량의 정도(precision)를 최대로 할 수 있는 최적의 표본배분방법은 다음과 같은 요인들의 영향을 받는다.

① 각 층을 구성하는 총 조사단위의 수 – 층의 크기

② 각 층을 구성하는 조사단위들 간의 변동 – 층내 분산

③ 각 층에서의 조사단위당 실사비용

표본의 배분 문제는 층화임의추출(모집단을 층화하고 각 층에서 단순임의추출법으로 표본을 추출하는 방법)의 경우에 발생하는데, 층화임의추출의 효율성은 표본의 배분방법에 크게 의존한다.

▌표본을 배분하는 기준

크기 n인 표본을 L개의 층으로 배분하는 기준에는 크게 고정표본 기준(fixed sample size)과 고정변동계수 기준(fixed CV) 두 가지가 있다.

고정표본 기준

주어진 정도를 만족시키는 전체 표본의 크기 n을 구한 다음에 이를 각 층에 적당한 비율로 배분하는 것이다.

고정변동계수 기준

주어진 정도를 만족시키는 표본의 크기를 각 층별로 구한 다음 이를 합쳐서 전체 표본의 크기를 구하는 것이다. 고정변동계수 기준은 추정치에 대한 전체 정도를 만족시킨다는 장점이 있지만 계산이 복잡하고, 표본의 크기가 처음 계획했던 것보다 커질 수 있다는 단점이 있다.

표본을 배분하는 방법

① 비례배분 방법(proportional allocation)

각 층별로 모집단의 구성비만큼 표본을 배분하는 방법이다. 모집단에 대한 정보가 거의 없을 경우 활용되며 추정식이 간단하고 단순임의추출에 비해 분산이 작다.

② 모집단 총합 비례배분 방법(Y-proportional allocation)

관심변수들의 분포가 비대칭이 되기 쉬운 사업체조사 등에 많이 이용된다.

③ $\sqrt{N}$ 비례배분 방법

각 층별로 추정치에 대한 정도를 개별적으로 관리할 때 유용하게 쓰인다.

④ $\sqrt{Y}$ 비례배분 방법(멱배분)

전체 추정치에 대한 정도보다 각 층별로 추정치에 대한 정도를 따로 관리하려고 할 때 유용하게 쓰인다.

⑤ 최적배분 방법(optimal allocation)

단위당 조사비용이 서로 다르고, 층별 분산 간의 변동이 존재하는 경우 이를 감안해서 표본을 배분할 때 유용하게 쓰인다.

⑥ 네이만(Neyman) 배분 방법

단위당 조사비용이 모두 같다고 가정했을 때 쓸 수 있는 방법이다.

6 표본추출법의 결정

표본추출법은 확률추출법(probability sampling method)과 비확률추출법(nonprobability sampling method)으로 구분되는데, 대부분의 표본설계에서는 추출틀 내의 각 추출단위가 표본으로 추출될 확률을 미리 정한 뒤 확률적 원리에 의해 표본단위를 선택하는 확률추출법을 사용한다. 표본에 입각한 모든 추정이론은 확률추출법으로 표본이 추출될 때라야 적용 가능한 것이므로 오늘날 과학적 표본조사라고 하면 당연히 확률추출법을 근간으로 한다.

대표적인 확률추출법으로는 단순임의추출, 층화추출, 계통추출(systematic sampling), 집락추출 등이 있으며 보통 이러한 추출법들을 결합하여 사용한다.

비확률추출법으로는 판단추출(judgement sampling), 편의추출(convenience sampling), 할당추출(quota sampling), 눈덩이추출(snowball sampling) 등이 있다.

▌단순임의추출법

크기 N인 모집단에서 n개의 추출단위를 뽑을 수 있는 모든 가능한 경우 각각에

표본으로 추출될 가능성을 동일하게 부여하는 방법이다.

그러나 현실적으로 이런 추출틀을 구하기가 어렵고, 추출된 조사단위들이 널리 퍼져 있다면 비용이 많이 들기 때문에 실제 표본설계에서 단순임의추출법이 단독으로 사용되는 경우는 거의 없다. 더 저렴하면서도 동일한 정도를 보장하는 다른 표본추출법들을 이용하는 경우가 많다.

층화추출법

층화(stratification)란 모집단의 단위를 어떤 특징에 따라 몇 개의 부분 집단으로 나누는 것을 말한다. 자료(측정치)에는 반드시 산포가 있다. 산포의 원인이 되는 인자에 관하여 층화하면 산포의 발생 원인을 규명할 수 있게 되고, 산포를 줄일 수 있으며, 나아가 품질향상에 도움이 된다. 일반적으로 행해지는 층화방법의 예는 다음과 같다.

- 기계 : 공정라인별, 위치별, 상표별 등
- 작업자 : 조별, 숙련도별, 남녀별, 연령별 등

층화추출법이란 모집단을 관심변수와 관련성이 많은 보조변수의 값이 유사한 추출단위들끼리 묶어서 만든 층(strata)들로 나누고, 각 층에서 단순임의추출법으로 표본을 추출하는 방법이다. 층화추출법으로 표본을 추출하려면 우선 모집단을 서로 겹치거나 누락되는 부분이 없도록 분할한 뒤 몇 개의 그룹(층)으로 나누어야

한다. 일반적으로 동질적인 추출단위들의 묶음이 되도록 모집단을 나누지만 때에 따라서는 관심을 갖고 통계를 산출해내려는 집단별로 나누기도 한다.

계통추출법

단순임의추출법으로 표본을 추출하려면 표본추출틀에서 각각의 추출단위에 번호를 부여하여야 하는데, 이 작업은 간단하지가 않다. 계통추출법은 이런 단순임의추출의 어려움을 해결하기 위해 표본추출과정을 더 단순하고 편리하게 변형한 추출법이다.

계통추출법은 추출틀에 수록된 처음 k개의 추출단위에서 하나를 무작위로 뽑고, 그 다음부터는 매 k번째에 해당되는 추출단위를 뽑는 추출법이다. 방법이 간편할뿐만 아니라 시간적 · 공간적으로 일정한 간격을 두고 추출단위를 뽑기 때문에 모집단 전체에서 추출단위가 골고루 뽑힌다는 장점이 있다.

집락추출법

서로 인접한 조사단위들을 묶어서 집락 또는 조사구를 만든 다음, 집락들 중에서 일부를 추출하고 추출된 집락에 속한 조사단위들의 일부 또는 전부를 표본으로 추출하는 확률추출법이다. 집락추출법은 모집단의 모든 조사단위들을 망라한 좋은 추출틀이 없거나, 추출틀을 얻는 비용이 많이 드는 반면 집락에 대한 추출틀은 쉽게 얻을 수 있는 경우에 주로 사용된다. 또한 조사단위들이 멀리 떨어져 있어 실

사 비용이 늘어나는 경우에 최소의 비용으로 최대의 정보를 확보하기 위한 표본추출법이다.

확률비례추출법

모집단을 구성하고 있는 집락의 규모가 서로 다를 경우 각 집락을 추출할 확률을 불균등하게 뽑는 추출방법이다. 규모가 큰 집락의 추출확률은 크게 하고, 작은 집락의 추출확률은 작게 배당한다.

7 가중

표본조사의 주된 목적은 추출된 표본자료를 통해 모집단의 특성을 추측하는 것이다. 그러므로 표본자료는 모집단을 대표하는 대표성이 있어야 한다. 즉, 표본으로 추출되지 않은 다른 단위들도 설명할 수 있어야 하는 것이다. 표본자료가 모집단에 대한 대표성을 확보하기 위해서는 표본자료에 대한 가중(weighting)을 고려해야 한다.

가중치

현실적으로 모집단을 대표할 수 있는 표본 확보가 어려운 경우가 종종 있다. 이 경우 표본의 일부 혹은 전부에 가중치를 부여함으로써 모집단에 대한 대표성을 확보할 수가 있다. 즉, 크기 n인 표본이 모집단 N을 대표하기 위해 표본 한 개당 N/n 가중치를 부여하는 것이다.

한편 아무리 훌륭한 설계에 의해 추출된 표본일지라도 쓸모가 없는 표본들이 종종 있다. 예를 들어 표본으로 선택된 응답자가 응답을 무성의하게 한 경우이다. 이 경우 추가 표본추출에는 시간과 비용이 많이 들어가므로 성실히 응답한 표본에 가

중치를 부여하여 표본의 대표성을 확보한다.

크기 N인 모집단으로부터 크기 n의 확률표본을 단순임의추출할 경우 모집단 총합에 대한 추정량은 $N/n \sum y_i$로 i번째 단위에 대한 가중치가 N/n이 되며 이를 i번째 단위에 대한 기본가중치라 한다.

가중의 장점은 다음과 같다. 첫째, 분산과 비용을 감소시킬 수 있다. 둘째, 영역별로 서로 다른 추출률을 할당하여 작은 영역의 추출률을 높임으로써 전체적인 비용을 감소시킬 수 있다. 셋째, 중복, 비포괄성 등의 불완전추출틀 문제를 해결할 수 있다. 넷째, 조사에서 발생하는 무응답 문제를 적절히 해결할 수 있다.

자체가중

가중 절차에서 추출단위에 일정한 값을 갖는 가중치를 고려할 수 있는데, 모집단 조사단위가 표본으로 뽑힐 확률이 0이 아닌 값을 갖는 경우로, 이와 같은 가중치를 자체가중(self-weighting)이라고 한다. 자체가중은 단순임의추출, 층화추출, 확률비례추출, 다단계추출 등에서도 자체가중표본을 추출하는 것이 가능하다.

자체가중방법은 첫째, 조사의 복잡성을 감소시키며, 둘째, 잘못된 가중치의 문제를 해결할 수 있으며, 셋째, 동일한 표본이 다양한 목적과 다양한 조사에서 사용될 수 있는 유연성과 편리성을 가지며, 넷째, 다양한 표본설계에 강건성(robustness)을 부여하며, 다섯째, 일반적인 가중 절차에 비해 이해하기가 쉽다는 장점이 있다.

가중의 절차

① 기본가중치 계산

기본가중치는 표본추출설계로부터 직접적으로 얻어지는 값이다. 임의의 모집단에서 적절한 크기의 표본을 추출할 때 단위가 표본에 포함될 확률의 역수로 기본가중치를 계산할 수 있다.

② 무응답 조정가중치 계산

무응답 조정가중치(nonresponse adjustment weight)의 중요한 역할은 무응답 편향을 제거하는 것이다. 무응답 편향은 무응답이 매우 높은 비율로 모집단의 추정치에 영향을 줄 때 발생한다. 무응답 조정가중치를 계산하는 방법에는 표본에 기초한 무응답 조정 방법과 외부정보를 이용한 무응답 조정 방법이 있다.

표본에 기초한 무응답 조정 방법 무응답 단위들의 기본가중치를 표본응답자들에게 배분하여 조정된 가중치의 합이 전체 표본 단위들에 대한 기본가중치의 합이 된다.

외부정보를 이용한 무응답 조정 방법 표본에 기초한 무응답 조정 방법으로 무응답 조정가중치를 계산한 뒤 외부자료를 이용하여 이 가중치를 조정한다.

③ 사후층화 가중치 계산

사후층화 조정은 추출틀의 불완전으로 인한 포괄성의 차이, 표본의 불균형, 또는 비대표성, 무응답에 의한 차이 등을 조정하기 위한 것으로, 표본응답자들의 가중치를 조정함으로써 가중된 표본분포가 모집단분포와 같아지도록 하는 것이다.

래킹비(raking ratio)

래킹비 조정은 2차원 분류표상의 각 셀 값을 반복적으로 조정해가는 방법으로, 기본가중치를 하나의 주변분포를 이용하여 조정한 후, 두 번째 주변분포를 다시 이용하여 가중치를 조정한다. 이 과정을 특정한 수렴조건을 만족할 때까지 반복적으로 수행한다.

보정

기본가중치 또는 추출가중치와 보정된 새로운 가중치 사이의 차이를 나타내는 일종의 거리함수(distance function)가 주어진 조건을 만족하면서 최소가 되도록 하는 것이다.

가중의 효과

만일 표본조사를 할 때 가중을 하지 않으면 추정치의 왜곡을 피할 수 없게 된다. 그러나 가중을 고려하면 추정치의 편향은 줄어들지만 가중의 효과 때문에 분산이 늘어나게 된다. 이런 현상은 각 단위에 부과되는 가중치의 변동이 매우 클 경우에 일어난다.

가중치의 절단(trimming of weights)

극히 작은 표본에 의해 극단적으로 커진 가중치는 추정치의 분산을 증가시키는 요인이 된다. 그러므로 가중치의 변동을 고려하여 최대값 수준에서 극단가중치를 절단하는 것이 필요하다. 일반적으로 가중치의 절단은 무응답에 대한 조정 후에 수행한다. 또한 가중치의 절단은 절단을 함으로써 발생하는 분산의 감소분보다 총 MSE(Mean Square Error) 상의 영향이 적도록 하는 것이 바람직하다.

8 추정

일반적으로 추정단계에서 가중치를 이용하면 모집단에 대한 특성치인 모수에 대한 비편향추정량(unbiased estimator)을 얻을 수 있다. 표본의 크기가 큰 대규모 조사에서는 추정량의 편향이 문제가 되기 때문에 추정과정에서 반드시 가중치를 사용해야 한다. 모집단의 특성치에 대한 추정은 가중치를 이용한 가중표본평균을 사용한다. 이처럼 모수 추정을 위해 사용하는 추정량의 정도(precision)를 계산하기 위해서는 추정량의 분산을 구해야 한다.

추정량의 분산추정치를 계산하는 방법

① 정확한 방법(Exact Method)

분산을 추정하는 최선의 방법이지만 실제 적용하는 데 한계가 있다. 왜냐하면 대부분의 표본추출설계는 매우 복잡한 형태로 이루어져있고, 분산추정식 또한 완전한 수식의 형태로 나타나지 않을 수 있기 때문이다. 이 추정방법은 표본추출설계에 의존하므로 대부분 가중치를 고려한 방법인 관심변수의 추정식을 사용한다.

② 최종 집락 방법

복합표본설계에 의한 표본에 근거한 분산추정방법이다. 이 방법에 의해, 집락은 다단계 설계에 의해 부차적인 표본추출이 이루어졌음에도 불구하고 1개의 PSU(1차 추출 단위:Primary Sampling Unit)가 전체 표본으로 구성된다. 분산추정치는 각 단계별 추출에서 분산성분을 계산하지 않고 PSU 총합들 간의 변동만을 계산한다.

③ 선형화 방법

대부분의 복합표본설계로부터의 추정량은 비선형추정량이다. 선형화 방법에서는 비선형추정량을 테일러 전개를 이용하여 선형화한 후 선형화된 함수의 1차항만 고려하여 이를 정확한 방법의 분산추정식에 대입하여 최종 분산추정치를 구한다. 실제적인 분산추정에 많이 사용되는 방법이다.

④ 반복적인 방법

표본자료로부터 다시 표본을 반복적으로 추출하는 방법이다. 이 때 추출되는 표본은 부차표본이라 하며 모집단으로부터 추출된 가중추정치를 새롭게 계산한 후 전체 표본을 이용한 추정치와 반복추정치 간의 변동을 기초로 분산을 계산한다.

9 표본의 사후관리

일회성조사가 아니라면 표본조사는 처음 조사를 기획했을 때의 개념이나 품질, 수준이 계속 유지될 수 있도록 체계적으로 관리해야 한다. 그렇지 않을 경우, 처음에는 아무리 많은 수고를 했다 하더라도 나중에는 신뢰하기 어려운 조사가 될 가능성이 높다. 조사담당자나 조사기관의 변화, 모집단이나 표본의 변화 등이 조사의 수준이나 품질 변화에 결정적인 영향을 미칠 수 있기 때문이다.

사후 관리의 핵심은 처음 조사를 기획 · 설계할 때의 개념, 원칙, 방법을 늘 동일한 수준으로 유지하는 데 있다. 담당자가 바뀔 경우 새로운 담당자가 조사의 여러 국면을 이해하는 데 시간이 필요한데, 그 과정에서 조사의 개념이 달라질 수 있기 때문에 언제, 누가 그 일을 맡아도 동일한 원칙을 유지할 수 있도록 조사의 전과정을 상세하게 체계화한 조사시스템을 구축할 필요가 있다.

조사시스템의 구축

체계적인 조사관리를 위해서는 조사할 때마다 최선의 노력을 기울이는 것 이상으로 늘 일정한 수준의 조사품질을 담보할 수 있는 조사시스템을 구축하는 것이

필요하다. 조사시스템의 핵심적인 내용으로는 예산, 조직, 인력, 구체적인 조사관련업무를 규정하는 조사 매뉴얼, 장비와 소프트웨어 등이 있다.

추출틀 및 표본 관리

처음 표본설계를 할 때는 대체로 그 당시의 모집단을 비교적 잘 포함하는 추출틀이 마련된다. 그러나 시간이 지남에 따라 모집단에도 변화가 생긴다. 새로운 조사단위가 생성되거나 기존의 조사단위가 소멸하기도 하며, 일부 조사단위는 특성이 변하기도 한다. 모집단의 변화를 반영하기 위해 매번 새로운 표본설계를 하는 것은 현실적으로 거의 불가능하다. 대개는 모집단의 변화를 추출틀에 계속 반영시켜 적절히 관리하고 그에 맞게 표본을 보완해주는 조치를 취하게 된다.

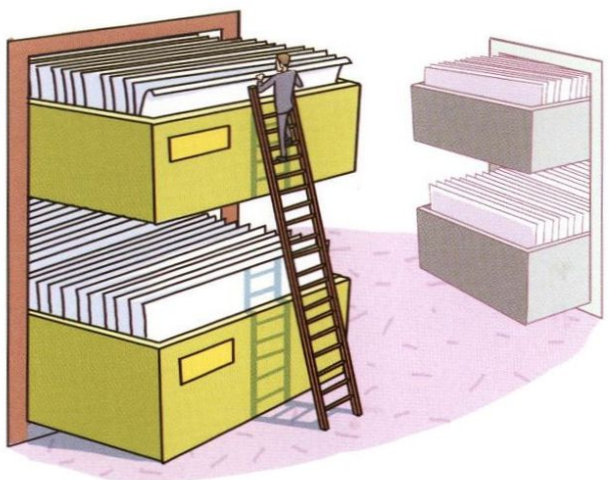
데이터베이스 관리 작업

계속조사에서 얻어지는 조사 데이터는 시계열 자료(time series data)가 되어 여러 가지 입체적인 분석을 위한 자료로 활용될 여지가 많다. 처음 조사된 자료를 통해서는 단순한 추정밖에 할 수 없다고 해도 시계열 자료가 모이면 보다 다양하고 방대한 분석이 가능해지기 때문이다. 따라서 조사된 데이터를 어떤 양식으로 저장, 관리할 것인가 하는 점도 중요한 문제가 된다. 앞으로 이 데이터를 이용해서 할 수 있는 다양한 종류의 분석들을 미리 고려하여 보다 효과적으로 활용될 수 있도록 데이터베이스를 관리하는 것이 필요하다.

무응답에 대한 대책

조사 과정에서 모든 조사관계자들이 아무리 노력한다고 해도 무응답은 생기게 마련이다. 이러한 무응답은 전체 조사의 일정 및 조사의 질에 영향을 미친다. 따라서, 조사원 품질을 일정 수준으로 유지하고 관리하기 위해 사전에 미리 무응답에 대한 대책을 세워두는 것이 필요하다.

무응답은 크게 단위무응답(unit nonresponse)과 항목무응답(item nonresponse)으로 구분된다. 단위무응답이란 응답자가 조사 자체에 불응하는 것이고, 항목무응답은 일부 조사 항목에 대해 응답하지 않은 경우이다. 무응답은 표본의 크기를 원래 목표보다 작아지게 함으로써 조사의 효율에 영향을 미치고, 무응답이 어떤 경향성을 띠게 되는 경우 추정값의 편향을 초래할 수 있다.



체크리스트

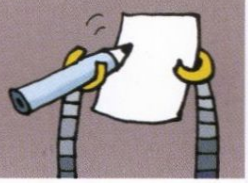
III

표본설계의 매 단계마다 반드시 확인해야 할 사항들을 정리한 체크리스트이다.

check list

체크리스트

표본설계, 이렇게 한다



모집단의 정의

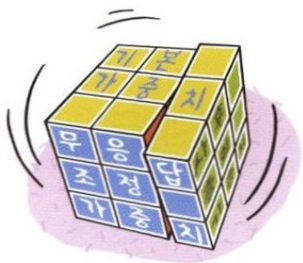
- □ 모집단에 대한 정확한 정의가 이루어졌는가?
- □ 모집단 정의 시 기준년도 등에 관한 설명은 이루어졌는가?
- □ 목표모집단과 조사모집단에 대한 조작적 정의가 이루어졌는가?

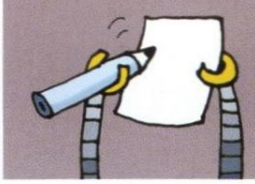
가구 단위 조사

- □ 가구 표본의 경우에는 어떠한 모집단을 대상으로 하고 있는가?
- □ 표본추출 대상 제외 가구에 대한 명시적 정의가 있는가?
- □ 가구 표본의 경우에는 표본 대상 가구원의 연령 등에 대한 제한점은 고려하고 있는가?

사업체 또는 기업체 단위 조사

- □ 사업체 또는 기업체의 분류 단위는 무엇인가?
- □ 매출액 또는 종업원 규모별 자료가 모집단 자료로 구성되어 있는가?
- □ 사업체 또는 기업체 표본의 경우에는 어떠한 모집단을 대상으로 하고 있는가?
- □ 사업체 또는 기업체 표본의 경우에는 모집단의 분포 형태를 고려하고 있는가?





check list

추출틀의 준비

- □ 추출틀은 현실적으로 구성이 가능한가?
- □ 추출틀의 포괄성은 어느 정도인가?
- □ 추출틀의 구성 형태는 어떠한가?
- □ 이용하고자 하는 추출틀은 리스트 프레임인가, 아니면 Area 프레임인가?

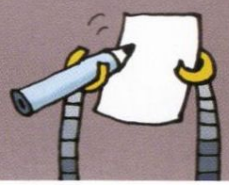
가구 단위 조사

- □ 가구 단위 조사인 경우 행정 프레임 또는 조사구 프레임을 사용할 것인가?
- □ 주 추출틀 이외에 사용 가능한 보조 프레임은 없는가?

사업체 또는 기업체 단위 조사

- □ 사업체 리스트 또는 기업체 리스트 프레임과 그 외 보조 프레임은 없는가?
- □ 조사구를 추출틀로 사용하고 있는가?

check list



효과적인 층화

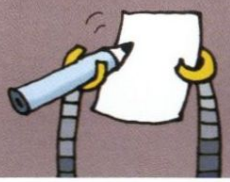
- □ 층별로 추정량의 산출이 필요한가?
- □ 층화추출설계가 가능한 모집단인가?
- □ 층의 개수는 표본의 크기와 비교해서 적당한가?
- □ 층별 표본배분방법은 어떠한 방법을 사용했는가?
- □ 고려한 층별배분방법은 적절한가?
- □ 주요 변수들의 변동계수를 고려하여 층화를 고려하였는가?

가구 단위 표본

- □ 가구 단위 표본에서는 층화를 어떤 기준에 의해 수행하였는가?
- □ 가구 단위 표본에 대한 층수는 적절한가?

사업체 또는 기업체 단위 표본

- □ 사업체 또는 기업체 단위 표본에서는 어떤 기준에 의해 층화하였는가?
- □ 사업체 규모별로 층화한 경우 전수층과 표본층을 구별하여 층화하였는가?



check list

표본의 크기 결정

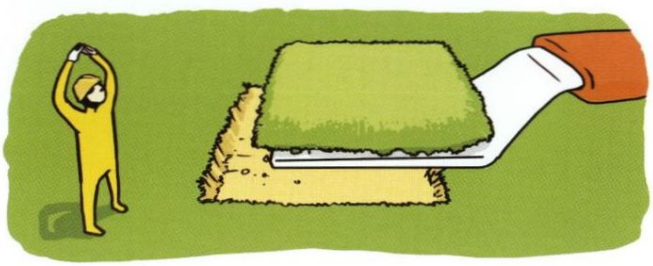
- □ 목표오차를 어느 정도로 고려하였는가?
- □ 주어진 예산과 현실을 적절히 반영한 표본규모인가?
- □ 표본크기의 계산식은 어떤 공식을 적용하였는가?
- □ 층화표본설계인 경우 층별로 배분된 표본의 크기는 적절한가?
- □ 결정된 표본크기와 더불어 예비 표본의 크기까지 함께 고려하였는가?
- □ 주요 변수의 목표오차를 조정할 수 있도록 표본크기를 고려하였는가?

가구 단위 표본

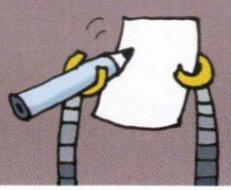
- □ 표본 지역 또는 표본 조사구는 몇 개로 할 것인가?
- □ 지역별 또는 조사구별로 몇 개의 가구를 표본으로 선정할 것인가?
- □ 최종 표본 가구 수는 몇 가구로 결정되었는가?

사업체 또는 기업체 단위 표본

- □ 산업별 또는 규모별 표본사업체 또는 기업체 수는 몇 개로 할 것인가?
- □ 최종 표본사업체 또는 기업체는 몇 개인가?



check list



표본추출법의 결정

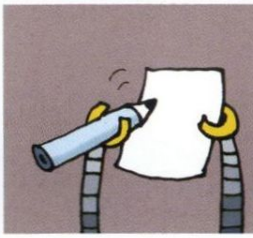
- □ 모집단을 적절히 대표할 수 있는 표본추출방법인가?
- □ 확률표본추출방법을 사용했는가?
- □ 집락추출인 경우 1차 추출단위(PSU : Primary Sampling Unit)에 대한 정의는 적절한가?
- □ 복합표본추출설계를 고려해야 하는가?
- □ 자체가중 표본추출설계는 가능한가?
- □ 불균등 확률추출방법을 사용했는가?

가구 단위 표본

- □ 가구 단위 표본추출의 경우 조사구를 1차 추출단위(PSU)로 고려했는가?
- □ 1차 추출단위(PSU)의 추출방법은 무엇인가?
- □ 최종 표본 가구의 추출방법은 무엇인가?

사업체 또는 기업체 단위 표본

- □ 최종 표본사업체 또는 기업체의 추출방법은 무엇인가?
- □ 표본추출 시에 사용한 보조정보가 있다면 무엇인가?



check list

가중

- □ 기본가중치는 계산하였는가?
- □ 복합표본추출방법을 적용했다면, 각 단계별 가중치는 고려했는가?
- □ 100% 완전응답표본이 아니라면, 무응답 가중치는 계산했는가?
- □ 사후층화 가중치는 계산되었는가?
- □ 사후층화 가중치 산정 시 고려된 방법은 무엇인가?
- □ 최종적으로 구한 가중치의 변동을 검토했는가?
- □ 가중치의 효과를 검토했는가?

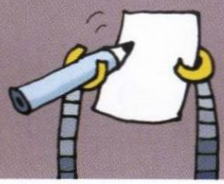
가구 단위 표본

- □ 지역별 1차 추출단위(PSU)의 추출확률은 고려하였는가?
- □ 표본 1차 추출단위(PSU) 내의 표본가구의 추출확률은 고려하였는가?
- □ 무응답 가구에 대한 가중치 조정 작업은 수행하였는가?
- □ 보조정보를 이용한 사후층화 가중치 조정은 수행하였는가?
- □ 최종 가구가중치의 변동을 고려하였는가?

사업체 또는 기업체 단위 표본

- □ 사업체 또는 기업체 추출확률은 고려하였는가?
- □ 층화 다단계추출인 경우 각 단계별 추출확률은 고려하였는가?
- □ 무응답 가중치 조정은 수행하였는가?
- □ 최종 표본사업체들의 가중치의 변동은 고려하였는가?

check list



추정

- □ 적용한 추정식은 표본설계를 적절히 반영하고 있는가?
- □ 복합표본추출방법을 적용했다면, 각 단계별 가중치를 고려한 추정산식인가?
- □ 사용한 추정량은 비편향성을 지니고 있는가?
- □ 추정치는 평균인가, 총합인가, 아니면 비율인가?
- □ 주요 관심 변수들의 추정치는 과거 추정치와 시계열성을 유지하고 있는가?
- □ 추정치에 대한 표준오차는 계산하였는가?
- □ 비선형추정량인 경우 다양한 근사적인 추정방법을 고려하였는가?
- □ 분산추정산식은 적절한가?
- □ 추정치의 계산 및 분산추정을 위해 사용한 프로그램은 무엇인가?
- □ 반복적인 분산추정방법을 사용하였는가?

가구 단위 표본

- □ 추정치는 가구당 평균 또는 총합인가?
- □ 가구의 추정치에 대한 표준오차는 계산되었는가?
- □ 표준오차의 계산식은 제시되었는가?

사업체 또는 기업체 단위 표본

- □ 산업별 추정치인가, 아니면 업체별 추정치인가?
- □ 각 추정치에 대한 표준오차는 제시하였는가?
- □ 표준오차의 계산식은 제시되었는가?

표본설계 사례

Ⅳ

구체적인 표본설계의 사례로, 사업체 표본조사의 경우 '도 · 소매업 통계조사'(통계청)와 가구 표본조사의 경우 '최저생계비 계측조사'(보건복지부)를 살펴보기로 한다.

1 사업체 표본조사의 표본설계 사례
– 도 · 소매업 통계조사(통계청)

1. 조사목적

전국의 도 · 소매업 및 숙박 · 음식점업의 경영실태 및 구조변화를 파악하여 각종 정책수립과 연구 · 분석을 위한 기초자료로 제공하기 위함이다.

2. 조사대상

한국표준산업분류상(제6차 개정 1991년 9월, 제8차 개정 2000년 1월)의 대분류 G, H 업종의 사업체 중 표본으로 선정된 약 29,000개 사업체를 조사대상으로 한다.

3. 조사주기

- 매년

4. 조사기준시점 및 조사대상기간

- 조사기준시점 : 조사대상년도 12. 31.
- 조사대상기간 : 조사대상년도 1. 1. ~ 12. 31.

5. 표본설계(2003년 기준 조사)

가. 추출틀

2002년 기준 사업체기초통계조사 결과 중 산업중분류 50(자동차판매 및 차량연료소매업), 51(도매업 및 상품중개업), 52(소매업 ; 자동차 제외), 55(숙박 및 음식점업)에 해당하는 모든 사업체 중에서 52820(노점 및 유사 이동판매업), 52899(기타 무점포소매업) 및 55223(이동음식점업)에 해당되는 사업체는 제외하였다.

나. 표본의 구성

① 전수조사 업종

- 백화점(52111), 기타종합소매업(52119), 호텔업(55101)
- 산업세분류 및 시 · 도별 모집단 사업체 수가 11개 미만인 업종

② 표본조사 업종

- 전수층 : 종사자 수가 일정 규모(절사점) 이상인 모든 사업체
- 표본층 : 사업체 수가 많아 전수조사가 불가능하여 사업체 단위로 표본사업체를 추출하여 일부만 조사하는 업종

다. 층화

산업세분류, 16개 시 · 도, 종사자 수 순으로 층화한 후 표본사업체를 선정하였다.

① 산업세분류

- 50 : 자동차판매, 수리 및 차량연료소매업(5개 산업세분류)

• 51 : 도매업(24개 산업세분류)

• 52 : 소매업 ; 자동차 제외(25개 산업세분류)

• 55 : 숙박 및 음식점업(6개 산업세분류)

② 시 · 도별 : 16개 시 · 도

③ 재층화

• 표본조사 업종의 경우 절사점을 기준으로 전수층과 표본층으로 재층화

– 전수층 : 종사자 수가 일정 규모(절사점) 이상인 사업체는 모두 표본으로 선정

– 표본층 : 종사자 수가 일정 규모(절사점) 미만인 사업체는 계통추출방법으로 표본 선정

• 표본층에서 매출액이 100억 이상이거나 업종 평균 매출액의 100배 이상인 사업체는 전수층으로 모두 선정

*절사점 : 시 · 도별로 종사자 수가 큰 사업체들의 누적비와 변동계수 및 상대허용오차와 신뢰계수에 의해 결정하며 전수층과 표본층을 나누는 경계점

라. 표본규모(크기) 결정

신뢰도 68%와 상대허용오차 14%로 표본규모를 결정, 서울의 경우는 신뢰도 68%와 상대허용오차 15%로 표본규모를 결정

마. 표본규모(크기) 계산

① 특성치 : 매출액(x)

• 총 표본규모 :

$$n_{hi} = {}_{c}n_{hi} + {}_{s}n_{hi}$$

• 표본층 표본규모 : $${}_sn_{hi} = \frac{\frac{z^2 \cdot (Q_{hi} \cdot CV_{hi})^2}{e^2}}{1 + \frac{z^2 \cdot (Q_{hi} \cdot CV_{hi})^2}{{}_sN_{hi} \cdot e^2}}$$

- 첨자 h : 산업세분류 i : 시 · 도 c : 전수층 s : 표본층
- 변수 n : 표본 수 N : 모집단 수 Q : 종사자 총합 중 표본층이 차지하는 비율
  CV : 표본층 변동계수 S : 표본층 표준편차
  e : 허용상대오차 z : 신뢰계수

바. 표본추출

3개 전수조사 업종에 대해서는 해당되는 모든 사업체를 조사하고, 표본조사 업종에 대해서는 사업체 단위로 절사법 표본설계방법을 응용하여 표본추출을 실시하였다.

- 전수층 : 절사점 이상인 사업체는 모두 선정
- 표본층 : 절사점 미만인 사업체는 주어진 표본규모에 따라 계통추출방법으로 표본을 선정

사. 표본오차

① 업종별 · 시도별 총량 추정시

• 분산 $$V(\bar{X}_{hi}) = {}_sN_{hi}^2 \left(\frac{{}_sN_{hi} - {}_sn_{hi}}{{}_sN_{hi}} \right) \cdot \frac{{}_ss_{hi}^2}{{}_sn_{hi}}$$

여기서, $${}_s s_{hi}^2 = \frac{1}{{}_sn_{hi} - 1} \left(\sum_j^{{}_sn_{hi}} {}_sx_{hij}^2 - \frac{(\sum_j^{{}_sn_{hi}} {}_sx_{hij})^2}{{}_sn_{hi}} \right)$$

j : 개별사업체를 나타냄

• 표준오차 : $SE(\hat{X}_{hi}) = \sqrt{V(\hat{X}_{hi})}$

• 상대표준오차 : $RSE(\hat{X}_{hi}) = \frac{SE(\hat{X}_{hi})}{\hat{X}_{hi}} \cdot 100$

② 업종별 전국 총량 추정시

• 분산 : $V(\hat{X}_h) = \sum_i V(\hat{X}_{hi})$

• 표준오차 : $SE(\hat{X}_h) = \sqrt{V(\hat{X}_h)}$

• 상대표준오차 : $RSE(\hat{X}_h) = \frac{SE(\hat{X}_h)}{\hat{X}_h} \cdot 100$

아. 모수 추정

표본조사 결과 계산된 업종별, 시 · 도별 합계에 2002년 기준 사업체기초통계조사 결과의 해당 사업체 수 또는 종사자 수를 기준으로 승수를 주어 총량 추정

① 업종별 · 시도별 총량 :

$$\hat{X}_{hi} = \sum_j^{n_{hi}} x_{hij} + w_{hi} \sum_j^{n_{hi}} x_{hij}, \quad w_{hi} = {}_sN_{hi} / {}_sn_{hi}$$

② 업종별 전국 총량 : $\hat{X}_h = \sum_{i=1}^{16} \hat{X}_{hi}$ (첨자 $\hat{X}$ 는 추정치)

2 가구 표본조사의 표본설계 사례
– 최저생계비 계측조사(보건복지부)

1. 조사목적

저소득층 가구의 가계수지 및 생활실태를 정확히 파악함으로써 건강하고 문화적인 삶을 영위할 수 있는 최저생계비를 추정하기 위한 기초 자료를 생산하기 위함이다.

2. 조사범위 및 대상

① 조사지역 : 전국

② 조사대상 : 가구조사를 기본으로 하되 복지서비스대상에서는 가구원조사도 병행한다.

3. 조사내용

① 최저생계비를 가구원(1인~7인) 규모별로 구분하여 계측하고, 이를 위한 표준가구설정 및 합리적인 가구균등화지수를 산출한다.

② 최저생계비를 가구의 주거점유 형태(자가, 전세, 월세 등)와 가구원의 인구경제학적 특성(성, 연령, 장애 여부, 질병 여부, 학생 등)별로 나누어 계측한다.

③ 최저생계비 계측 결과를 토대로 차상위 계층의 규모를 파악하고 이의 특성과 생활실태 및 복지요구를 파악한다.

4. 표본설계

가. 모집단 분석

현행 표본조사를 위한 주요 변수에 대한 이용 가능한 모집단 정보가 없기 때문에 이와 유사한 조사로부터 주요 변수에 대한 자료를 이용하는 것이 바람직하다. 그러나 2005년도 인구주택총조사 자료분석 결과를 조사설계 당시 이용할 수 없고, 2004년도 계측조사에서는 3개의 조사구를 결합하여 조사가 이루어진 관계로 현행 조사설계와 유사한 모집단 자료를 이용하기 어려운 점이 있었다. 따라서 가구소득과 지출에 어느 정도 관계가 있는 것으로 파악된 주택유형 및 조사구당 가구 수 등을 기준으로 모집단 분석을 수행하였다.

나. 조사구명부 작성

표본설계 당시 2005년도 인구주택총조사 90% 조사구에 대해 이용 가능한 자료는 조사구별 가구 수, 조사구 형태, 주택형태뿐이었기 때문에, 이를 바탕으로 517개 조사구의 기초자료를 집계하여 지역별, 조사구 유형별, 읍면동별, 주택형태별로 분류하여 분포를 파악하였다.

다. 추출틀

본 조사의 추출틀은 2005년 인구주택총조사 25만여 개의 조사구 중 90%조사구인 22만 3천여 개의 조사구를 추출틀로 하며, 이때 사용한 자료는 조사구 특성(일반조사구, 아파트조사구), 주택유형(단독, 아파트, 연립 및 다세대), 조사구당 가구수 등으로 구성된 리스트를 사용하였다.

라. 표본추출방법

2005년도 인구주택총조사구 25만여 개의 조사구 중 90% 조사구인 22만 3,000여 개 조사구 중 예비조사구를 포함하여 517개 조사구를 지역별 조사구 규모에 따라 층화추출하였으며, 층화의 주요 기준변수로는 지역(16), 조사구 형태(2), 주택유형(3) 등을 사용하여 총 96개 층으로 나누어 각 층별로 확률비례추출하였다.

각 조사구별로 평균 60가구 중에서 조사원의 업무 할당을 고려하여 일률적으로 51가구를 조사하도록 하였으며, 조사불능가구 또는 조사대상제외가구가 발생할 경우 가구명부의 순서에서 바로 다음 가구를 조사하도록 하였다. 조사구의 크기가 51가구 이하가 되는 조사구는 조사구 내의 모든 가구를 조사토록 하였으며, 조사불능이나 제외가구의 발생으로 51가구 이하가 되는 조사구에서는 조사가능가구만을 표본으로 산정하도록 하였다. 일부 조사구에 대해서는 예비조사구의 대체 가구로 대체조사하였다.

지역별 표본 가구 수

(단위 : 개)

	계			동부			읍면부		
	합계	단독	아파트	합계	단독	아파트	합계	단독	아파트
전 국	29,448	18,624	10,824	24,008	13,616	9,839	5,991	5,008	985
서 울	6,568	4,601	1,967	6,568	4,601	1,967	0	0	0
부 산	2,457	1,337	1,120	2,457	1,337	1,120	0	0	0
대 구	1,587	978	609	1,587	978	609	0	0	0
인 천	1,614	931	683	1,614	931	683	0	0	0
광 주	842	405	437	843	405	437	0	0	0
대 전	851	464	387	851	464	387	0	0	0
울 산	625	343	282	626	343	282	0	0	0
경 기	4,925	2,634	2,291	4,288	1,790	1,969	1,166	844	322
강 원	1,052	717	335	576	298	263	490	419	72
충 북	980	643	337	529	280	268	431	363	69
충 남	1,235	885	350	387	184	195	856	701	155
전 북	1,312	884	428	783	382	402	528	502	26
전 남	1,320	915	405	760	413	379	528	502	26
경 북	1,853	1,348	505	814	452	350	1,051	896	155
경 남	1,944	1,298	646	1,113	625	486	833	673	160
제 주	283	241	42	212	133	42	108	108	0

마. 표본규모

표본의 규모는 예비조사구를 포함하여 총 517개 조사구의 약 3만 가구를 표본으로 결정하였다. 이때, 2004년 계측조사와는 다르게 2007년 계측조사에서는 표본으로 선정된 조사구는 인근 조사구를 결합하는 방식이 아닌 1개 조사구씩을 표본으로 추출하였다.

과거의 조사자료를 근거로 주요 변수에 대한 상대오차를 이용하여 새로운 표본규모를 결정할 때는 다음과 같은 방법으로 간단히 구할 수 있다.

$$n' = n\left(\frac{CV_1}{CV_2}\right)^2 \quad \cdots\cdots\cdots\cdots (1)$$

여기서 n'은 새로운 표본규모이며, n은 기존의 표본규모, CV_1은 기존의 조사로부터 구한 상대표준오차, CV_2는 새로운 표본에 대한 목표 상대표준오차이다.

시도별 목표 정도

(단위 : 개, %)

시 도	2002년 가계조사			현행 표본설계			
	조사구수	가구수	목표정도	조사구수	목표정도	가구수	목표정도
전 국	697	5,089	1.62	500	1.91	29,448	0.67
서 울	127	995	2.87	110	3.08	6,568	1.12
부 산	70	518	3.45	41	4.51	2,457	1.58
대 구	49	357	5.62	27	7.57	1,587	2.67
인 천	50	377	3.09	27	4.20	1,614	1.49
광 주	52	340	5.07	14	9.77	842	3.22
대 전	50	357	6.17	14	11.66	851	4.00
울 산	23	190	4.11	10	6.23	625	2.27
경 기	50	384	5.73	91	4.25	4,925	1.60
강 원	27	191	7.85	18	9.61	1,052	3.34
충 북	27	193	5.81	16	7.55	980	2.58
충 남	30	184	6.96	21	8.32	1,235	2.69
전 북	29	191	6.31	21	7.42	1,312	2.41
전 남	29	209	5.75	22	6.60	1,320	2.29
경 북	32	230	4.67	30	4.82	1,853	1.65
경 남	34	253	5.57	33	5.65	1,944	2.01
제 주	18	120	5.48	5	10.40	283	3.57

바. 추정방법

① 전국 단위의 추정

소득, 지출 등의 주요 변수에 대한 전국 단위의 추정치를 계산하기 위한 공식은 다음과 같이 표본추출 과정을 고려하여 계산된다.

$$\bar{y}=\frac{\sum_{h}^{L}\sum_{i}^{n_h}\sum_{j}^{m_{hi}} w_{hij}y_{hij}}{\sum_{h}^{L}\sum_{i}^{n_h}\sum_{j}^{m_{hi}} w_{hij}} \quad \cdots\cdots\cdots\cdots (2)$$

여기서 h는 전국 6개 층으로서 지역별 조사구 유형을 나타낸다. $h=1, 2, \cdots 6$ 또는 전국 16개 광역시 및 시도와 동부 및 읍면부를 나타낸다. i는 표본조사구를 나타내는 첨자로서 $i=1,2,\cdots n_h$ 이다. 그리고 j는 표본조사구 내의 가구를 나타내는 첨자로서 $j=1,2,\cdots m_h$이다. W_{hij}는 i지역의 i번째 표본조사구 내의 j번째 가구에 부여된 가중치이다.

또한 $W_{hi}=\sum_{j}^{m_{hi}} w_{hij}$ 으로 h층의 i번째 조사구 내의 가구들의 가중치 합을 나타낸다.

표본평균 $\bar{y}$의 분산추정치는 다음과 같이 정의할 수 있다.

$$\hat{V}(\bar{y})=\frac{\sum_{h}^{L}\frac{n_h}{n_h-1}(1-f_h)\sum_{i=1}^{n_h}\left[W_{hi}(\bar{y}_{hi}-\bar{y})-\frac{1}{n_h}\sum_{s=1}^{n_h}W_{hs}(\bar{y}_{hs}-\bar{y})\right]^2}{(\sum_{h}^{L}\sum_{i}^{n_h}W_{hi})^2} \quad \cdots\cdots (3)$$

여기서 $\bar{y}_{hi}=\frac{\sum_{j}^{m_{hi}} w_{hij}y_{hij}}{\sum_{j}^{m_{hi}} w_{hij}}$ 는 h층의 i번째 조사구 내의 가구들의 소득, 또는 지출 등의 주요 변수들의 평균이다.

② 지역별(또는 층별) 추정– 지역별, 가구규모별, 가구유형별 소득

만일 층별 또는 부차모집단에 대한 추정치를 얻고자 한다면, 다음과 같은 추정공식을 사용할 수 있다. 즉, 관심 대상이 되는 h번째 부차집단의 평균의 추정치는

$$\hat{V}(\bar{y}_h)=\frac{\frac{n_h}{n_h-1}(1-f_h)\sum_{i=1}^{n_h}\left[W_{hi}(\bar{y}_{hi}-\bar{y}_h)-\frac{1}{n_h}\sum_{s=1}^{n_h}W_{hs}(\bar{y}_{hs}-\bar{y}_h)\right]^2}{(\sum_{i}^{n_h}W_{hi})^2} \quad \cdots\cdots\cdots (4)$$

으로서 만일 h층의 i번째 표본조사구의 j번째 가구가 어떤 특성을 가지면, y_{hij} = 1, 그 외에는 0이라고 할 때, 모비율의 추정치로 사용할 수도 있다.

층별 표본평균에 대한 분산추정량은 다음과 같이 정의된다.

$$\bar{y}_h = \frac{\sum_{i}^{n_h} \sum_{j}^{m_h} w_{hij} y_{hij}}{\sum_{i}^{n_h} \sum_{j}^{m_h} w_{hij}} \quad \cdots\cdots\cdots\cdots (5)$$

③ 가구유형별/점유형태별 추정량

특정층에 속한 비율을 추정하고자 할 때에는 식(3)을 변형한 다음과 같은 추정산식을 이용하면 된다. 이에 해당되는 추정대상은 주로 노인 가구, 장애인 가구 등의 비율 등이 될 수 있다.

$$\bar{y}_G = \frac{\sum_{h}^{L} \sum_{i}^{n_h} \sum_{j}^{m_{hi}} w_{hij} y_{hij} I[hij \in G]}{\sum_{h}^{L} \sum_{i}^{n_h} \sum_{j}^{m_{hi}} w_{hij} I[hij \in G]} \quad \cdots\cdots\cdots\cdots\cdots (6)$$

여기서 G는 1인 가구, 2인 가구, 노인 가구, 장애인 가구 등이 될 수 있으며, 또한 $I[hij \in G]$는 h층의 i번째 조사구 내의 j번째 가구가 어떤 특성을 가지면 1, 그렇지 않으면 0을 갖는 지시함수이다.

식(6)에 대한 분산추정량은 다음과 같다.

$$\hat{V}(\bar{y}_G) = \frac{\sum_{h}^{L} \frac{n_h}{n_h - 1} (1 - f_h) \sum_{i=1}^{n_h} \left[W_{hiG}(\bar{y}_{hiG} - \bar{y}_G) - \frac{1}{n_h} \sum_{s=1}^{n_h} W_{hsG}(\bar{y}_{hsG} - \bar{y}_G) \right]^2}{(\sum_{h}^{L} \sum_{i}^{n_h} W_{hiG})^2} \quad \cdots\cdots (7)$$

여기서 가중치의 합은 $W_{hiG} = \sum_{j}^{m_{hi}} w_{hij} I[hij \in G]$ 이고, 그룹별 평균은 다음과 같다.

$$\bar{y}_{hiG} = \frac{\sum_{j}^{m_{hi}} w_{hij} y_{hij} I[hij \in G]}{\sum_{j}^{m_{hi}} w_{hij} I[hij \in G]} \quad \cdots\cdots\cdots\cdots\cdots (8)$$

만일 조사무응답이 존재하는 경우에는 무응답을 조정한 가중치를 $W^*_{hij} = W_{hij} \times r_{hij}$라 하면 이 값을 위의 추정량의 W_{hij} 대신 대입하여 무응답 조정이 된 추정량을 구할 수 있다. 여기서 r_{hij}는 h층의 i번째 조사구의 j번째 가구의 응답률이다. 이때, 개별 가구의 응답률보다는 각 층 내의 조사구별 응답률을 이용하여 무응답에 대한 가중치 조정이 가능하다.

$$r_{hi} = \frac{h\text{층의 } i\text{번째 조사구에서 응답한 총 가구 수}}{h\text{층의 } i\text{번째 조사구 내의 총 가구 수}} \quad \cdots\cdots (9)$$

표본 품질관리 매뉴얼

V

매뉴얼을 통해 표본 품질관리의 개념과 순서를 정리해 본다.

Manual

표본 품질관리 매뉴얼

모집단의 정의

① 모집단에 포함되는 조사단위를 명확히 정의한다.

가령 20세 이상의 성인을 모집단으로 정의한 경우를 생각해보자. 우리나라 사람들의 연령 개념은 다소 모호하여 만 나이도 있고 일반 나이도 있다. 따라서 더 명확한 정의를 내리려면 "호적에 등재된 생년월일이 1983년 6월 31일 이전인 성인"과 같은 식으로 해야 한다. 또, 모집단의 지리적 · 공간적 · 시간적 범위를 명확히 밝혀야 한다. 지리적 · 공간적 범위를 밝힌다는 것은 모집단에서 제외되는 지역이 있으면 명확히 밝혀야 한다는 뜻이다. (예를 들어 시간과 비용이 훨씬 많이 들기 때문에 도서 지역을 제외하는 경우)

② 목표모집단과 조사모집단의 차이를 검토한다.

목표모집단과 조사모집단이 다른 경우 그 차이의 정도가 어느 정도인지를 비교 검토해야 하는데 이를 위해서는 필요한 자료를 확보해야 한다. 두 집단 간의 차이가 무시할 수 없을 정도라고 판단될 경우에는 조사모집단 확대를 검토할 수 있다.

참고 목표모집단과 조사모집단의 차이가 발생하는 요인

① 추적 불가능한 단위	② 결측단위
③ 일시부재	④ 부재
⑤ 응답불능	⑥ 응답거부
⑦ 기타 무응답	⑧ 중복 또는 외부단위 등

추출틀의 준비

① 조사목적에 적합한 추출틀을 정한다.

조사목적에 맞으면서 이용 가능한 다양한 추출틀을 확보하여 비교한 후 가장 적합한 추출틀을 마련해야 한다.

② 가능한 한 조사 관련 보조정보를 잘 갖춘 추출틀을 마련한다.

추출틀에 각 조사단위의 특성을 나타내는 보조정보가 많으면 표본설계의 효율을 높일 수 있을 뿐 아니라 추정 과정이나 무응답을 위한 조치를 할 때 보조정보를 적절히 활용할 수 있어서 매우 효과적이다. 따라서 가능한 한 조사단위에 관한 정보를 많이 얻을 수 있는 추출틀을 구하는 것이 바람직하다.

③ 필요에 따라서는 여러 개의 추출틀을 활용할 수도 있다.

다양한 추출틀이 존재하지만 불완전하거나 비용이 많이 드는 경우에는 현실적으로 활용 가능성이 높은 여러 개의 추출틀을 동시에 사용할 수 있다. 이 경우 가능하면 상호보완적인 성격을 지니는 복수의 추출틀을 활용하는 것이 바람직한데, 가능한 한 전문가의 도움을 받아 이론상으로 문제의 여지가 없는지를 검토한다.

④ 추출틀의 포함범위를 주기적으로 평가하고 보정한다.

추출틀과 조사모집단 사이의 괴리는 추정값의 편향을 초래할 수 있다. 따라서 추출틀이 조사모집단을 어느 정도 포함하는지를 주기적으로 평가하여 적절한 조치를 취하여야 한다.

⑤ 추출틀의 품질을 유지, 향상시키기 위한 시스템을 갖춘다.

한 번의 조사로 끝나는 것이 아니라 장기적으로 이루어지는 조사일 경우 추출틀의 품질은 항상 유지, 관리되어야 한다.

⑥ 동일 모집단에 대한 조사는 가능한 한 동일한 추출틀을 사용하여 일관성을 유지한다.

계속조사이거나 유사한 다른 조사와 동일한 모집단을 대상으로 조사하는 경우 가능한 한 동일한 추출틀을 이용하여 조사들 간의 일관성을 유지하는 것이 바람직하다.

효과적인 층화

① 설계변수와 밀접한 연관성을 갖는 변수를 선정한다.

층화변수는 일반적으로 설계변수와 밀접한 연관성을 갖는 변수를 선정해야 한다. 하나의 특정 변수만을 고려하여 층화를 하다보면 고려되지 않은 다른 변수들에 부정적인 영향을

끼칠 수 있다. 그러므로 다목적조사에서는 가장 중요하다고 생각되는 복수의 층화변수를 선택하는 것을 고려해야 한다. 이 때 좋은 층화변수를 찾는 것이 층화의 핵심적인 사항이다.

② 복잡한 조사, 대규모 조사일 경우 층화 다단추출법에서 1차 추출단위에 대한 층화를 최우선으로 고려한다.

일반적으로 복잡한 조사, 대규모 조사일 경우 층화 다단추출법을 사용한다. 이때에는 최초의 추출단위인 1차 추출단위(PSU : Primary Sampling Unit)에 대한 층화를 잘 하는 것이 가장 중요하다.

③ 관심영역에 대한 부분 통계의 생산을 원할 때에는 반드시 이를 반영할 수 있는 층화변수를 선정한다.

④ 층을 나눈 후 각 층 내의 모집단 단위들을 관심변수와 가장 관련이 깊은 보조변수의 크기 순으로 정렬한 다음 계통추출방법으로 표본조사단위들을 추출한다.

⑤ 층화를 할 때 가능한 모든 경우를 다 나눈 후 역으로 합쳐가면서 적절한 층의 개수를 정한다.

⑥ 층을 나눈 후 모든 층에 대해 획일적인 표본추출방법을 적용할 필요는 없으며, 층에 따라서 표본추출의 방법을 달리 하는 것도 고려한다.

⑦ 추출틀에 층화에 필요한 보조정보가 충분히 들어 있지 않은 때에는 이중추출(double sampling) 기법을 고려한다.

이중추출이란 일차로 대규모의 표본을 뽑아 층화변수로 사용할 수 있으면서도 응답이 간편한 층화변수 관측값을 구한 후 이를 근거로 1차 표본단위를 층화한 후 각 층에서 일부의 2차 표본을 추출하는 방법이다.

표본크기의 결정

추정량의 정도(precision) 결정시 유의할 점들

① 조사로부터 얻어진 추정량에 대하여 허용 가능한 오차는 어느 정도인가?

예를 들어 추정오차의 한계가 95% 신뢰수준에서 ±5%인가, 아니면 그 이상인가 이하인가를 결정한다.

② 조사모집단 전체를 대상으로 한 추정량과 부 모집단 영역(subpopulation) 추정량의 정도는 구별하여 확정하였는가?

조사모집단 전체를 대상으로 한 추정량에 요구되는 정도와 조사모집단 내 각 부 모집단에 대한 추정량에 요구되는 정도는 반드시 구별되어야 한다. 예를 들어 전국을 대상으로 한 조사에서 어떤 추정량에 대한 정도가 전국적으로는 3%일 수 있지만, 도별로는 5%, 군별로는 10%가 될 수 있다.

③ 추정치에 대한 표본분산의 상대적 크기는 어느 정도이어야 하는가?

정도는 추정치의 크기를 고려해서 결정해야 한다. 보통 추정치에 대한 표본분산의 상대적 크기는 추정치의 10%~20% 정도가 적당하다. 즉, 추정치의 크기가 10일 경우 이에 대한 추정오차의 한계(정도)의 크기는 1 또는 2가 적당하다.

④ 표본의 크기를 늘림으로써 정도가 얼마나 개선되는가?

정도(precision)는 표본의 크기를 늘림으로써 개선되지만, 그 개선의 폭은 표본의 크기에 선형비례하지는 않는다.

⑤ 가장 작은 오차의 한계를 갖는 가능한 한 가장 큰 표본을 선택하는 것이 최선의 방법은 아니다. 부 모집단의 추정량의 정도를 함께 고려해야 하는 경우도 발생하기 때문이다.

표본의 크기에 영향을 미치는 요소들

① 모집단 변동계수(CV : population Coefficient of Variation)

조사모집단에서 각 특성들의 변동(variation)은 서로 다르며, 변동의 크기는 표본의 크기에 영향을 준다. 일반적으로 표본조사에서는 서로 다른 변동을 갖는 여러 개의 특성들을 측정한다. 어느 한 특성에 대해 주어진 정도를 만족시키는 충분한 크기의 표본을 구했다 하더라도, 그보다 더 큰 변동을 갖는 특성에 대해서는 주어진 정도를 만족시키지 못할 수 있다. 따라서 주요한 특성들의 요구 정도를 만족시키는 표본의 크기를 구하기 위해서는 가장 큰 변동을 갖는 특성에 맞도록 표본의 수를 결정해야 한다.

② 모집단의 크기(population size)

조사모집단의 크기가 작을수록 표본의 크기에 더 큰 영향을 미친다. 모집단의 크기가 작을수록 집단 내에서 표본이 차지하는 비율이 커지므로, 작은 크기의 모집단의 경우는 표본조사가 아닌 전수조사를 하게 된다.

③ 표본설계에 따른 설계효과(design effect)

표본설계와 그에 따른 추정량은 정도(precision)에 많은 영향을 미친다. 일반적으로 주어진 정도를 만족시키는 표본의 크기는 층별 총합이나 집락별 총합과 같이 이용 가능한 정보가 없을 경우 단순임의추출하에서의 표본분산을 근거로 구해진다. 따라서 복합설계하에서 같은 정도를 갖는 표본의 크기를 구하기 위해서는 단순임의추출하에서 구한 표본의 크기에 설계효과를 곱해 표본설계에 맞는 표본크기를 결정해야 한다.

④ 조사응답률

추정치에 대한 원하는 정도를 얻기 위해서는 표본의 크기를 예상응답률로 조정해주는 작업이 필요하다.

표본의 배분

① 모집단 총합 비례배분과 $\sqrt{Y}$ 비례배분 방법을 이용할 경우 보조정보를 이용한다.

② 각 층에 배분되는 조사단위의 수는 적어도 2개 이상이 되도록 한다.

조사단위가 전혀 배분되지 않는 층이 생길 경우 해결책 중의 하나는 각 층에 각각 크기 2인 표본을 강제 배분하고 나머지를 배분방법에 의해 배분하는 것이다.

③ 여러 개의 변수에 근거해서 표본을 배분하는 방법을 고려한다.

표본추출법의 결정

① 적용 가능한 다양한 표본추출법들을 고려하여 가장 효율적인 추출법을 선택한다.

② 가능한 한 단순한 추출법을 사용한다.

다단추출법이나 집락추출법 등 다소 복잡한 추출법은 그 추출법을 사용함으로써 명백한 이점이 있는 경우에 한해 적용하도록 한다. 추출법이 복잡해질수록 추정식이나 관리가 까다로워진다는 사실을 명심하여야 한다.

③ 추출틀과 확보 가능한 보조정보를 고려하여 거기에 맞는 적합한 추출법을 결정한다.

추출틀 마련이 어려운 경우 1차로 추출틀 마련이 용이한 규모의 집락을 형성하는 다단추출법이 바람직하다. 이 때 집락을 어떻게 구성하는 것이 최적인지에 대한 연구가 필요하다.

④ 표본추출방법을 결정할 때 조사의 용이성, 조사비용 등을 함께 고려한다.

예를 들어 표본으로 뽑힌 조사단위들이 지리적으로 너무 떨어지지 않게 하려면 지리적으로

인접하는 마을이나 지역을 하나의 집락으로 하는 추출법을 생각할 수 있다. 또한 조사단위들 간의 조사비용이 서로 다를 경우 조사비용을 고려하는 표본추출이 이루어져야 한다.

⑤ 가능하다면 자체가중설계(self-weighting design)가 되도록 표본을 추출한다.

자체가중설계란 모집단에 속하는 최종추출단위(ultimate sampling unit)들의 추출확률을 동일하게 하는 방법이다. 자체가중설계인 경우 각 조사값들의 가중값이 같아지므로 나중에 분석을 할 때 매우 편리하다. 일반적인 통계 소프트웨어에서 제공되는 여러 분석 방법들은 조사값들의 추출확률을 같은 것으로 고려하기 때문에 분석 과정에서 편향이 생길 여지가 많다. 하지만 자체가중설계를 하는 것이 어려운 경우도 있으므로 반드시 자체가중설계를 해야 하는 것은 아니다.

⑥ 자체가중설계가 이루어지지 않은 때에는 표본추출방법에 따른 설계가중값을 반드시 명시하여 추정에 반영한다.

⑦ 조사가 주기적으로 반복되는 계속조사일 때에는 가급적 융통성 있게 표본추출설계를 고려한다.

향후 모집단이나 표본의 상황에 변동이 생길 수 있으므로 표본크기의 변화, 재층화, 추출확률의 수정 등이 가능하도록 한다. 또한 주기적 조사일 경우 응답자의 응답 부담도 사전에 고려하여 이를 반영할 수 있는 추출법을 선택하는 것이 중요하다.

⑧ 계속조사의 경우 표본의 품질을 지속적으로 모니터할 수 있는 절차를 개발한다.

품질이 크게 저하된 층이 생길 경우 수정 · 보완 또는 재설계를 위한 전략을 수립한다. 이를 위해 가능하면 모집단의 변동을 감지할 수 있도록 하는 방안을 강구하는 것이 바람직하다.

가중

① 표본설계 시 되도록 자체가중설계를 하도록 한다.

자체가중설계의 장점은 이해가 쉽고 다른 추출설계에 직접적으로 적용이 가능하다는 점이다. 따라서 표본추출 설계시 가능하면 자체가중설계를 고려함으로써 표본설계의 유연성을 확보할 수 있다.

② 자체가중설계가 어렵다면 가중치 조정단계에 따라 가중값을 부여한다.

만일 자체가중설계가 표본설계과정에서 고려되지 않았거나 고려하기 어려운 경우라면, 기본가중치와 무응답가중치 그리고 사후가중치 조정단계에 따라 가중치를 계산하여 추정치 산정에 반영해야 할 것이다. 또한 가중치 조정단계에서 이용 가능한 정보를 적극 활용하는 것도 고려해야 한다.

③ 가중치 조정과정에서 이용 가능한 보조정보를 적극 활용한다.

이용 가능한 정보는 표본과 표본 외부에 모두 존재할 수 있으며, 만일 외부정보를 활용할 수 있다면 이를 적극적으로 활용하는 것이 바람직하다. 사후층화 조정시에 이러한 외부정보를 사용함으로써 표본과 모집단 간의 일치성을 보장할 수 있을 것이다.

④ 가중치를 적용했다면 가중치 효과를 반드시 계산한다.

통상적으로 가중치를 사용하지 않으면, 추정량의 편향이 발생하게 되는 반면에 가중치를 사용하면 추정량의 편향은 줄일 수 있으나, 분산이 증가하는 문제가 있다. 따라서 무조건 가중치를 적용할 것이 아니라 먼저 분산팽창인자를 이용하여 가중효과를 산정해봄으로써 적절한 가중치를 적용해야 할 것이다.

⑤ 극단가중치는 추정치에 큰 영향을 주므로 적절한 기준하에 절단하여 사용한다.

극단가중치는 분산을 크게 함으로써 추정치의 정도를 떨어뜨리는 역할을 한다. 따라서 적절한 기준하에 극단가중치를 조정할 필요가 있다. 이는 총 MSE 차원에서 다루는 것이 합당한데, 왜냐하면 가중치를 조정하여 분산감소의 효과를 볼 수 있는 반면 추정치의 편향이 발생하기 때문이다.

추정

① 표본설계에 알맞은 추정식을 결정한다.

각종 모수의 추정은 층화, 추출률, 추출방법을 고려하여, 표본설계에 알맞은 가중치 작성 방법과 모수 추정방법을 제시한다. 모집단의 보조정보를 가중치 작성에 반영하여 모수를 추정할 수 있으며 정확한 보조정보가 있는 경우에 보조정보를 추정단계에서 이용하면 추정의 정도는 상당히 향상될 수 있다.

② 통상적으로 복합표본설계에 대한 추정식은 가중치를 고려한 추정식이므로 자체가중설계가 아닌 경우에는 가중추정식을 사용한다.

표본설계의 내용이 복합표본추출설계인 경우 단순한 추정식을 사용하게 되면, 추정량의 편향과 분산추정의 문제가 발생하게 된다. 따라서 표본설계의 내용이 복합표본설계인 경우 반드시 추출 단계별 가중치를 함께 고려한 추정식을 사용하는 것이 바람직하다.

③ 추정량의 분산 또는 상대표준오차를 제시한다.

발표된 통계의 신뢰도에 대한 지표로서 추정량의 분산 혹은 상대표준오차가 사용된다. 또한 사업체나 기업체 통계조사에서 발표된 통계자료가 이후 새로운 분석의 기초자료로 사용될 경우에 정확한 분산은 더욱 필수적이다. 사업체나 기업체 통계조사는 대개 복합표본설계에 의해서 표본이 추출되고, 추정단계에서 가중치를 이용한 추정방법이 사용된다. 따라서 분산추정에도 복합표본설계와 가중치를 반영하여 각 항목에 대해 추정치의 통계적 정확성을 평가할 수 있어야 한다.

표본의 사후관리

조사시스템의 구축

가. 예산, 조직, 인력

① 기대하는 품질수준의 통계를 생산하기 위한 예산과 인력을 배분한다.

통계 생산에서 적정한 수준의 예산과 인력이 배분되지 않을 경우 그 영향은 반드시 조사의 질 저하로 나타나기 마련이다. 신뢰할 수 없는 통계를 생산하는 것은 그 자체가 예산과 인력의 낭비라고 할 수 있다. 그러므로 무조건 예산과 인력을 줄이거나, 그 반대로 늘리는 것이 능사가 아니고 해당 통계 생산을 위해 가장 적정한 수준의 예산과 인력을 확보하는 것이 필요하다.

② 조사의 전 과정을 상세하게 이해하고 관리할 수 있는 전문적 조사관리자를 확보한다.

해당 조사에 정통한 전문적 조사관리자를 키우지 않을 경우 처음 조사를 기획하고 설계하는 단계에서 가장 중요하게 고려되었던 부분들이 시간이 갈수록 잊혀져서 원래 의도와 다른 모습으로 관리될 우려가 있다. 뿐만 아니라, 세월이 지나면서 모집단이나 표본에 새로운 상황이 발생하게 될 경우 이를 원래 설계 시의 의도에 부합하게 대처하지 못하게 되어 편향을 초래할 가능성도 있다. 담당자가 바뀔 경우에도 미리 대비를 하여 조사의 전 과정을 꿰뚫어 볼 수 있는 관리자를 두는 것이 중요하다.

나. 조사 매뉴얼

① 각 조사단계별로 최종적인 결정사항들을 명확히 기술하고 아울러 필요한 참조사항들을 자세히 기록한다.

처음 조사를 기획, 설계하는 과정에서 여러 가지 결정들을 내리게 된다. 이러한 결정을 내리는 과정에서 여러 상황들을 고려하여 특정한 결정을 내리는 것이 보통인데 그런 결정을 내리게 된 이유들이 있을 것이다.

② 조사관리자, 감독자, 조사원 훈련을 위한 매뉴얼을 마련한다.

조사는 결국 사람에 의해 이루어지므로 조사 관련 담당자들이 표준화된 개념, 방법으로 조사에 임하여야 한다. 이를 위해서는 그들을 체계적으로 훈련시키는 것이 필요하다. 조사 매뉴얼에는 조사 관련 담당자들의 훈련을 위한 지침도 포함되어야 한다.

③ 예비조사, 본조사, 편집, 자료분석 등 조사수행 과정에 필요한 운영사항들을 문서화한다.

표준화된 조사, 표준화된 자료처리 및 분석을 위해 필요한 운영사항들을 상세하게 문서화하는 것이 필요하다. 매번 조사 때마다 새로운 사항들이 발견되면 이를 계속 보완해갈 필요가 있다.

다. 장비 및 소프트웨어

① 각각의 시스템에 오류가 없는지 철저히 점검한다.

모든 시스템은 초기에 많은 오류가 발견된다. 따라서 충분한 경험이 쌓이기 전까지는 시스템을 무작정 믿지 않고 철저히 점검하여 수정, 보완하는 것이 필요하다. 관련 전문가들에게 거듭 검토를 받는 것 또한 필요하다.

② 전문적인 시스템의 도입이 어려운 경우 범용 소프트웨어를 활용한다.

큰 규모이면서 중요한 조사일 경우 전문적인 시스템을 도입하는 것이 바람직하다. 하지만 큰 규모의 조사가 아니고 예산상으로도 여유가 없는 조사일 경우에는 널리 일반화된 범용 소프트웨어를 사용하여 시스템을 구축할 수도 있는데 이때에는 개발 비용이나 시간 등을 많이 줄일 수 있다.

추출틀 및 표본 관리

가. 추출틀 관리

① 추출틀 보완주기를 미리 결정한다.

추출틀 보완주기는 모집단의 변화 양상에 따라 달라져야 한다. 변화가 극심한 경우에는 주기가 짧아야 하고 변화가 적은 편이면 상대적으로 주기를 길게 해도 된다. 조사의 관리 체계를 세울 때 추출틀의 보완주기도 미리 결정해두는 것이 바람직하다.

② 해당 조사의 추출틀 보완을 위해 참조할 관련 통계를 찾는다.

해당 조사의 추출틀을 무엇으로 하였건 간에 이것과 유사한 관련 통계를 찾을 수 있을 때가 있다. 정부의 여러 부서에서 매년 생산하는 행정통계 등이 있으므로 해당 조사의 추출틀 보완에 도움이 되는 통계들을 찾아두는 것이 좋다. 모집단에 뚜렷한 변화가 생겼는데 추출틀 보완이 여의치 않은 상황에서는 해당 부분에 대한 추출틀을 별도로 마련할 수도 있다.

③ 대표성 확보를 위해 여러 개의 추출틀을 동시에 활용할 수도 있다.

추출틀을 동시에 여러 개 활용하여 모집단에 대한 포함률을 높이는 표본설계에 관한 연구들이 있으므로 추출틀 보완을 위해 여러 개의 추출틀을 활용하는 방안도 생각할 수 있다.

나. 표본 관리

① 표본조사단위를 대치해야 할 때와 삭제해야 할 때를 구분하여 조치한다.

만일 조사단위가 존재하는데 불응이나 장기부재 등으로 인해 조사가 어려울 때에는 대치를 하는 것이 좋다. 그렇지 않고 조사단위 자체가 소멸되는 경우에는 대치를 하기보다 표본에서 삭제하는 것이 바람직하다.

② 표본조사단위가 추가 또는 삭제될 때에는 관련된 기록을 남긴다.

기존에 조사되던 표본이 추가되거나 삭제될 때에는 해당 기록을 남겨두면 추후에 모집단 변동에 대한 정보로 활용할 수 있다.

③ 표본의 추가나 삭제가 일어날 경우 이를 추정에 적절히 반영한다.

표본의 추가나 삭제가 일어날 경우 이에 따른 가중값 조정 등의 조치가 취해져야 하며 이는 추정식의 수정을 야기하게 되므로 추정 과정에서 이를 적절히 반영해주어야 한다. 그렇지 않으면 추정에서의 편향이 초래될 수도 있다.

④ 일정한 주기가 되면 표본을 전면적으로 개편한다.

계속조사에서 처음 표본이 아무리 잘 설계되었다고 해도 시간이 경과함에 따라 표본의 모집단에 대한 대표성은 떨어질 수밖에 없다. 모집단의 변동 상황을 파악하여 이를 표본에 반영하는 것은 일반적으로 매우 어려운 일이다. 따라서 일정 기간이 경과하면 표본을 전면적으로 재설계하여 개편하는 것이 필요하다. 표본 개편을 위해서는 좋은 추출틀의 마련이

전제되어야 하므로 일반적으로 인구주택총조사 등과 같은 총조사 시행 주기에 맞추어서 표본을 개편하는 것이 바람직하다.

데이터베이스 관리 작업

① 다양한 가능성을 충분히 고려하여 데이터베이스를 설계한다.

가장 단순한 형태의 데이터베이스 관리는 조사 수행 후 입력된 원본 데이터 파일을 그대로 보관하는 것이며 이것은 당연히 해야 할 일이다. 그러나 경우에 따라서는 원본 데이터뿐만 아니라 보고서에 발표된 통계들을 보관 · 관리하는 것이 필요하기도 하다.

② 새로운 통계 수요에 대비한다.

사회가 변화해감에 따라 통계의 수요도 달라져간다. 그러므로 현재의 수요뿐만 아니라 가능하다면 장래 요구될 가능성이 많은 통계 수요도 감안하여 데이터베이스를 관리하는 것이 바람직하다.

무응답에 대한 대책

① 조사의 전 과정에서 응답률을 극대화시킬 수 있는 방안을 마련한다.

일반적으로 응답률에 큰 영향을 미치는 요소로는 조사방법, 조사원의 능력, 조사원의 업무량, 조사주제, 응답부담, 조사표의 길이와 복잡성, 응답자 인센티브 등이 있다.

② 가능하다면 무응답에 대해 재조사(callback)를 실시한다.

무응답자에 대한 재조사는 응답률을 높이는 데 기여하는 동시에 무응답층의 특성을 파악하는 데 도움이 된다. 무한정 재조사를 실시할 수는 없으므로 재조사를 몇 회까지 실시할 것인지, 전체를 재조사할 것인지 아니면 일부만 재조사할 것인지에 대해 구체적인 지침을 마련하는 것이 필요하다. 재조사를 할 때 무응답으로 인한 편향이 클 것으로 생각되는 조사단위에 우선순위를 두는 것이 좋다.

③ 무응답의 원인을 기록하고 모니터 한다.

매번 조사 시마다 응답거부, 부재, 기타 무응답이 발생한 원인을 체계적으로 기록하여 관리한다. 이는 추후에 무응답에 대한 종합적인 분석 및 대책 마련을 할 때 중요한 정보가 될 수 있다.

④ 무응답 데이터에 대해 가중값 조정 또는 대체 등 적절한 조치를 취한다.

무응답 데이터를 삭제하고 그 영향을 고려하여 가중값을 제거하거나 아니면 보정방법을 사용하여 무응답을 보정하는 조치를 취한다. 무응답에 대해 취한 조치에 따라 나중에 추정과정에서도 이를 반영해주어야 하며 이를 명확히 밝혀야 한다. 특히 보정을 하는 경우 데이터 세트에서 보정값 여부를 나타내는 표시(flag)를 반드시 해주어야 한다.

⑤ 응답률 데이터를 공표한다.

모든 조사단위를 응답과 무응답으로 분류하여 표시하고 각 조사의 응답률을 공표하여 조사가 지니는 한계를 밝히는 것이 필요하다.

⑥ 무응답에 관한 정보들을 축적하고 체계적인 연구를 계속한다.

무응답에 관한 정보들이 축적되면 이를 이용하여 여러 유용한 정보를 얻을 수 있으므로 조사과정에서 무응답과 관련된 정보들을 체계적으로 수집해가는 것이 필요하다. 응답자와 무응답자 사이의 특성 차이 등이 밝혀지면 무응답으로 인한 편향 등을 추측하는 데 큰 도움이 된다.

표본 품질관리 매뉴얼

Statistics

발행 • 2007년 10월
인쇄 • 2007년 10월
발행인 • 이창호
발행처 • 통계청

총괄지휘 • 김해수 차장 · 제정본 국장
공동기획 • 김설희 품질관리과장 · 유상길 사무관 · 이지은 주무관
디자인 및 진행 • 예감기획(02-337-3810)

주소 • 대전광역시 서구 둔산동 920 정부대전청사 (〒302-701)
전화 • 042-481-2577
팩스 • 042-481-2463
홈페이지 • www.nso.go.kr ; http://quality.nso.go.kr

발간등록번호 • 11-1240000-000470-14

ISBN • 978-89-5801-109-5 93310