

'Introduction to Survey Quality'

조 사 품 질 개 론

PAUL P. BIEMER & LARS E. LYBERG 공저

김 영 원 & 강 명 욱 공역

머 리 말

조사연구는 세계적으로 번성하고 있는 산업이다. Rossi 등(1983)은 미국에서만 이 산업분야의 총소득이 대략 50억 달러이며 고용인이 60,000명 정도라고 추정하고 있다. 현재 상황에 대한 추정결과는 알 수 없지만, 조사업무 관련 시장은 지난 20년 동안 계속 성장하여왔다. 이러한 성장에 수반되어 조사방법론은 획기적으로 발전되었다. 인지심리학 분야가 조사 대상자에게 조사 참여를 부탁할 때 조사연구자가 취해야 할 접근방법, 설문지설계 방법, 그리고 조사결과의 해석 방법을 극적으로 변화시켰다. 컴퓨터 기술에도 눈부신 발전이 있었고 이로 인하여 자료를 수집하는 방법이 변화되었다. 또한, 새로운 조사자료 품질평가기법의 사용으로 조사결과의 타당성과 신뢰성에 관해서 이전에 가능하다고 생각되었던 것 보다 더 많은 정보가 제공되고 있다. 오늘날의 조사자료 수집은 일찍이 없었던 기술이면서 잘 발달한 과학이다.

동시에, 이 산업에서의 경쟁은 날로 심화되고 있다. 자료 이용자와 조사 의뢰자는 조사 기관이 낮은 비용으로 품질이 더 좋은 자료를 생산할 것을 계속 요구한다. 이런 요구에 답하여 조사 기관은 복잡하고 최적화된 정교한 자료수집과 자료처리 절차를 개발했다. 이 높은 수준의 기술적인 정교함과 복잡함으로 인해 모든 조사단계에 대한 최고의 조사 접근방법을 잘 알고 있는 동시에 이러한 접근방법을 실제 문제에 적용시킬 수 있는 조사업무 전문가들을 필요로 하게 되었다. 매우 적은 수의 조사 업무 종사자만이 대학에서 조사 연구에 대한 교육을 받았기 때문에 조사 기관들은 그들의 고용인들에 대한 대학 졸업 후 과정의 최첨단 조사방법론 교육을 추구하고 있다. 불행하게도 조사방법에 대한 대학 과정의 교육의 발전은 이 산업의 성장에 비해 뒤떨어졌다. 미국과 다른 곳에서 조사방법론에 대한 학위를 수여하는 프로그램이 얼마 되지 않고 조사방법에 대한 교과과정이 부족하며 많은 조사 종사자들이 쉽게 접근하기 어렵다(Lyberg, 2002 참조). 또한, 조사 연구의 실제적인 방법을 교육할 수 있는 다른 수단도 거의 없다.

대부분의 사람이 이용 가능한 한 가지 수단은 조사방법에 대한 문헌이다. 조사방법론에 대한 최근의 결과들이 수록된 전문 정기간행물들을 대학교와 대부분의 기업의 도서관에서 찾을 수 있다. 불행하게도 대부분의 문헌은 조사연구나 통계학에 대한 정식 교육을 받지 않은 조사 업무 종사자들이 이해하기 힘들 것으로 생각된다.

독학으로 지식과 경력을 향상시키기를 바라는 보통 수준의 조사 종사자에게는 문헌에 사용된 전문용어와 조사방법에 대한 이론은 커다란 장애가 된다고 할 수 있다.

우리는 기관 자체가 가지고 있는 이러한 지식의 결여와 그것을 채울 수단의 부족을 인식하고 조사방법론에 대한 입문 교재가 필요하다고 판단했다. 이 책은 초보학생이 조사방법 문헌을 읽을 때 자주 직면하는 광범한 용어, 개념, 방법들에 접하게 할 것이다. 또한, 조사방법론이나 통계학에 대한 정식 교육을 거의 받지 못한 조사업무 종사자가 접근하기 쉽도록 비표본추출오차, 평균제곱오차, 편향, 신뢰도, 타당성, 면접원 분산, 신뢰구간, 오차 모형화 같은 몇 가지 고도의 주제들을 이 책에서 비전문 용어로 다루고자 최대한 노력하였다.

따라서, 이 책은 조사방법 문헌을 읽으면서 직면하게 되는 개념, 전문용어, 표현법, 모형 등에 대한 비전문적이고 종합적인 입문서를 목표로 한다. 이 책의 구체적인 목적은 다음과 같다.

1. 특히 표본추출오차와 비표본추출오차에 중점을 둔, 조사 품질 측정을 위한 기초적인 원리와 개념의 개요를 제공
2. 조사방법론에 대한 문헌을 통하여 조사 품질 측정에 대한 지속적인 연구를 위한 배경을 개발
3. 조사업무에서 마주치는 조사 품질 측정의 개선과 관련된 이슈를 확인하고 그것을 해결하는 기본적 토대를 제공

이 책에서 대상으로 하는 독자층은 조사와 관련된 업무를 수행하는 사람이나 조사 자료를 다루는 사람일 수도 있으나 반드시 교육받은 조사연구자일 필요는 없다. 조사 프로젝트 총 관리자, 자료수집 책임자, 조사 전문가, 통계학자, 자료 처리자, 면접원, 기타 운영자이며, 이들은 표본추출오차와 신뢰구간, 타당성, 신뢰도, 평균제곱오차, 조사설계에서 비용과 오차의 균형, 무응답 오차, 추출틀 오차, 측정 오차, 특정화 오차, 자료처리 오차, 조사자료를 평가하는 방법, 그리고 조사자원의 최선의 사용으로 이 오차들을 줄이는 방법을 포함한 조사자료 품질의 개념에 대한 충분한 이해로 이득을 보게 될 사람이다.

조사연구분야 학생들도 이 책이 대상으로 하는 독자층이다. 이 책은 이론편, 석사 또는 박사 논문의 일부로 조사자료 수집을 필요로 하거나 나중에 연구자로 일하게 될 모든 학문분야의 학생을 위한 강의 교재로 사용할 의도로 집필되었다. 이 책

의 내용에 참고문헌의 읽을거리가 적절히 보충되면 학부나 대학원 수준에서 2 또는 3학점 강의를 위한 충분한 소재가 될 수 있다.

이 책은 어느 한 주제에 대한 심층 연구보다는 조사 품질 측정 분야를 소개하기 위하여 집필되었다. 이 분야에서 최근에 개발된 원리와 개념 외에도 이미 입증된 것들도 검토한다. 또한 아직 해결되지 않고 현재 조사방법 문헌에서 적극적으로 연구되고 있는 중요한 이슈들도 검토한다.

이 책은 조사과정을 통해서 수집된 자료의 품질에 대한 여러 주제를 다룬다. 평균제곱오차(*mean squared error*)와 그것의 구성 요소들로 측정되는 총조사오차(*total survey error*)가 조사자료의 품질을 평가하는 가장 중요한 척도이다. 1장에서는 조사연구의 유래를 알아보고 조사품질과 자료품질의 개념을 소개한다. 2장에서는 자료품질이 어떻게 측정되는지에 대한 비전문적인 논의와 조사설계를 비용과 적시성의 제한 조건 아래서 최적화하기 위한 기준을 제공한다. 여기에서는 이 책의 전체에 걸쳐 사용되는 자료품질의 본질적인 개념을 제시한다.

그 다음에는 조사 오차의 중요한 원인을 상세하게 논의한다. 특히, 우리는 (1) 각 오차 원인의 발단 (즉, 그것의 근본적인 원인), (2) 이 오차 원인에서 나오는 오차를 줄이기 위하여 제안된 가장 성공적인 방법, (3) 원인이 총조사오차에 미치는 영향을 평가하기 위하여 실제로 가장 흔히 사용되는 방법을 검토한다. 3장에서는 포함오차와 무응답오차를 다루고, 4장에서는 전반적으로 측정오차를, 5장에서는 면접원 오류를, 6장에서는 자료수집방식을, 7장에서는 자료처리오차를 다룬다. 8장에서는 자료 품질을 평가하기 위한 근본적인 접근방법들을 요약한다. 9장에서는 표본추출오차의 기본 원리를 다룬다. 마지막으로 10장에서는 이 책 전체에 걸쳐 사용된 많은 개념을 실제적인 조사설계를 위한 교훈들로 통합하였다.

이 책은 조사오차의 본질을 이해하기 위한 많은 개념과 아이디어, 조사품질의 개선을 위한 기법, 그리고 가능한 경우에는 그것들의 비용 문제와 진행 중인 조사 프로그램의 자료품질을 평가하기 위한 방법을 다룬다. 이 책의 중요한 주제는 독자에게 조사오차를 다루는 언어(*language*)나 전문용어를 소개해서 그들이 독학으로 또는 문헌에 나온 읽을거리를 통해 조사방법론연구를 계속할 수 있도록 하는 것이다.

이 책의 집필에는 4년이 걸렸다. 하지만 많은 내용은 두 저자 중 한 명인 P.P. Biemer이 여러 곳에서 강의한 단기 코스의 내용의 일부로 10여년에 걸쳐서 개발된 것이다. 이러한 장소에는 미시간대학교 조사연구센터(*University of Michigan*

Survey Research Center)와 메릴랜드대학교-미시간대학교 조사방법론 공동 프로그램(University of Maryland-University of Michigan/Joint Program in Survey Methodology)이 있다. 이 기간동안 많은 사람들이 책과 강의에 기여했다. 여기서 우리는 그들의 기여에 감사하고 노력에 진심 어린 사의를 표한다.

이 프로젝트의 전 기간에 걸쳐 저자의 소속 기관(RTI, International and Statistics Sweden)이 이해와 격려로 후원해 준 것을 진심으로 감사하게 생각한다. 4년 동안 밤과 주말에 걸친 이 책의 집필로 인하여 우리는 분명히 일상 업무에 소홀하였을 것이다. 특히 우리가 초고를 완결시키기 위해 질주하던 프로젝트의 끝 무렵에는 정상적 근무시간 외에는 일반 업무를 거의 하지 못했다. 우리는 RTI과 미국 NASS의 재정적 후원에 감사한다. 이것은 저자 중 한 명(P.P.B)이 사무실에서의 시간을 줄여서 책을 집필하는 것을 가능하게 했다. 또한 두 저자 모두에게 두 대륙을 오가는 데 드는 여행시 생활비 이외에도 유럽과 미국으로의 몇 번의 여행을 부분적으로 지원했다.

여러 사람이 이 책의 여러 장을 검토했고 개선을 위한 뛰어난 의견과 제안을 했다. 그들은 Fritz Scheuren, Lynne Stokes, Roger Tourangeau, David Cantor, Nancy Mathiowetz, Clyde Tucker, Dan Kasprzyk, Lynn Lepkowski, David Morganstein, Walt Mudryk, Peter Xiao, Bob Bougi와 Peter Lynn이다. 물론 그들의 기여로 책이 크게 개선되었다. 또한, Rachel Caspar, Mike Weeks, Dick Kulka, Don Camburn은 다양한 지원을 했다. 또한 책의 내용에 많은 영향을 주고 수업을 어떻게 개선할지에 대한 많은 제안을 한 많은 학생들에게도 감사한다.

마지막으로 이 기간동안 가족들의 희생에 감사한다. 책의 저술 작업이 계속 진행되어야 했기 때문에 우리가 가족의 여가활동과 가족 행사에 참석하지 못하거나 그들과 함께 할 수 없었던 많은 일들이 있었다. 처음에 생각했던 것보다 더 긴 기간이었음에도 불구하고 그들의 인내와 격려, 그리고 상황에 대한 이해를 고맙게 생각한다.

2002년 6월

미국 노스캐롤라이나 리서치 트라이앵글 파크에서 Paul P. Biemer

스웨덴 스톡홀름에서 Lars E. Lyberg

제 1 장 조사 과정 품질의 발전

통계학은 불확실한 상태에서 지식을 얻고 옳은 결론을 내리는 방법을 종합한 학문이다. 통계학은 관찰, 가설의 공식화, 예측, 검증과 같은 과학적인 연구의 모든 단계에 적용된다. 이런 통계적인 방법 중에는 기술통계학, 실험계획법, 상관분석과 회귀분석, 다변량분석과 다차원분석, 분산분석과 공분산분석, 확률과 확률 모형, 우연변동과 우연모형과 유의성검정 등이 포함된다.

이 책에서는 통계학의 한 분야인 조사 방법론(*survey methodology*), 보다 구체적으로 조사 품질(*survey quality*)을 다룬다. 1장에서는 이 책의 토대가 되는 조사와 조사 품질의 개념을 정의한다. 이를 위해 우선 조사란 무엇인가를 정의하고 1.2절에서는 오늘날 실제로 접할 수 있는 전형적인 조사들의 유형에 대해 설명한다. 사회·경제적 연구에서 조사 방법론의 발전의 역사에 대한 간단한 기술을 통해 조사와 관련된 사항을 1.3절에서 다루고자 한다. 그 다음 세 절에서는 정의하기 매우 어려운 개념인 품질, 특히 조사 품질을 다룬다. 최근 몇 해 동안 조사가 변천함에 따라 조사의 수행에 관련된 내용이 어떻게 발전되었는지를 통해서 품질의 의미를 간단히 설명한다. 이런 관점에서 1.4절에서는 품질 혁명(*quality revolution*)을 다루고, 1.5절에서는 통계기관에서의 품질의 의미를 설명한다. 1.6절과 1.7절에서는 각각 조사 과정 품질의 측정법과 개선방안을 다룬다. 마지막으로 1.8절에서는 이 장의 중요 개념을 요약한다.

1.1 조사의 개념

미국통계학회(*American Statistical Association; ASA*)의 조사연구방법분과(*Section on Survey Research Methods; SRM*)는 “조사란 무엇인가?”라는 제목의 10개의 소책자를 발간하였다(Scheuren, 1999). 이 소책자 시리즈에서는 주요 조사 단계와 조사를 수행하기 위해 필요한 구체적인 사안들을 다루고 있다. 이는 일반대중을 위해 발간된 것이고 이것의 궁극적인 목표는 조사에 참여하거나, 조사의 결과를 이용하거나 또는 그 분야에 대하여 알고 싶어 하는 사람들에게 조사에 대한 이해를 증진시키는 것이다.

Dalenius(1985)는 몇 가지의 필수적인 연구 전제 조건을 포함하는 조사의 정의를 제시하고 있다. Dalenius에 의하면 다음의 전제 조건들을 만족하여야만 연구 과제가 조사라고 할 수 있다:

1. **조사는 모집단을 구성하는 일련의 객체에 관한 것이다.** 다양한 형태의 모집단이 있을 수 있다. 우선 개인, 기업 또는 농장과 같은 유한 객체들로 구성된 모집단이 있을 수 있다. 다른 형태의 모집단으로 일정 기간에 발생한 특정 사건의 발생과 관련된 경우를 생각할 수 있다(예를 들어, 범죄에 의한 희생과 사고). 또한 토지의 활용 실태 또는 어떤 지역의 야생동식물의 서식 등과 같은 자연환경에서 일어나는 사건들과 연계된 모집단도 고려할 수 있다. 어떤 경우이든지 관심의 대상이 되는 모집단(목표 모집단; *target population*)은 항상 구체적으로 정의되어야 한다. 때때로 현실적인 또는 경제적인 이유로 조사대상을 제한할 필요가 있다. 예를 들어 조사의 대상이 되는 모집단에서 어떤 외딴 지역을 제외해야 될 지도 모르고, 경우에 따라 면접 조사상에 큰 어려움이 없는 연령대로 조사대상을 제한해야 될 지도 모른다. 가구 모집단 조사에서는 기관에 수용되지 않은(군대를 제외한 교도소 또는 병원 등과 같은 기관에 수용되지 않은) 15세에서 74세까지의 사람들로 특정 기간에 국내에 거주하는 사람만을 조사대상으로 하는 것이 일반적이다.

2. **조사의 대상이 되는 모집단은 한 가지 이상의 측정 가능한 특성을 가지고 있다.** 예를 들어 개인들로 이루어진 모집단에서 특정 시점에 어떤 사람이 현재 종사하는 직업은 측정 가능한 특성을 나타낸다. 또한 일정 기간동안 발생한 특정 범죄 건수는 사건발생과 관련된 모집단에서 측정 가능한 특성에 해당한다. 한편 환경 관련 모집단에서는 인구가 밀집된 지역의 비율과 같은 것들이 측정 가능한 특성이 될 수 있다.

3. **연구과제의 목표는 측정 가능한 특성으로 정의된 한개 이상의 모수로 모집단을 기술하는 것이다.** 이를 위해서는 모집단(모집단의 일부분인 표본)을 관측하는 것이 요구된다. 모집단에서 실업율, 일정 기간 동안 특정 산업분야의 총 사업 수익, 그리고 특정 시점에 어떤 지역에 서식하는 특정 동식물의 총수 등이 모수에 해당하는 예들이다.

4. **조사관측을 위해서는 모집단의 구성 객체에 접근을 가능하게 하는 추출틀(모집단을 구성하는 모든 객체들을 인지할 수 있는 목록 또는 지역의 지도와 같**

은 특정 형식의 틀)이 필요하다. 추출틀의 예로는 사업체 또는 주민등록 명부, 경계선이 명확하게 정해진 지역들을 구분하는 지도, 또는 전화번호를 개인에게 연계 시키는데 활용할 수 있는 모든 n자리 숫자들의 집합 등을 들 수 있다. 쉽게 활용할 수 있는 추출틀이 없는 경우에는 목록 작성 작업을 통해서 추출틀이 구성되어야 한다. 일반 모집단의 경우 적절한 표본을 추출하기 위해 필요한 이런 추출틀 작성은 매우 지루하고 방대한 작업일 수도 있다. 다단계 표본추출에서는 먼저 지도를 이용해서 몇 개 지역을 표본으로 추출하고 현지 직원이 표본으로 선정된 지역에 있는 모든 객체를 조사하여 목록을 만들어서 다단계 표본추출틀을 완성한다. 미국의 프 로야구 선수들의 모집단 같은 특수한 모집단의 경우에는 모든 팀의 선수명단을 합쳐 하나의 큰 명부를 구성하고, 이를 표본추출에 활용할 수 있다. 어떤 경우에는 여러 개의 불충분한 목록 또는 추출틀들을 함께 이용하기도 하는 데, 이런 경우 이런 불완전한 추출틀들을 체계적으로 종합해서 최종 추출틀을 구성하는 작업이 필요하다. Hartley(1974)는 이런 상황을 위해 다중추출틀이론(*multiple frame theory*)을 개발하였다.

5. 확률구조와 표본크기가 명확히 정해진 표본 설계에 따라 추출틀에서 표본 객체가 선택된다. 표본 조사 관련 문헌에는 다양한 상황에 사용될 수 있는 많은 표본설계들이 제시되어 있고, 기본적으로 다음 두 가지 사항이 설계 상황에 따라 고려되어진다. 첫 번째 사항은 2단계 이상에 걸친 표본 추출과 최종 단계에서 조사대상으로 확인된 객체만을 측정하는 과정이 필요하다면, 이런 조사과정을 좀더 손쉽게 수행할 수 있도록 하는 것이다. 여기서는 가능하면 명부작성 및 조사원의 이동에 따른 비용절감 등과 같은 조사편의를 주로 염두에 두게 된다. 두 번째 사항은 관심을 갖는 특성에 대한 모집단 분포를 표본설계에 고려하는 것이다. 예를 들어 모집단 분포가 한쪽으로 매우 치우친 경우 층화표본추출 또는 절단표본추출(*cutoff sampling*)을 적용하는 경우이다. 이 경우 모집단 분포를 반영하여 절단표본추출의 경우 가장 큰 몇 개의 객체들을 조사대상으로 하게 되며, 층화표본추출에서는 먼저 모집단 객체들을 크기 순서에 따라 나열하고 매 n 번째 객체를 표본으로 추출하는 계통표본추출법을 흔히 사용하게 된다. 어떤 표본추출 방법을 적용하더라도 모든 표본 추출설계에서는 반드시 선택확률과 표본 크기를 구체적으로 정해야 하며, 선택확률을 파악할 수 없으면 이런 것은 통계적으로 타당한 표본설계라고 볼 수 없다.

6. 표본의 관측은 측정과정에 의해 이루어진다(측정방법과 그것의 사용에 관한 규정). 관측은 자료수집방법(*data collection mode*)이라는 측정수단을 통해 이루어진다. 자료 수집은 여러 다양한 방법으로 시행될 수 있다. 특정 개인, 기업, 또는 지역이 관측단위가 될 수 있다. 관측은 어떤 기계적 장치(예를 들어 TV 시청자의 프로그램 시청 행위를 기록하는 전자 장치), 직접관측(예를 들어 항공사진에 나타난 야생동물의 숫자를 세는 것), 또는 조사면접원 또는 조사대상자 자신이 작성한 설문지(연구 목적에 따른 특정 개념을 반영하는 질문을 통해 사실과 행동을 관측)를 통해 수행된다.

7. 표본으로 모집단에 대한 추론을 하기 위해서는 모수의 추정값을 계산하기 위한 관측자료를 기반으로 한 특정한 추정과정이 적용된다. 관측을 통하여 자료가 만들어지고, 각각의 표본설계의 특성에 따라 유도된 자료를 기반으로 추정값을 계산할 수 있는 한개 이상의 추정량이 주어진다. 추정량은 단지 수집된 자료만을 토대로 할 수도 있지만 필요에 따라 다른 정보를 활용할 수도 있다. 모든 추정량은 표본 자료가 갖는 모집단에 대한 대표성을 보완하는 것을 목적으로 하는 표본 가중값을 포함하게 된다. 모집단 전체가 아닌 표본에서 얻어진 추정값을 사용함에 따라 발생하는 오차는 분산 추정량을 이용해서 관측된 자료에서 직접 계산 할 수 있다. 분산 추정량은 표준오차와 신뢰구간의 계산을 가능하게 한다. 그러나 조사 자료가 내포하고 있는 모든 오차가 분산에 반영되는 것은 아니다.

Dalenius가 제시한 7가지 기준(전제 조건)과 이와 관련된 간단한 참고사항을 <표 1.1>에 요약했다. 이들 7가지의 기준을 통해 조사의 개념을 정의할 수 있다. 이들 중 한 가지 이상이 이행되지 않을 경우 그 연구는 조사로 분류될 수 없고 이 경우 결과적으로 선택된 표본으로 목표모집단에 대한 올바른 추론을 내릴 수 없게 된다. 그러나 심각한 결점이 있고 추론결과의 신뢰성이 매우 의심스럽지만 조사라고 불리어지는 연구들을 우리는 흔히 접할 수 있다.

<표 1.1> Dalenius의 조사에 필요한 전제 조건

기준	참고사항
1. 조사는 모집단을 이루는 일련의 객체에 관한 것이다.	목표모집단을 정의하는 것은 추론적인 목적에서 뿐만 아니라 표본추출틀을 구성하는데 있어서 매우 중요하다.
2. 조사의 대상이 되는 모집단은 한 가지 이상의 측정 가능한 특성을 가지고 있다.	연구과제의 목표를 가장 잘 이룰 수 있도록 관심대상 특성이 선택 되어야 한다.
3. 연구과제의 목표는 측정 가능한 특성으로 정의되는 여러 개의 모수로 모집단을 설명하는 것이다.	관심 특성들이 주어졌을 때 평균, 총계, 백분위수, 또는 부차그룹에 대한 평균 등 다양한 모수들이 정의 될 수 있다.
4. 관측을 위해 모집단 객체에 접근하기 위해서는 추출틀이 필요하다.	목표 모집단을 완전히 포함하는 추출틀을 개발하기가 어려울 때도 많다.
5. 확률구조와 표본크기가 명확히 정해진 표본설계에 따라 추출틀에서 표본을 추출한다.	표본설계는 항상 조사와 관련된 실제 상황에 따라 결정된다.
6. 정해진 측정과정에 따라 표본에 대한 관측이 이루어진다.	자료 수집은 다양한 방법으로 시행될 수 있다. 흔히 여러 가지 방법이 사용된다.
7. 표본으로부터 모집단에 대한 추론을 위해 모수 추정값을 계산하는 위한 관측자료를 토대로 한 특정한 추정과정이 적용된다.	전체 모집단이 아닌 표본만을 관측함에 따라 발생하는 오차는 분산 추정량을 써서 계산할 수 있다. 여기서 얻어진 추정값은 신뢰구간을 구하는데 사용될 수 있다.

출처 : Dalenius(1985)

일반적인 조사연구에서 나타나는 대표적인 결함들을 정리하면 다음과 같다.

- 조사단위를 찾거나 접근하는데 생기는 어려움 때문에 조사 도중에 목표 모집단을 다시 정의 한다. 예를 들어 조사를 위한 이동상의 문제 또는 자료수집 비용 때문에 어떤 지역이나 연령대에 속하는 객체는 관측이 불가능할 수 있다. 따라서 이런 객체는 실제로는 연구대상에서 제외되지만 조사목표에 있어서는 이런 변경이 제대로 반영되지 않는다.

- 모든 선택된 조사단위의 선택확률을 알 수 없다. 예를 들어 최종표본이 미리 명확히 정해진 특성별 표본수에 따라 구성되도록 조사원이 자신이 원하는 응답자를 마음대로 선택하도록 하는 할당표본추출(*quota sampling*)을 사용하는 경우 선택확률은 알 수 없다. 이런 표본추출은 쇼핑몰 방문객이나 공항의 여행자를 조사할 때 흔히 사용된다. 또한 많은 연구에서 자발적 참여에 의해 구성된 자기선택(*self-selection*)표본이 사용되는 데 이도 이런 경우에 해당한다.

예를 들어 호텔서비스에 대한 고객조사를 위해 고객들에게 객실에 비치되어 있는 설문지를 작성하여 프린트에 제출토록 요청한다고 하자. 이런 경우 비교적 적은수의 손님들(아마도 10%이하)이 설문지를 작성할 것이며, 그럼에도 불구하고 호텔의 경영자측에서는 “조사에 의하면 우리 손님의 85%가 호텔 서비스에 만족 한다”는 것과 같은 보고서를 낼 수 있다. 여기서 백분율은 호텔 서비스에 만족하는 손님의 수(설문지 응답자 중)를 프린트에 제출된 총 설문지의 수로 나누어 계산된 것으로 설문지를 쓰지 않는 대다수의 손님들이 의견은 전혀 반영되지 않고 있다.

누가 설문지에 응답하고 안하고에 대해 통제가 전혀 이루어지지 않기 때문에 이런 조사에서 얻어진 추정결과는 편향되었을 가능성이 매우 높다. 다른 예로 신문이나 TV프로그램에 자주 인용되는 웹 또는 이메일을 통해 이루어지는 인터넷 조사이다. 이 경우 구독자나 TV시청자들은 인터넷에 접속해서 그들의 생각을 표현하도록 강요받고 있다. 그 결과는 대부분 특별한 설명 없이 공표되고, 대중은 그 결과가 모집단의 실제 상황을 나타낸다고 믿게 될 수 있다. 신문이나 TV에서 선전하는 인터넷조사의 경우에는 자발적 참여에 의한 자기선택표본이 최소한 다음 4가지 과정을 통해 선택된다. (1) 응답자가 응답을 할 기회가 있기 위해서는 구독자나 시청자이어야 한다. (2) 응답자는 인터넷을 이용할 수 있어야 한다. (3) 응답자는 인터넷을 통해 참여할 의향이 있어야 한다. (4) 응답선택문항에 “모름” 과 “의견 없음”은 거의 사용되지 않기 때문에 응답자는 자기의 의견을 무조건 피력해야 한다. 물론 이런 종류의 자기선택표본은 어떤 종류의 확률표본(*random sample*)에도 해당되지 않는다.

- 정확한 추정식이 사용되지 않았다. 어떤 조사에서 쓰이는 추정식은 정확한 표본가중값이 없거나 분산공식이 실제 적용된 표본설계를 제대로 반영하지 못하고 있다. 조사실무자들은 흔히 이와 같은 사용된 표본추출방법에 적합하지 않은 분

산계산 방법을 잘못 사용하곤 한다. 어떤 사람들은 상당히 복잡한 표본설계가 적용되었음에도 불구하고 가장 간단한 기본적인 표본설계에서 사용되는 분산계산 방법을 그대로 사용하는 오류를 범한다.

이런 것들이 조사의 기본적인 기준(전제조건)의 위반 사항에 해당한다. 이런 것들은 잘 계획된 과학적인 조사의 표본설계 및 실행에서도 표본조사의 불완전성에 기인한 불가피하게 발생하는 조사오차와 혼동되어서는 안 된다. 이 책에서는 후자의 조사오차에 대한 내용(오차원인, 오차구조, 오차 예방책, 오차 추정방법 등)을 다룬다. 오차(*error*)는 실수가 있었다는 것을 의미하기 때문에 오차라는 용어는 많은 사람들, 특히 조사 자료를 생산하는 사람들에게는 매우 부정적으로 들릴 수 있다. 따라서 어떤 사람들은 불확실성(*uncertainties*) 또는 자료의 불완전성(*imperfections*) 같은 보다 긍정적인 전문용어를 더 선호하지만, 그것들은 실제로는 우리가 사용하는 오차라는 용어와 같은 의미를 갖는다. 최근 몇 십년동안 품질(*quality*)이라는 용어는 자료 사용자가 중요하다고 믿는 조사결과의 모든 특징을 포함하기 때문에 널리 쓰이게 되었다.

조사에서 다음과 같은 문제점으로 인해 통계적 추론에 결함이 발생할 수 있다.

- 조사하는 동안 목표모집단의 정의를 바꾸는 것
- 알려지지 않은 선택확률
- 잘못된 추정식과 추론

1.2 조사형태 및 조사기관

다양한 형태의 조사와 조사모집단이 있다(Lyberg와 Cassel, 2001 참조). 많은 조사는 모집단의 특성 또는 행동과 태도를 측정하는 것을 목적으로 하는 일회성 조사이다. 한편 연속조사들은 일정 기간 동안 연속적으로 조사가 시행되며, 시간에 따른 변동을 추정하는 것을 주요 목적으로 한다. 어떤 경우 시간이 흐름에 따라 발생하는 모집단 변동에 대한 사용자들의 관심이 증대됨에 따라 일회성 조사로 시작되었

던 조사가 반복적으로 실시되고 결국 연속조사로 변하는 경우도 흔히 발생한다.

정부기관에서 발표하는 개인, 기업, 기관 또는 농업 경영체 등을 모집단으로 하는 공식통계들은 대부분 연속조사 프로그램에 의해 산출된다. 예를 들어 대부분의 나라들은 실업, 인구, 소매업, 가축, 농작물 및 운송 등과 관련된 통계를 생산하기 위한 조사 프로그램들을 갖고 있다. 전 세계 거의 모든 나라에는 정책 결정자나 다른 사용자들에게 이런 정보를 지속적으로 제공하기 위한 한개 이상의 정부기관(국가통계기관)이 있다. 이런 정부기관에서 생산된 자료를 일반적으로 공식통계(*official statistics*)라고 한다.

국제통화기금(*International Monetary Fund; IMF*), 유엔(*United Nations; UN*)과 그 산하의 식량농업기구(*Food and Agricultural Organization; FAO*) 또는 국제노동기구(*International Labour Office; ILO*) 그리고 모든 중앙은행들처럼 조사 자료를 수집하고 분석하는 일이 직무 중에 일부 포함된 대형 기구들이 있다. 한편 유럽연합(*European Union; EU*)에 속하는 모든 국가 통계기관들을 위한 중앙 기구 역할을 하는 Eurostat, 이와 유사한 기능을 하는 아프리카의 Afristat, 미국에서 다양한 자료 수집활동을 감시하고 허가하는 업무를 담당하는 미국 행정관리에산국(*Office for Management and Budget; OMB*) 등과 같은 기구들은 모두 자료수집활동을 통합 조정하고 감독하는 기능을 하고 있다.

또한 다른 형태의 자료 수집은 교육기관과 일반 조사회사에서도 수행된다. 정부기관이 적절하다고 판단하는 경우 때로는 이들 기관에서 정부의 공식통계를 생산하기도 한다. 이런 상황은 나라에 따라 차이가 있다. 어떤 나라에서는 국가통계기관이 아닌 다른 기관이 공식통계를 생산하는 것을 허용하지 않지만, 어떤 나라에서는 다른 조사기관이 국가통계를 생산하는 것이 허용되는 경우도 있다. 일반 조사회사는 개별 회사 또는 조직과의 계약을 통해 보통 시장조사, 여론조사, 태도 및 행동에 대한 조사와 특수한 모집단의 특징을 파악하기 위한 조사를 주로 수행한다. 미국에서만 130,000여명 이상의 사람들이 조사산업에 종사하고 있으며, 전 세계적으로는 조사산업에 종사하는 사람의 수는 이 보다 훨씬 클 것이다. 예를 들어, 유럽의 정부통계기관 중에는 적은 경우 6명 정도의 직원만을 고용하고 있는 경우(룩셈부르크)도 있지만, 반면 수천명의 직원을 고용하고 있는 경우도 있다.

조사연구를 실시하는 전 세계의 통계기관들이 처하고 있는 여건은 매우 다양하다. 한편에서는 모집단에 대한 통계를 산출하기에 매우 적합한 표본추출틀의 확보

가 가능하고, 조사방법 이론에 대한 전문적인 지식을 갖춘 충분한 인력을 확보하고 있으며, 아울러 컴퓨터의 도움을 받는 조사방법의 사용과 같이 발전된 과학기술을 마음껏 활용할 수 있는 여건을 갖추고 있는 국가들도 있다. 하지만 개발도상국과 과도기에 있는 나라들은 발전된 조사방법론의 적용, 컴퓨터나 전화 같은 과학기술의 이용 또는 충분한 전문지식을 갖춘 인력의 확보 등에 있어서 많은 제약을 가지고 있다.

예를 들어 대부분의 개발도상국에서는 적당한 표본추출틀을 확보할 수 없고 전화 보유율이 낮아 조사를 위해 전화를 사용하는 것이 어렵다. 따라서 개인 면접조사가 유일한 실현 가능한 조사방법이다. 아울러 조사실시를 위한 충분한 예산을 확보하는 것이 어렵다는 것도 개발도상국 뿐 아니라 세계의 많은 국가에서 체계적인 조사 연구를 수행하는 데 장애 요인이 되고 있다.

조사방법을 개발 향상시키고 관련 연구를 촉진시키는 것을 후원하기 위한 단체들도 있다. 미국통계학회(*American Statistical Association; ASA*)의 조사연구방법분과(*Section on Survey Research Methods; SRM*), 국제통계기구(*International Statistical Institute; ISI*)의 국제조사통계학회(*International Association of Survey Statisticians; IASS*)와 미국여론조사학회(*American Association for Public Opinion Research; AAPOR*) 등이 이런 단체들이다. 또한 여러 국가에는 조사와 관련된 연구를 하는 산하기구가 있는 통계학회들이 있다. 세계의 많은 대학교에서도 조사에 관한 연구를 하고 있으며, 이런 조사 관련 연구는 통계학과에서만 아니라 사회, 심리, 교육, 언론, 경영학과 등에서도 수행되고 있다. 최근 조사방법론이 학제간의 공동연구를 필요로 하는 학문이라는 인식이 확산되면서 조사연구에 대한 여러 학문분야의 공동연구가 증가되는 추세에 있다.

조사산업의 중요한 역할은 전 세계의 지도자들에게 의사결정을 위한 중요한 정보를 제공하는 것이기 때문에 주요 정책 결정의 기반이 되는 조사 자료는 높은 품질 수준을 갖추는 것이 필수적이다. 따라서 조사기관은 좋은 품질을 확보하는 방법들을 알고 이용할 수 있어야 한다. 그러나 현실은 불행하게도 그렇지 못한 경우가 많으며, 이것이 우리가 이 책을 쓰는 가장 중요한 동기이자 목적이다.

1.3 조사방법론의 간단한 역사

조사의 뿌리는 성경시대로 거슬러 올라가 찾을 수 있다. Mandasky(1986)는 “성서 인구조사”라고 부르는 구약성서에 나오는 센서스 보고를 제시하고 있다. 국가가 전쟁이나 과세목적으로 인구가 대략 얼마인지를 아는 것은 매우 중요하다. 센서스는 고대 이집트, 로마, 일본, 그리스와 페르시아 등에서 실시되었다. 국가의 인구는 바로 국력을 의미하는 것으로 여겨졌다. 예를 들어 1700년경만 해도 스웨덴의 인구조사에서 예상했던 것 보다 인구가 훨씬 적은 것으로 나타나자, 이 센서스 결과가 많은 부정적인 영향을 줄 것을 염려한 스웨덴정부는 그 결과를 비밀로 하였다. 스웨덴 정부의 가장 큰 걱정은 센서스에 나타난 적은 인구를 그대로 공개함으로써 다른 나라들로부터의 침략을 유발할지도 모른다는 두려움이었다.

조사표본추출은 여러 세기동안 직관적인 측면에만 의존하여(Stephan, 1948) 1900년경까지는 명확한 표본추출이론의 발전이 없었다. 전체 국민에 대한 센서스가 현실적으로 불가능하다고 판단하던 시절인 1900년 이전의 유럽에서는 인구를 추정하는 것이 많은 과학자들의 관심대상 이었다. 몇몇의 유럽 국가에서 사용된 정치계산(*political arithmetic*)이라는 방법은 1650년에서 1800년까지 영국의 Graunt와 Eden에 의해 성공적으로 사용되었다. 정치계산은 비율추정과 유사한 개념을 기초로 한다(9장 참조). 출생률, 가족수, 거처당 평균 거주자 수, 그리고 선택된 일부 지역에 대한 과학자들의 개인적인 관찰결과를 종합하여 전체인구를 추정하는 것이 가능했다. 이들 중 일부 추정값은 나중에 센서스에 의해 매우 정확한 것으로 확인되었다. 프랑스와 벨기에에서도 비슷한 시도가 있었다. 이런 것들에 대한 보다 자세한 논의는 Fienberg와 Tanur(2001)와 Bellhouse(1998)을 참조하기 바란다.

조사방법론의 과학적 원리는 수학, 확률론과 수리통계학에 뿌리를 두고 있다. 순열과 조합의 계산과 관련된 문제는 상당부분 이미 10세기에 풀렸다. 이런 연구결과는 확률론의 기초가 되었고, 1540년에는 Cardano가 확률(*probability*)을 “성공에 해당하는 경우의 수를 모든 가능한 경우의 수로 나눈 값”이라고 정의하였다. 이는 현재에도 많은 기초통계과목에서 사용되는 고정적인 확률의 정의로 사용된다. 17세기에는 Galiei, Fermat, Pascal, Huygens 와 Bernoulli가 확률론을 발전시켰다. 그 다음 150년 동안 Moivre, Laplace, Gauss, Poisson 같은 과학자들이 수학, 확률과 통계를 더욱 발전시켰다. 극한 정리와 분포함수는 이 시대의 가장 큰 공헌중의 하나이며, 오늘날 통계이론 중 많은 것들에 이들의 이름이 붙여져 있다.

19세기말과 그 후 몇 십년동안 표본조사가 전수조사 또는 센서스를 대체할 수 있다는 생각이 지배적이었다. 1895년에 노르웨이의 Kiear는 그가 “대표조사 (*representative investigation*)”라고 부른 방법에 대해 좀더 심층적인 연구가 필요하다는 제안을 ISI에 했다. 이런 방법이 관심을 끈 이유는 Graunt 등 다른 사람들도 이와 유사한 생각을 가졌기 때문이다. 비용 측면뿐만 아니라 상세한 조사를 하는데 필요한 엄청난 노력과 정성을 고려할 때 전수조사는 많은 경우 불가능하다. Bowley도 Kiear와 함께 “대표조사” 방법의 실용성에 대해 ISI를 설득하기 위해 노력했다. Kiear는 1897년, 1901년과 1903년 세 번의 ISI 회의에서 표본추출법에 대해 발표했다. 10년 후 Bowley(1913)가 통계이론과 조사설계를 연결시키려는 시도를 했다. 그는 몇 개의 논문에서 확률표본추출, 추출틀의 필요성과 일차 표본추출단위의 정의 등에 대해 논했다. 그는 유의선정에 대한 이론의 개요를 제시하고 조사설계를 위한 지침을 마련했다. Kiear이나 Bowley 어느 누구도 확률화의 필요성에 대해서는 주장하지 않았던 점에 특히 유의해야 한다. 그들은 랜덤선정과 유의선정을 통합하여 사용하는 방안을 처음으로 주장했다.

예를 들어 조사단위와 작은 집락은 랜덤하게 또는 아무렇게나(무계획적으로) 선정되어야 되고 큰 집락은 원하는 의도에 따라 유의적으로 선정되어야 한다는 것이다. 이런 노력들과는 별개로 층화확률추출에서 추정값 공식을 개발한 Tschuprow에 의해 매우 유사한 발전이 러시아에서도 일어나고 있었다. 마침내 1920년대 중반에 ISI가 이 방법에 대한 집중적인 연구를 권장하고 이런 기법의 활용에 동의했다. 하지만 전수조사 대신에 표본을 사용하면서 어떻게 대표성을 확보하고, 이와 관련된 불확실성을 어떻게 측정할 것인가 등에 대한 세부적인 사항은 전혀 명백하지 않았다. 어떤 나라에서는 표본추출이 과학적인 방법으로 완전히 받아들여지는데 몇 십년이 걸리기도 했다.

Tschuprow에 의해 제시된 몇 가지 방법들은 Neymen에 의해 좀더 발전되었다. Neymen이 유한모집단에서의 표본추출이론에 대한 개요를 발표하였을 때 그가 Tschuprow의 결과를 이용할 수 있었는지 여부는 명확하지 않다. 연구결과가 다소 중복됨에도 불구하고, Neymen은 1920년대 그의 초기의 연구결과를 발표하면서 Tschuprow라는 이름을 한번도 인용한 적이 없다.

그 뒤 몇 년 동안 표본조사의 발달이 매우 급진적으로 이루어졌다(9장 참조). Neymen(1934)은 “대표(표본)조사 방법에서 두 가지 견해에 대하여: 층화추출법과

유의선정방법”이라는 역사적인 논문을 발표했다. Neymen은 그의 논문에서 확률추출법의 중요성을 강조했다. 그는 또한 최적층화, 집락추출, 대표본에서 선형추정의 정규성 근사와 유의선정을 위한 모형을 다루었다. 그의 연구들은 매우 획기적인 것들이었지만 그의 연구결과가 탁월하다고 인정되기까지는 시간이 걸렸다. Neymen의 연구는 농업통계에서 출발했고, 로담스테드(Rothamsted)에서 Fisher에 의해 전개된 실험계획법에 관한 연구도 역시 그랬다. Fisher의 연구와 랜덤 실험에 관한 그의 아이디어는 표본추출법에 있어서도 매우 중요했다. 하지만 통계이론에 있어 가장 뛰어난 두 명의 공헌자인 Neymen과 Fisher사이의 큰 불화 때문에 애석하게도 과학분야로서 조사 표본추출법의 발달이 상당부분 지체되었을 것으로 추측된다.

1930년대와 1940년대에 오늘날에 사용되는 대부분의 기본적인 표본추출법들이 개발되었다. Fisher의 확률화 원리가 농업분야의 표본추출과 부차표본추출 연구에서 사용되고 검증되었다. Neymen은 신뢰구간이론, 집락표본추출법, 비추정법과 이중표본추출법을 제시하였다(9장 참조).

Neymen은 표본추출오차는 추정량의 분산을 계산하여 측정할 수 있다는 것을 보여 주었으며, 다른 오차 발생원인에 대해서는 특별히 언급하지 않았다. 다른 오차의 발생원인에 대한 추정방안을 정식으로 제시한 첫 번째 과학자는 인도의 통계학자인 Mahalanobis였다. 그는 농업자료를 수집하는 현장 조사원에 의해 발생할 수 있는 오차를 추정하는 방법을 개발했다. 그는 이 오차를 상호관입(*interpenetration*)이라는 방법으로 추정할 수 있었다. 이 방법은 조사원, 자료 입력자와 조사 감독자등이 초래하는 오차를 추정하는 방법으로 오늘날까지도 사용되고 있으며, 이는 각 조사담당자가 직접 관여한 관측결과들은 상당히 일관성을 유지하고, 이런 경향은 개인에 따라 차이가 있다는 점에 착안한 오차 계산 방법이다.

표본추출이론은 고전적인 통계학자들과 그들의 뒤를 잇는 Cochran, Yates, Hansen 등과 같은 통계학자들에 의해 개발되고 더욱 발전되었다. 1940년대에는 표본추출오차(*sampling error*)와 총조사오차(*total survey error*)는 동일한 것으로 해석될 수 없다는 것이 널리 받아 들여졌다. 예를 들어 현장조사원에 의해 발생하는 오차에 대한 Mahalanobis의 연구를 이미 언급했다. 1940년대에 미국 센서스국의 Hansen과 그의 동료들은 총조사오차의 측정을 위한 모형을 제안했다. 흔히 미국 센서스국 조사모형이라고 불리는 그 모형에서는 추정량의 총오차는 그 추정량의 평균 제곱오차로 측정한다. 그들이 사용한 모형에서는 분산을 추정하는 동시에 평균제곱

오차를 구성하는 편향의 발생요인을 추정하는 방법을 제시하고 있다. 이 모형은 표본추출분산은 단지 오차의 일부분에 해당하고 표본추출오차만을 토대로 조사오차의 추정하면 총조사오차를 과소추정하게 된다는 것을 명백하게 보여 주고 있다. 이 모형은 Hansen 등(1964)의 논문과 Bailar와 Dalenius(1969)에 의한 연구에 설명되어 있다.

수리 통계학자들은 자료에 있는 오차를 추정하고 그 오차를 조정하는 것에는 숙달되어있지만 일반적으로 조사연구에서 비표본추출오차(*nonsampling error*)를 제어하고 축소하거나 방지하는 것에는 익숙하지 않다. 비표본추출오차를 축소하기 위해서는 통계학, 사회학, 심리학, 언어학 등을 포함한 몇 개의 학문분야에 대한 지식과 이론을 결합한 세밀한 기획과 주의 깊은 조사설계를 필요로 한다. 많은 오차의 발생원인은 인식 또는 인식전달 체계와 관계가 있으므로 비표본추출오차를 설명하고 방지하기 위한 연구가 통계학이 아닌 다른 학문분야에서 많이 이루어지는 것은 전혀 놀라운 일은 아니다. [이런 학제간 연구들에 대한 것은 O'Muircheartaigh(1997)를 참조하라.]

표본추출이론의 초기 개발 단계에서는 비추정량과 관련된 “기술적인 편향(*technical bias*)” 같은 추정량 자체의 특성과 관련된 기술적인 문제 말고는 편향이 거의 관심대상이 되지 않았다(9장 참조). 초기의 통계학자들은 응답영향, 면접자와 응답자 사이의 접촉, 응답을 하거나 측정하는 일의 복잡성, 또한 변수의 현실성(즉 측정된 변수가 그것이 설명할 하려고 의도되었던 개념과 관계) 등과 관련된 모형들에는 특별히 관심이 없었다. 이런 사항에 대해서는 다른 분야에서 연구가 수행되었다. 예를 들어 Muscio(1917)는 질문에 사용된 용어에 따른 영향에 관한 초기 연구 결과를 제시하였다. 적절한 방법으로 태도를 나타내는 척도는 1920년에서 1950년 사이에 Likert를 포함한 다수의 학자들에 의해 개발되었다. 1940년대에는 다양한 설문지 설계의 장단점을 확인하기 위해 광범위한 학술적인 연구가 수행되었다.

O'Muircheartaigh는 또한 조사방법에 관한 논문에 나타난 면접자와 응답자의 역할에 관한 흥미 있는 사례들을 보여주고 있다. 면접자의 역할에 대한 초기의 시각은 단순히 정보는 면접자에 의해 수집이 가능하거나 또는 그렇지 않다는 것이었다. 수집이 가능하다면 그런 정보는 응답자로부터 쉽게 얻을 수 있다고 생각했다. 그래서 오래 전에는 응답자를 면접하는 기본적인 방법은 사실상 전적으로 대화에 의존하는 것이었다. 따라서 면접자가 대화할 때에 필기도 하지 않으면서 격식을 차리지

많은 대화를 나누고, 나중에 면접에서 얻은 정보를 요약하는 것이 하나의 면접을 통한 조사방식이었다. 한편 좀더 형식을 갖춘 면접 조사방식은 다른 형태는 미리 명시된 질문들을 가지고 면접자가 그것을 순서대로 물으면서 응답을 기록하는 것이었다.

응답에 대한 면접자의 영향은 원래는 관심사가 아니었다. 면접은 그 당시 사회조사를 위한 방법으로 통제와 표준화에 많은 어려움이 있어서 높게 평가를 받지는 못했다. 점차 표준화는 중요한 고려사항이 되었으며, 1942년에 Williams는 미국의 National Opinion Research Center(NORC)의 면접자들이 지켜야 할 기본적인 지침들을 제시했다. 1946년에 미국에서 개최된 학술회의에서 한 토론자는 이상적인 면접자를 “특별한 정치성향을 갖지 않고 조사와 관련된 지시사항을 이해하고 따를 수 있는 37세의 결혼한 여자”라고 주장했다. 적어도 여자라는 측면에서 면접자의 적합성에 대한 이런 주장은 오늘날까지도 조사기관에서 어느 정도 수용되고 있다.

초기의 문헌들에서는 응답자의 역할에 대해 별로 언급이 없었다. Cannell와 Kahn이 면접자와 응답자 사이의 상호반응을 코드화하는 방안을 발표한 1968년까지는 이런 상호반응에 대한 연구는 이루어지지 않았다. 사실상 응답자는 자료 수집과정에 있어 흔히 장애물로 간주되었고, 특히 법적으로 응답을 강제하는 조사를 실시하는 기관에서는 오늘날에도 이런 입장을 견지하는 경우가 있다. 이미 언급된 논문들 이외에도 Fienberg와 Tonur(1996), Converse(1986), Kish(1995), Hansen 등(1985) 그리고 Zarkovich(1996)와 같은 역사적인 논문들이 있다.

많은 새로운 연구 개발이 오늘날의 조사연구에 영향을 미쳤다. 예를 들어 오늘날에는 모형 기반(model-assisted) 표본추출이론이 폭 넓게 활용되고 있다. Sarndal등(1991)가 이런 새로운 표본추출이론을 정리한 대표적인 최신 교재이다. 또한 이제는 응답오차를 통계모형에 직접 반영하고, 면접 중의 인식과 관련된 문제들이 계속 연구되고 있다. 응답과정을 이해하려는 노력이 계속되고 있으며, 이런 새로운 이해를 통해 자료수집방법이 개선되고 있다. 또한 새로운 기술의 개발에 따라 컴퓨터를 이용하는 자료수집, 스캐닝 방법 또는 복잡한 언어로 표현된 응답을 자동으로 숫자 코드로 변환해 주는 소프트웨어의 사용 등이 실용화 되고 있다.

그렇지만 이런 끈임 없는 연구 노력에도 불구하고 조사연구와 관련된 근본적인 문제점들은 오늘날까지도 남아있다. 불가피한 조사오차의 발생, 오차에 대한 적절한 측정수단과 처리 방법의 미흡함, 오차가 추정값과 분석결과에 미치는 영향에 대한

조사 관련자(생산자, 사용자 및 의뢰자)의 이해 부족 등이 이런 근본적인 문제점들에 해당한다. 조사연구에서는 항상 끊임없는 품질의 향상이 요구된다.

조사 관련 역사에서 획기적인 사건들

- 조사설계에 대한 지침이 20세기 초반에 처음으로 개발되었다.
- “대표조사” 방법에 대한 Neyman의 역사적인 논문이 1934년에 출간되었다.
- 1940년대에 Mahalanobis가 조사 현장조사원에 의한 오차를 추정하기 위해 상호관입법을 개발했다
- 1960년대 초반에 Hansen 등에 의해 처음으로 조사모형이 개발되었다

1.4 품질혁명

지난 20년 동안 학자들이 사회적인 품질혁명(*quality revolution*)이라 부르는 많은 사건들이 있었다. Deming, Juran, Taguchi, Crosby, Ishikawa, Joiner 등이 품질 향상의 필요성과 방법에 대해 강조해왔다. 예를 들어 Deming(1986)은 14개의 논점과 7가지의 치명적인 병폐에 대해 발표했다. Juran과 Gryna(1980)는 품질의 지속적인 발전에 대해 논의했고, Taguchi(1986)는 변동을 강조하는 실험계획법을 개발하였다. Crosby는 문제점을 해결하는 것보다는 회피하는 것이 보다 바람직하다는 주장을 했고, Ishikawa(1986)는 7가지의 품질관리도구(자료수집, 히스토그램, 파레토도표, 생선뼈(fishbone) 도표, 층화, 그림 그리기와 관리도)를 제시하였다. 그리고 Joiner(Scholtes 등, 1994)는 삼각관계(품질, 과학적인 접근, 협력)를 강조했다. 그보다 훨씬 이전에 Shewhart는 관리도를, Dodge와 Romig(1944)는 합격판정추출법을 제안하고 이를 통해 통계적 공정관리(*Statistical process control*)방법의 개발을 시작했다.

오늘 날 여러 형태의 품질 산업이 이들 “선도 개발자들”에 의해서 시작되었다는 것은 의심의 여지가 없다. 이런 새로운 아이디어 또는 약간은 뒤떨어진 아이디어들이 지나치게 높이 평가됐다는 적절한 비평에도 불구하고(Brackstone, 1999; Scheuren, 2001), 이를 통해 품질관리를 위한 검사 규모의 축소, 담당자의 권한 강

화, 고객 중심의 사고 강화, 하향식의 경영과는 대조적인 팀워크 강화 등과 같은 바람직한 결과들이 얻어졌다.

1.5 품질의 정의와 통계기관에서의 품질

품질개선은 변화를 의미하고, 자동차 제조나 통계 생산의 과정과 같은 변화를 위한 과정이 있다. 성공한 기관들은 그들의 직무를 계속하기 위해서는 끊임없는 개선이 절대적으로 필요하다는 것을 알고 있으며, 개선을 위한 수단을 꾸준히 개발한다. 전형적으로 이런 기관들은 자신들의 핵심 가치, 즉 자신들이 건설적인 방향으로 변화하는 것을 도와줄 가치를 위한 몇 가지 전략을 채택했다.

조사기관은 끊임없는 개선의 필요성에 관해서는 다른 조직과 다르지 않다. 좋은 품질의 산출물을 필요로 하며, 사용자의 새로운 요구에 따라 과정을 조정 할 필요도 있다. 이런 의미에서 품질은 어떻게 정의되어야 하는가? 그것은 상당히 막연한 개념으로 다음과 같이 정의 될 수 있다. 가장 일반적이고 널리 인용되는 것이 Juran과 Gryna(1980)의 “사용을 위한 적합성”일 것이다. 그렇지만 어디에 사용되는가에 따라 (통계 결과의 경우) 사용자들이 다양한 형태의 품질을 나타내는 특성을 요구하기 때문에 사용을 위한 적합성의 개념은 매우 복잡해진다.

품질(*quality*)은 간단히 “사용을 위한 적합성”으로 정의할 수 있다. 조사의 관점에서 보면 조사 자료가 의도된 목적이 충족될 만큼 정확하고, 필요할 때에 이용할 수 있어야 하고, 그 조사결과를 필요로 하는 사람들이 얻기 쉬워야 한다. 따라서 정확성(*accuracy*), 적시성(*timeliness*), 접근성(*accessibility*)이 조사품질의 세 가지 요소이다.

또 하나의 정의는 설계의 품질과 적합의 품질을 구별하는 것이다(Juran 과 Gryna, 1980). 설계 품질(*design quality*)은 예를 들어 자료의 표현방식과 같은 것들이다. 통계자료를 표현하는 단색의 도표나 단순한 표보다 다채로운 색의 도표가 포함된 출판물이 훨씬 미적으로 나올 수 있다. 그러므로 설계품질이 후자의 경우 더

높다고 한다. 일반적으로 설계품질은 비용을 증가시킨다. 반면에 적합품질(*quality conformance*)은 생산물이 의도된 용도를 충족시키는 정도이다. 조사에서 적합성은 모수의 추정량에 대한 미리 정해진 오차범위 같은 것이다. 설계품질과 적합품질의 차이가 항상 뚜렷한 것은 아니다.

통계결과의 품질은 다차원적인 다시 말해 여러 가지 요소로 구성된 개념으로 설명된다. 자료의 품질은 정확성, 적시성, 충분한 내용, 접근성, 비밀 보호수준 등의 요소를 포함한다. 본장의 후반에서 공식통계에서 사용되는 품질을 위한 전반적인 틀에 대한 예를 볼 것이다. 전통적으로 조사품질은 조사오차(자료의 정확성)에 의해 결정된다고 강조되어 왔다. 하지만 다른 기업체들처럼 조사기관들도 품질에 대한 훨씬 폭넓은 정의를 연구 대상으로 할 필요가 있어졌다. 왜냐하면 사용자들은 단지 추정의 정확도에만 관심이 있는 것은 아니고, 적절하고 적시성과 일관성이 있으면서 얻기 쉽고 서로 비교할 수 있는 자료를 요구하기 때문이다.

어떤 사람들은 정확성이 최우선이라고 주장한다. 정확성 없이는 다른 측면에서의 품질은 비현실적일 수 있다. 그렇지만 그 반대일 수도 있다. 매우 정확한 자료가 만일 너무 늦게 발표되거나 적절하지 않으면 아무런 소용이 없다. 정확성 관점에서 품질의 개념과 관련된 몇 가지 문제점으로 인해 지난 10년 동안의 발전은 통계기관들의 변화를 촉진하기 시작했다.

1. 정확도의 측정은 어렵고 많은 비용을 필요로 하므로 대부분의 조사에서는 거의 시행되지 않는다. 정확도는 보통 총조사오차로 설명되지만 어떤 오차원인은 측정이 불가능하다. 따라서 바람직한 성질을 갖는 제품의 생산이 가능한 신뢰할 수 있는 과정을 수행함으로써 품질을 보장해야 한다. 기본적으로 제품의 품질은 공정상의 품질을 통해 이루어진다.
2. 사후조사에서 총조사오차의 측정은 비교적 품질개선에 미치는 영향이 제한적이다. 반복조사를 제외하고는 정확도의 추정은 품질의 개선에 거의 영향을 주지 못한다.
3. 조사 운영 과정에서의 코딩과 입력과 같은 기계적인 품질관리를 통하여 지속적인 품질개선이 이루어지기는 어렵다. 그 보다는 피드백과 학습 등을 통한 품질개선과정에 조사 종사자들이 자발적으로 참여함으로써 보완이 이루어져야 한다.
4. 정확도를 추정하는 것에 대한 지나친 집중은 설계 품질의 발전에 크게 기여할

수 없다.

이삼십년 전에 사용자는 조사과정의 주요 참여자가 아니었다. 대부분의 통계기관은 조사결과의 뚜렷한 사용자(예를 들어, 조사 의뢰자)가 있지 않으면 사용자와의 관계를 중요하게 여기지 않았다. 그 당시의 기술로는 자료의 신속한 공개와 올바른 발표가 어려웠다. 센서스 자료가 수집된 몇 년 후에야 공개되는 경우가 드물지 않았고, 사용자와 생산자 모두는 이와 같이 조사과정을 수행하는 데 많은 시간이 걸린다는 것을 받아들였다. 이는 아이러니컬하게도 조사결과에 대한 평가연구를 할 충분한 시간을 확보할 수 있다는 것을 의미한다. 이와 같이 정확성에 대한 추정결과를 조사 자료의 공개와 거의 동시에 발표하는 것이 가능했던 평가 연구 사례들이 있다. 또한 많은 기관들은 품질관리와 특정 조사과정의 정확성에 대한 연구를 할 수 있는 비교적 충분한 예산을 확보하고 있었다. 그러므로 만일 원하는 적시성과 접근성을 확보하는 것이 거의 불가능하다면 생산자 입장에서는 자료의 정확성에 노력을 집중하는 것이 당연했다.

오늘날에는 상황이 바뀌었다. 많은 국가통계기관의 경우 예산이 삭감되었고, 조사 시장에는 예전보다 훨씬 많은 관계자들이 있다. 동시에 기술의 발달이 보다 바람직한 설계품질 요소를 확보할 수 있는 여건을 제공하고 있다. 자료분석과 데이터베이스에서의 정보 취합 같은 부가가치적인 행위나 자료수집 같은 전반적인 자료처리과정이 오늘날 매우 신속하게 이루어진다. 통계기관들은 완벽한 완성품을 적시에 적절한 방식으로 발표하지 못하면 경쟁자에게 자신의 일을 넘겨줄 각오를 해야 한다.

이런 새로운 상황으로 인하여 전 세계의 많은 통계기관들이 그들의 연구를 개선하기 위하여 품질경영 모형, 직무탁월성 모형, 사용자 동향, 감사, 자체평가와 관련된 작업을 하고 있다. 이런 흐름에 뒤떨어지면 파산할 위험이 있으며, 국가통계기관 까지도 이런 위험에 노출되어 있다. 예를 들면 생산결과에 대해 대량으로 이루어지는 검사와 검증 또는 사후평가 대신에 생산과정 중의 공정관리로 관심이 옮겨지고 있다. 이런 움직임은 제품품질은 과정상의 품질을 통해 이루어진다는 믿음 때문이다. 조사 연구의 이런 과정의 중요성에 대한 시각(*process view*)은 조사기관에서 수행하는 거의 모든 과정에 확대 적용된다. 왜냐하면 조사연구를 뒷받침하는 많은 과정들이 통계 제품의 품질에 영향을 미치기 때문이다. 이런 과정들에는 훈련, 사용자 접촉, 제안서 작성, 벤치마킹, 계획안 작성, 자료 제공자와의 접촉 및 전략 수립 등

이 포함된다.

몇 개의 통계기관은 그들이 새로운 품질 요구에 따라 어떻게 일을 해야 하는가에 대한 지침을 문서로 제시하고 있다. 호주 통계국, 뉴질랜드 통계청, 네덜란드 통계청, 덴마크 통계청, 스웨덴 통계청, 영국 국립통계국, 미국 센서스국, 미국 노동통계국은 직무계획, 전략수립 또는 규정에 대한 지침서를 제시하고 있는 국가통계기관들이다. 예를 들어 뉴질랜드 통계청은 공식적 통계의 생산과 공개를 위한 몇 가지 실행 규약을 만들었다. 이 규약들을 정리하면 다음과 같다.

공식통계의 생산과 공개를 위한 뉴질랜드 통계청의 실행 규약

1. 조사의 필요성은 반드시 정당화되어야 하고 이 필요성은 자료를 수집하는 비용과 응답자 부담보다 커야 한다.
2. 많은 조사단계에 대한 예정표, 예산안과 품질 기준에 대한 계획과 함께 명백한 조사목표와 이와 관련된 품질기준이 개발되어야 한다.
3. 자료의 수집, 보안성, 사생활 보호, 자료의 공개를 규정하는 입법상의 책임이 반드시 따라야 한다.
4. 적절한 통계 방법론에 의해 조사의 설계가 이루어져야 한다.
5. 다른 출처에서 얻어진 자료와의 통합이 가능하고 개발비용을 줄이기 위해 표준화된 틀, 질문, 분류가 사용되어야 한다.
6. 조사표는 응답자가 정확하게 완성하기 쉽고 효율적으로 처리될 수 있도록 설계되어야 한다.
7. 응답자의 보고 부담이 가능한 최소화 되어야 한다.
8. 수집 자료를 분석하고 결과를 발표하는 데 있어서 객관성과 전문성이 유지되어야 하고 자료는 이해하기 쉬운 방법으로 공정하게 공개되어야 한다.
9. 수집 자료에 대한 주요 결과는 쉽게 이용할 수 있어야 하고 모든 사용자들에게 공평하게 이용할 수 있는 기회를 제공해야 한다.

또한 이와 유사한 많은 문서들이 있다. 예를 들어 UN이 공식통계를 위한 10가지 기본원칙(UN, 1994a)을 편찬했고, Frabchet(1999)는 국제 통계기관들을 위한 실행 지침을 제시하였다. 스웨덴 통계청은 품질 정책에서 객관성, 접근성과 전문성을 강

조하고 있다. 캐나다 통계청에는 몇 개의 정책 문서가 있는 데, 하나는 품질 지침에 대한 것이고 또 다른 하나는 협력 계획과 프로그램 감시 체계에 대한 것이다 (Fellegi와 Brackstone, 1999). 덴마크 통계청(2000)은 2000년에서 2005년까지를 위한 전략을 발표했다. 그 문서에는 기관의 공식적인 지위, 주요 목적과 전략, 그리고 일반 국민 및 직원과 기관의 관계를 명시하고 있다. 장기 전략계획서에서 미국 센서스국(1996)은 기관의 목적을 성취하기 위한 주요 전략을 정의하고 있다. 이 문서에는 또한 핵심 직무, 핵심 직원의 능력과 함께 목표 고객을 명시되어 있다. 미국노동통계국, 미국국립농업통계서비스국과 다른 미국 연방 통계기관들도 비슷한 문서를 갖고 있으며, 이들 문서들은 대부분 그 기관의 웹사이트에서 찾아볼 수 있다.

여기서 중요한 점은 통계기관에서 품질의 개념이 지난 10년 동안 많이 변화하였다는 것이다. 품질에 대한 오늘날의 지배적인 접근 방법은 1986년의 ISO8402 표준에 명시된 바와 같이 “진술되거나 암시된 요구를 만족시키는 능력에 영향을 미치는 생산물이나 서비스의 총체적인 특징과 특성”이라고 할 수 있다. 국제 표준화 기구인 ISO는 소재, 생산물, 과정이 목표에 부합되도록 기술적인 세부사항과 규칙, 지침, 특성의 정의 등에 대한 일관성 있는 정확한 기준을 제시하는 협정서를 개발하는 기구이다. ISO는 품질에 크게 공헌하고 있다. 따라서 통계기관에서도 정확성이 더 이상 품질을 나타내는 유일한 척도가 될 수 없다는 점에 주목해야 한다. 품질은 사용자의 요구를 반영하는 몇 개의 요소로 이루어져 있다. 이런 상황에서 품질은 이런 요소를 통해 설명할 수 있고 정확성은 이 중 하나의 요소일 뿐이다. <표 1.2>에서 보듯이 Eurostat의 품질 개념은 일곱 개의 요소로 이루어져 있다.

다른 기관들은 조금 다른 품질 요소를 사용한다. 캐나다 통계청(Brackstone, 1999)은 여섯 가지의 요소(적절성, 정확도, 적시성, 접근성, 해석가능성, 일관성)를 사용하고 스웨덴 통계청(Rosen과 Elvers, 1999)은 다섯 가지 요소(내용, 정확성, 적시성, 비교가능성/일관성, 가용성/명확성)를 사용한다. 각각의 요소는 몇 개의 부요소로 세분화 될 수 있다.

또 하나의 중요한 관심은 비용과 품질의 관련성이다. 조사비용 그 자체는 품질 요소가 아니지만 대안으로 다른 설계를 고려할 때에는 중요한 역할을 한다. 우리는 품질 요소(즉, 원하는 수준의 정확성, 일정, 완성도)와 관련된 제약조건하에서 가장 비용이 적게 드는 설계를 선택해야 한다. 또는 주어진 조사 예산으로 최적의 품질 요소들을 최대한 만족시키는 최상의 설계를 선택해야 한다. 그러므로 비용은 조사

설계와 관련된 효율성 판정기준을 구성하는 성분이다.

<표 1.2> Eurostat의 품질 요소

품질요소	의견
1. 통계개념의 적절성 (<i>relevance</i>)	통계결과가 사용자의 요구를 만족시키는 것이 적절성이다. 그러므로 사용자의 요구는 처음부터 정립되어야 한다.
2. 추정의 정확성 (<i>accuracy</i>)	정확성은 추정값과 모수의 참값과의 차이다. 비용상 또는 방법상의 한계 때문에 정확성을 평가하는 것은 항상 가능한 것은 아니다.
3. 결과 보급의 적시성 (<i>timeliness</i>)과 정시성 (<i>punctuality</i>)	경험상 이것이 가장 중요한 사용자의 요구사항 중 하나일 것이다. 왜냐하면 이런 요소가 결과의 효율적인 사용과 매우 관련이 깊기 때문이다.
4. 정보의 접근성 (<i>accessibility</i>)과 명확성(<i>clarity</i>)	쉽게 얻을 수 있고 사용자가 원하는 형태로 이용할 수 있으면 조사결과는 매우 높게 평가될 수 있다. 또한 자료 제공자는 사용자가 결과를 해석하는 데 도움을 주어야 한다.
5. 비교가능성 (<i>comparability</i>)	공간과 시간에 따른 비교는 필수적이다. 최근 국가간 비교의 필요성이 증대되고 있다. 앞으로는 문화적 차이를 조정할 수 있는 방법의 개발에 대한 새로운 요구가 있을 것이다.
6. 일관성(<i>coherence</i>)	기본적인 개념은 더 복잡한 것으로 결합될 수 있다는 점에서 출처가 동일한 통계는 일관성을 갖는다. 출처가 다른 경우, 특히 다른 조사시점을 기준으로 하는 통계들의 경우 정의, 분류방식과 방법상의 기준이 동일하다면 일관성을 갖는다고 할 수 있다.
7. 완전성(<i>completeness</i>)	사용자 집단의 요구와 우선순위를 반영하여 결정된 영역에 대해 활용 가능한 통계를 제공해야 한다.

출처 : Eurostat(2000)

Fellegi(1996)와 De Vries(2001) 같은 통계체계의 특성에 관한 연구들이 있다. 최근에 품질 분야의 선도 그룹은 유럽통계체계(*European Statistical System*)의 개선에 대한 연구보고를 발표했다(Lynberg 등, 2001).

이미 언급한 것과는 별도로 자료의 품질을 평가하는 몇 개의 틀이 있다. 국제통

화기금(Carson, 2000)이 개발한 것을 예로 들면 다음과 같다.

- 품질은 정확성을 포함한 몇 개의 요소에 의해 정의될 수 있다.
- 자료의 품질은 자료 생산과정의 품질관리를 통해 이루어진다.
- 자료 생산과정상의 품질은 기관에 적합한 체계와 절차에 의해 결정된다.

1.6 품질의 측정

품질을 정의하는 틀이 일단 정의되면 품질을 측정하는 것이 중요하다. 만일 우리가 조사항질을 주요소와 부요소로 정의해 보면 품질의 어떤 요소는 양적이고 다른 것들은 질적인 다차원적인 개념으로 품질이 설명된다. 정확도는 양적이고 다른 요소들은 대부분 질적이다. 전체 조사항질 척도(모든 요소를 고려한 하나의 품질 척도)가 산출된 사례는 찾아 볼 수 없다. 대신에 각 요소와 관련된 정보가 제공되면 이에 대한 품질보고서(*quality reports*)나 품질고지서(*quality declarations*)를 작성한다. 품질보고서에는 사용자의 인식과 만족, 표본추출오차와 비표본추출오차, 주요 생산 일자, 보급형태, 증빙문서의 가용성과 내용, 예비결과와 최종 결과의 차이, 연간 결과와 단기 결과, 그리고 연간통계와 센서스를 기반으로 한 품질에 대한 설명과 평가를 하는 것이 이상적이다.

여러 나라에서 표준품질보고서에 대한 작업이 진행되고 있다. 프랑스 공식통계를 위한 사업체조사보고서의 개발, 공식통계를 위한 모든 조사는 품질고지서를 곁들여야 한다고 명시한 스웨덴의 결정, 그리고 미국의 몇몇의 조사나 조사체계가 품질분석서를 제공하는 것들이 이런 사례에 해당한다(8장 참조). 품질분석서(*quality profile*)는 조사의 품질이나 조사의 체계에 대해서 알고 있는 모든 사항들을 종합한 것이다. 몇 가지의 예를 들어보면, SIPP(Survey of Income and Program Participation), AHS(Annual Housing Survey), 그리고 SSS(Schools and Staffing Surveys)의 경우 이런 품질분석서가 개발되어 있다. 품질분석서의 문제점은 그것이 적시에 발간될 수 없다는 것이다. 왜냐하면 이를 위해서는 품질 관련 연구결과를

모아 편집하는 것이 필요한 데, 이미 언급했듯이 이런 조사후 수행되는 작업들이 시간이 많이 걸리기 때문이다. 품질분석서, 품질고지서와 품질보고서에 대해서는 8장과 10장에서 더 자세하게 논의될 것이다.

많은 조사기관들은 이제 품질을 측정하는 새로운 접근방법을 채택하고 있다. 이런 접근방법은 개선을 위한 기준을 세우기 위해 기관의 업무성과를 평가하는 것이 특징이다. 이를 달성하기 위한 방법은 여러 가지가 있다. 그 중 한 가지 방법은 품질경영법과 전체품질경영(*total quality management; TQM*)이라는 원리를 근거로 한 직무탁월성모형을 사용한 성과 평가다. TQM은 한 기관의 확고한 핵심적인 가치를 토대로 연구하고 직무를 개발하는 방법이다. 이런 가치에는 고객지향, 모든 직원의 참여와 지도력, 과정적응, 과정변화의 측정과 이해 그리고 지속적인 개선 등이 일반적으로 포함된다.

TQM 그 자체는 실제 수행에 대한 지침을 제공하지는 않지만 결과적으로 더 구체적인 직무탁월성모형의 개발을 가능케 했다. Swedish Quality Award Guidelines, Malcolm Baldrige National Quality Award, European Foundation for Quality Management(EFQM) 모형이 이런 사례들이다. 이런 모든 모형들은 기관이 모형 지침에 실린 기준과 비교하여 자신을 평가함으로써 개발 된다. 예를 들어, Malcolm Baldrige Award는 지도력, 전략계획, 고객과 시장 중심, 정보와 분석, 인적자원 중심, 공정경영, 그리고 직무결과라는 기준을 가지고 있다. 이 모형에 따른 기준 만족에 대한 기관의 평가는 전문적인 외부 조사관의 도움이 바람직하기는 하나 본질적으로 자체적으로 수행되어야 한다.

Baldrige Award에서는 각각의 기준에 대해 해당 기관은 다음 세 가지 기본적인 질문에 답하도록 한다.

1. 구체적으로 기준을 검토하기 위해 무엇을 했습니까?
2. 기관 전체에서 이런 방법을 어느 정도 사용해 왔습니까?
3. 이 방법들이 어떻게 평가되고 계속적으로 개선됩니까?

이런 것들은 의미가 없는 질문들이라고 생각할지도 모르지만 그렇지 않다. 대표적인 시나리오는 모든 기관들이 품질을 향상시키기 위해 뭔가의 방법을 수행했지만 그것이 기관 전체에서 일률적으로 적용되지 않고 좀처럼 평가되지 못하는 것이다. 사실 많은 단체들은 개선할 때가 되면 임시적이고 국부적인 방법을 쓴다. 좋은 과

정이 기관 전체에서 항상 사용되는 것은 아니며, 성공한 과정이 자동적으로 기관에 퍼지는 것도 아니다. 따라서 이미 언급했듯이 변화 과정이 있어야 한다. 앞에서 기술한 직무모형의 기준에 대한 좋은 평가가 통계결과의 품질과 관계가 있기 때문에 다른 기관과 마찬가지로 통계기관에서도 이런 평가가 도움이 된다.

또 다른 유용한 평가 도구가 있다. 그것은 ISI 인증인데, 인증을 받기를 원하는 기관은 품질업무 조직, 권한의 분산, 처리 절차, 공정지시, 명세서와 계획의 검증에 대한 지침서를 제시해야 한다. 몇 개의 통계기업체들을 포함한 세계의 많은 단체들이 인증을 받았다. 몇몇 나라와 사업 분야에서는 계속 사업을 하고 싶어 하는 단체에게 인증은 필수적이다.

균형채점표는 기업의 네 가지 다른 직무영역(고객, 학습, 재정과 공정) 사이의 균형을 강조하는 또 하나의 도구이다(Kaplan과 Norton, 1996). 예를 들어, 핀란드 통계청이 이 도구를 사용하기 시작했다. 이런 채점표가 개발된 이유 중 하나는 아주 많은 기관들이 재정 부분에만 너무 많은 가중치를 두어서 다른 세 요소가 과소평가되거나 자주 무시되기 때문이다.

직무과정 리엔지니어링(*business process reengineering*)은 직무과정의 개선을 위한 완전히 다른 접근 방식이다(Hammer와 Champy, 1995). 이는 근본적으로 다시 시작해서 밑바닥에서부터 끝까지 다시 과정을 만드는 것을 의미한다. 기관을 리엔지니어링 하는 것은 낡은 체계를 버리고 새롭고 개선된 것들로 대체하는 것을 의미한다. 그것은 비용, 품질, 서비스 및 속도 같은 중요한 성과 척도의 극적인 개선을 목표로 근본적으로 다시 생각하고 업무과정을 급진적으로 재설계하는 과정을 요구한다. 근본적인(*fundamental*), 급진적인(*radical*)과 극적인(*dramatic*)이라는 단어를 포함한 것을 주목해라. 이것은 이 방법이 끊임없는 개선과 점진적인 변화를 추구하는 방법과는 매우 다르다는 것을 의미한다.

어떤 통계기관들은 최근의 직원 분위기조사, 고객조사, 간단한 점검표와 내부 품질감사를 시행하기 시작했다. 이런 방법들은 직원의 동기유발, 사기, 전문성을 정기적으로 평가하는 것의 중요성을 인정하는 것이다. 예를 들어 영국에서는 국립통계국(*Office of National Statistics; ONS*)이 직무, 직속상사, 전체 조직, 내부 의사소통, 교육과 개발에 대한 직원의 인식과 태도에 대한 정보를 얻기 위한 설문지를 개발했다. 스웨덴 통계청, 핀란드 통계청, Eurostat이 직원들의 분위기조사를 하고 있는 또 다른 기관들이다.

고객조사는 고객의 요구와 조사 기관의 과거 수행한 업무에 대한 평가결과를 제공하는 중요한 도구이다. 이는 고객들이 어떤 것들을 정말 중요하다고 생각하고, 그 기관이 제공하는 생산물과 서비스의 품질에 대한 고객들의 인식을 알아보는 데 사용될 수 있다. 다른 종류의 질문은 그 기관의 이미지가 다른 단체들의 이미지와 어떻게 비교가 되는가에 관한 것이다. Morganstein과 Marker(1997)가 지적했듯이 많은 고객 만족도 조사는 방법론에 문제를 가지고 있다. 예를 들어 많은 응답자들이 대부분 같은 수치(예를 들어 “매우 만족함”)을 고르는 결과가 나오게 하는 한정된 3점 또는 5점 척도를 사용한다. 많은 경우에 응답선택항목이 양극(매우 만족함과 매우 불만족스러움)으로 구성되며, 그 결과 중간의 구분이 응답자에게 불명확하다.

만족이라는 개념과 이것이 질문에 어떻게 표현되어야 하는지는 자주 발생하는 문제점이다. 정보가 충분하지 못하고 오해할 소지가 있는 응답 결과를 가지고 사용자나 고객 단체에서 올바른 해석을 하는 것은 쉽지 않다. 조사방법론에 대해 정식교육을 받지 않은 사람들에 의해 개발된 수많은 고객 만족도 조사(호텔, 항공회사 등)에서는 무응답률이 높고 열의가 없는 반응이 나타난다. 이런 분야는 전문적인 조사기관들이 앞장서서 통찰력이 있는 새로운 방법을 개발해야 할 분야이다.

다른 종류의 자체 평가는 조사 책임자가 기입할 수 있는 간단한 품질점검표를 사용하는 것이다. 뉴질랜드 통계청에서 이런 사례를 찾을 수 있다. 점검표에는 표시를 하고 의견을 적을 수도 있다. 조사 책임자는 각 질문에 “예” 또는 “아니오”라고 답하고 그 답을 상세하게 설명할 수 있다. 점검표에 나오는 항목의 예가 <그림 1.1>에 있다. 이런 종류의 점검표에는 언제(when)와 어떻게(how) 같은 주요 단어들이 들어있는 추적질문들을 추가하여 사용할 수도 있다.

마지막으로 자체 평가방법이나 외부 또는 내부에서 실시하는 감사가 있다. 외부 감사에서는 전문가가 과정, 조사 또는 기관의 일부나 전체를 평가하게 된다. 전문가들은 일반적으로 실제조사를 그들이 잘 알고 있는 높은 품질의 비슷한 조사와 비교를 한다. 만일 감사에서 기관의 업무성과가 대상이라면 앞에서 언급한 직무탁월성 모형 중 하나를 사용할 수 있다. 감사는 보통 업무개선을 위한 몇 가지 권장사항을 제시한다. 바람직한 업무처리 사례는 그 단체의 다른 부서에 전달된다.

내부감사는 그 기관의 직원이 실행한다. 모든 감사는 결과와 과정에 대한 내부 문서, 조직의 지침과 방침, 그리고 감사관의 관찰을 근거로 한다. 통계기관에서의 감사는 점차 증가하고 있다. 예를 들어 네덜란드 통계청은 내부 정규직 감사에 의

한 감사 체계를 가지고 있다. 스웨덴 통계청은 모든 조사가 5년 내에 최소한 한번 감사를 받는 5개년 프로그램을 최근 시작했다.

뉴질랜드 통계청의 품질점검표

- 주 사용자가 누구인지 그리고 새로운 감시자에 대한 충분한 이해가 있다.
- 설문지, 정의, 분류방식은 최신의 요구와 상황을 반영한다.
- 과정과 결과에 대한 문서가 완전하고 접근 가능하다.
- 전문적인 학회와 국제적인 발달에 기여한다.
- 자료 정의는 다른 조사 프로그램과 일치한다.
- 표본은 정기적으로 재설계한다.
- 장기 시계열을 활용할 수 있다.
- 계절 조정 분석을 수행한다.
- 출처와 방법에 대한 문서를 이용할 수 있다.
- 발표 날짜를 미리 공시한다.
- 정보는 사용자가 요구하는 형태와 매체로 이용할 수 있다.
- 요구에 부응하는 데 걸리는 시간에 대한 기준을 만족시킨다.
- 자료 수집의 용도에 대한 정보를 응답자에게 제공한다.
- 기밀성을 검토하고 공표한다.
- 품질 지침을 정기적으로 측정하고 감시한다.
- 자료는 보관소에 보관한다.
- 통계법령(공식통계 결과를 위한 법률 틀)의 요구조건을 만족한다.

<그림 1.1> 조사 단체를 위한 품질점검표의 예. [출처: 뉴질랜드 통계청]

품질의 측정과 보고 방법

- 품질 보고서를 표준들에 따라 개발한다.
- 품질 분석서를 개발하고 사용한다.
- 조직의 성과를 “탁월성모형”에 따라 평가한다.
- 직원 분위기조사를 실시한다.
- 고객 조사를 실시한다.
- 내부와 외부감사를 실시한다.

1.7 품질의 개선

1.6절에서 설명한 방법들은 개선이 필요한 분야를 확인할 수 있는 방법과 측정에 관한 예들이다. 품질문제는 때로는 쉽게 해결할 수 있다. 그것은 단지 어떤 필요조건을 더 잘 만족하도록 과정을 약간 바꾸는 문제다. 하지만 때로는 체계적인 개선 노력이나 프로젝트를 필요로 하는 광범위한 개선책이 요구되기도 하다. 개선은 제대로 역할을 하지 못하는 어떤 과정에 관한 것이다. 많은 국가통계기관이 생각하는 방안은 품질경영도구를 활용하는 특수한 팀을 창설하는 것이다. (예를 들어, 앞에서 언급한 Ishikawa(1982) 도구들)

품질개선 프로젝트 목표

- 직업 코딩의 품질과 효율을 강화한다.
- 토지사용 통계를 위한 과정을 능률적으로 한다.
- 에너지 통계의 에디팅 작업을 개선한다.
- 농장 등기부의 자료 수집을 단순화 한다.
- 면접 작업의 품질을 보증한다.
- 사용자 접촉의 질을 개선한다.
- 직원모집 과정을 개선한다.

<그림 1.2> 스웨덴 통계청의 품질개선 프로젝트 목표 사례

일반적으로 프로젝트팀에는 예를 들어 Scholtes 등(1994)에 나오는 것들과 같이 문서에 의해 충분히 입증되고 승인된 작업방침에 따라 도움을 주는 품질 담당자가 있을 것이다. 모든 참가자의 참여가 가능하도록 도구들은 되도록 간단한 것이 바람직하지만 경우에 따라 실험계획 같이 복잡한 통계 도구가 필요할 수도 있다. 스웨덴 통계청은 1993년에 품질개선에 대한 체계적인 연구를 시작한 이후로 이런 개선 프로젝트를 100회 넘게 실시했다. 이들 프로젝트에서 세웠던 목표들은 <그림 1.2>

에서 볼 수 있다. 이런 프로젝트에서 모든 직원이 그 팀을 대표한다는 것에 많은 가치를 두는 것이 필요하다. 과정 각 단계에서 일을 하는 사람들은 자신의 직무 개선에 책임을 져야 한다. 마찬가지로 만일 어떤 사람이 업무 개선을 원한다면 그는 개선을 위한 변화를 수용하는 데 보다 자발적으로 참여하는 것이 필요하다.

어떤 과정은 한 기관의 여러 부서에서 공통적일 수 있다. 이런 과정에는 설문지 개발, 코딩, 에디팅, 무응답 축소와 조정, 채용, 직원직무평가, 자료수집 등이 있다. 이런 공통된 과정이 한 단체내에서도 부서에 따라 매우 다른 방법으로 수행되는 경우를 접하는 것이 어렵지 않다. 접근방법의 차이는 일반적으로 최종결과의 특성상의 차이로 나타날 것이며, 모든 접근방법이 똑같이 효율적이지는 않을 것이다. 최고의 전략은 과정을 표준화해서 불필요한 차이를 제거하는 것이다. 현행최상방법(*current best method*; CBM)에 따른 접근방법이 바로 이런 것들 중 하나이다.

CBM을 개발하는 과정은 Morganstein과 Marker(1997)에 설명되어 있다. 적절한 업무수행이 이루어지는지 여부를 확인하기 위해서는 어떤 과정의 내부검사를 실시하는 팀을 선임해야 한다. 또한 이런 업무 수행절차들을 다른 단체의 것들과 비교하는데 이런 접근방법을 벤치마킹(*benchmarking*)이라고 한다. 그 다음에는 그 팀이 CBM 초안을 개발하고 이것은 더 많은 직원들에 의해 검토된다. 이런 검토과정엔 얻어진 의견과 제안들을 모아 CBM을 수정한다. 한번 받아들여지면 CBM은 수행되고 이것의 성과에 관한 자료가 수집된다. CBM은 전형적으로 매 4년 정도마다 개정한다. 스웨덴 통계청에서는 편집, 무응답 축소, 무응답 조정, 기밀성 보호, 설문지 설계와 프로젝트 수행에 대한 CBM이 개발되었다(10장 참조). 적절한 CBM을 가지고 있으면 이에 따라 관련 문서들을 이용할 수 있기 때문에 과정의 개선이 훨씬 쉬워진다. 또한 신규 직원들이 그들의 직무를 좀더 빨리 수행할 수 있도록 신규 직원 교육을 일관성 있게 할 수 있다.

최소 기준의 확립, 품질 지침과 권장 수행방법들과 같은 과정들의 표준화를 겨냥한 몇 가지 대책이 있다. 최소 기준은 기본적인 체면을 유지하는 수준을 확보해야 하고, 품질 지침은 어떻게 하는가 보다는 오히려 무엇을 하는가를 유도해야 하며, 권장된 수행방법들은 이들 중 선택해 사용할 수 있는 다양한 수행절차들을 제공해야 한다. 이런 도구들에 대한 보다 심도 있는 논의는 Colledge와 March(1997)에서 찾을 수 있다.

1.8 품질에 대한 요약

이 장에서 우리는 품질의 개념에 대해 알아보았고 품질은 여러 가지 요소로 구성된 다차원적인 개념이라는 것을 알게 되었다. 그것을 구성하는 한 요소는 총 조사 오차를 측정하는 정확성이다. 다른 요소들은 기관에 따라 다르게 부른다. 이 책에서는 품질향상을 지향하는 방법으로 조사 자료의 정확성의 관리문제를 다룬다.

오차를 방지하고 지속적인 품질 개선을 이루기 위해서는 과정을 제대로 볼 수 있어야 한다. 정확한 자료는 그 자료를 산출하는 정확한 과정을 통해 얻어진다(즉, 자료의 품질은 생산과정상의 품질관리를 통해 얻어진다). 부정확성은 수행된 과정의 결함에 의해 발생하므로 자료의 정확성 같은 조사 결과의 특성에 가장 큰 영향이 있는 중요한 생산과정과 관련된 변수들을 관리하는 것이 중요하다.

이 책의 각 장에는 조사 과정을 어떻게 통제하고 개선할 수 있는가에 대한 많은 예들이 나온다. 단지 좋은 조사방법을 사용하는 것뿐만 아니고 모든 직원이 개선 업무에 참여하게 하고, 집약된 의견 중에서 가장 좋은 의견은 채택하는 것이 중요하다. 우리가 높이 평가하는 몇몇의 도구는 팀에 의한 접근 없이는 실행될 수 없다. CBM들의 개발은 과정에 대한 실용적인 지식과 경험이 그 과정의 참된 개선을 야기할 도구를 생산하는 데 필수적이라는 것을 보여주는 하나의 예이다.

정말 많은 품질 요소가 있는데 왜 우리는 자료의 정확성에만 초점을 맞춰야 하는가? 이에 대한 해답은 정확성은 품질을 토대로 하고 정확성 없는 조사 자료는 별로 쓸모가 없다는 것에서 찾을 수 있다. 만일 자료가 잘못되었으면 적합성, 적시성, 접근성, 비교 가능성, 일관성과 완전성이 충족 되어도 별로 도움이 안 된다. 게다가 이 모든 다른 품질 요소가 중요하지만 우리는 이것들을 최적화해야할 속성으로 보기보다는 오히려 과정의 제약조건으로 본다. 예를 들어 우리는 조사를 완성하는 시간을 최소한으로 단축해야 하는 것이 아니라, 주어진 날짜에 그 자료를 이용할 수 있도록 해야 한다. 그런 의미에서 비용이 제약조건이듯이 적시성 또한 제약조건이다. 따라서 비용과 적시성에 대한 제약조건을 만족하면서 가능한 정확한 자료를 제공하는 것이 우리의 목표이다. 정확성이 아닌 모든 품질 요소는 같은 시각으로 해석할 수 있다. 그러므로 이 책은 정확성을 최대화하는 조사를 설계하는 것에 대한 것이다.

조사품질에 관련되는 많은 문헌이 있다. 그러나 불행하게도 알려진 모든 조사오차의 발생원인을 다룬 조사방법에 대한 교재는 없다. 그 중 일부를 다룬 것으로 Andson 등(1979), Groves(1989), Lessler와 Kalsbeek(1992)의 교재가 있다. 많은 대부분의 조사방법론 교재들은 표본추출이론은 상세하게 다루고 있지만 비표본추출오차는 한 두장 정도에서 간단히 다룬다. 설문지 설계, 조사면접과 무응답 같은 특수한 주제를 다룬 책들도 있다. 지난 10년 동안 조사방법론과 관련된 특정 주제에 대한 전공 논문을 편집한 총서들이 출판되었다. 이런 것들은 조사방법 관련 교재의 부족을 채우려는 노력의 일환이다. 지금까지 발표된 전공논문을 편집한 총서로는 패널조사(Kasprzyk 등, 1989), 전화조사방법론(Groves 등, 1988), 조사측정오차(Biemer 등 1991), 조사관측과 과정품질(Lyberg 등, 1997), 컴퓨터를 이용한 조사정보수집(Couper 등 1998)과 조사 무응답(Groves 등, 2002)이 있다. 조사품질과 관련된 복잡한 사안들에 대한 좀더 진전된 연구는 Platek과 Särndal(2001)에 의해 최근에 시작됐다.

제 2 장 조사과정과 자료의 품질

이 장에서는 조사과정을 재검토하고 그 과정의 각 단계와 관련된 오차의 주요 원인을 설명한다. 또한 평균제곱오차(*mean squared error*)라는 척도를 이용하여 조사자료의 오차수준을 정량화하는 방법을 생각해 볼 것이다. 이런 조사의 정확성을 나타내는 척도는 이 책 전반에 걸쳐 조사오차의 주요 원인들을 확인하고, 주어진 예산과 일정이라는 제약조건하에서 그것들을 가능한 줄이기 위한 노력의 가이드라인 역할을 할 것이다. 평균제곱오차는 최상의 방법을 선택하기 위해서 대안이 될 수 있는 다른 방법들을 비교하는 도구의 역할을 할 것이다. 그러므로 조사오차를 측정하는데 있어 평균제곱오차의 개념은 자료 품질 연구의 기본이다.

2.1 조사과정의 개요

1장에서 언급했듯이 오늘날 정부, 기업, 교육, 과학과 일상적인 개인 생활과 관련된 시기적절하고 정확한 정보는 매우 중요하다. 현재를 이해하고 미래를 위한 계획을 수립하기 위해서는 기업조직과 사회단체 같은 기관뿐만 아니라 사회 구성원의 선호도, 요구와 행동양식에 관한 자료가 필요하다. 많은 연구자와 정책 입안자에게 표본조사와 센서스는 이런 정보를 제공해 줄 수 있는 중요한 토대가 된다.

조사(*survey*)라는 용어는 연구 대상인 모집단에 속한 단위들의 표본(*sample of units*)에 해당하는 개인, 가구, 정부 기관 등의 일부분으로부터 정보를 모으는 방법으로 자주 사용된다. 예를 들어, 노동인력의 규모 측정을 위해 정부가 표본 구성원들에게 현재의 직업에 대해 질문을 할 수 있다. 기업은 조사에서 얻은 정보를 통해 그 기업의 제조비용과 다른 비슷한 기업들의 비용을 비교할 수 있다(1장 참조). 이 절에서는 표본 조사를 계획하고 실시하는 과정에 대한 개요를 소개한다. 조사과정을 이해하는 것이 조사품질을 측정하고 관리하는데 매우 중요한 역할을 한다.

<그림 2.1>에서 볼 수 있듯이 조사과정은 연구 목적을 결정하는 것부터 자료의 분석까지 대체로 순서대로 실행하는 몇 개의 단계로 이루어져 있다. 지금부터는 조사과정의 주요 과정을 어떤 단체가 새로운 법안을 만들거나 현행 법안을 평가하는 가설적 연구와 연계해서 설명한다. 개념을 보다 쉽게 설명하기 위해 해당 단체가

미국 Health Care Finance Administration(HCFA)라고 가정하자.



<그림 2.1> 조사과정: 조사과정의 계획 단계는 대부분 반복된다. 계획의 각 단계에서 설계의 실현 가능성에 관해 새로운 정보가 나타날 수 있다. 조사를 수행하기 전에 연구목적, 설문지, 목표모집단, 표본설계와 수행 전략이 여러 번 수정될 수 있다.

HCFA는 65세 이상의 시민과 장애가 있는 미국 시민에게 의료 보호 혜택을 제공하는 의료건강보험(Medical Health Insurance) 프로그램을 운영하는 기관이다. HCFA가 의료보호 대상자(Medical beneficiaries)라고 부르는 의료보호를 받는 사람들의 건강 상태를 계속하여 모니터하는 데 관심이 있다고 가정하자. 그들은 의료보호 혜택의 새로운 수혜자(즉, 최근에 65번째 생일을 맞은 사람)의 전반적인 건강상태를 측정하고, 아울러 이 모집단의 건강상의 특성이 시간이 지나면서 나이를 먹고 계속 의료 보호 혜택을 받으면서 어떻게 변화하는지를 측정하는데 특히 관심이 있다. 이 연구는 미국 의회의 요구에 의해 수행되었는데, 연구를 시작하는 2년간에 대한 조사 진행일정과 처음 3년 동안 초과해서는 안 되는 총 예산을 명확히 정하고 있다. 이 예를 사용해서 나이 많은 미국 시민 모집단의 건강에 대한 정보를 얻기 위한 조사를 설계하고 실시하는 과정을 생각한다.

연구목적의 결정

조사 과정의 첫 단계는 연구목적(즉, 조사 결과 또는 조사 자료의 분석에서 산출되는 기본적인 추정값)을 결정하는 것이다. 구체적인 연구목적은 조사과정의 중요한 요소이고 조사 설계에서의 많은 결정을 용이하게 할 것이다. 연구목적은 조사를 통해 응답할 수 있는 주요 연구문제(연구에서 관심을 갖는 핵심적인 사항)들의 확인을 통하여 잘 규정할 수 있다. 이런 작업은 보통 조사 의뢰자 또는 조사를 위임하는 연구원과 협력하여 공동으로 이루어진다. 여기서 조사 의뢰자는 HCFA이다.

의료보호제도 분석가에게 매우 중요하면서 매우 일반적인 질문은 의료보호프로그램이 보호 대상자의 건강과 복지에 기여하는지 그리고 어떻게 기여하는 지이다. 이런 추상적인 개념에 대한 정보를 얻기 위한 설문지를 설계하기 전에 순차적인 몇 가지 단계가 필요하다. 첫째로, HCFA가 주관하는 노인의 건강과 복지 전문가들이 참석하는 회의를 개최할 수 있다. 전문가들은 개념을 충분히 설명하고 평가하기 위해 측정되어야 하는 개념의 다양한 요소들을 결정해야 한다. 예를 들어, 그들은 음식 섭취, 운동, 의료진단, 삶의 질 등에 대한 자료를 수집하도록 결정을 내릴 수 있다. 그들은 또한 다른 연구에서 확인되어 이해하기 쉬운 이런 개념들을 평가하기 위해 기존의 척도, 도구, 또는 등급을 검토할 수도 있다. 이들 전문가들은 65세 이상의 모든 사람들이 의료보호를 받기 시작하는 시점에 이런 정보를 수집해야하고, 보호 대상자들이 나이를 먹으면서 이런 특성에 어떤 변동이 발생하는지 파악하기 위해 시간을 두고 주기적으로 정보를 다시 수집해야 한다고 결정할 수도 있다.

전문가들은 의사 방문; 처방 받은 약; 현재 건강상태; 키, 몸무게, 혈압, 응답자의 기능 상태(시력, 청력, 이동성, 정신 건강) 같은 개인 특징; 삶의 만족도, 우울증 빈도와 다른 정신 상태에 대한 자료를 수집하는 것을 권하기도 한다. 더 나아가서 그들은 의료 서비스에 대한 요구가 증가하면서 이런 특성들이 시간이 지나면서 어떻게 변화하는지 파악하기 위해 동일한 사람들을 표본으로 하는 조사를 반복할 것을 결정할 수 있다.

포괄적인 연구문제들의 개발에 쓰인 시간은 설문지 설계 단계에서의 시간을 절약해 준다. 왜냐하면 연구를 위해 제기된 각각의 문제들이 결국은 자료 수집 과정에서 수집될 한 개 이상의 자료 구성 요소와 연결될 수 있기 때문이다. 이 자료 구성 요소나 항목은 조사 설문지에 있는 한 가지 이상의 질문과 연결이 되어 있다. 과다하거나 조사목적을 위해 필요하지 않은 질문이 설문지에 나오게 되는 상황을 막기 위해 설문지에 있는 모든 질문이 최소한 하나의 연구문제와 상응하게 하는 것이 바

람직하다. 연구 목적과 조사 질문을 연결시키는 과정은 연구 목적에 부합되는 필요한 조사 질문들이 설문지에 포함되어 있는 지 여부를 확인하게 해 준다(<표 2.1> 참조). 이 장의 후반부에 설명되는 것처럼 이런 접근방식을 통해 결과의 특정화 오류(*specification error*)의 발생위험을 최소화 할 수 있다. 특정화 오차는 조사 질문이 연구문제(연구주제와 관련된 핵심적인 사항)에 대한 정보를 얻는 데 있어 필수적인 것들을 응답자에게 묻지 못할 때 발생하는 오차이다.

<표 2.1> 연구 문제와 조사 질문의 일치*

연구 질문	조사 질문						
	SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
							← SQ7은 불필요한 질문이다;
RQ1	v	v					삭제할 수 있음
RQ2	v		v				
RQ3				v			
RQ4					v	v	
RQ5							← RQ5에 초점을 맞출 질문사항이 없음

* 이 표는 설문지에서 연구목적에 부합되지 않거나 과다 또는 불필요한 질문을 파악하는데 유용하다.

목표 모집단의 정의

조사 과정의 다음 단계는 연구 대상 모집단이나 조사의 목표가 되는 모집단을 정하는 것이다. 목표 모집단(*target population*)은 연구 결과가 적용되고 조사 결과에서 얻어진 추론결과의 대상이 되는 사람 또는 다른 단위들의 집합이다. 의료보호 연구에서 목표 모집단은 “미국에 거주하는 65세 이상의 의료보호제도에 등록된 사람”이다. 이 정의는 미국 밖에 거주하는 의료보호제도 대상자나 의료보호제도에 등록하지 않은 65세 이상의 사람들은 포함되지 않는다는 점에 유의해야 한다. 하지만 의료보호 혜택의 수혜 여부에 관계없이 의료보호제도에 등록된 사람들은 포함된다. 목표 모집단에 누구를 포함하고 누구를 제외할 것인가에 대한 결정은 중요하다. 나중에 보겠지만 이 결정은 조사과정상의 모든 단계의 조사 설계와 관련된 중요한 결

정들을 하는데 영향을 주게 된다(3장 참조).

자료수집방법의 결정

조사 목적과 목표 모집단이 정해지면 다음 단계는 자료수집방법을 정하는 것이다. 여기서 우리는 우편조사, 전화조사, 면접조사 또는 어떤 다른 방법을 사용할 것인지를 생각한다. 자료수집방법에 따라 매우 다른 종류의 설문지가 필요하기 때문에 설문지를 설계하기 전에 이런 결정이 내려져야 한다. 또한 자료수집방법에 따라 조사에서 사용될 수 있는 표본추출법이 달라질 수 있다. 면접조사는 보통 집락화된 표본(즉, 인접한 지역에 거주하는 사람들이 하나의 집락을 구성하는 표본)을 필요로 한다. 이를 통해 면접조사원의 이동 비용을 줄일 수 있다. 전화와 우편 조사의 경우 조사원 이동 비용이 들지 않기 때문에 보통 지리적으로 분산되고, 집락화된 필요가 없다. 하지만 전화 조사는 목표 모집단에 있는 사람들에게 대한 거의 완벽한 전화번호명부나 목표 모집단을 대표할 수 있는 전화번호의 랜덤 표본을 산출해 주는 실용적이고 비용이 적게 드는 방법을 필요로 한다. 우편 조사는 목표 모집단에 있는 사람들에게 대한 거의 완전한 주소명부를 필요로 한다. 목표 모집단에 대한 주소명부가 없으면 우편조사는 불가능하다.

자료수집방법을 결정하는데 있어서 우리가 부닥치는 첫 번째 제약조건은 비용이다. 상당히 집락화된 표본이라도 면접조사를 하면 전화나 우편으로 자료를 수집하는 것보다 몇 배 더 비용이 많이 들 수 있다. 조사에 할당된 예산이 흔히 자료수집방법의 선택을 제한한다. 또 하나 중요한 고려사항은 조사해야 할 주제에 관한 것이다. 면접원은 민감한 주제에 대한 질문의 응답에 영향을 미칠 수 있으므로 만일 이것이 우려되면 자기기입식 우편 설문지 같은 더 비공개적인 방법이 어찌면 더 바람직할 것이다.

또한 면접하는 동안 응답자와 시각에 의한 정보교환이 얼마나 필요한지를 고려해야 한다. 예를 들어 응답자가 복용하는 약을 확인하는데 플래시 카드(flash card)를 사용하는 것이 필요한지 또는 응답자가 의료상의 문제점이나 절차에 대한 매우 긴 목록에서 응답을 선택해야 하는지 등이 고려되어야 한다. 조사 기간도 자료수집방법을 결정하는데 중요한 고려 사항이다. 만일 자료 수집을 하는 기간이 두 달이 안 된다면 우편 조사는 아마도 좋은 선택이 아닐 것이다.

다양한 검토를 한 결과 설문지는 면접원의 도움 없이 응답자 스스로 작성하기 충

분히 쉽게 구성될 수 있고, 설문지를 우편으로 부치는 것이 고려중인 다른 방법보다 더 저렴하므로 자기기입식 설문지를 이용한 우편조사가 실행이 가능하면서 비용이 적게 드는 가장 적절한 방법이란 결론을 의료보호 조사설계팀은 내릴 수 있다. 더욱이 모든 목표 모집단 구성원의 현주소가 의료보호 데이터베이스에 있으므로 설문지를 올바른 주소로 우송하는 데 있어 아무런 문제도 없다.

마지막으로 적절한 수준의 응답률을 확보하기 위해서(약 75% 정도가 미국 정부 조사에서 일반적으로 허용하는 최소 응답률이다.) 우편 무응답자에 대한 전화를 통한 추적조사를 자료수집과정에 포함해야 한다. 어떤 방법을 주 자료수집방법으로 사용하면서 또 다른 방법을 두 번째 또는 추적조사 방법으로 활용하는 경우를 자료수집 설계에서 흔히 접할 수 있다. 혼합형 자료수집(*mixed-mode data collection*)이라고 부르는 이런 전략은 흔히 조사의 응답률을 극대화하기 위해 필요하다. 최상의 자료수집방법을 결정하기 위해 고려해야 할 추가적인 사항들은 6장에서 논의하기로 한다.

설문지 개발

조사과정의 다음 단계는 설문지나 적절한 조사도구를 개발하는 것이다. 이 과정에서 사전에 설정된 연구목적이 조사에서 수집할 자료 구성요소(즉, 각각의 연구 문제에 초점을 맞추는데 쓰일 변수)를 결정한다. 각각의 자료 구성요소는 설문지에 있는 질문에 대한 한 개의 응답에 상응한다. 예를 들어, 생년월일이나 약물치료에 관한 질문에 대한 응답이 각각 하나의 자료 구성요소가 될 수 있다.

조사 후 자료 분석 과정에서 몇 개의 자료 구성요소들을 종합하여 새로운 자료 구성요소를 만들 수도 있다. 예를 들어 연구문제가 의료보호 대상자의 정신적인 복지와 시간이 흐름에 따른 이것의 변동과 관련이 있다고 가정하자. 구체적으로 말하면 연구원이 의료보호 대상자들이 정신적으로 우울한지, 만족하는지 그리고 나이가 들면서 이런 속성이 어떻게 바뀌는지 파악하기를 원할지 모른다. 이런 문제에 대한 근본적인 척도는 설문지에 있는 몇 개의 자료 구성요소들로부터 산출되어 요약된 점수이다. 이 점수는 사람의 정신상태에 대한 정보를 하나의 측정값으로 요약한다. 그 결과 정신 건강 상태 점수라는 것을 얻게 되는 데, 이 점수는 하나의 연속변수 이면서 행복감이 높아지면 증가하고 반면에 우울증 징후가 나타나면 감소한다. 하지만 이 척도는 단지 하나의 자료 구성요소가 아니라 개인 정신 건강에 대한 정보

를 제공하는 몇 가지 자료 구성요소를 종합함으로써 얻어진다는 점에 유의해야 한다. 그러므로 하나의 연구문제에 초점을 맞추는데 몇 개의 자료 구성요소가 필요한 상황이 것이 보기 드문 것은 아니다(<표 2.1> 참조).

앞에서 언급했듯이 주어진 자료수집방법으로 원하는 자료를 수집하기 위해서는 자료수집방법과 목표 모집단 구성원의 능력이 설문지 설계에 반영되는 것이 바람직하다. 예를 들어 만일 자기기입식 우편 설문지를 선택하였다면, 설문지 설계는 조사 대상자 중 가장 나이가 많은 응답자도 손쉽게 설문지를 완성할 수 있도록 충분히 큰 글자체를 사용해야 할 것이다. 이와 같은 사항과 기타의 조사도구 개발에 대해서는 4장에서 논의한다.

표본추출방법의 설계

목표 모집단과 연구목적을 정했고 자료수집방법을 결정했으면 조사과정의 다음 단계는 표본설계 즉 표본추출방법을 결정하는 것이다. 표본설계를 위해 고려해야 할 사항은 조사에 사용될 표본추출틀(즉, 모집단 구성원 명단), 추출틀에서 표본을 랜덤하게 뽑는데 적용할 방법, 그리고 필요한 표본크기의 결정이다. 표본추출틀(*sampling frame*)은 표본을 추출할 목표 모집단 구성원의 명단이다. 표본추출틀은 여러 개의 리스트를 결합한 것, 지도, 또는 표본을 뽑는데 쓸 수 있는 다른 형태의 틀일 수도 있다. 앞에서 언급했듯이 표본추출을 위해 선택하는 추출틀은 자료수집 방법에 따라 크게 달라진다. 의료보호 사례에서 논리적인 추출틀은 의료보호 명부에 있는 의료보호 프로그램에 등록된 모든 사람들이다. 이 추출틀의 포함률(즉, 추출틀에 포함된 목표 모집단 구성원의 비율)은 거의 100%이다. 이것은 목표 모집단의 모든 구성원이 조사에 뽑힐 가능성이 있다는 것을 의미한다. 또한 의료보호 추출틀에는 표본 구성원들에게 설문지를 우송하는 데 필요한 모든 정보가 수록되어 있다.

무응답자에 대한 2차적인 전화 추적조사를 동반한 우편조사를 통해 조사가 수행되므로 의료보호 조사에서 이동 비용은 고려할 필요가 없다. 따라서 표본을 집락화하려고 노력할 필요 없이 완전히 랜덤하게 추출할 수 있다. 하지만 추출틀에서 모집단 구성원을 동일한 성질을 갖는 것끼리 하나의 집단으로(예를 들면 나이로) 층화하는 표본추출법을 사용할 수도 있다. 왜냐하면 이런 설계를 통해 조사비용의 큰 증가 없이 조사 추정값의 정도를 높일 수 있기 때문이다.

마지막으로 연구에서 측정할 가장 중요한 모집단 특성의 추정값에 대해 요구되는 정도(precision)를 고려하여 표본크기를 결정한다. 의료보호 조사에서 필요한 표본크기를 결정하는데 있어서 응답 거부, 사망, 잘못된 주소 또는 다른 원인에 의한 무응답 때문에 불가피하게 발생하는 표본크기의 감소현상을 고려해야 한다.

<그림 2.1>에서 보여주듯이 이 시점까지의 과정은 상당부분 반복적으로 수행된다. 예를 들어 조사의 모든 목적을 달성하기 위해서 필요로 하는 설문지나 면접방법이 주어진 예산으로 수행할 수 없을 정도로 상황하거나 또는 표본으로 선정된 응답자에게 너무 부담이 돼서 조사에 참여하기를 거절할 가능성이 높은 경우에는 설문지를 개발하는 과정에서 조사목적에 다시 조정할 필요가 있다. 더욱이 선택된 자료수집방법으로는 모든 조사 목적들이 달성 될 수 없다는 결론이 얻어질 수 있다. 이런 경우 조사에서 몇몇 연구문제를 제외시키거나 자료수집방법을 재고할 필요가 있다.

마찬가지로 표본설계를 연구하는 단계에서 적당한 표본 추출틀이 존재하지 않거나 이를 새로 작성하는 데 비용이 너무 많이 드는 경우가 있을 수 있다. 이에 따라 한 개 이상의 표본 추출틀의 사용이 필요하거나 접촉하기 힘든 집단을 제외하도록 목표 모집단의 정의를 수정하는 것이 필요할 수도 있다. 비용 문제로 인하여 표본크기를 줄여야 하는 상황은 흔히 발생한다. 그러므로 최종 설계가 확정되기 전에 앞에서 말한 여러 단계의 설계 과정이 반복될 수 있다. 표본설계에 관한 추가적인 사항들은 9장에서 고려한다.

자료 수집과 자료 처리 계획의 수립

일단 초기의 기초적인 설계가 결정되면 자료의 수집과 처리를 위한 계획을 수립할 수 있다. 이 단계에서 현장에서 조사를 실시하고, 자료를 수집하고, 자료를 컴퓨터가 읽을 수 있는 포맷으로 변환시키고, 자료를 손으로 또는 컴퓨터로 에디팅하는 과정을 구체적으로 정해야 한다. 의료보호 조사에서 이 과정은 대상자의 흐름을 관리하고, 우편 반송을 확인하고, 전화 추적 작업으로 변경하고, 설문지 자료를 입력하거나 스캐닝하고, 우편조사와 전화조사에서 얻어진 자료를 통합하는 것을 포함한다. 또한 조사 자료를 에디팅하는(즉, 설문지 응답결과 중 논리적으로 맞지 않거나 부적절한 응답들, 설문지에 기재하거나 자료를 입력하는 과정에서 발생한 오류, 다른 응답항목과 일치되지 않는 잘못된 응답들 또는 다른 자료상의 오류를 교정하는)

방안을 개발해야 한다. 최종 자료 파일의 형태는 자료 분석이 용이하게 이루어 질 수 있도록 구성되어야 한다.

의료보호 조사 설계과정에서, 노년층이 정확하게 설문지 응답을 완료할 수 있을 것인지; 우편조사방법을 썼을 때 원하는 응답률을 확보할 수 있을 것인지; 시설에 수용된 사람들을 어떻게 효율적으로 취급할 것인지; 본인 대신 대리인에게서 얻은 정보를 받아들일 것인지 등에 대한 문제가 있을 수 있다. 이런 문제들을 처리하기 위해 조사과정과 설문지에 대한 사전조사를 통해 초기 설계가 적절한 것인지 검토되어야 한다. 사전조사는 설계 내용 중 어떤 사항은 적절하지 못하므로 설계의 수정이 필요하다는 것을 지적해 준다. 한 예로 설문지 설계에 문제가 있거나 전화 추적 과정에서 우편 무응답자의 전화번호를 확보하는 데 문제가 있을 수 있다(10장 참조).

자료 수집과 처리

조사의 다음 단계에서는 이전 단계에서 수립한 자료 수집과 처리 계획을 수행하는 것이다. 조사원을 모집하고 교육시켜서 자료를 수집하기 위해 현지나 전화 센터로 보내야 한다. 만일 조사를 우편으로 실시하면 설문지를 우송하고 무응답자에 대한 추적조사 계획을 이행해야 한다. 잘 계획된 조사에서도 계획에서 벗어나게 되는 뜻밖의 문제가 발생할 수 있다. 이런 잠재적인 문제점이 실제 큰 문제로 발정하기 전에 프로젝트 관계자가 주요 과정 변수들의 측정을 통해서 자료 수집 작업의 진행을 조심스럽게 감시 하는 것이 중요하다. 그러므로 자료수집계획에서 중요한 사항은 자료 수집 과정에 대한 꾸준한 감시와 감독자와의 의견교환을 위한 계획을 수립하는 것이다. 의료보호 연구에서 이것은 설문지를 우송하고, 응답된 설문지를 확인하는 절차를 개발하고, 설문지에 응답하지 않는 표본 구성원과 연락할 전화 조사원을 교육시키고, 우편설문지를 컴퓨터로 스캐닝하고, 이런 작업들이 계획한 대로 실시되었는가를 확인할 수 있는 품질관리계획을 실시하는 것을 의미한다.

자료가 일단 컴퓨터가 읽을 수 있는 형태로 변환되면 에디팅하고, 결점을 없애고, 추정과 분석을 위한 준비를 할 수 있다. 자료를 에디팅하는 것은 가능한 범위 밖의 또는 일관성이 없는 응답을 고치고, 필요에 따라 응답자에게 다시 연락해서 추가적인 정보를 얻고, 자료에서 상식적으로 인식할 수 있는 오류를 수정하는 것을 의미한다. 자유 서술형 질문(즉, 자유롭게 응답을 서술하도록 하는 질문)에서 얻어진 정

보는 흔히 응답자가 서술한 내용을 요약해 주는 코드로 변환하는 작업이 필요하다 (7장 참조).

추정과 자료 분석

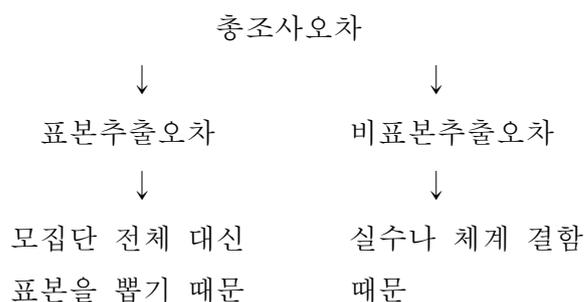
마지막으로 동일하지 않은 선택 확률, 자료의 결측, 추출틀상의 문제를 보정하기 위해 자료에 가중값(*weight*)을 주고 미리 개발한 추정과 분석 계획에 따라 추정값을 계산한다. 자료에 가중값을 주는 것은 본질적으로 표본 추정값이 참된 모수를 더 잘 나타내도록 각각의 관측값에 적당한 승수를 부여하는 것을 의미한다. 추정과 분석 계획은 자료 분석과정에서 다루어야 할 주요 연구문제, 계산할 추정값과 수행할 통계분석의 일람표를 만드는 작업을 의미한다. 여기에는 최종 추정값의 산출을 위한 가중값 적용방법과 무응답 보정방법에 대한 구체적인 사항들을 포함한다.

이 장의 나머지 부분에서는 조사 설계 과정에서 결정해야 되는 많은 사항들과 이런 것들이 어떻게 결정되는지를 이해하기 위한 전반적인 배경에 관하여 논의 한다. 상황에 따라 자료수집방법, 설문지 설계, 자료 수집 절차 등이 다르기 때문에 불행하게도 모든 상황에 적용되는 최상의 선택기준은 존재하지 않는다. 조사 설계는 오히려 과거 경험, 기본적인 이론, 그리고 우리가 접하는 각 상황에 따라 현명한 결정을 내릴 수 있도록 도와주는 다양한 설계 방법들의 장단점에 대한 분석결과를 토대로 결정되어야 한다. 따라서 특정 상황에 따른 특수한 지침 보다는 오히려 일반적인 상황에서 적용될 수 있는 바람직한 조사 설계 이론을 강조할 필요가 있다. 좋은 설계는 그 결과를 상당부분 예측할 수 있는 실제적이고 신뢰할 수 있는 과정을 사용하는 것을 전제로 한다. 따라서 여러 가지 특수한 상황에서 조사오차의 원인을 다루는 다양한 임시방편적인 방법들을 배우는 것 보다 조사오차의 근본적인 원인을 다루는 기본적인 기법들과 이런 기법들을 개발하기 위해 필요한 이론들을 배우는 것이 훨씬 유용하다는 것이 우리의 기본 철학이다.

2.2 자료품질과 총 표본추출오차

조사 자료의 많은 사용자들에게 자료품질은 단지 자료에 포함된 오차량의 함수이다. 만일 자료가 완벽히 정확하면 자료의 품질은 매우 좋다. 만약 자료가 많은 양의

오차를 포함하면 자료의 품질은 좋지 않다. 모수(평균, 총계, 비율, 상관계수 등)의 추정에서도 근본적으로 자료품질에서와 같은 기준을 적용할 수 있다. 모수에 대한 적절한 추정량을 사용했다고 가정하면 그 추정값의 근거가 된 자료의 품질이 좋으면 모수에 대한 추정값의 품질도 좋다. 반대로 만일 자료 자체의 품질이 나쁘면 추정값도 또한 품질이 나쁠 것이다. 하지만 추정값의 경우 추정값의 근거가 된 표본의 크기도 품질을 결정하는 중요한 요소이다. 자료의 품질이 좋다고 해도 너무 적은 수의 관측값을 토대로 한 추정값은 신뢰할 수 없고 잠재적으로 쓸모가 없다. 그러므로 모수 추정량의 품질은 총조사오차(*total survey error*)의 함수이며, 총조사오차는 비표본추출오차성분(*nonsampling error components*)이라고 하는 자료 수집과 처리과정과 관련된 오차 성분과 표본추출오차성분(*sampling error components*)이라고 하는 모집단에 대한 완전한 센서스를 실시하는 대신 단지 표본을 통해 결과를 산출함으로써 발생하는 오차 성분으로 구성된다.



<그림 2.2> 총조사오차: 총조사오차는 표본추출오차와 비표본추출오차로 구분할 수 있다.

지금부터 추정량(*estimator*)이라는 용어는 조사에서 추정값을 산출하는 공식이나 규칙을 말한다. 예를 들어 조사에서 어떤 특성에 대한 모집단 평균 추정량은 응답한 모든 표본 구성원에 대한 측정값의 합계를 응답한 표본 구성원의 수로 나누어 구한다. 어떤 조사 설계에 따라 얻어진 하나의 표본에서의 관측결과로 구한 추정량의 값이 22라고 하면 22는 모집단 평균에 대한 추정값(*estimate*)이라고 한다.

간단히 말하면 추정값의 총조사오차(*total survey error*: <그림 2.2> 참조)는 추정값과 실제 모수와의 차이를 나타낸다. 총조사오차의 개념을 설명하기 위해 5,000명의 근로자가 있는 작은 지역의 모든 근로자의 연평균 소득을 추정하는 매우 간단한 조사를 생각해보자. 이 경우 모수는 5,000명 모든 근로자 소득의 평균이다. 조사

설계자는 그 지역 모집단에서 랜덤으로 뽑은 400명의 근로자의 표본이 모집단 평균 소득에 대한 적당한 추정값을 산출하기에 충분하다고 결정했다. 또한 표본에 있는 400명 근로자의 소득으로 계산한 단순평균이 가장 바람직한 연평균 소득에 대한 추정량이라고 정했다. 그래서 표본을 뽑고, 조사원을 고용하여 교육시키고, 자료를 수집하여 조사 자료에서 표본평균을 계산했다.

표본에 있는 사람들의 연평균 소득이 \$32,981라고 하자. 이것은 모수에 대한 조사 추정값이다. 마지막으로 이 지역의 실제 모집단 평균소득(즉, 모수)이 \$35,181라고 가정하자. 물론 실제로는 이 값은 모른다. 그렇지 않다면 그것을 추정하는 조사를 할 필요가 없기 때문이다. 하지만 설명을 목적으로 표본 추정값의 오차를 계산할 수 있게 하기 위해 그것을 안다고 가정하자. 연평균 소득 조사 추정값과 그 지역의 알려지지 않은 모집단 연평균 소득의 차이가 연평균 소득 추정값의 총조사오차다. 이 경우 추정값의 총조사오차는 $\$32,981 - \$35,181 = -\$2,200$ 이다.

총조사오차(*total survey error*)는 모집단 평균, 총계 또는 다른 모수와 표본조사(또는 센서스)를 바탕으로 한 모수에 대한 추정값의 차이이다.

앞에서 언급했듯이 모수의 실제 값은 모르지만 경우에 따라 8장에서 나오는 방법으로 근사적인 값을 알 수도 있다. 그러므로 추정값의 총조사오차도 정확히는 알 수 없지만, 조사를 평가하는 특수한 방법으로 근사값을 파악할 수도 있다. 다음에는 왜 조사오차가 불가피하게 발생하는지 그 이유를 알아본다.

조사 추정값이 모수와 꼭 일치하지 않는 한 가지 큰 이유는 표본추출오차(*sampling error*) 때문이다. 표본추출오차는 전체 모집단인 5,000명의 근로자(즉, 완전한 센서스)를 사용하지 않고 그 지역에서 단지 400명의 근로자를 표본으로 사용함으로써 생기는 추정값과 모수의 차이이다. 다른 400명의 근로자를 표본으로 추출하면 이들 표본 근로자의 소득이 달라지므로 첫 번째 표본 추정값과 다른 추정값을 얻게 될 것이다. 추정량 산출과정에서 표본추출오차를 완전히 제거하는 방법은 완전한 센서스를 실시하는 것이다. 이 경우 “표본”에 있는 모든 5,000명의 근로자의 평균소득은 모집단 평균소득과 일치하게 된다.

하지만 우리가 표본추출오차 없이 실제 연소득을 추정하려고 그 지역 전체를 관

측할 수 있는 여유가 있더라도, 비표본추출오차(*nonsampling error*)라고 하는 다른 종류의 오차 때문에 추정값이 정확히 \$35,181이 되지 않는 것이다. 비표본추출오차는 조사과정의 각 단계에서 발생한다. 비표본추출오차는 자료수집, 자료처리와 추정 등을 하는 과정에서 생길 수 있는 다양한 종류의 표본추출오차를 제외한 모든 오차를 포함한다. 이 오차들의 누적된 결과가 총조사오차에서 비표본추출오차 부분이다. 앞의 예에서 비표본추출오차는 다음과 같은 원인에 의해 발생할 수 있다.

- 응답자(*respondent*): 응답자들은 그들의 실제 소득을 노출하고 싶지 않을 수 있고, 고의적인 것은 아니지만 응답과정에서 팁, 선물, 보너스, 상금 등과 같은 것들은 소득에서 제외할 수 있다.
- 면접원(*interviewer*): 면접원들이 조사용지에 정보를 기재하면서 실수를 하거나 응답자가 실수를 하게 할 수 있다. 예를 들어 어떤 것들이 소득에 포함되는지에 대해 응답자에게 부정확한 정보를 줄 수 있다.
- 참여 거부(*refusals to participate*): 조사를 위해 접촉한 400명 중 몇 명은 그들의 소득에 대한 응답을 거부하거나 면접에 참여하는 것 자체를 거절할 수도 있다.
- 자료 입력(*data entry*): 조사 설문지에 써놓은 소득을 자료입력과정에서 잘못 입력할 수 있다.

이들 오차 중 하나 또는 전부가 잘못된 소득을 기록하게 하는 결과를 가져올 수 있고 그 결과 연소득 추정값이 실제 모수에서 벗어나는 원인이 된다. 그러므로 비표본추출오차는 조사과정의 어느 단계에서나 일어날 수 있는 실수나 고의가 아닌 오차로 볼 수 있다. 그것을 피하기 위한 우리의 많은 노력에도 불구하고 비표본추출오차는 특히 대규모의 자료수집에서는 피할 수 없다. 반면에 표본추출오차는 우리가 표본크기를 조절해서 오차의 크기를 조정할 수 있다는 의미에서 상당부분의 도적인 오차이다. 표본크기가 1일 때 표본추출오차가 최고이고 표본크기를 모집단 크기(앞의 예에서는 5,000)로 커지면(센서스처럼) 표본추출오차가 0으로 완전히 없어진다. 그러므로 표본추출오차는 표본크기를 조정하여 원하는 만큼 작게 할 수 있다. 이 장의 후반부에서 표본추출오차에 대한 더 많은 실례들을 볼 수 있다. 표본추출오차는 9장에서 더 자세히 다룬다.

반면에 비표본추출오차는 예측할 수 없고 쉽게 통제되지 않는다. 예를 들어 표본 크기가 커짐에 따라 비표본추출오차의 수준은 실제적으로 상당히 커질 수 있다. 이것은 경험이 적고, 특정한 종류의 오류를 더 내기 쉬운 또는 적절한 감독을 받지 않는 더 많은 면접원을 고용함으로써 발생하는 결과일 수 있다. 조사의 규모가 커지면 품질관리의 체계가 일부 무너지거나 특정 오차들을 방지하는데 비효과적일 수 있다. 조사 자료의 처리과정에서도 비슷한 관리상의 문제가 발생하기 쉽고 그 결과 더 큰 자료 처리 오차가 생길 수 있다.

$\text{총조사오차} = \text{표본추출오차} + \text{비표본추출오차}$

최근 50년 동안 조사분야 연구자들은 조사에서 비표본추출오차가 표본추출오차보다 추정값에 훨씬 더 악 영향을 미치는 경우가 많다는 것을 인식하기 시작했다. 이미 말했듯이 이 책 전체뿐만 아니라 이 장의 중요한 목적은 어떻게 이런 오차가 발생할 수 있는지, 조사에서 표본추출오차와 같이 비표본추출오차를 통제하는 것이 왜 필요한지 설명하는 것이다.

비표본추출오차에 관해 많은 연구 논문들이 있다: 비표본추출오차의 출처와 원인, 최종결과에서 그것들을 최소화하는 조사 설계, 조사 결과에 미치는 영향을 평가하는 통계적방법과 모형, 추정값에 미치는 영향을 줄일 수 있는 사후 조사를 통한 조정 등이다. 조사 자료품질의 비결은 비표본추출오차의 근본원인을 이해하고 그것을 최소화 하는 방안을 강구함으로써 찾을 수 있다. 따라서 이 책에서는 이러한 측면을 다루도록 한다. 머리말에서 언급했듯이 우리의 목적은 특정한 주제를 깊이 있게 다루기보다는 광범위하게 다루는 것이다. 각 주제에 대한 더 깊이 있는 내용은 독자의 관심에 따라 조사오차에 관한 여러 분야의 연구 논문을 통해 얻을 수 있다. 이러한 논문은 이 책에 제공되어 있는 참고문헌에 나와 있다.

2.3 비표본추출오차의 구성 요소

모든 조사 설계의 목적은 조사의 예산과 다른 자원에 대해 주어진 제약조건하에

서 추정값의 총조사오차를 최소화 하는 것이다. 이 장의 후반에 볼 수 있듯이 조사 경비를 조절하여 비표본추출오차를 줄이는 것은 때때로 비표본추출오차의 주요 원인을 줄이기 위하여 표본추출오차를 늘이는 것(표본 크기를 줄여서)을 뜻한다. 최적의 조사 설계를 하는 것은 조사를 위해 배정된 예산으로 총조사오차가 가능한 작게 되도록 표본추출오차와 비표본추출오차의 균형을 찾는 것이다. 이를 위해선 오차의 주요 원인이 만족스러운 수준으로 통제되게 조사 자원을 조사 과정의 다양한 단계에 배분하는 것이 필요하다. 이것은 과정의 모든 단계를 경비를 고려하지 않고 가능한 한 정확하게 수행하는 것을 요구하는 것은 아니다. 왜냐하면 이런 접근방식은 조사 예산을 크게 넘어가게 하는 결과를 초래할 수 있기 때문이다.

주어진 예산이 한정되어 있기 때문에 조사원을 충분히 교육시키기 위해서는 자료 처리 단계에서 실시하는 품질관리 활동을 없애거나 제한해야 할 수도 있다. 조사의 응답률을 만족스러운 수준으로 높이기 위해서는 표본크기를 크게 줄이는 것 등이 필요할 수도 있다. 이런 결정을 어떻게 내려야 할까? 현명하게 균형을 맞추기 위해서는 비표본추출오차의 원인과 그것을 어떻게 관리 할 수 있는지를 이해해야 한다. 이 목적을 위해 다음 절에서는 비표본추출오차의 주요 원인을 자세히 알아본다.

2.3.1 비표본추출오차의 5가지 요소

<표 2.2>는 비표본추출오차를 발생원인에 따라 5가지 유형으로 정리한 것이다: 특정화 오차(specification error), 추출틀 오차(frame error), 무응답 오차(nonresponse error), 측정 오차(measurement error), 그리고 처리 오차(processing error). 이 책에서 다루는 모든 비표본추출오차는 이런 다섯 가지 분류 중 한가지에 해당한다.

특정화 오차

특정화 오차는 조사 질문이 뜻하는 개념과 조사에서 측정해야 되는 개념이 다를 때 일어난다. 이런 경우 조사에서 모수가 잘못 추정되므로 그 추정값을 토대로 한 추론은 오류를 범하게 된다. 특정화 오차는 보통 연구원, 자료 분석가, 또는 조사의뢰자와 설문지 설계자 사이의 잘못된 의사소통이 원인이다. 예를 들면, 농업 조사에서 연구원이나 의뢰자가 토지 한 구획을 공정한 시장가격으로 팔았을 때 그 가격

에 관심이 있다고 하자. 이 경우 적절한 질문은 “만일 이 토지를 오늘 팔려고 내놓으면, 그 토지 한 구획의 공정한 가격은 얼마가 되겠습니까?” 이다. 하지만 조사 질문을 간단히 “어떤 가격에 이 토지를 팔겠습니까?”라고 물을 수도 있다. 이 경우

<표 2.2> 다섯 가지 유형의 비표본추출오차와 발생원인

오차의 원인	오차의 종류
특정화 오차	개념 목적 자료구성요소
추출틀 오차	누락 잘못된 포함 중복
무응답 오차	전체 조사 단위 조사 항목 불완전한 정보
측정 오차	정보체계 배경 자료수집방법 응답자 면접 도구
처리 오차	에디팅 자료입력 코딩 가중처리 도표화

그 토지의 시장 가치를 측정하는 대신 질문이 그 토지 한 구획이 농지 소유자에게 얼마의 값어치가 있는 지를 대신 측정하는 것이 될지도 모른다. 농지 소유자는 시장 가치보다 훨씬 비싼 가격을 주지 않으면 그 땅을 매도할 의사가 없다면 그 가격은 시장가격에 비해 훨씬 높게 나타날 것이다. 조사 질문이 연구문제에서 제기된 개념과 맞지 않으므로 이 질문은 특정화 오차를 범했다고 한다.

이 예에서 한 걸음 더 나아가 조사 분석가가 울타리, 관개시설, 이착륙장, 저장고, 부속 건축물 등과 같은 그곳에 존재하는 부가적인 시설물의 가치를 포함하지 않는 순수 토지의 가치에만 관심이 있다고 가정하자. 하지만 조사 질문에는 이 부분에 대한 언급이 없을 수 있다. 예를 들어 간단히 “이 토지 한 구획의 현재 시장 가격이 얼마라고 생각합니까?”라고 묻는다면, 이 질문이 토지에 세워진 시설물의 가치를 명백히 제외하지 않고 있으므로 조사된 토지 가격은 연구원도 모르게 시설물의 가치만큼 부풀려져 조사될 수도 있다. 더 타당한 질문은 “이 토지 한 구획의 현재 시장가격이 얼마라고 생각합니까? 단, 울타리, 저장고, 관개시설 등과 같은 시설물의 가격은 포함하지 마십시오.” 일 것이다.

“이 토지 한 구획의 현재 시장 가격이 얼마라고 생각합니까?”라는 질문이 잘못 표현된 질문은 아니다. 단지 연구목적에 고려할 때 잘못된 질문이다. 연구목적에 분명하게 이해하지 못하고, 농업경제학자 또는 다른 자료 사용자가 토지 가격에 대한 자료를 어떻게 사용할 것인지 모르는 설문지 설계자는 이러한 특정화 오차를 파악할 수 없을 것이다. 이 때문에 연구 분석가나 연구목적에 확실하게 달성하기 위해 측정해야 할 개념을 정확하게 이해하는 사람이 질문들을 철저히 재검토해야 특정화 오차를 확인할 수 있다. 연구 분석가는 각 질문을 연구목적과 관계되는 본래의 의도와 관련해서 재검토하고 질문이 그 의도를 충분히 반영하나 확인해야 한다. 토지 가격의 예에서 농업 경제학자나 토지 가격에 대한 자료를 사용할 다른 분석가가 조사 설문지에 특정화 오차가 있는지를 검사 할 가장 적합한 사람들이다. 일반적으로 특정화 오차를 찾아내기 위해서는 연구목적에 달성하기 위해 자료를 분석하는 것에 책임을 지고 있으며, 조사에서 어떤 개념을 측정해야 하는지 가장 잘 알고 있는 연구원들이 조사 질문들을 재검토하는 것이 필요하다.

계량경제학과 같은 학문에서 특정화 오차는 잘못된 변수를 회귀모형 같은 모형에 포함하거나 모형에서 중요한 변수들을 무시하는 것을 뜻한다. 여기서 특정화 오차는 모형이 아니라 설문지에 있는 질문을 대상으로 하고 있다.

추출틀 오차

비표본추출오차의 다음 원인은 조사를 위해 표본추출틀을 구성하는 데서 일어나는 오차이다. 표본추출틀은 표본을 추출하는데 사용되는 목표모집단 구성원들의 목록이다. 위의 의료보호조사의 예에서는 추출틀이 의료보호제도에 등록된 모든 사람

의 명단이었다. 하지만 농업 토지가격에 대한 조사 사례에서처럼 지도에 그려진 토지 구획들을 랜덤하게 표본으로 선택하는 경우 추출틀은 연구대상 지역의 지도일 수도 있다.

표본추출틀이 물질적인 목록이 아니라 개념적인 목록일 수도 있다. 예를 들어 전화조사 표본은 흔히 랜덤번호전화걸기(*random-digit dialing*; RDD)라고 부르는 방법을 사용해서 선택된다. 미국과 캐나다에서 실시하는 RDD조사에서는 추출틀이 전화번호로 사용이 가능한 모든 10자리수 번호들로 구성된 개념적인 목록이다. 즉, 이 경우 실제적인 전화번호 목록 대신 개념적인 목록을 사용하는 것이며, 랜덤하게 10자리수를 발생시키는 알고리즘을 사용해서 필요에 따라 표본 전화번호를 랜덤하게 산출한다.

표본이 확실하게 모집단 전체를 대표하기 위해서는 추출틀에 모집단을 구성하는 모든 사람, 농장 경영자, 가구, 사업체 또는 기타의 조사단위들이 목록에 기입되어 있어야 한다. 또한 적절한 선택 확률을 구해 응답값에 가중값을 주기 위해서는 중복기재가 가능한 경우 추출틀에 있는 각 구성원이 목록에 기재된 횟수도 알아야 한다.

추출틀을 구성할 때 몇가지 일어날 수 있는 오차가 있다. 모집단 구성원이 일부 누락되거나 또는 중복 횟수가 알려지지 않는 경우가 있다. 또는 포함되지 말아야 되는 구성원이 추출틀에 포함되어 있을 수도 있다(예를 들어 농업 조사에서 농업에 종사하지 않는 가구 등). 완전한 추출틀을 만드는 데 비용이 너무 많이 드는 경우 잘못된 누락이 자주 일어난다. 우리는 조사비용 때문에 누락이 발생하는 것을 흔히 감수해야 한다. 한편 몇 개의 목록이 결합 되어 하나의 추출틀을 구성하게 되는 경우 구성원의 중복 등재 문제가 흔히 발생한다. 또한 각 추출틀의 구성원이 모집단에 속하는지 여부를 확인할 수 있는 정보가 충분하지 않으면 모집단에 속하지 않는 구성원이 추출틀에 잘못 포함되는 문제가 발생할 수 있다. 3장에서는 누락, 잘못된 포함과 중복의 문제들이 조사 추정값의 오차에 어떤 영향을 주는지 논의한다.

무응답오차

<표 2.2>에 나오는 무응답 오차는 단위 무응답, 항목 무응답과 불완전한 응답을 포함하는 상당히 일반적인 오차의 원인이다. 단위 무응답(*unit nonresponse*)은 표본 조사단위(가구, 농가, 사업체 등)가 설문지의 어떤 문항에도 응답하지 않을 때 일어

난다. 예를 들어 한 가구가 조사에 참여하기를 거절하거나 우편 설문지가 조사 대상 사업체로부터 반송되지 않는 경우 이다. 항목 무응답(*item nonresponse*)은 설문지가 부분적으로만 완성되었을 때 일어난다(즉, 답을 해야 하는 어떤 항목들을 건너뛰거나 빈칸으로 남겨둘 경우). 한 예로, 가구 조사에서 가구원 소득에 대한 질문들은 전형적으로 항목 무응답이 나오기 쉽다. 왜냐하면 응답자들은 설문지에 있는 많은 다른 질문들은 답할지라도 흔히 그들의 소득은 드러내기를 원치 않기 때문이다. 마지막으로 서술형 질문에 대한 불완전한 응답(*incomplete responses*)도 무응답 오차의 한 종류이다. 여기서는 응답자가 약간의 정보를 줄 수도 있지만, 응답이 매우 짧고 불충분한 경우이다. 예를 들어 “당신의 직업이 무엇입니까?” 라는 대부분의 노동력 조사에서 나오는 서술형 질문에서 이런 종류의 무응답이 나오기 쉽다. 응답자는 그의 직업에 대하여 어느 정도의 정보를 줄지 모르지만 아마도 그 정보가 나중의 자료처리 단계에서 직업 또는 산업 코드 번호로 변환하는 것이 가능할 만큼 충분한 정보가 아닐 수 있다. 이런 종류의 오차는 7장에서 상세히 다룬다.

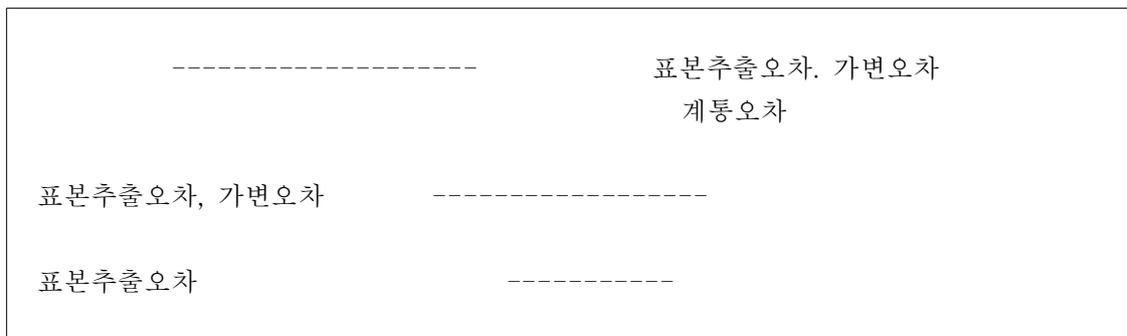
측정오차

측정오차는 널리 연구되어왔고 아마도 조사방법 연구 논문에서 다른 어떤 종류의 비표본추출오차보다 더 많이 논의 되었을 것이다. 많은 조사에서 측정오차는 악영향이 가장 큰 오차라고 볼 수 있다. 측정오차를 발생시키는 주요인은 응답자, 조사원과 조사 설문지이다. 응답자는 고의적으로 또는 무심코 틀린 정보를 줄 수 있다. 조사원들은 몇 가지 방법으로 오차를 일으킬 수 있다. 우선 그들은 자료를 왜곡하거나, 부적절하게 응답에 영향을 주거나, 응답을 잘못 기록하거나 또는 조사 절차에 따르지 않을 수 있다. 만약 설문지가 잘못 설계되었다면 그것이 오차의 주요 원인이 될 수 있다. 애매한 질문, 혼동되는 지시, 오해될 소지가 있는 용어 등이 사용된 설문지 때문에 측정오차가 발생할 수 있다.

또한 응답자들이 그들의 응답을 구체화하기 위해 사용하는 정보 체계에서 발생하는 오차를 생각해 볼 수 있다. 예를 들어 농장 경영자나 기업주가 잘못된 기록을 참고 하여 조사된 자료에 오차가 생길 수 있다. 또한 자료수집방법이 측정 오차에 영향을 미칠 수 있다는 것은 잘 알려져 있다. 예를 들어, 어떤 경우에는 전화 면접에서 얻은 정보가 대면면접에서 얻은 정보에 비해 덜 정확할 수 있다. 마지막으로 조사를 실시하는 주변여건이나 환경도 측정오차에 영향을 줄 수 있다. 예를 들어

약물사용, 성적인 행동, 생식력 등 같은 민감한 주제에 대한 자료를 수집하기 위해서는 집안의 다른 식구들이 있는 장소보다 남의 눈을 피할 수 있는 장소에서 면접을 하는 것이 정확한 응답을 얻는데 도움이 된다. 사업체 조사에서는 토지 사용, 손익, 환경 쓰레기 취급, 자금 배분 같은 주제도 민감할 수 있다. 이런 경우 비밀을 보장해 줌으로써 고의적인 잘못된 응답에 의한 측정오차를 줄일 수 있다.

이런 종류의 비표본추출오차들은 조사 추정값의 정확도에 엄청나게 큰 영향을 줄 수 있다. 이를 설명하기 위해 알려지지 않은 실제 평균소득이 \$35,181인 어떤 지역의 근로자 연평균 소득을 추정하기 위한 조사의 예를 다시 보자. 표본이 400명인 경우 표본추출오차에 의한 추정값 오차가 \$500정도로 예측된다고 하자(표본추출오차의 예측 방법에 대한 세부사항은 9장을 참조하라). 즉, 조사에서 나오는 추정값이 낮게는 \$34,681 높게는 \$35,681이 될 수 있다. 그렇지만 위에서 설명한 다양한 원인 때문에 발생하는 비표본추출오차의 영향으로 조사 추정값의 오차는 훨씬 더 커질 수 있다. 예를 들어, 오차가 표본추출에서 생기는 오차의 두 배인 \$1000 정도 된다고 예상하는 것이 전혀 근거 없는 것은 아니다. 이 결과 실제 모수가 \$35,181일 때 조사 추정값은 적게는 \$34,181 크게는 \$36,181이 될 수 있다는 것이다(<그림 2.3> 참조).



<그림 2.3> 표본추출오차, 가변오차와 계통적인 오차에 의한 표본조사에서 산출한 추정값 범위: 표본크기가 400일 때 평균 소득에 대한 가능한 추정값 범위를 보여준다. 표본추출오차만 있을 때는 범위가 훨씬 작고 체계적 비표본추출오차가 포함되면 가능한 추정값 범위가 실제 참값을 포함조차 하지 않을 수 있다.

소득을 과대하게 말하는 응답자의 오차가 소득을 과소하게 말하는 응답자의 오차와 균형을 이루지 않으면 추정값에 더욱 심각한 오차가 생길 수 있다. 즉, 만약 오

차가 한 쪽 방향으로만 치우쳐 발생하면 추정값을 편향되게 하는 경향이 생긴다. 예를 들어, 소득의 경우에 응답자들은 일반적으로 그들의 소득을 과대하게 말하기 보다는 과소하게 말하는 경향이 더 클 것 이므로 음의 오차가 주로 발생한다. 이런 상황에서는 추정값에 음의 편향(*negative bias*)이 생기게 되는데, 결과적으로 조사 추정값이 항상 실제 모수 보다 어떤 미지의 액수만큼 작아진다는 것을 의미한다. 이 경우에는 실제 모수가 \$35,181이지만 소득 추정값의 범위는 오히려 \$33,681에서 \$34,681에 더 가까울 지도 모른다. 편향이 발생하는 계통오차(*systematic errors*)와 비편향인 가변오차(*variable errors*)에 대해서는 다음 절에서 더 다룬다.

처리오차

<표 2.2>에 있는 마지막 다섯 번째 오차는 자료 에디팅, 자료 입력, 코딩, 조사 가중값 할당과 조사 자료의 도표화 등을 포함하는 자료처리 단계에서 발생하는 오차인 처리오차다. 에디팅 오차의 예로 자료 처리 담당자는 응답이 명확히 어떤 한 계를 넘으면 응답자에게 전화를 다시 걸어서 예산 항목을 확인하도록 지시를 받았다고 하자. 어떤 경우에는 담당자가 이런 규정을 올바르게 수행하지 않아서 자료에 오차가 생길 수 있다.

코드화될 서술형 항목에서는 코딩 오차가 또 한 종류의 처리 오차다. 자료를 코딩하는 직원이 실수를 하거나 규정된 절차를 벗어날 수도 있다. 직장, 직업, 종사하는 산업, 대학생들의 전공 등에 대한 응답에 코드 번호를 부여하는 경우 코드 부여 체계 자체가 불명확하여 실수하기 쉬울 수 있다. 이 결과 코드 번호가 일관성이 없거나 부정확하게 부여될 수도 있고, 결과적으로 심각한 코딩오차가 발생하게 된다.

균일하지 않은 선택 확률, 무응답 오차, 또는 추출틀 포함오차를 통계적으로 보정하는 조사 가중값이 잘못 계산될 수도 있고, 가중값을 계산하는 추정 소프트웨어에 프로그래밍오류가 있을 수도 있다. 자료의 도표화를 위한 소프트웨어에 있는 오류가 최종 결과표에도 악영향을 미칠지 모른다. 예를 들어 추정값을 계산하는데 쓰이는 스프레드시트(도표 계산용 소프트웨어)에 미처 발견되지 못한 셀-참조 오류가 있어서 가중치가 잘못 적용되고 결과적으로 조사 추정값이 잘못 계산될 수 있다. 이에 대한 좀더 자세한 사항은 7장과 8장을 참고하기 바란다.

2.4. 총조사오차 크기에 대한 평가

2.3절 에서 보았듯이 조사 설계를 개발하기 위해서는 조사 추정값의 총 오차에 영향을 미칠 수 있는 많은 사항들을 결정해야 한다. 이것들은 표본크기, 자료수집방법, 조사원 교육과 감독, 설문지 설계 등에 관련된 결정들인데 이런 것들이 조사 자료의 품질을 결정하게 된다. 이런 결정들은 다양한 옵션에 따른 비용과 조사기간에 미치는 영향을 고려한다. 우편조사가 방문조사보다 비용이 적게 들지만 주어진 조사기간이 짧은 경우 우편조사는 실현 가능성이 없을 수도 있다. 대면면접은 면접원 비용과 현지 조사비용 때문에 적절하지 않을 수 있기 때문에 주어진 시간 내에 조사를 수행할 수 있으면서 비용이 적게 드는 또 다른 방법을 찾아야 할 필요도 있다. 이런 상황에서 전화 면접은 비용도 적절하고 적시성도 만족할 수 있지만 어떤 항목의 경우 자료품질에 문제가 있을 수도 있다. 예를 들어 응답자에게 알약 카드 목록이나 잡지표지 같은 시각적 정보를 직접 보여주는 것이 필요한 질문들은 전화로 실행할 수 없다. 그러므로 조사의 설계를 결정할 때에는 일정, 예산, 그 밖의 제약조건을 충족시키며 자료 품질면에서 최고의 설계가 되도록 여러 요소를 동시에 고려해야 한다. 그 결과로 산출된 설계가 다수의 사용자와 자료의 용도를 우선적으로 고려한 적절한 타협점이 될 것이다.

올바른 설계의 결정은 여러 가지 품질, 가격과 적시성 등이 요소를 동시에 고려하고 총조사오차를 최소화함과 동시에 예산과 일정을 충족시키는 설계구성요소를 선택하는 것을 필요로 한다. 조사 과정에서의 총오차를 정량화하면 설계 과정에 많은 도움을 준다. 이런 정량화를 통해 다양한 조사 설계들을 비용과 적시성 뿐만 아니라 총조사오차의 각도에서도 비교할 수 있다.

한 예로 설계 A와 설계 B 라는 두 가지 조사 설계를 생각해 보자. 그리고 두 설계 모두 주어진 예산과 시간적 제한을 충족시킨다고 가정하자. 하지만 조사에서 측정할 주된 특성(예를 들어 소득)에서 보면, 설계 A의 추정값의 총오차는 $\pm \$3780$ 인 것이 반하여 설계 B의 총오차는 $\pm \$1200$ 밖에 안 된다. 설계 B의 오차가 훨씬 작기 때문에 다른 모든 조건이 같다면 이것이 선택될 것이다. 이와 같이 조사 과정에서 발생사는 총오차를 종합하고 정량화하는 방법으로 경쟁이 되는 설계 중에서 최선의 방법을 선택할 수 있다.

이런 척도를 이용한 접근방식은 또 다른 이점을 갖고 있다. 예를 들어, 설계 B에

의한 조사 과정의 대부분의 오차는 무응답 오차 때문이라는 것을 입증할 수 있다고 가정하자. 이것은 무응답이 설계 B의 가장 중요한 오차의 원인이라는 것을 나타내므로 집중적인 무응답오차의 축소를 통하여 설계 B에 의한 조사 자료의 품질을 더 향상시킬 수 있다. 무응답오차를 줄이기 위해 더 많은 비용을 사용하기 위해서는, 조사 과정의 다른 분야의 비용을 줄일 수 있는 조사 설계를 해야 한다. 이런 조정으로 인해 다른 오차 원인이 늘어날 지도 모르지만 무응답 오차의 대폭적인 감소를 통해 종합적으로는 조사오차를 줄일 수 있다. 이런 과정을 통해 설계 B와 관련된 총오차는 총조사 비용을 늘이지 않고도 줄일 수 있다.

조사 추정값의 총조사오차를 정량화하는 많은 방법이 있다. 그렇지만 조사연구 관련 논문에서 가장 흔히 쓰이는 척도는 총 평균제곱오차(*mean squares error*: MSE)이다. 조사 자료에서 계산될 각 추정값에 대해 모든 오차 원인들이 추정값에 미치는 영향을 반영하는 MSE를 고려할 수 있다. MSE는 총조사오차의 크기를 측정하는 데, 좀더 구체적으로 말하면 MSE는 총조사오차가 관심의 대상인 추정값에 미치는 영향의 크기를 측정하는 것이다. MSE가 작으면 총조사오차도 작고 조사가 잘 관리되었음을 뜻한다. MSE가 크다면 한 가지 이상의 오차 원인이 추정값의 정확도에 상당부분 나쁜 영향을 미친다는 것이다. 이미 언급했듯이 이런 정보는 유사한 조사가 반복되는 경우 향후의 조사방법뿐만 아니라 이 조사 자료를 사용하는 방법에도 영향을 줄 수 있기 때문에 중요하다.

MSE의 가장 중요한 용도 중 하나는 조사 자료의 정확성을 측정한다는 것이다. 불행하게도 보통 MSE를 조사 자료에서 바로 계산하는 것이 가능하지 않으며, 특히 비표본추출오차가 큰 경우 더욱 계산이 어렵다. 대부분의 상황에서는 실제 관심대상이 되는 조사에 추가로 특수한 평가 연구를 수행하여 총 MES를 추정하게 된다. 자료의 정확성의 측정은 조사 결과의 정확한 해석을 위해 중요하다.

한 예로 2000년 미국 인구센서스에서, 센서스에 누락된 사람 수를 추정하는 동시에 궁극적으로 누락된 사람 수의 추정값으로 최종 센서스 결과를 보정하기 위해 센서스에 잇따라 사후확인조사(*postenumeration survey*; PES)를 실시하였다. 센서스와 PES를 실시하는 기간동안 누락에 따른 과소계상을 조정한 경우와 이런 보정을 하지 않을 경우의 센서스 총계의 MSE를 비교하기 위한 특수한 연구를 실시하였다. 센서스 총계의 한 중요한 용도는 하원 의원수의 할당(*congressional apportionment*)이라고 부르는 과정에 활용하는 것인데, 미국 하원의 435석을 어떻게 50주에 분배

해야 하는지를 결정하는 과정이다. 이렇게 얻어진 총 MSE로 측정된 과소계상 보정에 따른 센서스 총계의 품질 개선 정도가 2000년도 하원의원 할당에 과소계상에 대한 보정방법을 사용하지 않는다는 결정을 하는데 중요한 고려사항이었다.

그러므로 총조사오차의 측정은 조사 설계와 개선이란 측면에서 매우 중요한 개념이다. 조사 설계의 가장 중요한 목적은 주어진 예산으로 예정된 일정에 맞게 주요 조사 추정값의 MSE를 최소화 하는 것이라고 간단히 말할 수 있다. 이 장 나머지에 서는 이런 중요한 개념을 이해하고 개발하는데 전념 할 것이다. 다음 절에서는 조사 과정에서 발생하는 비표본추출오차: 가변오차(*variable error*)와 계통오차(*systematic error*)를 분류하는 방법을 논의한다. 가변오차와 계통오차는 총 MSE의 필수 구성요소이다. 왜냐하면 전자는 조사 추정값의 분산에 의해 결정되고 후자는 편향에 의해 결정된다. 이 장의 후반부에서 MSE가 근본적으로 조사과정의 많은 오차 원인이 제공하는 분산과 편향 구성 요소의 합이라는 것을 볼 수 있다.

조사 설계의 가장 중요한 목적은(*primary objective of survey design*) 주어진 예산으로 일정에 맞게 수행함과 동시에 주요 조사 추정값의 MSE를 최소화 하는 것이다.

가변오차

여기서는 소득과 관련된 조사 설문지의 특정한 항목이나 질문을 생각해 보는 것이 유용할 것이다. 이러한 항목에서는 조사의 모든 다양한 오차 원인에서 발생하는 비표본추출오차는 응답결과에 누적되어 나타나며, 어떤 한 사람의 응답값은 참값보다 크거나 작을 수 있다. 바꿔 말하면 어떤 특정 관측에 대한 총오차의 누적된 영향은 양수일 수도 있고 음수일 수도 있다. 이것은 모든 관측값에 적용된다: 어떤 사람들에게 대한 오차의 누적합은 양수이고 어떤 사람들에게 대한 오차의 누적합은 음수일 수 있다. 표본 관측값의 평균(즉, 표본에 있는 관측값들의 합 나누기 관측값의 개수)으로 모집단의 평균 소득을 추정한다고 가정하자. 또한 모집단에 있는 사람들이 그들의 소득을 보고할 때 양의 오차를 낼 가능성과 음의 오차를 낼 가능성이 같다고 가정하자. 이 상황에서 음의 오차가 어느 정도까지 양의 오차를 상쇄하여 오차의 최종적인 영향은 대체로 매우 작을 것이다. 즉, 관측의 음수 오차가 양수 오차를

상쇄하는 경향이 있어서 비표본추출오차가 모집단 평균의 추정값에 근본적으로 아무 편향된 영향을 미치지 않을 것이다.

만일 소득 자료를 수집하는 조사 과정이 모집단에게 반복적으로 수행되면 매우 비슷한 결과가 일어날 것이다. 즉, 음의 오차가 양의 오차를 거의 상쇄할 것이다. 이런 종류의 오차를 발생시키는 근원을 가변오차원인(*variable error sources*)이라고 하고 여기서 발생하는 오차들은 가변오차(*variable errors*)라고 한다. 시끄러운 방에서는 사람들이 하는 얘기를 들을 수 있는 능력이 제한되듯이, 가변오차는 자료가 우리에게 무엇을 말 하는지를 이해하는 능력을 제한하기 때문에 자료의 가변오차가 높으면 흔히 그 자료에는 잡음(*noisy*)이 있다고 한다.

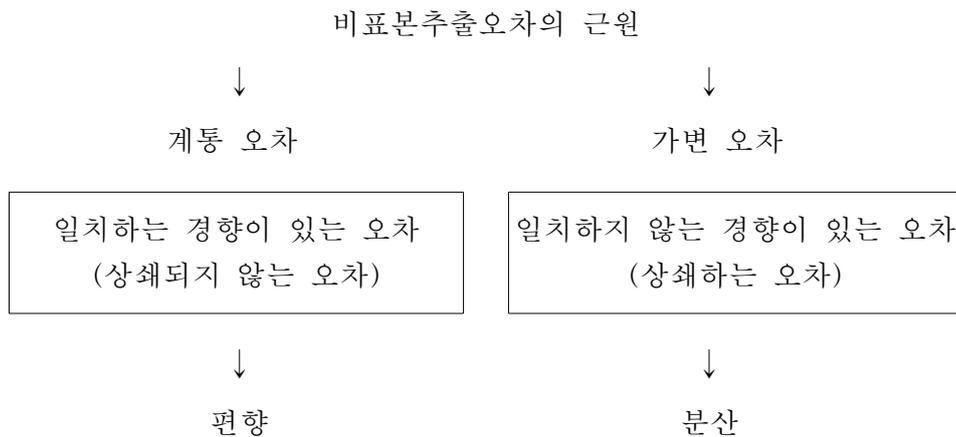
가변오차와 밀접하게 관련되고 논문에서 흔히 나오는 것이 자료의 신뢰도(*data reliability*)이다. 신뢰도(*reliability*)는 두 종류의 변동의 비로 정의된다: 모집단 구성원에 대한 참값들의 변동과 가변오차에 기인하는 부가적인 변동 뿐 아니라 참값들의 변동을 포함하는 총 변동. 이 두 가지 변동의 비를 신뢰도비(*reliability ratio*)라고 한다. 그러므로 신뢰도비는 0.0에서 1.0사이의 값을 갖는다. 완벽한 신뢰도(*perfect reliability*)는 자료에 가변오차가 없다는 것을 의미한다. 이 경우에는 비의 분자가 분모와 같고 신뢰도비는 1이다. 가변 오차가 늘어남에 따라 비의 분모가 늘어나고 신뢰도비가 줄어든다. 예를 들어 측정하는 특성의 신뢰도비가 0.5(50%) 일 때에는 모집단에 있는 어떤 특성의 참 값의 변동은 가변 비표본추출오차(*variable nonsampling error*)에 기인하는 관찰한 값의 변동과 같다는 뜻이다. 이런 경우 어떤 기준으로든 신뢰성에 문제가 있다고 판정될 것이다.

어떤 경우에는 신뢰할 수 없는 자료는 조사에 관련된 변수의 면밀한 검사를 통해 파악할 수 있다. 예를 들어 태도에 관한 조사에서 비슷한 태도에 관한 문제들에 대한 동일한 사람의 응답결과들이 상당부분 차이가 있다면 이는 신뢰성에 문제가 있다는 것을 의미한다. 하지만 대부분의 상황에서는 관찰한 자료가 신뢰할 수 있나 없나를 결정하기 위해서는 신뢰성을 평가하는 특수한 연구를 필요로 한다. 신뢰성을 평가하는 하나의 방법을 예로 들면, 처음의 면접 며칠 후에 표본가구에 대해 면접을 반복하는 재면접 방법이다. 첫 번째 면접이 두 번째 면접의 응답에 조금도 영향을 주지 않는다고 가정하고, 두 면접 결과의 비교를 통해 자료가 신뢰성이 있는지 여부를 판단할 수 있을 것이다. 만일 자료가 신뢰성이 있으면 첫 번째와 두 번째 응답이 일치할 것이고, 불일치하는 경우가 많으면 신뢰할 수 없는 자료를 의미

한다. 그러므로 신뢰도는 교육심리통계학분야에 뿌리를 둔 용어인 시험-재시험 신뢰도(*test-retest reliability*)라고 부른다.

계통오차

많은 상황에서는 양과 음의 오차가 정확하게 서로 상쇄되지 않는다. 예를 들어 양의 오차가 음의 오차보다 훨씬 더 우세할 수 있고 그 결과 관측값들의 평균이 실제 모수보다 훨씬 더 커질 수도 있다. 이런 경우 표본평균은 양의 편향을 갖는다고 한다. 반대로 음의 오차를 범하는 표본에 있는 응답자 수가 양의 오차를 범하는 수보다 훨씬 많을 수 있고 그러므로 평균의 추정값은 비표본추출오차가 없을 때 보다 작아진다(즉, 음의 편향이 생긴). 표본 관측값들의 평균을 구했을 때 합이 0이 되지 않는 오차는 계통오차(*systematic errors*)라고 부른다(<그림 2.4> 참조). 계통오차에서 음의 오차가 우세하게 되면 표본 평균은 음의 편향을 갖게 될 것이다.



<그림 2.4> 두 종류의 비표본추출오차: 모든 비표본추출오차 원인은 가변오차, 계통오차, 또는 둘 다를 만든다. 계통오차는 편향된 추정값이 산출되게 하고, 가변오차는 추정량의 분산에 영향을 준다.

비표본추출오차에 대한 논의에서 계통오차와 가변오차의 정의는 선택된 어떤 특정한(*particular*) 표본에서 일어나는 것을 기준으로 한 것이 아니라는 점에 유의해야 한다. 좀 더 정확히 말하면 근본적으로 같은 조사 상황에서 동일한 조사과정을 수많은 반복 수행하여 얻어진 표본들과 결과들을 종합한 결과를 기준으로 한 것이

다. 조사과정을 반복한다는 개념은 통계적 품질관리 분야에서의 가정과 흡사하다. 예를 들어 컴퓨터 칩 같은 어떤 제품의 제조를 위해 설계된 과정을 생각해 보자. 과정의 설계자에게 중요한 것은 수많은 반복을 통해 그 과정에서 생산된 칩의 품질이지 그 과정에서 생산된 특정한 하나의 칩이 아니다(물론, 그 칩을 구입한 소비자에게는 그 칩이 가장 중요한 관심사일 것이다.). 마찬가지로 조사도 자료를 생산하는 과정이다. 자료의 소비자 입장에서는 어떤 조사방법을 통해 얻어진 하나의 특정한 조사 수행결과에 주로 관심이 있지만, 조사 자료 품질 이론은 이 보다 그 과정을 반복적으로 수없이 수행했을 때 얻어지는 반복 결과들에 더 관심을 가진다.

예 2.4.1 조사 질문은 계통오차나 가변오차의 원인이 될 수 있다. 예를 들어 한 사람의 지난주 술 섭취량에 대해 묻는 질문을 생각해 보자. 응답자들은 그들이 정확히 섭취한 양을 생각해내기보다 그들의 섭취량을 추정할 것이다. 하지만 많은 응답자들은 창피하지 않기 위해 그들의 술 섭취량을 고의적으로 실제 이하로 응답할 것인데, 이러한 것을 연구문헌에서는 사회적 기대부응 편향 (*social desirability bias*) 이라고 한다. 이에 따라 모든 표본 구성원의 응답결과로 구한 평균 술 섭취량은 작은 쪽으로 편향 될 것이고, 일인당 평균 술 섭취량에 대한 추정값은 과소추정 될 것이다.

가변오차는 지난 6개월 동안 식료품점에 간 횟수같이 응답자가 정확한 응답을 하기 위해 스스로 추정과정을 거쳐야 하는 경우 일어난다. 지난 6개월 동안 식료품점에 간 횟수를 응답하기 위해 과거를 더듬어 실제 횟수를 세기보다는 많은 응답자들은 그 숫자를 추정하려고 어떤 방법을 쓸 것이다. 예를 들어 어떤 사람들은 그들이 보통 식료품점에 일주일에 두 번 정도 가니까 이 숫자에 대략 6개월에 해당하는 주의 수인 26을 곱해 계산할 것이다. 다른 사람들은 다른 추정방법을 사용할 수 있다. 결과는 어떤 응답자들은 실제보다 약간 작은 숫자로 응답할 것이고 어떤 경우 반대현상도 일어난다. 하지만 표본 전체를 생각하면 자료의 총계를 계산할 때 양과 음의 오차는 서로 상당부분 상쇄되기 때문에 식료품점에 간 평균 횟수는 실제 평균에 매우 가까울 수 있다.

계통오차와 가변오차가 추정결과에 미치는 영향

계통오차와 가변오차 둘 다 정확도를 떨어뜨리지만 어느 종류의 오차가 정확도에 더 해로운 것인가? 이 질문의 답은 무엇을 추정하나에 달려있다. 이미 언급했듯이 모집단 평균, 합계와 비율의 추정 같은 선형추정(*linear estimates*), 바꿔 말하면 표본의 관측값들의 합의 형태로 얻어지는 추정에서는 계통오차는 추정결과에 편향이 발생하게 하는데 반해서 가변오차는 서로를 상쇄하므로 흔히 편향이 생기지 않는다. 그러므로 선형추정에서는 편향 결과 때문에 계통오차가 가변오차보다 더 해로울 수 있다. 나중에 보겠지만 가변 비표본추출오차가 선형추정에 미치는 영향은 표본추출오차가 선형추정에 미치는 영향과 비슷하다. 즉 표본의 크기를 늘리면 표본추출오차와 가변 비표본추출오차를 둘 다 줄일 수 있다. 그러므로 가변오차가 선형추정에 미치는 영향을 줄여주는 한 가지 방법은 조사의 표본 크기를 늘리는 것이다. 하지만 표본 크기를 늘리는 것은 계통오차에는 아무런 도움이 되지 않는다. 앞에서 언급했듯이 표본크기가 늘어나면 조사의 수행 영역이 확대되고 비표본추출오차 원인의 관리가 어렵기 때문에 결과적으로 계통오차가 늘어날 수 있다.

상관계수 추정, 회귀계수 추정, 표준오차추정 등 같은 비선형 추정의 경우에는 어떤 종류의 오차가 더 해로운가 하는 질문의 답은 그리 간단하지 않다. 이런 종류의 추정에서는 계통오차와 가변오차 둘 다 편향을 발생시킬 수 있다. 예를 들어 가변오차가 존재하면 회귀계수의 추정값은 줄어든다는 것(즉, 0쪽으로 편향됨)을 증명할 수 있는데, 계통오차가 있는 경우에는 편향의 방향을 예측할 수 없다. 그러므로 비선형추정에서는 계통오차와 가변오차 사이에 우열을 가릴 수 없다. 그 이유는 이 책의 범위를 벗어난다. 이 주제에 관심이 있는 사람들은 Fuller(1987)과 Biemer와 Trewin(1997)을 참조하라.

많은 경우에 표본조사의 가장 중요한 목적은 목표 모집단의 평균, 총계와 비율을 산출하는 것이다(즉, 모집단에 대한 기술적인 연구). 이러한 이유로 총오차의 감소를 위한 조사 설계를 하는데 있어 계통오차의 주요 원인의 확인과 제거를 우선적으로 고려한다. 조사 설계의 목적은 두 가지 종류의 오차를 최소화하는 것이지만 조사 설계자는 흔히 어떤 종류의 오차가 가장 해로운가를 먼저 판단하고 다른 종류의 오차는 무시하거나 가볍게 관리하면서 보다 해로운 것을 주로 관리해야 한다.

예를 들어 설계자는 더 많은 예산을 설문지를 더욱 정교하게 만드는 것에 더 투자할 것인지 아니면 면접원 교육에 더 많은 배분을 하는 것이 효율적인지를 판단해야 한다. 또는 주어진 예산을 자료 입력 오차를 줄이기 위해 더 광범위한 품질관리

검사를 하는 데 쓰는 것 보다 조사 무응답자의 추적에 더 많은 조사 예산을 사용하는 것이 더 나은지 여부를 판단해야 할 것이다. 이런 상황에서는 계통오차나 가변오차를 가장 많이 발생시키는 오차 원인이 어떤 것인지를 판단할 수 있는 아이디어를 가지고 있는 것이 바람직하다.

대부분의 상황에서는 계통오차의 원인을 제거하는 것이 가변오차를 발생시키는 오차원인의 제거보다 우선적으로 고려되어야 한다. 이것에 대한 명확한 규칙은 없지만, 그래도 조사과정의 오차원인에 대한 다음 논의에서 다루어지는 각 오차원인 때문에 발생하는 계통오차의 위험을 이해하는 것이 유용할 것이다.

계통오차와 가변오차를 발생 시키는 오차원인

어떤 오차원인은 주로 계통적인 오차를 발생 시킨다. 예를 들어 무응답 오차는 가변오차라기보다 오히려 계통오차이다. 왜냐하면 특정한 추정값에 무응답 오차가 미치는 누적적인 영향은 추정결과를 편향되게 하기 때문이다. 3장에서 보겠지만 편향의 크기는 무응답율과 응답자와 무응답자의 특성상의 차이에 따라 달라진다. 추출틀에 모집단 구성원 중 일부 특정한 구성원이 빠지는 경우(예를 들면 농업조사 추출틀에서 작은 농장은 빠지는 경우)와 같이 불완전한 추출틀을 사용함으로써 조사과정의 반복 수행을 기준으로 조사결과는 어느 한쪽으로 잘못되는 경향이 있게 된다.

어떤 오차원인은 주로 가변적인 오차를 발생시킨다. 예를 들어 면 자료를 입력하는 사람들이 범하는 오차는 우연이거나 아무런 방향성이 없다고 볼 수 있기 때문에, 자료 입력에 따른 오차는 전형적인 가변오차이다. 자료를 입력하는 사람들이 추정값을 늘리거나 또는 줄이려는 경향을 갖는 경우는 흔치 않은 일이다. 오히려 이런 종류의 오차는 대부분 서로를 상쇄되는 경향이 있다. 앞에서 사용한 식료품점에 몇 번 가는가를 묻는 것 같은 응답을 위해 응답자가 어떤 추측과정을 거치는 예에서도 어떤 응답자들은 높게 추측할지 모르지만 다른 사람들은 낮게 추측할 수 있고, 따라서 결국은 그 추측의 평균은 거의 0에 가까울 지도 모른다.

다른 오차원인은 가변오차와 계통오차 둘 다 발생 시킬 수도 있다. 예를 들어 면접자가 범하는 오차는 가변오차와 계통오차를 동시에 야기할 수 있다. 소득문제를 다시 생각해보자. 어떤 면접원들은 그들의 태도, 복장, 면접 전에 하는 말 등으로 응답자가 실제 소득을 부풀려 응답하게 유도할 수도 있고, 다른 면접원의 경우 반

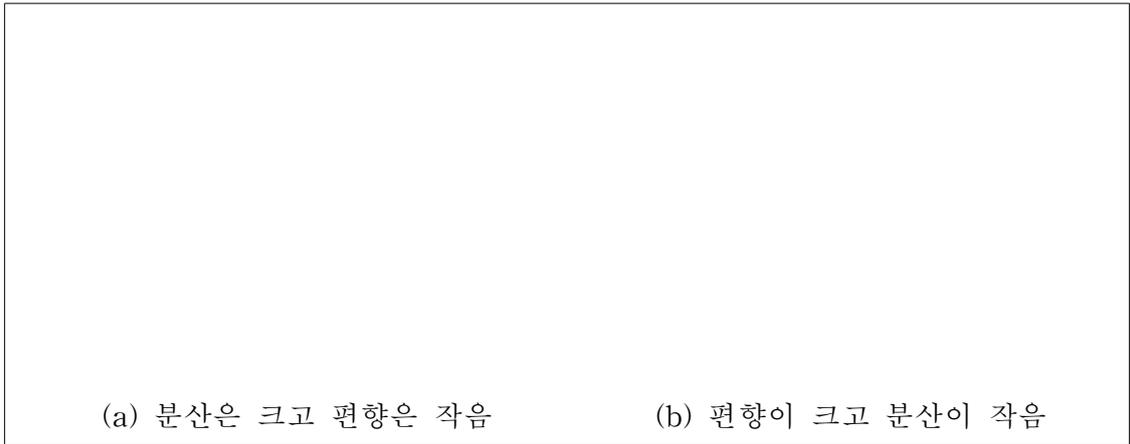
대의 영향을 줄 수도 있다. 또한 특히 저소득층의 응답자들은 그들의 소득을 부풀려 실제보다 더 유복하게 보이고 싶어 하는 경향이 강하다. 그러므로 종합적으로(조사 과정의 많은 반복수행을 고려할 때) 소득을 과장해서 응답하는 경향이 있으며, 면접원은 가변오차를 발생시키는 요인이기 때문에 소득 자료는 일반적으로 편향되고 신뢰성이 떨어진다.

다음 절에서는 평균제곱오차를 이용하여 조사의 가변오차와 계통오차가 추정결과에 미치는 영향을 종합 요약하는 방법을 생각한다.

2.5 평균제곱오차

사격과 표적에 비유한 설명

총조사오차가 어떻게 계통오차와 가변오차 또는 편향과 분산으로 분해 될 수 있는지를 잘 이해하기 위해 <그림 2.5>를 생각해 보자. 이 그림에서 조사를 이용하여 모수를 추정하는 과정을 설명하기 위해 표적에 총을 쏘는 과정을 생각해 보자. 표적의 중앙은 우리가 조사 자료를 가지고 추정하기를 바라는 모수를 나타낸다. 물론 실제로는 조사로 추정하고 싶은 모수가 많을 수도 있지만 우선은 모집단의 평균 소득 같은 특정한 하나의 모수만을 고려하자. 모수를 추정하기 위해 조사를 실시하는 것은 사격수가 표적의 정중앙을 조준하여 총을 쏘는 것과 유사하다. 만일 사격이 정확하면 표적의 정중앙을 맞추게 되고, 그렇지 않으면 어떤 거리를 두고 정중앙을 빗나가게 된다. 조사에서 나온 추정값과 모수 사이의 “거리”가 추정값의 총오차이고 이는 사격수의 사격에서 정중앙을 빗나감으로써 발생한 총오차와 유사한 개념이다.



<그림 2.5> 계통오차와 가변오차를 표적으로 표현한 것: 만일 이 표적들이 두 조사 설계의 오차를 나타낸다면 어느 조사 설계를 선택하겠습니까? (a)의 조사 설계는 분산이 크고 편향이 작은 추정결과를 산출하고, (b)의 조사 설계는 분산이 작고 편향이 큰 추정결과를 산출한다.

이제 사격수가 매번 정중앙을 조준하여 반복해서 총을 쏜다고 가정하자. 그것은 매번 모수를 추정하려고 동일한 상황에서 조사 과정을 몇 번 반복하는 것과 유사하다. 실제로는 모수를 추정하기 위해 조사 과정을 단지 한번 수행한다. 하지만 표적에 나타난 탄흔의 패턴이 사격수의 조준상의 문제점에 대해 무언가를 알려주듯이, 만일 조사를 여러 번 반복한다면 조사 추정값들에서 나타나는 변동은 조사과정의 총오차에 대한 정보를 제공해 준다.

사격수의 조준처럼 조사오차는 계통오차와 가변오차 요소로 이루어져 있다. 예를 들어 만일 사격수의 조준이 바르지 않다면 아마도 명중하지 않을 것이다. 만일 사격수의 조준이 일관성이 있지만 정조준은 못 한다면 각 탄흔은 한쪽으로 정중앙까지의 거리가 일정하게 나타날 것이다. 사격수의 조준 불량은 편향된 조사과정과 유사하다. 이런 조사과정은 그것을 수행할 때마다 매우 일관성 있는 결과를 산출할지 모른다. 하지만 추정값들은 모두 모수와 대략 동일한 방향으로 동일한 양만큼 차이가 날 것이다.

조준 불량 뿐 아니라 몇몇 다른 이유로 사격수는 정중앙을 놓칠지 모른다. 예를 들어 사격수의 조준이 일관성이 없다면 쏠 때 마다 표적의 정중앙을 예측할 수 없는 랜덤한 거리만큼 빗나갈 것이다. 때로는 정중앙의 왼쪽에 맞을 수도 있고 때로는 오른쪽으로 방향을 바꾸어 맞거나 정중앙의 위나 아래에 맞을 수도 있다. 바람

이나 날씨, 총구의 상태와 총기 자체 같은 다른 요소가 사격의 정확도에 예측할 수 없는 랜덤한 영향을 미칠 수도 있다. 이런 요소들은 조사에서의 가변오차와 유사하다. 조사 과정이 반복 수행될 때마다 수많은 요소들 때문에 생기는 랜덤 변동이 추정값의 정확도에 영향을 미칠 수 있고 총오차(탄흔과 정중앙과의 거리)에 반영된다.

<그림 2.5>에 있는 두 표적이 다른 두 개의 총기를 사용하는 두 명의 사격수에 상응한다고 하자. 오른쪽에 있는 표적의 탄흔 패턴은 계통오차가 문제가 되고 있는 것을 나타낸다는 점에 유의하라. 즉 바꿔 말하면 총기나 사격과정상에 총알이 의도한 지점에서 한쪽 방향으로 치우쳐 맞게 하는 어떤 영향이 근본적으로 존재한다는 것을 나타낸다. 왼쪽 표적의 패턴은 계통오차는 더 작지만 가변오차가 문제가 되는 경우이다. 즉, 사격과정과 관련된 많은 요소들의 누적된 영향이 사격수가 마치 랜덤한 방법으로 사격한 것과 유사한 패턴을 보여주고 있다.

이 표적들은 조사과정의 계통오차(편향)와 가변오차(분산)의 이해를 돕는데 쓰일 수 있을 뿐만 아니라 그것들을 측정하는 방법을 이해하는데도 도움이 된다. 동일한 조사과정을 똑같은 조건에서 여러 번 반복한다고 가정하자. 즉, 같은 표본추출 절차를 쓰고 (하지만 각 반복 실험에서 다른 표본이 추출됨), 같은 설문지, 동일한 조사원을 고용하여 같은 과정의 교육을 시키고, 같은 자료 수집 절차 등을 쓴다고 하자. 조사의 각 반복실험 결과는 표적상에 나타난 하나의 탄흔과 같이 모수에 대한 하나의 추정값을 의미한다. 오른쪽에 있는 표적은 모수를 추정하기 위한 어떤 조사과정과 상응하고 왼쪽에 있는 표적은 같은 모수를 추정하기 위한 또 하나의 조사과정을 나타낸다. 그러므로 그 두 조사 과정은 서로 다른 계통오차와 가변오차를 발생시킨다. 오른쪽 표적과 상응하는 조사에는 상당한 계통오차가 있지만 분산(가변오차)은 비교적 작다. 오른쪽 표적과 상응하는 조사에서는 큰 가변오차가 있지만 계통오차는 매우 작다.

이제 두 조사 과정이 거의 같은 비용이 든다고 가정하자. 모수를 추정하는데 어느 조사과정(즉 왼쪽 표적에 상응하는 조사 아니면 오른쪽 표적에 상응하는 조사)을 선택해야 하는가? 이 결정을 내릴 때에 계통오차와 가변오차의 영향을 하나로 결합시켜서 두 조사의 정확도를 정량화하는 어떤 방법, 바꿔 말하면 총오차를 하나의 숫자로 요약하는 방법이 있으면 매우 도움이 될 것이다. 그러면 더 작은 수준의 총조사오차가 발생하는 것이 보다 효율적인 조사과정일 것이다.

평균제곱오차를 사용한 추정량의 총오차 요약

특정 패턴을 발생하는 과정상의 오차 또는 보다 더 일반적으로 <그림 2.5>에서 표현한 조사과정의 총오차를 요약하는 많은 방법들이 있다. 하지만 바람직한 통계적인 성질 때문에 통계 연구 논문에서 보편적으로 쓰이는 것은 근접성을 탄흔의 위치와 표적의 정중앙 사이의 거리의 제곱으로 정의하는, 즉 탄흔에서 중앙까지의 평균 근접성을 나타내는 척도인 평균제곱오차(*mean squared error*)이다. 평균제곱오차를 계산하기 위해 각 탄흔과 표적의 정중앙사이의 거리를 측정하여 그 거리를 제곱하고, 이 제곱한 거리들의 평균값을 구한다.

예를 들어, 조사과정 A와 조사과정 B가 똑같은 조건에서 10번 반복 수행된 상황을 생각해 보자. <그림 2.5>는 이 20번의 수행에 대하여 조사과정 A(왼쪽 표적)와 조사과정 B(오른쪽 표적)의 추정값 오차를 보여준다. 오른쪽의 10개의 탄흔들에 대한 제곱한 거리의 평균이 0.15이고 왼쪽의 평균이 4.5이다. 그러므로 추정값의 평균 제곱오차가 더 작기 때문에 오른쪽 결과를 산출한 조사과정이 더 좋다.

사격에서 평균제곱오차(*mean squared error*: MSE)는 표적의 탄흔과 정중앙 사이의 거리의 제곱의 평균이다. 조사 추정량의 MSE는 조사과정의 가상의 반복수행에서 나오는 추정값(표적의 탄흔에 해당)과 모수(정중앙에 해당)의 차이의 제곱의 평균이다.

이 설명을 더 구체적으로 조사와 관련지어 설명하기 위해 오른쪽은 면접원이 응답자한테서 자료를 수집하는 조사방법(면접원 보조 면접법)을 나타내고 왼쪽은 같은 자료를 응답자가 직접 설문지에 기록해서 자료를 수집하는 조사방법(자기기입식 조사방법)을 나타낸다고 가정하자. 분산의 차이가 비표본추출 가변오차에 의해서 발생토록 하기 위하여 두 조사과정 모두 같은 표본 크기를 기준으로 한다고 가정하자. 면접원이 응답자에게 미치는 영향 때문에 면접원이 보조한 조사방법의 편향이 더 크다고 가정하자. 하지만 면접원이 준 도움 때문에 응답자의 이해가 증진된 결과 분산이 더 작아지고, 자기기입 방법이 면접원의 영향 때문에 생기는 편향을 제거하지만 응답자가 질문을 잘못 해석하는 결과로 인해 더 큰 가변오차를 낸다고 가

정하자. 그러므로 면접원이 보조한 조사의 오차는 <그림 2.5>의 오른쪽과 유사하고 자기기입법의 오차는 왼쪽과 유사하다. 이 예는 동일한 정보를 수집하는 것을 목표로 서로 다른 조사방법을 적용한 경우 두 조사과정이 편향과 분산이란 관점에서는 어떤 특성상의 차이가 있는지를 보여주고 있다.

위에서 정의한 평균제곱오차의 주요한 특징은 같은 조건 하에서 같은 모집단을 대상으로 동일한 조사과정이 여러 번 반복된다는 개념을 토대로 한다. 그러면 평균제곱오차는 과정의 각 반복 실험의 추정값과 모수의 차이의 제곱의 평균이다. 이 정의에는 두 가지 문제가 있다. 우선 매번 똑같은 조건 하에서 조사과정을 반복하는 것은 보통 불가능하다. 우리가 사는 세상은 동적이고 늘 변화하고, 조사 조건도 한 수행에서 다음까지 많이 바뀔 수 있다. 많은 요소(날씨, 정치 등)가 조사의 결과를 좌우하고 이 요소들은 시간에 따라 바뀐다. 모수 자체도 시간이 흐름에 따라 바뀔 수 있다. 게다가 MSE를 추정하는 목적으로 조사 과정을 여러 번, 또는 단지 두 번이라도 반복 수행하는 것은 비용측면에서 결코 현실적이지 않다. MSE의 요소들을 추정하는 더 효과적이고 효율적인 방법들이 있는데 이것은 8장에서 다룬다.

조사 추정의 MSE를 평가하는데 또 하나의 어려움은 추정의 오차를 정량화 할 수 있게 목표 모집단 모수의 참값을 파악하는 작업이다. 추정값이 매우 정확하게 산출 되도록 특수한 과정을 추가적으로 적절하게 수행하는 경우를 제외하고는 단 한번의 조사를 실시하는 것만으로는 정확한 모수를 추정하는 것은 불가능하다. 어떤 특성에 대한 모수의 참값을 파악하는 일반적인 방법들은 매우 정확한 행정자료를 활용하거나, 원래의 조사과정의 고유한 오차를 대부분 제거할 수 있는 비용이 많이 들고 추가적인 노력을 필요로 하는 특별한 측정 방법을 활용하는 것이다. 이런 방법들은 보통 원래의 조사 자료를 수집한 후 실시한다. 8장에서 이 방법들을 상세하게 다룬다.

MSE의 추정과정은 보통 복잡하고 비용이 많이 드는 과정이므로 총 MSE는 실제로는 거의 추정되지 않는다. 추정을 한다면 흔히 실제 MSE의 대략의 근사치 정도를 파악하는 수준이다. 또한 MSE를 구성하는 요소 중 가장 중요한 요소 몇 개만을 추정한다. 예를 들어 MSE를 계산하는 대표적인 접근 방법은 계통오차와 가변오차의 중요한 발생 원인과 관련된 몇 개의 편향 또는 분산 성분을 추정한 다음 평균제곱오차를 계산하기 위해 이런 오차 성분들을 결합시킨다. 앞에 설명한 계산 공식을 이용하여 이런 방식으로 평균제곱오차를 계산하는 것은 쉽지 않다. 다음 절에서는

이러한 결점을 보완하는 평균제공오차를 계산하는 또 다른 방법을 설명한다.

편향과 분산 성분에 의한 평균제공오차의 분해

조사 추정에서의 총오차를 보는 다른 시각을 논의하기 위해 <그림 2.5>로 다시 돌아가자. 각각의 표적에서 탄흔들은 표적에 집락을 형성하고 있다. 우선 오른쪽에 있는 표적의 탄흔 또는 추정값들의 대략적인 중앙(또는 중심)에 해당하는 점을 찾아내라. 집락의 중앙에 있는 점으로 조사과정의 평균 오차를 설명할 수 있다. 탄흔으로 구성된 집락의 중앙과 표적의 정중앙 사이의 거리가 조사과정의 오차 또는 편향이다. 더 나아가서 각각의 탄흔들과 집락 중앙 사이의 거리들은 조사 과정의 가변오차를 나타낸다.

그러므로 같은 조건에서 조사과정을 여러 번 반복해서 추정값의 평균을 계산하고, 이 추정값들의 평균에서 모수의 참값을 빼서 추정량의 편향을 계산할 수 있다. 조사에서 추정량의 분산은 반복된 각 조사 추정값과 평균 조사 추정값 사이의 거리의 제곱의 평균이다.

이런 개념을 중심으로 조사 추정량의 평균제공오차를 계산하는 2가지 방법을 제안한다.

[방법 1]

MSE = 탄흔과 표적의 정중앙사이의 거리의 제곱의 평균

[조사 용어로 설명하면]

= 여러 번 반복된 조사에서 나온 추정값과 모수의 참값 사이의 거리의 제곱의 평균

[방법 2]

MSE = 탄흔들의 중앙과 표적의 정중앙사이의 거리의 제곱

+ 탄흔들의 중앙에서 탄흔들까지의 거리의 제곱의 평균

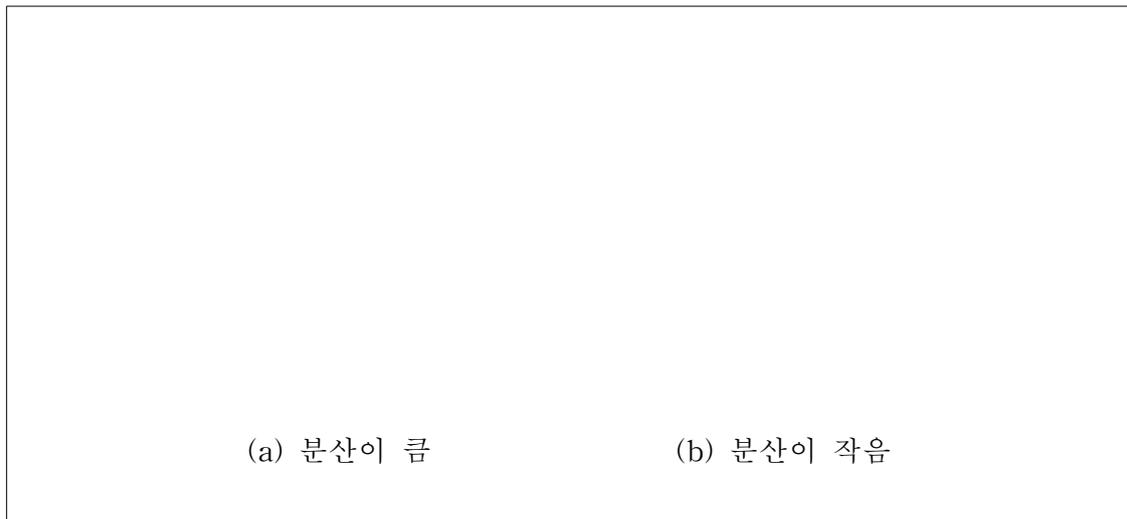
[조사 용어로 설명하면]

= 조사과정의 반복수행에 의한 추정값들의 평균과 모수의 참값사이의 거리의 제곱 + 반복수행에서 얻어진 추정값들과 추정값들의 평균과의 차이의 제곱의 평균

두 방법 어느 것을 사용해도 모두 같은 평균제곱오차가 계산된다. 하지만 방법 2가 평균제곱오차를 편향제곱과 분산 성분으로 분해하여 계산한다는 측면에서 선호된다. 편향의 제곱(*squared bias*)은 조사과정의 반복 수행을 통해 얻어진 추정값들의 평균과 모수 사이의 거리의 제곱이다. 분산(*variance*)은 반복 수행에서 얻어진 추정값들과 반복된 추정값들의 평균의 차이의 제곱의 평균이다. 그러므로 방법 2의 공식은 다음과 같다.

$$\text{MSE} = \text{편향의 제곱} + \text{분산} \quad (2.1)$$

추정값의 총 MSE는 편향의 제곱과 분산을 합한 것과 같다는 것을 이 공식을 통해 알 수 있다. 만약 추정량의 편향이 0이라면, MSE는 단순히 추정량의 분산이다. 하지만 편향이 0이 아니면 MSE의 추정값을 계산하기 위해서 편향을 추정해야 한다. 편향의 계산은 모수의 참값에 대한 지식을 필요로 한다. 그렇지만 분산은 모수의 참값을 알지 않고도 계산할 수 있다.



<그림 2.6> 진실은 알 수 없다: 모수의 참값을 모를 때 더 좋은 조사과정을 결정하는 것은 표적의 정중앙을 모르고 두 사격수의 정확도를 평가하는 것과 같다. 조사 추정값에 대해서 알려진 것은 (a)의 조사 설계는 분산이 크고 (b)의 조사 설계는 분산이 작다는 것뿐이다.

이런 상황을 명확히 보기 위해서 <그림 2.5>에 있는 표적의 탄흔들만을 남기고 표적지는 삭제한 <그림 2.6>을 생각해 보자. 여기서도 각 탄흔에서 탄흔들의 평균이라 할 수 있는 탄흔들의 중앙까지의 거리의 제곱인 분산은 계산할 수 있다는 점에

주목해라. 하지만 표적의 정중앙 또는 모수의 참값을 알 수 없기 때문에 편향은 계산할 수 없다. 이 그림은 또한 분산만을 기준으로 두 조사과정 중에서 적절한 것을 선택하는 것은 잘못이라는 것을 설명하고 있다. 분산만을 알기 때문에 그 설계의 편향과는 관계없이 오른쪽에 있는 조사과정이 왼쪽에 있는 조사과정보다 우수한 것으로 판단하고 항상 이것을 선택하게 된다는 점에 유의하기 바란다.

MSE를 계산하는 두 방법 모두가 같은 것을 보여주기 위하여 <표 2.3>과 <표 2.4>에 있는 자료를 생각해보자. <표 2.3>에서는 [방법 1]을 써서 <그림 2.5>에 있는 두 조사 설계의 MSE를 계산한 것이다. <그림 2.5>에 있는 각 탄흔에서 표적의 정중앙과의 거리를 측정하고 이 거리들을 제공했다. 각 표적의 거리 제공의 평균이 조사과정의 MSE이다. 왼쪽에 있는 조사과정에서는 이 평균이 21.8이고 오른쪽에 있는 과정에서는 21.4이다(<표 2.3> 참조).

<표 2.3> [방법 1]^a을 사용한 평균제공오차의 계산

왼쪽 표적			오른쪽 표적		
탄흔	탄흔과 표적의 정중앙의 거리	표적의 정중앙까지 거리의 제공	탄흔	탄흔과 표적의 정중앙의 거리	표적의 정중앙까지 거리의 제공
1	2.2	4.8	1	3.1	9.6
2	-3.6	13.0	2	3.7	13.7
3	-4.5	20.3	3	5.3	28.1
4	6.8	46.2	4	4.9	24.0
5	5.1	26.0	5	6.1	37.2
6	-7.2	51.8	6	4.4	19.4
7	-3.9	15.2	7	2.8	7.8
8	5.3	28.1	8	6.1	37.2
9	-1.8	3.2	9	4.5	20.3
10	3.1	9.6	10	4.1	16.8
평균 또는 중앙	0.15	21.8 (=MSE)	평균 또는 중앙	4.5	21.4 (=MSE)

a. 평균제공오차는 탄흔(또는 추정값)에서 정중앙(또는 모수)까지의 거리 제공의 평균이다. [방법 1]에서는 각 탄흔마다 탄흔에서 정중앙까지의 거리(또는 오차)를 계산하고, 그 결과를 제공하고, 10개 탄흔 전체에 걸쳐 평균을 구한 것이다.

<표 2.4>에서는 [방법 2]를 써서 MSE를 계산했다. 이 방법에서는 탄흔들의 중앙을 찾아내서 각 탄흔들과 탄흔들의 중앙 사이의 거리를 측정하고 그 결과를 제공한다

다. 이 제공한 거리들의 평균이(즉, 왼쪽 표적은 21.81 그리고 오른쪽 표적은 1.17) 추정값들의 분산이다. 왼쪽의 분산이 오른쪽의 분산보다 몇 배 더 크다는 것에 유의해라. 또한 탄흔들의 중앙과 표적의 정중앙까지의 거리를 측정하면 왼쪽 표적은 0.15 그리고 오른쪽 표적은 4.5가 나온다. 이미 보았듯이 오른쪽 표적이 왼쪽 표적보다 훨씬 더 큰 편향을 갖는다. 분산과 편향 요소를 함께 합치면 $MSE(\text{왼쪽 표적})=(0.15)^2+21.81=21.83$ 이고 $MSE(\text{오른쪽 표적})=(4.5)^2+1.17=21.42$ 이다. 이것은 <표 2.3>에서 계산한 MSE들과 일치한다. 이와 같이 MSE를 계산하는 두 가지 방법에 서 얻어지는 결과는 같다는 것은 이론적으로 증명될 수 있다.

<표 2.4> [방법 2]^a를 이용한 평균제공오차의 계산

왼쪽 표적			오른쪽 표적		
탄흔	탄흔과 탄흔들의 중앙까지의 거리	중앙까지 거리의 제공	탄흔	탄흔과 탄흔들의 중앙까지의 거리	중앙까지 거리의 제공
1	(2.2-0.15)=2.05	4.20	1	(3.1-4.5)=-1.40	1.96
2	(-3.6-0.15)=-3.75	14.06	2	(3.7-4.5)=-0.80	0.64
3	-4.65	21.62	3	0.80	0.64
4	6.65	44.22	4	0.40	0.16
5	4.95	24.50	5	1.60	2.56
6	-7.35	54.02	6	-0.10	0.01
7	-4.05	16.40	7	-1.70	2.89
8	5.15	26.52	8	1.60	2.56
9	-1.95	3.80	9	0.00	0.00
10	2.95	8.70	10	-0.40	0.16
평균	0.0	21.81 (=분산)	평균	0.0	1.17 (=분산)
편의=(0.15-0.0)=0.15	편의 ² =0.023		편의=(4.5-0.0)=4.5	편의 ² =20.25	
MSE = 편의 ² +분산 = 21.8			MSE = 편의 ² +분산 = 21.4		

^a평균제공오차는 두 단계로 계산할 수 있다. 먼저 각 탄흔에서 탄흔들의 중앙까지의 거리를 측정하고 그 결과를 제공한다. 10개의 탄흔 전체에 걸쳐 이 제공거리들의 평균을 구한다. 마지막으로 탄흔들의 중앙과 표적의 정중앙사이의 거리를 계산하고 그 결과를 제공한다. MSE는 이 두 값의 합이다. 이는 <표 2.3>의 결과와 일치하는 것에 유의하고, 두 방법이 같은 결과를 산출한다는 것을 볼 수 있다.

MSE의 주요 요소들

<표 2.2>에 있는 각 오차의 원인들은 총 MSE를 구성하는 편향과 분산 두 요소

에 영향을 준다. 하지만 어떤 오차 원인들은 편향에 더 많은 영향을 주고 어떤 것은 분산에 그리고 어떤 오차원인은 편향과 분산 둘 다에 크게 영향을 줄 수 있다. <표 2.5>는 각 주요 오차원인과 각각의 가변오차와 계통오차의 위험성에 대한 평가결과를 보여준다. 이런 위험성은 조사 설계와 조사할 모집단의 특성에 따라 좌우된다. 하지만 어떤 오차원인이 계통오차나 가변오차의 발생에 더 영향이 큰 것인지를 파악할 수 있는 명확한 규칙은 없다. 그럼에도 불구하고 <표 2.5>는 일반적인 조사에서 위험성 발생 상황을 보여 주고 있다.

<표 2.5> 주요 오차 원인에 따른 가변오차와 계통오차의 위험성

MSE 요소	가변 오차의 위험	계통오차의 위험
특정화 오차	낮음	높음
추출틀 오차	낮음	높음
무응답 오차	낮음	높음
측정 오차	높음	높음
자료처리 오차	높음	높음
표본추출오차	높음	낮음

예를 들어 적절항 확률추출법을 적용하는 일반적인 조사에서 표본추출오차에서 생기는 편향의 위험은 매우 작고 표본 분산은 피할 수 없다. 반대로, 조사 질문의 특정화에서 기인한 오차는 가변오차보다 계통오차를 일으키기 때문에 특정화 오차 때에는 주로 편향이 발생한다고 볼 수 있다. 무응답오차는 분산보다 비표본 편향을 발생시킬 위험이 높다. 무응답률이 꽤 높을 때에는 무응답 조정방법을 적용함으로써 본질적으로 분산에 영향을 줄 수 있다. 추출틀 오차, 특히 모집단 구성원이 추출틀에서 빠져서 생기는 오차는 주로 편향을 발생시키는 원인인 된다. 나중에 이 책에서 다루겠지만 측정오차와 자료처리오차는 조사 추정값에 편향과 분산 모두에 악영향을 줄 수 있다.

<표 2.5>를 이용하여 식 (2.1)의 MSE 공식을 확장하는 것을 생각할 수 있다. 편향 제곱 성분에서 계통오차를 야기할 위험이 높은 표에 나오는 모든 오차원인의 편향 성분(즉, 특성화 편향, B_{SPEC} ; 무응답 편향, B_{NR} ; 추출틀 편향, B_{FR} ; 측정 편향, B_{MEAS} ; 자료처리편향, B_{DP})이 포함되게 확장 할 수 있다. 이들 편향 성분들이 모

두 합해져서 편향(*bias*)이라고 부르는 총 편향성분이 된다는 것에 유의해라. 마찬가지로 분산 성분은 표에 나오는 분산을 발생시키는 주요 원인이 되는 성분들(즉, 표본 분산, Var_{SAMP} ; 측정분산, Var_{MEAS} ; 자료처리 분산, Var_{DP})의 합으로 설명할 수 있다. 이와 같이 편향과 분산의 모든 주요 원인을 포함하는 확장된 의미의 MSE는 다음과 같다.

$$\begin{aligned} \text{MSE} &= \text{BIAS}^2 + \text{Variance} \\ &= (\text{B}_{\text{SPEC}} + \text{B}_{\text{NR}} + \text{B}_{\text{FR}} + \text{B}_{\text{MEAS}} + \text{B}_{\text{DP}})^2 \\ &\quad + \text{Var}_{\text{SAMP}} + \text{Var}_{\text{MEAS}} + \text{Var}_{\text{DP}} \end{aligned} \quad (2.2)$$

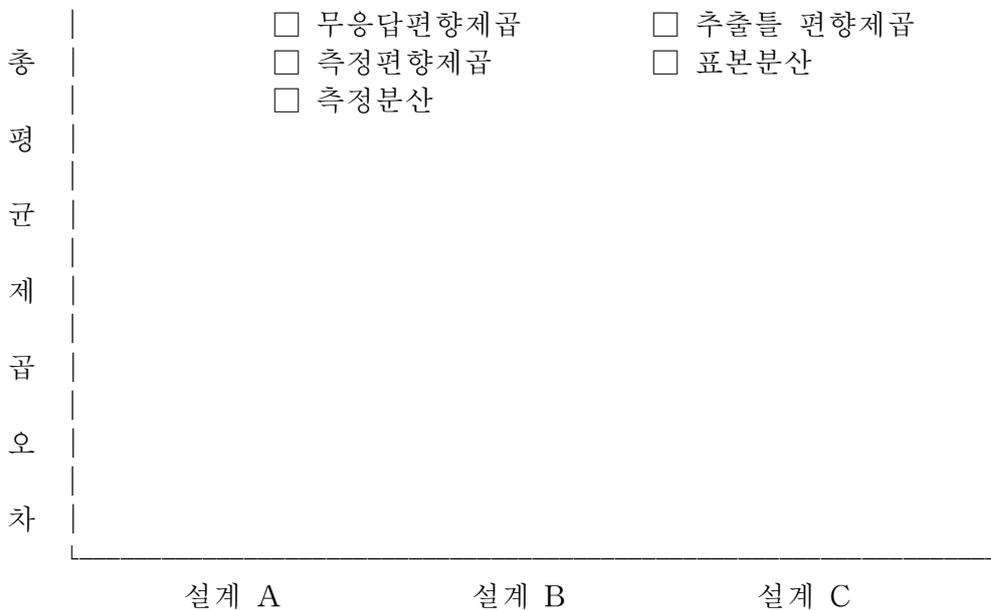
실제로 추정량의 MSE를 평가하는 한 방법은 식 (2.2)에 나타난 8가지 성분을 모두 추정하는 것이다. 마찬가지로 MSE를 줄이는 한 방법은 식 (2.2)에 있는 총 MSE를 구성하는 8가지 요소의 영향을 최소화하는 조사 설계를 개발하는 것이다. 이러한 조사 설계 접근법은 이 책 전체에 걸쳐 이야기하는 조사 설계의 기본적인 원리의 기반이 된다. 2.6 절에서는 여러 가능한 설계 중에 주어진 조사 예산으로 최선의 조사 설계를 찾아내는데 이런 접근방법이 어떻게 활용될 수 있는지를 보여준다.

2.6 개념의 해설

평균제곱오차가 조사 품질의 측정 도구로서의 유용함을 예를 들어 설명하기 위해 조사계획자가 A, B, C라고 부르는 동일한 조사목적을 만족시키기 위해 자료를 수집하는 세 개의 조사 설계 중에 하나를 선택해야 하는 시나리오를 생각해 보자. 설계 A는 대면면접으로 자료를 모으고, 설계 B는 전화면접으로, 설계 C는 자기기입법에 의한 우편조사로 자료를 수집한다. 각 처리 방법과 관련된 비용이 다르므로 각각의 총 자료 수집 비용이 같도록 설계들이 조정 되었다. 예를 들어 대면면접이 전화 또는 우편 자료수집보다 비용이 많이 들므로 같은 총 비용을 만족시키기 위해서 설계 A에서의 표본크기는 다른 두 설계의 표본크기보다 반드시 작아야 한다. 마찬가지로

설계 B의 일인당 면접 비용이 설계 C에서 보다 높으므로 설계 B에서의 표본크기가 설계 C의 표본크기보다 작아야 한다.

여기서는 각 설계의 다양한 조사오차의 원인과 관련된 편향과 분산에 관해서 거의 완전한 정보가 있다고 가정한다. 특히 무응답, 추출틀의 포함률, 측정 오차와 관련된 편향을 대략 안다고 하자. 설계 A가 가장 높은 응답률을 달성할 것으로 예상되기 때문에 무응답 때문에 생기는 편향이 이 설계에서 가장 낮을 것이라고 예상된다. 설계 C는 가장 낮은 응답률이 나올 것으로 예상하므로 무응답 편향이 가장 클 것이다.



<그림 2.7> 세 가지 조사 설계의 총평균제곱오차의 비교: 설계 A의 표본추출오차가 비록 가장 클지라도 총오차가 가장 작기 때문에 설계 A가 가장 바람직하다.

마찬가지로 설계 A에서는 지역 추출틀(*area frame*)을 이용한 표본추출 방법을 사용할 것인데, 이는 표본 지역에 있는 모든 주택 단위가 목록에 등재되고, 이들 모든 주택이 표본으로 추출될 수 있다는 것을 의미한다. 이러한 상세한 목록을 작성하고 여기서 표본을 추출하는 과정은 목표 모집단 모든 구성원에 대한 거의 완벽한 포함을 보증하므로 추출틀 편향은 0이 될 것이다. 그에 반해서 설계 B는 랜덤번호전화걸기(RDD) 추출틀을 사용할 것이다. 이 추출틀에서는 전화가 있는 모든 주

택 단위가 표본으로 선택될 가능성이 있게 하기 위하여 10자리수 전화번호를 랜덤하게 추출한다. 하지만 전화가 없는 주택 단위는 선택될 가능성이 없으므로 전화가 없는 주택 단위의 비포함 문제로 발생하는 작은 추출률 편향이 예상된다. 설계 C는 목표 모집단 구성원의 주소를 얻기 위해 결국 전화번호부 유형의 목록을 사용하므로, 전화가 없거나 목록에 올라있지 않은 전화번호를 가진 가구들이 이 추출률에서 빠질 것이다. 따라서 추출률 오차는 이 설계에서 가장 높을 것으로 예상된다.

세 가지 설계 모두에 존재하는 마지막 편향의 원인은 측정오차이다. 이미 언급했듯이 측정 편향은 설문지, 응답자, 면접원(설계 A와 B에서만), 조사여건과 자료수집 방법과 같은 많은 원인에 의해 발생한다. 계통오차의 원인들은 대면면접에서 덜 나타날 것으로 예상된다. 설계 B와 C에서는 더 큰 측정오차가 나타나게 될 것으로 예상된다. 하지만 그 두 설계에서의 측정오차 편향에 의한 영향은 총체적으로는 거의 비슷할 것으로 판단된다.

측정 분산과 관련해서는 면접원의 차이에 의해 발생하는 분산이 문제가 될 것으로 예상된다. 면접원 분산은 질문에 대한 응답자의 반응에 따른 면접원들의 태도, 응답자에 대한 그들의 예상, “만족스러운” 응답을 얻기 위한 일관성 없는 노력, 그리고 많은 다른 편향된 행동의 결과로 나타나는 면접원의 응답에 대한 영향력과 관련이 있다. 그 영향은 전화 면접에서 보다 대면면접에서 훨씬 클 것으로 예상되고 물론 우편조사에서는 존재하지 않는다. 그러므로 측정 분산은 설계 A에서 가장 높고 설계 C에서 가장 낮다.

마지막으로 표본 분산은 각 설계의 표본크기와 대략적으로 반비례할 것이다. 그러므로 표본크기가 가장 큰 설계 C의 표본 분산이 가장 작고, 설계 A는 비교적 작은 표본크기로 인하여 표본 분산이 가장 크다.

이런 정보가 주어졌을 때, 최선의 또는 최적의(*optimal*) 설계를 어떻게 선택해야 하는가? 각각의 설계에서 총 자료수집비용이 같도록 표본크기가 정해져 있고 자료수집을 완성하는데 필요한 시간은 중요한 기준이 아니라고 가정하면, 최선의 설계는 가장 작은 평균제곱오차를 갖는 것이다.

평균제곱오차를 계산하기 위해서는 <그림 2.7>에서처럼 간단히 제공된 편향과 분산 성분을 합하면 된다. 이 그림에서 보듯이 이 기준에 의한 최선의 설계는 표본크기가 가장 작은 설계 A이다. 만일 기준이 비표본추출오차는 무시하고 표본추출오차만을 최소로 하는 설계를 선택하는 것이었다면 설계 C가 최선이었을 것이라는

것을 알 수 있다. 하지만 비표본추출오차와 표본추출오차 성분을 함께 생각하면, 이 설계가 가장 비효율적인 것으로 평가된다.

<그림 2.7>로부터 다음과 같은 결론을 얻을 수 있다.

- 평균제곱오차는 조사 설계에 있는 모든 다양한 오차원인으로 발생하는 총 편향제곱에 분산을 합한 것이다.
- 비용, 적시성과 다른 품질 요소가 경쟁하는 설계에서 같을 때 최선의 설계는 가장 작은 평균제곱오차를 갖는 것이다.
- 비표본추출오차가 총조사오차에 미치는 영향은 표본추출오차보다 몇 배 더 클 수 있다.
- 조사 설계를 표본추출오차나 분산에만 근거해서 선택하는 것은 총 자료 품질 관점에서는 최선의 설계를 선택하는 올바른 방법이 아니다.

제 3 장 포함오차와 무응답오차

거의 모든 대규모 자료 수집에서는 미관측오차(*errors of nonobservation*)라고 부르는 결측 자료가 나타나는데 특히 이는 조사연구에서 문제점이 될 수 있다. 미관측오차는 두 가지 다른 이유로 일어날 수 있다:

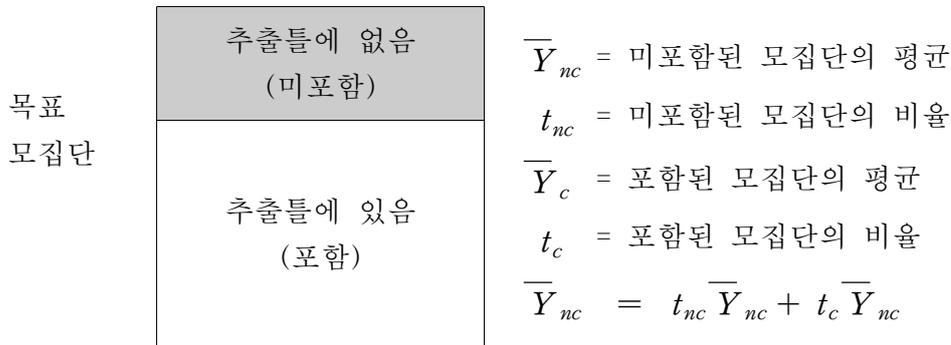
- 추출틀이 목표 모집단의 모든 단위를 포함하지 못하는 것(예를 들면 불완전한 목록, 오래된 지도 등)으로 이것은 추출틀 포함오차 (*frame coverage errors*)의 결과를 가져온다.
- 표본의 몇몇 단위의 측정값을 얻지 못하는 것으로 이 때문에 무응답오차 (*nonresponse error*)가 발생한다. 단위 무응답(*unit nonresponse*)은 한 단위에 대해 아무런 측정값도 얻지 못한 것인데, 항목무응답(*item nonresponse*)은 상당수의(대부분의) 변수에 대한 측정값은 얻었지만 모든 변수에 대한 측정값은 얻지 못한 상황을 가리킨다.

뒤에 언급하겠지만 포함오차와 무응답오차는 평균제곱오차에 거의 같은 종류의 영향을 미친다. 목표 모집단 추출틀에서 빠진 모집단 단위를 무응답의 한 유형으로 볼 수도 있다. 왜냐하면 두 경우 모두 이 단위에 대한 자료가 빠졌기 때문이다. 하지만 두 종류의 오차를 줄이는 데 이용할 수 있는 방법은 상당히 다르다.

이 장에서는 포함오차와 무응답오차를 검토한다. 각각의 오차의 영향에 관한 유용한 통찰력을 제공할 수 있는 간단한 모형을 생각해 본다. 흔히 포함오차와 무응답오차가 얼마나 큰지 우리는 전혀 모른다. 대부분의 경우 우리는 미관측의 정도(즉, 미포함률과 무응답률)에 대한 정보만 있고 어떤 경우에는 이 정보조차도 없다. 가정을 근거로 한 “가상적인 상황”에 따른 분석을 통해 오차 문제가 발생할 소지가 있는 모집단 특성을 확인하도록 한다. 단지 미관측율(즉, 무응답률 또는 미포함률)을 근거로 한 판단을 하지 않는 것이 중요하지만 불행하게도 이런 접근방식이 많은 조사자료 사용자들에게 흔한 관행이다. 나중에 보겠지만 미관측 유형에 따라, 미관측오차가 편향에 주는 영향은 미관측률과 결측된 단위들과 관측된 단위들의 특성상의 차이의 함수로 설명된다. 이 오차들을 방지하고 줄이는 몇몇 방법을 살펴보자.

3.1 포함오차

포함오차의 특성을 논의하기 전에 1장과 2장에서 다룬 몇 가지 개념을 복습하자. 자료의 수집과 추론의 대상이 되는 단위들의 집합을 목표 모집단(*target population*)이라고 부른다. 실제로 표본추출 대상이 되는 단위들의 집합은 추출틀 모집단(*frame population*)이라고 한다. 추출틀(*frame*)은 모집단 구성원을 정의하는데 쓰이는 요소들로 이루어져 있는데, 단위들의 목록, 지도, 또는 가능한 전화번호의 목록으로 이루어질 수 있다. 근본적인 포함 문제는 추출틀 모집단이 목표 모집단과 일치하는 정도와 관련된다. 이들 사이에 일대일의 대응이 성립 되지 않으면 잠재적으로 문제가 발생할 수 있다. <그림 3.1>을 보면, 큰 정사각형은 목표 모집단 전체를 나타낸다. 정사각형에서 어둡게 표시되지 않는 부분은 모집단 중 추출틀 목록에 실려 있는 부분을 나타내고, 어둡게 표시된 부분은 추출틀에서 누락된 부분을 나타낸다. 이 그림에 있는 수학적 부호는 뒤에 다시 거론하기로 한다.



<그림 3.1> 근본적인 포함 문제: 상자의 어둡게 표시되지 않은 부분만 표본 추출틀에 포함이 되어있다. 어둡게 표시된 부분은 포함되지 않았고 그러므로 표본에서 누락된다. 상자의 오른쪽에는 미포함 오차 때문에 생기는 편향을 구성하는 요소들을 정리한 것이다.

이 그림은 목표 모집단과 추출틀 모집단 사이의 일치도와 관련된 여러 가지의 상황을 설명한다. 이상적인 경우는 물론 추출틀과 목표 모집단 단위사이 일대일의 대응이 성립하는 경우이다(즉, 각 목표단위가 추출틀에 단 한번만 포함되어 있고 추출틀은 목표 모집단에 없는 단위는 포함하지 않는다). 이런 경우가 그림에서 어둡게 표시 되지 않은 부분에 해당한다. 이상적인 경우에서 벗어나는 상황은 여러 가지 이유 때문에 발생할 수 있다.

- 첫 번째로 목표모집단의 단위가 추출틀 모집단에서 빠질 수 있다. 따라서 추출틀은 다양한 유형의 누락 때문에 불완전해진다. 이 경우를 추출틀 과소포함(*frame undercoverage*) 또는 미포함(*noncoverage*)라고 부르고 이것이 주요한 포함상의 문제가 된다. 누락된 단위들은 표본에 포함 될 확률이 사실상 0이다. 보통 미포함의 정도는 지역에 따라 다르고 모집단 그룹 사이에서도 차이가 있다. (센서스에서는 개인, 농장, 사업체 등의 모집단 구성원 전부를 계상하지 못하고 실제보다 적게 계상하는 것이다.)

예 3.1.1 추출틀 미포함의 한 예는 추출틀 모집단을 모든 가능한 전화번호로 정의하고 목표 모집단은 모든 가구로 정의했을 때 일어난다. 미국 가구의 6%정도가 전화가 없기 때문에 조사를 오직 전화로만 실시하면 그런 가구는 누락되게 된다. 스웨덴에서는 유사한 비율이 단지 1%이지만 몇몇 다른 나라에선 그 비율이 50% 또는 그 이상이 될 수 있다. 미포함률이 늘어나면 전화 추출틀을 전화 없는 가구를 포함하는 어떤 다른 추출틀과 병합하여 사용해야 한다. 개발도상국에서는 전화 없는 가구 비율이 너무 커서 잠재적인 포함 편향 때문에 전화조사를 실행할 수 없다.

- 이상적인 경우에서 벗어나는 다른 경우는 추출틀 모집단에 있는 몇몇 단위가 목표 모집단의 구성원이 아닐 때 일어난다(즉 그들은 부적격이거나 조사 범위를 벗어나는 경우이다). 이런 단위들은 조사 연구원에게 관심의 대상이 아니므로, 표본을 추출하기 전에 확인하고 추출틀에서 제거해야 한다. 그러나 이런 단위를 미리 관찰하지 않고는 조사에 대한 적격성을 결정하는 것은 불가능하다. 이 경우에는 자료 수집할 때 부적격한 단위를 확인해서 조사 자료에 편향이 발생하지 않도록 제거해야 한다. 최악의 경우는 조사의 범위를 벗어난 몇몇 단위가 적격한 단위로 잘못 판단되고, 따라서 목표 모집단과 다른 특성을 가진 단위들이 잘못 포함된 만큼 이러한 조사 결과를 근거로 한 추정값들이 편향될 수 있다.
- 추출틀의 세 번째 문제는 추출틀상의 두 개 이상의 단위가 목표 모집단에서 동일한 단위와 대응이 될 때 일어난다(즉, 추출틀에 중복된 단위가 존재). 이러한 경우를 과대포함(*overcoverage*)이라고 부른다. 물론 이 문제에 대한 명백한

개선책은 추출틀에서 중복을 확인해서 표본추출하기 전에 제거하는 것이다. 만약 가능하다면 표본에 있는 단위가 추출틀에 등재된 횟수에 대한 정보를 자료 수집 과정에서 확보하는 것이 필요하다. 예를 들어 RDD 전화조사에서 면접원이 가구에 연결된 음성 전화로 사용되는 전화번호의 회선수에 대한 정보를 수집하는 것이다. 이런 정보는 가중값을 계산하는데 사용되는 선택 확률을 조정하여 과대표함에 따른 추정값 편향의 보정을 위해 추정단계에서 사용할 수 있다.

- 이상적인 경우에서 벗어나는 마지막 경우는 추출틀에 있는 하나의 단위가 여러 목표 모집단 단위와 대응 될 때 일어난다. 이것의 한 예는 추출틀이 거주 단위의 주소로 이루어져 있고 목표 모집단은 개인으로 이루어져 있을 때이다. 이 경우 각 거주 단위의 주소에 한 명이나 그 이상의 거주자가 대응이 된다. 여기서 거주 단위에서 어떤 사람들을 표본으로 추출할 것인지 결정하려면 중복과 누락 없이 거주자들의 명부를 작성하는 것이 필요하다. 하지만 가구 구성원의 프라이버시나 거주자 중 조사 적격자의 판별 등이 이 과정에서 문제가 될 수 있다. 예를 들어 조사 시점에 거주지에 머물고 있는 사람들 또는 거주지에 없는 사람들이 실제로는 이 거주 단위에 살고 있거나 그들 자신의 다른 거처를 갖고 있을 수 있다. 또한 응답에 따라서는 어떤 사람이 가구원에 해당하는지 여부를 결정해야 한다.

추출틀 미포함(*frame noncoverage*) 때문에 어떤 목표 모집단 단위에서는 선택 확률이 0인 결과를 가져오고, 다양한 형태의 과대표함(추출틀 중복) 때문에 의도된 것 보다 더 큰 선택 확률이 적용될 수 있다. 만일 적절히 보정이 이루어지지 않으면 이 두 문제 모두가 추정값이 편향되게 한다.

추출틀과 목표 모집단 사이의 대응 문제만이 포함과 관련된 문제는 아니다. 추출틀은 추출틀에 포함된 유용한 정보의 양과 질에 따라 그 효율성에 차이가 난다. 이런 추출틀에 포함된 정보는 단위를 확인하고, 분류하고, 접촉하고, 연계하는 데 활용될 수 있다.

표본추출틀은 행정상, 재정상 목적으로 생산된 사용 가능한 많은 정보를 기초로 작성될 수 있다. 어떤 나라에서는 예전부터 행정 목적으로 주민등록부를 갖고 있다. 이런 등록부를 통계적인 목적으로도 쓸 수 있다. 때로는 사람, 사업체, 자동차, 토지, 금융거래 등과 관련된 등록부의 품질이 아주 좋아서 아주 적은 미포함 문제만 감수하면 표본 추출틀로 활용할 수 있다.

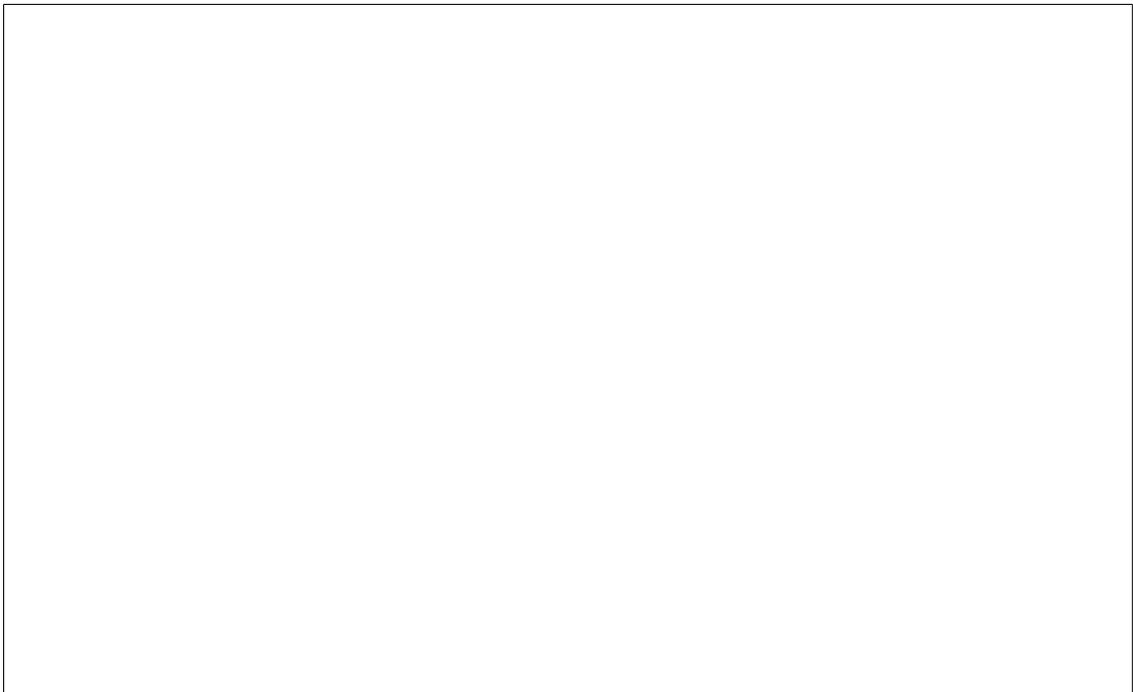
예를 들어 스웨덴 총인구 등록부는 한달에 두 번 갱신된다. 이 등록부의 과소포함률은 본질적으로 0이라고 볼 수 있다. 여기서 가장 중요한 포함 문제는 외국에서 태어나서 이곳에 거주하다 당국에 신고하지 않고 출신 나라로 돌아가는 사람들이다. 흔히 이런 포함 문제는 무응답 문제로 처리할 수 있다. 표본 추출된 사람을 연락할 수 없다면 이는 무응답을 나타내지만 이들의 경우 중 몇몇은 이미 외국으로 이주했으므로 더 이상 목표 모집단의 구성원이 아니다.

두 번째 예로 스웨덴 통계청은 스웨덴 국세청으로부터 스웨덴에 있는 사업체의 “개업”과 “폐업”에 대한 정보(즉, 모집단에 새로 들어오는 사업체와 모집단에서 빠져 나가는 사업체와 관련된 변화)를 한 달에 두 번 받는다. 모든 사업체들은 사업체통계를 위한 표본추출틀로 사용할 수 있는 사업체 등록부에 실려 있다. 일반적으로 사업체 추출틀이 총인구 추출틀보다 훨씬 더 복잡하다. 사업체 추출틀은 같은 등록부를 쓰는 조사들이 서로 연계되어 공동 작업을 필요로 하기 때문에 보통 계속적으로 갱신할 수 없다. 하지만 스웨덴 공식통계에 해당하는 사업체조사를 위한 추출틀은 일년에 네 번 갱신된다. 그렇지만 많은 나라들이 사업체 추출틀을 그렇게 정기적으로 정확하게 갱신할 수 있는 입장에 있는 것은 아니다.

흔히 추출틀은 다양한 출처를 통해 얻어진 목표 모집단 구성원에 대한 정보를 근거로 하여 개발되어야 한다. 완전한 추출틀은 얻기 위한 좋은 접근방법 중 하나는 몇 개의 불완전한 추출틀을 근거로 이들을 합치고 중복을 배제하여 새로운 추출틀을 작성하는 토대로 활용하는 것이다. 하지만 모든 모집단 구성원을 목록에 실는 것이 가능하지 않을 수도 있기 때문에 결과적으로 일정 부분의 모집단 미포함을 수용해야 하는 경우가 많다.

특히 농업과 토지 사용조사에서는 추출틀로 지도를 사용하는 경우가 많다. 조사 목적에 따라서 지도가 나라 전체 또는 나라의 일부 지역을 포함할 수 있다. 일차추출단위(primary sampling units)라고 부르는 분명하게 정해진 경계선이 있는 지역을 정의하여 나라 전체(또는 어떤 지리적 관심 대상)를 일차추출단위로 분할한다.

그 다음에는 지역 표본(일차추출단위)을 추출하고 표본 지역에 대해 이단계 추출 계획을 나타내는 새로운 지도를 작성한다(표본 추출에 대한 논의는 9장을 참조하라). 이런 과정은 최종 추출 단계에 이를 때 까지 계속된다. 이런 과정을 통해 최종 지역 표본을 추출하게 된다. 이 최종 지역 표본내에서 현지 직원들이 최종 조사단위들의 목록을 만든다. 이 목록을 만드는 과정에서 누락, 잘못된 포함, 중복과 같은 앞에서 논의된 오차들을 범할 수 있다. 수년에 걸쳐 조사단위 목록을 만드는 좋은 방법이 개발되었는데 이는 낱말이 세는 것을 토대로 하는 센서스에 사용되는 방법과 비슷하다. <그림 3.2>에 지역 추출틀 지도의 예가 있다.



<그림 3.2> 지역 추출틀 지도

포함 문제는 원래 정의한 목표 모집단에서 의도적으로 그리고 명백한 목적을 갖고 목표 모집단의 일부를 제외하는 경우와는 구별되어야 한다. 예를 들어 면접을 하는 데 드는 비용, 단위들에 접근하는 데 발생하는 어려움, 어떤 지역에 자격이 있는 적절한 면접원을 확보할 수 없는 문제 등에 기인하는 실제적인 사항 때문에 이런 의도적인 제외가 이루어진다. 어떤 경우에는 고의적이든 아니든 추출틀에서 몇몇 단위를 제외하도록 목표 모집단 자체를 재정의 하는 경우도 있다. 이렇게 하여 제외된 단위를 조사 부적격 대상으로 선언하여 추출틀 미포함 문제를 근본적으로

해결하는 것이다.

하지만 이런 방안은 추론을 위한 조사 결과의 적절성과 조사의 가장 중요한 목적을 달성한다는 측면에서 많은 문제들을 야기할 수도 있다. 만일 목표 모집단을 표본 추출틀과 더 잘 일치하게 하기 위해 재정의의 했다면 조사 결과를 사용하는 데 있어서 오해와 잘못된 추론을 막기 위해 새 목표 모집단을 명백히 그리고 분명하게 명시하는 것이 중요하다. 하지만 결국은 조사결과의 사용자만이 이런 절차가 그의 요구를 만족시킬 수 있는지 여부를 판단할 수 있다.

예를 들면 미국에서 실시되는 개인을 대상으로 한 조사에서는 병원, 감옥이나 면회가 제한된 다른 시설에 수용된 사람들을 모집단에서 제외하는 것은 일반적이다. 따라서 목표 모집단을 시설에 수용되지 않은 사람들의 집단으로 정의한다. 대면면접을 필요로 하는 개인을 대상으로 한 민간기관에서 실시하는 미국의 많은 조사에서는 추출틀에서 알래스카와 하와이를 제외하고 목표 모집단을 48개주에 사는 사람들로 정의한다. 이런 제한이 앞에서 언급한 이유로 정당화될지는 모르지만 적절성 같은 품질 요소에 영향을 줄 수 있다는 점에 유의해야 한다.

포함 개념과 추출틀 개발과 관련된 문제에 대해 Groves(1989)가 가구조사, Colledge(1995)가 사업체조사, 그리고 유엔의 식량농업기구(1996, 1998)가 농업조사에 관해서 유용한 정보를 제공하고 있다.

3.2 포함 편향의 측정

이 절에서는 미포함 오차가 조사 추정값에 미치는 잠재적인 영향을 상세하게 설명한다. <그림 3.1>에서 미포함 편향과 관련된 몇 가지 특성을 정의했다. 이 값들이 포함오차에서 생기는 편향과 상대 편향을 설명하기 위한 간단한 공식을 개발하는 데 쓰일 것이다. 그림과 관련된 기호는 다음과 같이 정의된다.

\bar{Y}_c = 추출틀에 포함된 사람들의 모집단 평균

\bar{Y}_{nc} = 추출틀에 미포함된 사람들의 모집단 평균

t_c = 추출틀에 포함된 목표 모집단의 비율 (포함률)

t_{nc} = 추출틀에 미포함된 목표 모집단의 비율 (미포함률)

\bar{Y}_p = 목표 모집단 평균

여기서 마지막 기호인 모집단 평균이 조사에서 추정하기를 원하는 모수이고, 미포함 오차에 의해 발생하는 이 모수의 조사 추정값의 편향을 나타내는 식을 고려해보자. 우선 모집단 평균을 \bar{Y}_p 이라고 표시하면, 모집단 평균은 다음과 같이 추출틀에 포함된 경우와 미포함된 경우의 모집단 평균의 가중 결합으로 표현될 수 있고

$$\bar{Y}_p = t_c \bar{Y}_c + t(1-t_{nc}) \bar{Y}_{nc} \quad (3.1)$$

이를 다시 말로 표현하면 다음과 같다.

$$\begin{aligned} \text{모집단 평균} &= (\text{포함률}) * (\text{포함된 모집단의 평균}) \\ &+ (1-\text{포함률}) * (\text{미포함된 모집단의 평균}) \end{aligned}$$

포함오차에 의해 모집단 평균 \bar{Y}_p 를 추정할 때 발생하는 편향을 나타내는 식을 이 관계에서 유도할 수 있다. 이 식은 이런 오차의 영향을 평가하려고 할 때 많은 도움이 된다. 이 식의 유도에 대한 수학적 과정은 Groves(1989), Lessler와 Kalsbeek(1992)에서 찾을 수 있다. 한편 미포함 오차에 의한 모집단 평균의 추정값의 편향은 다음과 같고

$$B_{nc} = (1-t_c)(\bar{Y}_c - \bar{Y}_{nc}) \quad (3.2)$$

이를 다시 말로 표현하면 다음과 같다.

$$\begin{aligned} &\text{미포함에 의한 편향} \\ &= (\text{미포함률}) * (\text{포함된 모집단의 평균과 미포함된 모집단의 평균의 차이}) \end{aligned}$$

이 공식이 의하면 미포함 편향이 평균제곱오차에 미치는 영향을 측정하려면 두가지 값을 알아야 한다. 그 두 가지 값은 추출틀에서 빠진 목표 모집단의 비율, $t_{nc} = 1 - t_c$ 과 추출틀에 있는 집단의 평균과 추출틀에 포함되지 않은 집단의 평균의 차이 $\bar{Y}_c - \bar{Y}_{nc}$ 이다.

예를 들면 모집단의 20%가 추출틀에서 누락되었다고 가정하고 소득과 같은 어떤 특성j에 있어서 추출틀에 있는 사람들과 누락된 사람들의 차이가 대략 매년 \$10,000인 것을 알고 있다고 하자. 그러면 연간 평균소득을 추정할 때 미포함 때문에 생긴 편향은 $0.20 \times \$10,000$ 또는 \$2000. 즉 이 추출틀에서 뽑은 표본은 목표 모집단의 연간소득을 \$2000 만큼 과대 추정하는 경향이 있을 것이다.

또 하나의 유용한 척도는 미포함 오차 때문에 생기는 상대 편향(*relative bias*)이다. 이 척도는 편향을 모집단 모수에 대한 상대적 비율 또는 백분율로 표현한다. 즉, 식(3.2)에 나오는 편향을 \bar{Y}_p 로 나누어 계산한다. 예를 들어 위에서 모집단의 연간 평균소득이 \$40,000인 것을 안다고(혹은 이전의 조사에서) 가정하자. 여기서 편향의 추정값이 \$2000이므로 미포함 때문에 생기는 상대 편향이 $\$2000/\$40,000 = 0.05$, 또는 5%이다. 바꿔 말하면 이 추출틀을 사용한 연간 소득의 추정값은 미포함 편향의 결과 약 5% 과대 추정한다는 것이다. 상대 편향의 계산공식은 다음과 같고

$$RV_{nc} = (1 - t_c) \frac{\bar{Y}_c - \bar{Y}_{nc}}{\bar{Y}_p} \quad (3.3)$$

이를 다시 말로 표현하면 다음과 같다.

미포함에 의한 편향 = (미포함률)

*(포함된 모집단 평균과 미포함된 모집단의 평균의 차이) / 모집단 평균

포함 편향과 상대편향에 대한 식에서 몇 가지 사항에 유의할 필요가 있다. 우선 만일 미포함률이 0에 가까우면(즉, 만일 t_c 가 1에 가까우면) 포함된 모집단과 미포함된 모집단의 차이가 아무리 클지라도 편향은 작을 것이라는 점에 주목해라. 미포함률이 늘어 날수록 그 결과로서 생기는 편향의 크기도 늘어나지만 그것이 얼마나

빨리 늘어나는 지는 포함된 모집단과 미포함된 모집단의 차이에 따라 결정된다. 만일 추출틀에 포함된 단위의 평균이 추출틀에 포함되지 않은 단위의 평균과 매우 가까우면 미포함 때문에 생기는 편향은 별로 없을 것이다.

예 3.2.1 랜덤번호전화결기(RDD) 가구조사(1장 참조)와 관련된 포함편향을 생각해 보자. RDD조사는 전화가 있는 가구만 포함한다. 이것은 현재 미국에서의 모든 가구의 약 94%를 의미한다. 그러면 $1 - t_c$ (즉 전화가 없는 가구의 비율)는 0.06이다. 모집단 가구의 평균 소득을 추정하는 것을 목표로 하는 조사를 생각해 보자. Massey와 Botman(1998)에서 나온 자료를 사용해서 전화가 있는 가구의 경우 \bar{Y}_c 가 \$18,700로 추정되고, 전화가 없는 가구는 \bar{Y}_{nc} 가 \$11,500로 추정되었다. 그러므로 (3.1)의 공식을 쓰면 모든 가구의 평균은 $0.94 \times \$18,700 + 0.06 \times \$11,500$ 또는 \$18,268이다. (3.3)의 미포함에 의한 상대 편향의 공식을 적용하면 다음 결과를 얻는다.

$$RB_{nc} = \frac{0.06(18,700 - 11,500)}{18,268} = 0.024 \text{ 또는 } 2.4\%$$

따라서 상대편향은 2.4%이다. 편향이 양수이기 때문에 추출틀 미포함에 의해 과대 추정이 되고 있다. 바꿔 말하면 그 추출틀에서 뽑은 랜덤포본으로부터 얻은 추정값은 실제 연간 평균소득을 2.4% 과대평가하는 경향이 있다. 음의 편향이나 음의 상대편향은 추정값이 모집단 모수를 과소평가하는 경향이 있다는 것을 의미한다. 간단하지만 이 포함 오차 모형이 다양한 종류의 모집단에 대해 포함편향이 언제 문제가 되는지에 대한 중요한 단서를 제공할 수 있다.

2.4%의 상대편향이 조사에 문제가 되는가라는 질문의 답은 조사의 목적과 자료가 어떻게 쓰일 것인지에 따라 달라진다. 만일 2.4%의 상대편향이 분석가들이 자료를 토대로 내리는 결정에 상당한 영향을 준다면 2.4%는 매우 중요한 편향이다. 그렇지만 만일 분석가들이 추정값에서 10%이상이 되는 차이에 관심이 있다면 2.4%의 편향은 중요하지 않게 여겨질 수도 있다. 그 경우 2.4% 편향은 수용이 가능하다. 하지만 이를 수용할 수 없다면 몇 가지 옵션을 생각할 수 있는데, 이에 대해서 생각해

보도록 하자.

만일 가능한 방법이 전화조사 밖에 없다면 분석가는 추출틀에서 전화가 없는 가구의 제외 때문에 생기는 포함편향을 감수하는 수밖에 없을 것이다. 만일 전화 없는 사람들에 대한 좋은 주소 목록이 존재하면 우편을 통해서 전화가 없는 가구로부터 자료를 얻는 것도 하나의 선택일 수 있다. 또 하나의 대안은 지역추출에 의한 대면면접조사를 병행해서 수행하는 것이다. 이런 방법을 이중추출틀조사(*dual frame survey*)라고 한다. 이 사례에서는 RDD 조사에 병행하여 조사단위를 대면면접하는 지역추출틀 조사를 함께 실시하는 것이다. 각 면접에서 응답자들에게 그들의 집에 유선 전화가 있는지 여부를 묻는다. 이렇게 함으로써 전화추출틀에 그 가구가 포함됐는지 여부를 확인할 수 있다. 하지만 대면면접이 보통 전화면접보다 비용이 훨씬 더 많이 들기 때문에 이에 다른 상당한 추가비용을 감수해야 한다.

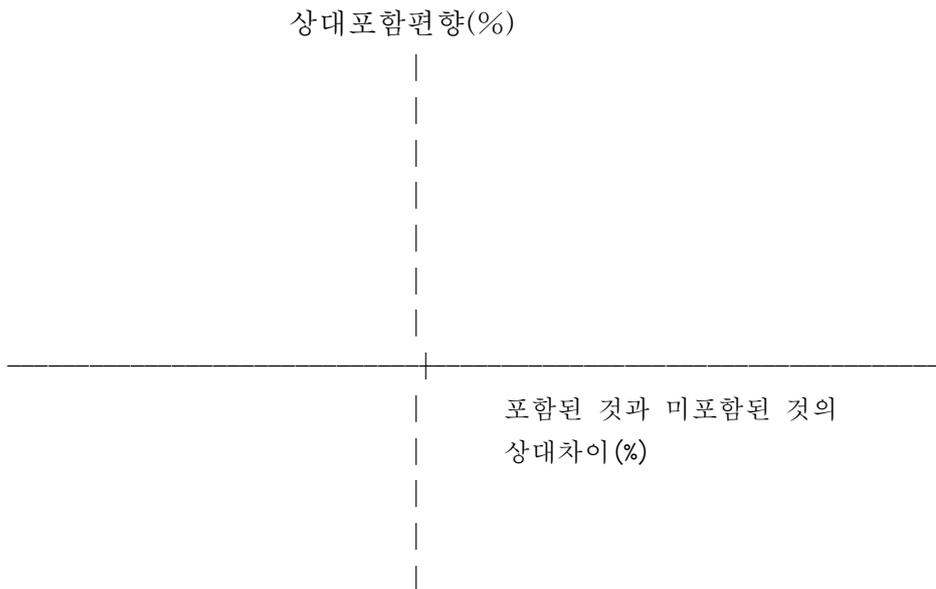
세 번째 옵션은 미포함편향의 추정량을 보정하는 것이다. 표본조사 관련 연구는 문들에는 이런 종류의 조정을 하는 많은 기술들을 설명하고 있다. 그렇지만 전화가 없는 포함편향의 보정에 대한 연구 논문들은 이런 포함 조정이 대부분의 경우에 그렇게 성공적이지 않다고 지적하고 있다. 사후 조사 보정이 약간의 포함편향은 제거할 수 있지만 상당 부분의 편향은 그대로 남고, 따라서 유일한 선택은 전화가 없는 가구에 대한 자료를 수집하는 것이다. 사후조사 보정은 최근 센서스나 다른 주요 조사에서와 같이 어떤 외부자료의 활용을 필요로 한다.

지역에 있는 모든 가구의 목록(예를 들어 우편 업무 관련 목록) 같은 다른 유형의 추출틀을 토대로 한 자료수집 방법에서는 목록에 없는 가구를 찾기 위해 추출틀 포함여부를 확인하는 특별한 방법을 수행할 수 있다. 포함오차 개선을 위해서는 더 경험이 있는 직원을 활용하거나 추출틀에 누락된 단위를 채우기 위한 더 세밀한 방법을 사용해서 목록의 포함상태를 개선하려는 노력이 요구된다. 포함오차 개선방법의 적용에 따른 비용 때문에, 관련 연구에서는 추출틀 포함오차를 추정하고, 가능하면 이를 보정하기 위해 표본을 대상으로 한 포함여부에 대한 추가적인 정보를 수집한다. 결과적으로 포함 오차는 일반적인 방법을 적용하는 경우와 포함 오차를 줄이기 위해 좀더 개선된 방법을 적용하는 경우의 차이라고 말할 수 있다.

예 3.2.2 예 3.2.1에서는 전화 보유여부에 따른 가구의 특성과 전화추출틀의 미포함률에 대한 정보가 주어졌기 때문에 전화추출틀 사용에 따른 포함 편향에 대한 상

당히 정확한 점 추정값을 구했다. 하지만 어떤 경우에는 이런 수량적인 정보가 부족하다. 그렇다면 앞에 말한 공식들을 사용해서 “가정” 분석을 하는 것이 필요하다.

어떤 지역 또는 관심의 대상인 부모집단 추출들의 미포함률이 12%보다 크지 않다는 것 외에는 정확히 아는 것이 없다고 가정하자. 그리고 포함된 모집단과 미포함된 모집단의 평균 소득의 차이가 \$10,000와 \$12,000사이인 것으로 알고 있다. 이런 시나리오에서 가능한 최대 편향을 계산해서 이 상황에서 포함오차가 문제가 되는지를 판단하기 위해 상대 편향공식을 이용할 수 있다. 예를 들면 미포함률이 12%이고 공식에 있는 두 평균의 차이가 \$12,000라고 가정하자. 식 (3.2)를 사용하면 미포함편향의 추정값이 \$1440이 된다. 만약 모집단 평균이 \$18,268이면 상대편향은 $\$1440/\$18,268 = 0.079$ 또는 7.9%이다.



<그림 3.3> \bar{Y}_c 와 \bar{Y}_{nc} 의 상대차이와 t_c 의 함수로 나타낸 포함 편향

우리는 최악의 경우를 가정한 시나리오를 고려했기 때문에 만약 이런 편향을 받아들일 수 있으면 이 추출틀을 받아들일 수 있다. 만약 이 편향을 받아들일 수 없다면 미포함편향의 영향을 가능한 더 정확히 평가하기 위해 더 많은 정보가 필요하다. 이러한“가정”분석은 조사 설계자가 포함 오차가 특정 모집단 특성과 조사 자료의 활용에 문제가 되는지를 결정하게 해준다. <그림 3.3>은 이러한 목적에서 그려졌다. 그림에서 상대 포함편향은 모집단 중 추출틀에 포함된 부분과 포함되지 않은

부분의 차이의 함수로 설명된다. 즉, 그래프의 X축은 $(\bar{Y}_c - \bar{Y}_{nc})/\bar{Y}_p$ 이고 Y축은 RB_{nc} 이다. 그래프의 선들은 포함률 t_c 의 여러 가지 수준을 의미한다. 첫 번째 선은 t_c 가 0.9 또는 90% 포함을 나타내고 두 번째는 70% 포함, 세 번째는 50% 포함을 나타낸다. X축은 모집단의 포함된 부분과 포함되지 않은 부분의 평균의 차이를 총 평균의 백분율로 표시한 것이다. 다음 예에서는 추출틀에 대한 정확한 정보가 알려지지 않은 상황에서 “가정”분석을 실시하기 위한 이 도표의 사용방법을 설명한다.

예 3.2.3 \bar{Y}_c 와 \bar{Y}_{nc} 의 차이가 \bar{Y}_p 보다 30% 높다고 하자. 그러면 X축에서 30을 찾아서 이 점에서 포함률과 가장 가까운 선까지 올라간다. 예를 들면 만일 추출틀이 모집단의 90%를 포함하면 Y축에서 가리키는 상대 편향은 대략 2.5%이다(0%와 5%사이에서 보간법 적용) 하지만 만일 포함이 70%에 가까우면 상대 편향은 약 9%로 훨씬 높아진다. 50% 포함에서는 상대편향이 대략 17% 정도까지 올라간다. 이 도표를 이용해서 추출틀에 포함된 집단과 미포함된 집단의 특성의 다양한 상대적인 차이와 다양한 포함률의 변화에 따른 몇 개의 시나리오를 검토하는 것이 가능하다. 물론 포함편향의 요소들에 대한 정보가 적을수록 포함편향의 가능한 범위가 넓어진다. 추출틀 포함이 70%와 90% 사이 어딘가라는 것만 알면 앞의 경우와 동일한 상대적인 차이일 때, 상대 포함편향은 2.5%에서 9%사이에 있다. 이런 이유로 이용을 계획하고 있는 어떤 추출틀의 수용여부를 결정하기 위해서 더 많은 연구를 해야 될 경우도 있다.

예 3.2.4 미국 국립농업통계국(*U.S. National Agricultural Statistics Service; NASS*)의 농장 조사에서 농장을 선택하는 데 사용한 NASS의 목록 추출틀을 생각해 보자. <표 3.1>은 NASS 목록 추출틀과 1977년 미국 농업 센서스를 위해 개발한 추출틀을 비교한 것이다. NASS 목록 추출틀에서 농장의 경우 포함률이 56.3%이지만 농장 토지 면적의 경우 77.6%로 훨씬 높다. 농업 센서스는 훨씬 더 많은 농장인 89.2% 그리고 대부분의 토지인 98.6%를 포함한다. 낮은 농장 포함률(56%)이 주어졌을 때 농장 조사에서 이 추출틀만 사용하는 경우에는 포함편향에 따른 문제가 있다. 여기서 NASS 추출틀에서 전체 농장의 44%가 빠졌는데도 불구하고 농장

토지의 약 78%를 포함하고 있다는 것에 주목하기 바란다. 이는 추출틀에서 빠진 것은 대부분 토지 면적이 작은 농장인 것으로 추측되고, 대부분의 주요 농업 경영체는 NASS 추출틀에 포함되어 있다고 볼 수 있다.

<표 3.1> NASS 목록 추출틀 포함 현황

	NASS 목록 추출틀 (%)	1977 농업센서스 (%)
농장	56.3	89.2
농장 보유 토지	77.6	98.6

출처: National Agricultural Statistics Service 자료

이제 포함 편향 공식들을 NASS 표본추출틀의 편향을 결정하는데 적용해 보자. 포함 편향은 포함률과 추출틀에 포함된 부분과 포함되지 않은 부분의 특성의 차이의 함수이므로, 우리는 가장 중요한 조사 특성들에 대해서 소규모 농장과 대규모 농장에 어떤 차이가 있는지 판단하기를 원한다. 소득, 판매량, 재고량, 생산량과 같은 특성에 있어서는 뚜렷한 차이가 있을 것이다. 하지만 어느 특정한 계절의 밀, 옥수수 등의 총 생산량의 추정에 관심이 있다면 만일 소규모 농장들이 총 생산량에 미치는 영향이 총체적 아주 작다면 소규모 농장들이 추출틀에서 빠진 것이 별로 중요한 사안이 아닐 수 있다. 하지만 다른 항목에 대해서는 추출틀 미포함이 농업 추정값에 중요한 영향을 미칠 수도 있기 때문에 NASS가 이 영향을 평가하기 위한 연구를 수행하였다. 농장 대신 농장 보유 토지를 대상으로 하는 경우 포함 편향에 따른 위험은 상당히 완화될 것이다. 한편 농업 센서스에서는 농장 토지의 약 1%밖에 누락되지 않았다. 그러므로 t_c 가 매우 작기 때문에 \bar{Y}_c 와 \bar{Y}_{nc} 의 차이가 매우 클지라도 포함편향이 매우 작을 것이다.

예 3.2.5 이제 미국의 보건면접조사에서의 한 예를 생각해 보자. 1980년대 중반에 미국 국립보건통계센터(National Center for Health Statistics; NCHS)가 보건면접 조사를 실시하는 데 RDD를 사용하였다. 그 당시 RDD 추출틀은 미국에 있는 가구들의 약 93%만 포함했다. 보건면접조사를 RDD로 실시하는 것에 대한 평가에서 NCHS가 알고자 했던 것은 조사에서 나오는 주요 추정값에 포함편향이 미치는 영

항이었다. 조사에 있는 항목 중 하나는 응답자들의 흡연여부였다. 이 문항에 대한 포함률 t_c 는 92.8%, \bar{Y}_c (전화가 있는 가구중에 흡연자의 비율)는 28.8%이었고, \bar{Y}_{nc} (전화가 없는 가구 중에 흡연자의 비율)는 49.6%이었다. 그러면 (3.1)의 \bar{Y}_p 공식을 적용하면 모집단 평균을 다음과 같이 얻을 수 있다.

$$\bar{Y}_p = 0.928 \times 0.288 + 0.072 \times 0.496 = 0.30$$

식 (3.2)에서 편향을 계산하면 다음과 같고,

$$B_{nc} = 0.072 (0.288 - 0.496) = -0.015$$

따라서 상대편향 RB_{nc} 는 $-0.015/0.30 = -0.05$ 또는 -5.0% 이다. 이것은 현재 담배를 피우는 사람의 비율이 RDD를 사용하면 5% 과소평가 된다는 것인데, 이것은 자료 수집 방법에 대한 검토에서 무시할 수 없는 차이이다.

예 3.2.6 또 다른 예는 직업에서 받는 정신적 스트레스와 관련된 전화 조사에 대한 것이다. t_c , \bar{Y}_c 와 \bar{Y}_{nc} 의 백분율은 각각 92.8%, 17.6%과 17.0%이다. 여기서는 직업이 정신적 스트레스의 원인이 되는지 여부에 있어서 추출틀에 포함된 집단과 포함되지 않은 집단 사이에 차이가 별로 없다. 다시 말해 전화가 있는 가구의 사람들과 전화가 없는 가구의 사람들이 모두 비슷하게 스트레스를 받는다는 것이고, 따라서 미포함 때문에 발생하는 상대 편향은 매우 작다. 이 경우에는 단지 0.0025이었다.

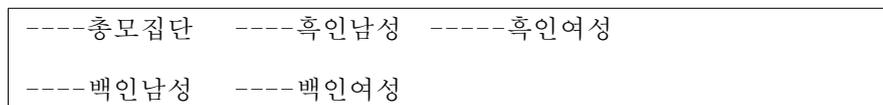
예 3.2.7 마지막으로 개인 건강보험에 관련된 예를 생각해 보자. NCHS는 개인 건강보험을 가지고 있는 사람의 비율을 추정하기를 원한다. 이 기관은 이것을 빈곤층과 부유층 두 그룹 모두에 대하여 알고 싶어 한다(<표 3.2> 참조). 빈곤층에서는 전화 포함율이 단지 72.6%이고, 부유층에서는 전화 포함율이 96.3%로 증가한다. 따라서 빈곤층 부모집단의 상대 편향은 16.3% 이고, 반면에 부유층 부모집단의 경우

1.2% 이다. 이를 통해 포함 문제가 어디에서 큰 편향을 발생시키는지 파악할 수 있다.

<표 3.2> 건강보험조사에서의 상대 포함편향(%)

	t_c	\bar{Y}_c	\bar{Y}_{nc}	상대편향
빈곤층	72.6	34.5	16.8	16.3
부유층	96.3	87.1	59.3	1.2

미국에서 10년마다 하는 센서스 추출틀은 목표 표본을 설정하고 표본을 다양한 층으로 배분하는 목적으로 미국 조사 연구에 흔히 사용된다. 많은 미국의 조사기관들은 가구조사를 실시하므로 가구조사를 실시하는 사람들에게는 센서스국에서 10년마다 하는 센서스에서의 모집단 포함 현황에 대해 매우 관심이 높다. <그림 3.4>에 있는 자료는 1940년에서 1990년까지의 센서스에서 과소계상(net undercount; 실제보다 적게 셈) 현황을 나타낸다. 이 그래프의 아래에서 두 번째 굵은 선은 전체 모집단을 나타낸다. 여러 해에 걸친 미국센서스국의 집중적인 노력으로 과소계상 또는 센서스에서 빠진 사람들의 수는 감소했다. 하지만 1990년 센서스에서는 과소계상이 약간 늘어났다. 1980년에 백인여성의 과소계상 비율이 음수가 되었는데 이것은 사실상 과대계상이 되었다는 것을 의미한다. 이 그림에서 흑인 모집단이 백인 모집단보다 더 많이 누락되고 흑인남성이 흑인여성보다 더 많이 누락되며 일반적으로 남성이 여성보다 더 많이 누락된다는 것을 볼 수 있다.



<그림 3.4> 미국 인구 및 가구 센서스에서 추정된 과소계상 비율

10년마다 하는 센서스를 통해 얻은 미국 센서스국의 이러한 경험은 다른 미국 조사기관이 가구 모집단 조사를 실시하면서 얻은 경험과 다르지 않다. 조사기관들은 거의 같은 사람들을 누락 시킬 것이고 어찌면 미국 센서스국이 센서스에서 이 사람들을 누락하는 것 보다 더 높은 비율로 빠뜨릴 것이다. 이런 조사들에서도 흑인이 백인보다 더 많이 누락될 것이고 남성이 여성보다 더 많이 누락될 것이다. 조사 기관들이 사용하는 조사과정이 미국 센서스국이 10년마다 하는 센서스에서 사용하는 과정과 크게 다르지 않으므로, 모든 인구통계학 관련 조사에는 포함오차가 발생하게 된다. 물론 다른 나라의 경우 각 나라마다 누락되기 쉬운 경향이 있는 특유의 모집단 그룹 패턴이 존재할 수 있다.

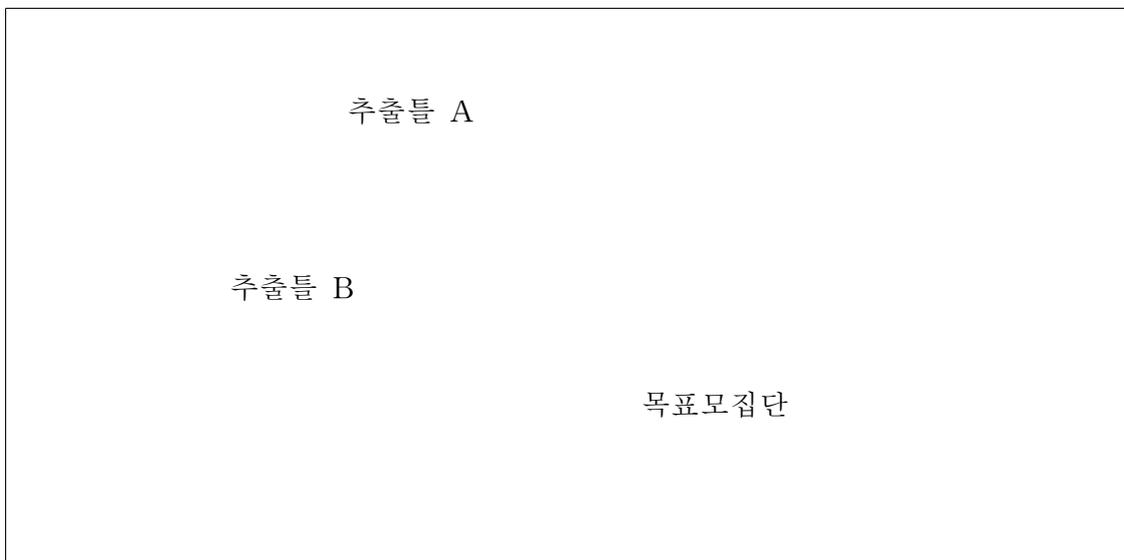
포함편향은 미포함율과 포함된 모집단과 포함되지 않은 모집단의 관심대상인 특성의 차이의 함수이다.

3.3 포함 편향의 축소

이 절에서는 다양한 표본추출틀과 관련된 포함오차를 줄이거나 제거하는 몇 가지 방법을 생각해본다. 이미 언급했듯이 가장 효과적인 편향 축소 방법은 아마도 중복된 것과 잘못 포함된 것을 제거해서 추출틀을 보완 하는 것과 추출틀에서 누락된 단위를 추가하기 위해 현장조사를 통해 누락된 것을 확인하는 것이다. 이것은 비용이 많이 드는 과정이지만 많은 조사기관들은 자신들의 인구통계학적 조사와 사업체 조사를 위해 이런 체계를 세우려고 한다. 이들은 오랜 기간 동안에 걸친 많은 조사에 이를 활용함으로써 결과적으로 이런 과정에 투여된 많은 비용을 충분히 보상 받을 수 있다는 입장을 취하고 있다. 물론 작은 조사기관의 경우 이런 접근법은 평균 제곱오차를 줄이는 효율적인 방법이 아닐뿐더러 가능하지 않은 방법일 수도 있다. 한편 다중추출틀(*multiple frames*; <그림 3.5> 참조)의 활용은 작거나 큰 기관 모두

에게 비용 효율이 더 높을 수 있다.

<그림 3.5>에서 목표 모집단은 원으로 나타나 있고, 이 원이 덮고 있는 어둡게 표시된 두 직사각형은 각 추출틀의 포함상황을 나타낸다. 이 예에서는 추출틀 A와 추출틀 B가 있다. 이 추출틀 중 어느 하나도 단독으로는 목표 모집단을 포함하지 못한다. 하지만 그림에서 보여주듯이 두 추출틀이 결합한 것이 어느 한 추출틀 자체보다 훨씬 더 바람직하다. 중간의 작은 직사각형은 두 추출틀의 중복 부분을 의미한다. 만약 이 중복 지역에 있는 단위들이 중복되는 것을 제거하지 않고 추출틀을 결합시키면 중복된 단위들은 이 지역 밖에 있는 단위들 보다 선택의 확률이 높을 것이다. 중복된 부분에 있는 단위들은 양쪽 추출틀에 나타났기 때문에 그것들은 기본적으로 선택될 가능성이 두 배인데 반해서 둘 중 한 개의 추출틀에만 있는 단위들은 선택될 가능성은 이에 비해 1/2이 된다. 만일 조사과정에서 우리가 조사하고 있는 단위들이 어떤 추출틀(A 또는 B) 또는 둘 다에 포함되는지를 알 수 있다면 이것은 문제가 되지 않는다. 왜냐하면 이런 중복여부가 파악되면 표본 추출단위에 대한 정확한 포함 확률을 산출하여 추정과정에서 이것의 역수를 가중값으로 적절히 반영할 수 있기 때문이다.



<그림 3.5> 두 개의 추출틀을 사용한 모집단 포함 상황: 추출틀 A도 추출틀 B도 모집단 전체를 완전히 포함하지 않는다. 하지만 두 추출틀의 결합은 모집단 전체를 거의 포함한다. 두 추출틀의 중복부분을 추출틀의 결합이나 추정과정에서 검토하는 것이 필수적이다. 그렇지 않으면 추정값이 불균등한 표본 포함 확률에 의해 편향 될 것이다.

예를 들어 추출틀 A가 RDD 전화번호 추출틀이라고 가정하자. 이 추출틀에서 고른 모든 가구는 전화가 있는 가구이다. 추출틀 B는 운송용 고객 명부같은 전화소유를 토대로 하지 않은 어떤 다른 종류의 추출틀이라고 가정하자. 이 추출틀에서는 전화소유에 대한 정보는 알 수 없다고 가정하자. 만일 우리가 표본을 추출하는 데 이 추출틀들을 결합해서 사용하면 표본에 있는 모든 단위의 선택 확률을 알기 위해서는 추출틀 B에 있는 모든 단위의 전화소유 여부를 알아야 할 것이다. 그렇지 않으면 표본 자료에 가중값을 올바르게 줄 수 없다(9장 참조).

표본으로 선택된 특정한 단위가 몇 개의 근원 추출틀에 속하는지를 알기 어렵게 결합 추출틀이 만들어졌을 때에는 문제가 발생한다. 일반적으로 몇 개의 추출틀이 결합되는 경우 다양한 추출틀 사이에서 단위들이 중복되지 않도록 조정하는 것이 최선이다. 이런 작업은 추출틀들이 이름, 주소, 또는 서로 확인 가능한 정보를 공유하면 이를 비교하여 중복된 것을 확인할 수 있는 컴퓨터 프로그램으로 상당히 쉽게 해결할 수 있다. 하지만 이 과정에서 약간의 오차가 생길 것이고 다소의 중복이 발생할 것이다. 또한 같은 추출틀에 있는 다른 단위들이 잘 못 대응될 가능성이 있고 어떤 단위들은 추출틀에서 잘못 제외될 것이다. 그럼에도 불구하고 하나의 추출틀을 완성하기 위해 소모적인 노력을 하기 보다는 이런 접근방식이 조사의 포함 오차를 줄이는 훨씬 더 비용 효율이 높은 오차를 줄이는 방법일 것이다.

추출틀의 미포함율을 추정하고 미포함 편향을 보정하는 다양한 기술들이 있다. 이 방법들은 주로 센서스 포함 오차를 검사하기 위해 개발되었다. 기록 검사(record check)에서는 외부 기록에 있는 목표 모집단의 일원들이 센서스에서 빠졌나를 보기 위해 외부 기록을 검토한다. 한 살 미만의 어린이 같은 모집단의 일부 집단에 대해서는 이런 기록들이 존재할 수도 있다. 현행 출생기록이 있다고 가정하고 한 살 이하인 N명의 유아 가운데 n명의 유아를 표본추출하여, 그들이 센서스에 계상되었는지 여부를 확인해 보았다고 하자. 그러면 n명 중 센서스에 계상되지 않은 유아들의 비율이 센서스에서 누락된 비율에 대한 추정값이다.

관심 모집단에 대해 대조 기준이 되는 신뢰할 수 있는 총계 추정값이 있다면, 그 두 추정값을 비교하는 일반적인 방법에 해당하는 포함비율(coverage ratio)를 활용하는 것인데, 이는 조사에서 나온 추정값을 “선호되는” 추정된 총계(즉, 별개의 대조 기준 모집단 총계)로 나누어 계산한다. 포함비율에 대한 예는 <표 3.3>에 제시

되어 있다. 추정값과 대조 기준 통계사이의 편차는 포함 문제가 아닌 다른 요소들 때문에도 얼마든지 발생할 수 있기 때문에 조사 추정값을 모집단 대조 통계에 비교하는 것은 포함 오차를 확인하는 상당히 불완전한 방법이다.

<표 3.3> 미국 CPS(Current Population Survey) 포함비율

나이	비 흑인		흑인		모든 사람		
	남	여	남	여	남	여	합계
0-14							
15							
16-19							
20-29							
30-39							
40-49							
50-59							
60-64							
65-69							
70 이상							
15 이상							
0 이상							

출처: 미국 센서스국(1992)

추출틀은 조사 과정에서 매우 중요한 역할을 한다. 만약 추출틀이 적당하지 않으면 큰 포함 오차가 생길수도 있고, 그것을 줄이거나 보완하는 데 비용이 많이 들거나 불가능할 수도 있다. 많은 정부 기관이나 대규모 조사 기관들은 그들이 수행하는 모든 조사에 사용될 수 있는 통합된 추출틀 개발을 위해 노력한다. 이런 경우에 출발점은 사업체등록부, 농장등록부나 개인등록부이다. 이 등록부들을 이용해 농업 조사와 같은 개별 조사를 위한 추출틀 데이터베이스를 구축하고, 이를 통해 추출틀 정비 및 등록된 단위와 관련된 보조 정보의 보완작업 등을 단순화할 수 있다.

표본 추출틀에서 흔히 얻을 수 있는 보조정보(*auxiliary information*)는 추정뿐만 아니라 다른 설계특성을 결정할 때 고려할 매우 중요한 요소다. 보조정보는 비추정(9장 참조)과 같은 모집단 모수의 추정방법을 개선하는데 사용될 수 있는 모집단 단위들에 대한 추가적인 정보를 의미한다. 사실상 실용적인 가치가 있으려면 추출

틀은 단지 모집단 단위들의 목록에 그치는 것이 아니라 보다 더 많은 정보를 포함해야 한다. 9장에서 설명되겠지만 효율적인 층화와 추정을 위해서 사용될 수 있는 정보가 필요하며, 특히 사업체조사나 우편 가구조사에서는 단위 내에서의 접촉 대상자에 대한 사항이 추출틀에 포함되어야 할 중요한 정보이다. 사업체조사 추출틀에는 사업체의 사업규모, 주요산업, 생산 제품 등에 대한 정보가 포함될 수 있다. 이런 정보는 표본을 추출하기 위해서 뿐 아니라 면접을 위한 첫 접촉 대상자를 확인하는데 중요할 수 있다.

포함오차의 축소와 추정을 위한 몇 개의 더 심층적인 방법들이 있다. 중복표본추출법(Sirken, 1970; Johnson, 1995), 다중추출틀방법(Hartley, 1974; Kott and Vogel, 1995), 추출틀 구성원을 추출틀에 포함되지 않은 목표 모집단의 구성원에 연결하는 방법(Groves, 1989)과 추출틀과 등록부(Colledge, 1995) 등이 이에 해당한다. 미국 센서스국은 10년 마다 실시하는 인구 및 주택 센서스와 관련된 포함오차에 대한 매우 광범위한 연구를 여러 해에 걸쳐 수행하였다. 센서스 자료를 위한 포함오차 모형은 Wolter(1986)가 논의하였고, 발생 가능한 포함오차를 평가하고 조정하는 것을 목적으로 하는 사후확인조사에 대한 내용은 Hogan(1993)이 소개하고 있다.

포함오차를 방지 또는 줄이기 위한 방법

- 추출틀에서 중복된 것과 잘못 포함된 것의 제거
- 두 개 이상의 추출틀의 사용
- 목표 모집단 구성원의 추출틀 누락을 확인하기 위한 외부 기록 확인

3.4 단위 무응답 오차

이 절에서 우리는 두 번째 종류의 미관측 오차인 무응답을 고려한다. 무응답은 표본으로 선택된 단위에 대한 자료 수집을 실시할 수 없을 때 일어난다. 표본 단위에 대한 면접이나 설문지 전체가 없어졌거나 어떤 응답도 얻지 못한 경우를 단위 무응답(*unit nonresponse*)이라고 부른다. 어떤 경우 완성된 설문지를 얻었더라도 설문지에 있는 모든 항목이 완성되지 않았을 수 있는 데, 이와 같이 일부 항목에 대

한 응답이 누락된 경우를 항목무응답(*item nonresponse*)이라고 부른다.

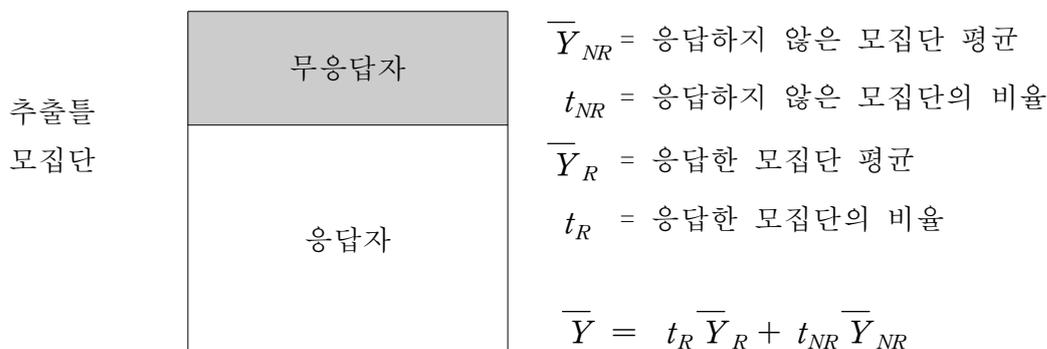
어떤 조사단위의 경우 조사항목들 중 상당 부분에 대해 응답을 했을 때, 단위 무응답과는 대조가 되는 응답자로 분류하기 위한 최소 기준을 설정하기 위한 뚜렷한 지침은 없다. 기술적으로 말하면 만일 단위가 설문지에 있는 한 개의 질문에 답을 하면 응답자로 부를 수 있다. 하지만 실제로 응답 단위가 되기 위한 필요조건은 보통 그것 보다는 훨씬 더 엄격하다. 설문지가 완성된 것으로 볼 수 있는지 여부를 판정하기 위해서는 사전에 설문지에 있는 몇 개의 또는 어느 중요 항목들이 완성될 필요가 있는지 등에 대한 응답자 분류 기준이 필요하다. 그러면 이 기준을 만족시키지 못하는 설문지는 불완전하다고 간주하고 그 조사단위는 무응답으로 분류된다. 상당한 항목무응답이 있는 경우, 이러 조사단위를 응답한 단위로 분류하기에 충분한 것으로 볼 수 있는지 판정하는 것은 전적으로 연구자에 달려있다. 항목무응답은 3.6.10절에서 더 논의한다.

단위무응답은 표본단위와 접촉할 수 없기 때문에 또는 접촉에 성공했지만 그들이 조사에 참여하기를 거부하기 때문에 일어난다. 이러한 경우들은 각각 비접촉(*noncontact*)과 거부(*refusal*)의 경우로 부르고, 두 그룹은 무응답 추적과정과 추정단계에서 모두 별도로 구분하여 처리될 수 있기 때문에 이들을 구별하는 것이 필요하다. 하지만 비접촉과 거부의 차이는 항상 뚜렷한 것은 아니다. 예를 들어 우편조사에서 응답하지 않는 사람들과 접촉되지 않은 사람들은 실제로는 모두 응답거부에 해당한다. 조사기관이 설문지 사본을 여러 번 보내거나 또는 표본 단위를 전화로 접촉하려고 노력했지만 성공하지 못했을 수 있다. 많은 기관들은 이런 경우를 비접촉으로 분류하지만 어떤 비접촉 단위들은 분명히 조사시도를 알고 있지만 반응하지 않기로 결정한 경우로 이것은 사실상 참여하기를 거부한 것이다.

가구 전화조사에서 전화를 받는 가구 구성원이 면접원에게 거짓으로 표본으로 선정된 사람이 집에 없다고 한다면 이것도 숨겨진 거부의 한 경우이다. 이런 경우 진짜 비접촉인지 또는 숨겨진 거부인지를 알 수 없다는 것이다. 항목무응답은 질문들을 고의로 또는 무심코 건너뛰기 때문에 일어난다. 우편조사에서는 좋지 않은 질문 배치 때문에, 응답자가 질문의 의미에 대해 확실치 않게 때문에 또는 그가 단순히 질문에 답을 하고 싶지 않기 때문에 일부 질문들을 그냥 넘어갈 수 있다. 면접에서는 면접원이 지시를 놓치거나 건너뛰거나 응답자가 몇몇 질문들에 답하기를 거부하여 항목무응답이 생긴다.

포함오차에서처럼 단위무응답에 대처하기 위해서도 두 가지 주된 전략이 있다. 하나는 무응답률을 줄이는 것을 목표로 하는 것이고, 다른 전략은 무응답을 조정하는 것이다. 항목무응답은 설문지의 표현방법과 질문의 배치 등을 보완함으로써 일부 예방할 수 있다. 이런 노력에도 불구하고 생기는 항목무응답을 보정하기 위해 대체(imputation)를 위한 다양한 기법들이 일반적으로 사용된다(3.6.11절과 7장 참조). 여기서는 단위무응답의 요소, 무응답편향의 의미 그리고 무응답을 줄이는 몇 가지 방법들에 대해 논의한다. 무응답 보정방안들은 기술적으로 어려워서 여기서는 이런 기법들에 대해 상세하게 논의 하지 않는다. 이런 기법을 적용하기 위해서는 충분한 기술적인 능력을 갖춘 통계학자의 전문적인 지식을 필요로 한다.

무응답 때문에 발생하는 편향에 대한 이해를 돕기 위해 <그림 3.6>을 참고하자. 이 그림은 근본적으로 미포함오차(<그림 3.1> 참조)를 논의 했을 때 고려했던 것이다. 여기서는 무응답문제에 적용하기 위해 일부 라벨을 다시 붙였다. 나중에 보겠지만 우리가 미포함 오차를 검토했을 때 사용했던 것과 같은 종류의 추론이 무응답오차에도 마찬가지로 적용된다. 앞에서와 같이 큰 직사각형이 표본 추출틀에 나타난 모집단을 나타낸다. 이 모집단에서는 조사에 응답하는 단위(응답자들)와 그렇지 않은 단위(무응답자들), 두 종류의 표본단위가 존재한다. 처음에는 응답자와 무응답자를 표본 차원이 아닌 모집단 차원에서 이야기 한다는 것이 혼란스러울지 모른다. 그 이유는 표본에 추출된 단위에 대해서만 우리가 조사에 응답하라고 요청하기 때문이다. 하지만 이런 접근방식은 새로운 기호나 기술적인 개발 없이 포함편향을 위해 사용된 일반적인 공식의 사용을 가능하게 하기 위한 것이다. 이렇게 미포함 오차를 논의 하는데 사용했던 같은 원리를 사용해서 무응답오차가 어떻게 조사 오차에 영향을 미치는지를 설명하는 것이 쉽게 이해하는 데도 도움이 될 것이다.



<그림 3.6> 무응답 문제: 상자의 어둡게 표시된 부분은 현재의 방법으로 조사에 응답하지 않을 모집단의 비율을 나타낸다. 조사 무응답률은 추출틀 모집단에서 이 영역의 비율에 대한 추정값으로 생각할 수 있다. 어둡게 표시되지 않은 부분은 응답을 할 모집단의 비율을 나타낸다. 표본에 있는 응답자들은 이 영역에서 선택된 것이다. 이 그림과 <그림 3.1>의 유사함에 주목하라. 무응답편향과 포함편향의 개념은 상당히 밀접한 관련이 있다.

모집단에서 표본을 추출할 때, 표본 단위와 접촉을 시도하고 그들에게 참여해 달라고 요청할 때까지는 응답자와 무응답자를 구별할 수 없기 때문에 표본 추출과정에서는 응답자와 무응답자를 구분할 수 없다. 모집단에서 무응답에 해당하는 단위를 표본으로 선택할 때 마다 그 표본은 응답을 하지 않을 것이므로 한 개의 무응답이 발생하게 된다. 결국 응답 자료가 있는 표본들은 응답자 모집단에서 뽑힌 것들이다. 따라서 어둡게 표시된 부분에 있는 직사각형의 비율이 무응답률에 해당한다.

이 두 그룹의 개념적인 모집단 구분은 매우 복잡한 현실을 매우 간단한 결정적 모형으로 설명한 것이다. 이 모형은 다른 것처럼 현실을 정확하게 표현하지 못한다. 실제로는 응답자와 무응답자 사이의 구분은 그렇게 간단하지 않다. 더 정확한 설명을 하기 위해서는 서로 다른 조사방법은 모집단을 응답자와 무응답자로 다르게 구분할 것이라는 것을 참작해야 한다. 무응답률을 최소화하기 위해 다른 조사방법을 고려한다는 것이 이런 현상을 염두에 둔 것이다. 더 복잡한 모형은 표본 단위에 대한 응답 확률의 변동 가능성(즉 상황에 따라 추출된 단위가 때로는 응답자가 때로는 무응답자가 될 수 있음)을 가정에 포함하는 것이다. 간단한 결정적 모형에서는 응답 확률이 1 또는 0 중에 하나라고 가정한다. 하지만 간단한 모형이 그림에도 불구하고 무응답이 조사 추정값에 미치는 영향을 이해하는데 매우 유용하다. 포함오차를 설명하기 위해서 유사한 기호를 사용했기 때문에 여기서 사용된 기호는 익숙할 것이다.

<그림 3.6>에서 t_R 이 조사에 응답할 모집단의 비율을 표시하고(어둡게 표시되지 않은 부분), \bar{Y}_R 은 응답자 모집단 평균이라고 하자. t_R 은 우리가 조사에서 기대하는 응답률로 해석할 수 있다는 것에 유의하라. t_{NR} 은 조사에 응답하지 않을 모집단의 비율을 표시하고 (어둡게 표시된 부분), \bar{Y}_{NR} 은 무응답자 모집단 평균이라고 하

자. 마지막으로 앞서처럼 \bar{Y} 는 모집단 전체 평균을 표시한다.

3.2절에서의 접근 방법과 유사하게 모집단 전체 평균은 응답자와 무응답자 각각의 평균을 가중 평균하여 다음과 같이 구할 수 있다. 여기서 가중값은 응답자 모집단의 크기(즉, 조사에서의 “기대” 응답률)의 함수이다.

$$\bar{Y} = t_R \bar{Y}_R + (1 - t_R) \bar{Y}_{NR} \quad (3.4)$$

이것을 다시 말로 표현하면 다음과 같다.

$$\begin{aligned} \text{추출틀 모집단의 평균} &= (\text{기대 응답률}) \times (\text{응답자의 평균}) \\ &+ (\text{기대 무응답률}) \times (\text{무응답자의 평균}) \end{aligned}$$

또한 무응답 때문에 발생하는 편향과 상대편향을 나타내는 식은 다음과 같다.

$$B_{NR} = (1 - t_R)(\bar{Y}_R - \bar{Y}_{NR}) \quad (3.5)$$

이것을 다시 말로 표현하면,

$$\begin{aligned} \text{무응답편향} &= (\text{기대 무응답률}) \\ &\times (\text{응답자의 평균과 무응답자의 평균의 차이}) \end{aligned}$$

이고, 상대편향은 다음과 같다.

$$RB_{NR} = (1 - t_R) \frac{\bar{Y}_R - \bar{Y}_{NR}}{\bar{Y}} \quad (3.6)$$

여기서 무응답 때문에 생기는 상대편향은 편향을 추출틀 모집단 평균으로 나눈 것임을 보여주고 있다.

예 3.4.1 전화조사에서 응답률이 75%이고, 따라서 식 (3.6)에서 $t_R=0.75$ 로 가정한다. 무응답률은 $t_{NR} = 1 - t_R$ 이다. 또한 응답자의 평균소득이 $\bar{Y}_R = \$107,000$ 이고 무응답자의 평균소득이 $\bar{Y}_{NR} = \$89,000$ 이라고 가정하자. 그러면 $\bar{Y} = 0.75 \times \$107,000 + 0.25 \times \$89,000 = \$102,500$ 이다. 식 (3.6)에 있는 상대 편향공식을 이용하면 무응답 때문에 생기는 상대 편향은

$$RB_{NR} = \frac{0.25(107,000 - 89,000)}{12,500} = 0.044 \text{ 또는 } 4.4\%$$

이다. 따라서 25% 무응답률의 결과 4.4%의 상대편향이 발생했다고 말한다.

앞에서처럼 만약 관심의 대상이 되는 특성에 대하여 무응답자의 평균에 대한 정보가 없거나 응답자와 무응답자의 상대 차이를 다른 방법으로 계산할 수 없다면, 포함편향에 대한 불충분한 정보가 주어진 경우에 그랬듯이, “가정”분석을 할 수 있다. 만약 응답률을 늘릴 수 있거나 만일 그것이 더 낮으면 또는 응답자와 무응답자의 소득의 차이가 더 크면 우리는 어느 정도의 편향이 생기는지 알고 싶어 한다.

이런 목적으로 <그림 3.7>에 나온 무응답 때문에 생기는 상대편향과 응답자와 무응답자의 상대차이의 관계를 사용 할 수 있다. <그림 3.3>에서 그랬듯이 <그림 3.7>에서 X축은 응답자와 무응답자의 상대적인 차이를 그리고 Y축은 상대 편향을 나타낸다. 그림에 있는 각 선은 다른 응답률과 상응한다. 한 예로 응답률이 70%이고 응답자와 무응답자의 상대 차이가 20%에서 30%사이라고 생각되는 상황에서 무응답 때문에 생기는 상대편향을 그림에서 측정하자. 상대 차이가 작은 경우, X축에서 20% 지점을 찾고 70% 응답률 선으로 읽어 올라간다. 직각으로 Y축에서 해당하는 위치를 보면 상대편향이 대략 6%로 추정된다. 상대 편향의 상한을 얻기 위해서는 비슷하게 30%로 이 과정을 반복하면 약 9% 정도의 상대편향을 추정한다. 그러므로 상대편향의 범위는 6%에서 9% 사이이다.

상대 무응답편향(%)

|
|
|



<그림 3.7> \bar{Y}_R 와 \bar{Y}_{NR} 의 상대차이와 응답률 t_R 의 함수인 상대 무응답편향

<그림 3.7>은 또한 무응답 때문에 생기는 편향은 단지 응답률만의 함수가 아니라라는 것을 강조하는 역할을 한다. 조사에서 응답률을 상당히 강조하고 응답률이 낮은 어떤 조사들은 이에 따른 편향에 대한 구체적인 평가 없이 낮은 응답률 때문에 조사가 중단되기도 하는데, 편향은 관심 대상이 되는 특성에서 응답자와 무응답자의 차이에도 영향을 받는다. 실제로는 응답률이 예상보다 설사 낮다 할지라도 편향은 만족스러울 수도 있다.

무응답 편향 공식에서 볼 수 있듯이 무응답 편향은 응답자와 무응답자의 차이와 무응답률에 따라 좌우된다. 만일 우리가 측정하는 특성에 대하여 무응답자가 응답자와 많이 다르지 않으면 응답률이 비록 낮을 지라도 응답 편향이 상당히 작을 수도 있다. 그러므로 무응답 편향을 평가하는 데에는 두 가지 정보가 필요하다. 응답률을 알아야 하고 또한 응답자와 무응답자의 특성의 차이를 알아야 한다. 많은 조사에서 후자에 대해서는 파악이 어려운 데, 그 경우라면 무응답률이 높은 조사에 대해서는 조사의 가치를 의심하게 된다.

무응답편향(*nonresponse bias*)은 무응답률과 관심 대상인 특성의 응답자와 무응답자의 차이의 함수이다.

3.5 응답률의 계산

앞에서 살펴본 무응답률 개념은 모든 사람이 이해하고 동의할 수 있다. 하지만

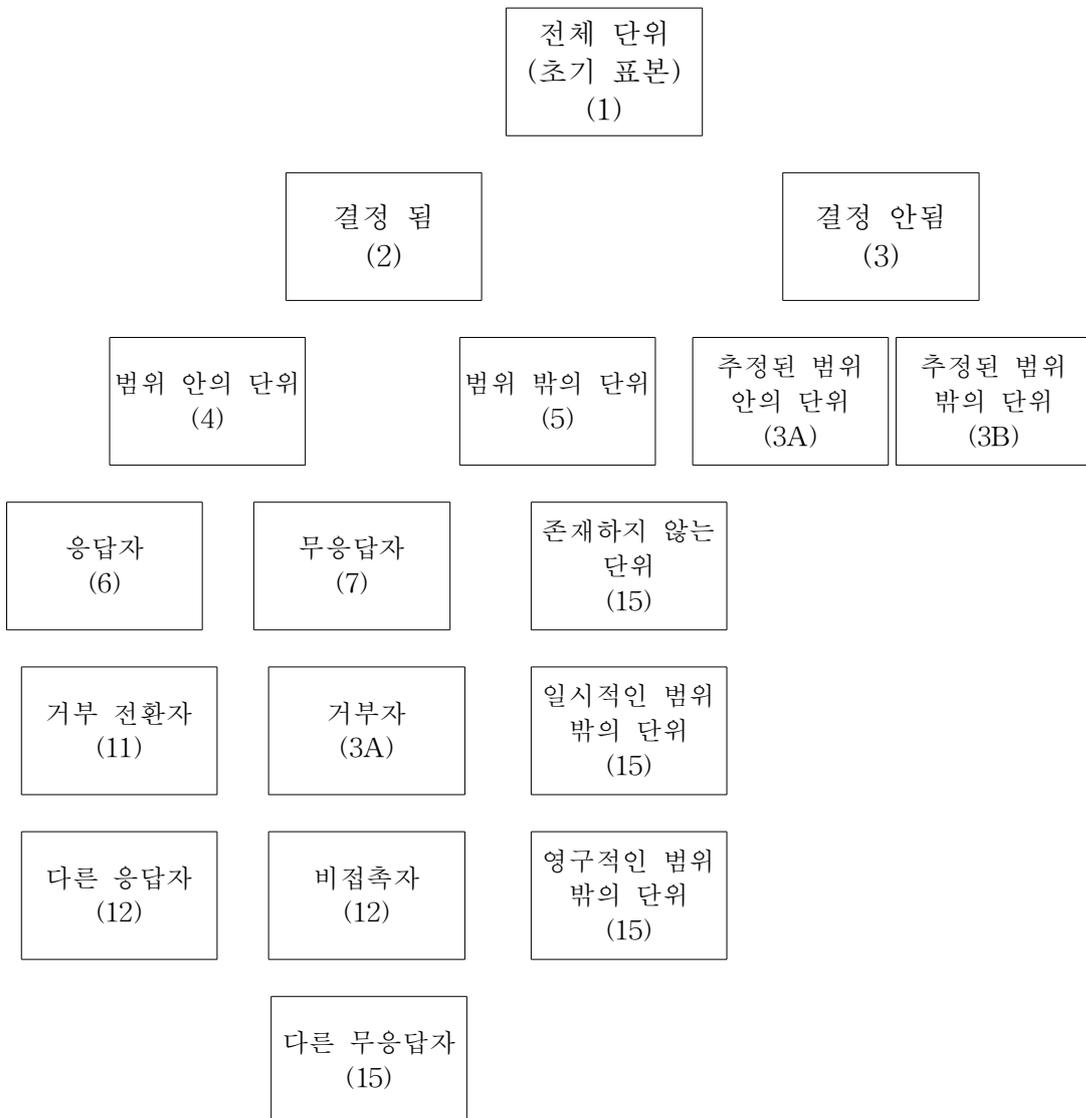
무응답 때문에 표본에 있는 모든 단위들에 대한 적격성을 판정할 수 있는 것은 아니기 때문에 보편적으로 인정된 응답률에 대한 명확한 정의는 없다. 이것은 조사들의 응답률을 비교할 때 신중할 필요가 있다는 것을 의미한다.

응답률 이외의 자료 수집 과정의 여러 가지 양상을 알아보는데 유용한 몇 개의 척도를 자료 수집을 위한 접촉 결과에 대한 자료를 이용해서 계산할 수 있다. 이런 척도 중에는 자료 수집 과정의 여러 가지 양상과 관련되고 중요한 자료수집 과정변수에 대한 척도의 역할을 하는 것도 있다. 이러한 척도를 가지고 있음으로서 자료수집과정을 관리하는 것이 가능해지고 또한 과정을 개선하여 무응답률을 감소시키는 결과를 가져오며 많은 경우에는 무응답편향을 감소시킨다. 그러므로 이러한 비율들에 대해 알아보고 이들의 산출 방법에 대해 살펴보자.

주요한 과정변수들에는 예를 들어 접촉률, 협조율, 거부율, 거부전환률 등이 있다. <그림 3.8>은 Hidrogou 등(1993)에서 인용하였는데, 그들은 응답률을 계산하기 위한 모형과 캐나다 통계청에서 사용되는 다른 자료 수집 과정 통계에 대해 논의하고 있다. 이 그림에서 흐름도의 맨 위 첫 상자에 있는 전체표본은 결정된 단위와 결정되지 않은 단위의 두 부류로 나뉜다. 결정됨(*resolved*)이라 함은 단위가 범위안에 있는지(목표 모집단의 구성원이고 조사에 포함되어야 함) 아니면 범위 밖에 있는지(목표 모집단의 구성원이 아님)를 결정했다는 것을 의미한다. 예를 들어 농장 조사에서 농업에 종사하지 않는 가구들은 목표 모집단의 구성원이 아니기 때문에 조사범위 밖에 있다. 어떤 경우에는 단위가 범위안에 또는 밖에 있는지 알 수 있는 정보가 충분하지 않을 수 있다. 예를 들어 조사 기준에 의해 농업에 종사하는 것으로 분류될 수 있는 가구가 결정에 필요한 정보를 제공하기를 원하지 않아서 결정되지 않은 것으로 분류될 수 있다. 이러한 상황은 통화를 시도한 전화번호 중에 실제 전화를 받지 않는 경우가 많은 RDD 전화조사에서 매우 흔히 발생한다. 그 번호의 누구하고도 통화를 하지 못하고 그 번호에 대해 아무 다른 정보(예를 들어 그것이 가구 또는 사업체와 관련 있는지 아니면 공중전화 또는 다른 종류의 전화인지)도 알 수 없기 때문에 그 번호를 가구로 또는 조사의 범위 밖으로 분류할 수 없다. 그러므로 그 번호는 결정되지 않은 것으로 구분된다.

어떤 응답률 척도에서도 분모는 조사범위 안에 있는 단위들의 수이기 때문에 응답률을 계산하기 위해서는 결정되지 않은 단위 중에서 어떤 비율이 범위 안에 있는지 추정해야 한다. 결정되지 않은 단위에 대해 어떤 정보가 있는 지에 따라 이를

위한 몇 가지 방법이 개발되었다. 만약 가구 단위에 거주하는 것으로 보이거나 전 화번호 단위가 사용되는 것으로 보이는 것 이외에는 아무 정보가 없으면, 결정된 단위 중 조사범위 안에 있는 결정된 단위의 비율을 이용하여 결정되지 않은 경우의 범위 안에 있는 단위의 수를 추정(비록 부정확하지만) 할 수 있다. 결정된 경우에 $p_{resolved}$ 가 조사범위 안에 있는 비율을 나타내고 $n_{unresolved}$ 가 결정되지 않은 수를 나타낸다고 하자. 그러면 결정되지 않은 경우 중 범위 안에 있는 단위의 수는 결정되지 않은 단위의 수에 결정된 경우 중에서 범위 안에 있는 단위의 비율을 곱해 구할 수 있다(즉 $p_{resolved} \times n_{unresolved}$).



<그림 3.8> 응답과 무응답요소에 의한 최종 표본의 배치[Hidioglou 등(1993)에서 인용]

이제 조사범위 안에 있는 단위를 생각하면 이것들도 응답 단위 또는 무응답 단위의 두 가지 종류로 분류될 수 있다. 나아가서 응답 단위는 한번도 거부하지 않은 단위와 거부전환(refusal conversions)이라고 하는 처음에는 거부했지만 결국은 응답자로 전환한 것으로 구분될 수 있다. 무응답 단위는 다시 최후거부, 한번도 연락 또는 접촉을 못한 단위, 다른 종류의 무응답 단위의 세 그룹으로 나눌 수 있다. 여기서 “다른” 그룹은 연락은 되었지만 질병, 면접기간 동안 만날 수 없어서, 언어장벽 등의 이유로 조사에 참여 할 수 없는 단위로 이루어진다. 범위 밖에 있는 단위도 또한 세 그룹으로 다시 나눌 수 있다. 한 그룹은 존재하지 않는 단위(*nonexistent units*)로 이루어진다(즉 추출틀에는 나타나지만 더 이상 존재하지 않는 단위). 존재하지 않는 단위의 예는 해체된 가구 단위 또는 RDD 조사에서 더 이상 사용하지 않는 전화번호 등이다. 만일 표본 추출틀이 상당히 오래된 것이라면 존재하지 않는 단위수는 상당히 클 수 있다.

일시적으로 범위 밖에 있는 단위(*temporary out-of-scope units*)는 한 때에는 범위 안에 있었지만 지금은 범위 밖에 있는 단위다. 이것은 추출틀을 만들 때는 농장의 정의를 만족시켰지만 농업판매가 줄어서 더 이상은 조건을 만족시키지 못하는 농장일 수 있다. 이 단위는 언젠가 다시 농장의 정의를 만족 시킬 수 있다. 마지막으로 영원히 범위밖에 있는 단위(*permanently out-of-scope units*)는 범위밖에 있고 그것들이 범위 안에 다시 들어올 가능성이 없는 단위이다. 이것들은 사업을 매각하고 다시는 농가의 일원이 될 의도가 없는 농장 같은 것들이다.

조사 프로젝트를 진행하는 동안에 자료수집 책임자가 자료 수집의 과정을 모니터하기 위해 여러 가지 비율에 관심이 있을 수 있다. 응답률을 물론이고 또한 협조율, 언어 장애율, 접촉률, 결정되지 않을 비율 등이다. 1장에서 자료 수집 과정을 계속적으로 개선하기 위해 조사의 주요 과정변수와 관련된 비율을 모니터하는 것의 중요성에 대해 논의 했다. 조사 설계자는 원하는 개선방향에 다른 모니터링을 위해 적절한 변수와 비율을 선택할 수 있다.

하지만 응답률은 조사의 품질을 판단하기 위해 가장 널리 보고 되고 비교 되는 조사과정 관련 통계량이지만 이에 대한 표준화가 필요하다. 그러나 응답률 계산방법의 표준화에 대한 진전은 느리게 진행되고 있으며, 여러 조사에 대한 응답률들을 비교할 때 각 조사에서 응답률의 계산에 사용된 방법을 정확히 이해하는 것이 논리

적으로 매우 중요하다. 많은 국가 통계기관들은 우리가 방금 논의한 캐나다 통계청
에서와 같이 개념화를 토대로 한 응답률을 계산하는 표준을 갖고 있다. 하지만 조
사를 실시하는 여러 가지 행정상의 요인들로 인하여 개념화하는 것이 기관들과 나
라들에 따라 상당히 다르다.

최근에 미국여론조사학회(AAPOR, 2001)의 연구자들이 RDD 전화조사와 개별 가
구조사를 위한 완성률을 계산하는 지침을 제공했다. AAPOR의 지침에는 여섯 가지
의 응답률을 계산하는 공식이 있다. 또한 비교대상이 되는 비율들의 계산방법을 이
해하는 것은 매우 중요하다는 것을 밝히고 있다. 현재 RDD 전화조사와 개별 가구
조사와 관련된 무응답 문제에 대한 논문의 발표에서 조사연구 관련 정기 학술간행
물의 편집위원들은 AAPOR 지침을 따르도록 적극 권장하고 있다.

<그림 3.8>에서 계산할 수 있는 몇 가지의 유용한 비율에 대해 살펴보자. 우선
조사의 응답률부터 생각해 보자. 응답률은 기대응답률 t_R 의 추정값인 것을 상기하
라. 만일 우리가 t_R 을 정확히 추정할 수 있으면 무응답 편향의 가능성에 접근하기
위해 이용할 수 있는 최고의 정보를 갖고 있는 것이다. 전에 언급했듯이 물론 편향
에 대한 완전한 파악을 위해 응답자와 무응답자의 차이를 아는 것이 필요하다. 그
럼에도 불구하고 이 정보가 없을 때에는 t_R 의 추정이 목표가 된다.

간단히 말하면 조사 응답률은 단지 응답자 수를 조사의 범위 안에 있는 표본구성
원의 수로 나눈 것이다. 균등 확률로 각 표본단위를 추출했을 때 이것이 t_R 에 대한
최량(best)의 추정량이다. 가중된 추정량(weighted estimators)이라고 하는 다른 추
정량은 나중에 논의 하겠지만 불균등 확률 표본에서 t_R 을 추정하기 위한 것이다.

응답률을 계산하는 표준화된 기준이 일반적으로 인정되지 않는 중요한 이유 중
하나를 응답률의 분모를 계산하는 데 따른 어려움 때문이다. 이런 문제는 표본에
있는 결정되지 않은 단위의 존재 때문에 발생한다. 응답률(response rate)을 계산하
는 타당한 방법은 상자(6)의 응답한 단위 수를 그림에서 상자(4)와 (3A)에 해당하는
범위 안에 있는 것으로 알려진 것과 추정된 범위 안에 있는 단위수를 포함하는 모
든 범위 안에 있는 단위 수로 나눈 것이다. 따라서 응답률은 $[6]/([4]+[3A])$ 이다. 응
답률의 계산에서 상자(3A)에 해당하는 추정된 범위 안에 있는 단위 수를 제외하는
오류를 범하기 쉽다. 응답률을 계산하는 사람이 결정되지 않은 단위는 모두 범위
밖이라는 잘못된 가정을 하는 경우가 있다. 예를 들어 RDD 전화조사에서 결정되지

않은 전화번호가 조사기간 중에 휴가이거나 다른 이유로 연락되지 않았던 사람의 것일 수도 있다.

결정되지 않은 단위 중 범위 안에 있는 단위의 수를 추정하기 위한 아무런 정보가 없을 때, 간단하고 널리 인정된 방법은 결정된 단위 중 범위 안에 있는 단위의 비율을 추정해서 결정되지 않은 전체 단위 수에 곱하는 것이다. 바꿔 말하면 상자(4)에 있는 경우의 수 나누기 상자(2)에 있는 경우의 수는 결정된 경우에 범위 안에 있는 비율이다. 그러면 이 수량 곱하기 상자(3)에 있는 수가 상자(3A)) 들어갈 수가 된다. 이것은 다음과 같이 표현될 수 있다.

$$[3A] = \frac{[4]}{[2]} \times [3]$$

또한 $[3B] = [3] - [3A]$ 이다.

조사 연구에서 자주 계산하는 또 하나의 비율은 협조율(*cooperation rate*)이다. 협조율은 조사에서 접촉된 단위들이 조사에 참여하는 것에 동의한 정도를 나타낸다. 따라서 그래서 상자(6)의 응답한 단위 수를 상자(6)과 상자(13)에 있는 수의 합인 연락된 총 단위수로 나눈 것이다. 그러므로 협조율은 $[6]/([6]+[13])$ 이다. 계산할 수 있는 다른 비율들에는 거부율(*refusal rate*), 무응답률(*nonresponse rate*)과 비접촉률(*noncontact rate*) 등이 있다. 이들 비율을 요약하면 아래와 같다.

응답률 :	$\frac{[6]}{[4]+[3A]}$
협조률 :	$\frac{[6]}{[6]+[13]}$
거부율 :	$\frac{[13]}{[4]}$
무응답률 :	$\frac{[7]+[3A]}{[4]+[3A]}$
비접촉률 :	$\frac{[14]+[3A]}{[4]+[3A]}$

지금까지 우리는 가중값을 주지 않은 응답률만 이야기해 왔다. 그렇지만 앞에서

말했듯이 서로 다른 선택확률을 사용해서 표본을 추출할 때에는(9장 참조) 가중된 응답률이 t_R 에 대한 보다 더 적절한 추정값일 수 있고, 이는 평균제곱오차에서 무응답편향의 잠재적인 영향에 대한 더 좋은 척도이다. 표본의 선택 확률이 단위마다 상당히 다르면 가중된 경우와 가중되지 않은 경우의 응답률이 상당히 다를 수 있다. 예를 들어 지난해 체포된 사람을 지나치게 과대 포함한 표본을 뽑는다고 가정하자. 예를 들어 이 부차모집단은 모집단 전체의 3%를 차지하지만 그것이 표본의 50%를 차지한다. 이 그룹의 응답률은 40%이고 나머지 표본의 응답률은 80%라고 가정하자. 그러면 가중되지 않은 응답률은 이 두 응답률의 평균인 60%이다. 하지만 가중된 응답률은 피체포자 응답률에 모집단에서 피체포자의 비율인 3%, 피체포자가 아닌 그룹에 97%의 가중값을 줄 것이다. 따라서 가중된 응답률은 $0.03 \times 40 + 0.97 \times 80 = 78.8\%$ 로 18.8%포인트의 차이가 난다. 가중된 응답률의 계산은 표본의 가중된 비율을 추정하는 것과 근본적으로 같다. 이는 9장에서 다룬다.

가중값을 적용하지 않은 응답률은 표본 구성원에게서 응답을 얻어내기 위한 현장 조사의 성공의 척도이기 때문에 역시 중요한 과정변수이다. 그 점에 있어서는 가중값을 주지 않는 응답률이 현장조사과정을 모니터하는데 그리고 현장조사가 진행중 인 동안 대처할 수 있는 무응답 문제를 확인하는데 유용하다. 하지만 이것이 반드시 무응답이 추정값의 정확도에 주는 영향에 대한 척도는 아니다. 이런 관점에서는 가중된 응답률이 더 적절하다. 왜냐하면 가중된 응답률은 <그림 3.6>의 어둡게 표시된 부분에 있는 모집단의 비율인 t_R 의 추정값이다. 이미 언급했듯이 무응답편향은 t_R 과 응답자와 무응답자의 평균의 차이의 함수이다. 그러므로 t_R 은 무응답이 편향에 미치는 영향의 척도라고 볼 수 있다. 불균등 확률 표본에서는 가중된 응답률이 t_R 의 추정량인데 반해서 가중되지 않은 응답률은 적절한 것이 아니다.

우리가 사용한 비율들은 캐나다 통계청이 사용하는 동시에 Hidiroglou 등이 제시한 활용 가능한 방법들이다. 그렇지만 여러 기관에서 개발되고 사용된(AAPOR가 개발한 것 같은) 몇 개의 이런 방법들이 있고, 이런 기관들은 이와 관련된 표준화된 기준을 세우기 위한 노력을 시작했다. 또한 가능한 여러 가지의 비율이 조사 세계의 복잡함을 명백히 보여준다. 우리는 여기에서 더 복잡한 시나리오에 대해 이야기 하는 것을 피하였다. 예를 들어 학교조사에서는 표본 설계가 계층적일 수도 있다. 첫 단계에서 학교가 선택되고 선택된 학교 내의 교사를 선택한다. 여기에서 무

응답은 두 단계에서 모두 발생할 수 있다. 이런 다 단계의 조사에서 무응답률은 결과적으로 증가하게 된다.

또 하나의 복잡한 예는 표본 단위를 시간의 흐름에 따라 계속 추적 조사하는 다시점 조사이다. 다시점 조사에서는 추적조사 또는 조사과동(*survey wave*)에 따라 무응답 또는 표본마모(*sample attrition*)가 시간이 흐름에 따라 계속 증가하게 된다. 또한 다시점 조사 표본에서는 단위들이 가구크기의 변화와 같이 시간에 따라 바뀔 수도 있어 상황을 더 복잡하게 한다. 미국 연방통계방법위원회(*U.S. Federal Committee on Statistical Methodology*, 2001)가 이런 복잡한 상황에서 무응답률을 계산하는 방법을 검토하였다.

3.6 무응답오차의 축소

무응답률은 무응답 편향의 두 가지 결정요소 중 하나다. 무응답률이 관심대상 변수에 대한 응답자와 무응답자의 관심의 대상이 되는 변수의 값의 차이인 두 번째 결정요소보다 일반적으로 줄이기 쉽다. 그러므로 응답률을 증가시키는 조사 방법을 개발하기 위해 많은 노력을 기울였고 그 결과 조사의 종류에 따라 차이는 있지만 응답편향의 위험을 상당히 줄일 수 있다.

자료 사용자와 생산자는 만족스러운 응답률은 얼마인가? 라는 질문을 자주한다. 이미 무응답편향의 공식을 분석했으므로 우리는 상황에 따라 좌우된다고 쉽게 답할 수 있다. 이러한 답이 별로 도움이 되지는 않겠지만 응답률만 알고 무응답 편향 공식의 두 번째 항에 대해 충분히 알지 못하면 이 질문에 대한 만족스러운 답은 있을 수 없다. 무응답률이 낮은 것이 바람직하고, 반면에 무응답률이 높으면 이에 따른 편향이 커진다. 그럼에도 불구하고 무응답자의 특성에 대한 자료가 일반적으로 없기 때문에 많은 기관들은 무응답 편향의 위험을 제한하는데 효과적이고 달성이 가능한 수용할 수 있을 수준의 응답률 기준을 설정하고 있다.

무응답편향(*nonresponse bias*)은 무응답이 있고 \bar{Y}_R 과 \bar{Y}_{NR} 이 다를 때 발생한다. 무응답 편향을 감소시키려면 t_{NR} 뿐 아니라 $t_{NR}(\bar{Y}_R - \bar{Y}_{NR})$ 을 줄여야 한다. 따라서 $\bar{Y}_R - \bar{Y}_{NR}$ 에 대해 알지 못하면 만족스러운 무응답률에 대해 말하는 것은 의미가 없다. 하지만 무응답률은 무응답 편향의 가능성을 나타내는 유용한 척도이다.

3.6.1 무응답의 원인과 영향

여러 요인들이 무응답률에 영향을 미칠 수 있다. 개인과 가구조사에서 이러한 요인들의 예로는 동기부족; 시간부족; 공공기관에 등록되거나 문서에 기록되는 것에 대한 두려움; 질병, 휴가, 직무, 시기상 문제; 자동응답기; 언어장벽 등이 있다. 이러한 요인들은 사업체조사에서도 나타난다. 그 이유는 사업체 응답자들도 사람이기 때문이다. 사업체조사에서는 추가적인 요인들이 있는데 이런 요인들의 예로는 지정 응답자 결정과 자료수집에 대한 허가 취득의 어려움; 직원변동; 조사 참여에 제한적인 사업체의 정책; 지식부족; 응답자의 낮은 우선권; 응답에 따른 사업체 부담 비용 등이다.

조사기관 내부에도 영향을 미치는 요인들이 있다. 이런 요인들의 예로 면접원 업무량, 너무 짧은 자료 수집기간, 사용하기 부담스럽고 어려운 설문지 등이 있다. 이 밖에 많은 요인들이 있을 것이다. 그 중 어떤 요인들은 다른 것들 보다 훨씬 쉽게 관리할 수 있다. 내부 요인들은 예를 들어 표본 단위의 위치를 찾아내는 것보다 더 다루기 쉬울 것이다. 무응답의 원인이 되는 요인들은 무응답의 일반적인 분류방식인 접촉할 수 없음, 응답 거부, 다른 무응답(질병, 시설에 수용됨, 언어문제 등)으로 구분될 수 있다.

경험에 비추어 보면 무응답률을 줄이는 방법들은 무응답 종류에 따라서 다르기 때문에 무응답을 종류에 따라 분류하는 것은 자료수집을 위해 상당히 유용할 수 있다. 예를 들면 연락하기 힘든 단위들에 대해서는 효과적인 추적 수단을 적용하고 접촉 시도를 효율적으로 반복함으로써 무응답을 줄일 수 있다. 거부는 대부분 설득력 있는 권유와 응답을 가능한 줄임으로써 대처한다. 언어 장벽, 질병 등 같은 무응답의 경우는 대리 응답자(표본으로 선정된 사람 대신에 응답을 해 줄 수 있는 사

람)의 활용으로 처리할 수 있다. 이런 분류를 통해 무응답률을 줄이기 위해 자원을 효율적으로 배분할 수 있을 것이다.

무응답 분류가 도움이 되는 또 하나의 이유는 만일 시간경과에 따른 무응답 패턴의 변화를 기록하면 조사여건(*survey climate*)이라는 모집단 구성원이 조사에 참여하려는 의지와 태도의 변화를 설명하는 개념의 변동을 파악할 수 있다. 이러한 개념의 조사여건은 프라이버시 침해와 응답자 부담 등에 대한 공개적인 토론 등과 같은 특정한 상황이 발생하고 시간이 흐름에 따라 바뀔 수 있다. 이런 요인들은 조사 참여요청에 대한 자발적인 참여 의향에 일시적으로 영향을 미칠 수 있다. <표 3.4>는 무응답의 가장 흔한 여러 이유와 무응답 형태를 정리한 것이다.

<표 3.4> 무응답의 흔한 이유와 그것으로 인한 무응답의 형태

무응답의 이유	대표적인 무응답 형태
의욕이 없음	거부
시간이 없음	거부
등록이 되는 것에 대한 두려움	거부
여행, 휴가 중	접촉할 수 없음
시간이 좋지 않음	거부
목록에 없는 전화번호	접촉할 수 없음
자동 응답기, 전화번호 표시기	접촉할 수 없음
잘못된 주소 또는 잘못된 전화번호	접촉할 수 없음
질병 또는 장애	기타
언어문제	기타
회사 직원 변동	접촉할 수 없음
회사 소유자 변동	접촉할 수 없음
회사의 개편	거부, 접촉할 수 없음, 기타
조사가 너무 어려움	거부
조사에 참여 안하는 회사 방침	거부
낮은 우선권	거부
비용이 너무 많이 듦	거부
민감한 질문	거부
지루한 주제	거부
조사원 업무량 과다	거부, 접촉할 수 없음
자료 수집기간이 너무 짧음	거부, 거부, 접촉할 수 없음
차단	거부
좋지 않은 질문 또는 설문지	거부
이사함	접촉할 수 없음

주된 두 가지 무응답 종류인 접촉할 수 없음(*unable to contact*)과 거부(*refusal*)에 대해 좀더 자세히 생각해보자. 예를 들어 현장면접조사에서 비접촉에 대처하기 위해서는 가구나 농장이나 사업체에 연락이 될 때까지 다른 시간대에 그리고 다른 요일에 반복해서 접촉을 시도하는 것이 바람직하다. 만일 면접원이 조사 질문에 응답할 수 있는 집에 있는 사람을 찾을 수 없다면, 최소한 그 가구의 거주자 현황과 가구의 특성에 대한 정보를 줄 수 있는 이웃을 찾는 것이 가능할 것이다. 전화조사에서도 거의 같은데 매 요일마다 하루에 여러 차례 응답자에게 전화를 거는 것이다. 비접촉으로 인하여 무응답률이 증가하는 것을 막고 무응답 편향을 줄이기 위한 주

된 전략은 결국 포기해야 될 때까지 계속해서 표본 단위와의 접촉을 시도하는 인내력이다.

반면에 거부는 해결하기 훨씬 더 어려운 문제이다. 사람들이 조사 참여를 거부하는 이유에 대한 많은 연구가 있다. 만일 우리가 그 이유들을 충분히 잘 이해할 수 있으면 사람들이 조사에 참여하도록 더 잘 설득할 수 있을 것이다. 사람들이 조사에 참여하고 싶어 하지 않는 이유에 대한 많은 정보가 있는데, 많은 조사들은 이런 정보를 정기적으로 수집한다. 즉, 조사원들은 조사단위들이 거부할 때 왜 거부하는지를 묻는다. 비록 사람들이 참여하는 것을 거부할 지라도 이유를 말하는 것은 꺼리지 않을 수도 있다. 예를 들어 RTI(Research Triangle Institute)가 실시한 연구에서 거부자의 42.7%가 그들은 관심이 없다고 응답했다. 만일 모집단의 대부분이 주제가 지루하거나 따분하다고 생각하면 이에 대해서 조사설계자가 취할 수 있는 대책이 별로 없기 때문에 이런 연구결과는 흥미롭기는 하지만 별로 유용하지 않다. 표본 단위가 응답을 거부하는 더 복잡한 근본적인 이유가 중요하다. 면접원이 조사 참여에 관심을 가지게 하지 못하거나 면접원의 말이 지루하게 들려서 표본단위가 관심을 갖지 못할 수도 있다. 또한 표본 단위가 조사가 무엇에 대한 것인지 완전히 이해하지 못해서 관심을 갖지 못할 수도 있다. 표본 단위가 관심이 없다고 말하는 것이 조사에 참여하기를 거부하는 쉬운 방법일 수도 있다.

사람들이 왜 조사에 참여하고 싶지 않아하고 그들의 마음을 어떻게 바꿀 수 있는지에 대한 이해를 높이기 위해 우리가 할 수 있는 것이 더 있다. 최근 몇 년 동안 조사 참여에 대한 이론의 개발에 관한 중요한 연구가 본격적으로 착수됐다. 그 중 중요한 연구는 Groves 등(1992)과 Groves와 Cooper(1998)에 의한 것들이다. 그들의 연구는 개인과 가구조사에 대한 것이지만 연구결과 중 일부는 사업체 조사에도 관련이 된다. 그 이유는 사업체 조사에서도 사람을 통해서 조사 연구자가 찾는 정보를 전달받기 때문이다(현지 조사원이 직접 관찰을 하는 경우나 TV 시청상황을 조사하는 기계적 장치를 사용하여 자료를 기록하는 경우 등은 해당 안됨). 대부분의 사람들은 자기 자신에 대해 응답하건 사업체를 대신해서 응답하건 비슷하게 반응을 나타낸다. 하지만 사업체조사는 개인과 가구조사와 다른 점이 많다. 이 차이에 대한 유익한 논의는 Riviere(2002)에서 찾을 수 있다.

Groves 등(1992)은 조사방법론에 대한 많은 연구논문에서 논의된 거부에 영향을 미치는 요인들을 제시하였다. 조사 설계, 자료수집 방법, 응답자 규정, 면접의 길이,

면접기간, 조사 주제, 그리고 설문지 설계 모두가 거부에 영향을 미칠 수 있다. 응답자의 특성도 거부율에 영향을 미친다. 나이, 성별, 소득, 응답자의 보건, 거주지역(도시/시골), 거주지역의 범죄율, 응답자 문맹여부가 그들의 응답하는 경향을 좌우할 것이다. 면접원 특성도 응답률을 좌우한다. 그들의 경험, 응답을 꺼리는 응답자를 다루는 방법, 자신감, 기대와 선호 이외에 나이, 성별, 인종, 면접원이 감지한 소득 등도 응답률에 영향을 미칠 수 있다. 면접원들은 사람들에게 친절히 다가가면 조사에 쉽게 응답할 것이라고 믿는가 아니면 많은 경우에는 이것도 소용없는 것이라고 생각하는가? 영향을 주는 다른 요인들은 면접에서 그들이 보여주는 자세와 그들의 최근 경험들이다. 응답거부가 많았던 날에는 자신감을 일시적으로 잃을 수 있다. 낮은 조사급여나 그들의 감독자가 높은 응답률의 중요성을 충분히 강조하지 않았기 때문에 참여를 하게 하려고 하는 동기 부여가 낮아질 수 있다. 마지막으로 사회적 책임과 조사 목적의 정당성과 같은 사회적 요인도 거부율에 영향을 줄 수 있다.

무응답을 초래하는 하나의 요인이나 유사한 요인들의 그룹에 대해 무응답을 줄이는 한 가지 이상의 수단이나 기술이 있다. 이 기술 중 몇 개를 검토해 보겠지만 적절한 응답률을 확보하기 위해서는 단순히 한 가지 방법만을 활용하는 것으로는 충분하지 않다는 것을 처음부터 강조하고자 한다. 이러한 방법들 중에서 각각의 방법은 무응답에 기여한다고 생각되는 많은 요인 중 한 가지 이상에 대처할 수 있어야 한다. 또한 이 방법들은 서로 다른 조사들에 있어서 동일하게 적용될 수 없기 때문에 특정한 조사의 필요성에 맞춰 조정되어야 한다. 여러 가지 방법이 쓰여야 하는 또 하나의 이유는 서로 다른 방법들은 여러 모집단 그룹에게 다르게 작용하기 때문이다. 이론적으로는 예를 들어 젊은 사람들이나 중간 규모 농장들에 대한 응답률을 훨씬 높이기 위해 특정한 한 가지 수단을 사용하는 데에 주안점을 두는 것은, 동시에 응답자와 무응답자의 특성상의 차이를 증가시킬 수 있기 때문에 무응답 편향을 줄이는 데 도움이 안 될 수도 있다.

응답자와 무응답자의 차이에 대한 정보가 부족할 때, 높은 무응답률이 자료 품질에 미치는 영향이 명확하지는 않지만 그럼에도 불구하고 조사 과정에서 무응답을 절대 가볍게 받아 들여서는 안 된다. 첫째로, 무응답 때문에 우리가 얻은 표본크기가 계획됐던 것이나 또는 만일 응답률이 더 높았다면 가능했을 표본크기보다 작아진다는 것을 의미한다. 9장에서 거론되듯이 표본크기가 작아지면 추정의 정도가 낮아질 것이고, 그것은 관심 대상인 모수의 신뢰 구간이 넓어진다는 것을 의미한다.

둘째로, 조사에서 최소한 몇 개의 항목에 대해서는 응답자와 무응답자의 특성에 차이가 있다. 예를 들어 1991년에 스웨덴 노동력조사에서 무응답 편향 때문에 실직한 사람 수가 8% 과소평가되었다. 그 이후에 보조정보를 이용해서 새로운 추정량이 개발되었고 그 결과 편향을 상당히 줄였다. 셋째로, 무응답은 통계 생산 과정에 혼란을 일으키고 비용의 증가와 조사를 지연 시키는 결과를 가져올 수 있다. 넷째, 무응답은 많은 조사 이용자와 의뢰자가 잘 알고 있는 조사품질을 나타내는 특성이고, 그들은 양질의 조사에서는 무응답률이 높게 나타나지 않는다고 직관적으로 알고 있다. 자료 사용자들은 자료가 어떻게 수집되었고, 자료 품질을 개선하기 위해 모든 단계에서 이루어진 노력에 대해 거의 모른다. 그러므로 단순히 조사 응답률을 2장에서 논의한 비표본추출오차의 모든 5가지 근원을 포함한 조사과정 전체의 품질에 대한 척도로 보는 경향이 있다. 사용자들에게는 무응답률이 조사기관의 능력을 암시 하는 것일 수도 있다. 높은 응답률은 효율적인 양질의 자료수집 과정의 수행과 같은 뜻으로 이해된다. 그래서 계속 조사 업무를 수행하려면 무응답률을 계속 낮게 유지해야 할 현실적인 이유가 있다. 그렇지만 긴 안목으로 보면 조사 공동체는 조사품질 척도에 관한 조사 의뢰자와 사용자의 이해를 돕기 위한 교육에 힘써야 한다.

한 가지 수단만으로는 적절한 수준의 응답률을 얻기에 충분하지 않다. 무응답 원인이 되는 하나의 또는 여러 가지의 요인들에 대처할 수 있는 방법들의 모음(*collection*) 또는 일련의 방법(*battery of methods*)이 필요하다.

3.6.2 조사 참여 이론

조사 참여 이론의 발달에는 두 가지 부류의 접근방식이 있다. 하나는 면접조사와 관련이 있으며 조사 참여를 설명하기 위해 심리적인 개념을 사용하는 것이다. 또 하나는 가구방문 면접에서 면접원과 응답자 사이의 면접을 개선하기 위한 사회성 능력(*social skill*)에 대한 분석이다(Morton-William, 1993). 우편 또는 다른 자기기입식 자료수집방법에 대해서는 면접조사와도 밀접한 관계가 있는 사회교환이론(*social exchange theory*)을 토대로 한 Dillman(1978, 1991)이 주도한 연구분야가 있

다.

조사 참여의 기초적인 이론은 Cialdini(1990)가 개발했고, Groves와 Couper(1998)가 더욱 발전 시켰다. 이 이론은 사회 인구통계학적인 그리고 조사 설계 요인에 의한 영향과 면접원과 응답자 사이의 상호작용에 따른 심리적인 요소에 따른 명확하지 않은 영향을 통합한 것이다. 이와 관련된 요인들은 면접원과 표본단위의 사회적 수준, 조사 설계의 속성, 표본으로 추출된 사람의 특성, 면접원의 속성, 그리고 응답자와 면접원의 상호작용 등이다. <그림 3.9>는 조사 참여에 영향을 미치는 것으로 생각되는 여러 요인들을 보여준다(Groves와 Couper, 1993 참조). 이 조사 참여 개념의 체계에서는 면접원의 행동은 매우 표준화 되어야 한다는 전통적인 생각을 어떤 면에서 부정한다. 실제적으로 그들의 이론의 핵심은 목적에 맞추는 기술이 면접원 교육의 일부가 되어야 하고 개인이나 가구의 면접조사에서 조사 도구나 요소는 표본에 속한 사람의 특정한 속성에 맞춰져야 한다는 것이다. 목적에 맞추는다는 (*tailoring*) 것은 면접원들이 특정한 표본단위에 접근할 때 그 대상 표본단위의 특성에 맞게 모든 것을 하라는 것을 의미한다. 면접원들은 조사요구에 대한 응답자의 수락을 용이하게 하는 것과 관련된 특성에 집중하려고 노력해야 한다.

Cialdini(1984)는 거부율에 영향을 미칠 수 있는 여섯 가지의 심리적인 요인들을 밝혔다. 이러한 심리적인 요인들 중 첫 번째는 보상(*reciprocation*)이다. 기본적으로 Cialdini가 제안한 것은 만일 사람들이 조사 참여를 선물, 보수, 그들에게 주어진 다른 보상에 대한 보답으로 보거나 그것이 그들에게 도움이 된다고 생각하면 사람들은 조사에 응답하고 싶은 생각이 더 들 것이다. 이것이 응답자에게 인센티브를 주는 과학적인 논거다. 관련 연구에서는 사전에 주는 인센티브가 사후에 약속된 인센티브보다 더 효과적이라는 것을 보여준다(Berk 등, 1987). 이 결과 많은 우편조사는 설문지와 함께 \$2에서 \$5의 인센티브를 동봉한다. 이것이 사전 인센티브(*prepaid incentive*)의 의미이다(즉, 응답자들은 그들이 응답하기 전에 보상을 받는다).

연구자가 통제 할 수 없는 요인

연구자가 통제할 수 있는 요인

- 사회적 배경**
- 정치적
 - 경제적
 - 사회적
 - 조사에 임하는 여건
 - 주변 특성

- 조사설계**
- 응답자 부담
 - 응답자 선택
 - 조사주제
 - 조사운영방법

- 응답자**
- 가구구조
 - 사회인구통계학적인 특성
 - 주제에 대한 지식
 - 조사경험
 - 정서상태
 - 심리적인 경향

- 면접원**
- 사회인구통계학적인 특성
 - 경험
 - 예상
 - 정서상태

응답자-면접원 상호작용

참여 또는 거부 결정

<그림 3.9> 가구 조사에서 참여 또는 거부 결정에 영향을 미치는 요인들. [Groves와 Couper(1993)에서 인용]

인센티브에 대한 연구논문들은 응답자들이 약속된 인센티브를 서비스에 대한 보수로 본다고 말한다(Singer 등, 1999 참조). 어떤 응답자들은 그들의 시간에 대한 보수로 충분하지 않을 수 있기 때문에 약속된 인센티브는 사전 인센티브만큼 효과적이지 못한 경향이 있다. 또한 보상이론에 의하면 만일 표본구성원이 조사가 그들에게 이롭다고 느끼면 그들은 조사에 응할 것이다. 경험이 많은 면접원들은 때로는 이런 측면을 이용하기 위해 응답자에게 이 조사가 응답자가 속한 그룹에게 얼마나 중요한지를 설명한다. 예를 들어 노인의 의료보호 납입과 관련된 조사에서 면접원이 이 조사의 결과 의료보호제도를 담당하는 기관에게 서비스, 지불방식, 또는 그 프로그램의 다른 특성을 개선할 수 있는지에 대한 정보를 주어서, 결과적으로 이 조사가 어떻게 의료보호대상자를 이롭게 하는지 응답자에게 알림으로써 참여를 유

도할 수 있다.

두 번째 심리적 요인은 일관성(*consistency*)이다. 이 요인은 조사 요구가 승낙되기 위해서는, 표본구성원의 지위, 의견, 태도 또는 가치관과 일치하도록 조사해야 한다는 것이다. 예를 들어 만일 면접원과 응답자가 서로 인사하는 중에 응답자가 환경에 대해 매우 관심이 많다고 한 경우, 만약 조사에 환경과 관련된 사항이 있다면 면접원이 일관성 있게 이 조사 참여가 환경에 어떻게 도움을 줄 것인가를 응답자에게 알릴 수 있다. 표본구성원이 환경에 대한 그들의 관심을 이미 나타냈을 경우 이 수법은 효과적이다. 많은 사람들은 단지 그들의 복지에 관심이 있다. 이 경우에 응답자들에게 조사 결과가 그들의 인생과 사회를 나아지게 하는데 어떻게 사용될 것인지를 알려주는 것이 유용할 수 있다.

세 번째 심리적인 요인은 사회적 검증(*social validation*)인데 만약 표본구성원들이 다른 사람들도 응할 것 이라고 믿으면 그들도 응할 가능성이 높다는 것을 말한다. 예를 들어 집집마다 다니며 제품을 주부에게 팔려고 하는 판매원이 많은 사람들이 제품에 대해 만족스러워 한다는 것을 당신에게 납득시키기 위해 이웃에 있는 많은 사람들이 그 제품을 매입 했다고 말할 수 있다. 따라서 주부는 그 제품의 특성에 대하여 모르더라도 무엇이 그 제품을 그렇게 인기 있게 하는지 알고 싶어 할 수도 있다. 다른 예로, 어린 소년이 자전거 헬멧 쓰기를 싫어하지만 이웃에 있는 나이가 더 많은 소년이 쓰고 있다고 알려주면 쓰기로 결정하게 된다. 비슷한 유형의 심리학이 응답자가 조사에 참여할 것을 설득하는데 효과적으로 쓰일 수 있다. 면접원이 이웃에 있는 많은 응답자들이 조사의 중요성을 깨닫고 참여를 원해서 이미 조사에 참여했다고 말할 수 있다.

네 번째 심리적인 요인은 권위(*authority*)이다. 만약 조사 참여를 권위 있는 기관에서 한다면 승낙할 가능성이 더 높다. 조사 연구논문에 나오는 보고에 의하면 일반적으로 정부가 후원하는 조사가 개인사업체에서 실시한 조사보다 거부율이 더 낮다. 특히 시장 조사는 실시된 모든 조사 중에 협조율이 가장 낮으며 이는 권위가 그 이유라고 볼 수 있다. 정부는 가장 권위 있는 기관에 해당한다.

예를 들어, 미국에서는 센서스국, 미국 국립농업통계국(NASS)과 다른 연방기관은 국가 이익을 위해 조사 자료의 수집 권한을 갖고 있다. 따라서 농장 운영자에게 조사 협조를 부탁하기 위해 접근할 때에 NASS는 정당한 권위를 갖고 있는 기관에 해당한다. 이런 이유로 NASS 조사의 경우 미국 센서스국 또는 다른 정부 기관들에

서 실시하는 조사들처럼 협조율이 상당히 높다.

개인 조사기관이 정부기관을 위하여 조사를 실시할 때에는 정당한 권위의 표시로 그 정부기관의 전용 편지지나 그 기관에 있는 저명한 사람이 서명한 편지를 사용할 수도 있다. 정부 조사를 실시하지 않는 시장조사 사업체들조차도 정당한 권위의 중요성을 이해하고 표본구성원들이 그들의 조사 요청에 따르게 하도록 한다. 어떤 시장조사단체들은 편지지가 정부기관의 것처럼 보이도록 만든다. 하지만 일부 국가에서는 정부기관을 특별히 권위 있는 것으로 보기보다는 오히려 의심을 가지고 본다는 점에 특히 주의해야 한다.

다섯 번째 심리적인 요인은 희소성(*scarcity*)이다. 기회가 드문 경우 다른 것들보다 더 소중하게 여긴다. 표본구성원은 조사 요청을 받는 것이 드문 일이고 이번 기회를 이용할 필요가 있다고 생각되면, 조사 요청에 기꺼이 따를 수도 있다. 면접원들은 표본구성원들에게 그들이 자기 의견을 말할 수 있는 기간이 얼마 남지 않았거나, 오늘이 조사의 마지막 날이라거나, 2,000명 중 한 명에게만 연락했다고 하거나, 표본 단위의 응답이 지금 필요하고 만일 그러지 않으면 그들이 참여할 기회를 잃을 것이라고 하면서 이런 요인을 이용 할 수 있다.

유대감(*liking*)이 여섯 번째 심리적인 요인이다. 표본구성원들은 만일 면접원이 그들에게 매력적이거나 면접원의 가치, 생각, 언어, 사회적 배경이나 개인 스타일 등이 그들과 비슷해 보이거나 들리면 더 자발적으로 협조한다. 이것이 많은 면접원들이 저소득 지역에서는 캐주얼 차림을 하고 고소득 지역에서 면접할 때는 정장을 입는 이유 중에 하나이다. 어떤 면접원들은 표본 구성원에게 칭찬을 할 수도 있고 또는 그들의 집, 어린이, 애완동물, 화원 등에 대하여 칭찬을 할 수도 있다. 그들은 응답자의 말하는 방법까지도 흉내 낼 수 있다.

<표 3.5>는 이 여섯 가지의 심리적인 요인과 그것들의 관련성을 요약한 것이다. 언급한 방법들은 조사가 어디에서 실시되었나에 따라 더 또는 덜 영향력이 있을 수 있다. 우리가 방금 이야기한 대부분의 심리적인 요인은 표본구성원들에게 조사에 참여해 달라고 부탁하기 전에 우리가 그들의 특성을 미리 파악하는 것을 필요로 한다. 일관성은 표본구성원의 믿음, 태도나 가치에 대해서 아는 것을 필요로 하고, 유대감의 원칙은 응답자들이 면접원으로부터 무엇을 원하는지에 대해 아는 것을 요구한다. 보상을 하기 위해서는 응답자가 조사를 어떻게 보고 우리가 조사에서 응답자에게 어떤 도움을 줄 수 있는지 알 필요가 있다. 만일 우리가 전혀 모르는 사람들

대상으로 면접하면 이런 원칙들을 과연 적용할 수 있을까?

<표 3.5> 여섯 가지 심리적 요인과 적용의 예

심리적 요인	의미	조사함축의 예
보상	승낙을 응답자에게 주는 선물, 보수, 특혜, 이익에 대한 보답으로 본다.	조사를 요청할 때 표본 단위에 인센티브를 준다. 우편에 자료 브로셔가 포함된다.
약속과 일관성	승낙은 응답자의 믿음, 태도, 가치 등과 일치된다.	응답자: 세금은 조사에 쓰여서 안 된다. 면접원: 결정은 자료를 토대로 하는 것이 중요하고, 조사는 자료를 산출한다.
사회성 검증	만일 다른 비슷한 사람들도 응할 것이라고 믿으면 더 잘 응한다.	면접원: 대부분의 사람들은 이 조사에 참여하고 그것을 즐기는 것 같다.
권위	잘 알려진 정부기관 또는 권위 있는 기관의 요청은 승낙할 수 있다	면접원은 가능하면 기관의 명성을 강조한다.
희소성	드문 기회라고 인식되는 요청에는 잘 응한다.	면접원은 기회를 강조한다. • 다른 몇 천명을 대표한다. • 자료 수집기간이 곧 끝나니까 지금 참여해라
유대감	좋아하고 매력적인 면접원의 요청에 더 잘 응한다.	면접원은 이것을 이용하기 위해 옷, 태도, 말하는 방법과 스타일을 선택할 수 있다.

여기에서 지속적 상호작용(*prolonged interaction*)이란 개념이 매우 유용하게 사용될 수 있다. 지속적 상호작용은 면접원이 특정 응답자에게 알맞게 맞추는 (*tailoring*) 접근법을 적용할 수 있도록 단서를 확인하는 목적으로 표본 구성원과 대화를 지속하는 것을 의미한다. 응답자가 참여하고 싶지 않은 이유를 대는 것을 듣고 효과적인 설득(<표 3.5>에 있는 심리적 원리에 기초한)으로 그 이유들을 반박하는 것이다. 때로는 응답자에게 알맞게 맞추는 작업을 연기하는 것이 현명하다. 예를 들어 전화면접에서 표본구성원이 현재 매우 마음이 내키지 않는 것 같을 때는 구성

원이 나중에 참여하도록 설득할 수 있는 어떤 단서를 얻어 낸다. 왜냐하면 이를 통해 다음 시도에서 협조를 얻을 가능성을 흔히 높일 수 있기 때문이다.

분명히 이들 원칙 중 몇 개는 Cialdini, Groves, Couper 등이 명확하게 논의하기 전에도 조사 설계자와 면접원들이 오랫동안 이해하고 있던 것들이다. 이해의 많은 부분은 상식을 근거로 했고, 이러한 원칙이 협조를 얻기 위한 방법의 개발에 직관적으로 적용되었다. 예를 들어 보상원칙의 경우 인센티브, 자료 브로셔와 표본들이 소중히 할지 모르는 다른 것들을 사용해 왔다. 또한 보상은 면접원에 의한 보답일 수도 있다. 어떤 경우에는 면접원이 응답자에게 한 시간짜리 면접을 부탁 할 수 있고 응답자는 거부한다. 만일 설계가 허락하면 면접원이 양보하여 단지 몇 개의 주요 조사 변수의 측정을 위한 15분짜리 면접을 요청할 수 있다. 더 큰 요청 후에 더 작은 요청을 하면, 그 작은 요청이 더 쉽게 승낙 된다는 것을 보여주는 연구결과가 있다.

일관성원칙은 조사 소개와 관련된 면접원 지시문을 개발할 때 사용할 수 있다. 조사와 응답의 가치를 설명하기 위해 면접원 지시문이나 사전편지에 “사실을 근거로 결정할 수 있게 하기 위하여 이 조사를 실시한다.” 같은 문장을 넣는 것을 의미할 수 있다. 권위와 희소성원칙은 조사 소개문을 작성하는 데 일반적으로 사용된다.

앞에서 언급한 원칙들은 일반적인 관행 보다 일관되게 사용될 수 있다. 조사 참여를 수락 받는다는 측면에서 모든 면접원들이 같지 않은 이유는 그들의 사회성 능력, 단서를 찾는 능력, 그 단서에 빠르게 그리고 성공적으로 반응하는 능력 등에 있어 큰 차이가 있기 때문이다. 비록 다른 지역에서 일하는 경우라도 협조를 얻는데 항상 대단히 성공적인 면접원의 예가 있다. 그들의 성공은 표본구성원이 제공하는 단서를 파악하고, 해석하여 대처하고, 그것들에 효과적으로 대응하는데 있어 노련하다는 것에서 찾을 수 있다.

많은 경험 있는 베테랑 면접원들은 그들이 매일하는 일에 이런 원칙들을 적용한다. 하지만, 그중에 극소수만 그들이 사용하는 기술을 조사 참여 이론의 일부라고 생각해왔다. 그러므로 경험 유무에 상관없이 모든 면접원은 우리가 방금 다룬 심리적인 요인들, 특히 지속적인 접촉과 목적에 알맞게 맞추는 전략을 통해 도움을 받을 수 있다. 협조율이 높은 면접원들이 실제로 그들의 직무를 어떻게 실행에 옮기느냐에 대하여 더 많이 알 수 있게 그들을 모니터하고 더 자주 관찰하여야 한다. 지속적인 접촉과 개별 응답자에게 알맞게 맞추는 방안에 대한 교육과 함께 좋은 면접

관행에 대한 사항들을 면접교육에 포함시킬 수 있다.

우편 조사에서 사전편지 또는 동봉한 편지의 문구를 개선하는데 쓰이는 여러 원칙들이 있다. 우편조사에서 응답률을 최대화하기 위해 여러 가지 단계들의 결합 방안을 개발하기 위해 사회교환이론을 근거로 한 좀더 개선된 과정이 사용되어 왔다. 아래에서 이런 결합에 대해 논의한다.

3.6.3. 방법의 결합 사용

무응답의 원인은 많고 다양하기 때문에 무응답을 줄이는 완전한 전략은 각각의 원인을 목표로 하는 특별한 기술들을 포함하는 광범위한 것이어야 한다. 응답자와 관련된 모든 조사를 위한 전반적인 규칙은 다음과 같다. (1) 설문지는 응답자가 작성하던지 면접원이 하던지 관계없이 완성하기 쉬워야 한다. (2) 조사가 사전편지나 면접원에 의해 흥미롭고 전문적인 방법으로 소개되어야 한다. (3) 재요청, 거부 전환 시도와 추적활동을 포함한 추가절차를 위한 계획과 이에 따른 여유 예산이 따로 있어야 한다. (4) 응답거부를 줄이고 접촉하기 힘든 표본구성원을 접촉하기 위한 기술이 면접원들에게 철저히 교육되어야 한다.

더 특수한 방법의 적용은 자료수집방식에 따라 좌우된다. Dillman(1978, 2000)은 우편조사에서 높은 응답률을 얻기 위한 표준화된 단계적인 방법의 결합인 사회교환이론에 의한 전략을 개발했다. 이 전략은 많은 경우에 매우 성공적이었고, 이 전략의 최신판은 9단계로 이루어져 있다.

이전의 7단계 전략에서는 만약 그 단계 중에 어느 것이라도 빠뜨리면 무응답률이 증가한다는 연구결과가 있다(DeLeeuw and Hox, 1988 참조). 이런 전략의 최종결과는 해당 국가, 주제, 모집단 등에 따라 차이가 있으며, 각 단계와 그 다음 단계 사이의 시간은 적절히 조정되어야 한다. 예를 들어 사업체조사에서 한 가지 중요한 단계는 그 회사나 조직에서 “적절한” 정보 제공자를 찾는 것이다. 대부분의 국가에서는 정부조사에서 우표를 사용하는 것이 어렵기 때문에 전략 중 우표의 사용을 생략하는 수정이 필요할 것이다.

우편조사 응답률 향상을 위한 Dillman의 9단계 전략

- 1단계: 표본단위들에게 미리 편지를 보낸다.
- 2단계: 가능하면 우송료로 우표를 사용해서 설문지를 보낸다.
- 3단계: 어느 정도 시간이 지난 후에 조사에 따라서 감사/재요청 겸용 카드를 보낸다.
- 4단계: 3단계에도 응답 하지 않은 사람들에게 두 번째 설문지 사본을 보낸다.
- 5단계: 4단계에도 응답 하지 않은 사람들에게 마지막인 세 번째 사본을 특별 우편 서비스로 보내거나 전화 추적조사를 실시한다.
- 6단계: 표본단위들에게 우편물을 보낼 때에는 가능하면 우표가 부착된 반송봉투도 함께 송부한다.
- 7단계: 표본단위와 하는 모든 서신 왕래에서는 응답자 개인의 이름을 사용한다. (즉, 사전편지 윗부분에 표본단위의 이름을 프린트 한다.)
- 8단계: 모든 응답자가 완성하기 쉽게 설문지가 설계되어야 한다.
- 9단계: 작은 금전적인 인센티브를 사용한다.

출처: Dillman(1978, 1991, 2000)

면접조사에서는 면접원들이 교육이 잘 되어있고 의욕적인 것이 중요하다. 그들은 협조를 얻는데 유용한 방법들에 대하여 잘 알고 있어야 한다. 이는 면접원들이 조사 참여 이론을 잘 알아야하고, 표본단위를 찾아내는 데 효율적인 자료를 활용해야 하고, 효율적인 추적조사 방안을 적용해야 한다는 것 등을 의미한다. 모든 수단을 다 사용한 면접원보다 그렇지 못한 면접원의 경우 응답률이 더 낮아진다는 것은 스웨덴 통계청에서 입증했다.

응답자들이 가구원의 활동, 구매 같은 것들에 대한 기록을 매일 해야 하는 일지 조사에서는 일지 자체는 응답자 자신이 관리하지만 면접원이나 다른 직원에 의해 이루어지는 추적관리가 매우 효과적이다. 예를 들어 Dillman(2000)은 2주간의 가구 생활비조사에서 일지 기록자와 5에서 7번의 연락을 취하는 것이 적절하다고 하였다. 또한 조사 참여 의지가 저하되는 것을 막기 위해 일지 기록자에게 조사 기간이 곧 끝난다는 메모를 보내는 것이 상당히 효과가 있을 수 있다.

때로는 자료수집에 관련된 일반적인 노력과 전략으로는 충분하지 않을 수도 있다. 어떤 조사들은 특별한 방법을 필요로 할 수 있다. 왜냐하면 그 조사들에서 무응답의 원인이 특별하기 때문이다. 무슨 방법과 방법을 결합하는 것이 효과적인지를 결정하는 유일한 방법은 무응답의 원인에 대한 자료를 수집해 보는 것이다. 이런 자료는 정보 내용의 수정, 대규모 회사의 조사를 위한 특별 지원 또는 한 가지 이상의 자료수집방법의 도입과 같은 설계 조정을 위한 정보를 제공한다.

3.6.4 프라이버시와 기밀성

프라이버시와 기밀성은 조사자료의 보호와 관련된 개념이다. 프라이버시(*privacy*)는 개인과 사업체가 그들에 대한 어떤 정보를 제공할 수 있는지를 결정하는 권리이다. 기밀성(*confidentiality*)은 이미 수집된 자료가 권한 밖의 사용으로부터 보호되는 정도를 나타낸다. 대부분의 나라들은 이런 문제를 규제하는 자료관련 법령과 법률상의 체계가 있다. 그럼에도 불구하고 프라이버시와 기밀성은 측정오차와 무응답 문제와 관련되어 있기 때문에 프라이버시와 기밀성에 대해 통계기관 및 다른 자료수집 기관들이 관심을 갖는다.

프라이버시는 질문의 민감성뿐만 아니라 개인 또는 자료 제공자의 프라이버시 때문에 항목무응답과 관련이 된다. 무엇이 민감한지는 시대에 따라, 문화와 모집단에 따라, 그리고 개인에 따라 다를 수 있다. 하지만 연구대상인 모집단의 상당 부분이 어떤 질문을 민감하다고 생각한다면 항목과 단위 무응답을 피하기 위하여 그 질문을 조심스럽게 취급해야 한다. 민감한 질문 때문에 생기는 무응답 문제를 피하는 가장 좋은 방법은 질문과정 초기에 그것에 대해 묻는 것을 피하는 것이다. 그렇지만 정부정책에 필요한 대부분의 정보들은 민감한 질문을 필요로 한다. 그러므로 이런 질문은 거의 선택의 여지가 없지만, 가능하면 그런 질문을 배제하는 것이 좋다. 특히 설문지 설계에 따른 영향을 감안하는 것이 더 현실적이다. 어떤 주제에 대해서는 자유서술형질문이 정해진 선택항목을 사용하는 질문보다 바람직할 수 있다.

예를 들어 “당신은 술을 얼마나 자주 마십니까?” 하는 질문이 “당신은 술을 매일, 하루걸러, ...?”라고 하는 것보다 더 막연하기 때문에 사람들에게 덜 민감하다. 소득 같은 경우는 일반적인 질문이 더 민감할 수도 있다. 소득에 대한 직접적인 질문보다는 선택할 수 있는 선택항목이 가능한 많은 것이 더 응답률이 높아진다고 알려져

있다. 경우에 따라 “여러해 전에는 ‘_____’의 문제를 공공연히 이야기하지는 않았지만 오늘날에는 이런 주제를 내놓고 이야기한다.” 또는 “...는 이례적이지 않다” 같은 표현을 질문 도입부에 사용하여 민감성을 제거하는 것이 가능하다. “당신은 하루에 담배를 몇 대 피웁니까?”라고 단순히 묻는 것이 “당신은 흡연자입니까?”라고 물은 다음에 “당신은 하루에 담배를 몇 대 피웁니까?”라고 묻는 것보다 흡연은 비난 받을 행위라는 뜻을 덜 함축한다(Delanius, 1988 참조). 이런 질문과 관련된 전략은 항목응답률에 긍정적인 영향이 있을지 모르지만 반면에 측정오차에는 부정적일 수 있다. 질문 표현법에 대한 조정이 끝나면, 초반부의 민감한 질문들 때문에 발생하는 단위무응답을 피하기 위해서 민감한 질문들은 설문지 거의 끝에 두는 것이 바람직하다.

만일 민감한 질문들이 매우 중요하거나 혹시 조사 전체가 민감하다고 여겨지면 응답자에게 가장 많은 수준의 프라이버시가 보장되는 방법이 선택되어야 한다. 4장에서 보겠지만 우편조사는 대면조사나 전화조사 방법보다 덜 위협적이고 사회적 기대부응(*social desirability*) 편향의 경향이 적다. 사회적 기대부응 편향은 표본단위들이 그들이 바람직하지 않은 사회적 특성을 가지고 있다는 것을 드러내기를 싫어하는 데에서 생기는 조사오차다. 그들은 사회적으로 바람직하다고 인정하는 방향으로 응답하던지 아니면 전혀 응답하지 않는다. 통상적인 면접조사에서 민감한 부분에 대해서는 자기기입식 방법을 결합하는 등의 다양한 방법이 있다(4, 5, 6장 참조).

민감한 정보의 수집을 위해서 몇 가지 특별한 방법이 개발되었다. 지금은 컴퓨터 보조(*computer-assisted*)방법으로 대체되었지만, 예전에 쓰였던 방법은 민감한 질문들을 다루는 대면면접 부분을 자기기입식 설문지로 대체하는 것이다. 담당 면접원이 알지 못하게 하기 위해 응답자가 그 설문지를 면접원이 제공한 밀봉된 봉투에 넣으라고 한다. 또 다른 방법은 확률화응답(*randomized response*)기법을 사용하는 것이다(4장 참조).

기밀성의 보증이 약하다고 응답률에 큰 영향을 미치는 것은 아니며, 항상 두려움 그 자체가 문제다. 기밀성 보증 유무는 모집단이나 사업체의 단지 일부에게만 영향을 미친다. 기밀성의 보장과 무응답률의 증가 사이에 상당한 관계가 없더라도 통계자료나 다른 자료수집 단체들은 자료가 부당하게 공개되지 않게 보호할 윤리적, 법률적 의무가 있다. 그러므로 이런 공개의 가능성을 제한하는 방법의 개발을 위하여 많은 연구가 실시되었다. 개인 단위에 대한 자료를 보호하기 위해 설계된 방법 중

에 잘 알려진 것들이 아래 설명되어 있다. 이 방법들은 특정 표본단위에 대한 개인 자료가 공개되는 위험을 줄이는데 목적이 있다. 개인자료의 공개는 응답자의 감소와 원상회복하는데 긴 세월이 걸릴 수 있는 공공 및 사용자로부터의 신뢰도 하락 등 조사기관에 상당히 부정적인 영향을 줄 수 있다. 어떤 조사기관도 이런 문제를 가볍게 취급할 수 없으므로, 대부분의 단체들은 자료에 대한 사회의 필요성과 적당히 균형을 유지하는 프라이버시 침해를 규제하는 프로그램을 갖고 있다. 기밀성과 관련된 방법에 대해서는 Fienberg와 Willenborg(1998)을 참조하라.

조사자료의 기밀성을 보증하기 위해 사용되는 방법들

신분을 확인하지 않는 것(*de-identification*) 또는 익명화(*anonymization*). 기록에 있는 사람을 또는 사업체를 직접적으로 확인할 수 있는 것(예를 들어, 사회보장번호나 다른 신분증명번호)은 제거한다.

은폐(*suppression*). 자료 파일에서 몇몇의 또는 모든 표본구성원의 특정 변수 값을 삭제한다. 그 이유는 경우에 따라 자료파일에 있는 단위 수가 너무 작아서 개인 단위와 특정 변수값에 대한 연상이 가능하기 때문이다.

구간 값(*interval values*). 본래값 대신에 본래 값을 포함하는 구간을 사용한다. 이 접근 방법의 변형이 상단부 코딩(*top coding*)인데 이는 모든 큰 값을 특정값으로 대체하는 것이다. 예를 들어, \$100,000이상의 소득은 모두 >\$100,000로 코딩한다.

잡음추가(*adding noise*). 본래값 X 를 $X+a$ 로 대체한다. 여기서 a 는 작은 난수이다.

암호화(*encryption*). 원래 코드로 변환하기 힘든 암호화 알고리즘을 써서 자료를 뒤섞는다. 이것은 자료를 조사 현지에서 정보처리를 하는 곳으로 우편 또는 온라인으로 보낼 때 일반적으로 사용한다.

3.6.5 자료수집방법의 선택과 응답률

자료수집방법과 관련된 오차는 6장에서 다룬다. 그렇지만 무응답률을 줄이는 데는 자료수집방법의 선택이 중요하다. 여기서는 간단히 자료수집방법과 무응답의 관계를 설명한다. 서로 다른 자료수집방법은 서로 다른 무응답수준을 나타낸다는 것을 보여주는 많은 논문들이 있다. 일반적으로 인정되는 기대 응답률순으로 배열하면 면접, 전화, 우편, 그리고 기타 웹조사와 일지조사 같은 자기기입식방법인데, 대면면접 방법이 가장 높은 응답률을 보여준다. 하지만 방법선택을 오직 기대 응답률만으로 할 수는 없다. 비용, 포함성, 적시성, 측정 같은 여러 요인을 생각해야 한다(더 자세한 논의는 6장 참조). 자료수집방법과 무응답에 대한 지식을 이용하여 모든 설계 조건들을 참고로 주된 자료수집방법을 선택할 수 있다. 만일 필요하다면 응답률을 늘리기 위해 제2 또는 제3의 방법까지도 그것과 결합할 수 있다. 서로 다른 방법은 왜 다른 수준의 무응답률로 나타나는지 확실하지는 않다. 응답자들이 만일 선택권이 있다면 그들이 선호하는 자료수집방법이 나타날 것이다. 관련 연구에 의하면 대부분의 응답자들은 면접을 하는 것보다는 우편 설문지에 기입하는 것을 더 선호한다. 하지만 다른 차원에서는 만일 응답자에 따른 적절한 접근방법에 대한 충분한 지식을 갖춘 유능한 면접원이 조사 참여요청을 하는 경우 응답자가 이를 거부하기가 일반적으로 쉽지 않다.

우선 기본적인 전략은 설계와 관련된 모든 측면에서 가장 적절한 주된 조사방법을 선택하는 것에서 출발한다. 무응답과 관련해서 주된 방법을 미리 정한 전화시도 횟수 같은 특정한 제약조건하에서 최대한 사용한다. 다음으로 응답률을 높이기 위하여 다른 방법을 추가적으로 사용한다. 이런 전략은 방법에 대한 응답자의 다양한 선호도, 방법에 따라 응답자의 위치를 찾는데 다양한 가능성을 갖고 있다는 측면, 또한 조사가 중요하다라는 사실을 응답자에게 알릴 수 있다는 점에서 매우 효과적이다. 혼합방식(*mixed mode*)이라고 부르는 이런 전략은 절충안에 해당한다.

만일 조사에 가장 적당하다고 여겨지는 주된 자료수집방법이 있으면 다른 방법을 사용하는 것은 부가적인 오차를 생성할 수 있다. 예를 들어 우편조사를 위해서 개발된 설문지에는 질문에 따라 다수의 선택항목이 있어도 된다. 만일 그 설문지가 전화 추적조사에 사용되면 선택항목의 수가 많으면 전화 면접원이 읽은 것을 응답자가 못 듣거나 구별하지 못할 위험이 있다. 그래서 어떤 경우에는 방법을 바꾸는 것이 품질을 저하시킬 수 있다. 반면에 면접원이 참여하는 방법을 사용해서 개선할 수 있는 다른 품질적인 측면도 있다. 품질에 있어서 우리는 이득과 손실이 동

시에 발생하는 상황을 직면한다. 여러 가지 제한이 주어졌을 때 우리의 관심은 전체 품질을 가능한 향상시키는 것이다. 만약 응답률을 높이는데 두 가지 이상의 방법을 사용하는 것이 전체 품질에 향상을 가져오고, 만일 비용과 적시성 같은 다른 사항도 충족해주면 그런 방법을 사용해야 한다. 그렇지만 대부분의 경우 이에 따른 이등과 손실에 대한 정확한 정보가 없다. 물론 매우 신중히 계획해서 실행된 혼합 방식 조사들의 사례가 있긴 하지만, 대부분의 혼합방식 설계는 최소한의 관리를 필요로 하는 임시적인 방법이다.

혼합방법조사에서의 일반적인 관행은 우편조사와 같은 비용이 덜 드는 방법으로 시작해서 응답률을 높일 수 있는 전화면접조사와 같은 비용이 더 드는 것으로 방법을 추가하는 것이다. 비록 저비용에 의한 방법이 측정오차가 작아지는 등 중요한 설계요소와 일치할 수도 있지만, 이 상황에서는 비용이 가장 중요한 설계요소이다.

예를 들어 어떤 민감한 질문에 대한 응답은 면접을 하는 것 보다는 우편조사를 사용하는 것이 더 정확할 수 있다는 것을 우리는 잘 알고 있다. 게다가 우편조사가 면접보다 덜 비싸므로 그것을 사용하면 비용 측면에서도 도움이 된다. 더 우선의 방법이 비싼 것일지라도, 덜 비싼 것으로 조사를 계속하는 것은 정당화될 수 있다. Carrol 등(1986)이 실시한 주택 에너지 소비연구 사례를 살펴보면, 세 차례의 대면 면접을 통해 각각 76.6%, 80.4%, 89.9%의 축적 응답률을 얻었고, 추가적인 두 차례의 우편수집 조사를 통해 응답률을 4.7%포인트 더 증가시킬 수 있었으며, 이들 중에 75%가 대면면접 단계에서는 거부자였다.

3.6.6 응답자 부담

무응답에 관계되는 한 가지 중요한 사항은 응답자가 설문지를 완성하고 조사에 참여하면서 느끼는 부담감이다. 만일 조사 요구사항이 흥미롭고 쉽다고 인식되면 협조를 얻을 가능성이 높아진다는 것은 널리 알려졌다. 그렇지만 응답자 부담은 매우 일반적인 개념이다. 개인과 가구조사에서는 응답자가 조사를 힘들고 시간이 걸리는 것으로 인식하는 정도이다. 하지만 사업체조사에서 부담은 흔히 조사참여와 관련된 비용을 흔히 반영한다. 부담을 줄이기 위해 조치를 취하면 일반적으로 응답률은 높아진다.

대표적인 부담에는 설문지 또는 면접의 길이, 시간과 노력면에서 보았을 때 응답

자의 작업량, 질문에 직면했을 때 응답자가 느낄지도 모르는 정신적 압박, 어느 특정한 기간 동안 응답자가 받는 조사요청의 건수 등이 있다. 응답자 부담 문제에 대처하기 위해 두 가지 일반적인 방법이 사용된다. 하나는 응답자가 그것이 덜 부담스럽다고 인식하도록 조사 요청을 수정하는 것이고, 다른 것은 인센티브를 사용해서 응답자에게 보상하는 것이다.

응답자 부담의 축소는 설문지를 완성하는데 필요한 시간이 최소화되어야 한다는 것을 말한다. 미리 입증된 분석계획과 상응하는 질문이 아니면 설문지에 포함하지 말아야 한다(2장에서 이와 관련된 계획안이 제공됐음). 설문지가 개발되면 응답자 부담을 줄이기 위해 각 질문에 대해 다음과 같은 연속적인 사항들에 대한 검토가 필요하다.

- 질문이 필요합니까?
- 응답자는 누구를 위하여 답을 하고 있습니까? (자기 자신, 가구, 사업체)
- 응답자는 적당한 노력으로 질문에 답을 할 수 있습니까?
- 대부분의 응답자가 답을 제공할 것 같습니까?
- 답을 할 수 있는 질문이가요? 예를 들어, 만일 설문지의 설계자에게 같은 질문을 한다면 답을 하는데 어려움이 없겠습니까?

만일 이 항목 중에서 한개 이상에 문제가 있으면, 그 질문을 포함할지를 다시 고려해야 한다. 질문에 대한 일반적인 부담과 관련된 문제는 어려운 용어, 표현, 개념, 응답자한테 일어난 일들과 그 일이 일어난 때를 기억해내라고 하는 상황, 응답자가 답을 얻기 위해 다른 자료를 찾아보거나 계산하지 않을 수 없는 상황 같은 것들이다. 말을 하면서 생각하는 프로토콜과 포커스 그룹(8장 참조)과 같은 인지연구(*cognition research*)를 통해 이런 많은 문제점들을 밝혀낼 수 있고 질문들을 바꾸거나 제거할 수 있다.

문제를 제기할 수 있는 것은 질문의 개수나 특성만이 아니다. 특히 자기기입식 방법에서는 설문지의 구성도 중요하다. 자기기입식 조사는 이해하기 쉬워야 하는 것이 중요하고, 한 질문에서 다른 질문으로의 이동이 명백해야 한다. 이런 설문지를 묘사하기 위해 Dillman 등은 “응답자 위주”(respondent friendly)라는 표현을 사용했다(Jenkins와 Dillman, 1977; Redline과 Dillman, 2002). 심리적인 그리고 사회학적

인 이론들을 근거로 한 새 이론이 이 분야에서 생기고 있다. 여기서는 자기기입식 설문지를 위한 설계 원칙을 대상으로 말로 나타내거나 또는 말을 쓰지 않는 언어와 관련된 이슈들을 다루고 있다(물론 이 중에 많은 원칙들은 면접 조사에도 적용될 수 있지만, 이 경우 면접원이 지각과정에서 주요한 역할을 하기 때문에 그것이 무응답에 미치는 영향이 면접 조사에서는 아마도 그렇게 크지 않을 것이다.) 응답자의 일을 더 쉽게 하고 자기기입식 조사의 오차를 줄이는 것을 겨냥한 설문지 설계 원칙의 예가 아래에 있다. 보다 완전한 원칙은 Dillman(2002)을 참조하라.

응답자 위주의 우편 설문지를 위한 설계 원칙

- 밝기, 색, 모양, 위치선정 등의 시각적 요소를 사용해서 설문지를 완성하기 위한 응답자 이동로를 설정하라.
- 설문지 중간에 형식상의 변동이 있는 경우 이를 응답자가 인지할 수 있도록 두드러진 시각적 표시를 사용하라.
- 지시문은 사용될 곳에 잘 볼 수 있도록 배치하라.
- 정보를 이해하기 위해서 응답자가 떨어진 위치에 있는 정보를 참고할 필요가 없도록 하는 정보 제시방식을 적용하라.
- 질문을 한 번에 하나만 해라(즉 이중목적(*double-barreled*)의 질문을 피해라).

출처: Dillman(2002)

응답자 부담을 줄이기 위해 할 수 있는 것을 다한 다음에 응답률을 더 높이려면 인센티브를 사용해야 한다. 인센티브의 사용이 지난 몇 십년동안 많은 연구의 중심이었다. 중요한 주제이고 비용, 측정오차, 조사윤리 같은 많은 다른 설계 양상에도 관련이 있다. 인센티브는 항상 응답률이 높아지게 한다고 말하는 것은 당연 할 수 있다. 많은 마케팅연구소들은 표본단위가 자진해서 참여하는 비율을 높이기 위해 복권, 출판물, 달력 같은 작은 인센티브를 정기적으로 제공한다. 전형적인 시나리오는 일단 응답자가 참여하면 그에게 선물을 줄 것을 약속한다. 때로는 인센티브가 마음이 내키지 않는 응답자의 마음을 움직이게 하는 수단으로만 사용되는데 그것은

윤리상의 이슈가 된다. 물론 정부조사는 거부전환(거부상태에서 응답으로 전환)을 위한 인센티브를 써서는 안 된다. 모든 표본구성원이 인센티브 제공으로 이익을 받을 가능성이 같아야 한다. 그렇지만 이에 대해서는 논란의 여지가 있다.

인센티브에 대한 연구논문에서 강하게 제안했듯이 응답률 측면에서 보면 미리 제공하는 사전 인센티브가 조사 후에 제공을 약속하는 사후 인센티브보다 더 효과적이다. Berk 등(1987)은 “인센티브가 없는“ 경우에 오히려 사후 인센티브보다 더 높은 응답률을 산출했다는 연구결과를 보여주고 있다. 3.6.2절에서 왜 인센티브가 효과가 있나, 그리고 특히 왜 미리 제공하거나 미리 배달된 인센티브가 응답을 늘릴 수 있는지 설명하는데 도움이 되는 교호작용원칙과 다른 이론들에 대해 논의 했다. 사회교환이론을 근거로 하면 응답자는 참여 요청을 받기 전에 받은 선물에 보답하는 경향이 있다. 많은 경우에 참여하는데 보내는 시간의 가치측면에서 보면 조사 참여가 제공된 선물보다 가치가 훨씬 높다. 만약 인센티브가 사후에 약속된 것이면 경제교환이론이라는 다른 이론이 작동하기 시작한다. 여기서는 표본 단위가 조사요청을 “요청 대 보상”이라는 상황 판단에 따라 받아들이거나 거부할 수 있는 사업상의 제의로 보게 된다.

인센티브가 때로는 응답률을 많게는 10%포인트 늘일 수 있다는 증거가 있지만 그것의 실용성은 무응답을 줄이는 다른 수단과 비교 검토 되어야 한다. 또한 예상할 수 있겠지만 인센티브는 모집단과 모집단 그룹에 따라서 다르게 작용하고 측정 그 자체에 미치는 영향에 대해서는 별로 알려지지 않았다. 사업체 조사에서의 인센티브는 개인과 가구조사에서 사용되는 것과 다르다. 사업체 조사에서는 조사에서 얻어진 다른 사업체와의 비교 결과의 제공이 조사 참여를 유도할 수 있다. 이런 경우 규모가 비슷하거나 또는 같은 지역에 위치한 다른 사업체와의 비교 결과에 대해 특히 관심을 보일 것이다.

3.6.7 사전통보 편지의 사용

많은 조사 연구자들은 사전 통지 없이 표본 단위와 접촉을 했을 때, 조사 참여를 유도하는 것이 어렵다는 것을 잘 알고 있다. RDD 조사에서는 모든 표본 단위와 미리 접촉을 하는 것이 불가능하다. 모든 RDD 조사가 그러지는 않지만 전화번호와 주소를 대응시켜서 상당한 비율의 주소를 알아내는 방법들이 있다. 다른 조사방법

에서는 주소를 얻기가 더 수월하다. 미리 발송한 편지는 보통 응답률에 긍정적인 결과를 가져온다. 이 결과를 확증할 수 있는 연구결과는 Sykes와 Collins(1988)에서 찾을 수 있다. 그렇지만 이 규칙에 예외가 있다. 사전 통보는 어떤 표본 단위들에게는 오히려 경고를 주어 실제 조사를 위해 접촉했을 때 단호한 거부를 준비하게 한다.(Lyberg와 Lyberg, 1991). 많은 나라에서는 법률상 공식 통계 조사를 위해 선택된 표본 단위에게 사전 통보를 해야 한다고 명시하고 있다. Lupperts(1994, 1995)가 사전통보 편지에 포함되어야 할 내용에 대한 점검목록을 정리했다. 그의 연구는 여러 나라에서 실시한 몇 개의 소비조사에서 사용된 사전편지에 대한 검토 결과를 바탕으로 했다. 최근 스웨덴에서 실시한 실험에서 해설자로 잘 알려진 대중 인물을 이용한 비디오테이프 형태의 사전통보가 효과적인 것으로 입증되었다(Ahtiainen, 1999).

효과적인 사전통보 편지를 작성하는 몇 가지 규칙

- 편지는 이해하기 쉬워야 하며 응답자에게 생소한 용어는 피해야 한다.
- 편지의 내용은 몇 개의 짧은 절로 구성되어야 한다.
- 편지의 어조는 호의적이어야 한다.
- 여러 가지 다양한 정보는 자연스러운 순서에 따라 나열해야 한다.
- 잠재적으로 논쟁의 여지가 있는 표현은 피해야 한다.

3.6.8 면접원과 응답률

5장에서는 면접원과 면접과정에서 생기는 오차를 다룬다. 이 절에서는 면접원과 무응답의 관련성에 대해 살펴본다.

면접 조사에서는 면접원이 설문지와 표본 단위를 연결한다. 면접원에게는 두 가지 주요 역할이 있는데, 협조를 얻는 것과 일단 협조를 구했으면 면접을 실행에 옮기는 것이다. 이 역할들은 두 가지의 매우 다른 기술을 필요로 하는데 면접원들은 보통 두 가지 기술을 모두 가지고 있지 않다.

면접원의 특성과 그의 작업 수행이 무응답률에 상당한 영향을 미친다. 특히 아주 중요한 것은 첫 접촉인데, 접촉의 첫 순간이 매우 결정적이다. “문앞(*doorstep*)”전략이 아주 결정적이고, 우리가 살펴본 심리적인 이론을 근거로 한 새로운 아이디어가 또한 중요하다. 면접원은 각 상황을 검토 평가하고 첫 접촉에서 얻었던 사전 정보를 근거로 개별 응답자에게 알맞게 맞추어 접근해야 한다. 중요하다고 입증된 몇몇 요인들은 다음과 같다.

추적하는 기술, 목표에 알맞게 맞추는 방법, 거부 전환 방법 등을 이용하는 면접원은 그렇게 하지 않는 면접원보다 더 좋은 결과를 얻는다. 최근의 경험에 의하면 면접원 교육은 개별 응답자에 알맞게 맞춘다는 측면에서 새로운 방법의 장점을 이용하여 훨씬 개선될 수 있다. 또한 표본단위를 접촉하는데 있어 최상의 면접원이 사용하는 방법을 검토하고 이를 교육 자료에 포함시키는 것에 대해서는 아직 별로 연구된 것이 없다.

높은 응답률과 관련된 면접원의 특성

경험(*experience*). 경험이 많은 면접원은 경험이 적은 면접원보다 응답률을 높이는 경향이 있다.

면접원의 인구통계학적 특성(*interviewer's demographical characteristics*). 면접원의 성별, 나이 등의 특성은 응답률에 영향을 주는 데, 응답자와 비슷한 특성을 갖게 하고 조사 주제에 맞게 하는 것이 바람직하다.

면접원의 목소리, 말씨, 면접 방식(*interviewer's voice, accent, and interviewing style*). 목소리와 말씨는 대면 조사에서 보다는 전화조사에서 더 중요하다. 개인 면접방식이 응답자에게서 더 많은 정보를 얻어낼 수 있지만 더 큰 면접원 오차를 낼 수도 있다.

면접원 기대감과 선호(*interviewer expectations and preferences*) 긍정적인 태도를 가진 면접원이 협조를 얻기 힘들 것이라고 생각되는 면접원 보다 협조를 얻을 가능성이 높다.

태도와 자신감(*aptitude and confidence*). 예를 들어 연속된 여러 번의 응답거절은 면접원이 다음 표본단위에 접근하는데 영향을 준다.

3.6.9 추적조사

우리는 접촉하기 힘들거나, 참여하기를 싫어하거나, 조사를 거부하는 구성원과 다시 접촉하는 다양한 효율적인 방법들을 알아보았다. 이러한 추적 재조사는 다양한 형태로 이루어진다. 우편 조사에서는 응답하지 않는 표본단위에게는 재요청 편지를 보낸다. 우리는 Dillman의 총 설계 방법과 같은 방안에 대해 이미 알아보았다. 이러한 재요청 편지는 그 전의 것과 같아 보이지 않도록 잘 정립된 패턴에 따라 작성된다. 재요청에 따른 결과 자료를 수집하여 이 자료를 토대로 추적조사를 어떤 시점에 종료할 것인지 결정해야 한다. 일반적으로 이런 자료에는 재요청 사이의 간격, 각 재요청 결과에 대한 자료, 표본 그룹에 따른 재조사 응답패턴의 차이 등이 포함된다.

전화 면접 조사와 전화 추적조사가 추가된 대면 면접 조사에서 면접원은 끈기가 있어야 한다. 통화스케줄 알고리즘(*call-scheduling algorithms*) 같은 효율적인 전화 통화를 실시에 대한 이론이 개발되었다. 만일 조사 설계자에 의한 접촉시간 배분에 대한 관리체계(CATI의 응용으로 가능함, 6장 참조)가 없으면, 스케줄은 면접원의 선호에 따라 결정된다. Swires-Hennesey와 Drake(1992)에 의하면 만약 면접원에게 전화 스케줄을 정하는 것을 맡기면 그들은 주중 편리한 때에 전화를 걸기 시작해서 더 필요하면 주말이나 저녁에 계속 전화를 건다고 한다. 통화스케줄알고리즘을 개발할 때에 기본적인 전제는 표본단위들에게 연락이 가능할 때 전화를 걸어야 한다는 것이다. Kulka와 Weeks(1998)는 조건부 확률 접근법을 사용하여 매긴 3-통화 알고리즘의 순위를 제시하고 있다. 가장 높은 순위에 있는 3-통화 배합은 다음과 같다.

1. 주중 저녁, 일요일, 일요일
2. 일요일, 주중 저녁, 주중 저녁
3. 일요일, 주중 저녁, 주중 아침

그리고 가장 안 좋은 배합은 다음과 같다.

1. 주중 오후, 주중 오후, 주중 오후
2. 주중 오후, 주중 아침, 주중 아침

아무런 제한이 없다면 가장 안 좋은 배합이 개인 면접원이 가장 먼저 선택할 지도 모르는 것 중 하나라는 것을 쉽게 볼 수 있다. 만일 정식 통화스케줄알고리즘을 개발할 충분한 시간이 없으면, 대다수의 통화시도는 저녁과 주말에 하는 것이 좋다는 일반적인 전략을 채택할 수도 있다.

비용 때문에 현지에서 하는 대면 추적조사는 흔하지 않다. 원칙적으로는 전화 접촉 전략을 현지에서도 적용할 수 있다. 실제로는 응답자와 시간약속을 할 때 전화 접촉 전략을 사용한다. 다른 경우에는 면접원이 무응답자들과 같은 동네에서 다른 면접을 실시할 때 무응답자를 추적조사 하도록 교육을 받는다.

3.6.10 항목무응답

항목무응답의 원인은 질문이나 설문지에 관한 문제로 볼 수 있다. 모든 항목에 답을 하지 않은 이유로는 다음과 같은 것들이 있다.

- 질문이 답하기 어렵거나 민감하기 때문에 응답자가 고의적으로 응답하지 않기로 결심한다.
- 설문지가 복잡하고 만일 방식이 자기기입식이면 응답자가 생략하라는 지시나 특정 질문을 놓칠 수 있고, 또는 응답과정이 지루하거나 거부감을 주거나 시간이 많이 걸리기 때문에 응답과정을 그냥 중단할 수 있다.
- 설문지의 서술형 질문은 항목무응답의 위험이 높다.
- 응답자나 면접원이 기술적인 오류를 범해서 특정한 질문에 대한 답이 지워진다.
- 설문지가 너무 길다.

어떤 상황에서는 항목무응답이 측정오차와 밀접한 관계가 있다. 응답자는 민감한 질문에 정확한 정보를 주지 않는 방법으로 답을 할 수 있다. 예를 들어 만일 질문에 “모름”이나 “의견 없음”같은 대안이 있으면 이런 대안은 그것이 없으면 질문에

답을 하지 않고 남겨 두었을 지도 모르는 응답자들의 탈출구 역할을 할 수 있다.

항목무응답은 질문과 설문지의 설계와 관련된 예방대책으로 잘 관리 할 수 있다. 좋은 설계 전략은 민감한 질문을 덜 사용하고, 최상의 생략 지시문을 사용하고, 응답자에게 너무 부담이 되는 질문을 피하는 것이다. 항목무응답이 있을 때에는 분석이 복잡해지므로, 수년에 걸쳐 결측자료를 적절한 방법으로 다른 자료로 대체하는 몇 가지 대체방법에 대한 연구가 수행되었다(3.6.11절 참조).

소득과 사회적으로 비난받을 행동이나 상황에 대한 질문에 대해 항목무응답이 많이 발생한다고 잘 알려져 있다. Christianson과 Tortora(1995)가 언급했듯이 대부분의 통계기관은 항목무응답을 완전히 또는 체계적으로 보고하지 않는다. 보고가 될 때 항목무응답률은 그 항목에 응답할 자격이 있는 표본단위 수와 응답할 자격이 있는 단위 중 응답한 표본단위 수의 비로 계산된다. 항목무응답률을 주요 질문별로, 표본 그룹별로, 면접원 별로 계산해 보는 것이 유용하다. 이렇게 하면 항목무응답률이 매우 중요한 과정변수의 역할을 할 수 있다(Madow 등, 1983; U.S. Federal Committee on Statistical Methodology, 2001 참조).

3.6.11 무응답의 보정

무응답을 최소화하려는 노력에도 불구하고 일반적으로 표본의 100% 사람이 조사에 응답을 하는 것은 아니다. 자료 수집이 완성되고 모든 자료가 입력된 후에도 무응답이 추정값에 미치는 영향을 줄이기 위해 사용되는 방법들이 있다. 이 방법들은 무응답을 위한 사후조사 보정이라고 부른다. 자료에 대한 이런 조정을 실행하기 위해 이용할 수 있는 몇 가지 통계적인 방법이 있지만 그것들은 크게 “가중과 대체”라는 두 개의 일반적인 구분으로 분류될 수 있다. 이 사후조사 보정방법들을 설명하고 증명하는데 필요한 수학은 복잡하고 이 책의 범위를 벗어난다. 최근의 관련 연구결과는 Groves 등(2002)과 Brick와 Kalton(1996)를 참조하라.

많은 조사 사례에 있어서 유감스럽게도 무응답을 보완하기 위해 많은 노력을 기울이지 않았다. 이런 경우 조사 추정값은 응답자와 무응답자의 차이를 보완하는 조정이 없이 단지 응답자를 근거로 산출된다. 대부분의 경우 무응답 보정의 근거로 활용될 수 있는 무응답자에 대한 약간의 정보가 있다. 그런 경우 무응답에 따른 보정을 하지 않는 것은 응답자만을 근거로 한 추정값이 무응답자를 포함한 전체 모집

단을 잘 나타낼 수 있다고 가정하는 것이기 때문에 무응답 보정이 반드시 필요하다. 이것은 사실상 분석가가 $\bar{Y}_R = \bar{Y}_{NR}$ 을 가정한다는 것을 의미하는데, 이것이 사실이라는 명백한 증거가 있는 것이 아니라면 이론적으로 이는 옳은 방법이 아니다. 단위 무응답을 보완하기 위해서는 조사 가중값(9장 참조)에 가중조정(*weight adjustments*)이라고 하는 요소를 곱하는 것이 일반적인 방법이다. 이 조정 요소들은 \bar{Y}_R 의 추정값을 모집단 평균 \bar{Y} 에 가깝도록 보정한다.

대부분의 가중법은 추정량의 표준오차가 늘어나게 하지만, 무응답 보정의 결과 줄어드는 편향이 결과적으로 MSE를 더 작아지게 한다. 모든 가중절차는 가정들을 근거로 한다. 만일 이 가정이 맞지 않으면 무응답 편향이 예상되는 정도로 줄어들지 않을 것이다.

대체는 항목무응답을 처리하는데 사용되지만 어떤 경우에는 단위무응답에도 사용될 수 있다. 대체는 결측 자료를 가공의 또는 “모형화 된” 자료로 대신하는 절차이다. 결측값을 대체하는데 사용할 수 있는 적당한 모형을 결정하는데 상당한 통계적인 작업을 필요로 하기 때문에 대체는 중요한 조사 특성에 대해서만 시행한다. 각각의 특성에 대해서 별도의 다른 모형이 필요하다.

대체가 처음 도입되었을 때 이 방법은 완전한 자료를 만드는 수단이었다. 이를 통해 전산장비가 덜 발달되었던 시기에 계산상의 문제를 해결하였다(Ogus 등, 1965 참조). 대체방법의 사용을 허용할 수 있는 항목무응답 수준에 대한 엄격한 기준(많아야 5%)이 예전에는 있었다. 완전한 자료를 얻는 것이 아직도 대체의 목적이지만, 오늘날에는 전산장비의 발전으로 계산상의 문제는 없다. 또한 “5% 기준”은 더 이상 적용되지 않는다. 오늘날 결측값 대체는 무응답률이 훨씬 더 높은 자료에 대해서도 사용한다.

결측값을 대체하는 많은 방법들이 있다(Dillman 등, 2002). 대표적인 세가지 예는 다음과 같다.

- 핫덱대체(*hot deck imputation*)에서는 결측 변수값을 응답자 중 한 명의 값으로 대체한다. 흔히 무응답자와 비슷한 특성이 있는 응답자 중에서 무작위로 선택된 응답자의 관측값으로 결측값을 대신한다.
- 최근방대체법(*nearest neighbor imputation*)에서는 적당한 거리 척도를 기준으로

로 무응답자와 가장 “가까운” 응답자로부터 대체값을 얻는다. 명백하게 이 기법은 핫덱과 비슷하지만 핫덱은 랜덤한 과정이고 최근방대체법은 결정적인 과정이다.

- 직접모형화(*direct modeling*)에서는 예측 변수로 결측된 특성과 관련이 많은 것으로 알려진 조사 항목을 사용해서 결측된 특성을 예측하는 통계 모형을 개발한다. 예를 들어, 응답자의 교육수준, 나이, 성별, 또는 설문지의 다른 특성을 이용해서 결측된 소득을 예측할 수 있도록 소득에 대한 대체모형은 구현한다.

무응답 보정작업은 상당히 복잡할 수 있고 전문가가 실행해야 한다. 조사 책임자는 조사 추정절차의 일부로 보정 방법이 포함되도록 해야 한다.

무응답 보정 방법(*methods for nonresponse adjustment*)의 기본적인 두 가지 방식은 가중과 대체이다.

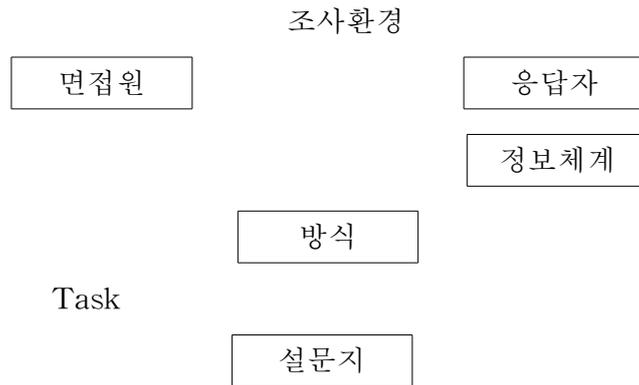
제 4 장 측정과정과 설문지 설계

이 장과 5장, 6장에서는 비표본추출오차의 다양한 요소들에 대해 생각해 본다. 조사에서 비표본추출오차의 발생원인 중 가장 복잡한 것이 측정과정이며 이는 다양한 요인들로 구성된다. 이 장에서 설문지에 대한 응답자와 면접원(만약 있으면)의 상호작용을 다룬다. 우선 측정과정을 구성하는 다양한 요소를 소개한다. 설문지는 이 요소들 중에서 중요한 부분을 차지한다. 이 장에서는 질문에 응답하는 과정을 살펴본다. 응답자들은 조사 질문을 이해하고, 질문에 답하는데 필요한 적절한 정보를 기억하거나 추론하고, 주어진 응답선택 범주를 참고해서 어떻게 응답을 할 것인지를 생각하고, 마지막으로 응답한다. 뒤에서 보겠지만 응답자가 응답에 이르는 동안 많은 장애요인이 있을 수 있다.

4.1 측정오차의 요소

<그림 4.1>에서 볼 수 있듯이 측정과정은 조사의 측정오차에 원인이 되는 여섯 가지 기본적인 요소로 이루어져 있다. 그것은 면접원, 응답자, 자료수집 방식, 설문지, 정보체계, 면접 환경이다. 자료수집 방식은 면접에 사용되는 전달 수단을 가리킨다. 그것은 전화, 대면 면접, 또는 면접원이 없는 자기기입식 설문지일 수 있다. 정보체계는 응답자가 응답을 하는데 이용될 수 있는 대량의 정보를 가리킨다. 예를 들어 그것은 물리적인 기록, 가구에 있는 다른 사람, 또는 자신의 기억력이 될 수 있다. 마지막으로 조사환경은 면접이 실시되는 장소인 집, 교실, 야외, 병원 등의 조사여건을 의미한다.

어떤 측정과정이든지 중요한 요소는 조사 정보를 수집하는 데 쓰일 도구이다. 이것은 물리적인 측정에 사용되는 기계적 장치(예를 들어, 혈압계), 또는 사람으로부터 자료를 수집하거나 기록되어 있는 정보를 문자로 바꾸는 데 사용되는 도구인 종이 또는 컴퓨터화된 설문지 일 수 있다. 조사 연구에서 측정과정은 주제, 응답자, 또는 과정에서 모아지는 정보를 제공하는 사람과 자료수집자, 면접원, 또는 도구를 사용하는 동시에 측정이나 응답 결과를 기록하는 사람 등을 필요로 한다.



<그림 4.1> 측정오차의 잠재적 원인

면접원과 응답자 사이의 상호작용의 본질은 도구가 적용되는 방법인 자료수집 방식(mode of data collection)에 의해 크게 좌우된다. 예를 들어, 면접을 전화 또는 대면으로 실시할 수 있고 면접원이 종이와 연필을 사용하는 설문지[지필면접(paper-and-pencil interviewing; PAPI)이라고 부른다] 또는 컴퓨터화된 설문지[컴퓨터보조면접(computer-assisted interviewing; CAI)라고 부른다]를 사용할 수 있다. 자기기입식 방법에서는 응답자가 면접원의 도움없이 설문지를 완성한다. 혼합방식조사도 흔히 있다. 예를 들어, 민감한 주제에 대한 조사들에서는 민감한 질문들에 대한 응답은 자기기입식으로 수집하고 면접의 다른 부분은 면접원이 수집한다. 우편조사에서는 무응답자를 추적하는데 전화나 대면면접 또는 모두를 사용할 수 있다. 또는 어떤 조사에서는 참여를 촉진하기 위해 응답자들에게 자료수집 방식을 선택하게 할 수도 있다.

직접관측(direct-observation) 조사에서는 자료를 응답자가 제공하지 않고 면접원이나 관측자가 직접 수집한다. 예를 들어 농작물 수확량 조사에서 농업관리인(관측자)이 어떤 작물의 산지에서 직접적인 물리적인 측정과 함께 시각적으로 예상 수확량을 추정할 것이다. 이 경우 응답자는 없다. 마찬가지로 어떤 조사에서는 면접원이 응답자, 응답자의 주거지, 또는 이웃, 면접하는 동안 식구들의 태도 등을 시각적으로 조사를 하는 것이 필요하다. 다른 자료수집에서는 자료수집자가 회사나 기구의 기록에서 정보를 종이나 컴퓨터화된 형태로 바꿔써서 필요한 조사 정보를 얻을 수 있다.

정보체계는 설문지를 완성하는데 도움이 될 수 있는 다양한 정보의 근원을 가리

킨다. 예를 들어 응답자가 외부 정보 자료의 도움 없이 대답할 수도 있고 가구나 사업체에 있는 다른 사람들, 회사 데이터베이스 또는 기록 체계, 또는 달력 메모, (수표책 등에서) 떼어 주고 남은 부분, 영수증 등 집에 있는 증거물에서 얻은 정보를 참고로 할 수 있다. 마지막으로 면접 환경은 면접이 실시되는 환경이다. 소음이 많은 사무실이나 집, 응답자가 트랙터에 앉아 있는 야외, 집 앞, 학교 식당 등이 면접을 할 수 있는 장소가 될 수 있다.

<그림 4.1>에 있는 오차 요소 사이에 있는 화살표는 요소들은 서로 밀접한 관계가 있고 측정과정에서 서로 영향을 미친다는 것을 시사한다. 그러므로 한 요소로 인한 오차는 과정의 다른 요소에 의해 영향을 받거나 변경될 수도 있다. 한 예로 면접원과 응답자 사이의 상호작용은 면접 방식, 조사환경, 설문지설계에 영향을 받는다. 전화면접에서는 응답자와 면접원 사이의 접촉이 대면 면접보다 더 짧고 덜 사교적일 수 있다. 대면 면접은 면접원과 응답자 사이의 접촉이 더 길고 환경이 더 사교적인 경향이 있다. 각 질문에 정확하게 대답하려는 응답자의 능력과 마음도 조사환경의 영향을 받을 수 있다. 조용한 방에서 면접을 하는 것이 응답자의 아파트 밖의 복도 같은 불편하고 시끄러운 환경에서 면접하는 것보다 응답자가 더 심사숙고하여 정확하게 응답하게 해 줄 수 있다.

응답자가 접근하는 기록 체계는 응답 정확도에 엄청나게 큰 영향을 미칠 수 있다. 회사나 가구 기록은 조사환경, 면접 방식, 면접 시기 등 때문에 정확하지 못하고 오래되고 완전하지 않거나 이용하기 어려울 수도 있다. 한 예로 사업체조사 면접이 면접에 필요한 기록을 활용하기 어려운 직원의 좁은 작업대에서 실시 될 수 있다. 더욱이 작업대는 공간이 좁아 응답자가 질문에 정확하게 대답하기 위해서 필요한 자료와 파일을 펼쳐 놓을 공간이 부족할 수 있다. 가구조사에서는 응답자가 면접원을 집안에 들이기를 거부하기 때문에 문 앞에서 다소 긴 면접을 실시해야 할 수도 있다.

측정오차의 요소 사이의 복잡한 상관관계의 결과 특정한 오차 원인이 응답에 미치는 영향을 예측하지 못할 수도 있다. 그것은 그 오차 원인과 다른 오차 원인 사이의 상호작용이 응답 과정에 미치는 영향을 변화시킬 수 있기 때문이다. 조사방법 문헌에 보면, 조사되는 모집단, 면접원의 특징 등에 따라서 특정 면접 방식이 조사 응답에 매우 상이한 결과를 가져오는 몇 가지 예가 있다. <그림 4.1>에서 볼 수 있듯이 특정한 측정오차의 원인이 응답에 미치는 영향을 파악하기 위해서는 다른 오

차 요소와의 상호작용을 고려해야 한다. 그러므로 과정의 한 요소가 자료 품질에 미치는 영향이 과정의 다른 요소의 변화함에 따라 어떻게 바뀔 수 있는지를 생각해야 한다.

이 장 나머지는 측정과정의 세 가지 중요한 요소인 응답자, 설문지 또는 도구, 정보의 출처를 중점적으로 다룬다. 우리의 연구에서 특별히 흥미를 끄는 것은 응답자가 응답에 이르기까지의 과정이다. 5장에서는 면접원 오차에 대해 다룬다. 6장에서는 자료수집 방식과 조사환경에서 발생하는 오차를 다룬다. 과정의 요소를 이해하는 것이 측정과정을 용이하게 하고 품질을 높일 수 있는 조사 설계와 관련된 소중한 통찰력을 제공할 것이다.

4.2 설문지 설계에 따른 오차

설문지 연구의 기원

조사연구자들은 질문의 다른 표현법이 특히 견해와 관련된 질문에서 사람들의 응답에 미치는 영향의 가능성에 대해 오랫동안 인정해 왔다. Rugg(1941)에서 자주 인용하는 유명한 예가 있다 “미국이 민주주의에 반대하는 대중 연설을 금하는 것이 옳다고 생각합니까?”라고 물었을 때 응답자의 54%가 “예”, 그것을 금해야 한다고 했다, 미국이 민주주의에 반대하는 연설을 허용하는 것이 옳다고 생각합니까?“라고 물었을 때 응답자의 75%가 ”예“ 라고 했는데 이것은 단지 25%만 대중연설을 허용하지 않겠다는 것을 의미한다. 이 결과는 단순히 질문의 표현을 변경한 것이, 즉 금하다(*forbid*)를 허용하다(*allow*)로 대신한 것이 여론조사의 결과를 바꿀 수 있다는 것을 보여준다. 조사를 다른 표현의 질문을 사용해서 다른 조사 기관이 다시 했을 때, 질문의 표현에 따른 차이는 뚜렷하게 나타날 수 있으며 이는 견해에 대한 질문에서 뿐만 아니라 행동에 대한 질문에서도 나타난다(Sudman 등, 1974 참조).

같은 질문을 다른 두 가지 표현방식으로 표현했을 때 나타나는 응답의 차이를 연구하기 위해 자주 쓰이는 실험법이 분할투표실험(*split-ballot experiment*)이다. 가장 간단한 형태로는 표본을 반으로 랜덤하게 나누고 각각의 다른 표현 형식의 질문을 받게 한다. 같은 방법을 한 표본을 3그룹 이상으로 나누어 적용할 수도 있다. 표본의 분리는 랜덤하게 하기 때문에 각 부차표본은 평균적으로 같은 추정값을 가져

을 것이다. 그러므로 각 분할된 표본에 설문지만 다르게 조사된다면 추정값의 차이를 주로 설문지 차이로 해석할 수 있다.

Seymour Sudman과 Norman Bradburn은 조사에서의 질문 응답 효과에 대한 연구의 선구자들이다. 1970년대 초반에 이 방법론 연구자들은 조사관련 문헌에 나타난 900개 정도의 분할투표와 측정오차 연구들에 대한 메타분석을 실시했다(Sudman 등, 1974). 메타분석(*meta-analysis*)은 잘 명시된 특정한 연구 질문에 대한 몇 개의 독자적인 연구들에서 나온 양적인 결과들을 결합시키는 통계적 접근 방법이다. 그것의 목적은 어떤 논점에 대한 (발표되거나 또는 발표되지 않은) 모든 연구에서 공통적으로 다루어지는 연구 논점에 대한 일반적인 답을 얻는 것이다. Sudman과 Bradburn의 분석은 <그림 4.1>에 있는 대부분의 측정오차의 요소를 고려했지만 주로 설문지 길이, 어려운 단어 사용, 서술형(즉, 응답자한테 말 그대로의 응답을 얻는 질문)과 선택형(즉, 응답자가 선택해야 하는 응답종류를 제공하는 질문) 같은 질문의 형태, 설문지에서 질문의 위치, 응답자에게 있어 질문의 중요도, 응답자가 정보를 기억하는 데 도움이 되는 도구의 사용 등과 같은 설문지 설계 변수를 중점적으로 다루었다. 그들의 연구에서는 다양한 설문지 설계 특징이 어떻게 조사 응답에 영향을 미치나에 대한 포괄적이고 통합된 양상을 제시하고 있다.

인지 심리학자들과 조사 연구자들이 서로 다른 분야의 연구에 기여할 수 있다는 것을 인식한 것 자체를 지난 10년 동안 일어난 놀랄만한 발전으로 볼 수 있다. 인지이론은 조사 응답자의 직무에 대한 이해와 질문의 배경과 같은 설문지의 양상이 조사 응답에 어떻게 영향을 미치나에 대한 이해를 쉽게 해 주었다(인지 방법의 설명은 8장을 참조하라).

설문지 설계의 목표

설문지 설계의 목표는 세부분으로 이루어진다. 첫째, 2장에서 언급했듯이 조사에서 다루는 연구 질문은 모집단에서 측정해야 할 개념을 암시한다. 설문지에는 이러한 개념들에 대해 면접원이나 응답자가 정보를 제공할 수 있도록 만들어진 조사 질문들이 있다. 그러므로 설문지 설계의 목표 중 하나는 연구가 의도한 의미를 정확히 전달할 수 있는 질문들을 만드는 것이다. 둘째, 응답자에게서 정보를 이끌어내기 위해 유리하면서, 가장 정확한 대답이 나오게 설계된 방법으로 질문들을 설문지에 기술해야 한다. 바꿔 말하면 설문지는 주어진 예산과 조사의 제약조건 아래서 계통과

가변오차를 최소화할 수 있게 설계되어야 한다. 또한 조사의 분석적인 목적을 달성하는 조건을 만족한다면 가능한 설문지를 완성하는데 필요한 시간과 응답자 부담을 최소화해야 한다. 왜냐하면 이를 통해 조사 대상자의 참여 가능성을 높이고, 장기간의 응답에 따른 피로와 소홀함 때문에 생기는 응답오차의 발생 가능성을 줄일 수 있기 때문이다. 마지막으로 자료수집 비용이 예산을 넘지 않도록 설문지를 설계해야 한다. 많은 다양한 질문이 있는 복잡한 조사에서는 이러한 목적을 이루는데 상당한 능력과 경험이 필요하다. 다음 예가 설문지 설계에서 발생할 수 있는 몇 가지 복잡한 상황을 설명한다.

예 4.2.1 학생들의 건강 또는 체육 관련 교육시간 수를 추정하는 것을 목표로 하는 학교 교사들에 대한 연구를 생각해 보자. 자료는 초등학교(첫 6년의 교육), 중학교(7, 8학년), 고등학교(9학년에서 12학년)로 구분된 3단계의 교육과정에 대하여 모두 필요하다. 각 단계에서 전형적인 12년의 교육과정을 마친 학생들이 특정 건강 관련 주제에 대해 교육받은 “노출”시간 수에 대한 정보를 수집해야 한다.

예를 들어 한 가지 연구 질문은 현 교육제도에서 얼마나 학교에서 실시하는 육체적인 운동이 강조되는지를 파악하는 것이라고 하자. 조사연구자는 그 개념을 간단히 “각 학년의 학생들은 얼마나 자주 학교에서 주관하는 단체 운동에 참여하고, 이런 운동을 몇 시간동안 하는가?”라고 말할 수 있다. 설문지 설계자의 일은 이 연구 질문을 교사가 개념을 이해할 수 있고, 그래서 그것에 정확히 대답할 수 있도록 각 학년에 어울리게 조사 질문들을 구체적으로 작성하는 것이다.

모든 학년에서 잘 실시될 수 있는 설문지를 설계하려면 설문지 설계자들은 면접원이 각 학년의 교사들을 면접할 때 마주칠 다양한 시나리오를 질문들이 반영하거나 설명할 수 있도록 학교가 운영되는 방법에 대한 약간의 지식이 있어야 한다. 여러 개의 설문지를 개발해야 될 필요가 있을 지도 모른다. 또한 모집단에 대한 상당한 지식과 연구자의 의도를 필요로 하는 몇 개의 문제점이 생길 수도 있다. 다음과 같은 예들을 생각해 보자.

- 조사를 봄학기에 한다고 가정한다면, 질문은 어느 기간을 기준으로 조사되어야 하는가(즉, 조사 기준기간(*reference period*)이 무엇인가)? 지난 가을학기, 현재 봄학기(면접 당시에 아직 지나지 않은 기간을 포함해서), 또는 둘 다에 적용되

는 것인가? 아니면 이 질문이 이전 학년을 의미하는가?

- 특정한 주제의 “노출” 시간의 추정값이 얼마나 상세해야 하는가? 예를 들어 교사가 어떤 주제를 가르친 시간을 답하려고 할 때 “0 또는 없음”, “1에서 10 시간” 또는 “10시간 이상”의 선택범주를 제시하는 것이 충분한가 아니면 더 상세해야 하는가?
- 어떤 기억 보조물이 도움이 될까? 예를 들어 이 질문의 대답을 하는 데 도움이 될 교안 또는 강의계획서에 따라 교사가 응답해야 하는가? 만일 그렇다면 이것을 정확하게 따를 것인가? 아니면 그것에서 꽤 벗어날 있는 것인가? 이것이 초등학교, 중학교와 고등학교에 따라 다른가?
- 육체적 운동에 대한 질문에 접할 때, 연구자가 “운동”으로 여기는 것이 무엇이며 이 정의가 어떤 교사들에 의해 잘못 해석될 수 있지 않은가? 한 예로 고등학교 수준에서 학생들은 체육 과목을 수강할 수 있는데, 초등학교 수준에서는 체육에 대한 정식 과목이 없을 지도 모른다. 이 두 단계의 교사들이 육체적 운동에 대한 매우 다른 인식을 가지고 있을 가능성이 있고, 어느 쪽도 연구자가 의도한 의미와 일치하지 않을 수도 있다.

이 문제점들에 대한 목록에서 볼 수 있듯이 분명하게 이해가 되고 쉽게 대답할 수 있는 질문들의 개발은 조사방법론 전문가와 조사연구자 또는 관련 주제 전문가들 사이의 공동 작업을 필요로 한다. 관련 주제 전문가는 학교 체계와 교육 관행에 대한 필요한 지식을 제공하는데 최고일 것이고, 조사방법론 전문가의 장점은 자료 수집상의 제약조건하에서 측정오차를 가능한 줄일 수 있는 조사 질문들을 개발할 수 있다는 것이다.

조사방법론 전문가와 관련 주제 연구자 사이의 긴밀한 협력을 통해 보통 이런 협력 없이 개발된 설문지보다 특정화오차와 측정오차가 훨씬 덜 발생하는 설문지를 작성할 수 있을 것이다. 그렇지만 흔히 이상적인 상황에서 설계된 설문지조차도 중요한 결점이 있을 수 있으며, 추가적인 수정과 개선이 필요하게 될 것이다. 예를 들어 설계과정에서 예상하지 못한 상황 때문에 어떤 질문들은 여전히 응답자에게 혼란스러울 수 있으며, 질문순서가 서투르거나 부자연스러울 수 있고, 또는 응답 범주가 너무 제한적일 수도 있다. 설문지 길이가 예산상 허락하는 길이를 크게 넘어서서 어떤 질문들은 제거되어야 할 필요가 있을 수 있다. 주제와 조사방법 전문가들

이 설계한 설문지일지라도 실제 자료수집에 있어서 심각한 문제를 불러일으키는 것은 보기 드문 일이 아니다.

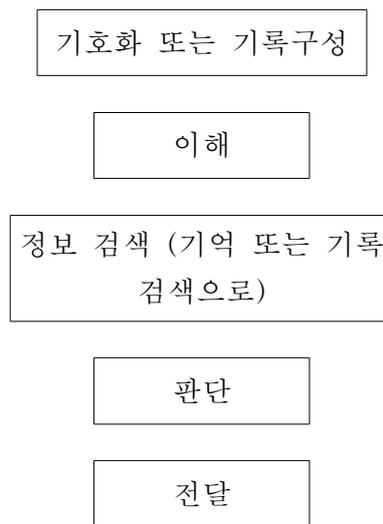
이런 문제들은 조사 이전에 설문지에 대한 사전 검사와 다른 평가를 실시하여 확인할 수 있다. 흔히 사용되는 방법은 조사 대상인 목표 모집단을 대표하는 몇 명의 응답자와 면접을 실시하는 것이다(예를 들어, 예 4.2.1에서 학교교사들). 면접은 실제 조사에서와 같은 방식으로 실시할 수 있고, 아니면 면접과정 중에 발생할 수 있는 문제들을 더 잘 확인하기 위해서 특별하게 설계한 인지면접기법(*cognitive interviewing techniques*)을 적용할 수도 있다. 이런 기법들을 사용해서 인지 면접원이 “그 답이 어떻게 떠올랐습니까?”, “---/---”라는 용어는 당신에게 무슨 뜻입니까?”, 또는 “그 숫자를 생각해 내는 데 사용한 과정을 말씀해 주세요.” 같은 질문들을 이용해서 응답자의 개념에 대한 이해를 시험할 수 있다. 설문지 문제점을 확인하기 위한 몇 개의 비면접방법 뿐 아니라 인지 면접기법과 다른 사전 검사 방법들은 8장과 10장에 자세하게 나와있다.

지난 15년 정도 동안 조사설계자들은 질문을 만들고 설문지를 설계하는 일에 인지 이론을 적용해 왔다. 설문지 평가에 널리 사용된 한가지 이론은 이른바 조사 응답 과정에 대한 인지이론이다. 다음 절에서 이 이론을 검토하고 설문지 설계에서의 문제점을 확인하고 고치는데 어떻게 사용될 수 있나를 보여준다. 8장에서 보겠지만 응답과정에 대한 이런 인지이론은 몇 개의 설문지 평가 방법의 근간을 이룬다.

4.3 응답과정의 이해

이 절에서는 설문지의 설계와 사전검사에서 널리 사용된 Kahn과 Cannell(1957)이 제안한 조사 응답 과정의 모형을 생각해 본다. 이 과정의 더 완전한 설명은 Tourangeau 등(2000)을 참조하라. 이 모형에서는 응답자가 한 개의 조사 질문에 대답하면서 5개의 뚜렷한 인지 단계를 순서대로 거쳐서 나아간다. 그것은 (1)기호화 또는 기록구성, (2)질문을 파악하거나 이해하는 것, (3)판단내리거나 평가하는 것, (4)질문에 대한 응답을 구성하는 것, (5)응답을 편집하고 전달하는 것이다. 이 단계들은 <그림 4.2>에 흐름도로 나타냈다. 이 응답 모형은 응답 과정을 다소 이상적으로 묘사한 실례를 반영하지만, 그림에도 불구하고 질문의 설계에 대해 생각하고 설

문지의 잠재적인 문제를 확인하는데 꽤 유용하다. 이 모형은 주로 개별면접을 위해 개발 되었지만 Edwards and Cantor(1991)과 Sudman 등(2000)에서 기업 조사에 전문적으로 사용할 수 있게 한 이 모형을 찾을 수 있다. 다음은 Biemer and Fesco (1995)에 설명한 한 모형의 간단한 해설인데 가구와 기업체 조사 모두에서 적용할 수 있다.



<그림 4.2> 응답과정의 5단계

조사 질문에 대한 대답을 명확하게 말하기 위해서는 응답자가 우선 타당한 대답을 하기 위해 필요한 지식, 믿음 또는 태도를 갖추어야 한다. 만약 요구한 정보가 회사 데이터베이스 같은 물리적인 기록을 참고하는 것이면, 정보를 검색할 수 있도록 면접시 그 자리에서 기록이나 자료가 입력되어야 한다(기호화/기록 구성; *encoding/record formulation*). 두 번째로 질문 전체 뿐 아니라 질문에 있는 각각의 단어들에 관해 연구자, 면접원과 응답자 모두가 동일한 의미로 받아들여야 한다(이해; *comprehension*). 다음에는 과거에 일어났던 일이나 행동에 관한 질문에 대하여 대답하려면 응답자는 기억력에서 필요한 정보를 검색하려고(*retrieve the required information*)할 것이다. 만약 요구되는 정보가 한 회사의 과거 성과에 대한 어떤 특징이면 응답자는 회사의 파일에서 정보검색을 시도할 수도 있다. 물론 태도, 믿음, 견해에 대한 질문들은 정보를 생각해 낼 필요 없이 면접을 하는 동안 응답자의 판단만을 필요로 한다. 하지만 이런 상황에서조차도 응답자는 전에 형성한 판단을 생각해내려고 할 수도 있다.

일단 정보를 검색하거나 판단하면 응답자는 그것을 어떻게 면접원에게 전달할 것인가를 결정해야 한다. 그러기 위해서는 그는 그것이 응답자가 선택해야 할 대답 종류에 들어맞게 하기 위해 대답을 구성(*format the response*)해야 할 필요가 있을 수도 있다. 마지막으로 응답자는 정확하고 정직하게 대답하는 것의 손익을 참작하며 면접원에게 대답을 전달해야(*communicates*)한다. 또는 만약 자기기입식 설문지 방식이면 대답을 용지에 기록한다. 이 단계에서 사회적 기대부응의 영향과 공개의 두려움, 또는 묵인(즉, 동의 또는 동의안함 형식의 질문에 내용에 상관없이 무조건 “예”로 응답하는 경향) 때문에 응답자는 그들의 대답을 교정하거나 수정하기로 결정하고, 이에 따라 자신의 알고 있는 것이 확실치 않은 것으로 간주하고 다르게 대답할 수 있다. 이 개념들은 아래에서 더 자세히 이야기 한다.

이 모형의 중요한 장점은 이 모형이 조사의 설계와 평가에서 별도로 다룰 수 있게 응답과정을 작게 분해시킨다는 것이다. 이 모형의 한 가지 단점은 조사설문지를 완성하는 데에서의 복잡함을 피하기에 모형이 너무 단순하다는 것이다. 예를 들어 실제로는 응답자가 용어를 이해하거나 옳은 정보를 생각해 내려고 하지 않고 질문에 대답할 수도 있다. 응답자가 피곤하거나 무관심하거나 그냥 재미로 이런 것이 일어날 수 있다. 이런 응답 과정을 짧게 하는 것을 때로는 만족스러운(*satisfying*) 행동이라고 한다(Krosnick, 1991 참조). 설문지가 길고 단조롭거나 응답자가 좋은 대답을 제공할 의욕이 없으면 응답자는 최소한의 필요조건을 추구하거나 정확하게 답하려고 노력하지 않고(즉, 최적으로(*optimizing*) 하지 않고) 대답할 수도 있다. 앞으로 언급 되겠지만 우리가 이 책에서 생각하는 다른 간단한 모형들처럼 이 모형은 여러 다양한 상황에서 설문지설계와 관련해 다소 중요한 통찰력을 제공한다는 면에서 여전히 유용하다. 다음에는 이러한 응답모형이 조사에서의 측정오차의 통제와 측정에 어떻게 도움이 되는지를 알아본다.

기호화와 기록생성

정보의 기호화는 정보를 습득해서 기억에 저장하는 과정이다. 사건이나 경험을 측정과정에서 상기하거나 검색해 내려면 그것의 기록을 만들어야 한다. 예를 들어 조사에서 식구의 행동에 관한 질문을 응답자가 정확하게 대답하려면 그 행동을 우선 관찰하고 면접 중에 기억해 낼 수 있도록 암기해야 한다. 이와 같이 만약 면접 중에 사업체의 데이터베이스에서 사업상 거래나 숫자상의 자료 항목을 생각해 내야

하면 그 정보가 사업체의 데이터베이스에 우선 기록이 되어야 한다.

기호화 또는 기록구성에서는 지식을 얻고 처리하고 기억에 저장하거나 물리적인 기록을 만든다.

응답과정에서 기호화나 기록생성단계가 측정 과정을 착수하기 여러 달이나 여러 해 전까지도 일어날 수 있는 유일한 단계이다. 그럼에도 불구하고 만일 자료가 기억에 기호화 되거나 물리적인 기록을 보관한 적이 전혀 없다면 응답자들이 사실, 사건, 다른 자료를 생각해 내기를 기대할 수 없기 때문에 기호화나 기록생성단계는 과정의 중대한 부분이다.

정보를 기호화하는 것에 실패하는 것은 조사에서 오차의 중요한 원인이다. 예를 들어 설문지에서 요구한 정보를 대리 응답자가 주는 것을 허용하는 조사를 생각해 보자. 대리 응답자(*proxy respondent*)는 다른 사람에 관한 질문에 응답을 제공하는 사람이다. 예를 들어 나이가 아주 많은 사람들에게 대한 조사는 표본구성원 자신이 응답할 수 없을 때 표본구성원을 돌보는 사람들이 질문에 답하는 것을 허용할 수 있다. 질문의 주체가 되는 사람들이 항상 접근하기 쉽거나 만날 수 있거나 자발적이거나 자신이 직접 참여할 수 있지 않기 때문에 무엇보다 먼저 이 질문들에 자기 응답(*self-response*)이라고 하는 자신들 만이 대답하는 것이 허용되면 무응답 때문에 생기는 조사에서 빠진 자료의 분량이 타당한 추정값을 제공하기에 너무 클 수 있다. 게다가 어떤 상황에서는 대리 응답이 자기 응답보다 더 정확할 수 있다. 유일한 대안이 무응답이라면 대리응답이 더 낫게 여겨진다.

한 예로 건강 조사에서 “지난 30일 동안 의사를 몇 번 방문했습니까?” 라는 질문을 각 가족 구성원에게 묻는다. 만일 이 질문을 자기응답으로 대답하면 (즉, 질문의 대상 인물이 대답을 하면) 이 정보는 의사를 방문할 때 기억에 기호화했기 때문에 응답 과정의 기호화 단계는 문제가 되지 않을 것이다. 물론 의사를 여러 번 방문했는데 그것들을 분명히 기억하지 못하면 기억오차가 여전히 문제가 될 수 있다. 그렇지만 만일 그 질문을 대상자의 의사 방문에 대해 모를 수도 있는 대리 응답자가 대답하면 그 정보가 한 번도 기호화된 적이 없기 때문에 기호화오차가 문제가 될 수 있다. 이런 경우에는 가장 바람직한 결과는 대리응답자가 이 정보를 모르는 것

을 고백하는 것이다. 불행하게도 대리응답자들은 그들이 필요한 정보를 기호화했는지와 관계없이 응답을 추측하거나 응답을 하려고 할 수도 있다. 이런 종류의 오차를 기호화오차(*encoding errors*)라고 부른다.

미국 노동력조사(*Current Population Survey; CPS*)에서는 기록하는 대리인의 기호화오차의 실례를 보여준다. CPS는 미국 센서스국이 미국 전체 노동 인구 모집단의 취업, 실업과 그 밖의 특징에 대한 추정값을 제공하기 위해 매달 실시하는 가구 표본조사이다. 이 조사에서 Roman(1981)은 자기응답이 대리인이 보고했을 때보다 실업률이 더 높았다는 것을 발견했다. 이것에 대한 그럴듯한 설명은 많은 대리응답자들은 직장이 없는 대상자가 그 전주에 직장을 구하는 중이었다는 사실을 모를 수도 있다. CPS규칙에 따르면 직장을 구하고 있었던 사람들은 실업자로 분류되는데 직장을 구하고 있지 않았던 사람들은 노동 인구에 있지 않은 것으로 분류된다. 그러므로 왜 대리 응답자들 보다 자기 응답자의 더 높은 비율의 실직을 보고 했는지를 기호화오차가 부분적으로 설명할 수 있다.

Moore(1988)는 대리보고에 대한 뛰어난 비평이 실린 논문을 발표하였다. Moore가 지적하듯이 거의 모든 대리연구(?)는 선택편향(*selection bias*)이라고 하는 중요한 문제가 있다. 즉, 대리보고자와 자기보고자의 응답을 비교할 때 많은 연구들은 간단히 자료를 보고자의 종류에 따라 분류하여 두 집단의 평균과 비율을 비교한다. 하지만 자기보고를 하는 사람들은 대리보고를 해 주는 사람들과 매우 다른 특징을 가지고 있을 수 있다. 대리인이 대신 보고하는 사람들 보다 자기보고자들이 더 접근하기 쉽고 만날 수 있고 협조적이어서 두 집단의 실제 특징이 다르다. 그러므로 보고 종류와 관련된 측정오차는 실제 집단 차이와 얽히거나 혼동된다. 이 이유로 대리보고와 자기보고의 정확도를 비교하는 많은 문헌은 설득력이 없다.

기호화오차는 자기보고에서도 마찬가지로 문제가 될 수 있다. 미국 국립질병통제 센터(CDC)에서 실시한 연구는 아동 면역 조사에서 기호화오차의 좋은 실례를 제공한다. 미국 국립면역연구(*National Immunization Study; NIS*)는 어린이들이 2살 때 까지 맞은 예방 접종에 대해 부모들로부터 정보를 수집한다. 이 나이의 어린이들은 5가지 다른 백신을 최소한 14회 맞았어야 하므로 세심한 부모들도 아이의 예방접종을 정확하게 보고하는 것을 어려워한다. 이유를 자세히 조사하기 위해 <그림 4.2>에 있는 인지 응답모형을 지침으로 연구들을 실시했다. 처음에는 주사 기록이나 다른 기억보조들의 도움이 없는 부모의 보고는 어린이들이 받은 예방접종을 줄여 말

하는 경향이 있다고 NIS가 관찰했기 때문에 기억오차가 잘못 보고하는 가장 중요한 원인이라는 가설을 세웠다. 그렇지만 또한 가설은 문제가 기호화라는 것이다 (즉, 예방 접종을 투여했을 때 부모들이 그것에 대하여 잘 알지 못하기 때문에 부모들의 보고는 잘못되었을 수 있다).

두 가설을 독립된 연구에서 시험해 보았다. 기호화오차를 시험하기 위해 7세 이하의 어린이들에 대한 연구를 소아과에서 실시했다. 어린이들과 병원 방문한 부모들에게 자기 아이들의 의료 방문에 대해 병원을 나가면서 짧은 면접을 완성하라고 부탁했다. 놀랍게도 예방접종을 투여 받은 직후에도 대부분의 부모들은 자기의 아이들이 그 날 받은 예방 접종에 대해 조금밖에 몰랐다. 가장 일반적인 실수는 주사를 맞지 않고서도 맞은 것으로 보고하는 것(즉, 틀린 긍정적 보고)보다는 오히려 주사를 맞은 것을 모르는 것이었다(즉, 틀린 부정적 보고). 틀린 부정 비율은 거의 50%인데 틀린 긍정 비율은 단지 18%였다. 이 연구는 부모의 보고 오차는 기호화오차이고 보고의 정확도를 높이기 위해서 기억단서나 기억 보조물의 사용은 효과가 없다고 결론을 내렸다(Lee 등, 1999 참조).

또 한 가지 종류의 기호화오차는 응답자가 한 질문 주제에 대해 완전하지 않고 왜곡되고 정확하지 않은 정보만 가지고 있을 때 발생한다. 예를 들어 농장 경영자의 조사에서 농민들에게 농업에 사용되는 그들이 소유하는 땅의 특정한 구획의 가치를 추정하라고 물었다. 그들의 땅을 팔 의향이 없는 어떤 농민들은 그들의 땅의 가치에 대해 추측조차도 하지 않았다. 그렇지만 어떤 응답자들은 그들이 추정을 근거로 할 아무 정보가 없어도 추정을 할 수가 있었다. 근처에 있는 땅이 어떤 액수, 이를테면 1에이커에 \$10,000에 팔리고 있다고 들었을 수 있고 그들의 땅도 그만큼 가치가 있다고 추측할 것이다. 그렇지만 그들이 가지고 있는 정보는 정확하지 않던지 아니면 땅의 실제 가치를 나타내지 않을 수도 있다. 이것이 응답자의 대답이 조사 질문 주제에 대한 정확하지 않거나 완전하지 않은 정보에 의해 어떻게 잘못 전해질 수 있는지의 한 예이다.

사업체 조사에서는 사업체의 기록이 빠졌거나 완전하지 않거나 조사에서 요구되는 것과 모순될 때 응답 과정의 기록구성 단계에서 생기는 오차가 일어날 수 있다. 조사용지에서 요구한 정보가 사업체의 데이터베이스에 저장된 자료와 비슷하지만 실제로 매우 다른 것은 보기 드문 일이 아니다. 정보를 조사 요구와 일치되게 재포맷하기 보다 응답자가 정보를 사업체의 데이터베이스에서 바로 제공할 때 설문

지와 정보의 출처사이의 불일치가 오차의 원인이 된다.

미국노동통계국에서 실시한 미국 고용조사(*Current Employment Survey; CES*)에서 이런 종류의 오차의 한 예가 발생했다. CES 자료의 품질평가에서 Poinkowski와 Meily(1989)는 기업의 59%가 실제의 정의에 충실하지 않은 것을 발견하였다. 가장 큰 문제점은 조사 설문지가 보수 없이 휴가중인 고용인들은 급여지불명부(*payroll*)에 포함시키지 않을 것을 요청했지만 많은 회사들은 이 고용인들을 포함시켰다. 왜 이런 실수를 했냐고 물었을 때 실수를 범한 약 40%의 응답자는 그 원인이 조사요구와 그들의 회계 체계의 불일치라고 했다. 그 사업체들은 조사요구를 만족시키기 위해 급여 지불 명부 자료를 재구성하기(*reconstruct*)보다는 오히려 회사 데이터베이스에서 쉽게 이용할 수 있는 숫자를 간단히 주었다.

그러므로, 개인 특징과 행동에 대하여 묻는 설문지를 설계할 때에 주요 결정 사항은 대리응답의 허용 문제이다. 이 전략은 한편으로 어떤 항목에 대하여 정확하지 않은 정보를 얻는 위험과, 다른 한편으로는 자료가 빠지는것에 대한 비교 검토를 하여야 한다. 어떤 조사에서는 전혀 자료가 없는 것보다는 정확하지 않은 자료라도 얻는 것이 더 나을 수 있다. 대리 응답 규칙을 사용하지 않거나 단위 무응답을 피하기 위한 마지막 수단으로 대리 응답 규칙을 사용되어야 하는 상황들이 있다. 예를 들어 여론이나 태도질문에서는 대리응답을 받아들일 수 없다. 만일 대리 답이 허용되면 면접원들이 부닥칠 다양한 상황에서 적당한 정보제공자를 확인하는 결정 규칙을 명기해야 한다.

예를 들어 가구조사에서는 이상적인 정보제공자(즉, 조사를 위해서 정보를 제공하는 사람)는 보통 질문의 대상자(*reference person*)에 대하여 가장 잘 아는 가구에 속한 사람이다. 조사에서는 보통 배우자나 부모나 어린이들을 돌보는 사람이다. 그렇지만 흔히 질문의 주제와 가구구성원 사이의 관계에 따라서 이상적인 정보 제공자가 바뀐다. 조사에 있는 질문과 사람마다 이상적인 대리인을 면접하려고 하는 것은 실행 불가능한 일일 수 있다. 만일 각 가구에서 한 명만 면접한다면 가장 좋은 전략은 정보제공자에 조사의 주요 항목을 위하여 종합적으로 누가 최고인지를 확인하는 것일 수도 있다.

마찬가지로 사업체조사에서는 회사가 제공하는 정보의 정확도는 그것을 제공하는 사람에 의해 크게 좌우된다. 만일 회사의 운영비와 다른 지출에 대한 정보를 요구하면, 이상적인 응답자는 회사의 재무담당 최고 책임자일 수 있다. 하지만 만일 그

가 받아들일 수 있는 유일한 응답자면 이 조사의 거부율은 상당히 높을 수 있다. 협조률을 늘리기 위해 회사의 다른 직원들이 이 정보를 제공할 수 있도록 허용하게 응답자 규칙이 특정한 범위내에서 충분히 융통성이 있어야 한다. 특히 자기기업식 설문지와 회계, 전 직원, 경영, 생산을 포함한 다양한 주제에 대한 정보를 요구하는 조사들에서는 이것을 관리하기 어려울 수 있다. 이런 넓은 범위에 걸친 주제에 대한 자료를 얻는 것은 회사에서 한명이 아닌 여러 사람들과 면접을 하거나 협력하는 것을 필요로 할 수 있다.

이해(질문을 이해하는 것)

응답과정의 두 번째 단계는 이해 또는 질문을 이해하는 것이다. 이 단계에서는 응답자는 질문을 읽거나 듣고 어떤 정보를 요구하는지 이해하려고 한다. 그러므로 좋은 질문을 개발하기 위한 중요한 목적은 응답자가 쉽게 이해할 수 있는 말로 정확히 어떤 정보가 필요한지를 응답자에게 설명하는 것이다. 응답자는 질문에 정확하게 대답하기 위하여 이 단계는 없어서는 안 된다. 이 단계에서 응답 과정에 몇 가지 종류의 오차가 소개될 수 있다.

질문을 이해(*comprehend* 또는 *understand*) 하려면 응답자는 질문을 숙고하고 어떤 정보를 요구하나 이해하려고 해야 한다.

첫째, 질문의 표현법이 복잡할 수 있고 또는 생소한 용어를 포함할 수 있다. 예를 들어 “몇 년도에 이 대학에 입학이 허가 되었습니까?”의 질문은 어떤 학생들은 이해를 못할 수 있다. “몇 년도에 이 대학교에 처음 등록했습니까?”는 더 쉽게 이해한다. 질문의 문자 그대로의 의미를 넘어서 연구자가 의도했던 질문의 해석도 응답자에게 정확하게 전달되어야 한다. 예를 들어 “당신은 자동차를 소유합니까?”라는 질문에는 생소하거나 복잡한 단어가 없지만 그래도 응답자는 어떤 정보가 요구되는지 아직도 이해를 못할 수 있다. 자동차 소유의 의미가 무엇인가? 한 사람이 자동차를 구입했지만 아직도 자동차의 할부금을 지불하고 있다고 가정해 보자. 구입이라기보다는 오히려 3년 기간 동안 리스를 하고 있는 경우도 있다. 남편과 아내가 공동으로 소유하는 자동차일 수도 있고 합법적으로 양도한 적은 없지만 아들이나 딸만 운

전하는 자동차 일 수도 있다. 이런 상황에서 “소유”로서 자격이 있는 것인지 명백하지 않기 때문에 응답자는 갑자기 떠오르는 자신의 해석에 따라 응답할 수 있고, 그래서 가변오차나 신뢰할 수 없는 응답을 만들어 낸다.

이 단계에서 발생할 수 있는 또 하나의 문제는 문맥효과(context effects)이다. 설문지에 나온 이전의 질문들, 질문 앞에 나온 부분제목, 질문에 답하는 것을 위해 제시한 지시 등 같은 설문지에 나오는 다른 정보에 의해 질문의 해석이 영향을 받을 때 문맥효과가 일어난다. 문맥효과의 가능성 때문에 설문지에 있는 질문의 위치까지도 응답자들이 질문에 있다고 생각하는 의미에 영향을 미칠 수 있다. 예를 들어 “당신은 당신의 의료보험에 얼마나 만족하십니까?”하는 질문은 “당신은 당신의 의사에 얼마나 만족하십니까?”라는 질문이 앞에 나오지 않을 때와 나올 때 매우 상이한 응답을 이끌어 낼 수 있다.

흔히 설문지설계에서 설문지에 나온 사전의 질문이나 정보가 함축된 문맥의 질문의 이해를 용이하게 하는데 꽤 효과적일 수 있다. 한 가지 주제를 다루는 질문들을 설문지의 한 부분에 모으고 응답자에게 그 부분의 문맥을 명백하게 하면 질문에 대한 이해는 높아진다. 예를 들어, 농업조사에서 설문지의 한 부분 전체가 농장 운영 전체의 특징에 대하여 묻고 설문지의 다른 부분들은 농장안에 있는 특정한 땅의 구획하고만 관계된다고 하자. 그러면 설문지에서 각 질문전에 어느 질문들이 농장 운영 전체와 관련됐고 어느 질문들이 특정한 땅의 구획과 관련됐나를 표시하는 지시를 할 필요가 없다. 질문들은 모으는 것과 설문지 부분들의 문맥 때문에 응답자들은 어떤 부분들은 농장의 일부에 대한 것이고 다른 부분들은 운영전체에 대한 것이라는 것을 이해한다. 각 부분에 나오는 질문들은 그것들의 문맥으로 명확하게 적혀 있다.

그렇지만 질문의 문맥 때문에 흔히 응답오차의 결과를 낳는 질문들의 오해가 생길 수 있다. 만일 응답자가 실제로는 그렇지 않지만 설문지의 한 부분에 있는 모든 질문들이 같은 객체에 관계됐다고 실수로 믿으면 문맥효과가 일어날 수 있다. 한 예로 앞에서 설명한 설문지의 전체 농장 부분에 질문 중 하나가 농장 안에 있는 경작지나 한 구획 같은 농장의 일부하고만 관계됐으면 응답자는 문맥의 변화를 알아차리지 못하고 농장 전체에 대한 대답을 할 수도 있다.

만일 문맥이 바뀐 것을 변환지시문, 절제목, 볼드체 글자 등을 사용해서 더 알기 쉽게 만들 수 있으면 이런 문맥효과는 피할 수 있다. 그렇지만 응답자가 앞에 나오

는 질문들의 영향을 받을 수 있고 이런 영향을 예방하려는 노력은 효과적이지 않기 때문에 문맥효과를 모든 경우에 통제할 수 있는 것은 아니다.

일반적인 모집단 조사에서 전문용어나 모집단의 일부만이 그 의미를 이해할 수 있는 단어도 조사에서 이해오차를 가져올 수 있다. 예를 들어 미국 국가보건면접조사(National Health Interview Survey; NHIS)에 있는 한 질문은 “지난 12개월 동안 위염, 대장염, 장염, 계실염(憩室炎)에 걸렸었는가?”를 묻는다. 여기에서의 전략은 만일 응답자가 용어를 모르면 그는 그 질병을 앓지 않았을 것이라는 것이다. 하지만 전문용어는 보통 흔히 쓰는 용어로 대신할 수 있다. 예를 들어 “중이염(otitis media)”보다는 “귀가 감염됨(ear infection)”을 사용하는 것이 좋다.

이해오차는 예를 들어 영어에서 불어로의 번역과 같은 한 언어에서 다른 언어로 번역할 때 발생할 수 있다. 만일 번역을 문자 그대로 하고 모집단에 있는 불어를 하는 사람의 문화적인 차이나 의미의 미묘한 차이를 무시하면 번역이 전문적으로 정확하기는 하나 이해오차를 가져올 수 있다. 그러므로 질문의 엄밀한 축어적인 번역을 유지하려고 하기 보다는 알맞은 의미를 전달하게 새 언어로 질문을 바꾸어 말할 필요가 있을 수 있다.

마지막으로 응답범주 자체가 이해문제를 가져올 수 있다. 응답자들은 상당히 자주 응답범주를 질문을 해석하는 것을 돕는데 사용한다. 예를 들어 “당신은 자동차를 소유합니까?”라는 질문에 “명백히 소유함”, “구입중”, “리스하는 중” 등 같은 응답범주이 자동차 소유가 무엇을 의미하는지를 분명하게 하는 것을 돕는다. 질문은 명료하게 기술되었지만 응답범주에서 복잡하거나 의미가 애매한 용어를 사용하거나 질문과 잘 상응하지 않을 수 있다. 겹치거나 상호배반이 아닌 항목은 문제가 있지만 응답자들은 여전히 한 항목을 고르게 되어있다. 다음 예들이 면접 과정의 이해 단계에서 일어날 수 있는 어려움을 설명한다.

예 4.3.1 1977년 미국경제센서스(US Economic Censuses)에 뒤이어 미국 센서스국이 생산센서스(Census of Manufacturers)의 자료품질평가를 실시했는데 조사에서 이해오차의 증거를 찾았다. 연 급료지급금액으로 제공된 액수에서 잘못된 포함되거나 제외된 부분의 합이 연 급료지급금액의 센서스 총계의 약 2%인 \$37억이 된 것을 발견했다. 이 오차의 3분의 1 정도가 일년간 휴직하거나 휴가중인 고용인의 급료를 제외한 것에 원인이 있다고 했다. 어떤 사업체의 데이터베이스는 휴가 급료

를 포함하지 않고 연간 급여 지급금액을 입력했을 수도 있기 때문에 사업체 내에서 자료 기록에 오차가 있었을 수도 있다. 하지만 사업체 내에 대한 더욱 자세한 조사와 응답자와의 재 면접 후에 실제적인 문제는 조사 응답자가 휴가 급여를 포함해야 하는 것을 알고 있지 못했다는 것을 센서스국이 발견했다. 만일 질문에서 분명하게 휴가 급여를 포함해야 하는 것이라고 말했더라면 그들은 그것을 쉽게 포함할 수 있었을 것이다.

이 예는 사업체조사에서 상당히 흔히 있는 문제이다. 회계나 금융자료를 보고할 때 응답자는 어떤 정보(전직원, 지출, 봉급 등)를 항목에 포함해야 하는지 흔히 이해하지 못한다.

예 4.3.2 이 예는 Groves 등(1991)이 실시한 연구에서 제공했다. 응답자들에게 다음 두 질문을 하였다.

1. 당신 자신의 건강이 전반적으로 아주 좋음, 좋음, 보통, 좋지 않음 중 어떻다고 생각합니까?
2. 당신의 건강에 대하여 바로 전 질문에 대답했을 때 무슨 생각을 했습니까?
 - a. 당신 또래의 다른 사람들과 비교했을 때 당신의 건강.
 - b. 당신이 훨씬 젊었을 때 당신의 건강과 비교했을 때 당신의 현재 건강.
 - c. 최근에 비교했을 때 지난 몇 년 동안의 당신의 건강.

<표 4.1>은 질문의 각 해석을 표시한 응답자의 퍼센트를 보여준다. 응답자들이 그들의 전반적인 건강에 대하여 그들이 질문을 해석하는 방법에 있어서 상당한 불균형을 보여주므로 질문에 대한 그들의 격차가 있다(즉, 가변오차). 그러나 남자와 여자가 이 질문을 다르게 해석 할 수 있다는 사실에 근거하여 질문을 응답하는데 계통오차 또는 편향의 증거를 이 연구에서 찾았다. 예를 들어, 1번 질문의 답에서 여자의 단지 28%가 그들의 건강이 “아주 좋음”이라고 한데에 반하여 남자의 43%가 그렇게 대답했다. 그렇지만 성별비교가 같은 해석(즉, a, b 또는 c)을 한 응답자들로 제한되었을 때는 차이가 훨씬 작았고 해석 b에서는 사라지기까지 했다. 전반적으로 남자나 여자는 이 질문에 대하여 서로 달리 해석을 했고 그 결과 남자와 여자의 비교는 편향될 수 있다.

<표 4.1> 응답자들이 자신의 전반적인 건강에 대한 질문을 해석하는 방법

다른 사람과 비교	22%
더 젊었을 때와 비교	43%
지난 몇 년과 최근의 비교	42%

출처: Groves 등(1991)

정보검색

응답자가 질문을 이해한다면, 응답자는 이제 질문에 답하는데 필요한 어떤 정보라도 생각해 낼 준비가 되어있다. 정보기억단계에서 질문에 대한 응답을 구체화하기 위해 응답자는 필요한 정보를 생각해 낸다. 이 과정은 기호화 단계에서 장기기억에 저장한 정보를 상기하는 것, 컴퓨터 데이터베이스 또는 가구나 개인파일 같은 외부출처에서 자료를 생각해 내는 것, 또는 요구한 정보를 가지고 있는 가구나 사업체에 있는 사람과 의 상담을 필요로 한다. 의견이나 사고방식에 대한 질문이나 기본적인 인구통계학적 특징에 대한 질문들은 사실에 근거한 자료의 기억을 요구하지 않는다(즉, 사건, 날짜, 자서전적인 정보). 그렇지만 기억에서 느낌, 관점, 논란에 대한 입장 등 형태의 정보를 여전히 기억해 낼 수 있다. 또한 이 단계는 사고방식, 소신, 또는 의견에 도달하기 위하여 질문이 제기한 이슈를 심사숙고하는 반영과정을 포함한다.

정보검색(information retrieval)은 기억 또는 다른 가족 일원, 직장 동료, 회사 데이터베이스나 가구파일 같은 외부 출처에서 검색한 정보를 가리킨다.

<그림 4.2>에 나타냈듯이 요구한 정보를 두 가지 이상에서 입수할 수 있으면 응답과정의 이 단계에서는 자료 출처를 선택해야 한다. 예를 들어 질문이 바로 전 과세연도에 대한 총소득에 대하여 물을 수 있다. 이 정보는 전해의 소득세 신고와 개인 기억에서 입수할 수 있다. 만일 응답자가 최고의 답을 제공할 의욕이 있으면 그는 기억에 의존하기 보다는 오히려 전해의 소득세 신고를 사용할 수 있다. 마찬가지로

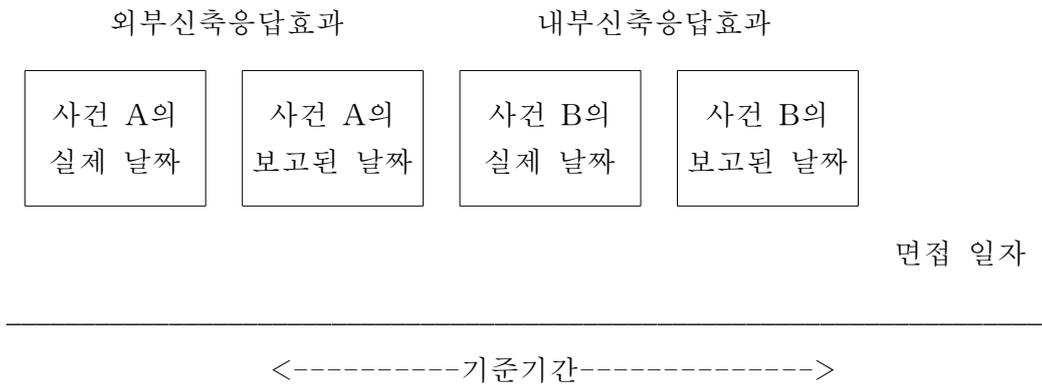
지로, 사업체 조사에서 질문이 응답자의 기관에서 일하는 고용인의 수에 대하여 물을 수 있다. 응답자는 그 수를 대략 알고 있고 그 수를 제공할 수 있고, 또는 회사의 인사 기록을 참고하기로 결정해서 더 정확한 숫자를 제공할 수 있다. 각 경우에 최종적으로 참고한 출처는 더 정확한 정보를 제공하는 데 필요한 부담, 질문에서 요구하거나 비쳐진 정확도(*degree of accuracy*), 설문지에 있는 다른 질문들을 근거로 얼마만큼의 정확도가 필요한가에 대한 응답자의 평가 등에 따라 결정된다.

누락오차나 기억오차는 기억에서 정보를 생각해내는 과정 중에 있어날 수 있다. 정보를 기억하지 못하는 두 가지 흔한 이유는 잊는 것과 신축응답효과이다. 망각(*forgetting*)은 장기기억에서 기억하는 것을 필요로 하는 질문에서 발생할 수 있다. 대체로 먼 과거에 일어난 일들이 가까운 과거에 일어난 일들 보다 더 잊혀질 가능성이 많다. 이것의 예외는 사랑하는 사람의 죽음이나 아기의 출생 같은 특히 두드러진 일 들이다.

신축응답효과오차(*telescoping error*)에서는 사건은 기억하지만 사건이 일어난 날짜가 정확하지 않다. 전진 신축응답효과오차(*forward telescoping error*)는 사건들이 면접 날짜에 더 가까이 일어난 것으로 기억될 때 발생하고 후진신축응답 효과(*backward telescoping*)에서는 사건들이 면접 날짜에서 더 멀리 일어난 것으로 기억된다. 외부 신축응답효과오차(*external telescoping error*)는 기준기간 밖에서 일어난 사건을 기준기간 안에서 일어난 것으로 잘못 보고하는 것을 말한다. 예를 들어 응답자가 사실은 의사 방문이 기준기간 전이었는데 그가 두 주의 기준기간 안에 의사에게 갔었다고 보고 할 수 있다. 이 경우 응답자는 그 일을 시간적으로 앞으로 신축했고 그 결과 그것이 기준기간 동안 일어나는 것으로 잘못 간주되었다. 내부 신축응답효과오차(*internal telescoping error*)는 사건들을 시간적으로 앞이나 뒤로 신축하는 것 때문에 기준기간 안에 일어난 일들의 시기가 틀렸을 때 발생한다. 예를 들어, 한 달 기준기간동안 일어난 도시 밖으로 나간 여행을 그것이 정말 일어난 것보다 더 오래전에 일어난 것으로 보고한다.(즉, 시간적으로 뒤로 신축되었다)(<그림 4.3> 참조).

망각은 사건을 실제 이하로 보고하게하고 그러므로 계통오차로 분류할 수 있다. 기준기간 동안 자주 일어난 사건을 응답자에게 사건의 건수를 세어보라고 했을 때 잊어버리는 것이 흔하다. 잊어버리는 것의 결과는 사건의 수를 실제 이하로 보고하는 것이다. 예를 들어, 응답자에게 지난 한 달 동안의 기간에 상관없는 자동차 여행

의 수를 세어보라고 한다고 가정하자. 만일 그 기준기간동안 여행을 많이 한 응답자에게 그것을 하나씩 세어보라고 하면 그는 잊어버리는 결과 실제 이하로 보고할 가능성이 높다. 이 망각의 결과는 여행을 많이 한 응답자 보다 여행을 몇 번 안한 응답자가 아마 더 적을 것이다. 그러므로 망각의 결과는 응답자가 한달에 간 평균 여행의 수의 추정값이 음으로 편향 되었을 수 있다.



<그림 4.3> 외부와 내부 신축응답효과. 사건 A는 기준기간 밖에서 일어났지만 기준기간 내에 일어난 것으로 보고 된다. 이것은 전진 외부신축응답효과의 예이다. 마찬가지로 사건 B는 전진 내부신축응답효과의 예이다. 후진 내부신축응답효과는 면접 날짜에 더 가까운 기준기간 중에 일어나는 사건들이 기준기간중이면서 실제보다 더 과거에 일어나는 것으로 보고 될 때 발생한다.

외부신축응답효과 오차도 편향의 결과를 가져 올 수 있다. 그렇지만 외부신축응답효과 편향이 보통 잊어버리는 방향과 반대 방향이다. 왜냐하면 보고된 사건 수가 실제 일어난 것 보다 더 클 경향이 있기 때문이다. 그 결과 외부신축응답효과의 결과 일어난 사건 총수의 추정값은 과대평가되는 경향이 있을 수 있다. 응답자에게 오래 지속되는 인상을 남길 수 있는 폭력적인 범죄 피해자이거나 큰 자동차사고를 목격하는 것 같은 매우 두드러지거나 감정적이거나 드문 사건에서 외부신축응답효과가 특히 문제가 된다. 응답자들은 이런 사건들은 실제로 일어난 것보다 더 최근에 일어난 것으로 기억할 수 있다. 특히, 드문 사건들에서는 응답자들이 면접원을 도움을 주기 위해 또는 “그 사건의 이야기를 하기 위해서는” 그 사건이 상세히 기록되어야 한다고 느끼기 때문에 기준기간 밖에서 일어난 사건을 기준기간 안에서

일어난 것으로 보고할 수 있다.

기준시간내에 있는 사건의 수를 세는 것의 대안으로 그 수를 추정할 수 있다. 예를 들어, 응답자에게 지난 30일 동안 승용차로 한 여행 횟수를 세어보라고 하기보다, 조사에서 지난 주 동안 또는 대표적인 한 주 동안의 여행 횟수를 세어보라고 해서 그 숫자를 대략 4로 곱해서 월간 추정값을 구한다. 사실 모든 사건을 기억하려고 하는 것보다는 이 방법이 인식하기 더 쉽기 때문에 세어야 할 사건이 많으면 응답자들은 세기보다는 이런 종류의 추정을 쓸 수도 있다. 만일 여행 횟수가 많으면 응답자들은 사건 횟수를 세려는 것이 헛된 시도라고 생각할 수 있고 추정을 더 정확한 답에 이르는 방법으로 쓸 수 있다.

사건을 세는 것은 잊어버리는 것과 신축응답효과 때문에 흔히 계통오차의 결과를 가져오는 데에 반해서 추정값은 어떤 응답자에 대해서는 실제보다 높고 다른 응답자에 대해서는 실제보다 낮을 수 있기 때문에 사건 횟수를 추정하는 것은 흔히 가변오차를 낳는다. 그러므로 추정을 하면 기준기간안의 사건 수를 과소평가하는 응답자들이 다른 응답자들의 과대평가 한 것을 상쇄하는 경향이 있다. 응답자의 표본에 걸쳐 추정값의 평균이나 전체는 편향이 조금 있거나 없을 수 있다. 그렇지만 추정과정의 오차의 결과 실제 센 것 보다 응답자에게서 얻은 추정해서 센 것이 더 변동이 많은 경향이 있다. 평균, 총계와 비율의 추정값에 대한 편향보다 가변오차가 덜 해롭기 때문에(2장 참조) 세는 데에서 생기는 편향이 클 것으로 예상되면 추정이 세는 것보다 더 좋을 수 있다.

만약 사건의 빈도가 작은 시간 간격에 걸쳐 어느 정도 고르지 않으면 짧은 시간 간격의 추정값을 더 긴 것으로 외삽법으로 산정하는 것의 추정오차는 상당히 심할 수 있다. 예를 들어, 모집단의 매달 피우는 담배의 수를 매일 또는 매주의 수를 외삽법으로 산정하여 추정하는 것은 상당히 정확할 수 있다. 그렇지만 만약, 어떤 음식섭취가 계절적인 경향이 있으면, 모집단에서 특정한 종류의 음식을 먹은 일년간의 빈도를 그 전달에 먹은 음식으로 추정하는 것은 상당히 가변성이 많을 수 있다. 또한 한달보다 드물게 일어나는 사건에 대해서는 일 개월 보다 더 긴 기억이 필요할 수 있다. 예를 들어 연간 여행 수를 추정하는 것을 주어진 달 동안에 민간항공기를 타고 한 여행 수를 외삽법으로 산정하는 것은 상당한 가변오차를 가지는 추정값을 가져올 것이다.

응답자들은 세거나 추정하기 보다는 간단히 사건수를 어림잡거나 대략의 추정값

을 제공하기로 결정할 수 있다. 이런 종류의 행동은 작은 성과로 만족하는 또 하나의 형태다. 응답자들이 답이 정확한 것에 가까울 수 있지만 더 많은 인지 노력으로 그들은 한 층 더 정확한 응답을 제공할 수 있다. 전에 언급했듯이, 응답자들이 정확한 응답을 제공할 의욕이 없거나 조사요구가 과중한 부담일 때 작은 성과로 만족하는 일이 발생한다. 추정처럼, 어렵잡는 것은 추정값에서 더 늘어나는 양의 가변오차로 이끌 수 있다. 그렇지만 추정할 때 보다는 어렵잡을 때 변수가 훨씬 더 클 것 같다.

마지막으로 정보기억 단계에서 특히 사업체조사에서 꽤 흔한 또 하나의 문제는 오래되었거나 정확하지 않은 기록의 사용이다. 예를 들어, 조사에서 응답자에게 회사급료 지불명부(종업원 명부)에 있는 현재 고용인 수를 달라고 하고 응답자는 몇 개월 된 정보를 제공할 수 있다. 그 결과로서 응답자가 잘못된 숫자를 보고한다. 다음에는 기억과정에서의 실제 오차의 예를 몇 개 제공한다.

예 4.3.3 첫 번째 예는 미국 센서스국에서 실시하는 모든 소매사업체에 대한 센서스인 Census of Retail Trade이다. 1977년 센서스에 대하여 센서스국이 센서스 결과의 품질을 평가하기 위해 재면접 연구를 실시했다. 이 연구에서는 센서스에서 전문직원들이 센서스 오차의 평가에 도움이 될 정보를 사업체들의 표본에게서 얻기 위해 그들을 재방문했다. 가능한 경우에는 재면접원들이 기록에서 정보의 검색을 필요로 하는 질문 항목에 대하여 응답자들에게서 “장부가격”을 얻기 위하여 사업체의 파일을 확인하라고 부탁했다. 이 연구에서 한 가지 발견한 것은 보고된 고용인 숫자에 상당한 측정오차가 있었다는 것이다. 추가 분석에서는 조사보고에 있는 약 75%의 오차는 응답자들이 정확한 값을 얻기 위하여 그들의 기록을 들춰 보기보다 오히려 고용인 수를 추정하거나 어렵잡는 응답자들 때문이다. 아마 정확한 숫자를 얻기 위하여 회사의 기록을 확인하는 것에 대한 부담은 응답자가 자진해서 떠맡을 수 있는 부분보다 더 컸다. 그 결과 “충분히 근접한” 숫자를 제공하는 정도에서 만족했다.

예 4.3.4 검색오차의 두 번째 예는 NCVS(US National Crime Victimization Study)를 위해서 실시한 평가 연구가 제공한다. NCVS는 미국 센서스국이 US Bureau of Justice Statistics를 위하여 실시하는 정기적인 연구이다. 조사 설계는 응

답자들을 6개월 간격으로 면접하는 매달 순환하는 패널 조사다. 즉 매달 새 가구 표본이 조사에 추가되어 면접을 하게 된다. 6개월 전에 면접을 한 가구들도 다시 면접을 한다. 한 가구가 6개월 간격을 두고 7번 면접하면 그 가구는 조사에서 “은퇴 한다.” 즉 더 이상 면접을 안 한다는 의미이다. 이 설계의 표로 된 설명은 <표 4.2>에 있다.

<표 4.2> 전형적인 한 달의 NCVS^a의 순환하는 패널 설계

표본요소	면접한 달		
	6개월 전	대표적인 달(M월)	6개월 후
부표본 1	아직 완성 안됨	면접 1 (경계)	면접 2
부표본 2	면접 1 (경계)	면접 2	면접 3
부표본 3	면접 2	면접 3	면접 4
부표본 4	면접 3	면접 4	면접 5
부표본 5	면접 4	면접 5	면접 6
부표본 6	면접 5	면접 6	면접 7
부표본 7	면접 6	면접 7	더 이상 면접하지 않음

^aNCVS의 설계에서는 매달 7개의 독립적으로 선택한 표본들을 (첫 열에서 보여줌) 면접한다. 각 표본은 전에 다른 횟수의 면접을 했다. 예를 들어 표에서 M이라고 표시한 한 대표적인 달에 부표본 1이 그것의 경계면접(첫면접)을 위하여 면접을 하고 부표본 2는 그것의 2번째 면접을 하고 부표본 3은 그것의 3번째 면접을 한다. 각 표본은 6개월 간격을 두고 경계면접을 포함하여 총 7번 면접한다. 예를 들어 부표본 2는 M월 6개월 전에 경계면접을 위해 소개되었다. M월에 2번째로 면접을 했고, M월 6개월 후에 3번째로 면접을 했고 등이다. 부표본 7은 M월에 마지막으로 면접을 하는 것을 주목해라. 이 패턴은 매달 반복된다. 각면접에서 지난 6개월 동안 일어난 범죄와 희생(victimizations)에 대한 질문들을 묻는다.

각 면접에서 응답자들에게 지난 6개월 기간동안 그들이 목격하거나 경험한(피해자로) 폭행, 절도, 강도, 자동차 절도 등과 같은 범죄 행위와 관련된 사건들을 기억하라고 부탁한다. 그 자료가 US Bureau of Justice Statistics에서 발행한 범죄희생보고의 기초가 된다. 기억기간의 길이 때문에 망각, 신속응답효과와 다른 기억력 오차가 문제가 될 수 있고 따라서 범죄희생보고가 왜곡되고 추정된 희생률에 계통오차와 가변오차가 생긴다. 절도, 강도와 다른 종류의 범죄의 빈도가 높은 심각한 범

죄 지역에서는 언제 범죄가 일어났는지를 정확히 기억하는 것이 상당히 어려울 수 있다. 줌도둑과 사소한 폭행 같은 범죄도 만일 그것이 비교적 드물다면 기억하기 어려울 수 있다.

조사에서 외부신축응답효과의 거의 제거하기 위하여 NCVS는 7번의 연속면접에서 첫 번째 면접을 경계면접(*bounding interview*)으로 사용한다. 즉, 첫 번째 면접은 두 번째 면접을 위한 기억기간의 시작을 정하기 위하여 사용된다. 첫 번째 면접에서는 응답자에게 지난 6개월간 일어났던 희생에 대하여 묻는다. 그렇지만 외부신축응답효과 때문에 7개월이나 더 이전에 일어났던 희생도 보고 될 수 있다.

첫번째 NCVS면접을 근거로 한 희생 추정값은 상당히 위로 편향된 것으로 알려졌기 때문에 오래전부터 센서스국은 첫 번째 면접을 근거로 한 희생 자료를 희생률을 추정하는 목적으로는 사용할 수 없다고 결정했다. 대신, 예를 들어 첫 번째와 두 번째 면접에서 보고된 희생을 대응비교하고 두 번째에 있는 범죄 중에 전에 보고된 범죄의 제거를 통하여 첫 번째 면접에서 보고한 희생을 두 번째 면접에서의 신축응답효과를 제거하는데 사용할 수 있다. 마찬가지로 두 번째 면접은 세 번째의 경계면접의 역할을 할 수 있다. 이와 같은 방법으로 남은 모든 면접에서 그렇게 할 수 있다. 두 번째에서 일곱 번째 면접은 자료면접(*data interviews*)이라고 부르는데 경계면접과 다르게 이 면접에서 나온 자료는 국가 범죄 희생률의 추정에 사용된다. 그러므로 두 번째 면접이 사실은 첫 번째 자료면접이고 세 번째 면접에 두 번째 자료면접이 된다.

경계면접이 신축응답효과를 줄일 수 있지만 희생보고에서 방각의 가능성을 검토하지 않는다. 잊어버리는 것을 줄일 수 있는 한 방법은 NCVS 면접을 더 잦은 간격으로 실시해서 기준기간을 단축하는 것이다. 1980년대 초반에 센서스국이 NCVS에서 6개월의 기억기간을 3개월로 하는 것의 자료 품질에 미치는 영향을 평가하기 위해 연구를 실시했다. 이 연구에서는 NCVS표본의 일부는 3개월마다 면접을 했고 나머지는 6개월마다 면접을 하는 분할표본설계(*splits sample design*)를 사용했다. 따라서 연속된 두 개의 3개월 기억 면접에서 보고된 범죄를 합하여 하나의 6개월 기억기간에 보고된 범죄 수에 직접 비교할 수 있는 추정값을 생각해 낼 수 있다. 예를 들어 3개월 기억 설계에서 1월-3월과 4월-6월에 기록된 총 범죄수와 6개월 기억 설계에서 1월에서 6월까지 보고된 총 수와 같아야 한다.

두 설계의 희생률은 <표 4.3>에서 비교할 수 있다. 별표가 있는 것은 그 차이가

통계적으로 유의하다는 것(*statistically significant*)을 나타낸다. 즉 차이의 규모가 우연히 일어날 것으로 합리적으로 기대되는 것보다 더 크다. 표는 3개월 기억이 6개월 기억 기간보다 항상 주목될 만큼 더 높은 보고를 제공한다는 것을 분명하게 보여준다. 만일 추정값의 유일한 오차가 잊어버리는 것이었다면 3개월 기억 기간이 더 높은 추정값을 가져온다는 사실은 3개월 기억오차가 잊어버리는 오차를 덜 내고 그러므로 적게 편향 되었다는 것을 암시한다. 그렇지만 경계면접 설계에도 불구하고 자료의 작은 양의 외부 신축응답효과오차가 여전히 있을 수 있다. 그러므로 신축응답효과 오차 때문에 6개월과 3개월 기억 추정값이 둘다 다소 양으로 편향되었을 수 있다. 그렇지만 왜 3개월 기억 설계가 6개월 기억 설계보다 신축응답효과 편향이 더 커야하는 지 그럴듯한 이유는 없는 것 같다. 그러므로 3개월 기억 추정값이 일반적으로 더 크고 그래서 6개월 기억을 근거로 한 추정값보다 덜 편향되었다.

<표 4.3> NCVS의 3개월과 6개월 기억의 비교 (12세 이상 100명당)

범죄종류	6개월 기억	3개월 기억	차이
총 개인 범죄	12.8	15.5	-2.7*
폭력범죄	3.5	4.3	-0.8*
도둑질 범죄	9.4	11.2	-1.8*
총가구범죄	23.0	26.4	-3.4*
강도	8.5	9.7	-1.2*
절도죄	12.7	15.1	-2.4*
자동차 도둑질	1.8	2.1	-0.3

출처: Bushery(1981)

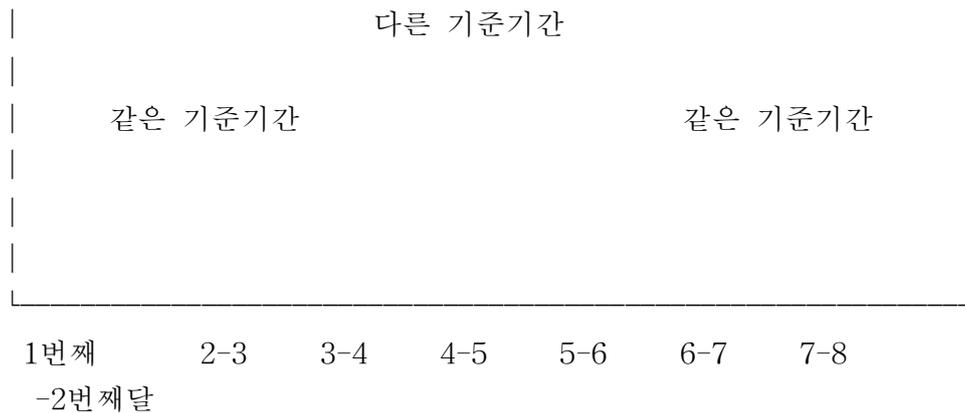
*5% 유의 수준에서 통계적으로 유의한 차이

3개월 기억이 범죄 희생 추정값에서는 기억편향에 생기기 쉽기는 하나 고정 비용 설계에서는 평균제곱오차 전체를 고려했을 때 6개월 기억을 근거로 한 추정값이 여전히 더 정확할 수 있다. 3개월 설계에서는 응답자들과 더 자주 면접을 해야 하기 때문에 6개월 설계와 같은 조사비용을 유지하기 위해 3개월 설계의 표본 크기는 반드시 더 작아야 한다. 그러므로 측정오차는 줄었지만, 3개월 설계의 표본 분산은 6개월 설계의 표본 분산의 두 배까지 될 수 있다. 따라서 편향제곱의 합 더하기 분산인 총 평균제곱오차가 3개월 기억을 사용하면 실제로 더 클 수 있다.

3개월 기억 기간으로 바꾸는 것을 결정하는 데 비용 외에 다른 고려해야 할 사항

들이 있다. 예를 들어 설사 추정값의 전체 평균제곱오차를 줄다 할지라도 만일 희생물의 표준오차를 극적으로 늘어나면 이용자가 어떻게 반응 하겠습니까? 희생물의 편향은 보고 되지 않기 때문에(전에 설명했듯이 특수한 평가 연구 없이 기억 편향은 추정할 수 없다), 편향의 어떤 감소도 자료 이용자의 눈에 띄지 않고 가치를 인정받지 못할 수 있다. 사실 자료 이용자에게는 추정값의 표준오차가 늘어나면서 자료 품질이 실제로 나빠지는 것처럼 보일 수 있다. 이런 고려해야 할 사항들이 기억 편향을 줄이는 데 3개월 기억의 이점에도 불구하고 센서스국이 NCVS에서 6개월 기억 설계를 유지하게 하였다.

SIPP 고용 상태를 총계의 퍼센트로 나타낸 월별 변화



<그림 4.4> 접합효과. 근접한 달 사이에 고용 상태를 바꾼 사람들의 퍼센트는 10.2%로 뛰는 4번째와 5번째 달을 제외하고 모집단의 4%와 5%사이이다. 이유는 접합 효과이다. 첫 번째에서 4번째 달의 고용 상태는 한번의 면접에서 얻었는데 5번째와 8번째 달의 상태들은 4개월 후에 다른 한 면접에서 얻었다. 그러므로 4번째와 5번째 달은 대략 4개월 떨어진 두 다른 어떤 시점에서 수집되었다.

패널 조사에서 자주 부딪히는 또 한 가지 문제는 두 개의 다른 기준기간에 대해 양쪽으로 걸치는 근접한 달 보다 같은 기준기간 안에 있는 근접한 달들 사이에 더 많은 월별 변화가 관측되는 현상인 접합효과(seam effect)이다. <그림 4.4>는 미국에서 실시한 Survey of Income and Program Participation(SIPP)에서 측정한 결과를 설명한다. 이 결과의 가장 뚜렷한 이유는 기억오차다. 응답자들은 그들의 고용상

태를 면접 4개월 전 달에 보다 면접 전달에 더 기억할 것이다. 그 경우 접합(즉, 그림에서 4번째와 5번째 달) 사이의 변화는 잘못됐고 실제 변화는 접합이 아닌 다른 달들에서처럼 4%와 5%사이일 것이다.

그렇지만 기준기간 안에서 월별 변화는 줄여 말했다는 것이 또 다른 설명이 될 수 있다. 즉, 응답자들이 4개월 기준기간 전체 동안에 그들의 고용 상태에 변화가 없었다고 최소한의 필요조건을 추구하는 것의 형태로 응답할 수 있다. 예를 들어, 그들은 그 변화에 대해 부가적인 질문을 피하고 싶을 수도 있다. 접합효과의 가장 그럴듯한 설명은 기준기간 안의 변화를 줄이고 기준기간 사이의 변화를 늘리도록 행동하는 요인들의 결합 때문이라는 것이다.

응답의 구성

정보검색 다음에 일어나는 응답과정의 단계는 판단과 응답을 구성하는 (*judgement and response formatting*) 단계이다. 응답과정의 바로 이 단계에서 전단계에서 검색했던 정보를 평가하고 질문에서 요구한 서식에 따라 응답을 구체화하는 것이다. 흔히 조사질문들은 응답자가 응답범주목록에서 답을 선택하는 것을 요구하는 선택형질문이다. 응답자를 위해 이미 적당한 답이 제공 됐기 때문에 선택형질문들은 흔히 면접에서 시간을 절약하고 응답자 부담을 줄인다.

서술형 질문들은 응답자에게 답을 자기의 말로 표현하고 면접원이 그 답을 기록하는 것이다. 언어 실력이 약한 응답자들이 서술형질문에 대해 유용한 답을 제공하지 못할 것이 걱정 될 때에는 선택형 질문이 사용될 수도 있다. 그렇지만 “지난 한 해 동안 의사를 몇 번 찾았습니까?”라는 질문과 같이 질문이 요구하는 형태가 뚜렷할 때에는 서술형질문이 더 선호된다.

한 예로 Sudman 등(1996)은 행동 빈도를 얻을 때에는 서술형 질문 형태를 사용할 것을 권장했다. 뒤에 나오는 예 4.3.5에서 보겠지만 응답자들은 때로는 응답범주의 숫자 범위를 그들 자신의 행동 빈도를 추정하는데 참고 틀로 사용하고 이것은 계통오차의 결과를 가져올 수 있다. 응답은 숫자로 표시됐기 때문에 만일 원하면, 이런 답을 컴퓨터로 코딩하는 데에는 특별한 어려움이 없다. 그렇지만 앞에서 서로 다른 기억과 추정 전략에 대한 검토에서 언급했듯이 서술형형태의 질문들은 응답과정에서 다른 오차의 원인에 의해 여전히 편향될 수 있다.

서술형과 선택형질문의 절충으로 선택형 질문에서 “그 밖”이라는 응답범주를 사

용하여 응답범주들이 충분하지 않을 때 응답자가 자진하여 답하는 것을 허용한다. 이 접근 방법은 모든 가능한 응답을 포함하기 위해 어떤 응답범주를 제공할지 뚜렷하지 않은 질문들에서 추천된다. 또한 그것은 주어진 응답범주 중에서 선택하지 않을 비교적 작은 소수의 응답자가 제공할 수 있는 미지의 가능한 응답 수를 포함하기 위해 포괄적인 항목으로 사용될 수 있다.

그러므로 선택형 질문에서는 응답을 구성하는 단계는 응답자가 전 단계에서 얻은 정보를 제공된 응답범주에 따라서 구성하는 것이다. 서술형 질문에서는 그 질문에 대처하는 응답을 어떻게 구성하느냐를 결정한다.

예를 들어 질문이 “당신 회사의 컴퓨터 지원 서비스는 집중식입니까 아니면 비집중식입니까?”라고 물을 수 있다. 응답을 구성하는 단계는 이 두 가지 응답 선택 중 어느 것이 회사의 컴퓨터 지원 서비스에 가장 잘 맞는지를 결정하는 과정을 의미한다. “당신은 몇 살입니까?” 또는 “당신의 소득은 얼마입니까?”처럼 질문이 만일 서술형이면, 응답자는 그 정보를 얼마나 정확하게 보고할까를 결정해야 한다. 예를 들어, 상당히 정확한 숫자 또는 대략 또는 반올림 된 숫자를 제공할까를 결정해야 한다.

응답을 구성할(formatting a response)때에는 정보를 평가하고 질문에서 요구한 서식에 따라 응답을 구성한다.

응답과정의 이 단계에서 몇 가지 오차가 발생할 수 있다. 선택형 질문에서 응답자가 그의 응답으로 전달하고 싶은 것을 응답범주가 충분히 묘사하지 못할 때 흔히 오차가 발생한다. 예를 들어 바로 전 예에서 응답범주가 집중식 서비스와 비집중식 서비스 두 가지밖에 없다고 가정하자. 그렇지만 응답하는 회사에서 인터넷 지원과 e-메일 같은 어떤 서비스는 집중식이지만 PC지원과 기술적인 소프트웨어 지원 같은 다른 서비스는 비집중식이었다. 이 경우 올바른 응답은 집중식과 비집중식 컴퓨터 지원의 혼합된 것이다. 그러므로 어느 응답범주도 적당하지 않기 때문에 응답자가 어떻게 답을 할지라도 그의 응답은 현실을 정확하게 설명할 수 없다.

너무 제한적일 뿐 아니라 응답범주들이 때로는 전형적인 응답이나 질문에 대한 다른 해석을 하게 할 수 있다. 예를 들어 “당신의 회사는 얼마나 자주 전략계획을

합니까?” 라는 질문은 만일 응답범주가 “절대 안함”, “매년”, “2-3년에 한번”, “4년마다 또는 그 이상” 이면 만일 응답범주가 “절대 안함”, “매달”, “ 일년에 여러 번”, “일년에 한번이나 매년”, “매년보다 덜 자주” 일 때와는 다른 해석을 취할 수 있다. 앞의 경우에는 응답자가 전략계획을 많은 부서들이 포함되는 회사 전체의 대규모 계획을 의미하는 것으로 해석할 수 있다. 뒤의 경우는 응답범주가 관심 대상의 전략계획이 더 적은 수의 사람들이 해당되는, 따라서 더 자주 접하게 되는 소규모의 계획이라고 생각할 수 있다. 또 하나의 예를 뒤에 설명한다.

세 번째 문제는 즉석응답(*top-of-the-head responses*)이라고 하는 응답을 구성하는 단계에서 깊이 생각하지 않고 빠른 응답을 하도록 응답자들이 재촉 받고 강요당할 때 발생한다. 이 오차는 어떤 방식에서도 발생할 수 있지만 우편이나 대면 조사보다는 전화조사에서 더 자주 발생하는 경향이 있다. 전화조사에서는 대화 중에 이야기가 길게 중단될 때 거부할 수 있고 아마도 그 이유로 그들은 빨리 대답하는 압박을 느낄 수 있다. 대면조사에서는 대화가 중단된 동안 무슨 일이 일어나고 있는지에 대한 정보를 시각적인 의사전달로 제공할 수 있고 따라서 그 침묵을 채우는 것에 대한 응답자의 압박이 덜 하다.

대체로 전화조사가 다른 면접 방식보다 즉석응답의 결과를 가져올 경향이 더 있다는 것이 연구에서 입증되었다. 더구나 서술형 질문에 대한 대답은 본인이 직접 했을 때보다 전화로 했을 때 더 짧은 경향이 있다. 이것은 응답자들이 대면면접에서보다 전화면접에서 이야기를 덜 하는 경향이 있다는 것을 나타낸다.

응답자들은 항목목록에서 응답항목을 선택할 때, 특히 항목들이 정돈되지 않았거나 명목상의 목록이라면 간단히 필요조건만을 충족시키면서 답할 수 있다. 예를 들어 초등학교 교사들에 대한 자기기입식 조사에서 교사들에게 그들이 사용할 수 있는 10개의 교육 보조물의 목록을 주었다. 그들에게 초등학생들을 가르치는데 가장 유용한 보조물을 선택하라고 부탁했다. 그 보조물들은 비록 아무 특별한 순서 없이 기재되어 있었지만 그 목록의 위쪽 반 항목이 아래쪽 반에 있었던 항목들보다 거의 두 배나 자주 선택되었다. 이것은 응답자가 목록전체를 읽고 최고의 응답을 고르기보다는 일단 그들이 마음에 드는 응답을 찾으면 그들은 목록을 읽는 것을 중단했다는 것은 응답자들이 최소한의 필요조건만을 충족시켰다는 것을 암시할 수 있다.

최소한의 필요조건을 추구하는 것은 서술형 질문에서도 심각한 문제가 될 수 있다. 예를 들어, “당신은 직장에서 보통 어떤 종류의 일에 종사합니까?”라는 질문은

응답자를 힘들게 할 수 있다. 여러 가지 활동을 수반하는 직업을 가진 사람은 이것들을 생각해 내기 힘들 것이고 어느 것들이 전형적인지를 결정하기 위해 그것들을 구분하고 그 목록을 답할 것이다. 이렇게 제공된 정보는 연구 목적에 적당하지 않을 위험이 있다.

예 4.3.5 응답자 때로는 응답범주로부터 모집단에 대한 연구자의 인식이나 질문에서 기대되는 전형적인 응답 정보를 얻을 수 있다. 이러한 정보는 응답자가 응답을 하는데 사용 될 수 있다. 어떤 경우에는 응답자가 응답범주를 통해서 드러난 현실에 대한 연구자의 가정에 맞게 그의 응답을 수정할 수 있다.

예를 들어, 특정한 행동의 빈도를 평가하기 위한 응답범주에 대한 연구에서 응답자들은 응답범주들이 모집단 행동의 분포를 반영한다고 가정할 수 있다는 결과를 보였다. 구체적으로 말하면 중앙에 있는 항목들은 전형적인 행동을 나타내고, 맨 끝에 나오는 항목들은 드물거나 “보통과 다른” 행동을 나타낸다. 이런 가정은 응답에 다양한 방법으로 영향을 준다. 응답자들은 어떤 경우에는 응답범주의 범위를 그들의 자신의 행동의 빈도를 추정하는데 참고로 쓸 수 있다. 만일 그들의 행동이 전형적이라고 보면 그들은 진정한 행동의 빈도를 더 정확하게 검토 평가하려고 시도하지 않고 중간 항목 근처에 있는 항목을 고를 수도 있다.

이것의 설명을 위해 Schwarz 등(1985)의 TV 시청에 관한 연구의 결과가 <표 4.4>에 나온다. 이 연구에서는 응답자의 확률표본의 반에겐 표 왼쪽에 있는 낮은 빈도 항목들을 제시했고 다른 반에게는 오른쪽에 있는 높은 빈도 항목들을 제시했다. 표에서 보여주듯이 낮은 빈도 항목이 주어진 응답자의 16.2%가 2시간 반 이상 매일 시청하는 것으로 보고 했는데 높은 빈도 항목이 주어졌을 때는 37.5%가 그러했다. 즉, 높은 빈도 항목이 제시된 질문들은 같은 질문들을 낮은 빈도 항목들을 사용해서 얻는 추정값보다 두 배 이상인 추정값을 나타냈다. 이것에 대해 몇 가지 설명이 가능하다.

한 가지 설명은 많은 응답자들은 TV를 얼마나 자주 보는지를 생각하기 보다는 응답범주에 주어진 “전형적인” TV 시청 빈도에 대한 정보를 사용해서 빈도를 간단히 추정했다. 낮은 빈도 항목이 제시된 경우 그들은 전형적인 사람은 1-2시간 정도 (즉, 낮은 빈도 척도의 중간) TV를 본다고 가정했다. 높은 빈도 항목이 제시된 경우에는 전형적인 사람이 하루에 3-4시간 정도 본다고 가정했다. 어느 경우든 응답

은 그들의 TV 시청이 보통 사람의 TV 시청과 어떻게 비교되는지에 대한 그들의 인식을 나타낸다.

<표 4.4> 두 가지 응답 종류의 따른 TV 시청 질문에 대한 응답

낮은 빈도의 항목		높은 빈도의 항목	
응답범주	보고율(%)	응답범주	보고율(%)
30분까지	7.4	2시간반까지	62.5
30분-1시간	17.7	2시간반-3시간	23.4
1시간-1시간반	26.5	3시간-3시간반	7.8
1시간반-2시간	14.7	3시간반-4시간	4.7
2시간-2시간반	16.2	4시간-4시간반	0.0

출처: Schwarz et al.(1985) 의 자료.

이 결과에 대한 또 하나의 설명은 사회적 기대부응 편향의 존재다. 이 연구에서 표본은 과도한 TV 시청을 재미없는 사회생활을 하는 인기 없는 사람들의 특징으로 생각하는 대학생들로 구성되었다. 따라서 TV를 과도하게 보는 것은 사회적으로 바람직하지 않다. 사회적으로 바람직하지 않은 생활 방식이 드러나는 것을 피하기 위해 응답자들은 척도의 중간범위가 전형적인 것과 더 일치하고 모집단에서 더 사회적으로 받아들일 수 있는 행동이라는 전제하에서 척도의 중간범위에 있는 응답범주를 선택할 수 있다.

세 번째 설명은 응답자들이 질문 때문에 혼동됐다는 것이다. 만일 질문이 “당신은 TV를 하루 평균 몇 시간 봅니까?”하고 묻는다면 응답자들은 “TV를 보다”라는 용어를 주어진 응답범주에 따라 다르게 해석할 수 있다. 높은 빈도 항목이 주어질 때에는 그들이 TV를 얼마나 주의 깊게 시청하는지와 관계없이 TV가 켜 있는 동안 TV와 같은 방에 있는 것을 연구자가 의미하는 것으로 응답자들이 가정한다. 낮은 빈도항목들이 주어지면 응답자들은 “TV를 보다”라는 말을 적극적인 TV 시청을 의미하는 것으로 해석할 수 있다.

어떤 설명이 올바른지와 상관없이 한 질문에 주어진 응답범주들은 그것들이 특정 행동에 대하여 시사하는 정보에 의해 질문의 응답에 큰 영향을 미칠 수 있다는 것을 이 예에서 명백히 볼 수 있다.

교정과 전달 (Editing and Communication)

응답과정의 마지막 단계는 교정과 전달이다. 이전 단계들에서 응답자는 질문을 이해하고, 질문에 답하는데 필요한 정보를 검색하고, 질문에 대한 그의 응답을 가장 잘 설명하는 응답항목의 값이나 대답을 결정했다. 이제 마지막 단계에서 응답자는 그의 응답을 수정할 것인지 아닌지를 결정한다. 즉, 가장 정확한 응답을 제공할지 또는 사회적 기대부응이나 공개의 두려움 때문에 바꾼 것을 제공할지를 결정하고 응답을 면접원에게 전달하거나 적당한 응답항목을 선택한다. 이 단계에서 여러 종류의 오차가 발생할 수 있다. 그것들은 사회적 기대부응 오차, 공개의 두려움 오차와 묵인인데 이것들은 아래에서 설명한다.

마지막 단계인 교정과 전달(editing and communication)에서는 응답을 그것이 나타난 그대로 또는 응답자가 교정을 한 후에 연구자에게 전달한다.

위에서 설명했듯이 사회적 기대부응 오차는 응답자가 그의 응답이 사회적으로 받아들일 수 없다고 결정할 때 그것을 사회적으로 받아들일 수 있는 것으로 바꾸는 것을 뜻한다. 예를 들어 술을 과도하게 마시는 응답자들은 음주량에 대한 면접원의 반감을 피하기 위하여 면접원에게 그들의 음주량을 고의적으로 실제 이하로 보고할 수 있다. 이 결과, 자료에 계통오차가 발생하고, 모집단의 음주량이 실제 이하로 보고 된다. 사회적 기대부응 편향(*social desirability bias*)이라고 부르는 이 편향은 사회적으로 용납할 수 없는 성행위, 마약사용, 세금을 덜 내는 것과 응답자가 면접원에게 그들의 실제 활동을 들어내기를 꺼리는 불법 행위와 같은 민감한 자료수집에서 상당히 자주 발생한다.

면접원들이 사회적 기대부응 오차의 가장 중요한 촉매이기 때문에 이 오차는 보통 자기기입식 조사 보다 면접 조사에서 훨씬 더 크다. 따라서 조사에서 민감한 자료를 수집하는 경우 자기기입식 자료수집이 면접원에 의한 시행 방법보다 더 정확한 좋은 방법으로 여겨진다. 여러 문헌들에서 사회적 기대부응 편향을 받기 쉬운 자료를 수집하는데 전화조사가 대면면접 보다 더 좋다는 증거가 제시 되었다. 6장에서 보겠지만, 이것은 항상 사실인 것은 아니다.

노출 공포 오차는 응답자들이 정확한 조사 보고를 제공하는 것에 대한 결과를 두

려워해서 그들의 응답을 수정할 때 발생한다. 예를 들어, 사업체 응답자는 회사의 경쟁자들이 조사에서 요구되는 소유자정보에 어떻게 해서든지 접근할 것을 염려 할 수 있다. 고소득자인데 그의 소득세 용지를 속이는 사람은 조사에서 사실을 말하는 것은 조세당국과 문제를 일으킬 것을 두려워 할 수 있다. 그러므로 노출 공포 오차는 면접원의 유무에 의해 반드시 영향을 받는 것이 아니다. 오히려 제공된 정보가 익명으로 기밀하게 유지되지 않을 수 있는 것이 걱정의 원인이 된다.

사회적 기대부응 오차와 같이 노출 공포 오차도 자료에 계통오차를 가져오고 추정값에 편향을 일으킨다. 노출 공포 편향을 피하는 한 가지 방법은 그것이 사실이고 가능하다면 응답자에게 그들의 응답은 익명으로 기밀하게 유지될 것이라는 것을 응답자에게 확신시키는 것이다. 조사 응답이 응답자의 신원과 연결될 수 없도록 조사에서 각별한 예방조치를 취하는 것도 공개의 두려움의 편향을 피하는 방법이다. 그렇지만 경우에 따라서는 이런 수단들이 적당하지 않다. 한 예로, 농민들은 농업노동력 조사에서 적절한 이주서류가 없는 계절적인 현장노동자들을 보고하기를 주저할 수 있다. 그들이 개인 보고의 기밀성보장에 대한 신뢰는 있을지 모르지만 조사 결과로 인해 농민들이 정식서류를 갖추지 않은 근로자들을 고용하는 것이 증가한다는 것을 나타낼 것이고 이것을 방지하기 위한 당국의 조치가 늘어나게 할 수 있다. 그러므로 그들은 이 근로자들에 대한 정직한 공개는 결국은 그들이 농장 노동을 위해 지불할 비용을 증가시킬 것이다.

목인은 응답 수정과 전달 단계 중에 발생할 수 있는 잠재적인 문제다. 이 오차는 응답자가 정확하게 보고해야 하는 것 보다 조사설계자나 면접원이 응답자들이 보고하기를 원한다고 믿는 대로 보고할 때 발생한다. 한 예로, 고객만족조사에서는 실제보다 제품과 서비스에 대한 응답자의 의견이 더 긍정적인 평가를 제공하는 경향이 있다. 만족감을 나타내는 응답을 원한다는 것을 응답자는 잘 알고 있고 그래서 그 방향으로 목인하는 경향이 있다. 이런 종류의 편향을 피하기 위하여 조사설계자들은 표현과 말투가 중립적이고 긍정과 부정에 관해서는 균형 잡힌 고객만족 조사 설문지를 설계하려고 애써야한다. 더구나 응답자들은 면접원이나 조사의뢰자의 편견에 의해 영향을 받을 수 있으므로 만족도 조사들은 흔히 자기기입식과 중립적이고 공정하게 간주되는 조사 의뢰를 해야 한다.

예 4.3.6 US National Household Survey on Drug Abuse(NHSDA)는 사회적 기

대부응 편향과 노출 공포 편향의 위험을 최소화하기 위해 설계된 조사의 한 예이다. NHSDA는 모집단의 현재와 과거의 불법적으로 남용된 약물복용 위를 측정하기 위하여 설계된 가구조사이다. 목표모집단은 가구에 거주하는 12세 이상인 사람이다. 면접단계에서 면접원과 자기기입식 도구를 사용해서 각 응답자로부터 약물복용과 인구통계학적 자료를 수집했다. 면접을 완성하는데 평균적으로 1시간 정도 걸린다. 응답자의 현재와 과거의 담배와 다른 형태의 흡연에 대한 자료를 수집하기 위해 설계된 면접원 기입식 질문으로 면접을 시작한다. 이런 초기의 질문들은 응답자가 NHSDA의 형식에 익숙해지게 한다.

설문지 나머지는 관심대상의 약물인 술과 진정제, 정신 안정제, 흥분제, 진통제의 비의학적 사용, 그리고 대마초, 흡입제, 코카인, 정제 코카인, 환각제, 헤로인의 부분으로 나누어진다. 각 부분에서 면접원은 응답자에게 답안용지를 주고 그들의 응답을 기록할 것을 부탁한다. 답안용지의 복잡함에 따라 면접원이 응답자에게 질문을 읽어주던지 만일 응답자가 원하면 그들이 질문을 읽을 수 있다. 응답자가 답안용지를 완성하면 면접원이 응답을 보지 못하게 하면서 답안용지를 봉투안에 넣을 것을 요구한다. 이 방식으로 면접을 실시하는 것에 대한 동기부여는 응답자가 질문들을 이해하고 설문지의 주요부분을 실수로 건너뛰지 않을 것을 책임지는 것과 더 중요한 것은 응답에 대한 기밀성을 보증하는 것이다.

대부분의 응답용지는 한번도 특정한 약물을 복용해 본적이 없는 응답자조차도 약물에 대한 각 질문에 답을 하도록 설계했다. 약물 사용자와 비 사용자가 근본적으로 같은 수의 질문에 응답할 것을 요구하기 때문에 면접원은 응답자가 답안용지를 완성하는데 걸리는 시간을 바탕으로 사용자인지 비사용자인지를 추측하기 힘들다. 이것은 응답자의 사생활을 보호하기 위해 설계된 조사의 또 하나의 특징이다. 직접적인 질문을 하면 한번도 약물을 사용해 본적이 없다고 표시한 응답자들이 나중에는 그 약물에 대한 간접적인 질문에 그 약물의 사용을 나타내는 대답을 한다. 중복 질문은 직접적인 질문에서 실제 이하로 보고하는 것의 보상으로 사용될 수 있는 약물 사용에 대한 추가적인 정보를 제공한다.

예 4.3.7 <표 4.5>는 조사에 포함될 수 있는 다양한 주제에 대한 노출 공포의 위험이나 사회적 기대부응 편향을 설명한다. Bradburn 등(1979)은 응답자들이 민감하게 느끼고 조사 주제로는 너무 사적일 수 있는 주제를 확인하기 위하여 연구를 실

시하였다. 이런 주제에는 약물, 알코올 섭취, 소득, 성행위, 도박, 음주, 스포츠 같은 것 들이 있다. 그들은 응답자에게 그 주제들을 그것들이 “대부분 사람들”을 얼마나 거부하게 하나에 따라서 매우 거부, 꽤 거부함, 약간 거부함, 전혀 거부하지 않음의 4점 척도로 평가하게 했다. 응답자에게 주어진 항목의 목록과 함께 조사에서 그 주제들에 대하여 이야기하기를 “매우 거부함”이라고 표시한 응답자의 퍼센트가 <표 4.5>에 제공되었다. 표에서 보듯이 성행위와 약물은 목록의 아래에 있고 전자는 흔히 창피하고 어떤 경우에는 사회적으로 용납되지 않고 후자는 불법이기 때문이라고 이해가 된다. 스포츠와 여가 활동은 실제로는 응답자들이 이야기하기를 즐기는 주제들이다.

<표 4.5> 조사에서 주제에 대하여 이야기하는 것을 거부하게 느끼는 사람의 비율

주제	대부분 사람들을 거부하게 만듦(R의 수준)
스포츠 활동	1
여가시간과 일반활동	2
사회활동	2
직업	3
교육	3
행복과 복지	4
맥주, 와인이나 독한 술	10
친구들과 도박	10
소득	12
애무와 키스	20
취하는 것	29
홍분제나 진정제의 사용	31
대마초나 인도대마의 사용	42
성행위	42
자위행위	56

출처: Bradburn et. al(1979) 참조

예 4.3.8 확률화응답기법(randomized response technique)이라고 부르는 사회적 기대부응 편향과 공개의 두려움편향 모두를 극복하기 위한 기법의 한 예로 이 장을 마무리한다. 이 방법의 한 예는 응답자에게 “당신은 1월에 태어났습니까?”와 “작년

과세과정에서 당신의 모든 소득을 보고 했습니까?” 같은 두 가지 질문을 한다. 한 질문은 민감하지 않고 다른 것은 민감할 수 있다는 것에 주목하라. 만일 두 질문의 답이 같으면(둘 다 “예” 또는 “아니오”) “그들은 같다”로 답하도록 했다. 그렇지 않다면 응답자는 “그들은 다르다”라고 답하도록 했다. 만일 모집단에서 1월에 태어나는 확률을 알 수 있으면 (대부분의 경우 센서스자료나 다른 모집단 기록에서 확인할 수 있다) 탈세를 혁신적인 통계연구법을 사용해서 추정할 수 있다. 이 방법으로 응답자들은 탈세에 관련된 직접적인 질문에 대한 그들의 정직한 응답을 드러내는 것을 피할 수 있다.

Warner(1965)가 처음 출판한 확률화응답기법은 초기에는 민감한 자료조사 수집에 획기적 발전이라고 여겼다. Danermark와 Swensson(1987)은 이 방법의 변형을 학교에서의 약물사용을 추정하는데 성공적으로 적용한 예를 제시하였고 다수의 다른 적용을 제시한 문헌들도 있지만 이 방법은 1960년대 후반에 예상했던 것처럼 실제적인 조사작업에서 사용되지는 못했다. 예를 들어 응답자들은 항상 그들의 응답이 정말로 보호될 것이라는 것을 알지는 못했다. 더욱이, 때때로 사용된 확률화장치(소형 롤렛 바퀴, 카드 등)들은 다양한 적용에서 표본구성원들에 의해 조사연구의 일부로 인식되지 못했다. 또한, 단지 몇 개의 조사 질문을 위하여 확률화응답과정을 시행하는 것은 혼란을 일으키고 비현실적일 수 있다.

설문지 설계문제와 이 장에서 설명한 많은 원리를 실행하는 설문지 설계 방법들의 포괄적인 취급을 위하여 Converse와 Presser(1986), Sudman과 Bradburn(1982), Bradburn 등(1979), Dillman(2000), Schwarz와 Sudman(1996), 그리고 Tanur(1982)를 추천한다. 조사에서의 측정오차를 광범위하게 다루는 저서로는 Biemer 등(1991), Groves(1989), Groves 등(1988), Lyberg 등(1977), Rossi 등(1983), 그리고 Turner와 Martin(1984)을 추천한다.

제 5 장 면접원과 면접에 의한 오류

조사설문면접원이 응답자에게 조사질문들을 묻고 응답자의 응답을 설문지에 기입하는 면접원기입식 또는 응답자가 질문을 읽고 응답자가 그의 응답을 면접원의 도움 없이 설문지에 직접 기입하는 자기기입식으로 설계할 수 있다. 이러한 두 종류의 조사운영이 결합된 방식도 드물지 않다. 예를 들어, 미국 NHSDA(National Household Survey on Drug Abuse)에서 면접원들이 면접의 일부를 처리하고 면접의 나머지는 약물사용에 대한 매우 민감한 정보를 수집을 필요로 하기 때문에 자기기입식이다. 6장에서는 조사질문을 응답자에게 전달하는 다양한 수단을 이용하는 몇 가지 자료수집 방법이나 방식에 대하여 이야기하고 각각의 장단점을 다룬다. 이 장에서는 응답자와 대면으로 또는 전화로 이야기를 나누는 조사면접원이 처리하는 조사에 초점을 맞춘다. 조사에서 면접원의 역할, 그 역할을 수행하면서 그가 범할 수 있는 오류들, 오류의 정도에 영향을 미칠 수 있는 요인들, 면접원 오류들을 평가하고 통제하기 위한 방법들에 대하여 이야기한다.

면접원들에 대하여 물어볼 수 있는 첫 번째 질문은 조사자료를 수집하는데 그들이 필요한 이유이다. 사실은 많은 상황에서 면접원이 필요하지 않다. 예를 들어, 우편, 인터넷, 전자메일 같은 자료수집 방식에서는 설문지를 응답자에게 보내고, 그들은 면접원의 도움 없이 설문지를 완성한다. 면접원들이 응답자들을 방문하는 것을 요구하는 조사에서도 설문지 내용의 민감성 때문에 조사에서 면접원의 역할을 되도록 적게 할 수 있다. 예를 들어, 1990년대의 NHSDA에서는 면접의 역할은 기본적으로 조사설문지를 전달하고, 응답자가 그것을 자기기입식으로 완성할 때까지 기다리고, 완성된 설문지를 처리와 분석을 하도록 연구기관에 전달하는 것이다. 컴퓨터 기술이 향상되면서 컴퓨터소프트웨어가 만들어낸 “가상(virtual)”면접원이 응답자와 사람이 접촉할 필요 없이 면접을 실시할 수 있다. 면접원을 이용할 것인가에 대한 질문은 2장과 10장에서 설명하듯이 설계과정의 초기단계에서 검토해야 할 것이다.

하지만 상당수의 자료수집에서 면접원은 조사과정에서 없어서는 안되고 매우 중요한 역할을 한다. 사실 면접원은 단순히 응답자를 면접하는 일 보다 훨씬 더 많은 것을 한다. 예를 들어, 면접원들은 지역의 가구단위들의 명부를 만들어서 그것으로부터 가구단위의 표본을 뽑을 수 있도록 표본추출과정을 도울 수 있다. 많은 대면 조사에서는 표본을 뽑을 추출틀에 이전 주소가 많이 있을 수 있기 때문에 면접원의

중요한 역할은 표본구성원을 찾는 것이다. 많은 유동성이 있는 모집단에서는 표본구성원의 현주소를 확정하는 것이 상당히 어려울 수 있다.

조사면접원들의 가장 중요한 의무 중 하나는 표본구성원과 연락을 해서 그들이 조사에 참여하도록 설득하는 것이다. 가족의 협조를 얻은 후에는 가족의 누가 조사에 책임인가를 결정하기 위하여 짧은 적격심사면접을 실시할 필요가 있을 것이다. 면접이 시작되면 면접원은 응답자에게 질문을 읽어주고, 필요에 따라 질문의 의미를 해석하거나 명백히 설명해주고, 응답이 애매하거나 명확하지 않을 때는 탐색조사 질문들을 하고 응답들을 도구에 기록할 수 있다. 또한, 면접원은 응답자, 가족, 또는 방문한 이웃들에 대하여 관측을 하고 기록할 수 있다. 만일 면접원의 이런 기능들이 조사목적을 달성하는데 중요하면, 자기기입식 방식을 배제하고 면접원 지원 조사 계획을 세워야한다.

응답자들을 면접하기 위하여 면접원들이 사용하는 스타일이나 예절은 문헌에서 논쟁의 주제가 되어왔다. 전통적으로 면접원은 연구자의 중립적인 대리인으로 교육받고 응답자들에게 질문들을 매우 표준화된 방법으로 제시해왔다(Fowler와 Mangione, 1990). 하지만 Suchman과 Jordan(1990)과 그 이후의 연구자들은 서로 더욱 상호적이거나 “대화적인” 접근 방법을 지지하며 표준화된 접근 방법에 이의를 제기해왔다. 전자의 접근 방법은 가장 정확한 응답들을 얻기 위하여 필요에 따라 면접원에게 응답자와 이야기를 나누는 것을 훨씬 더 많이 허용한다. 다음절에는 이 두가지 접근 방법들에 대하여 생각하고 각각의 사용에 대하여 지침을 제시한다.

5.1 면접원의 역할

Suchman과 Jordan의 논문이 1990년에 발표된 이후 조사방법론자들은 조사 면접에서 면접원의 역할에 대한 토론을 해왔다. 논쟁의 한쪽에는 실험자가 반응-자극 실험에서 처리를 표준화시키듯이 면접원과 응답자사이의 상호작용의 표준화를 시도하는 표준화 면접(*standardized interviewing*)방법이 있다. 다른 쪽에는 응답자에게 질문들의 의미의 표준화를 시도하려고 면접원과 응답자 사이에 훨씬 높은 수준의 상호작용을 요구하는 대화식면접 (*conversational interviewing*)이 있다. 이 두 시각의 차이를 강조하기 위하여 우리는 각 기법의 가장 극단적인 형태를 설명한다. 그

렇지만 실제로는 대부분의 조사방법론자들은 이 두 극단적인 시각을 타협한 면접 기법을 주장한다.

표준화 면접 (Standardized Interviewing)

표준화 면접의 관점은 1950년대부터 조사연구에서 널리 사용되어 왔다. 이 관점은 면접에서 면접원의 역할은 응답자에게 질문을 써진 그대로 읽어 주는 것이다. 면접원은 질문을 할 때와 찾고자하는 정보에 대하여 완전히 중립을 지켜야한다. 만일 응답자가 질문에 대한 해석을 묻는 경우 면접원은 답해 줄 수 있더라 답해서는 안 된다. 질문을 반복하고, 응답항목들을 다시 읽고 응답자가 자기의 해석을 하도록 해야 한다. 만일 용어가 스크립트로 되어 있으면 모든 면접원은 면접의 일부로 그 용어의 정의를 제공해 줄 수 있다. 면접원은 명확하게 하려고 중립을 지키면서 추가 질문을 할 수 있지만 조사 절차에서 지시한 것이 아니라면 응답자에게 반응해서는 안 된다.

그러므로 표준화 면접의 목적은 응답이 면접원에 의해 조금도 영향을 받지 않게 하기 위하여 모든 응답자에게 같은 질문을 정확히 같은 방법으로 하는 것이다. 이론상으로는 만일 설계한대로 수행하면 표준화를 사용한 조사에서 발생하는 면접원 오류는 만일 교육을 잘 받은 한 면접원이 모든 응답자에게 같은 접근을 사용해서 조사의 모든 면접을 실시했을 때 생기는 오류와 비슷하다. 바꿔 말하면, 표준화 면접은 많은 면접원들이 조사를 실시할 때 발생할 수 있는 오류의 변동을 제거하는 것을 의도한다.

표준화 면접 기법은 설문지가 잘 설계됐고, 거의 모든 종류의 조사 상황에 적절하고, 질문들이 잘 표현됐고 특정한 응답자들을 위해서 바꿔 말할 필요가 없고, 면접 중에 면접원들이 마주치는 상황은 일반적으로 설문지에서 사용된 개념의 정의와 잘 대응되고, 설문지에서 다른 개념을 더 분명하게 해 줄 필요가 있는 응답자는 거의 없다고 가정한다. 또한, 이 상황의 예외는 설문지의 설계에서 대부분 예상할 수 있고 그래서 특수한 지시, 명확히 하는 것, 추가 질문하는 것 등은 원고로 쓰고 면접원들이 끈임 없이 전할 수 있다. 따라서 표준화 면접에서는 원고에서 벗어날 이유가 거의 없고 면접원들을 질문에 대한 예상하지 않은 어느 문제도 완전히 일치된 방식으로 처리하게 교육시킬 수 있다. 이 표준화의 중요 이점은 면접원 때문에 생기는 오류의 변동인 면접원 분산 (*interviewer variance*)이 없는 것이다. 면접원 분

산은 조사 추정값에 상당히 해로울 수 있어서 그것을 제거하려는 시도를 해볼 가치가 있다.

불행하게도 몇 가지 이유로 표준화 면접에서 그것의 이상적인 목적을 이루지 못하는 경우가 있다. 첫째, 길고 복잡한 내용에 대한 질문은 상당한 사전 검사와 수정을 한 후에도 표현이 어색하고 혼란스러울 수 있다. 간단한 것 같은 질문 개념들이 상당히 복잡하고 해석의 여지가 있을 수 있으며, 응답자의 상황이 조사에서 사용된 용어의 공식적인 정의에 쉽게 대응되지 않는다. 이런 경우에 응답자는 혼동되고 면접원한테 아무 도움도 없이 질문의 해석을 추측해야 한다. 해석이 틀릴 수 있기 때문에 답도 역시 그럴 수 있다.

또한 명단에 대한 정보나 사건 기록과 같은 종류의 질문에서는 면접 진행이 조직적이지 않을 수 있다. 응답자는 가구 구성원에 대한 정보를 제시하거나 직장 경력 같은 사건을 기억할 때 표준화된 방법의 가정한 순서에 의하지 않고 아무렇게나 기억해 낸다. 이런 상황에서 제공된 정보를 기록하는 더 융통성 있는 방법을 위하여 표준화 면접인 구성이 있는 방법을 포기하는 것이 좋다.

표준화 면접은(*standardized interviewing*) 면접원이 질문을 나타낸대로 묻고, 면밀히 조사하고, 의견을 제공하고, 모든 면접에 걸쳐 일치된 방법으로 응답자하고 작용하는 면접계획안이다. 면접원 때문에 발생하는 응답의 측정오차를 줄이기 위하여 면접원의 행동을 표준화시킨다.

대화식 면접 (*conversational Interviewing*)

표준화 면접의 대안으로 최근 조사방법론자들은 더 융통성 있는 대화식 면접 방법을 사용할 것을 주장했다. 가장 극단적인 형태의 대화식 면접은 자연스럽게 대화하고 낯선 두 사람 사이의 정상적인 사회 접촉과 비슷한 방법을 위하여 표준화 면접 방법을 근본적으로 포기한다. 이 방법에서는 면접원은 질문의 표현법을 바꿀 수 있다. 예를 들어, 응답자의 상황에 맞추기 위하여 질문의 의미와 그 질문이 응답자의 특정한 상황에 어떻게 적용되는지 확실히 하기 위하여 필요한 어떤 방법으로도 응답자를 자유로이 도울 수 있다. 따라서 대화식 면접은 면접원의 행동을 표준화하

기보다는 정의를 분명하게 설명하고 실제로 질문을 하는 연구자가 무엇을 묻고자 하는지 응답자가 이해하기 위해 필요할 수 있는 정보를 제공한다. 이 기법의 제안자들은 대화식 면접이 면접원 보다는 응답자 때문에 생기는 오류 변동을 줄인다고 주장한다.

한 예로 질문이 “가구에 몇 명이 살고 있습니까?”라고 하자. 표준화 면접원은 그 질문을 쓰인 그대로 읽는다. 대화식 면접 방법을 사용하는 면접원은 그 질문을 “당신과 당신의 남편 외에 여기 사는 다른 사람들이 있습니까?”로 바꾸어 말하거나 적당한 다른 방법을 사용한다. 대화식 면접원은 응답자가 질문을 잘못 해석했을 수도 있는 단서를 찾고 필요에 따라 질문을 분명하게 하도록 교육을 받았다. 예를 들어, 응답자는 “우리과 3개월 동안 함께 산 언니가 있는데 그러면 나와 내 남편 외에 한 명이 더 있습니다.”라고 답을 할 수 있다. 표준화 면접원은 언니의 주거 특징을 더 조사해 보는 것이 표준 면접 과정의 일부가 아니라면 이 답을 받아들일 수 있다. 대화식 면접원은 가장 정확한 답을 제공하기 위해 이 문제를 더 조사해 보는 것이 필요한 지를 결정하는데 완전한 자율성이 있다. 예를 들어, 설문지에서 지시되지 않았더라도 면접원은 “당신의 언니는 그 해에 다른 곳에 살았습니까?”와 “그녀가 집이라고 부르는 다른 데가 있습니까?”라고 물을 수 있다.

대화식 면접은 면접원들이 조사에서 각 질문에 대한 연구자의 의도를 충분히 잘 이해해서 그 의도를 응답자에게 전달하도록 교육시킬 수 있다고 가정한다. 이 점에 있어서 이 기법이 표준화 면접에서 보다 면접원에게 훨씬 더 높은 기대를 한다. 대화식 면접원은 질문의 기초가 되는 면접원의 의도를 알아야 한다. 예를 들어, “당신의 기관은 내년에 고용인 성취도를 평가하는 방법을 바꿀 계획이 있습니까?”라는 질문을 생각해 보자. 응답자는 여기서 무엇을 묻는 것인가에 대하여 많은 의혹이 있을 수 있다. 기관은 무엇을 뜻하는가? 내 부서(department) 또는 과(division)인가, 아니면 회사 전체인가? 바꾼다는 것은 무엇을 뜻하나? 작은 변화도 포함되나 아니면 변화가 총 점검(overhaul)에서처럼 큰 변화를 의미하나? 성취도 평가 과정의 어떤 양상이 연구자에게 관심인가? 질문은 일년에 한번 하는 검토 과정을 가리키나 아니면 중간의 검토도 중요한가? 대화식 면접원은 이런 질문들을 연구자의 시각에서 답할 수 있어야 하지만 표준화 면접원에게는 이런 기대를 하지 않는다.

면접원의 역할에 대한 논쟁의 중요한 부분은 면접원이 설문지에 있는 모든 질문에 대하여 연구자가 의도했던 대로 질문의 의미를 응답자에게 설명할 수 있을 만한

충분한 지식이 있기를 기대하는 것이 현실적인가에 대한 것이다. 일부 조사방법론자들은 대부분의 조사 기관의 거의 최하 급여로 필요한 연구 경력과 사회 조사에 나오는 많은 질문들의 복잡한 뉘앙스에 정통한 책임감 있는 몇 백 명의 면접원들을 고용할 수 있기를 기대하는 것은 현실적이지 않다고 믿는다. 대화식 면접 방법에 정복한 능력이 있는 충분한 수의 근로자들을 모으려면 면접원 급여를 상당히 높여야 한다는 데에는 의견이 일치한다. 대화식 면접을 위하여 면접원들을 적절히 교육시키는 것은 쉬운 일이 아니며 질문 뒤에 있는 목적에 관련된 필요한 지식을 알리는데 걸리는 시간은 표준화 면접에서 보다 상당히 더 걸린다.

이 방법의 또 하나의 문제는 운영 시간이다. 대화식 방법의 면접은 조사하는 개념의 복잡성과 주어진 특정 상황하에서 응답자가 최선의 응답을 결정하는 것의 어려움에 따라 표준 방법을 사용한 것 보다 1.5에서 3배 더 걸릴 수 있다.

대화식 면접의 제안자들은 비용의 증가는 정확한 자료를 수집하는 비용이라고 주장한다. 그렇지만 그 방법의 비판자들은 면접원에게 질문을 제시하는 방법 너무 큰 영향을 주기 때문에 이 방법에서 면접원 분산의 위험이 더 크고 그래서 그 질문의 답들도 분산이 크다고 주장한다.

대화식 면접 (*conversational interviewing*)은 면접원이 응답자와 자유롭게 상호 작용하고 응답자의 상황에 맞게 질문을 수정하고 적합시키고, 응답자들이 답을 구체화하는 것을 돕는 것을 허락하는 면접 방식이다. 면접원의 오류 뿐 아니라 모든 측정오차 원인을 최소화해서 가장 정확한 응답을 얻으려고 시도한다.

기타 면접 방식

면접원의 역할에 대한 최근 문헌은 비용과 자료 품질 면에서 두 방법이 결합된 것이 최고 일 수 있다고 제의한다. 한 예로 Schober와 Conrad(1997)가 제안한 방법은 질문하는 것은 표준화하지만 질문을 쓰인 대로 읽은 다음에는 면접원이 필요에 따라 즉흥적인 방식으로 사후 점검하게 한다. 면접원은 응답을 명백히 하기 위해 탐색조사하고, 질문에 대한 응답자의 이해를 결정하기 위해 사후 질문들을 하고, 질

문의 분명한 해석 차이를 바로잡고, 응답자의 응답을 근거로 상식적인 추론을 하게 용기를 북돋아 준다.

이해를 돕기 위하여 Schober와 Conrad에서 이야기한 U.S. Current Population Survey(CPS)에서의 다음 예를 생각해 보자:

면접원: *지난주에 당신은 파트타임, 저녁 또는 주말 일을 포함해서 직업이 한 가지 이상 있었습니까?*

응답자: *그것은 형편에 좌우되지요. 저는 여러 사람을 위해 아기를 돌보는데 그게 하나의 직업입니까 아니면 한 가지 보다 많은 것입니까?*

만일 면접원이 표준화 방법을 사용하면 면접원은 응답자에게 답을 하지 않고 응답자 자신이 질문을 해석할 것을 요구할 것이다. 하지만 그들이 부르는 회화적인 융통성 있는 대화식 면접 방법(*conversationally flexible interviewing approach*)에서는 면접원이 응답자에게 이 질문에서는 U.S. Bureau of Labor Statistics가 한 명 이상의 고용주를 위해서 아기를 돌보는 것도 한 가지 직업으로 친다고 설명해 줄 수 있다. 그래서 이 방법에서는 면접원이 질문을 쓰인 대로 읽지만 그 후에는 조사 설계자가 의도한 대로 응답자가 질문을 해석하는지를 확실하게 하기 위하여 면접원과 응답자를 도울 수 있다.

Conrad와 Schober는 질문들이 답하기 쉬운 경우에는 회화의 융통성이 표준 면접과 근본적으로 같은 결과를 준다는 증거를 제공한다. 그렇지만 더 어려운 질문에서는, 예를 들어, 응답자가 매우 복잡한 상황을 조사 질문에 있는 똑같이 복잡한 정의에 대응하는 것이 요구되는 질문들에서는 회화적으로 융통성 있는 방법이 더 정확한 자료를 준다. 하지만 이 방법은 Suchman과 Jordan이 제안한 대화식 면접 방법에서와 같이 면접원 모집, 교육, 면접시간과 관련된 많은 제한을 받는다.

표준화 면접과 대화식 면접의 가장 좋은 요소를 결합하는 또 하나의 가능성은 대부분의 조사 질문에서는 표준화 면접을 사용하고 대화식 면접은 표준화 방법이 다루기 힘든 몇 개의 중요한 질문들에서 사용한다. 후자의 예로는 가구 명부 정보의 수집, 지난주의 외출 날짜, 기간, 목적지 등, 지난 10년 동안의 경력, 그 밖의 설문지에 행렬이나 격자 형태로 잘 기록할 수 있는 정보가 있다. 이런 정보의 수집은 흔히 조직적이 아니기 때문에 순서와 질문의 표현을 표준화하는 것은 매우 어렵고

정보를 기억해낼 수 있는 응답자의 능력을 방해할 수 있다. 이 종류의 자료수집은 대화식이나 융통성 있는 대화식 면접 방법의 이상적인 응용이 될 수 있다.

이런 혼합된 면접 방법은 질문을 하는 것은 표준화하지만 응답자에게서 최고의 답을 얻기 위해 상당한 융통성을 면접원에게 허용한다. 그 점에 있어서는 연구 목적에 대한 면접원의 지식에 대한 요구는 대화식 면접에서와 거의 같다. 그렇지만 대부분의 경우에 응답자는 면접원의 도움 없이 질문이 원래 쓰인 대로 답을 할 수 있기 때문에 면접원 분산의 위험은 작을 것이다. 하지만 다양한 상황에서 이 면접 방식들 중에 어느 것이 가장 좋은가는 아직도 논란의 여지가 있다.

네덜란드의 Dijkstra와 van der Zouwen(Dijkstra, 1987; Dijkstra와 van der Zouwen, 1987, 1988)는 공식적인과 인간적인으로 구분하여 부르는 2가지 유형의 면접을 실험했다. 공식적인(*formal*) 또는 직무 지향적인 면접 방식(*task-oriented interviewing style*)에서는 면접원은 본질적으로 면접에 대해 초연한 것처럼 행동하고, 면접을 하는 동안 감정적인 반응과 개인적으로 주고받는 것을 자제하도록 교육을 받는다. 인간적인(*personal*) 또는 인간 지향적인 방식(*person-oriented style*)에서는 면접원이 자연스러운 인간 지향적인 행동을 하는 것을 허용한다. 예를 들어 면접원은 “안 됐네요,” “당신에게는 좋은 일이네요,” 또는 “저도 비슷한 생각이예요” 같은 말을 할 수 있다.

두 가지 방식 모두에서 표준화 방법에서처럼 면접원들은 질문을 쓰인 대로 읽도록 교육을 받았다. 그렇지만 한 방식은 분석적이고, 냉정하고 어떤 점에서는 이상할 정도로 사무 지향적이고 다른 방법은 훨씬 더 인간답고 우호적인 것으로 의도되었다. 분할 표본 실험에서 Dijkstra와 van der Zouwen은 양 방식에서 면접원과 응답자 사이의 상호작용을 코딩을 통하여 공식적인 방식에 비해서 인간적인 방식이 응답자들이 더 적당한 응답을 하게 자극하는 경향이 있다는 것을 알아냈다. 그렇지만 인간적인 방식은 지시적인 탐색조사와 부적당한 답을 해석하는 것과 같은 면접원이 응답을 편향할 수 있는 행동을 쉽게 할 수 있게 해준다. 따라서 인간적인 방식에서는 면접원이 질문과 답변의 과정에 인간적인 시각을 더하지 않도록 교육시키는 것이 중요하다. 반면에 공식적인 방식의 문제는 자연스러운 인간 지향적인 방식을 잇는 것이다.

5.2 면접원 변동

5.2.1 면접원 변동의 의미

면접원에 의해 일어나는 오류를 설명하는데 다양한 용어들이 사용되어 왔다. 문헌에서 사용되는 용어에는 면접원 변동, 면접원 분산, 상관 응답 분산(correlated response variance), 상관 면접원 오차(correlated interviewer error), 면접원내 상관(intra-interviewer correlation), 면접원 효과, 면접원 설계 효과가 있다. 우리는 이 잘에서 이런 용어들을 정의하고, 면접원 변동의 근본적인 원인에 대하여 이야기하고, 면접원이 조사 추정값에 미치는 효과의 예와 설명을 한다.

2장에서 계통오차와 가변오차를 정의하고 조사자료에서 그것들이 어떻게 발생하느냐를 설명했다. 면접원 오류는 이 두 개념과 관련이 있다. 면접원은 가변오차를 일으킬 수 있다 (즉, 응답자마다 바뀌고 추정값을 구하기 위하여 같이 합했을 때 서로 상쇄된다)(2장 참조). 면접원은 다소 계통적이면서 면접원에게 배정된 과제에서 모든 응답자에 걸쳐 응답에 같은 방법으로 영향을 미치는 경향이 있는 오류를 일으킬 수 있다. 그러므로 오류들은 상쇄되지 않고 편향이 생긴다.

가변 면접원 오류는 조사를 실시하는 과정에서 때로 면접원들이 하는 직접적인 관측 때문에 생길 수 있다. 예를 들어, 면접원에게 한 동네에 있는 각 집의 가격을 추정하라고 할 수 있다. 면접원이 어떤 집들의 시장가치를 과대평가하고 어떤 집들은 과소평가한다. 그렇지만 동네에 있는 집들의 평균 가격은 실제 평균 가격과 매우 비슷할 수 있다. 이 경우에는 오류들이 서로를 상쇄되는 경향이 있다. 하지만 어떤 면접원은 주택의 현재 가격을 알지 못할 수 있다. 면접원이 집을 거래한지 여러 해 되었으면 이 면접원의 평가는 너무 낮을 경향이 있다. 따라서 그 면접원의 동네 전체 집값에 대한 평가의 오차가 일반적으로 너무 낮을 것이고 그래서 그 동네의 주택의 평균가격은 하향 편향될 것이다. 이것은 계통 면접원 오류(systematic interviewer error)의 한 예다.

계통 면접원편향도 면접원에 따라 다를 수 있다. 예를 들어, 면접원 A는 그의 할당된 과제에서 주택의 가격을 -5% 과소평가 할 수 있고, 면접원 B는 이 가격을 +7% 과대평가 할 수 있고, 면접원 C는 +20% 과대평가 할 수 있고, 면접원 D는 -12% 과소평가 할 수 있다. 면접원들의 계통 편향사이의 변동은 때로는 면접원 변

동 (*interviewer variability*) 또는 면접원 분산 (*interviewer variance*)이라고 부른다.

5.2.2 면접원 변동이 모집단 추정값에 미치는 효과

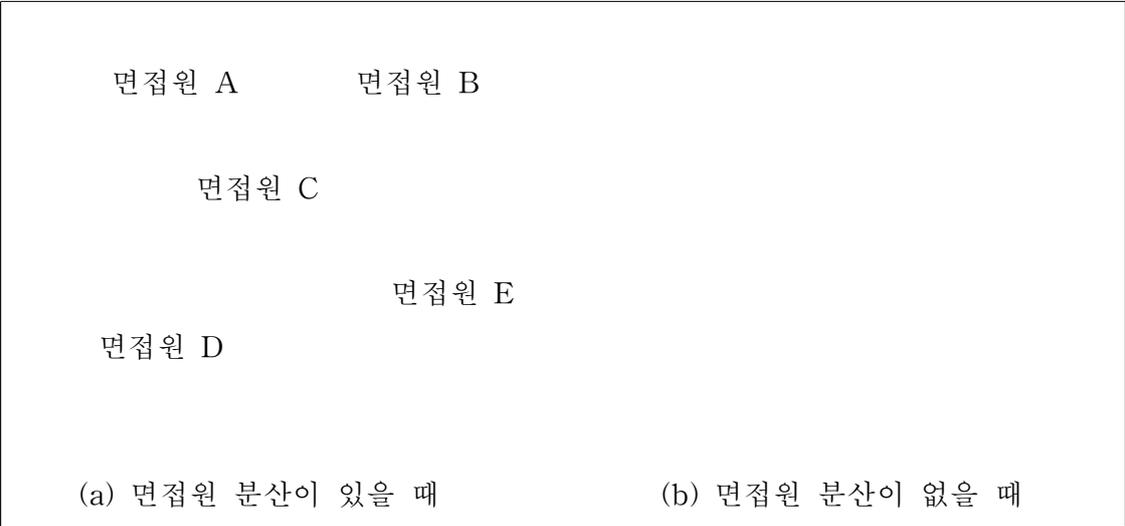
<그림 5.1>에서는 편향, 분산과 평균제곱오차의 개념의 이해를 돕기 위하여 2장에서 사용한 사격수와 표적의 설명으로 돌아가자. 그것은 면접원 분산의 개념과 왜 면접원 분산이 표본 평균과 다른 통계량의 변동을 크게하나를 이해하는 데에도 유용하다. 이것을 설명하기 위하여 한 지역에서 면접을 실시할 수 있는 면접원 A, B, C, D, E라고 부르는 5명의 면접원이 있다고 가정하자. 각 면접원에게 같은 수의 가구나 사업체를 배정하므로 각 면접원은 그 지역 표본의 5분의1에 대한 자료를 수집한다.

표본 단위로의 방문을 포함하는 전형적인 조사 운영에서 면접원에게 표본단위를 배정할 때에 자료를 수집하는데 드는 여행 비용과 그 밖의 비용을 최소화하도록 노력해야한다. 즉, 전형적으로 면접원 배정은 각 단위가 다른 것들과 지리적으로 아주 가까운 지리적 집락으로 구성한다. 표본 단위로의 여행비용과 시간을 절약하도록 한 집락을 가장 가까이 사는 면접원에게 배정한다. 그렇지만 집락들은 관심대상의 특성이 상당히 다를 수 있기 때문에 이 배정방법은 면접원 분산을 추정할 수 없게 한다. 예를 들어 어떤 집락들은 주로 고소득가구들만으로 이루어져있고 다른 것들은 중간이나 저소득가구들로 이루어져있다. 따라서 집락들간의 평균소득의 차이는 가능한 어떤 면접원 편향 이외에도 집락의 지리학 지역이나 동네 모두의 함수이다.

이런 이유로 면접원 분산을 관측하기 위하여 현지 조사에서 배정하는 방법을 바꿀 필요가 있다. 배정을 집락으로 하는 것 보다 배정을 무작위로 한다. 즉, 면접원 5명 모두가 조사할 지역의 랜덤 표본인 과제를 받을 때 까지 단위의 5분의1은 면접원 A, 단위의 5분의1은 면접원 B, 등으로 무작위하게 선택해서 면접원 연구과제를 구성한다. 이 배정방법은 흔히 상호관입(*interpenetration*)이라고 부른다.

상호관입 면접원 배정은 (*interpenetrated interviewer assignments*) 각 배정이 그 지역의 표본의 랜덤한 부표본을 이루도록 한 지역에 있는 표본 단위들을 그 지역에서 일하고 있는 면접원에게 무작위로 배정하여 구성한다. 이 종류의 실험설계는 현장에서 일하는 사람의 분산을 추정하기 위한 방법으로 Mahalanobis(1946)가 처음 사용했다.

상호관입 표본으로 개인 소득과 같은 조사의 한 특정한 관심대상의 변수에 면접원이 미치는 효과를 관측할 수 있다. <그림 5.1>에 있는 표적들은 다섯 면접원들이 조사한 가능한 소득의 시각적 개념이다. 왼쪽에 있는 표적은 (<그림 5.1a>) 자료에 상당한 양의 면접원 분산이 있을 때의 결과를 묘사하고 오른쪽에 있는 표적은 (<그림 5.1b>) 면접원 분산이 없거나 매우 작은 상황을 묘사한다. 면접원 분산이 없을 때에는 표본값은 표적에 한 개의 집락을 형성한다. 만약 다른 편향하는 요인이 없으면 집락은 표적의 정중앙에 위치한다.



<그림 5.1> 면접원 분산이 있을 때와 없을 때의 표본값의 분포. 표적의 정중앙에 있는 탄환들은 어떤 특징에 대한 표본값들을 나타낸다. (a)는 면접원 분산에 의한 높은 수준의 표본 변동을 보여준다. 각 집락은 면접원 배정내에 있는 값들을 나타낸다. 집락 안에 있는 작은 정사각형들은 집락의 평균값을 나타낸다. 면접원 분산이 있으면 계통 면접원 오류 때문에 집락 평균들이 상당히 다르다. 이것은 집락들이 표적의 정중앙을 가로질러 사방으로 흐트러지고 높은 면접원 분산의 결과를 가져온다. 면접원 분산이 없으면 (b)에서처럼 모든 집락들의 평균이 같다. 결과는 훨씬 줄어든 표본값의 분산이다.

그러지만 면접원들이 그들에게 배정된 단위들의 소득에 대한 질문의 응답에 영향을 미치면 표본에 있는 값들은 표적 곳곳에 집락을 형성하는 경향이 있다. 그림에서 보여주듯이 어떤 배정의 값들은 표적의 정중앙에 집중되어 있지 않고 이것은 값들이 편향되었다는 것을 가리킨다. 이것은 배정을 받은 면접원 때문에 생기는 편향이다. 두 집락의 중심간의 차이는 두 면접원의 편향의 차이와 같다. 그러므로 만일 두 면접원의 편향이 같으면 그들의 집락은 중복 될 것이다. <그림 5.1a>에서 보여주듯이 면접원 A의 배정에 있는 값들은 표적의 한 쪽에 있고 면접원 E의 것은 반대쪽에 있는데 이것은 이 두 면접원들은 소득 질문에 대한 응답에 매우 다른 영향을 미친다는 것을 말한다. <그림 5.1a>의 표적에 표시된 값은 꽤 흩어져 있다는 것에 주의하라. 그림에서 면접원 분산이 없는 경우 표적에 있는 값이나 “탄흔들”의 흩어짐을 면접원 분산이 있는 경우와 비교하면 면접원 계통 편향의 변동이 (즉, 면접원 변동) 소득 질문에 대한 응답의 분산을 증가시키는 것을 알 수 있다.

그러므로 면접원 변동의 효과는 표본에 있는 관측값의 분산을 늘리는 것이다. 이 표본에 있는 관측값의 분산의 증가는 모집단 모수의 추정값의 분산을 늘리는 것을 의미한다. 다음 절에서 보겠지만 평균, 총계, 비율의 추정값의 증가량은 면접원이 일으킨 관측값의 변동량과 정비례한다.

5.2.3 면접원 변동이 조사 추정값의 정확도에 미치는 효과의 정량화

면접원 분산을 위한 단순 모형

<그림 5.1>은 면접원이 관측값들을 표적의 정중앙에서 멀리 옮겨놓거나 관측값에 편향을 더하는 방식으로 응답에 영향을 미칠 수 있다는 것을 제시한다. 면접원 분산은 이 편향들이 면접원들 사이에 상이 할 때 일어난다. 이 개념은 면접원 변동이 관측과 이 관측에서 나온 추정값에 미치는 영향을 정량화하는데 사용할 수 있는 간단한 수학적 모형을 제시한다. 이 절에서는 이런 모형을 제시하고 이것이 암시하는 면접원 오류가 추정값에 미치는 효과에 대해 생각해본다.

<그림 5.1>은 다음과 같이 조사오차에 대한 매우 간단한 모형을 제시한다. y 는 표본 구성원에서 나온 한 관측값을 나타내고 i 는 표본 구성원을 나타낸다. 그러므로 y_i 는 i 번째 표본 구성원에서 나온 관측값을 나타내고 i 는 총 표본 크기인 1에서 n

까지 아무 값이 될 수 있다. 따라서 우리는 y_i 를 표적에 있는 한 개의 탄흔을 나타낸다고 생각할 수 있다. 모집단의 평균 또는 표적의 정중앙을 그리스 문자인 μ 로 그리고 μ 에서부터의 y_i 의 편차는 e_i 로 나타낸다. 따라서

$$\begin{aligned} \text{관측값} &= (\text{실제 모집단 평균}) \\ &+ (\text{실제 모집단 평균으로부터의 편차}) \end{aligned}$$

라고 쓰거나 또는 우리가 방금 정의한 기호로는 다음과 같다.

$$y_i = \mu + e_i, \quad i=1, \dots, n \quad (5.1)$$

<그림 5.1>에 있는 양 표적에서 e_i 로 나타낼 수 있는 편차의 합은 대략 0이다. 즉, y_i 값들은 표적의 정중앙 μ 주위에 집중되어 있어서 전체 표본에서 y_i 의 평균은 약 μ 이다.

그렇지만 <그림 5.1a>에 있는 표적에서는 각 면접원이 배정내에 있는 관측값을 대략 같은 어떤 양으로 옮겨놓은 듯하다. b_j 가 이 이동량(또는 계통 면접원 오류)을 나타내고 j 가 면접원(즉, $j=A, B, C, D, E$)을 나타내고 b_j 는 면접원 j 의 계통 편향이다. 또한, k 는 면접원의 배정 안에 있는 단위를 나타낸다. 예를 들어 만일 100개의 단위가 있는데 각 면접원은 20개를 배정 받았으면 k 는 $1, 2, \dots, 20$ 이 된다. 마지막으로 그리스 문자 ε 은 관측값과 실제 평균과 면접원 편향의 합의 차이를 나타낸다. 그러면 y_{jk} 는 면접원 j 의 배정에서 k 번째 단위의 관측값을 나타내고 ε_{jk} 는 $\mu + b_j$ 로부터 y_{jk} 의 편차를 나타낸다. 그러므로

$$\text{관측값} = (\text{실제 모집단 평균}) + (\text{면접원 편향}) + (\text{이 응답자의 편차})$$

라고 쓰거나 우리가 정의한 기호로는 $j = A, B, C, D, E$ 와 $k = 1, 2, \dots, 20$ 일 때, 다음과 같다.

$$y_{jk} = \mu + b_j + \varepsilon_{jk} \quad (5.2)$$

식 (5.2)의 해석을 <그림 5.1a>의 각도에서 설명해보자. 그림에서 면접원 A라고 표시된 탄흔들의 집락을 생각해보자. 그러면 각 탄흔은 k가 1에서 탄흔 수일 때 y_{Ak} 로 표시할 수 있다. 식 (5.2)에서 $y_{Ak} - \mu = b_A + \varepsilon_{Aj}$ 이기 때문에 집락에 있는 각 탄흔은 탄흔에서 표적의 정중앙까지의 편차가 $b_A + \varepsilon_{Aj}$ 이다. 그림에서는 각 탄흔들의 집락의 중심은 정사각형(■)으로 나타냈다. 이것은 면접원의 편향으로 해석할 수 있다. 따라서 표적의 정중앙에서 정사각형까지 면접원 A의 편차는 b_A 다. 마지막으로 집락 A안에 있는 각각의 탄흔과 집락 A의 탄흔의 중심에 있는 정사각형 간의 편차는 가변오차 ε_{Ak} 이다. 이 해석들은 <표 5.1>에 요약되어 있다.

<표 5.1> <그림 5.1>의 각도에서 본 모형 요소들

모형 요소	해석	기호
면접원 A의 관찰값	면접원 A 집락에 있는 탄흔	y_{Aj}
실제 모집단 평균	표적의 정중앙	μ
면접원 편향	면접원 A 집락의 중심	b_A
변수	면접원 A 집락 안의 탄흔과 집락의 중심과의 편차	ε_{Aj}
총오차	표적의 정중앙과 면접원 A 집락안의 탄흔과의 차이	$b_A + \varepsilon_{Aj}$

모형의 추가적인 설명을 위하여, 면접원들이 농장 관리자에게 농장에 있는 다양한 구획의 시장 가치를 추정하라고 의뢰하는 상황을 생각해 보자. 면접원은 관리자의 시장 가치 추정값을 자산상의 개량 없이 그리고 농장이 개발이 아닌 농업지로 판다고 가정하고 조사해야 한다. 그렇지만 어떤 면접원들은 지시를 잘못 알아듣고 농장주에게 자산상의 개량을 포함한 가치나 땅을 땅 개발자에게 팔았을 때의 가치를 제공하라고 부탁한다.

또한 어떤 응답자들은 그들의 땅을 파는 것에 대하여 한번도 생각해 보지 않았을 수 있고 그 땅이 어떤 가치가 있는지 전혀 모를 수도 있다. 어떤 면접원들은 항목

무응답 보다는 응답을 얻기 위하여 응답자가 추정값을 내게 이끌 수 있다. 그 결과 땅의 가치의 추정값은 면접원이 생각하는 땅의 가치에 상당히 영향을 받을 수 있고 땅의 가치 질문에서 면접원 변동의 위험은 매우 높다.

한 지역에 있는 구획의 가치가 약 한 에이커당 \$1500라고 가정하자. 또한 한 면접원은 평균적으로 한 에이커당 대략 \$500보다 상향 편향된 가치를 얻는다는 것을 우리가 안다고 가정하자. 현실로는 면접원의 편향은 알지 못한다. 하지만 설명의 목적으로 가정하자. 자기의 배정에 있는 5명의 농장 관리자한테서 면접원은 \$4800, \$6200, \$4400, \$6700, \$5900의 값을 얻는다. 그러므로 식 (5.2)에 있는 모형의 요소들의 값은 <표 5.2>에 주어진 대로이다. 모형이 이 값을(즉, $y_{jk} = \mu + b_j + \varepsilon_{jk}$) 만족한다는 것을 입증할 수 있다. 또한 가변오차 ε_{jk} 는 합해서 0이 된다.

<표 5.2> 면접원 모형의 요소의 값

j	y_{ij}	μ	b_i	ε_{ij}
1	4800	5100	500	-800
2	6200	5100	500	600
3	4400	5100	500	-1200
4	6700	5100	500	1100
5	5900	5100	500	300

이런 면접원 오류에 대한 모형은 면접원들이 조사자료에 영향을 미칠 수 있는 복잡한 방식들을 다소 지나치게 간소화한다. 예를 들어 면접원이 특정한 질문에 대하여 얻는 모든 응답에 일정한 같은 편향을 더한다고 가정한다. 물론 이 모형은 범주형 자료 항목들(예를 들어, “예” 또는 “아니오” 응답) 보다는 나이, 소득 같은 연속 자료 항목들에 적용된다. 면접원은 어떤 응답들에는 다른 것들 보다 더 큰 영향을 미칠 수 있기 때문에 연속 항목에서 조차도 그것은 너무 간단하다고 할 수 있다. 예를 들어 응답자가 면접원의 영향을 받는 정도는 질문 주제, 응답자의 특성과 면접원의 특성에 좌우될 수 있다는 것을 나중에 볼 것이다. 모형이 현실을 지나치게 간소화하는 또 다른 면들도 있다. 그럼에도 불구하고 모형은 면접원 변동의 효과와 관련된 통찰(insight)을 제공하는데 매우 유용하다. 그것은 또한 면접원 효과의 규모를 추정하려고 하는 연구 설계에 사용할 수 있다. 그래서 모형이 간단할지라도 면

접원 오류의 연구를 위한 매우 중요한 도구가 된다.

면접원 설계 효과

면접원 분산이 표본의 평균에 미치는 효과의 정도를 요약한 척도를 제안한 몇 개의 문헌이 있다. 이 척도 중에 가장 널리 쓰이는 것은 Kish(1962)가 제안한 “면접원”을 나타내는 첨자 “int”가 있는 그리스 문자 ρ 로 나타낸 면접원내 상관계수 (*the intra-interviewer correlation coefficient*)라고 부르는 것이다. Kish는 ρ_{int} 를 수학적으로 정의한다. 하지만 우리는 이 책에서 수학적인 취지를 최소화하려고 하기 때문에 ρ_{int} 를 <그림 5.1>에 있는 표적으로 정의할 것이다.

Kish는 ρ_{int} 를 두 분산의 비율 또는 변동의 척도로 정의한다. 분자는 면접원간의 분산이고 분모는 면접원간의 분산과 면접원내의 분산의 합인 총분산이다. <그림 5.1a>로 보면 면접원 평균은 정사각형 (■)으로 표시되었다. 면접원간의 분산 (*between-interviewer variance*)은 간단히 표적의 정중앙 주위에 이 정사각형들의 흐트러짐의 척도이다. 면접원 모형에서 j번째 면접원에 대하여 표적의 정중앙에서 이 정사각형들까지의 거리가 b_j 라는 것을 상기해라. 그러므로 ρ_{int} 의 분자는 b_j 들의 분산이기도하다.

ρ_{int} 의 분모는 분자에 있는 수량과 면접원내 분산의 합이다. 면접원내 분산 (*within-interviewer variance*)은 <그림 5.1a>에서 각 면접원 집락내에 있는 탄흔들간의 변동이다. 그것은 또한 <그림 5.1b>에서 탄흔들의 흠어짐에 대한 변동이다. 즉, 자료에서 면접원 편향을 제거한 후의 표본값의 분산이다. <그림 5.1b>는 근본적으로 <그림 5.1a>에서 면접원 영향 (즉, b_j)을 제거한 것이다. 마지막으로 ρ_{int} 의 분모를 계산하는 세 번째 방법은 <그림 5.1a>에 있는 모든 탄흔들의 변동을 모든 면접원에 걸쳐 계산하는 것이다. 그러므로 ρ_{int} 는 <그림 5.1a>에서 보여준 면접원 배정들의 평균의 변동 나누기 그 수량과 <그림 5.1b>에 있는 탄흔들의 변동의 척도의 합이다.

예 5.2.1 <그림 5.1a>에서 다섯 면접원의 b_j 값은 430, 445, 10, -435, -450이라고 가정하자. 이 수들의 분산은 다음에 있는 오차의 표본의 분산 공식을 사용해서 계

산할 수 있다.

$$\begin{aligned} \text{분산} &= \frac{b_1 + b_2 + b_3 + b_4 + b_5}{5} \\ &= \frac{430^2 + 445^2 + 10^2 + (-435)^2 + (-450)^2}{5} = 154,150 \end{aligned}$$

이 공식은 단지 기초통계학 과목에서 접하는 잘 알려진 분산 공식을 편향의 제곱의 합 나누기 5로 간단히 한 것이다. 이것은 다섯 편향의 평균이 0이기 때문이다. 또한 평균 면접원내 분산은 <표 5.2>의 y_{jk} 열에 있는 수들의 분산인 935,000이라고 가정하자. 이 설명을 목적으로 면접원 A의 면접원내 분산은 모든 면접원에 걸친 평균 면접원내 분산과 같다고 가정하자. 그러면 다음과 같다.

$$\begin{aligned} \rho_{int} &= \frac{\text{면접원간 분산}}{\text{면접원간 분산} + \text{면접원내 분산}} \\ &= \frac{154,150}{154,150 + 935,000} = 0.142 \end{aligned} \quad (5.3)$$

이 책의 범위를 벗어나지만 위의 면접원 오류 모형 하에서 ρ_{int} 가 면접원의 배경 내에 있는 두 관측값 사이의 상관이다. 면접원 효과의 또 다른 해석은 면접원 오류가 일으킨 응답 사이의 상관이다. 그래서 면접원내 상관계수가 0.142라고 말한다. 대신에, 면접원 분산과 총분산의 비율이 0.142라고 할 수 있다. 나중에 보겠지만 ρ_{int} 의 이론적인 값은 0과 1 사이지만 ρ_{int} 의 추정값이 마이너스가 될 수 있다. 이런 마이너스 추정값을 다루는 흔한 방법은 그것들을 0으로 대체하고 면접원 분산이 없는 것으로 해석한다.

어느 특정한 조사 항목에 대한 기호 ρ_{int} 는 그 항목의 면접원내 상관계수 (*intra-interviewer correlation coefficient*)이다. 조사 항목에 대하여 면접원이 조사 응답에 미치는 영향의 정도의 척도로 흔히 사용된다. ρ_{int} 값이 클수록 면접원 분산도 크다. ρ_{int} 가 0인 것은 면접원 분산이 없다는 것을 뜻한다.

이제 ρ_{int} 를 계산했으니 다음의 질문이 생긴다. 그것을 가지고 무엇을 합니까? <그림 5.1a>와 <그림 5.1b>를 비교하면서 보았듯이 면접원 변동은 표본 응답의 총 분산을 늘린다. Kish(1962)에서 이 증가량은 $deff_{int}$ 라고 표시한 면접원의 설계 효과(*design effect for interviewers*)와 관련이 있다.

$$deff_{int} = 1 + (m-1)\rho_{int} \tag{5.4}$$

여기서 m을 면접수를 보았을 때, 이는 조사를 위한 평균 면접원 작업량이다. 수량 $deff_{int}-1$ 또는 $(m-1)\rho_{int}$ 는 면접원 변동의 결과 생기는 평균, 총계, 또는 비율의 분산의 증가이다.

예를 들어 위에서 계산했듯이 $\rho_{int}=0.142$ 라고 가정하자. 또한 면접원들은 조사에서 평균 50명을 면접한다고 가정하자. 그러면 $deff_{int}$ 값은

$$deff_{int} = 1 + (50-1)\times 0.142 = 7.96 \tag{5.5}$$

이고 $deff_{int}-1$ 은 6.96이다. 즉, 조사응답에 미치는 영향의 결과 $\rho_{int}=0.142$ 인 어떤 특성의 평균의 분산이 거의 7배가 늘어났다. 특히 면접원내 상관계수가 0.142밖에 안된다는 것을 고려하면 이것은 분산의 엄청난 증가이다. 몇 개의 예를 통해 보겠지만 0.1이상의 ρ_{int} 는 매우 큰 ρ_{int} 이다. 사실은 대부분의 설문지 항목에서 ρ_{int} 는 전형적으로 0.0과 0.05의 범위 안에 있다.

$deff_{int}$ 는 m의 증가 함수라서 평균 면접원 배정 크기가 크면 클수록 면접원 때문에 생기는 추정량의 분산의 증가가 크다는 것에 주목하라. 전형적으로 m은 대면면

접에서는 20에서 80, 그리고 전화면접에서는 80에서 150이상 사이에서 변동한다. <그림 5.2>에서 총 분산의 증가 퍼센트는 m 의 함수인 $(deff_{int}-1) \times 100\%$ 를 m 은 20에서 150 그리고 ρ_{int} 는 0.0에서 0.1 범위에서 그렸다. 이 그림은 상관된 면접원 오류의 결과 생기는 분산의 상당한 증가를 확실히 보여준다.

면접원 상관(x축에 표시)이 늘어남에 따라
 표본평균의 분산도 늘어난다.
 그 효과는 더 면접원 배정크기 m 이 클수록 더 크다.

면접원 상관

<그림 5.2> 면접원내 상관계수 ρ_{int} 와 평균 배정크기 m 의 함수로 나타낸 상관 오차 때문에 생기는 표본평균의 분산의 증가. (출처: Biemer and Trewin(1977), 그림 27.1)

면접원 설계 효과 (*interviewer design effect*) $deff_{int}$ 는 면접원 분산 때문에 생기는 단순확률표본의 평균분산의 증가의 척도이다. 예를 들어, $deff_{int}=1$ 은 분산의 증가가 없는 것을 나타내고 $deff_{int}=2$ 는 면접원 분산의 결과 분산이 두배가 된다는 것을 나타낸다 (즉, 100%의 분산의 증가).

예 5.2.2 6000명을 대상으로 40명의 전화 면접원이 작업량을 대략 같게 나누어서 수집한다고 가정하는 전화면접을 생각해 보자. 그러면 이 조사의 m 은 대략 $6000/40$ 또는 150번의 면접이다. 조사에 나온 한 질문의 ρ_{int} 값이 대략 0.013으로 추정된다

고 가정하자. 상관된 면접원 오류의 결과 이 특성의 평균의 추정값에 대한 분산의 증가는 무엇인가?

식(5.4)의 $deff_{int}$ 공식을 적용하면

$$deff_{int} = 1 + (150 - 1) \times 0.013 = 2.94$$

이다. 따라서 분산은 $(2.94 - 1)$ 또는 1.94% 늘어났다.

효율적인 표본크기

조사의 실제 표본크기가 n 이지만 면접원 분산 때문에 관심의 대상인 조사 특성에 대하여 우리가 얻는 정보량은 표본크기가 시사하는 것 보다 훨씬 작을 수 있다. 예를 들어, 만일 면접원 분산이 단순확률표본의 표본평균의 분산은 잘 알려진 σ^2/n 이다. 그렇지만 우리가 보았듯이 면접원 분산은 평균의 분산을 인수 $deff_{int}$ 로 늘어나게 한다(즉, 평균의 분산은 σ^2/n_{eff} 이고 n_{eff} 는 $n/deff_{int}$ 이다). 바꿔 말하면 1보다 큰 $deff_{int}$ 를 가진 특성에 대하여 평균을 계산하면 평균의 분산은 표본크기 n_{eff} 인 평균의 분산과 같다. 그러면 표본크기는 사실상 n_{eff} 이고 n_{eff} 가 그 특성에 대한 유효 표본크기이다.

n_{eff} 라고 나타낸 조사 항목의 유효 표본크기(*effective sample size*)는 표본 크기 n 나누기 면접원들의 설계 효과인 $deff_{int}$ 이다. 그것은 면접원 오류의 결과 생기는 정확도의 손실(또는 표본 정보의 손실)이다.

위의 예에서, 6000명을 대한 전화조사는 면접원 오류가 전혀 없는 6000/1.94 또는 3092명의 조사에서와 조사항목에 대한 같은 정보를 제공한다. 따라서 만일 우리가 이 특성에 대한 면접원 분산을 제거하는 방법을 찾을 수 있다면(즉, $\rho_{int}=0$ 을 만들면) 소득의 추정값에 대한 정확도의 관점에서 표본에 거의 3000명을 더하는 것과 거의 동등하다.

면접원 분산과 관련된 또 하나의 중요한 사실은 조사를 할 면접원 수가 늘어남에 따라 $deff_{int}$ 의 크기가 줄어들고 유효 표본크기는 늘어난다. 예를 들어, 만일 전화조사에서 면접원수를 40에서 100으로 늘리면 평균배정크기는 150명에서 60명으로 줄어든다. 식 (5.5)를 사용해서 $deff_{int}$ 를 다시 계산하면 $deff_{int}$ 가 2.94에서 1.77로 줄어드는 것을 볼 수 있다.

그러므로 만일 조사의 어떤 중요한 주제에 대하여 면접원 분산이 문제가 될 것으로 예상되면 면접원을 100명 고용하는 조사가 40명을 고용하는 조사보다 더 정확할 것이다. 그렇지만 이러한 결과는 조사에서 면접원 수를 늘이는 데에서 생기는 실제적인 문제들을 무시한다. 예를 들어 더 많은 면접원을 고용하고, 교육시키고 모니터 하는데 드는 비용으로 인하여 면접원 수를 늘려서 얻는 분산의 감소가 가치가 없을 수 있다. 특히 만일 각 면접원에게 면접할 대상이 단지 몇 개로 주는 경우에 그럴 것이다. 또한 면접원에 배정된 수가 너무 작으면 면접원들은 그들의 일을 끝내기 전에 면접하는 경험을 별로 얻지 못할 것이다. 따라서 150명 보다 40명의 면접원이 하는 조사의 p_{int} 가 훨씬 높을 수 있다. $deff_{int}$ 를 다시 계산하는데 있어서 양쪽 상황에 같은 p_{int} 를 적용했다고 가정했다. 실제로는 그렇지 않을 수 있다.

6000명 단위 표본에서 최대 면접원수는 6000이라는 것에 주목해라. 이것은 면접원 1명당 한 개의 표본단위를 의미한다 (즉, $m=1$). 이것이 결국 자기기입식 설문지에서 일어나는 상황이다. 왜냐하면 각 응답자가 설문지를 완성하는데 면접원의 역할을 한다. 그래서 자기기입식 설문지에서는 $m=1$, $deff_{int}=1$ 이고 유효표본크기는 $n=6000$ 이다 (즉, 면접원 분산 결과로 발생하는 분산의 증가가 없다).

마찬가지로 조사에서 최소 면접원수는 1명이다. 예를 들어, 전화조사 예에서 면접원 1명이 6000명 전체 표본에 대한 면접을 실시했다고 가정하자. 물론 이것은 비현실적이지만 그 경우 면접원 분산이 어떻게 되는지가 흥미롭다. 여기에서 식 (5.4)에 의하면 $deff_{int}$ 가 최대값인 77.7이다. 면접원수를 1로 줄여도 p_{int} 는 바뀌지 않는다고 가정하는데 이것은 아마도 가능성이 별로 없다. 그럼에도 불구하고 p_{int} 를 원래 값의 10분의 1로 줄인다해도(즉, 0.013에서 0.0013으로) $deff_{int}$ 는 그래도 크다:

$$deff_{int} = 1 + 5999 \times 0.0013 = 8.8.$$

일반적으로 한 면접원이 조사 전체를 위한 자료를 수집하는 것은 매우 좋지 않은

조사 설계이다. 왜냐하면 그 설계에서 면접원 오류의 효과가 최대화된다. 그렇지만 조사의 최적의 면접원수를 결정하는 것은 면접원 분산의 관점에서 뿐만 아니라 조사비용과 면접원 모집, 교육, 감독 같은 현지 운영에 관련된 다른 요인들도 근거로 해야 한다.

면접원 분산과 미국 인구주택센서스

1950년대에서 1970년대까지 면접원 분산은 미국 센서스국의 중심 연구분야였다. 1950년 미국 인구주택센서스(*Census of Population and Housing*)에서 Morris Hansen과 그의 동료들은 면접원 분산이 센서스 결과에 미치는 영향을 평가하기 위한 실험을 실시했다. 그들은 많은 센서스 항목에 대한 면접원 분산이 상당히 높고 어떤 항목에서는 ρ_{int} 가 0.1을 넘었다. 이 결과는 1960년에 우편으로 부치고 우편으로 돌려주는(mail-out/mail-back) 센서스를 실시하기로 결정하는데 중요한 요인이었다(Hanson 등, 1961 참조).

조사에서 ρ_{int} 의 추정

이미 언급했듯이 현지면접조사에서 면접원 배정을 정하는 관행은 표본단위로 이동하는데 드는 과도한 비용을 가능한 피하기 위하여 면접원의 작업량을 지리적으로 그리고 면접원의 집 가까이에 집중시킨다. 예를 들어, 면접원을 두 명을 고용하는 큰 도시에서 하는 조사에서 시의 남쪽에서 거주하는 면접원에게는 남쪽에 집중된 단위들을 배정할 수 있고 북쪽에 사는 면접원은 그 부근에 배정 받는다. 면접원들의 표본들 간의 평균적인 특성의 차이는 도시의 북쪽과 남쪽 거주자의 차이로 간주하기 때문에 이러한 배정은 면접원 편향을 추정하는 것을 불가능하게 한다.

통상적인 배정 대신에 두 면접원의 배정을 상호관입했다고 가정하자. 각 면접원의 배정은 같은 모집단에서 나온 확률표본이라는 것을 뜻한다는 것을 이미 언급하였다. 예를 들어, 남쪽에 사는 면접원은 북쪽 표본의 대략 반을 가질 것이고 북쪽에 사는 면접원도 반대로 그렇다. 이렇게 하여 두 배정의 모집단 특성의 모든 차이가 제거된다. 그러면 만일 두 면접원이 비슷한 방식으로 실행하면 두 배정이 비슷한 응답 패턴을 가져올 것으로 기대할 수 있다(확률표본추출오차 범위안에서). 만일 두 배정의 평균적인 특성, 응답률이나 다른 요약 척도가 유의적으로 다르면 그 차이를 면접원 수행과 관련된 무언가를 탓으로 할 수 있다 (즉, 차이의 원인으로 모집단 차

이는 배제할 수 있다).

대면조사의 면접원 분산을 추정하는 실험에서 최소한 표본의 일부는 상호관입배정을 사용하고 표본의 나머지는 보통의 배정 배분하는 방법을 사용한다. 늘어나는 이동비용과 현지 직원을 조화시키는데 더 복잡한 것 때문에 상호관입표본에 필요로 하는 여분의 비용 때문이다. 전화조사에서는 이동 비용은 문제가 안 되고 가장 기본적인 자동화전화관리체계를 사용해도 면접원에게 전화번호를 랜덤하게 배정하는 것은 자연스럽게 쉽게 관리할 수 있기 때문에 상호관입설계는 상당히 실현 가능하다.

조사에서 면접원 오류에 대한 통제의 중요함과 그것을 추정하는데 드는 비교적 적은 비용을 생각하면 면접원 효과의 평가가 모든 중앙집중식(centralized) 전화조사 운영에서 관례적인 과정이 되어야하지 않을까요? 사실 상관 면접원 오차는 전화조사 운영에서 거의 추정하지 않는다. 평균제곱오차의 이러한 중요한 요소에 대한 주의가 부족한 이유는 면접원 오류의 해로운 효과에 관한 연구자들의 인식이 부족하기 때문이다. 또한 면접원 오류 평가는 조사 사후의 품질 척도이다 (즉, 그것은 자료수집이 완성된 후에 계산하되 면접 품질 측정이다). 따라서 한번만 실시하는 조사에서 ρ_{int} 추정에 대한 가치는 계속되는 조사에서 보다 적을 수 있다. 왜냐하면 후자의 경우에는 그것을 계속되는 품질개선과정에 포함 시킬 수 있다.

ρ_{int} 의 추정값은 매우 쉽게 변하기 때문에 두 가지 면접 방법에서 면접원 분산의 안정된 추정값을 얻는 것은 어렵다. 사실 평균작업량이 50 사례 이하인 20명이하의 면접원들을 고용하는 조사에서는 추정값의 표준오차가 너무 크기 때문에 좋은 ρ_{int} 추정값을 얻는 것은 보통 불가능하다. ρ_{int} 의 추정값은 두 분산 추정값들의 차이를 가지고 계산하기 때문에 추정값들이 마이너스가 될 수 있다. 그렇지만 전에 언급했듯이 ρ_{int} 는 두 플러스 수량의 비율이기 때문에 이론적으로 플러스 값만 되어야 한다. 따라서 이 모형에서는 마이너스 ρ_{int} 는 해석할 수 없고 보통 추정값이 불안정 상태라는 것을 의미한다.

이러한 문제점에도 불구하고 상관관계가 있는 면접원 오차에 대한 여러 연구가 발표 되었고 그 중 많은 결과는 Groves(1989, 8장)에 요약 되었다. <표 5.3>은 Groves가 제시한 두개의 표를 다시 정리한 것이다. 여기에는 10개의 대면조사와 9개의 전화조사의 평균 ρ_{int} 값이 포함되는데 ρ_{int} 값의 범위는 Health in America

Study의 0.0018에서 높게는 Lesotho의 World Fertility Survey의 0.102까지 이다. ρ_{int} 값의 대부분은 0.005에서 0.06사이에 있고 표에서 중앙값은 대략 0.01이다. 중앙값 아래에 있는 대부분의 ρ_{int} 값은 전화조사에서 나온 것이고 위에 있는 대부분의 ρ_{int} 값은 대면조사에서 나온 것이다. 대면조사의 평균 ρ_{int} 값은 약0.03이고 전화조사의 평균은 약 0.01이다.

<표 5.3> 면접원 분산 연구논문에서 나온 ρ_{int} 값

ρ_{int} 를 보고하는 연구	면접 방식	ρ_{int} 의 평균값
육체 노동자 연구 (Kish, 1962)	대면	
연구 1		0.020
연구 2		0.014
Canadian Census, 1961 (Fellegi, 1964)	대면	0.008
Canadian Health Survey (Feather, 1973)	대면	0.006
Study of Mental Retardation (Freeman and Butler, 1976)	대면	0.036
World Fertility Survey (O'Muircheartaigh and Narcjwardt, 1980)	대면	
페루, 본조사		0.050
페루, 재조사		0.058
레소토, 본조사		0.102
Consumer Attitude Survey (Collins and Butcher, 1982)		0.013
Interviewer Training Project (Fowler and Mangione, 1985)	대면	0.005
대면조사의 평균 ρ_{int}		0.0312
Study of Telephone Methodology	전화	0.0089
Health and Television Viewing	전화	0.0074
Health in America	전화	0.0018
1980 Post 선거 연구	전화	0.0086
Monthly Consumer Attitude Survey	전화	
1981년 11월		0.0184
1981년 12월		0.0057
1982년 1월		0.0163
1982년 2월		0.0090
1982년 3월		0.0067
전화조사의 평균 ρ_{int}		0.0092

출처: Groves(1989), 8장.

연구 문헌을 보면 전화조사가 대면조사보다 상관 면접원 오차의 경향이 적은 듯하다. 지속적인 감독이 중앙집중식 전화면접의 가장 중요한 장점 중의 하나인 점을 고려하면 이 결과는 놀라운 것이 아니다. 이러한 특징은 중앙집중식 전화면접의 초

창기부터 모니터를 하는 것과 면접원의 행동의 표준화를 통하여 면접원 분산을 줄일 수 있는 가능성으로 인식 되었다. 또한 중앙집중식 전화면접이 대면면접보다 면접하는 직원간에 접촉할 더 많은 기회가 제공되기 때문에 면접행동에 더 많은 동질성이 생기는 것으로 믿어진다.

대면면접에서는 면접원의 업무수행을 평가하기 위하여 감독자가 면접원과 동행하는 특수한 상황을 제외하고는 일반적으로 면접을 모니터 하거나 관측하지 않는다. 대면방법은 응답자와 면접원의 상호작용을 장려하는데 이것은 편향될 수 있다. 또한 전달이 청각적 이외에 시각적일 수 있으면 전화방법에서 보다 응답에 면접원 영향의 가능성이 더 크다. 무의식적인 몸짓(body language), 얼굴 표정과 다른 몸짓들은 응답자에게 많은 것들을 전달할 수 있다.

그러나 면접원의 작업량은 전형적으로 대면면접에서 보다 전화조사에서 더 많다. 모든 첫 번째 면접을 대면면접으로 하는 CPS는 면접원 작업량은 대략 50가구이다. 이 조사에서 평균 대면 값으로 $\rho_{int}=0.03$ 을 가정하면 CPS는 평균 $deff_{int}$ 가 약 2.5이다. 중앙집중식 전화면접에서 $\rho_{int}=0.01$ 을 가정하고 2.5의 $deff_{int}$ 에 대응하는 평균전화 면접원 작업량을 계산해 보면 150개의 면접이 나온다(즉, $m=50$ 이고 $\rho_{int}=0.03$ 인 대면조사의 $deff_{int}$ 는 $m=150$ 이고 $\rho_{int}=0.01$ 인 중앙집중식 전화조사의 $deff_{int}$ 와 동일하다). 그래서 면접원당 150개 이하의 면접원 작업량이 있는 전화조사가 50개 이상의 평균 작업량이 있는 대면조사보다 일반적으로 분산에 더 작은 면접원 기여를 한다.

면접원 오류에 의해 조사 추정값의 정확도가 극적으로 줄어들 가능성이 있다는 사실에 근거하여 조사방법론자들은 면접원 오류의 원인, 면접원 영향을 가장 받기 쉬운 응답자와 조사상황 종류, 면접원 교육과 감독, 모니터와 평가 이외에 더 나은 설문지 설계를 통해 면접원 영향을 줄이고 통제하는 방법을 알아내는데 많은 노력을 했다. 다음절에서는 면접원 효과에 영향을 미칠 수 있는 몇 개의 설계 요인을 알아본다.

5.3 면접원 효과에 영향을 미치는 설계 요인들

면접원 오류의 원인을 확인하고 그것을 통제하는 것을 목표로 특정한 설계요인을 면접원 효과와 관련지으려고 시도하는 몇 가지 연구가 있다. <그림 5.3>에서 보여 주듯이 이 연구는 조사설계의 네 가지 일반적인 분야(면접원의 특징/행동, 응답자의 특징, 설문지와 일반적인 조사 상황과 환경)에 집중되었다. 즉, 이 절에서는 문헌에서 발견한 것들을 요약하고 현지 조사 작업 계획과의 관계를 이야기한다. 그러나 면접원 효과의 원인과 영향에 대한 연구는 부족하고 이론적으로 면접원 오류에 기여하는 많은 요인들은 탐구되지 않았다. 이런 경우에 우리는 이런 요인들의 가능한 기여에 대하여 추측하지만 그것들의 중요함에 대해서는 아무 증거도 제시할 수 없다.

- 면접원**
- 특성(나이, 인종, 성별, 교육)
 - 외모
 - 동기
 - 믿음/태도
 - 지각
 - 예상
 - 행동
 - 기술
 - 지식

- 응답자**
- 특성(나이, 인종, 성별, 교육)
 - 지식
 - 관심/동기부여
 - 신뢰
 - 신념의 강도
 - 예상

면접원 오류

- 설문지**
- 명백한 정의
 - 전문용어/특수용어
 - 질문 형태
 - 지시
 - 질문 표현법
 - 질문 주제

- 조사 상황과 환경**
- 면접 방식
 - 표준화
 - 면접원 교육
 - 면접원 감독
 - 모니터/관측

<그림 5.3> 면접원 영향을 설명할 수 있는 설계 요인들

5.3.1 면접원 효과와 면접원과 응답자의 특징

인구통계학적인 특징

면접원의 특징과 행동이 면접원 효과에 영향을 미친다고 생각된다. 예를 들어, 응답자에 따라서는 면접원의 나이, 인종, 성별, 사회 계층과 교육 수준이 어떤 질문들의 응답에 영향을 미칠 수 있다. 예를 들어 성숙한 면접원은 그들 자신이 성숙한 응답자들보다는 어린이 응답자에게 더 큰 영향을 미칠 수 있다. 마찬가지로 대졸의 훌륭한 교육을 받은 것으로 보이는 면접원들이 교육을 받지 못하고 지식이 많게 보이는 면접원들 보다 정식 교육을 받지 못한 응답자에게 더 큰 영향력을 미칠 수 있다. 이것은 특히 면접원이 확실한 견해를 가지고 있는 의견과 태도에 대한 질문일 때 해당된다. 그렇지만 두 가지 가설 어느 것도 논문에서 입증된 적은 없다.

인종은 항상 중요한 면접원 특징일 것이다. 면접원과 응답자의 인종이 일치 되었을 때의 응답 패턴은 그렇지 않았을 때와 다른 결과를 가져온다는 것이 몇 건의 문헌에 나타나있다(Williams, 1964; Schuman and Converse, 1971 참조). 하지만 이 효과는 조사의 주제가 민감한 경우에만 중요하다는 증거가 있다. 면접원의 인종의 가장 큰 효과는 인종 문제를 다루는 질문에서 볼 수 있다(Hatchett와 Schuman, 1975; Wilson과 Olesen, 2002 참조).

질문이 성별의 역할에 대한 것일 때에는 면접원의 성별이 응답 패턴에 영향을 줄 수 있다는 증거를 제시한 논문도 있다. 예를 들어, Nealon(1983)은 농장 여성의 조사에서 성별 효과를 조사했는데 일에 대한 참여도, 농장에서 내리는 결정, 재정적인 타협, 농장 프로그램의 만족도에 대한 질문들에서는 남성과 여성 면접원에 따라 응답 패턴이 다르다는 결론을 내렸다. 여성의 권리와 역할에 대한 질문에서는 여성들이 여성보다 남성 면접원에게 더 페미니스트적인 응답을 했다(Ballou and de Boca, 1980). 또한 나이가 든 응답자들에게서는 면접원 변동이 더 크다는 증거도 있다.

이 결과들과 다른 연구들에서 면접원과 응답자 특성의 상호작용에 관련된 패턴이 제시된다. 조사의 주제가 특징에 관련된 것이라면 응답효과가 더 잘 드러날 것으로 예상된다. 한 예로, 10대 청소년의 성에 대한 견해에 대한 조사에서 더 젊은 면접원과 나이가 든 면접원은 서로 다른 응답을 이끌어낼 것으로 생각할 수 있다. 마찬가지로 부유한 사람에 대한 사고방식에 관련된 조사에서 비싼 옷을 입고 고가의 차를 운전하는 면접원은 훨씬 더 수수하게 보이는 면접원과 매우 다른 응답을 얻을 수 있다. 그러므로 자료수집원을 고용할 때 허용되는 범위내에서 면접원과 응답자의 시각적인 특징을 (나이, 인종, 성별, 사회 경제적인 지위) 조화시키는 것이 대부분의

면접에서는 좋은 전략일 것이다. 그 이유는 이러한 전략을 사용했을 때 응답들이 더 정확하다는 여러 가지 증거에 근거한다.

면접원 기대

연구자들은 면접원들이 응답자의 예상되는 대답에 대해 가지고 있는 기대와 특정한 질문에 답을 할 마음이 있는지 또는 능력이 있는지에 대한 기대감이 조사 응답에 큰 영향을 미친다 생각한다. 응답자가 질문에 부정적으로 반응할 수 있다는 면접원의 염려가 면접원이 그 질문을 고쳐 말하던지 간단히 그것을 건너뛰게 할 수 있다. Singer and Kohnke-Aguire(1979)의 연구가 이러한 가설을 시험했는데 질문의 어려움에 대한 면접원 기대와 응답자 행동 사이에 강한 관계를 찾지는 못했다. 비슷한 연구들도 면접원 기대 효과에 대하여 매우 약한 증거만을 찾았다. 면접원의 기대, 면접원 행동과 질문 주제 사이의 관계는 알려진 것이 별로 없다.

예를 들어, 응답자의 대답에 대한 면접원 기대가 일치성 없는 탐색조사의 원인이자. 응답자의 상황에 대한 선입관과 조사의 다양한 측정에 대한 상태를 근거로 면접원은 언제 철저한 조사를 할지 또는 설문지 개념을 명확하게 할지를 결정할 수 있다. 예를 들어, 만일 조사의 주제가 범죄 희생에 대한 것인 경우 절도 대한 질문의 “아니오”라는 답은 부유한 지역에서보다는 우범 지역에서 면접원이 더 엄밀한 탐색조사를 하고 싶어질 수 있다. 이처럼 전적으로 면접원의 재량에 의한 일치성 없는 탐사행동은 면접원 분산의 중요한 원인으로 믿어진다. 그렇지만 지금까지는 이 효과에 대한 체계적인 연구가 없었다.

경험이 많은 면접원들이 경험이 별로 없는 면접원들 보다 더 높은 단위 응답률을 얻는다는 증거가 있다. 그렇지만 항목 무응답에 대해서는 그 반대가 사실일 수 있다. Stevens and Bailer(1976)는 경험이 많은 면접원과 경험이 적은 면접원들의 CPS의 항목 무응답률을 비교했는데 소득 질문과 같은 질문에서는 경험이 적은 면접원의 빠진 자료 비율이 다소 낮다는 것을 알아냈다. 이것은 경험이 많은 면접원들은 이런 질문들에 대해 응답자들이 부정적으로 반응할 것이라는 기대의 결과로 질문을 건너뛰거나 거부를 너무 일찍 받아들인다. 경험이 별로 없는 면접원들은 이런 질문에 대한 응답자 반응에 관련해 더 순진하고 그래서 응답을 얻으려고 더 많은 노력을 할 수 있다.

무응답에 미치는 면접원 효과

면접원들이 조사의 응답률에 (협조률과 접촉률에도) 상당한 영향을 미칠 수 있다는 명백한 증거를 제시한 조사방법 문헌이 있다. Groves and Couper(1998)는 가구 조사에서 무응답에 미치는 면접원의 영향에 관련된 문헌의 포괄적인 검토를 제공한다. 그들의 분석은 다음 결론들을 얻었다.

- 경험이 많은 면접원이 경험이 적은 면접원보다 더 높은 협조률을 얻는 경향이 있다.
- 협조를 얻는 자신의 능력에 대하여 확신을 가진 면접원들이 더 높은 협조률을 달성하는 경향이 있다.
- (특히 전화면접에서) 조사에 대한 소개를 표준화하는 것은 협조를 얻는데 효과가 없다. 대본에 없고 가구원의 거부나 관심에 대하여 적응성 있는 소개를 하는 것이 더 효과적인 경향이 있다.

이것에 대한 타당하고 뚜렷한 설명들이 있다. 경험이 많은 면접원들은 설득력이 있고 시간을 효율적으로 활용하는 능력이 있을 것이다. 자신감을 가진 면접원들은 3장에서 언급한 조사 참여의 6가지 심리적인 원칙을 효율적으로 잘 적용할 수 있을 것이다. 그 원칙들은 보상, 일치성, 사회적 확인, 권위, 희소성, 유대이고 특히 후자의 원칙에 대해서 더욱 그렇다. 융통성 있는 소개는 면접원이 조사 참여의 요청을 개인의 특수한 관심이나 유보에 맞추어 할 수 있도록 해준다. 우리가 3장에서 배웠듯이 이러한 융통성은 초반의 참여하기를 꺼려함이 확고한 반대나 거부로 가는 것을 막아 준다. 면접원들은 경험, 자신감, 사회적 능숙함 (adroitness)에 있어서 상당히 다르기 때문에 조사에서 단위 무응답에 대한 면접원 분산이 보통 상당히 크다.

Bailer 등(1977)은 1970년 U.S. Census of Population and Housing에서 조사의 항목 무응답에 대한 면접원 분산의 증거를 찾았다. 전에 언급했듯이 다른 연구에서 면접원의 경험에 따라 항목 무응답에 차이가 있는 것을 보았다. 이것은 면접도구에 익숙해짐에 따라 줄어드는 건너뛰는 오류나 “슬쩍 지나침(slips)” 때문이다. 응답자로부터 전형적으로 부정적 응답을 이끌어내는 특정한 질문들을 묻기를 싫어하는 경험 많은 면접원들의 행태도 그러한 이유가 된다. 만일 면접원들이 조사에서 그 질문들을 해야 할 시점에 그 질문들이 “부적당”하거나 “전후 관계를 무시”한다는 느

낌이 들면 면접원들이 때로는 질문을 한두개 건너뛰는 증거가 스웨덴의 조사에서 발견된다. 그들이 면접하다가 나중에 다시 물어보려고 하지만 보통 잊게 된다.

CATI와 CAPI같은 컴퓨터 지원 면접 (CAI; computer-assisted interviewing) 방법들은 한 질문에서 다음으로 옮겨가는 것이 자동화되었기 때문에 전자의 오류가 줄었다. 현재의 질문에 대한 입력을 하기 전에는 CAI에서는 다음 질문으로 나아가는 것이 불가능하기 때문에 후자의 효과에서도 어느 정도는 같은 것이 사실이다. 그렇지만 면접원들은 만일 그들이 틀린 항목을 입력하기를 꺼리지 않으면 면접원들은 여전히 질문을 하는 것을 거절할 수 있다. 항목 무응답과 달리 면접원의 이런 행동을 찾아낼 가능성이 적다. 5.4절에서는 면접원 변조(falsification)를 찾아내는 방법들을 이야기한다.

5.3.2 면접원 효과와 설문지 설계

1920년대 이전에는 면접원 변동은 작물의 생산량 연구, 주택 평가와 그밖에 유사한 관측 연구들과 주로 관련되어 있었다. Rice(1929)는 사회적인 문제에 대해 매우 다른 의견을 가진 두 면접원들이 관여한 빈민들의 연구에서 면접원 효과를 보고했다. 주류 제조 판매 금지주의자인 한 면접원은 그들의 빈곤을 음주의 탓으로 하는 응답자들을 찾는 경향이 있었다. 사회주의자인 다른 면접원은 당시의 사회와 경제 상태를 탓하는 빈곤한 사람들을 찾는 경향이 있었다. 그 이후로 연구자들은 면접원의 행동, 매너리즘, 외모, 음성이 응답자에게 영향을 미칠 가능성이 있다는 증거를 찾았다. 면접원이 질문을 할지 또는 어떻게 할지, 분명한 설명을 요구할지, 피드백을 제공할지, 응답자의 응답을 어떻게 해석할지에 대한 면접원에 따라 다른 결정은 자료를 바꿀 수 있다. 면접원과 응답자와의 관계와 면접원과 응답자사이의 상호작용도 응답에 영향을 미칠 수 있다. 이러한 점에서 표준화된 도구와 면접 절차 대폭적으로 사용하게 했다. 표준화의 근본적인 목적은 응답자와 면접원사이의 상호작용을 동일하게 하려는 시도를 통하여 면접원의 태도, 믿음, 행동과 매너리즘이 조사 응답에 영향을 미치는 것을 제한하는 것이다. 그렇지만, 면접원 변동의 가장 중요한 원인은 면접원 그 자신이 아니라 설문지 설계, 조사 질문과 관련된 질문들이라고 상당한 증거를 제시한 문헌이 있다.

인구 통계학적인 특징과 잘 정의되고 응답자가 쉽게 접근할 수 있는 사실에 입각

한 정보를 요구하는 명확하고 혼동이 되지 않게 써진 다른 항목에 대해서는 낮은 ρ_{int} 값을 기대할 수 있다고 증거를 제시한 문헌이 있다. 태도나 의견에 대한 질문, 서술형 질문, 민감하거나 감정적인 질문과 면접원이 명백하게 하기 위해 엄밀한 탐색조사를 할 필요가 있는 어려운 항목들에서는 높은 ρ_{int} 값이 나온다. 그러나 문헌에 있는 조사 결과는 일치성이 없고 이러한 일반적인 지침은 설문지와 면접원 사이의 훨씬 더 복잡한 상호작용을 지나치게 간소화하는 경향이 있다. 대화 면접에 대한 문헌이 이 문제를 더 밝힐 수 있을 것이다.

Schober와 Conrad(1977)는 표준화 면접의 한 가지 문제점은 응답자의 상황이 질문에서 말한 정의에 쉽게 “대응” 않는 경우가 있다고 하였다. 이런 경우에는 표준화된 면접원은 면접 원고에서 벗어나서 응답자가 질문에 답하게 도와주도록 원고에 없는 설명과 지원을 주도록 압력을 받을 수 있다. 이런 것을 잘하는 면접원과 잘하지 못하는 면접원이 있는데 대부분의 표준화된 면접원들의 그러한 행동을 전혀 하지 않고 응답자가 이런 차이들을 그들 자신이 해결하도록 하는 표준화된 방법을 따른다.

한 실험에서(Schober과 Conrad, 1997) 41명의 응답자에게 다양한 상황을 설명하는 시나리오(또는 삽화(vignettes))를 주고 표준화 면접과 대화 면접 모두를 사용해서 시나리오에 대하여 면접을 했다. 시나리오의 반에서는 시나리오에서 설명한 상황과 질문에서 사용한 용어의 운영상 정의(응답자들이 이것을 직관적으로 알지 못했다)는 상당히 쉽게 대응되었다. 다른 반에서는 대응이 훨씬 더 복잡했다.

예를 들어, 응답자들에게 건축 평면도를 보여주고 욕조가 없는 욕실이 이 집에 몇 개 있습니까?” 라고 물었다. 방에 두 개의 시설(변기, 세면대)이 있을 때에는 어느 면접 방법에서도 응답자는 그것을 욕조가 없는 욕실로 알아보는 데에는 어려움이 없었다. 하지만 방에 그러한 시설이 한 개 밖에 없을 때에는 두 방법 사이에 옳은 응답의 수에 두드러진 차이가 있었다. 진행 중인 미국 정부 조사에서의 선택된 12개의 질문들에 걸쳐 시나리오의 대응이 합리적으로 정의되면 표준화 면접과 대화 면접의 정확도가 모두 거의 완벽한 97퍼센트와 98퍼센트였다. 그러나 대응이 복잡하면 표준화 면접의 정확도 28퍼센트는 대화 면접 방법의 87퍼센트에 비해 겨우 28퍼센트로 떨어졌다.

이 결과들은 면접원 분산이 운영상의 정의로 대응 시키는 것이 상당히 복잡한 상황을 면접원이 다루는 방법에 대한 변동이라는 것을 말한다. 면접원 분산은 추정량

의 분산을 늘리지만 응답에 더 높은 정확도를 달성하는 면접 방법들은 평균제곱오차의 편향요소를 상쇄하는 긍정적인 효과가 있을 수 있다. 그러나 어떤 절차가 상당한 면접원 분산을 야기할 것이라는 우리가 알고 있음에도 불구하고 그것들이 더 낮은 응답편향을 가질 것이라는 기대를 갖고 받아들여서는 안 된다. 오히려 더 나은 설문지 설계, 면접 방법과 교육 과정이 편향과 분산 모두를 관리하는데 필요하다. 많은 경우에, 면접원들이 조사 정의를 적용하는데 부닥치는 많은 복잡한 상황을 예상하는 것이 가능하지 않을 수 있다. 그런 상황에서는 응답자가 운영상의 정의를 특별한 상황에 적용하게 도울 수 있는 더 융통성 있는 면접 방법이 최고의 방법일 수 있다.

5.3.3 면접원 효과와 조사 환경

면접원 효과의 규모에 잠재적으로 영향을 미칠 수 있는 조사 설계와 관련된 몇 가지 다른 요인들이 있다. 그것들에는 면접 방법; 고용, 교육, 감독, 면접원 업무수행을 모니터하는 방법과 면접하는 방식 (예를 들어, 표준화, 대화 또는 혼성 면접 방법); 일반적으로 조사 환경이라고 부르는 프라이버시와 조사에 대한 주의 깊은 자세와 같은 면접을 하는 상황들을 나타내는 요인들이 있다. 이 장에서는 이러한 요인들이 면접원 오류에 미치는 영향들에 관해서 생각한다.

앞에서 면접 방식이 면접원 분산에 미치는 영향에 대해 이야기했다. 또한 전화와 대면 면접이 면접원 오류에 미치는 영향도 다루었다. 항목 무응답에 대한 면접원 효과를 거론하면서 CAI가 면접원 오류에 미치는 영향도 어느 정도 다루었다. 면접에서 컴퓨터가 건너뛰는 오류를 사실상 제거했고 그러므로 건너뛰는 오류 때문에 생기는 면접원 변동을 제거했다고 언급했다. 그렇지만 CAI 면접원의 기록의 정확도와 응답의 입력에서 생기는 오류 때문에 발생하는 랜덤 오차 문제에 대한 염려가 남아있다. 면접원에 의한 입력과 컴퓨터 속달은 면접원에 따라 다르므로 기록 오류도 면접원 분산의 원인이 될 수 있다.

여러 연구에서 CATI와 CAPI 면접원 모두에 있어서 자료 입력(keying)은 조사에 있어 오류의 중요한 원인은 아니라고 하였다. Lepkowski 등(1998)은 기록의 정확도는 지필면접(PAPI) 보다 CAI가 더 높다는 증거를 보여 주었다. 예를 들어 CAPI 면접원이 입력에서 범한 오류의 수는 PAPI 면접원이 응답을 적으면서 범한 오류의

수 보다 훨씬 적다. Tourangeau와 Smith(1998)는 몇 개의 민감한 질문의 측정에서 PAPI 면접원들 보다 CAPI 면접원들의 ρ_{int} 값이 더 작은 것으로 보고했다. 컴퓨터화를 통하여 면접원에게 특정 수준의 표준화를 하도록 하여 면접원 효과를 줄이는 듯하다. 그렇지만 우리가 알고 있는 한에서는 CAPI와 PAPI 개별면접의 면접원 분산을 비교하는 연구는 없었다.

Couper(1996)의 논문은 CPS에서 지필면접에서 CAPI로 바꾸는 것이 면접의 상황에 미치는 영향과 그 변화가 자료 품질과 비용에 미치는 영향을 조사했다. 그는 CAPI로의 변화가 CPS에서 실시한 문전 면접(즉, 면접원이 문밖에 서서 실시하는 면접)의 수에 큰 영향을 미쳤다는 것을 알아냈다. 그는 CAPI 면접은 응답자의 집안에서 노트북 컴퓨터를 놓고 전기사용이 가능한 콘센트 근처에 있는 탁자에 면접원이 앉아서 실시할 가능성이 훨씬 더 크다는 것을 알아냈다. 한 예로 CPS에서 지필 면접의 58%가 응답자의 집안에서 실시된 데 반해서 CAPI 면접의 75% 이상이 이런 환경에서 실시되었다.

문전 면접은 덜 만족스러운 자료 품질과 관련지어 생각할 수 있다. 예를 들어 면접시간이 짧아지는 경향이 있는데 이것은 응답자와 면접원이 집안에서 편안하게 앉아있을 때보다 문 앞에서는 더 서두르게 될 수 있다는 것을 말한다. 이것은 면접을 끝내기 위하여 즉석에서 나오는 응답이나 더 작은 성과로 만족하는 응답이 나올 수 있다 (Krosnick, 1991). 면접원이 문앞에서 노트북 컴퓨터를 들고 있으면 응답자나 면접원 자신의 업무수행보다 컴퓨터의 성능에 대하여 더 걱정할 수 있다(배터리가 방전되고, 기계를 드는데서 오는 피로, 입력하거나 화면을 보는 것에 대한 어려움).

그렇지만, 집안에서 앉아서 면접하는 것의 요구는 결점이 될 수도 있다. 응답자들은 면접원이 집안에 들어오는 것을 거절할 수 있고 이것은 단위 무응답이나 면접을 “갑자기 그만둘” 수 있다. 예를 들어, Couper는 지필면접 보다 CAPI 면접의 거부야간 더 많았다고 한다. 그렇지만 이 증가의 얼마가 면접원 환경 때문인가는 확실치 않다. 또한 집에서 실시한 면접은 더 오래할 가능성이 있고 그 결과 조사기관에서 면접원 비용이 더 든다.

Cooper는 CAPI가 일으킨 조사여건의 변화가 자료품질을 개선했다고 말하기가 어렵다고 결론을 내린다. 왜냐하면 그것은 어쩌면 더 낮은 응답 오류와 더 높은 무응답오차와의 교환에 따라 결정되기 때문이다. 그래도 아직 그의 분석은 기술의 변화

가 조사자료의 수집과정과 자료품질에 미치는 영향을 고려하는 것의 중요함을 설명한다.

Fowler와 Mangione(1985)은 면접 기간이 ρ_{int} 값에 미치는 영향에 관련해 중요한 연구 결과를 얻었다. 그들은 네 가지 교육기간(½, 2, 5, 10일)에 대해 조사를 하였다. 가장 높은 ρ_{int} 는 가장 짧거나 가장 긴 교육기간을 거친 면접원들에게서 추정되었다. 후자 집단의 발견에 대한 한 가지 그럴듯한 이유는 교육의 결과에서 오는 과잉 자신감이다. 면접원들은 표준화 면접 방법으로 교육을 받고 있었지만 긴 기간의 교육이 실은 표준화에 거꾸로 작용했을 수 있다. 5.1절에서 보았듯이 응답자의 상황을 운영 정의로의 복잡한 대응을 필요로 하는 경우에는 교육을 더 잘 받은 면접원들이 응답하는 일에 더 개입해서 응답자들이 답을 찾는데 도움을 주는 경향이 있었을 수 있다.

<표 5.4>를 보면 Fowler-Mangione의 연구에서의 나온 추가적인 몇 가지 사실을 알 수 있다. 불충분한 교육의 좋지 않은 결과는 표에서 뚜렷이 나타난다. 충분한 교육을 받지 못한 면접원들은 질문들을 쓰인 대로 읽고, 탐색조사를 하고, 답을 기록하고, 중립의 대인 관계 행동을 유지하는 것 같은 면접의 기본적인 역할을 하는데 준비가 충분하지 못하다. 이 모든 역할은 부가적인 교육으로 개선될 수 있다. 탐색하는 행동은 면접원들이 질문에 내포되어 있는 연구 목적에 관련된 더 많은 지식을 습득하는 긴 교육으로 개선 될 수 있을 것이다.

<표 5.4> 교육 기간이 면접원 업무수행 측정에 미치는 효과^a

면접원 행동	교육 프로그램 기간			
	½일	2일	5일	10일
질문을 쓰인 대로 읽는 것	30	83	72	84
선택형 질문의 탐색조사	48	67	72	80
서술형 질문의 탐색조사	16	44	52	69
선택형 질문의 기록	88	88	89	93
서술형 질문의 기록 ^b	55	80	67	83
편향되지 않은 대인 관계의 행동	66	95	85	90

출처: Fowler와 Mangione(1990), 표 7.4.

^a우수하거나 만족되게 평가된 면접의 퍼센트다.

^bn.s., all others $p < 0.01$.

그들의 연구에서 부가적으로 발견된 것은 다음과 같이 요약할 수 있다.

- 최소한의 교육을 받은 면접원들도 교육을 더 잘 받은 면접원들과 마찬가지로 협력을 얻거나 응답자와 마음을 통하는 것 같은 대인 관계의 면은 잘 처리할 수 있다. 이런 능력의 일부는 교육을 통해서보다는 경험으로 늘릴 수 있다는 사실 때문이라고 저자들은 제안한다.
- 더 많은 교육을 받은 면접원들은 직무 지향적이고 교육을 덜 받은 면접원들은 대인관계에 지향적이었다.
- 너무 많은 교육은 비생산적일 수 있지만 “너무 많은”의 정의는 조사에 따라 다르다.

5.3.4 현지 작업의 실제 결과

조사 면접의 세 가지 요소인 면접원, 응답자, 설문지 모두가 면접에서 일어나는 일과 결과로 나오는 자료를 결정한다. 그렇지만 면접원이 응답자에게 조사 질문들을 말하는 방법과 설문지에 최후로 기록되는 것은 전적으로 면접원에게 달려 있다. 면접원이 설문지 원고에 엄격히 충실해야 한다는 전통적인 관념은 설문지와 교육 지침이 가능한 모든 면접 상황을 적용된다는 것을 가정한다. 그렇지만 면접원들이 직면할 가능한 모든 특수한 경우를 예상한다고 가정하는 것은 현실적이지 않다. 설사 계획될 수 있다 할지라도 모든 면접원이 무수한 상황을 규정된 같은 방식으로 다룰 수 있도록 확실하게 하기 위해 필요한 교육은 비용면에서 실현 가능하지 않다. 이런 이유로 교육 지침은 면접원들에게 행동의 원칙을 주고 (예를 들어, 질문들을 쓰인 그대로 읽은 다음에 중립적이고 지시적이지 않은 방식으로 한다) 그들이 그 원칙들을 일어나는 다양한 상황에서 적용하게 한다.

이러한 접근 방법의 위험성은 면접원들이 설문지 개념을 제시하고 조사 응답자와 상호 영향을 미치는 방법의 차이의 결과 면접원 오류가 늘어날 것이다. 그렇지만 이런 해로운 결과로부터 보호하기 위하여 몇 가지 전략을 채택할 수 있다. 권고 할 만한 방법은 다음과 같다.

- 뛰어난 대인 관계 능력, 설득력, 조직력이 있고, 조사할 주제가 주어졌을 때 눈

에 보이는 면접원의 특징이 그들에게 배정된 표본 구성원의 특징과 잘 어울리는 면접원을 모집하라.

- 가능하면 항상 경험이 많은 면접원을 고용하라. 경험이 별로 없는 면접원을 쓰면 교육에서 그들이 접할 많은 상황에 면접 방법을 적합시키고 맞추는 것에 대한 수업과 연습 기회를 제공해야 한다. 이것은 응답자한테서 협조를 얻을 때에는 특히 그렇다.(면접원 교육방법에 대해서는 Groves와 McGonagle, 2001, Campanelli 등, 1997a.를 참조하라).
- 자신감과 조사에 대해서 긍정적인 태도를 가진 면접원들을 자신이 없어 보이고 조사가 성공하지 못할 거라고 느끼는 면접원들보다 더 선호된다. 어떤 경우에는 이런 태도들을 교육하는 동안 그리고 근무 중에 피드백할 때 생길 수 있다.
- 면접원을 교육시키는 것은 면접의 기술 이외에도 조사 질문에 내포되어 있는 개념과 목표를 포함해야 한다. 탐색조사의 기술도 어느 정도 깊이로 다뤄야하고 교육 기간은 면접원이 다양한 응답 상황에서 탐색 조사하는 것에 숙달되도록 충분해야 한다.
- 면접원이 조사하는 일에 충분한 숙달을 보여줄 때까지 면접원 업무수행을 처음에는 면밀히 모니터 해야 한다. 그 후에는 업무수행을 모니터하는 것은 자주하지 않아도 다. 가능한 빨리 면접원에게 모니터한 결과에 대한 긍정적인이거나 부정적인 피드백 모두 주어야 한다.

다음 절에서는 면접원을 모니터하고 평가하고 그들의 면접 업무수행에 대한 피드백을 제공하는 방법들을 다룬다.

5.4 면접원 업무수행 평가

면접원이 된 모든 조사에서 현지 작업을 하는 동안 면접원 업무수행에 대한 평가는 필수적이다. 이것은 면접의 품질을 관리하기 위해서 뿐만 아니라 설문지나 면접원 지시와 같은 조사과정의 다른 영역에 잠재하는 문제를 확인하기 위함이다. 면접원 업무수행 평가는 면접원의 직무에 대한 다양한 양상에 대한 자료와 정보를 수집하고, 이 자료를 평가하고, 이 평가를 토대로 어떤 행동을 취하는 것이다. 평가는

보통 다음의 네 개의 업무수행 분야를 목표로 삼는다:

1. **왜곡된 정보의 발견과 예방**(*Detection and prevention of falsified information*). 모든 조사 관리자는 면접원에 의해 자료가 왜곡되게 전해지는 위험성에 대해 준비하고 그것을 발견하거나 방지하는 방법을 찾아야 한다. 왜곡(*falsification*)은 조사방법론에 관한 문헌에서 위조(*fabrication*), 속임수(*cheating*), 테이블 꾸미기 (*table-topping*), 면접원 위조(*curbstoning*)라고도 부른다(마지막 용어는 면접원이 표본 가구 단위 밖의 보도 가장자리에 앉아서 설문지를 위조하는 모습을 근거로 한 것이다). 이런 종류의 오류는 면접 전체를 거짓으로 꾸며내거나, 어떤 질문들을 고의적으로 건너뛰거나, 가구 단위가 비어 있는지 거주하는지와 같은 정보를 그릇되게 전하는 것을 말한다. 면접원들이 왜곡을 할 이유는 많다. 아마도 조사에서 높은 응답률을 달성하거나 현지 작업을 촉진시키는 압력이 너무 클 것이다. 면접원들은 그들이 할당받은 일을 적절히 수행하는데 방해가 되는 개인적인 문제가 있을 수 있고 일을 끝내는 방법으로 사기(부정한 방법)에 의지할 수 있다. 험한 지역에 들어가기 두려워하는 면접원들은 면접을 지어낼 수 있다. 만일 면접원의 부정한 행위가 광범위하다면 조사 기획 전체의 결과를 무용지물로 만들 수 있다. 그런 이유로 면접원 왜곡은 어떠한 이일이 있더라도 관대히 취급해서는 안 된다(6장 참조).

2. **교육에서 설명한 면접규칙과 지침의 이행**(*Compliance with the rules and guidelines for interviewing set forth in training*). 면접지를 관리하기 위한 지시와 지침에 대한 이행이 되지 않으면 면접원 분산이 증가한다. 어떤 면접원들은 지시를 충실하게 따르고 어떤 면접원들은 그러지 않는다. 그 결과 자료의 품질은 면접원에 따라 달라진다. 이미 앞에서 보았듯이 이것은 응답자의 상황이 설문지의 운영 정의에 쉽게 들어맞지 않을 때 문제가 될 수 있다. 만일 면접원들의 불이행이 만연되어 있다면 면접원들 보다 지시와 지침이 잘못되었을 수 있다. 설계가 잘못된 절차는 면접원과 응답자에게 똑같이 부담이 되고 면접원들은 절차를 “고치려고” 시도 할 수 있다. 만일 그것이 사실이라면 절차를 수정해야한다.

3. **관리 활동같은 면접이외의 업무수행**(*Performance on noninterview tasks such as administrative activities*). 면접원들은 면접을 준비하고, 면접의 결과를 기록하고, 표본추출을 위해 현지에서 정보를 수집하는 등 면접 이외의 많은 일을 행

한다. 이 일들은 조사 결과의 정확도에 직접적인 관련은 없지만 조사자료 품질에 간접적으로 영향을 미칠 수 있으므로 평가와 함께 면접원에게 피드백을 해 주어야 한다. 예를 들어, 전화 면접원들은 각 통화의 결과를 기록해야한다(통화 중, 벨 울렸지만 받지 않음, 약속, 거절, 등). 이 과정에서의 오류는 응답률, 통화 시도의 성공률, 부적격률 등과 같은 편향된 과정 통계량을 제공할 수 있다. 현지 면접원들은 비용을 최소화하고 그들의 가정 방문의 효과를 늘리기 위해 그들의 경로를 계획해야 한다. 이 과정에서의 비능률적인 점이나 오류는 비용을 늘릴 수 있고 응답률까지도 줄일 수 있다.

4. 면접원과 설문지간의 의사소통 문제의 확인(*Identification of problems in the interviewer-questionnaire interface*). 종이와 연필을 사용하거나 또는 컴퓨터를 사용하거나 조사 설문지 설계의 한 가지 목적은 면접원과 설문지간의 인간공학적 의사소통을 제공하는 것이다. 조사 도구와 관련된 유용성 (usability)은 면접원들이 PAPI, CATI, CAPI 또는 다른 설문지와 얼마나 쉽게 서로 작용하나를 말한다. 만일 명령들이 복잡하고, 혼동되고, 사용하기 불편하면 면접 품질은 나빠질 것이다.

이런 면접원 업무 수행 분야들을 평가하기 위해 조사관리자가 할 수 있는 몇 가지 방법이 있다. <표 5.5>에서는 이 방법들과 이 방법들이 목표로 삼는 업무 성과 분야를 수록했다.

모든 면접원 업무 수행 평가 방법의 한 가지 공통된 설계 특징은 경험이 적은 또는 새 면접원을 중점을 두는 것이다. 예를 들어, 면접원들은 조사의 후반부에서 보다 조사가 착수된 시기에 더 자주 평가를 받는다. 새로 고용된 면접원은 경험에 많은 면접원 보다 더 자주 평가받을 수 있다. 그렇지만 모든 면접원들을 그들의 수행 기간이나 이미 증명된 기능이나 능력과는 관계없이 평가하는 것이 흔히 있는 일이다. 면접원 업무수행과 경험 사이에 곡선의 관계가 있다는 증거가 있다. 예를 들어, 항목 무응답과 비축어적인 질문을 하는 것은 경험이 많은 면접원들 사이에 더 많을 수 있다. 이 때문에 면접원 평가와 개선은 계속되는 과정이어야 한다.

<표 5.5> 면접원 업무 수행 평가 방법

평가 방법	주요 성과 분야 ^a			
	왜곡	면접 업무수행	면접외 업무	유용성
재면접	•	•		
확인 재접촉	•			
관측		•	•	•
음성 녹음(행동 코딩 있음/없음)	• (행동 코딩 없음)	• (행동 코딩 있음)		•
모니터	•	•		
설문지 검토	•	•	•	•
업무수행과 결과 측정	•	•	•	
키스트로크/ 추적 파일 분석	•			•
가상 면접/ 지식이나 연습 시험		•		•

^a 큰 점(•)은 그 방법의 가장 중요한 목적을 나타내고 작은 점(•)은 두 번째 목적이나 그 방법이 성과의 간접적인 성과의 측정을 제공한다. 표에 있는 사항의 일부는 아직 정의되지 않았지만 이것들은 이 장 뒤에서 이야기할 것이다.

면접원 업무수행을 개선하기 위해 설계된 또 하나의 방법은 면접원에게 피드백을 구두로 또는 인쇄된 피드백 용지를 사용해서 주는 것이다. 피드백이 효과적이기 위해서는 시기적으로 적절해야하고 개선이 필요한 분야이외에도 업무수행에 대한 긍정적인 메시지도 포함해야 한다.

5.4.1 재면접 조사와 확인 재접촉

재면접 조사와 확인 재접촉은 대면면접 업무수행을 평가하는데 널리 쓰이는 방법이다. 품질 평가와 관련해서 재면접 조사는 조사 응답자의 비교적 작은 확률표본을 선택해서 최초 면접 후 짧은 기간 내에 그들과 다시 접촉하여 최초 면접에 포함되었던 똑같은 주제들에 대하여 다시 그들을 면접하는 것을 의미한다. 재면접 조사는 어떻게 설계되었냐에 따라 비표본오차 요소, 면접원 업무수행 또는 모두에 대한 정보를 제공할 수 있기 때문에 조사의 비표본오차를 평가하는 중요한 방법이기도 하다. 8장에서는 재면접 조사를 통하여 비표본오차 요소의 추정에 대하여 생각해 본

다. 여기에서는 면접원 업무수행 평가에 대하여 이야기한다.

미국 센서스국은 1950년대부터 면접원 업무수행을 재면접을 사용해서 평가해왔다. 예를 들어, CPS에서 매달 면접원들을 랜덤하게 선택해서 그들의 할당된 일의 삼분의 일 정도를 재면접한다. 이 표본의 약75%는 신뢰성을 추정하는데 사용되지만 25%는 면접원 업무수행을 평가하기 위하여 설계된다. 이 작은 표본의 가장 중요한 목적은 면접원 위조(curbstoning)를 찾아내고 방지하는 것이다. Biemer와 Stokes(1989)가 보여줬듯이 이런 종류의 설계로 왜곡을 찾아낼 확률은 매우 작다. 그들의 할당된 일에 왜곡이 있는지를 조사 받을 것으로 알고 있는 면접원들은 위조할 가능성이 적기 때문에 이 방법의 실제 가치는 방지에 있다. 비용을 절약하기 위해 재면접은 가능하면 전화로 실시하고 대면면접은 마지막 수단으로 쓴다. 매우 낮은 응답률 때문에 재면접을 우편 설문지를 사용하는 시도는 성공적이지 않았다. 바르게 설계된 재면접 조사는 면접 전체를 꾸며낸 면접원들을 확인하는데 효과적인 방법이다. 최초 면접에서 했던 몇 개의 질문들을 다시 묻는 것에 더하여 응답자에게 직접 면접원이 면접을 실시했나를 물어볼 수 있다. 예를 들어, 재면접원은 이런 것들을 물을 수 있다:

- 면접 기간 동안 면접원이 전화를 하거나 방문했는지 여부
- 만일 그랬다면 통화나 방문의 기간
- 설문지에서 특정한 질문들을 했는지 여부
- 인센티브 제공 여부 (만일 조사에서 인센티브가 사용되었으면)

어느 정도까지는 재면접이 응답자 선택절차의 오류, 자기 보고를 요구했는데 대리인 보고를 받아들인 일, 전화 면접의 부적당한 이용과 거주하는 가구 단위를 비어있다고 잘못 분류하는 것 같은 다른 종류의 면접원 오류를 찾아내는데 사용될 수 있다. 더욱이 조정과 결합하면 재면접은 면접내에서 면접원 업무수행에 대하여 정보를 제공할 수 있다.

조정(*reconciliation*)은 조사 질문에 대한 재면접 응답을 최초의 응답과 비교하여 응답자의 모든 불일치를 해결하는 것을 의미한다. 조정은 어떤 질문들을 건너뛰고, 필요하고 적당할 때에 탐색조사를 하지 않고, 응답자에게 잘못된 정보를 전달하는 면접원들을 확인하는 효과적인 방법일 수 있다. 이 방법의 중요한 결점은 면접원이

응답자에 대한 보고를 흔히 거부하므로 면접 업무수행 문제점의 결정적인 증거를 거의 제공하지 않는다는 것이다. Forsman과 Schreiner(1991)는 면접원 성과 평가와 비표본추출오차 요소의 추정 모두를 위한 대안의 재면접조사 설계의 포괄적인 검토를 제시한다.

조사 기관에서 널리 사용되는 방법은 확인 재접촉이다. 재면접 조사처럼 확인 면접은 주 표본의 10에서 20%가 되는 작은 부차표본 면접을 주로 전화로 실시한다. 그렇지만 접촉의 목적은 단지 면접을 적절한 방법을 사용해서 적절한 응답자와 실시했다는 것을 검증하는 것이기 때문에 확인은 본래의 면접에서 했던 질문들을 다시 묻지는 않을 수 있다. 예를 들어, 확인 면접은 “우리 기관에서 면접원이 당신과 당신 택에서 지난주에 면접을 실시하였습니까?”를 물을 수 있다. 또한 이 접촉은 면접하는 동안 면접원의 품행(예의바름, 융통성, 조사에 대한 지식, 외관 등)에 대한 응답자의 견해에 대한 정보를 얻는데 사용될 수 있다.

패널조사와 반복하는 조사 설계들은 추가면접을 위해서 응답자에게 정기적으로 다시 연락하는 것을 필요로 한다. 예를 들어, CPS에서는 응답자와 4개월 연속으로 연락하고 그 다음 8개월은 접촉이 없고 다시 연속된 4개월의 추가 접촉이 있다. Survey of Income and Program Participation과 National Crime Victimization Survey같은 미국 패널조사는 덜 잦은 접촉을 요구한다. 면접원 업무수행을 평가하기 위하여 재면접을 사용하는 것에 대한 한 가지 염려는 재면접 접촉이 가져오는 응답자 추가적인 부담이 일부의 응답자들이 나중에 정기적인 패널조사 접촉을 하는 동안 참여하기를 거절하는 원인이 될 수 있다는 것이다. 하지만 미국 센서스국에서 실시한 연구에서는 재면접이 패널조사 응답률에 미치는 영향은 작거나 존재하지 않는다고 제시한다.

5.4.2 음성 녹음, 모니터링과 그 밖의 관측들

면접 업무를 수행하고 있는 동안 면접원들의 행동을 관측할 수 있는 몇 개의 방법이 있다. 대면 면접에서 면접원의 감독자가 면접원이 다양한 직무, 특히 면접을 어떻게 하는지를 관측하기 위하여 면접원의 임무 수행에 동반할 수 있다. 중앙집중식 전화 조사에서는 참견하지 않는 통화 모니터링이 일반적으로 사용된다. 감독하는 직원이나 모니터로서 이중의 역할을 맡은 면접원일 수도 있는 교육을 받은 모니

터들이 전화 면접의 표본의 전부나 일부를 듣고 면접원 업무수행의 긍정적이거나 부정적인 속성을 메모한다. 면접은 녹음기나 컴퓨터 음성녹음 면접(CARI) 기술이 설치되어 있는 노트북 컴퓨터를 사용해서 녹음할 수 있다. 이런 녹음한 것들은 다른 관측자에 의해서 반복적으로 검토 될 수 있기 때문에 관측이나 통화 모니터링보다 면접원 성과에 대하여 더 자세하고 신뢰할 수 있는 정보를 준다.

대면 면접에 대한 감독 관찰은 특정한 면접원의 업무수행에 대하여 다른 어느 방법보다 광범위한 정보를 제공할 수 있다. 감독자는 면접원의 면접하는 방법에 대하여 뿐만 아니라 면접원의 조직 능력과 접촉과 업무의 여러 가지 다양한 양상들을 계획하고 일정을 잡는 능력에 대해서도 정보를 얻을 수 있다. 이 방법의 한 가지 결점은 면접을 하는 동안 감독자의 존재가 면접원을 억제할 수 있어서 관측된 행동이 면접원의 보통 전형적인 행동의 아니라는 것이다. 예를 들어, 면접원은 감독자가 있을 때에는 그의 “최선의 행동”을 할 수 있다. 대부분의 경우에는 각색된 “좋은 행동”은 오랜 기간동안 지속하기 쉽지 않고 결국에는 면접원들이 제대로 수행하지 못하는 것이 발견된다. 그럼에도 불구하고 감독 관찰은 면접 지침의 고의적인 위반을 찾아내는데 이상적이지 않다.

전화 면접원의 통화 모니터링은 대부분의 중앙집중식 전화 설비가 있는 곳에서는 조사 운영의 관례적인 일이다. 통화 모니터링 체계에는 기본적으로 세 가지 요소가 있다. 모니터 해야 할 면접원과 면접을 선택하는 것, 모니터링 기간동안 기록할 정보, 건설적인 피드백의 형태로 면접원과 도구의 설계자에게 모니터 결과를 제공하는 방법이다. 면접원의 선택은 랜덤하거나 면접원들이 예측할 수 없는 어떤 다른 방법이어야 한다. 추출률은 면접원의 능력과 경험, 조사의 예산, 모니터 운영의 품질 목적에 따라 변할 수 있다. 지속적인 모니터의 전형적인 목표는 면접의 최소한 10%이고 각 면접원을 매주 모니터 해야 한다(Couper 등, 1982 참조).

많은 중앙집중식 전화시설에서 모니터들은 모니터를 한 후에 곧바로 면접원에게 피드백을 제공한다. 의견은 관측자의 관심과 태도에 따라 다르지만 흔히 주관적이고 체계적이지 않다. 더 객관적이고 체계적인 피드백을 제공하는 행동 코딩 체계가 개발되었다. 면접원과 응답자의 행동을 연구하면 흔히 면접원과 응답자가 조사 질문에 대하여 가지고 있는 문제점들을 드러내는데 이것은 조사 질문의 표현법이나 설계의 결함의 신호이다. 행동 패턴을 연구하는 것은 조사 관리자에게 문제되는 질문들을 지적할 수 있고 그것들을 개선방안을 제의할 수 있다.

면접원 업무수행을 평가할 때에 행동 코딩은 전형적으로 면접원 변동에 영향을 미치는 네 가지에 중점을 둔다. 그것들은 질문 전달, 탐색조사를 하는 행동, 응답자에게 하는 피드백과 속도 또는 명확성이다. <그림 5.4>는 Mathiowetz와 Cannell(1980)에서 인용한 어떤 중앙집중식 전화 시설에서 사용되는 코딩 체계의 한 예이다. 이 간단한 체계는 면접을 코딩하는데 사용되었는데 코딩이 면접과 동시에 일어난다. 실시간 코딩에서는 작성자의 판단의 신뢰도가 문제가 될 수 있다. 이것을 상쇄하기 위해서 코딩 절차를 더 극단적으로 단순화되는 경향이 있다. 응답자의 질문에 대한 이해의 문제들을 포착하는 항목들을 더하여 녹음된 면접에서 코딩을 하는 더 복잡한 체계가 고안되었다. 예를 들어, 응답자들이 질문을 반복해 달라고 부탁했는지, 명확하게 해 달라고 요구했는지, 답하기를 거절했는지 등을 코더들이 기록할 수 있다. 전화 조사를 위해 행동 코딩 체계를 설계하는데 유익한 지침에 대하여 Cannell과 Oksenberg(1988)를 참조하라.

질문 전달

피드백

- | | |
|------------------|--------------------|
| 11 질문을 쓰인 대로 읽는다 | 31 피드백을 적절하게 전달한다 |
| 12 주요하지 않은 표현 변경 | 32 피드백을 부적절하게 전달한다 |
| 13 주요한 표현의 변경 | 33 피드백을 전달하지 않는다 |
| 14 질문을 읽지 않는다 | |

탐색조사 행동

속도/속도조절

- | | |
|-------------------|------------------------|
| 21 적당하게 탐색조사를 한다 | 41 너무 빨리 또는 너무 느리게 읽는다 |
| 22 부적당하게 탐색조사를 한다 | 42 항목 사이의 시간이 너무 빠르다 |
| 23 반복하지 않는다 | 43 항목 사이의 시간이 너무 느리다 |

<그림 5.4> 전화 통화 모니터의 간단한 코딩 체계.

면접 과정을 평가하기 위한 행동 코딩의 이용의 실례로서 <표 5.6>에 제공되었다. 두 가지 연구의 결과가 이 표에 요약되어 있다. Rustemeyer(1977)이 실시한 연구는 200개가 넘는 가상면접를 근거로 했고 세 집단의 면접원들을 대상으로 하였다. 세 집단은 최소한 3개월의 현장경험이 있는 “유경험” 면접원, 단지 두 세 개의 과제를 완성한 “교육된” 면접원, 그리고 바로 교육을 마친 “신임” 면접원들이다. 세 집단 사이에 쓰인 그대로 읽는 질문들의 퍼센트에는 근본적으로 차이가 없다. 가장

큰 차이는 건너뛴 질문에서 발생한다. 여기에서는 경험이 많은 면접원들이 최고이고 새 면접원들이 그 뒤를 잇는다. 이 결과는 건너뛰는 오류는 주로 경험이 없는 것과 지필설문지의 navigational 특징의 함수다. 경험이 많은 면접원들은 주요한 표현 바꾸기의 퍼센트가 가장 높은 것에 주목해라. 이것은 설문지에서 표현이 어려웠다는 것을 말한다. 경험이 많은 면접원들은 이것들에 대하여 더 잘 알고 있고 표현을 바꿔서 어려움을 해결하려고 시도할 수 있다.

<표 5.6> 면접원과 질문 수행에 대한 행동 코딩 결과의 사용^a

	면접원 종류 (Rustemeyer, 1977)			질문 종류 (Mathiowetz and Cannell, 1980)	
	유경험	교육된	신입	서술형	선택형
질문전달					
쓰인 그대로	66.9	66.4	66.9	95.8	95.4
주요하지 않은 변경	22.5	17.9	19.9	71.9	3.7
주요한 변경	5.2	3.6	3.9	0.5	0.4
읽지 않음	3.3	8.9	6.0	1.8	0.5
탐색행위					
적절한 탐색	80.7	86.0	80.3	79.2	85.6
부적절한 탐색	19.4	14.1	19.6	20.8	14.4

출처: Groves(1989), 표 8.3.

(?합이 100%가 아님=>수정?)

^a code된 면접 행동을 근거로 한 종류별의 퍼센트다.

Rustemeyer의 연구에서는 탐색조사를 하는 행동에 대한 정보를 얻었다. 이것들은 적당한 응답항목을 배정할 수 있도록 응답과 질문의 의미를 명백히 하는 것을 겨냥하는 면접원과 응답자 사이에 말을 주고받는 것이다. 올바른 탐색은 독립적이고 응답자로부터 특정한 응답으로 이끌지 않으면서 응답을 명백히 하려고 시도한다. 옳지 않은 탐색은 이끌거나 지시적인 것이다. 이 행동에서는 경험과 정확함의 관계가 곡선형이고 가장 좋은 결과는 중간 집단에서 일어난다. Mathiowetz와 Cannell(1980) 연구에서처럼 행동 코딩은 면접원 업무수행을 질문 종류의 함수로 보는데 사용할 수 있다. 이 결과들은 우리가 앞에서 말한 서술형질문의 면접원 분산의 위험에 대한 의견과 일치한다.

Groves(1989)에서 언급한 여섯 개의 연구 중에 조사에 따라 면접원 행동의 변동은 꽤 두드러진다. 예를 들어, 쓰여진 그대로 전달되는 질문들의 비율은 대략 57%

와 96%사이에서 변동한다. 또한, 같은 조사 안에서 면접원 사이에 상당한 변동이 있다. Cannell와 Oksenberg(1988)는 그들의 전화 면접원 조사에서 만족스럽지 못한 질문 전달이 최고의 면접원들 사이에서 보다 가장 못한 면접원들 사이에서 최소한 일곱배가 더 일어난다는 것을 알아냈다.

5.4.3 면접원 업무수행을 평가하는 다른 방법들

몇 가지 그 밖의 기술이 조사에서 면접원 업무수행을 평가하는데 일반적으로 사용되고 있다. PAPI조사에서 사용되는 한 방법은 적절하기 못하게 완성된 용지들의 문제점을 파악하기 위해서 완성된 설문지를 검토하는 것이다. PAPI와 CAPI조사에서 무응답률, 코딩할 수 없는 정보, 완성된 일의 시간상의 분포(예를 들어, 시각, 요일, 주 등)같은 결과와 업무수행의 측정, 그리고 조사 과정에 대한 다른 통계량은 면접 과정의 품질에 많은 도움이 된다. 또한 그것들은 저렴하고 쉽게 실시할 수 있다. CPS에서 실업률이 낮은 것으로 알려진 지역에서 높은 비율의 실업자들을 보고하거나 예상된 여행 총 마일 수보다 훨씬 낮거나 예상된 빈 집 비율보다 훨씬 더 높은 비율을 보고한 면접원들을 평가 재면접을 위하여 선택하여 면접원 왜곡을 찾아낸 몇 가지 경우가 있다.

CAI 설문지의 광범위한 사용 때문에 CAPI 환경에서 전통적인 PAPI 면접원의 업무수행 측정은 이제 충분하지 못하다. CAI를 사용하면 읽기 쉬운지, 완성이 되었는지, 답이 범위 안에 있는지 등을 위한 용지의 검토는 불필요하다. 그것은 많은 이런 기능들이 이제 자동화되었기 때문이다. 이제는 키 누르기(keystroke)와 추적 파일분석이 조사의 지필조사를 대신한다.

키 누름 파일(*keystroke files*)은 면접원이 면접을 하는 동안 그가 CAI조사 도구를 통해 입력하는 모든 keystroke의 기록이다. 추적 파일(*trace files*)은 system이 실행하는 기능만을 포착한다. 두 파일 모두에는 많은 분량의 체계적이지 못한 자료가 있다. 이 자료들은 업무수행의 문제점을 드러내는 추적파일의 패턴을 확인하기 위하여 통계적 방법으로 분석되어야한다. 그렇지만 면접원, 면접 또는 질문의 다양한 code와 keystroke을 표로 만들어서 면접원들과 설문지를 평가할 수 있다.

Couper 등(1997)은 keystroke 파일 분석의 예비평가를 CAPI를 면접원 업무수행을 평가하는 도구로 제안한다. 그들은 이 도구를 인간과 기계간의 의사소통의 문제

들을 확인하는 방법으로 사용하는 것을 강조한다. 이것은 CAI 설문지의 유용성 (*usability*)라고 알려지게 되었다. 70세 이상의 성인을 대상으로 한 국가적 연구인 U.S. Study on AHEAD(Asset and Health Dynamics of the Oldest Old)에서 나온 keystroke 자료를 두 가지 방법으로 분석했다. 하나는 행동 코딩과 비슷한 방식으로 각 keystroke 파일의 자세한 검토와 코딩을 수반했다. 다른 방법은 단순히 선택한 keystroke 행동을 면접원들과 면접에 걸쳐 모았다.

그들의 분석은 키 누르기에 대한 분석이 없으면 확인하기 어려운 면접원들이 범하는 몇 가지 비능률적인 점들과 오류를 찾아냈다. 한 예로 면접원들의 86%가 최소한 한번 틀린 기능 키를 사용했고 이들 중 58%는 면접원들이 원하는 기능 키의 옆에 있는 키를 실수로 쳤다. 이 결과는 면접원들이 업무를 수행하는 다양한 기능들을 실행하는 키의 선택에 개선을 가져왔다. 키 누름 파일, 타임스탬프, 자세한 통화 기록 자료가 면접원 성과를 개선하는데 사용될 수 있는지 아직 알려지지 않았지만 면접 과정과 설문지 개발 과정을 개선하는 과정자료로 구실을 할 수 있다.

제 6 장 자료수집 방식과 관련된 오류

자료수집 방식(mode of data collection)은 표본구성원을 접촉하고 조사질문에 대한 그들의 응답을 얻는데 사용되는 수단을 가리킨다. 현재 세 가지 차원인 응답자와 접촉의 정도, 자료수집자와 면접원의 참여 정도, 컴퓨터 보조 정도에 의한 몇 가지 분류방법들이 사용되고 있다. 이 분류 체계를 사용한 다양한 자료수집 방식이 <표 6.1>에 나와 있다. 이 표에서 컴퓨터 보조 면접(CAI) 방식의 가지 수에 주목하라. 응답자들과 집, 직장 그리고 여행 중에 이야기를 나누고 상호 접촉을 하기 위한 새로운 기술들이 개발되면서 조사 자료를 수집하는데 컴퓨터의 사용이 계속 증가되고 있다(Nicholls 등, 1997). 이 절에서 일부 기술들에 대해 자세하게 설명한다.

<표 6.1> 자료수집방식: 자료수집자 참여, 응답자 접촉, 컴퓨터보조 정도의 함수^a

	자료수집자의 높은 참여		자료수집자의 낮은 참여	
	종이	컴퓨터	종이	컴퓨터
응답자와 직접 접촉	대면 (PAPI)	CAPI	일기	CASI, ACASI
응답자와 간접 접촉	전화 (PAPI)	CAPI	우편, 팩스, 이메일	TDE, 이메일, 웹, DBM, EMS, VRE
응답자와 접촉 없음	직접 관측	CADE	행정 기록	EDI

^aACASI, 음성 CASI; CADE, 컴퓨터 보조 자료입력; CAPI, 컴퓨터 보조 개별면접; CASI, 컴퓨터 보조 자기면접; CATI, 컴퓨터 보조 전화면접; DBM, 우편에 의한 디스크; EDI, 전자자료 교환; EMS, 전자우편조사; PAPI, 지필면접; T-ACASI, 전화 ACASI; TDE, touch-tone 자료입력; VRE, 음성인식 입력.

기본적으로 세 가지 중요한 자료수집 방식이 있다. 그것은 면접하는 동안 면접원과 응답자가 실제적으로 참석하는 대면(또는 개별 방문)조사, 면접원이 전화로 면접을 실시하는 전화면접, 그리고 설문지를 표본구성원에게 우편으로 보내고 우편으로 되돌려 받는 우편조사이다. 대면과 전화조사는 면접원 운영방식이고 우편조사는 자기기입식이다. 우리가 다룰 그 밖의 자료수집 방식에는 관리등록부(administrative registers)와 직접관측을 사용하는 것이 있다.

이 장에서는 이런 방식들을 자세히 설명하고 그것들의 주요 특징들을 검토한다.

자료수집 방식은 선택할 수 있다. 각 방식마다 장점과 단점이 있다. 많은 경우 방식의 선택은 복잡하고 조사의 목적, 방식의 특징, 다양한 설계 문제, 그리고 이용 가능한 방법론적 그리고 재정상의 자원과 관련이 된다. 이 장에서는 몇 가지 방식에 대한 장점과 단점들을 제시하고 방식의 선택이 자료 품질에 어떻게 영향을 미치는지를 설명한다.

6.1 자료수집 방식

이 절에서는 몇 개의 자료수집 방식들을 검토한다. 각 방법을 설명하고 일반적으로 알려진 그것의 특징들에 관해 알아본다.

6.1.1 대면 면접

대면 면접은 가장 오래된 면접 방식으로 최신 통신 기술에 의존되지 않는다. 대면 면접은 면접원과 응답자 사이에 최대의 의사전달과 상호작용이 제공되기 때문에 대부분의 조사 주제에 대해 자료품질면에서 선호되는 자료수집 방식으로 조사 연구원들에게 인식되어왔다. 민감한 주제와 관련된 측정오차 때문에 최근 몇십년 동안 이러한 시각에 이의가 제기되었다. 면접원 효과에 대한 연구가 진행되면서 면접원의 존재에 의해 좋지 않은 방향으로 영향을 받을 수 있는 주제가 늘어나고 있다. 대면 면접 방식의 몇 가지 장점과 단점들을 생각해보자.

장점과 단점

대면 면접은 면접원이 집이나 직장으로 응답자를 방문하거나 공공장소에서 만나야 하기 때문에 대개 예산/비용이 많이 드는 조사이다. 대부분의 경우 이동이 이 방법의 고비용 요소이다. 어떤 경우에는 면접원들이 장거리를 이동해야하며 하루에 단지 한 개의 면접만 마칠 수도 있다. 일반적으로 대면 면접이 다른 자료수집 방식보다 비용이 더 많이 든다.

대면 면접은 어떤 종류의 질문에 대해서는 사회적 기대부응 편향을 일으키는 것으로 알려졌다. 4장에서 언급했듯이 사회적 기대부응 편향은 민감한 질문들에서 일어

날 수 있는 현상이다. 대면 면접에서는 주제에 따라 응답자가 정확한 답을 하는 것 보다는 면접원에게 그들이 어떻게 보이는지에 대해 더 염려하는 경향이 있다. 이 현상은 다양한 방법으로 나타난다. 대면 면접에서 술 소비, 마약 사용, 또는 성행위 상대자수 같은 사회적으로 비난을 받는 행위를 포함하는 조사 주제에 대해서는 대체적으로 실제 이하로 보고 된다. 사고방식에 대한 질문에서 응답자는 정치에 관한 선호나 인종 문제에 관련된 극단적인 태도를 공개하지 않기 원한다. 또한 응답자들이 사실은 그렇지 않으면서 사회의 시사 문제에 대하여 박식하고 최신 정보에 대하여 잘 아는 것으로 보이고 싶어 하는 경향이 있다는 증거가 제시된 논문이 있다.

그렇지만 사회적 기대부응 편향만이 면접원 보조 조사에서 염려되는 측정오차의 원인은 아니다. 5장에서 언급했듯이 면접원들 자신도 중요한 오류의 원인이 된다. 면접원들은 각자마다 서로 다른 방법으로 응답자에게 영향을 미치는 경향이 있다. 각 면접원은 자기 자신의 행동, 작업 과정, 질문 전달과 질문 표현의 기술, 탐색 기술, 언제 탐사할 것인가에 대한 전략, 속도와 서술형 질문에 대한 응답을 기록하는 방법들을 가지고 있다. 이런 서로 다른 면접 방식 때문에 면접원들이 집단으로 상관관계가 있는 면접원 오류를 총조사오차에 제공한다. 5장에서 말했듯이 이 오류 요소는 추정값의 표준오차를 상당히 증가시킬 수 있고, 표본조사에서 계산한 분산 추정값에 통상적으로는 반영되지 않는다.

면접원 효과를 최소화하는 한 가지 접근 방법은 표준화된 면접과정이다. 그것은 면접원들이 전부 거의 같은 방식으로 수행하도록 교육시켜서 면접원 분산의 가능성을 제거하는 것이다. 그렇지만 5장에서 언급했듯이 표준화된 것에 비교해서 대화적인 면접 그리고 다른 면접 방식에 관하여 논쟁이 있다. 조사에서 면접원들을 이용하고 그들의 효과를 줄이는 방법들은 조사 방법론 분야에서 여전히 적극적으로 연구되고 있다.

면접원 분산과 사회적 기대부응 편향에 대한 이런 염려에도 불구하고 대면 면접이 가장 융통성 있는 방법으로 계속해서 여겨지는 몇 가지 이유가 있다. 이 방법은 한 시간 이상 지속될 수 있는 길고 복잡한 면접을 가능하게 한다. 면접원은 응답자와 함께 있기 때문에 협조를 얻기 위해 많은 책략을 사용할 수 있고, 면접을 하는 동안 직접 관측을 통해 추후 분석을 위해 기록해 둘 수 있고, 응답자의 응답이 다른 사람의 존재에 의해 영향을 받지 않을 수 있고, 더 완벽하고 정확한 답을 위하여 필요하면 탐색조사를 할 수도 있다.

전에 언급했듯이 대면 면접은 민감한 질문에 대해서는 시각적 보조물과 응답 카드를 줄 수 있고 응답자들과의 접촉에서 친밀하고 조화된 관계와 자신감을 쌓을 수 있다. 아마도 가장 가치 있는 특징은 응답률을 비교적 높은 수준으로 유지할 수 있다는 사실이다. 초기에, 그리고 필요하면 거부전환을 통해서 (처음에 안 한다고 했다가 다시 한다고 바꾸는 것) 대면 면접원들이 협조를 얻을 가능성은 대면 접촉이 없는 방식과는 비교가 안 된다.

대면 면접의 또 하나의 장점은 지역 확률표본 같은 표본설계에서 일반 모집단의 높은 포함이다(high coverage). 지역 표본은 지도나 사진 같은 지역 추출틀에서 표본을 추출하는 것을 수반하고, 이렇게 표본추출된 단위에 대한 자료를 수집하는 최종 단계에서는 대면 면접이 적격이다(3장과 9장 참조). 면접원들이 면접을 하는 동안 다른 사람들의 존재를 통제할 수 있음에도 불구하고, 다른 사람들에 의하여 영향을 받게 되는 것은 응답자의 집 또는 직장이나 공공장소 모두에서 문제가 될 수 있다. 다른 사람들의 존재는 응답자가 특정한 종류의 질문들에 숨기지 않고 정직할 수 있는 답하는데 방해가 될 수 있다.

아마도 대면 면접의 가장 심각한 현실적인 문제는 그것의 사용에 드는 비용이다. 면접원들이 단위들이 있는 곳으로 이동을 해야 하기 때문에 전화면접에서보다 작업을 할 수 있는 양이 훨씬 적다. 그러므로 이를테면 전화면접에서 보다 한 건을 완성하는데 더 많은 시간과 인적자원을 필요로 한다. 또한 교육 활동의 운영과 면접원 성과의 관측도 시간과 비용이 많이 든다.

마지막으로, 대면 면접에서의 중요한 걱정은 면접원이 면접을 속일 수 있다는 것이다. 이런 행위를 위조, 속임, 테이블의 꾸미기, 면접원 위조(*curbstoning*)라고 부른다(5장 참조). 이 행위는 모든 조사 기관들이 알고 있지만 찾아내기 매우 어려울 수 있다. 면접원은 면접 전체를 속이거나 시간과 응답자 부담을 줄이기 위하여 면접에서 질문들을 간단히 건너뛰고 나중에 자료를 가짜로 만들어서 건너뛴 질문들에 적어 넣을 수 있다. 이러한 현상에 대한 연구는 Biemer와 Stokes(1989)에서 찾을 수 있다. 이러한 행위는 재면접이나 확인면접과 응답분포의 분석을 통해서 발견할 수 있다(8장 참조). 면접원들이 면접을 속이는 한 가지 이유는 특정 모집단의 특정 부류에 접근하는 어려움 때문일 수 있다. 예를 들어 범죄가 많은 지역이나 경비 직원들이 보호하는 아파트 건물들이다.

컴퓨터 보조 개별면접

컴퓨터 보조 면접(CAI) 체계의 등장으로 컴퓨터 보조 개별면접(*computer-assisted personal interviewing*; CAPI)이라고 부르는 대면 면접의 컴퓨터화된 형태가 점차 많이 사용되고 있다. 면접은 노트북 컴퓨터에 저장된 프로그램으로 실시한다. 각각의 응답자를 위하여 정해진 순서대로 화면에 조사 질문들이 나타나게 프로그램이 되어있다. 면접원들은 질문을 하고 응답자가 제공한 답을 입력한다.

컴퓨터 보조 면접은 처음에 컴퓨터 보조 전화면접(CATI)이라고 부르는 중앙집중식 전화면접 환경에서 사용되었다(6.1.2절 참조). 집중식 CATI 시설에서 면접원들은 100개 이상의 면접장소가 있는 전화시설(*telephone facilities*) 또는 전화통화센터(*calling centers*)라고 하는 큰 방에서 같이 근무한다. 사례는 일반적으로 수요가 있을 때마다 중앙집중식 데이터베이스에서 면접원들에게 배정된다. 면접원이 통화를 완료하면 통화 결과가 데이터베이스에 입력된다. 만일 완전한 면접이 실시되지 않으면 그 사례는 다른 면접원에게 배정되어 완성할 수 있도록 한다. 보통 모든 면접의 일부(예를 들어, 5에서 10%)는 품질관리 목적을 위하여 감독자가 모니터한다.

이론적으로는 면접의 관리에 컴퓨터 프로그램을 사용하는 것이 약간의 오류를 제거할 것이다. 예를 들어, 만일 설문지에 건너뛴 것들이 있으면 컴퓨터 프로그램은 이미나온 답을 근거로 바른 길을 선택할 수 있다. 또한, 만일 가구조사의 설문지가 두 명 이상의 가구구성원에 대한 질문이 있으면 질문에서 묻은 그들의 이름을 컴퓨터가 입력할 수 있어서 이 정보는 자동적으로 질문의 일부가 된다. 그러므로 컴퓨터 소프트웨어는 자동 입력을 사용해서 설문지를 “개별화” 할 수 있다. 이것은 설문지에 흔히 나오는 일반인칭을 면접을 받는 특정한 응답자를 이름으로 대신한다. 가구에서 소유하는 모든 차들에 관한 질문들, 상세하게 조사하고 있는 사건들, 그리고 사업체의 모든 작업장에 관한 질문들도 이 방식으로 다룰 수 있고 그것에 의하여 면접원과 응답자 모두의 혼동을 피한다.

온라인 편집과 코딩 같은 특정한 자료 처리 활동도 면접 중에 일어날 수 있다(7장 참조). Groves와 Tortora(1998)에서 제시 되었듯이, CAI와 관련된 이런 이론적이고 논리적인 장점들은 방법론연구에서 나온 자료에 의해 그 차이가 항상 통계적으로 유의하다고 입증되지는 못했다. 그럼에도 불구하고 연구들은 측정오차의 명백한 감소를 보통 보여준다.

자료품질에 미치는 효과

CAPI가 자료품질에 미치는 효과에 대한 논문은 드물다. 대면 지필면접 (PAPI)과 CAPI를 건너뛰는 오류율, 항목무응답, 범위 밖의 응답과 면접원 분산과 같은 과정 변수와 관련해서 비교하는 몇 개의 연구가 있지만 상당한 차이는 없었다. 그보다는, 조사기관과 조사의뢰자들은 자료품질이 아닌 다른 이유에서 CAPI를 택한다는 것 증거를 제시한다. 비용과 시간의 절약은 중요하고 일부 기관들은 이런 이득을 보았다. 다른 기관들은 CAPI가 실제로 PAPI보다 비용이 더 든다고 보고하지만 CAPI가 PAPI보다 훨씬 많은 기능이 있기 때문에 단지 비용만을 근거로 두 방식을 비교하는 것은 타당하지 않을 수 있다. 예를 들어, CAPI는 PAPI보다 훨씬 더 복잡한 면접을 실시하고, 이전에 수집된 자료에서 얻은 응답들이 질문 표현의 일부가 될 수 종속면접(*dependent interviewing*)을 사용하고, 이를테면 온라인 편집에서 면접원과 응답자간의 고도의 상호작용을 사용하는 더 많은 기회가 제공된다. 그러므로 CAPI가 PAPI보다 비용이 다소 더 들더라도 이러한 기능 때문으로 얻어진 면접의 결과에 더해진 부가가치는 추가비용의 충분한 가치가 있다.

PAPI에서 CAPI 환경으로의 전환

조사기관에서 컴퓨터화되지 않은 환경을 컴퓨터화 하는 것은 비용에 관계된다. 이것은 하드웨어(컴퓨터의 기계 설비), 프로그램과 면접원 교육을 위한 초기비용이다. 어떤 경우에는 CAPI 프로그램 오류들이 면접이 시작된 후에 발견한다. 많은 기관들은 CAPI 적용의 문제를 찾아내어 고치는 지속적인 문제가 있을 수 있고 오류 때문에 사례를 잃는 경험을 할 수도 있다.

따라서 PAPI에서 CAPI로 옮겨야 할 몇 가지 이유가 있지만 CAPI는 모든 조사와 모든 조사기관에서 사용할 수는 없다. 큰 기관이라도 때로는 초기비용이 너무 클 수 있고 어떤 CAPI 시스템에서 다른 시스템으로 이동하는 것이 대단히 부담스러울 수 있다. 작은 기관에서는 영업용으로 이용할 수 있는 CAPI 시스템조차도 아직 경제적으로 실행할 수 없을 수 있다.

CAPI 시스템을 이미 가지고 있는 기관들도 일회성 조사에서 그 시스템을 항상 사용할 가치가 있는 것은 아닐 것이다. 만일 신속히 조사를 시행해야 한다면 적합하지 않을 수도 있다. 기술적인 작업을 하고 면접원들을 교육시키기에 시간이 충분

하지 않을 수 있다. 그래서 대부분의 큰 기관들은 상황에 따라서 PAPI와 CAPI를 둘 다 실시할 준비가 되어있어야 한다.

대면 면접은 (*face-to-face interviewing*)은 비용이 많이 들지만 융통성이 있고, 높은 응답률을 제공하지만 사회적 기대부응 편향과 면접원 분산을 일으킬 수 있다.

6.1.2 전화면접

전화 면접은 사회와 경제 연구의 자료수집 방식으로 항상 받아들여지지는 않았다. 1936년 미국 대통령 선거에서 전화번호부 목록을 근거로 한 전화조사 결과에서 Landon이 Roosevelt를 제치고 압도적 대승리를 거둘 것으로 예측했고 이것은 널리 발표되었다. Roosevelt가 이겼을 때 전화 방식은 잘못된 예측으로 비난을 받았다 (Katz와 Cantril, 1937). 사실은 그 당시 미국 전체 가구의 35%만이 전화가 있었고, 이런 전화 인구는 불균형적으로 공화당원이었던 것이 전화 가구가 Landon을 선호하는 것을 설명한다.

불행하게도 전화조사가 정당한 면접 방식으로 미국의 사회과학자들 사이에 받아들여지기 시작하는데 40년이 걸렸다. 그 방식에 대해 관심이 늘어난 두 가지 이유는 대면조사보다 낮은 자료수집 비용과 미국과 세계적으로 늘어난 전화 사용 모집단 때문이다. 예를 들어, Groves와 Kahn(1979)은 RDD와 지역추출틀 대면 면접의 비용과 오류 요소의 자세한 검토를 하였다. RDD 전화조사는 대면조사와 비슷한 자료 품질을 훨씬 낮은 비용에 제공할 수 있다는 것을 보여준다. 그렇지만 오늘날에는 독점적인 자료수집 방식으로서의 전화조사가 도전받고 있다. 이것의 가장 중요한 원인은 조사에서 응답률이 낮아지는 경향 때문이다. 그러나 전화는 특히 혼합 방식 조사(*mixed-mode surveys*)라고 하는 다른 방법과 결합한 많은 조사에서 아직도 사용되고 있다.

다음에는 이 자료수집 방식의 단독 사용과 다른 자료수집 방식과 결합하여 사용하는 방식의 장점과 단점을 이야기한다.

장점과 단점

응답자와 연락을 취할 때 대면 면접과 전화면접 모두 면접원을 이용하므로 두 면접의 특징이 매우 비슷한 것을 발견하는 것은 놀라운 것이 아니다. 두 방법 모두 조사결과에 면접원 분산과 사회적 기대부응 편향을 불러일으킬 가능성이 있다. 그렇지만 연구 문헌들에 의하면 이런 결과들은 전화 방식에서 다소 적다고 한다(4장과 5장 참조).

대면 면접에서 가능한 것의 많은 부분이 전화로는 할 수 없다는 것을 보면 전화 면접은 대면 면접보다 다소 융통성이 적다. 예를 들어, 시각적 보조물은 면접 전에 응답자에게 우송되지 않으면 사용할 수 없는데 이것은 매우 비현실적인 절차이다. 또한 전화로 하는 질문들은 너무 복잡해서는 안 되고 응답항목이 6가지를 넘어서는 안 된다.

응답자에게 기록을 조사하여 계산하라고 부탁하는 것과 같은 복잡한 것들은 전화 면접환경에서 하기는 대단히 어렵다. 면접원의 익명성 때문에 사회적 기대부응 편향이 대면 면접에서 보다 더 작을 수 있지만 바로 이 익명성이 친밀하고 조화된 관계를 발전시키고 응답자가 참여하도록 설득하거나 면접 중에 최선을 다하도록 하는데 불리하다. 이런 결과, 더 많은 목인과 극단적인 응답(처음 또는 마지막 등급이나 첫 번째나 마지막에 제공된 응답항목을 선택하는 것), 그리고 일반적으로 깊이 생각하지 않은 응답을 하는 경향이 있다.

예를 들어, 대면조사에서보다 전화로 “모름”과 “답 없음(?)”의 응답 수가 더 많은 경향이 있고 더 빠른 면접속도는 더 짧은 답을 하게 한다. 또한 전화면접원들은 대개 대면 면접에서 할 수 있는 것 보다 더 많은 작업량을 배정 받는다는 사실 때문에 면접원 분산이 대면 면접에서보다 클 수 있다. 전형적으로 응답률은 비교할 수 있는 종류와 크기의 대면 조사에서보다 전화 조사에서 더 낮다.

전화조사는 대면 면접보다 상당히 더 짧은 경향이 있다. 한 시간의 평균 면접 지속시간을 가지는 성공적인 전화조사의 예가 있기는 하지만 이것은 일반적인 사람들의 조사에서는 꽤 드물다. 대부분의 전화 접 설계자들은 면접시간을 30분 이하로 제한하려고 노력한다. 전화조사들은 더 빨리 시작할 수 있고 대면 조사보다 덜 비싸다.

중앙집중식 전화면접은 일반관리, 감독, 피드백과 교육을 쉽게 할 수 있는 특수한

장점들을 가지고 있다. 어떤 나라에서는 낮은 전화보급률((침투율(*penetration rates*)이라고 한다)) 때문에 전화를 주된 자료수집 방식으로 사용하는 것이 불가능할 수 있다. 전화방식과 관련된 중요한 문제점은 모든 표본구성원들이 전화를 가지고 있지는 않다는 것, 자동응답기와 발신자표시장치로 표본 단위로의 접근을 막을 수 있다는 것, 그리고 설문지가 너무 길 수 없고 시각적 보조물과 많은 응답항목들 같은 기능을 이용할 수 없다는 것이다.

컴퓨터 보조 전화면접

요즘에는 흔히 CAI 기술을 사용해서 전화면접을 실시한다. 사실 CATI는 약 30년 전에 등장한 첫번째 컴퓨터 보조 자료수집 방식이고 이제는 많은 기관들에서 일반적인 쓰고 있다. 이것은 더 이상 새로운 기술로 분류되지 않는다. CATI의 이익과 결점은 6.1.1절에서 이야기한 CAPI의 경우와 매우 비슷하다. 또한 단순히 도구의 컴퓨터화의 덕분으로 CATI와 CAPI 모두가 측정오차를 상당히 줄인다는 사실이 일반적으로 널리 인정 되었다(Couper 등, 1998). CATI의 중요한 장점은 복잡한 종이설문지에서 실수로 건너뛰는 해를 피할 수 있는 점이고 그러므로 CATI의 항목무응답률은 일반적으로 PAPI보다 낮다.

대면 면접과 비교한 전화면접의 장단점	
장점	단점
<ul style="list-style-type: none"> • 덜 비싸다 • 면접원 분산이 대면조사보다 더 작다. • 사회적 기대부응 편향이 대면조사보다 더 작다. 	<ul style="list-style-type: none"> • 대면 조사보다 융통성이 적다. • 시각적 통신매체를 사용할 수 없다. • 면접이 대면조사보다 짧아야 한다 • 전화가 없는 단위는 포함이 안된다

6.1.3 우편조사

우편 조사에서는 각 표본 구성원이 우편으로 보내지는 설문지를 완성하여 조사기관에 돌려보낸다. 자료를 수집하는 사람과의 접촉이 없기 때문에 이 방식은 자기기입식(*self-administered*)이라고 한다. 설문지는 그 자체로 응답자가 설명 없이 이

해할 수 있어야 한다(즉, 외부의 도움을 받을 수 없으므로 질문과 지시는 응답자가 쉽게 이해할 수 있어야 한다). 면접원 보조 방식들 보다 우편조사 결과로 생긴 자료의 품질은 설문지 설계의 품질에 훨씬 더 큰 영향을 받는다.

3장에서 자세히 말했듯이 우편조사의 응답률은 조사직원의 능력과 지식으로 인하여 조사기관에 따라 엄청나게 변할 수 있다. 우편조사를 성공적으로 수행하는 데 주의가 필요한 것은 분명하다. 이 절에서는 다른 자료수집 방식과 비교한 우편조사의 장점과 단점을 설명한다.

장점과 단점

지난 20년 동안 우편조사는 부침이 심했다. 면접원 보조 방식의 비용이 상당히 증가하면서 우편조사는 매력 있는 선택이 되었다. 우편조사의 수행은 상당히 저렴해서 적은 예산으로 하는 조사에서 선호되는 방식이다. 또한 자기운영식이라는 점에서 우편조사는 사회적 기대부응 편향 위험이 감소하기 면접원 보조 방식보다 품질면에서 유리할 수 있다. 이 방식에서는 시각적인 전달 방법을 사용할 수 있기 때문에 응답자에게 시각적 보조물을 제공하는 것이 필요한 상황(예를 들어, 국립공원으로의 최근의 여행에서 방문한 다양한 장소들을 표시하기 위하여 지도를 사용하거나 이를테면 6가지가 넘는 응답항목이 있는 질문)에서는 전화조사보다 이점이 있다.

응답자가 설문지를 가지고 있기 때문에 우편조사는 응답자가 심사숙고하여 답을 하고 기록을 찾아볼 시간도 준다. 마지막으로, 우편조사는 면접원에 의해 발생하는 분산을 일으키지 않고, 질문 순서나 응답항목 순서에 의해 발생하는 오류를 줄인다(Dillman, 2000).

그렇지만 우편조사는 면접원 보조 방식들보다 응답률이 더 낮다고 알려져 있다. 그러므로 그것은 면접원들이 실시하는 사후조사와 결합시킨다(CATI 또는 CAPI 혼합 방식 조사). 몇 십년전에는 우편조사가 면접보다 못하다고 여겨졌다. 이것은 낮은 응답률과 조사기관이 응답과정을 거의 제어할 수 없다는 사실 때문이다. 예를 들어, 의도한 sample person이 설문지를 완성하는지 또는 그가 다른 사람들하고 적절하게 조사설계에서 의도된 대로 협력하는지를 확실히 아는 것이 가능하지 않다. 설문지가 반송되지 않으면 이것이 설문지가 sample person에게 도달하지 않아서인지 아니면 sample person에게 도달했지만 그가 설문지를 되돌려 보내지 않아서인지 알 수 없다. 어떤 경우에는 sample person이 설문지를 완성하기 자격이 없을 수도

있지만 조사원들이 그 결정을 하기 위한 아무런 정보가 없다. 응답률을 계산하고 무응답의 추정값을 조절하는데 문제가 있다. 또한 방식과 관련된 상당한 항목 무응답률의 큰 위험이 있다.

가구조사에서 표본 대상자를 가구내에서 랜덤하게 선택해야 할 때에는 우편조사에는 문제가 있다. 최근 생일 방법(9장) 같은 안을 시도해 보았지만 이런 방법들은 확률표본을 산출하지 않을 염려가 있고 엄밀한 과학연구에서는 사용되어서는 안 된다. 그러므로 우편조사는 설문지를 특정 이름의 응답자에게 직접 우송할 수 없는 상황에서는 권하지 않는다.

우편조사는 만족스러운 응답률을 얻기 위하여 긴 현지 기간을 필요로 한다. 발송 우편물을 우송하는 것으로부터 마지막 반송까지 보통 8주 이상이 필요하다. 또한 응답자가 개념을 이해하고 설문지를 문제없이 진행할 수 있어야한다는 의미에서 설문지는 쉬워야한다. 물론 응답자들은 질문과 지시를 읽고 쓸수 있는 능력 수준이 되어야한다. 미국에서는 일반 모집단에 대한 정부 조사에서 사용되는 우편설문지에 답하기 위해서는 읽기 능력이 5학년 수준이 되어야한다는 일반적인 규칙이 사용된다.

최근 연구

우편조사 사용의 재기의 상당한 부분은 미국 워싱턴주 풀맨에 있는 워싱턴주립대학교의 Don Dillman의 연구 덕분이다(Dillman, 1978, 2000 참조). Dillman은 우편조사 자료수집을 위한 단계적인 계획으로 이루어진 효율적인 전략을 개발했다. 만일 이것을 따른다면 다른 문서화된 전략들에서보다 더 큰 응답률을 가져올 것이다. Dillman의 전략들은 사회교환이론(*social exchange theory*)을 기본으로 사용하며 3장에서 설명했었다. 그는 또한 면접 조사에서 협조를 증가시키기 위하여 자기기입식 방식의 이론을 개발하기 시작했는데, 그것은 그래픽 언어, 인식, 시각적 지각의 이론과 Cialdini(1984)와 Groves와 Couper(1998)의 이론의 결합을 근거로 했다. 3장에서 이야기한 사회적 그리고 심리적인 영향(보상, 서약과 일치성, 사회적 확인, 권위, 희소성, 유대감)도 우편조사의 설계와 운영에 사용될 수 있다.

그러므로 사용되는 원리가 두 가지 있다. 하나는 설문지가 자체로서 완비되게 개발하여 설문지에 있는 정보의 구성과 응답자에게 제공되는 진행 가이드를 통하여 설문지가 완성될 수 있도록 한다. 이 시각에서 보면 이런 기능들은 면접원들이 면

접원 보조 방식에서 응답자들을 가이드하는 것과 같은 방법으로 응답자를 돕는다.

두 번째 원리는 자료수집의 실행을 지시하는 단계적인 방식을 포함한다. 이것은 사전편지, 여러 번의 재권유, 반송 봉투에 붙인 실제 우표, 이름을 적은 개별화된 발송 우편물, 응답자에게 연속적으로 계속 재촉하는 요청과 응답률을 늘리기 위한 다른 전략들의 사용을 수반한다. 이 원리들은 3장에 설명되어 있고 더 자세한 설명은 Jenkins와 Dillman(1997), Dillman(2000), 그리고 Redline과 Dillman(2002)에 있다. 이 원리들은 수정되어 왔고 실험에 입각한 연구를 통해서 새로운 정보를 얻으면서 계속 수정될 것이다.

우편조사(*mail survey*)는 면접원 분산을 일으키지 않고 민감한 종류의 자료를 수집하는데 적당하다. 저렴하긴 하지만 우편조사는 수행하는데는 시간이 걸리고, 응답률이 비교적 낮고, 항목 무응답의 위험이 높다.

6.1.4 일지 조사

일지조사(*diary surveys*)는 사건들에 대한 정보를 회고적으로 수집하는 목적으로 사용된다. 가구 구매, 음식 섭취, 일간 이동, TV 시청, 그리고 시간의 이용 같은 자주 일어나는 일들에 대한 정보를 응답자가 입력한 일지로 구조적인 설문지를 대신한다. 기억오차를 피하기 위하여 응답자들에게 그 사건들이 일어난 후 곧 정보를 입력하라고 부탁한다. 이것은 흔히 응답자들이 정보를 매일 또는 더 자주 기록해야 한다는 것을 의미한다. 그러므로 일지의 성공적인 완성은 응답자가 정보를 기록하는데 매우 적극적인 역할을 하는 것을 필요로 한다.

우편조사와는 달리 일지조사는 면접원들이 응답자들에게 최소한 두 번 연락을 해야 한다. 첫 번째 접촉에서는 면접원이 일지를 전달하고, 응답자의 협조를 얻고, 자료를 기록하는 절차를 설명한다. 두 번째 접촉에서는 면접원이 일지를 걷고 회수하고 만일 완성이 되지 않았으면 그것을 완성하도록 응답자를 돕는다. 응답자가 일지를 정확하게 유지한다는 확실한 약속이 필요하기 때문에 일지 기록기간은 대체로 하루에서 2주사이의 상당히 짧은 기간이어야 한다.

예를 들어, 조사의 목적이 가구 지출에 대한 자료를 수집하는 것이면 구매를 한 후 즉시 물품의 가격과 함께 기록할 것을 응답자에게 부탁한다. 이런 자세한 정보는 면접으로 모을 수 없다. 이것은 응답자가 특히 작은 구매를 며칠이상 기억하지 못할 수 있기 때문이다. 어떤 종류의 구매에 대해서는 응답자들이 혼자 제공하는 정보보다 면접원의 도움을 통하여 더 정확한 정보를 제공할 수 있기 때문에 면접원은 이 방식의 중요한 요소이다.

자동차나 배와 같은 특히 큰 구매에서는 신축응답(*telescoping*)효과의 위험이 높다(4장 참조). 이런 구매는 많은 응답자들에게는 두드러진 일들이고 이런 일들은 보고기간에서 앞으로 신축되는 경향이 있다는 증거가 있다. 어떤 응답자들은 청구서의 납부는 구매로 여기지 않는다. 그 결과 어떤 지출 조사에서는 신용카드납부는 구매의 범위 밖이라고 생각한다. 대신 응답자에게 각자의 구매에 대해 기록해 달라고 부탁한다.

면접원의 접촉, 인센티브, 독촉장 등으로 협조를 자극하려는 노력에도 불구하고 응답자들이 점차적으로 기록에 관심을 잃어가는 분명한 경향이 있다. 이것에 대한 증거로 수집기간의 초반보다 후반에 구매의 보고가 적어진다(Silberstein과 Scott, 1991, 참조).

또한 응답자 쪽에서도 조건반사의 위험이 명백히 있다. 조사에 참여하는 결과 응답자는 그의 구매 행태를 바꿀 수 있다. 예를 들어, 응답자가 특정한 상품에 쓴 돈의 액수를 깨달으면 구매 패턴을 일시적으로 바꿀 수 있다. 자료수집이 끝나고 얼마 후에 구매 패턴은 다시 전으로 돌아가고 그 결과는 편향된 추정값이 된다.

기억오차로 인하여 많은 품목들에 대한 구매를 모두 기록하지는 않게 되고 따라서 과소평가를 가져온다. 응답자들이 외식을 하거나 휘발유를 넣고 항상 영수증을 보관하는 것은 매우 어렵다. 그래서 실제 이하로 신고하는 것이 이 분야에서의 큰 문제이다. 자료수집 기간의 길이에 따라서 기억오차는 늘어난다.

어떤 설계에서는 대리보고가 문제가 된다. 만일 가구당 일지응답자가 한명이면 실제 이하로 보고하는 일이 일어날 것이다. 물론 개별일지는 대리보고효과를 방지한다. 사회적 기대부응 편향은 면접에서보다 일지조사에서 더 작지만 만일 가구당 일지가 한 개밖에 없으면 문제가 될 수 있다. 소비조사에서 술과 담배의 구입은 대체로 실제 이하로 신고한다. 또한, 모든 지필식 자기기입식 방식에서처럼 읽고 쓰는 능력과 교육수준은 좋은 품질의 자료를 얻는데 중요한 필수 조건이다.

일지의 설계가 중요한 오류의 원인이 될 수 있다는 증거가 있다. 기본적으로 두 가지 설계가 있다. 하나는 대체로 빈 페이지로 이루어져있고 다른 것은 미리 인쇄된 제목 또는 구분이 있는 더 구조적인 것이다. 둘 사이의 비교에서 결정적인 결론을 얻지 못했다(Tucker, 1992). 최근에 일지조사의 설계에 대한 두 가지 문제를 광범하게 다룬 문헌이 발표되었다. 하나는 기록기간의 최적의 길이에 관한 것이다. 과거에는 구매에 대한 가구소비조사에서 4주간의 일지를 사용했다. 그렇지만 이에 따른 엄청난 응답자 부담과 긴 기간동안 응답의 완전도가 감소한다는 사실 때문에 최근에는 이런 관행이 바뀌었다. 그 결과 요즘은 더 짧은 수집기간의 일지가 흔하다.

다른 문제는 일지를 사용해서 다룰 수 있는 사건들의 종류에 관한 것이다. 최근 추세는 응답자의 부담을 줄이기 위해 조사의 범위를 제한하는 것이다. 오직 식품품목에 대한 자료만을 수집하는 소비조사의 예가 있다. 그렇지만 이 추세는 모든 나라에서 나타나지는 않는다. 예를 들어, 네덜란드에서 사용되는 한 설계는 두가지 일지 요소가 있다. 첫 번째는 2주간의 완전한 일지이고 두 번째는 특정 액수 이상을 기록하는 1년간의 일지이다.

일지조사의 특수한 특성은 일지의 배치가 규정된 패턴을 따른다는 것이다. 모든 종류의 구매나 활동의 수집이 가능하도록 표본은 대체로 한 해 전체에 걸쳐서 시간적으로 분산시킨다. 만일 이런 조사를 한 해의 일부로 제한하면 계절적인 패턴 때문에 어떤 일들의 발생은 매우 과소평가될 것이다. 그러므로 자료수집은 미리 지정된 날짜에 기록을 시작하는 부표본으로 표본을 나누어서 실시한다. 미리 지정한 요일까지도 고른 분포의 표본이 되는 것이 중요하다. 실제적인 이유로 인하여 배치계획의 충실한 이행을 어렵게 한다. 따라서 대부분의 기관들은 비교적 중요하지 않은 계획의 변경은 인정한다. 그렇지만 표본구성원이 설계에서 정한 날짜가 아닌 날짜에 기록을 시작하도록 허용하는 것은 이런 결정으로 비록 실제 무응답은 피할 수 있을지라도 “nonresponse in time” 때문에 발생하는 측정오차를 가져올 것이다 (Lyberg, 1991).

6.1.5 기타 자기기입식 방식

자기기입의 유용성을 높인 컴퓨터화된 자기기입식 방식들이 있다(Ramos 등, 1998). 전자 자료 교환(EDI; *electronic data interchange*)는 표본 추출된 기업들의

자체 컴퓨터에 저장된 기록에서 경제와 다른 자료를 직접 얻는 기술이다. 아직 초기 단계이지만 조사 기관들의 장기 목적은 이 기록들에서 자료를 수집하고 그것에 의하여 일부 분야에서 전통적인 조사의 필요성을 줄이는 것이다. EDI의 광범위한 사용 전에 해결되어야 할 몇 가지 문제가 분명히 있다. 통계적 자료를 수집 기관으로 전달하는데 사용할 수 있는 표준 메시지 포맷을 사용하도록 기업들은 설득하는 것이 필요하다. 경험에 의하면 기업들 뿐 아니라 회계 소프트웨어의 개발자들도 깊이 관련되기 때문에 표준화 하기는 어려울 수 있다. 그렇지만 기업의 데이터베이스를 수집하는 기관과 연결하는 바로 그 아이디어는 매력적이고 기업들은 항상 문서 업무를 줄이는 것에 관심이 있기 때문에 전망은 밝다(Keller, 1994).

우편에 의한 디스크 (*disk by mail; DBM*)에서는 디스켓을 응답자에게 우송한다. 그 디스켓에는 자기기입식 설문지가 들어있고 응답자들은 자기 자신의 컴퓨터에 그 설문지를 설치하고 응답한다. 응답들은 디스크에 저장하고 조사 기관에 우편으로 돌려보낸다. DBM은 질문의 흐름을 조정하고, 지시를 제공하고, 편집 점검을 수행하므로 DBM은 거의 전산화한 면접처럼 작동한다. DBM은 분명히 비용을 절감하고 오류를 줄이는 가능성이 있고 만일 면접원 보조 방법의 대안으로 본다면 면접원 분산이 수반되지 않는다.

전자 우편 조사(*electronic mail survey; EMS*)는 전산화된 자기기입식 설문지를 전자 우편으로 응답자에게 보내고 돌려받거나 인터넷이나 월드 와이드 웹(WWW)으로 전달된다. DBM과 EMS는 많은 공통된 요소가 있다. 근본적인 차이는 웹 조사는 DBM과 전자 메일이 가지고 있지 않은 동시(同時)의 또는 잠재적인 특성들이 있다. WWW의 그래픽과 멀티미디어 기능으로 설계의 선택은 거의 제한이 없으며 조사 방식 연구자들은 웹 조사 설계에 대한 지침의 개발에 집중하게 되었다.(Couper 등, 2001 참조). 현재 두 가지 개발 계통이 있다. 하나는 대형 조사 기관이 주도하며 조사 질문에 답을 하는데 있어 모든 사람들이 컴퓨터를 이용할 수 있는 않고 또한 이용하고 싶어 하지는 않는다는 사실을 인정하는 조심스러운 점진적인 개발이 특징이다. 그러므로 DBM과 EMS는 보통 대형 조사 기관들이 다른 수집 방법의 대안으로 제시된다. 다른 개발 방침은 조사 업무를 하지 않는 기관들이 주도한다. 후자의 예는 자기 선택을 근거로 하는 다량의 “조사”로 귀착되는 “오늘의 웹 질문” 이라고 불리는 조사를 실시하는 신문과 TV 네트워크들이다.

DBM과 EMS 대한 몇 가지 제한이 Ramos 등(1998)에 의해 요약되었다.

1. 대부분의 CAPI와 CATI 프로그램에서 볼 수 있는 건너뛰기와 에디팅 검사, 응답 항목들의 확률화와 같은 성능들을 사용하는 조사 기관들에 알맞은 손쉽게 이용 가능한 전자 설문지용 소프트웨어가 없다.
2. 설문지 설계 기능이 제한되어 있고 이것은 응용의 범위가 제한된다.
3. 전자 우편은 표준화되지 않은 소프트웨어를 근거로 하며 표준화가 곧 이루어질 전망도 없어 보인다.
4. 모뎀과 통신 소프트웨어에 대한 응답자의 이용이 제한됐다.
5. 전자우편 주소는 자주 바뀐다.
6. 기밀성과 안전에 관련된 많은 문제점들이 있다. 기밀성이 보장되어야 하고 통신 패키지의 설치가 바이러스 공격을 받지 않을 것이라는 보증이 되어야 한다.
7. 제대로 설치된 PC로의 제한된 응답자의 접근성과 제한된 모뎀과 인터넷의 접근성 때문에 포함오차가 염려된다.

이러한 제한에도 불구하고 전통적인 우편 조사에서 DBM과 EMS가 측정오차를 줄일 것이라는 몇 가지 이유가 있다. 그것은 전통적인 면접을 개선하는 CATI와 CAPI 배후에 있는 이유들과 관련되어 있다. DBM과 EMS는 더 통제된 수집 환경을 제공한다. 제어된 전달과 장착된 에디팅은 측정오차와 항목 무응답을 줄인다. 또한 DBM과 EMS의 이용에서 응답들이 더 사려가 깊다는 증거가 있다.

푸시버튼식 자료 입력(*touch-tone data entry; TDE*)은 응답자가 컴퓨터에게 전화를 걸고 컴퓨터가 묻는 질문에 답하는 수집 방법이다. 응답 과정은 응답자가 푸시버튼 전화의 키패드를 사용해서 자료를 입력한다. 응답을 입력한 다음, 컴퓨터가 응답자에게 제공한 응답을 확인할 것을 부탁한다.

TDE는 현대사회에서 매우 흔한 통신 기술이며 주문, 거래와 전화통화 연결에 사용된다. 지금까지 TDE는 상당히 짧고(이른다면 5에서 10분) 자료 입력이 간단한 조사에서만 사용되어 왔다. 이것은 요구되는 정보가 숫자이거나 숫자 코드로 전환할 수 있는 상황에 알맞다. TDE는 우편 설문지나 면접의 대안이고 그것의 발달은 비용과 관련된 문제에 의해 시작되었다. 또한 TDE 서비스는 일주일에 7일, 하루 24시간 열려있기 때문에 응답자들이 면접 시간을 선택하게 한다.

또 다른 이점은 많은 TDE 이용에서 수작업 방법과 비교하여 응답자당 수집 시

간이 단축되었다. 수집 기관으로서는 면접원 관여를 제거하는 것이 비용측면 뿐 아니라 응답자가 24시간 이용할 수 있다는 실제적인 이유 때문에 유리하다. 분명히 불리한 점은 제한된 적용 범위이다. TDE는 짧고 숫자로 하는 반복적인 입력과 관련된 자료 수집 상황에 가장 잘 작용하기 때문에 그것이 성공적으로 사용될 수 있는 조사는 얼마 되지 않을 수 있다.

음성 인식 입력 (*voice recognition entry; VRE*)은 TDE의 변형이다. 응답을 입력하기 위하여 키패드를 사용하는 대신 응답자가 답을 전화에 말하고 컴퓨터는 응답을 되풀이하여 말해서 확인한다. 이 기술은 매우 간단한 정보교환을 하는 사회의 몇 가지 상황에서 상당히 흔하게 사용된다. VRE 적용의 몇 가지 예로 전화로 하는 은행업무, 자동화된 주문 입력, 자동화된 전화 연결, 그리고 음성 우편사서함이 있다. 말로 하는 답은 쉽게 오해되거나 컴퓨터가 인식을 하지 못하기 때문에 VRE를 이용하기가 어렵다. 그 반면, 이 기술은 어떤 종류의 전화로도 할 수 있고 응답자들은 이것이 TDE보다 더 간단하다고 생각하는 듯하다(Clayton과 Werking, 1998 참조).

Dillman(2000)이 말했듯이 TDE와 VRE는 한정된 기술들이다. 응답자들은 그들이 우편 조사에서 있는 개요나 면접원들이 면접 조사에서 주는 도움을 받지 못한다. 대부분은 모든 상황을 예상해야 하고 이러한 상황들을 대한 적절한 메시지를 명확히 말해야 되기 때문에 TDE와 VRE의 설계는 복잡하다. 모든 격려를 완전히 똑같은 방법으로 하는 것은 긍정적인 면인데 “실시간” 면접에서는 이것이 가능하지 않다.

지난 10년 동안 특히 민감한 주제를 다루는 조사에서 새로운 자료 수집 방식과 그것의 변형들이 인기를 얻었다. 이 방법은 컴퓨터 보조 자기 면접 (*computer-assisted self-interviewing*)(CASI) 이다. 이 장 앞에서 이야기 했듯이 면접이 민감한 주제나 사회적 기대 부응 편향의 경향이 있는 주제들을 다룰 때에는 면접원 보조 방법을 일반적으로 권하지 않는다. CASI는 대면 면접, 자기기입식과 컴퓨터 보조의 이점을 결합하는 컴퓨터 과학 기술이다. CASI에서는 면접의 일부를 (예를 들어, 민감하지 않고 면접원 보조로 이익을 얻을 수 있는 질문들) 면접원이 실시할 수 있다. 그렇지만 민감한 주제가 발생하면 컴퓨터를 응답자를 향하게 문자 그대로 돌리고 컴퓨터가 면접을 인계 받아 응답자에게 직접 민감한 질문들을 하게 한다.

이 문제에 대한 초기의 낮은 수준의 해답이 확률화응답기법이다(*randomized*

response (RR) technique)(Warner, 1965). 컴퓨터 보조 수집 방법의 도래로 확률화 응답기법은 면접의 일부나 전부를 위하여 응답자에게 노트북 컴퓨터를 본질적으로 넘겨주는 기법으로 대신하게 되었다. CASI (때로는 문자만의 CASI(*text-only CASI*))에서는 응답자가 컴퓨터 화면에서 질문들을 보고 컴퓨터 키패드를 사용해서 그들의 답을 입력한다. 이 형식의 CASI는 응답자들이 질문들을 읽을 수 있는 식자 수준을 가질 것이 요구된다. 이러한 요구 조건은 음성 CASI(*audio CASI*) (또는 ACASI)에서는 제거된다. 응답자는 ACASI를 사용해서 헤드폰으로 질문들을 듣고, 화면에 있는 질문들도 읽을 수 있고, 표시된 키를 눌러서 그들의 답을 입력한다. (문자만의) CASI와 ACASI 모두 면접원이 도구를 CAPI 노트북 컴퓨터에 넘겨주고 응답자가 질문에 답하는 일을 끝낼 때까지 컴퓨터를 회수하기 위하여 기다리는 것이 요구된다. 현실적이고 비용 효율이 높은 대안은 전산화 자기 면접을 전화 환경에서 실시하는 것이다. 전화 음성 CASI(T-ACASI; *Telephone audio CASI*)는 TDI와 거의 똑같고 응답자가 전화를 걸거나 또는 면접원이 전화를 거는 것을 통하여 할 수 있다. 근본적인 차이는 T-ACASI에서는 면접원이 먼저 전화를 걸고 그 다음에는 면접의 나머지는 면접원이 컴퓨터와 응답자에게 넘긴다. TDE에서는 보통 면접원은 참가하지 않는다. 통화는 보통 응답자가 컴퓨터로 하거나 컴퓨터가 응답자에게 한다.

민감한 속성의 측정을 훌륭히 하기 위해서 CASI의 변형들이 생기기 시작했다. 그러한 관점에서 보면 CASI의 변형들은 보고의 증가를 가져오기 때문에 성공적인 것 같다(Tourangeau와 Smith, 1998; Turner 등, 1998). CASI 발전은 CATI나 CAPI와는 근본적으로 다르다. 후자 기술이 면접원의 역할을 더 표준화하였다. CASI는 면접원과 응답자간의 상호작용을 완전히 바꾸고 면접원은 더 관리적인 역할을 맡는다. 그리고 수집 형식은 그래도 우편 형식에 비하면 이 방법의 모든 이점을 가지고 있는 “면접”이다. Turner 등(1998)과 Tourangeau와 Smith(1998)가 이야기 했듯이 음성 CASI는 대면 면접 뿐 아니라 자기기입식 종이 설문지로 측정된 특정한 성행위를 적게 신고하는 것을 줄인다.

6.1.6 행정기록

조사에서 어떤 자료 수집의 목적으로 행정 기록을 사용하는 것은 비용 효율이 높

은 고품질의 대안이다. 행정 기록(*administrative records*)은 결정이나 결심을 하거나 기록의 각 주제에 영향을 미치는 행동을 취하는데 사용되는 정보를 포함하는 기록이다. 대부분의 국가들은 면허 발급, 세금 부과, 단속, 임금 및 대금 등의 지불 등과 관련된 조치를 취할 수 있도록 기록을 보관한다. 물론 이런 자료를 통계적 목적으로 사용하는 것이 매우 흥미를 끌만하고 기록 보존의 전통을 가지고 있는 나라들에는 이런 등록을 바탕으로 한 여러 개의 수집 기록들이 있다. 어떤 나라들에서는 행정기록을 통계적 목적으로 사용한 긴 역사를 가지고 있다. 1장에서 언급했듯이 초기의 센서스들은 행정목적으로 만들어졌고 유럽의 일부 생명에 관한 기록 체계들은 300년이 넘었다. 컴퓨터 기술과 기록 연계 방법의 최근의 발달은 기록을 기본적인 수집 목적에서 뿐 아니라 일부 추정 과정에서 보조 정보로 사용이 가능하도록 하였다.

상황은 보통 다음과 같다. 몇 나라에서는 당국과 기관들은 사람, 기업, 또는 다른 실체의 집단들에 대하여 행정기록들을 만든다. 그것의 목적은 필요한 경우 각각의 실체들에 대하여 조치를 취하기 위해서이다. 과세기록부에는 모든 개인과 기업체가 기록되어 있고 각 기록은 직업, 소득, 나이, 업종, 종업원 수, 주소, 사업장의 수, 총 수입 등 같은 몇 가지 변수에 관련된 정보가 들어있다. 등록부를 관리하는 기관은 개인이나 기업의 정확한 세액을 계산하는 것과 같은 각각의 당면과제에 대하여 올바른 조치를 취할 수 있도록 당면과제에 대한 정보를 필요로 한다. 이런 기록을 이용하여 각각의 실체와 관련된 상황에서 실체들의 집단 전체에 대한 관심의 대상인 특징과 속성의 상황으로 옮겨가는 것이 가능하다(즉, 등록부나 기록 자료를 근거로 한 통계). 이를테면 직업 분포에 대한 표를 제공하거나 기록 연계를 통하여 조사 자료와 직업등록자료를 대응시키는 것이 가능하다.

행정자료와 다른 방법으로 수집한 자료와 비교했을 때 오류의 구조는 원칙적으로 차이가 없다. 그러나 통계전문가 사이에 행정 자료의 품질에 근거가 없는 신뢰를 하는 경향이 있다. 조사 자료의 평가에서 행정기록자료를 금본위(*gold standard*)와 같은 절대기준의 품질을 가지고 있다고 믿는다. 행정 자료는 유사한 조사 자료와는 완전히 다른 상황에서 수집되었기 때문에 이것이 정당화 될 수 있다. 하지만 행정 기록의 자료품질이 다른자료보다 떨어질 수도 있다.

행정기록의 자료품질에 대한 특징은 다음과 같다.

- 대부분의 경우 행정자료를 사용하는 조사연구자는 행정자료의 생성, 최신화, 코딩에 영향을 미치지 않는다. 예를 들어, 납세자들이 기재한 직업을 기록 보존자들은 대개 그대로 인정한다. 왜냐하면 그 변수에 대해 기록 보존자는 관심이 별로 없거나 더 정밀한 조사를 위하여 단지 모집단의 특정 부분을 가려내는데 그 변수가 사용되기 때문이다.
- 행정기록을 담당하는 사람들은 일반적으로 통계적인 예비지식이 없어서 적어도 어떤 변수들은 불완전하게 수집되고, 기록되고, 관리된다.
- 통계적인 적용과 행정상의 사용에는 대체적으로 개념의 차이가 있다.
- 행정자료의 품질은 거의 평가하지 않는다.

행정자료의 사용과 관련된 장점은 다음과 같다.

- 기록과 자료는 이미 존재하고 사용 비용도 저렴하다.
- 많은 적용에서 등록 단위와의 접촉이 필요 없으므로 응답자 부담도 없다.

행정자료의 사용에 대해서는 Zanutto와 Zaslavsky(2002)와 Hidiroglou 등(1995b)에 더 자세히 나와 있다.

6.1.6 직접관측

자료수집 방식으로서의 직접관측(*direct observation*)은 양이나 사건을 관측자의 감각(시각, 청각, 미각)이나 물리적인 측정장치를 사용해서 기록하는 것이다. 이 방식에서 관측자는 응답자의 역할을 한다. 직접 관측은 조사 연구에서 보기 드문 것이 아니다. 사용의 예로 수확물을 추정하는 것, 좌석 안전벨트를 매지 않는 운전자의 수를 세는 것, 전자 장치로 TV 시청을 측정하는 것, 기계적인 기구를 통해 공해를 측정하는 것, 대기의 사진을 해석해서 토지 사용을 검토 평가하는 것이 있다. 직접관측은 인류학과 심리학에서 오랜 전통이 있다. 그 전통의 일부는 응답자와 면접원의 행동을 관측을 통하여 연구하는 최근 인식조사연구로 옮겨갔다.

직접 관측의 오류는 두 종류가 있다. 관측자들이 그들의 감각을 사용하면, 나타나는 오류패턴은 면접원들이 내는 것과 매우 흡사하다. 관측자들은 기록해야되는 정

보를 잘못 인식할 수 있고, 그들 내 또는 그들 사이에서 변동이 생길 수 있고, 행동이 시간이 지나면서 바뀔 수 있다. 면접원들에게 사용되는 것과 같은 종류의 관리 프로그램이 관측자들에게도 사용될 수 있다. 두 번째 종류의 오류는 측정에 기계장치를 사용할 때 나타난다. 이런 장치를 사용하는 연구는 계통오차를 확실하게 방지하기 위하여 타당성 검사를 위한 프로그램과 도구의 계수조정이 있어야 한다 (Fecso, 1991 참조).

6.2 방식의 결정

모든 조사에서 어떤 자료수집 방식을 사용할 것인지 결정해야 한다. 이러한 결정과정에서는 몇 가지 요인을 고려해야 한다. 그러한 요인에는 원하는 수준의 자료품질, 조사예산, 조사설문지의 내용(질문 종류, 응답항목 수, 질문 수, 복잡함과 시각적 보조물의 필요성), 자료수집기간, 모집단 종류, 모집단에 대한 표본추출틀의 정보가 있다. 또한 자료수집을 표본 단위로의 접근방법과 가능한 자료처리 대안과 같은 조사과정의 다른 단계들과 함께 생각해 보는 것이 필요하다.

하지만 많은 대안들은 대개 비용면에서 비현실적이거나 특정 연구에 실용적이지 않기 때문에 이런 경우 방식의 결정은 하는 것은 수월하다. 오직 하나의 대안 방법만 있는 상황의 예로 여러 가지 종류의 지출에 대한 자료수집과 수확물에 대한 자료수집이 있다. 전자의 경우 여러 문헌에서 한결같이 일지방식이 제안되었다. 그 이유는 구매를 기억하기 위해 응답자가 구매를 계속적으로 기록하여 기억효과로 인한 중대한 문제점의 방지하는 것이 필요하기 때문이다. 일지방식을 사용하면 응답자가 구매를 하고 바로 기록하는 것이 가능하다. 집, 자동차, 세탁기, 투자 등의 구매를 제외하고는 응답자들이 특정 기간에 구매한 모든 것을 기억할 수 없기 때문에 어떤 다른 수집 방식도 적당하지 않다. 수확물에 관한 후자의 경우에는 관측자들이 눈으로 추정하는 방법이나 표본지역의 작물을 수확하는 것이 유일한 수집 방식이다.

여러 대안이 있을 때 방식의 결정은 더 복잡해진다. 어떤 결정 요인은 비교적 평가하기 쉽다. 예를 들어, 비용은 비교적 쉽게 예측할 수 있고 표본추출방법과 무응답추적 체계와 같은 적절한 설계 요인들이 확립되었을 때도 무응답률은 쉽게 예측할 수 있다. 설계자가 측정오차를 중요하게 여긴다면 결정은 더 어려워진다.

측정오차를 최소화하는 관점에서 보면 어느 자료수집 방식이 최선인지 결정하는 두 가지 방법이 있다(Lyberg와 Kasprzyk, 1991, de Leeuw와 Collins, 1997 참조). 하나는 각 방식과 관련된 일반 원칙 또는 “경험 법칙(*rule of thumb*)”의 적용할 수 있다. 예를 들어, 만일 조사에서 표본의 크기가 엄격히 정해지면 우편이나 대면 면접은 사용하기 어렵다. 이것은 응답률을 과대 또는 과소평가한 경우 조정이 쉽지 않기 때문이다. 조사 예산이 적을 때에는 우편조사만이 선택할 수 있을 것이다. Branches와 skips가 있는 복잡한 설문지는 우편조사에 알맞지 않다. 만일 사회적 기대부응 편향이 걱정되면 면접원 보조 방식은 문제가 있다. 응답항목이 많은 질문은 면접 환경에서 잘 운영되기 힘들지만 대면 면접에서는 플래시 카드(즉, 응답자에게 응답항목을 보여주는 카드인데 그것에 의하여 “자기 기입 상황”을 만든다)를 사용할 수 있다. 만일 응답률이 높아야 한다면 대면 면접이 더 좋다. 만일 지리적인 분산이 요구되면(즉, 표본 단위의 지리적 집락이 매우 적은 표본들) 대면 면접은 비용이 너무 많이 든다. 면접원 관리는 집중식이 아닌 전화면접이나 대면 면접 보다 중앙집중식 전화면접 환경에서 시행하기 쉽다. 일반적인 경험법칙의 예는 계속 다룬다.

일반적으로 방식의 선택은 연구자가 평가하는 각 방식이 가지고 있는 다양한 장단점의 우선순위를 근거로 한다. 방식 비교에 대한 사전 연구는 선택에 많은 도움이 된다. 연구대상의 특정 현상에 어떤 자료수집 방식이 선호되는지에 관한 정보가 거의 없다면 그 문제를 검토하기 위해 연구를 실시해야 한다.

방식에 대한 비교연구를 설계하는데 있어서 어떤 종류의 방식효과를 측정하는지를 선택해야한다. 그것은 순수 방식효과 또는 방식체계효과이다. 순수 방식효과(*pure mode effect*)는 방식에 원인이 있는 측정편향이다. 순수 방식효과를 정확하게 측정하기 위해서는 두 가지 수집의 유일한 차이가 방식인 수집들의 결과를 비교해야 한다(즉, 두 가지 수집에서 같은 질문을 하고, 같은 특징을 측정하고, 특정한 기준기간 동안 같은 모집단을 연구한다). 두 가지 수집에서 나온 응답분포의 차이는 방식간의 차이의 추정값이 된다(Biemer, 1988; Groves, 1989).

하지만 이러한 방식은 실제에서는 더 복잡해진다. 설명한 방법의 추정값은 연구실 실험이나 현실성이 불확실한 현장 시험에 의존한다(10장 참조). 모든 설계 요인들을 그대로 두고 방식만을 바꾸는 상황이 가능하거나, 실행할 수 있거나, 타당한 경우는 매우 드물다. 이것은 복잡한 대규모 자료수집의 현실과는 거리가 있고 매우

간단하고 비현실적인 적용에서만 가능하다. 더욱이 조사전문가의 주된 관심은 순수한 방식효과에만 있는 것이 아니라 방식과 다른 설계 요인들의 결합된 효과에도 있다. 다음 예가 실제 상황에서 순수 방식효과를 실험하는 것에 대한 어려움과 바람직하지 않음을 설명한다.

예 6.2.1 우리가 전화면접과 대면 면접을 비교하기를 원한다고 가정하자. 대면 면접을 시행하는 기관의 면접원의 능력은 대체로 면밀하게 모니터하는 중앙집중식 전화면접기관과는 아마도 다를 것이다. 전화조사와 대면조사의 조사모집단은 다를 수 있다. 그 이유는 전화로 모든 사람들에게 연락을 할 수 없기 때문이고 이 결과 포함의 차이가 생긴다. 면접원이 보여주는 시각적 보조물을 사용하는 대면설문지는 전화면접에서 사용될 수 있는 것으로 쉽게 바꿀 수 없다. 그러므로 만일 조사전문가가 대면 면접에서 전화면접으로 변환하는 것을 생각하면 방식만이 바뀌는 것이 아니다. 전화면접의 공정한 시험을 실시하기 위하여 면접원, 설문지, 실시기간, 성과 모니터링 등과 같은 설계요인들을 전화 방식에 맞도록 최적화해야 한다. 면접 방식 외에 다른 많은 설계요인들이 두 방식 사이에 달라야한다.

방식의 비교에 대한 이러한 심각한 문제에서 벗어나는 방법은 방식체계효과(*mode system effect*)를 대신 측정하는 것이다. 방식의 효과를 분리시키려고 하기보다 자료수집의 전체 체계를 비교하는데 방법이다. 방식체계효과는 순수 방식효과보다 측정하기 더 쉽고 현실적이다. 여기서, 체계는 특정한 방식으로 설계된 자료수집 과정 전체를 의미한다. 체계는 면접원 고용, 면접원 교육, 면접원 감독, 설문지 내용, 전화를 다시 거는 시도의 수, 거부전환 전략, 표본추출 체계와 표본추출틀 포함 같은 설계 요인들을 포함한다. 두 체계의 차이는 응답분포, 단위무응답률, 항목무응답률, 주어진 응답률을 성취하는데 필요한 시간, 그리고 자료수집 비용과 요소들 중 몇가지 측면에서 측정할 수 있다. 높은 응답률이 낮은 응답률 보다 좋고, 낮은 비용이 높은 비용 보다 좋고, 짧은 생산기간이 긴 것보다 낫기 때문에 이들 요소 중 일부는 분명한 비교 수단이 된다. 측정편향을 반영하는 크기와 같은 특정한 항목의 응답분포에 대한 문제인 경우 두 결과를 비교할 수 있는 금본위(*gold standard*)와 같은 절대기준이 있는 것이 아니라면 체계 차이를 평가할 수 없다. 어떤 특징에 대해서는 한 응답분포의 결과가 다른 것이 비해 우수하다고 말할 수 있다. 예를 들어, 우리는 특정한 행동이나 사건의 더 잦은 보고는 더 낮은 기억오차나 더 낮은 사회

적 기대부응 편향의 표시로 여긴다. 그런 경우에는 “더 많은 것이 더 낫지만” 평가 연구를 실행한 경우가 아니면 아무도 확신할 수 없다(Moore, 1988; Silberstein과 Scott, 1991).

요즈음은 혼합 방식 조사가 일반적이기 때문에 자료수집체계는 대개 한 가지 방식으로만 이루어지지 않다. 예를 들어, 우편 같은 비교적 저렴한 방식으로 시작해서 몇 번의 추적 시도 후에 우편 무응답자들에 대하여 전화면접을 사용하는 경우가 흔하다. 어떤 조사에서는 전화가 없는 우편 무응답자들을 추적하려고 대면 면접을 한다. 또한 면접원 보조 방식으로 시작하고 우편 무응답자들을 추적하는 우편조사를 실시하는 조사의 예도 있다. 어떤 응답자들은 면접을 하는 것 보다 오히려 우편설문지를 받는 것을 더 선호한다 (Carroll 등, 1986).

순수 방식효과 문제가 비록 비실용적으로 보일 수 있지만 위의 상황을 생각하면 이것은 중요하다. 그러므로 대부분의 조사는 처음부터 두 가지 이상의 방식을 이해해야 한다는 사실을 감안하면 방식효과는 고려되어야 할 중요한 설계요인이다. 첫단계는 중요한 모든 설계요인들을 참작하여 조사 상황을 가장 잘 수용할 수 있는 주된(*main*) 방식을 선택하는 것이다. 주된 방식을 미리 정한 재통화 시도나 독촉의 제한 회수까지 최대한 사용하고 그 후에 응답률을 늘리는 중요한 목적으로 다른 방식을 사용한다. 이것은 필요하고 가장 실용적인 접근 방법이지만 타협적인 전략이다. 만일 주된 방식이 현재의 조사측정 상황에서 적당한 방식으로 간주되어 선택되었으면 다른 방식에서의 전환은 품질이 떨어지게 할 수도 있다. 그러므로 보완 방식의 효과는 되도록 작게 하는 것이 중요하다.

어떤 조사들에서는 동일한 설문지가 모든 방식에서 사용될 수 있기 때문에 방식효과가 작다. 대부분의 문제는 우편방식과 면접원 보조 방식의 결합에서 발생한다. 설문지가 매우 간단해서 면접에서 쉽게 사용될 수 있는 경우 이외에는 우리는 곤경에 빠지게 된다. 응답항목이 많은 우편조사는 전화면접에는 적당하지 않다. 이를테면 만일 특정한 질문에 대하여 8개의 응답항목이 있으면 응답자가 전화 추적에서 그것을 다 기억하고 구별하기 매우 힘들다. 또한 면접원이 모든 응답항목들을 읽기도 힘들고 불편할 수 있다. 이런 경우에는 응답자가 앞에 나오는 몇 개의 응답항목에서 선택하거나 면접원이 상당히 적절한 항목을 읽자마자 응답자가 면접원이 읽는 것을 중단시키고 면접원이 순조롭게 면접을 하도록 남은 항목들을 읽는 것을 건너

뛰게 하는 위험이 있을 수 있다.

만일 일부 질문들이 민감해서 우편설문지를 선택했다면 면접원 보조 방식으로의 전환은 위험할 수 있다. 전환은 더 큰 측정오차를 가져올 수 있고 이러한 응답 품질의 손실을 응답률의 증가로 생기는 이익과 비교 검토해야 한다. 특정 방식이 그 방식의 측정 특성 때문에 선택되었다면 예를 들어 응답률을 늘리기 위한 목적으로 다른 방식으로의 전환이 조사계획 단계에서 다루어지기는 매우 어렵다. 흔히 이러한 trade-off 상황을 다루기 위하여 통제를 최소화하는 특별한 전략들이 채택된다. 그러나 설문지 형태에 대한 계획은 별로 어렵지 않다. 우편 설문지에 나오는 질문은 면접방식과 자기기입식 상황에도 적절하게 작성될 수 있다. 그러나 이렇게 작성된 질문은 불행히도 측정 능력이 감소하게 된다(4장 참조).

전화면접에서 대면 면접으로의 전환은 별다른 문제가 없다. 대면 면접원은 거부 전환의 선택권과 응답자를 유도하고 도와줄 기회가 전화면접원보다 더 많다. 또한 대면 면접원은 시각적 보조물을 사용할 수 있는데 이것은 전화면접에서는 가능하지 않다.

더 높은 응답률을 얻기 위하여 한 방식을 한두 가지의 다른 방식들과 결합시킨 혼합 방식 전략이 사용되는 상황은 이미 이야기 했다. 혼합 방식 전략이 적절한 다른 경우들도 있다. 모집단에서 혼하지 않은 어떤 속성을 가진 부모집단을 가려내는데 한 가지 방식을 사용하고, 두 번째 방식은 부모집단의 그러한 속성과 그것과 관련사항을 측정하는데 사용한다. 예를 들어, 전화조사가 스웨덴에 상품을 수출하는 미국에 있는 기업들을 확인하는데 사용될 수 있다. 그 결과로 얻은 부모집단을 우편설문지나 대면 면접 방식으로 조사한다.

혼합 방식 전략의 세 번째 적용은 전략의 목적이 비용을 줄이는데 있는 패널 조사에서 흔히 볼 수 있다. 패널 연구는 응답자의 시간에 따른 변화를 측정하기 위하여 주기적으로 같은 표본 단위로부터 새로운 측정값을 수집하는 조사이다. 각 새 면접 round는 “파동(wave)”를 이룬다. 미국에서 실시한 Current Population Survey에서 대면과 전화 면접의 혼합된 방식이 사용된다(미국 노동부와 미국 센서스국, 2000). 패널 조사에서 혼합 방식 전략은 각 방식의 장점들을 활용한다. 패널 조사의 첫 파동에서는 대체로 대면 면접이 사용된다. 이것은 차후의 패널 round에서 협조와 응답을 유지하기 위한 응답자와의 관계를 수립하면서 동시에 전화가 없는 단위의 불완전한 포함의 문제를 해결한다(Kalton 등, 1989). 전화 면접(아마도 비집중식,

즉, 면접원들의 집에서 전화 면접을 하는 것)은 응답자가 대면 방식을 선호하거나 전화를 사용할 수 없을 때를 제외하고는 차후의 과동에서 사용된다.

지난 20년 동안 조사 자료 수집 방식은 과학 기술의 발전에 따라 변화되어 왔다. 이제 대부분의 경우 컴퓨터를 이용하는 방식으로 자료가 수집된다. 전화가 아직도 지배적인 방식이지만 자기기입식은 몇 가지 이유로 부침을 겪었다. 첫째, 자기기입식은 사회적 기대부응 편향이 작기 때문에 민감한 자료의 수집에서 최선의 방식이다. 둘째, 자기기입식은 컴퓨터를 이용할 수 있게 되었고, 셋째, 면접 방식에 비하여 저렴하다.

하지만 많은 나라에서는 문맹률이 너무 높아서 자기기입식의 사용이 불가능하다. 또한, 조사연구자가 모집단의 구성원을 조사할 때에 대면 면접에 의존해야하는 개발도상국들을 포함한 많은 나라의 전화보급률은 아직 낮다. 어떤 환경에서는 선택할 수 있는 자료수집 방식이 조사 주제에 관계없이 오직 대면 면접 한가지 밖에 없다. 이와 같이 나라와 조사 대상의 모집단에 따라 조사의 환경이 매우 다르다. <표 6.2>에는 주된 방식의 선택을 결정하는 요인들의 예가 제공되어 있다.

6.3 방식효과의 예

순수 방식 효과를 추정하기 위한 연구는 설계하기 어렵고 반드시 적절하지만은 않다는 것을 보았다. 대신에 방식 비교 연구들은 보통 자료수집 체계들 전체를 비교하도록 설계되었다(즉, 모든 요소가 있는 각 체계의 결과를 적절한 평가기준에 의하여 평가한다). 이런 기준의 예는 단위와 항목 무응답률, 보고의 완성도, 응답 분포의 유사성, 사회적 기대부응 응답의 존재, 서술형 질문에 대한 응답의 길이, 수집기간의 길이, 그리고 수집 비용이다.

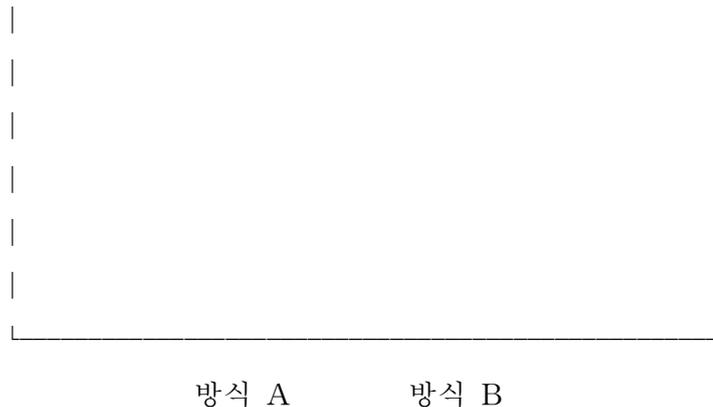
<표 6.2> 자료수집 방식을 결정하는데 고려할 요인

요인	방식선택의 implication
측정할 개념	시각적인 전달 방법이 필요하면 전화 조사는 제외한다. 개념이 복잡하면 면접원 도움이 필요하다.
조사할 목표모집단	전화 비사용인구를 무시해도 되면 전화조사를 고려한다. 식자수준: 우편방식은 국가의 평균이나 그 이상의 식자율이 필요하다.
추출틀에서 얻을 수 있는 연락 정보	이름과 주소이면, 우편이나 대면 면접을 고려한다. 전화번호이면 전화면접을 고려한다.
주제의 중요점	만족스러운 응답률을 얻기 위하여 설득이 필요하다면, 우편조사는 제외시켜야한다.
완성 속도	빨리 필요하다면, 전화가 최선이다. 몇 주 내에 필요하다면 우편 조사도 가능하다.
표본의 규모와 크기	지역조사는 세 가지 주요 방식이 모두 가능하다. 전국적인 조사는 우편이나 전화조사를 제안하는데 비용이 지배적인 요인이다.
표본 산포	산포가 크면 우편이나 전화조사를 제안한다. 대면조사에서는 약간의 집락은 거의 항상 필요하다.
목표 모집단의 추출틀	빈약한 포함추출틀만 이용 가능하면, 대면조사, RDD, 또는 혼합방식을 사용한다. 좋은 포함은 방식 선택에 있어서 융통성을 제공한다. 주소를 입수할 수 있으면 사전편지와 미리 지불하는 인센티브를 가능하게 한다.
무응답	면접 방식은 자기기입식보다 높은 응답률을 낸다. 망설이는 표본 단위를 설득하는 능력은 매체의 풍부함에 달려있다(예를 들어, 우편조사에서는 자극은 쓰여진 것으로 제한되어 있다). 무응답은 우편과 전화조사에서는 포함 문제와 혼동되어 있다. 표본단위는 우편설문지를 필요없는 우편물로 간주하고 버릴 수 있다. 우편설문지는 식자문제 때문에 모집단의 일부는 완성할 수 없다.
면접원	면접원들은 사회적 기대부응 편향 같은 응답오차를 일으킬 수 있다. 면접원 보조 방식은 민감한 정보의 수집에는 좋지 않다. 시각적 보조물과 탐색조사를 위해서는 면접원이 필요하다. 중앙집중식 전화면접은 비집중식 면접과 비교해서 비용과 오차를 줄인다. 전화면접원들은 이동 부담이 없기 때문에 더 많은 작업량이 주어진다.
응답자	응답자들은 자기기입식 조사를 선호한다는 증거가 있다. 자기기입식 방식들은 민감한 정보를 수집에 적당하다. 응답이 어려우면 면접원 도움이 필요하다.
도구	우편설문지는 비교적 간단해야하지만 응답항목이 많이 포함된 질문들의 경우에도 사용이 가능하다. 복잡한 도구는 면접 방식을 필요로 한다. 혼합방식은 모든 방식에서 사용할 수 있는 설문지를 사용해야한다.
비용	만일 우편조사를 할 수 있는 여유만 있으면 다른 모든 방식은 부차적이다. 전화면접이 대면방식보다 덜 비싸다.

De Leeuw와 van der Zouwen(1988)은 대면과 전화 면접을 비교하는 28개의 연구의 메타분석을 실시했다. 대면과 전화 면접을 비교한 결과를 마치 하나의 큰 연구를 실시한 것처럼 요약했다. 가장 큰 효과는 서술형 질문에 대한 응답의 길이에서 얻었고 그리고는 대면 면접이 가장 우수하다. 또한 전화로는 사회적 기대부응 편향의 경향이 적다. 민감한 변수들에 대하여 상당한 방식효과가 있다. 그리고 만일 시각전달이 필요하지 않으면 일반적인 주제들에 대해서는 전화 면접과 대면 면접의 품질은 대체로 같았다.

두 방식체계 A와 B대한 비교 연구에서 결국은 두 개의 추정값이 나온다. 하나는 방식체계 A의 결과이고 다른 하나는 방식체계 B의 결과이다. 설사 평가 기준이 있다고 해도 어떤 방식의 품질이 더 나은 것으로 평가되기는 어렵다. 그 방식이 가장 작은 측정오차를 내는 것일까, 가장 높은 응답률을 내는 것일까, 아니면 다른 평가 기준에서도 우수한 것일까? 해답은 연구자가 중요하다고 생각하는 것이 무엇인지에 따라 달렸다. 품질 요소들은 흔히 서로 상충하고 설사 우리가 자료 품질에 대하여 높은 관심을 가지더라도 비용과 조사결과를 내는데 걸리는 시간 같은 요인들은 무시될 수 없다(<그림 6.1> 참조).

두 가지 방식의 가설적 실험의 결과



<그림 6.1> 어느 방식이 더 좋을까? 방식비교연구의 결과는 결론이 나지 않을 수 있다. 이 가설적 예에서 두 방식은 주목할 만하게 다른 모수 추정값들을 가져오지만 이 정보만으로도 더 나은 방식을 결정하기 어렵다(즉, 실제 모수값에 더 가까운 추정값을 가져오는 방식). 그렇지만 만일 연구 대상인 특징에 대한 편향의 방향에 대한 정보를 입수할 수 있으면 더 나은 방식을 확인할 수 있다. 예를 들어, 만일 편향이 마이너스 경향이 있다는 것을 알면 (즉, 추정값들이 모수를 과소평가하는 경향이 있다), 방식 A가 더 높은 추정값을 가져오기 때문에 더 좋다. 만일 편향이 플러스인 경향이 있으면 방식 B가 더 좋다.

방식 연구에서 가장 큰 문제는 우리가 측정 편향에 대하여 많이 못한다는 것이다. 정보를 얻는 한 가지 방법은 6.1.6절에서 이야기했듯이 추정값 A와 B를 절대기준과 비교하는 것이다. 이런 평가 연구를 실시하는 것이 가능하긴 하지만 평가 연구는 비용이 많이 들고 따라서 제한된 규모로만 실시할 수 있기 때문에 결과의 신뢰도가 떨어진다. 그 대안은 측정 편향을 방향을 가망성이 높은 쪽으로 가정하는 것이다. 일반적으로 민감한 항목들은 실제 이하로 보고 되고 사회적 기대부응 편향이 제거된 방식들에서 더 많은 보고가 있다는 압도적인 증거가 있다.

이 현상의 실례가 <표 6.3>과 <표 6.4>에 나와 있다. 여기에서는 더 많은 보고가 더 나은 보고라고 가정하며 민감한 변수에 대해서는 이것이 보편적인 사실로 보인다. 그렇지만 어떤 규칙에도 예외가 있다. 무엇이 민감한 변수인가에 대한 생각이 나이와 성별에 따라 다른 것은 당연한 일이다. 예를 들어, 10대의 남자 청소년들은 성경험에 대하여 물었을 때 실제 이상으로 보고한다는 증거가 있다. 그룹에 따라 보고 패턴이 다를 수 있는 민감하게 생각하는 다른 변수들도 있다. 사회적 기대부응 편향이 플러스인 소득이 그것인데 특정 소득 수준의 구성원들은 대개 소득을 실제 이상으로 보고한다. 그러므로 평가연구에 의하여 그것을 추정해야 측정 편향이 확실해진다.

<표 6.3> 여성 건강 연구에서의 방식효과

실험 그룹	평균 성 상대자 수			불법 약물 사용 인정률(%)
	최근1년	최근 5년	평생	
자기기입식 질문(SA)	1.72	3.88	6.54	40.9
대화식 SA	1.56	3.37	6.88	42.5
컴퓨터 보조 SA	1.89	4.40	6.25	39.3
면접원 보조 질문(IA)	1.44	2.82	5.43	40.7
대화식 SA	1.56	2.86	4.58	39.3
컴퓨터 보조 SA	1.36	2.79	6.27	42.2

출처: Tourangeau와 Smith(1998)과 Tourangeau 등(1997)

<표 6.4> 방식에 따른 민감한 속성의 추정값

A. 한번 이상의 낙태를 한 미국 여성의 퍼센트

면접원기입식 설문지에 나온 최초의 보고	ACASI에 나온 보고		
	없음	1회	2회 이상
없음	95.4	3.5	1.0
1회	1.8	92.5	5.8
2회 이상	0.4	2.2	97.4

B. 남성간의 성접촉을 보고한 15-19세의 미국 남성의 퍼센트

방식	남성간의 성 접촉 보고율 (%)
자기기입식 PAPI	1.1
ACASI	4.7

C. 미국 인구의 민감한 행동의 Prevalence의 추정값

방식	한정된 성경험 (%) (최근 6개월간 성경험 없음)	가장 최근의 성관계가 6개월 미만 지속됨 (%)
CAPI	1.5	5.8
T-ACASI	8.0	21.3

출처: Turner 등(1996, 1998)

질문 순서 효과가 전화와 대면 면접 사이에 다를 수 있다는 것이 Cannell 등 (1987)의 연구에서 발견 되었다. 또한, 건강에 대한 보고는 대면 면접에서 보다 전화면접에서 항상 높았다. 그렇지만 방식들 사이의 실제 차이는 작았다. Sykes와 Collins(1988)는 영국에서 대면과 전화 면접을 비교하는 세 건의 실험 연구를 실시 했다. 특히 서술형질문에 대한 답들은 전화방식에서 더 짧은 경향이 있었고 이것은 전화방식이 더 빠르게 진행된다 사실로 설명될 수 있다. 방식 문제는 복잡하기도 하고 그렇지 않기도 하다. 순수 방식 효과가 평가하기 매우 어렵다는 면에서 보면 방식의 선택은 복잡하다. 그러나 비용이나 주제와 같은 설계 제한 때문에 실제로는 선택이 간단해진다는 것을 생각하면 방식의 선택은 복잡하지 않다. 흔히 만족스러운 응답률과 비용 효과를 감안하여 방식들을 결합시켜야 하지만 일부 주제와 질문 들에서 적당하지 않은 방식들도 있다.

제 7 장 자료처리: 오류와 오류의 관리

자료처리는 조사 자료를 자료수집 과정에서 얻은 가공되지 않은 결과를 분석, 발표, 보급에 사용될 수 있도록 가공되고 교정된 상태로 변환시키는 작업이다. 이 과정에서 정확도를 개선하는 목적의 몇 가지 작업에 의해 자료가 고쳐질 수 있다. 조사관리자가 “사용하기에 알맞다”고 결과를 만족할 때까지 자료는 점검되고, 비교되고, 교정되고, 컴퓨터에 입력되거나 스캔되고, 코드화되고, 표로 만들어 질 수 있다. 자료처리 단계의 일련의 작업에는 단순한 것부터(예를 들어, 자료 입력) 에디팅, 대체, 가중 등과 같은 복잡한 것까지 있다. 자료 입력, 에디팅, 코딩과 같은 작업은 비용과 시간이 많이 든다. 따라서 자료처리 작업은 점점 더 자동화 되어가고 있다. 최근의 과학기술의 혁신은 수작업 요원의 수요를 축소시켜서 작업 수행 방법에 큰 영향을 미쳤다. 과학 기술은 자료처리와 조사 과정을 더욱 통합시켰다. 예를 들어, 자료처리의 일부 단계는 자료수집 단계에서 할 수 있고 그 결과 자료의 정확도가 향상되고 비용과 총 소요시간이 줄었다.

자료처리를 수작업으로 하면 사람에 의한 오류가 생기기 쉽다. 수작업의 의존을 줄이면 자동화가 수작업 처리에서 생기는 자료의 오류는 줄이지만 사용된 기술 특유의 다른 종류의 오류를 범할 수 있다. 그러므로 자료처리 단계에서는 사람뿐만 아니라 기계로 인한 오류를 관리하는 다양한 조치를 해야 한다.

일부 조사에서는 자료처리 단계에서 생기는 오류들이 무시되는 경향이 있다. 일반적으로 조사 기관은 자료처리와 그것과 관련된 오류에 대한 지식이 매우 부족하다. 때로는 아무런 품질 관리를 위한 노력을 하지 않고 작업을 하고, 국가적인 매우 중요한 자료를 제외하고는 MSE로 측정되는 전체 정확도에 오류가 미치는 영향을 알지 못한다. 이 장에서는 주된 자료처리 작업, 그것의 오류의 구조, 그리고 이 오류들을 관리하고 줄이기 위해 취할 수 있는 조치를 설명한다.

7.1 자료처리 단계

자료처리 단계는 조사의 자료수집 방식과 자료처리에 이용되는 기술에 따라서 다르다. <표 7.1>은 PAPI 설문지의 자료처리 단계들을 보여준다. 자료 입력 단계를

생략할 수 있는 것을 제외하고는 CAI 설문지에서도 비슷한 단계들을 사용한다. PAPI의 처리 단계는 다음과 같이 이루어져 있다.

<표 7.1> PAPI 설문지의 자료처리 단계

처리단계	설명
1. PAPI 설문지	설문지를 현장에서 수집하고, 작업 단위를 구성한다.
2. 스캔 에디팅	빗나간 표시, 모호함 등의 자료 입력 문제들을 피하기 위하여 항목들을 검사한다.
3. 자료 입력	설문지 자료를 입력, 스캐닝 또는 광학상의 감지를 통해서 등록시킨다.
4. 에디팅	논리적인 에디팅을 등록된 자료에 실행시킨다. 대체와 같은 수정 조치가 에디팅의 일부이다.
5. 코딩	서술형 질문들을 코딩한다.
6. 파일 준비	자료에 가중치를 가하고 검사한다. 파일들을 공공 및(또는) 고객의 사용을 위해 준비한다.
7. 결과물 파일	결과물 파일을 분석하고, 문서화하고, 고객에게 전달한다.

1. PAPI 설문지(PAPI questionnaire)는 연구 대상인 변수들의 정보 수집에 사용된다. 이러한 변수들에 해당하는 질문들 중에는 면접원(또는 응답자)이 응답항목을 나타내는 상자에 표시하는 것을 요구하는 선택형이 있다. 예를 들어, 혼인 상황은 “1=미혼,” “2=기혼,” “3=이혼,” “4=미망인/홀아비,” “5=결혼한 적 없음”으로 코딩될 수 있다. 만일 질문에 답을 하지 않았으면 “98=빈칸”으로 코딩하거나 만일 응답자가 면접원 보조 방법에서 질문에 답하는 것을 거부했으면 “99=거부”로 코딩할 수 있다. 서술형 질문에서는 설문지의 공란에 자유 형식으로 응답을 적는다. 예를 들어, “당신의 직업이 무엇입니까?”라는 질문에서 “저는 SAS 항공사 승무원입니다” 또는 “배관공”이 응답될 수 있다.

2. 스캔 에디팅(scan edit)은 몇 가지 단계를 가지고 있는 예비 작업이다. 첫째, 조사 기관들은 설문지를 받으면 바로 수납 관리 부서로 보내서 표시하고 빈 페이지나 설문지가 사용 가능하려면 필히 완성되어야 하는 중요 항목이 빠져있는 뚜렷한 문제가 없는지 점검해야 한다. 면접원 보조 방법에서는 이 과정에서 완전하지 않아서 거부당한 설문지들을 면접원들에게 완성하도록 돌려보내질 수 있다. 우편 방법에서는 설문지를 완성하기 위하여 전화추적 과정으로 보내질 수 있다. 무응답 추적

이 없는 경우에는 설문지가 자료 입력에게 전달될 수 있지만 결국은 완전하지 않은 자료 때문에 무응답자로 코딩될 수 있다. 다음 단계에서는 설문지들을 그 이후의 처리 단계에 용이하도록 작업 단위(*work units*)라고 하는 작은 묶음으로 나눈다. 마지막으로 이 작업 단위들은 그것들이 다음 단계로 진행될 준비가 되어 있는지를 확인하는 마지막 검사를 받는다. 추가적인 검사가 과정의 후반에 다시 실시되므로 스캔 에디팅은 신속하게 진행된다.

3. **자료 입력(*data entry*)**은 설문지 자료를 컴퓨터가 읽을 수 있는 형태로 변환시키는 과정의 단계이다. 자료는 입력하는 기기를 사용하여 수작업으로 입력되거나 스캐너나 광학인식 도구를 사용하여 자동으로 입력될 수 있다. 이러한 과정의 단계를 때로는 자료 저장(*data capture*)이라고 부른다.

4. **에디팅(*editing*)**은 자료로 저장된 응답들은 타당한지 지 않으면 자료를 바로잡는 방법이다. 각 변수 또는 변수들이 조합에 대하여 에디팅 규칙이 만들어진다. 에디팅 규칙들에 의해 변수나 변수의 조합의 가능한 값들이 허용 가능한 값의 구간으로 구체적으로 지정된다. 에디팅을 통하여 확실히 잘못된 것으로 의심이 가는 명백한 오류와 응답이 드러난다. 그것들의 일부는 수정을 위한 조치가 취해진다. 때로는 면접원이나 응답자에게 다시 연락을 해서 빠진 값들을 채워 넣거나 기록된 값들을 고친다. 다른 경우에는 기록에 있는 다른 정보를 근거로 올바른 값을 추론하는 방법으로 값을 채워 넣거나 바꾼다. 대부분의 에디팅은 특별히 설계된 소프트웨어로 자동으로 시행된다.

5. **코딩(*coding*)**은 서술형 응답을 숫자나 알파벳으로 이루어진 코드번호로 식별되는 미리 정해진 범주로 분류하는 절차이다. 예를 들어, “당신의 직업은 무엇입니까?”라는 서술형 질문에 많은 수의 다른 답이 있을 수 있다. 이 정보를 차후의 분석에서 사용할 수 있도록 각 응답에는 특정 직업 범주를 식별하는 작은 수(이른바 400에서 500)의 코드번호 중의 하나가 지정된다. 서로 다른 기관에서 실시하는 여러 조사에서 직업 범주가 일치하도록 하기 위하여 표준직업분류(*standard occupation classification; SOC*) 체계가 사용된다. 전형적인 SOC 코드북에는 수백가지의 직업명 및(또는) 직무 내용의 설명이 각각 대응하는 세 자리 코드번호와 같이 있다. 대부분의 분류 기준에서 첫 자리는 넓은 또는 주된 범주를 나타내고 두 번째와 세 번째 자리들은 더 상세한 범주를 나타낸다. 따라서 “승무원”이라는 응답에 대해서 코더가 SOC책을 참고하여 승무원의 코드번호를 찾아본다. 코드번호가

712라고 가정하자. “7”은 “식당 서비스 근로자”라는 주된 범주에 대응시킬 수 있고, 71은 “웨이터와 웨이트리스”에, 그리고 712는 “승무원”에 대응시킬 수 있다. 자동화된 코딩(*automated coding*)에서는 대부분의 경우에 컴퓨터 프로그램이 코드번호가 배정되고 컴퓨터에 의해 정확한 코딩되기 어려운 경우에는 수작업으로 코딩된다.

6. 파일 준비(*file preparation*) 단계에는 몇 가지 작업이 포함된다. 먼저 표본단위의 가중치를 계산해야 한다(9장 참조). 가중치는 흔히 3단계로 정해진다. 첫째, 표본 구성원이 선택될 확률의 역인 기본 가중치를 구한다. 둘째, 무응답오차를 보정하기 위하여 기본 가중치를 조정한다. 셋째, 외부 정보를 입수할 수 있는지에 따라서 추출률 포함오차를 조정하기 위해 가중치를 추가 조정을 해야 할 수 있다. 이러한 조사 후 조정은 추정값의 정확도의 추가 개선을 위해 실시된다. 또한 파일 준비에서는 공개 회피와 기밀성을 목적으로 자료의 은폐를 할 수 있고, 빠진 자료나 잘못된 자료는 대체(*imputed*)할 수 있다(즉, 잘못된 또는 빠진 값을 자료 파일의 다른 응답을 바탕으로 하는 통계 모형에서 예측한 값으로 대신한다). 파일 준비 단계는 결과물 파일의 역할을 하는 분석 파일을 만들어내고 결과물 파일은 통계를 내고 분석을 하는데 사용된다.

7. 자료 분석(*data analysis*) 단계는 조사 자료의 생산자, 고객이나 조사 의뢰자, 또는 다른 자료 사용자가 실시할 수 있다. 이들의 어떤 조합이 실시할 수도 있다. 생산자가 일련의 분석을 실시하고, 다른 사람들이 분석 파일을 사용하여 다양한 부차적인 분석을 실시할 수 있다. 결과물 파일이 바르게 기록되어 있는 것이 중요하다. 한 가지 흔한 문서화는 흔히 자료 기록 설명서(*data record description*)라고 하는 포괄적인 자료 구조의 제공이다. 이것은 파일에 어떤 변수들이 있는지, 그리고 각 변수와 관련된 모든 코드번호들이 상세히 기록된다(예를 들어, 결혼 상황 변수는 1=“미혼,” 2=“기혼,” ..., 98=빈칸, 99=거부).

새로운 과학기술로 인하여 자료처리 단계들의 통합 가능성이 열렸다(Couper와 Nicholls, 1998 참조). 그러므로 어떤 조사에서는 단계의 순서가 위에서 설명한 것과 매우 다를 수 있다. 예를 들어, 자료입력저장과 코딩을 한 단계로 통합하는 것이 가능하다. 마찬가지로, 자료입력저장과 에디팅을 코딩과 통합할 수도 있다. 또한 에디팅과 코딩을 CAI 기술의 사용을 통해서 자료수집과 통합하는 것도 가능하다. 이런 종류의 통합의 이점은 자료의 불일치나 코딩을 하기에 충분하지 않은 정보는 응답

자와 즉시 해결할 수 있다는 것이다. 이것은 추적 비용을 줄이고 응답자로부터 더 나은 정보를 얻을 수 있다. 자료처리 단계들의 또 다른 다양한 결합도 가능성이 있다. 통합의 목적은 자료 품질을 개선하면서 운영의 효율을 증진시키는 것이다.

앞에서 분명하게 설명했듯이 자료는 자료처리에서 광범하게 수정될 수 있고 따라서 비표본추출오차의 중요한 원인이 될 수도 있다. 자료 품질의 개선하려고 에디팅을 하지만 그럼에도 불구하고 많은 오류들을 놓치거나 새로운 오류를 일으킬 수도 있다. 자동화는 수작업 처리에서 생기는 일부의 오류를 줄일 수 있지만 새로운 오류를 일으킬 수도 있다. 예를 들어, 광학적으로 인식하는 자료입력저장 운영에서 인식 오류는 모든 숫자와 문자에 걸쳐 균일하지 않는데 이것은 계통오차를 일으킬 수 있다(즉, 편향).

7.2 자료처리 오류의 특징

자료처리 오류와 관리에 대한 연구 문헌은 측정오차나(특히, 응답자 오류와 설문지 효과) 무응답에 대한 것에 비해 별로 없다. 유감스럽게도 코딩과 같은 처리 단계들은 오류가 발생하기 매우 쉽다. 특히 직업, 산업, 교육 조사에서 학문 분야와 다양한 종류의 의학 코딩 같은 복잡한 개념의 코딩이 그렇다. 특히 코딩 직원들이 잘 교육되어 있지 않으면 코딩 오류율 또는 코딩 불일치율은 어떤 변수에서는 20%수준에 이를 수도 있다.

조사 결과에 영향이 미치는 가능성에도 불구하고 많은 조사방법론자들은 자료처리 오류를 비교적 가볍게 여긴다. 그 이유는 아마도 인지 모형과 사회학 이론들은 무응답과 설문지 설계와는 달리 자료처리 운영에 직접 적용될 수 없기 때문이다. 이것이 자료처리 주제에 대한 문헌의 부족의 설명이 될 수 있다. 조사 연구에서 자료처리 오류의 중요함에 대한 충분한 증거가 있지만, 관련된 오류의 구조는 근본적으로 알려져 있지 않고 조사되지도 않았다.

이 장에서 우리는 자료처리 작업이 어떻게 조사 자료의 계통오차와 가변오차의 원인이 되는지 알아본다. 또한 자동화가 어떻게 계통오차를 초래하는 경향이 있고 수작업은 계통오차와 가변오차 모두를 초래할 수 있는지 검토한다. 예를 들어, 코더와 편집자가 범한 오류들은 면접원들이 범한 오류들과 최소한 이론적인 공통점이

있다. 면접원들처럼 에디터와 코더, 그리고 다른 자료처리자들은 그들에게 할당된 모든 단위들을 같은 방법으로 편향하는 계통오차를 일으킬 수 있다. 면접원들에 대한 이런 종류의 계통오차를 상관 면접원 오차라고 부르고 5장에서 면접원 분산의 개념을 설명했다. 마찬가지로 코더, 에디터와 다른 처리자 상관오차 때문에 발생하는 분산 요소(코더 분산, 에디터 분산 등)가 생길 수 있다. 예를 들어, 특정 코더가 기술자와 엔지니어를 코딩하는 지시를 잘못 해석할 수 있어서 많은 엔지니어들을 기술자로 잘못 코딩한다. 우리는 이런 계통오차가 추정값에 미치는 효과를 코더에게 할당된 일을 상호관입에 의해 측정할 수 있다. 만일 모든 코더가 할당받은 일에 대략 같은 수의 기술자와 엔지니어가 있으면 이 특정한 코더는 다른 코더보다 더 높은 비율의 기술자를 코딩하게 될 경향이 있다. 그러므로 5장에서 보았듯이 이 코더 단계의(coder-level) 계통오차는 모집단에 있는 기술자 수의 추정값 분산을 극적으로 늘릴 수 있다. 더욱이, 이 분산의 증가는 표준오차의 추정값에 완전히 반영되지 않고 직업에 대한 추정값의 표준오차는 작게 보고될 수 있다. 상관분산과 편향의 크기는 특별히 설계된 실험 방법으로 추정되어야 하지만 실제로는 거의 실행되지 않는다. 더 효율적인 접근 방법은 더 나은 감독, 재교육과 코더 분산을 줄이거나 억제하는 지속적인 품질 개선 기술 같은 방법을 사용하는 것이다.

자료처리 작업은 대체적으로 총 조사 예산의 매우 큰 부분을 차지한다. 일부 조사에서는 에디팅에만 총 조사예산의 40%까지를 배정한다(U.S. Federal Committee on Statistical Methodology, 1990). 그렇지만 조사의 총 재원 중에서 자료처리의 몫은 CASIC기술의 새 혁신에 의해 상당히 줄었다. 또한 자료수집과 자료처리와 관련된 총조사오차는 이런 변화들의 영향을 받았다. 이렇게 비용이 줄어들고 총오차가 잠재적으로 감소하는 이유는 Lyberg와 Kasprzyk(1997), Bethlehem과 van de Pol(1998), Speizer와 Buckley(1998), Biemer와 Casper(1994), 그리고 Lyberg 등(1998)을 포함한 몇 개의 논문에 상세히 나와 있다. 다음은 이런 발달과정을 설명한다.

- 자료처리의 자동화의 새로운 발전으로 수작업의 필요성이 많이 줄었다. 수작업 환경에서의 자료처리는 매우 노동집약적이고 많은 시간을 요한다. 자료처리 중앙 시설에서 작업량이 처리 한도에 이르면 스케줄에 맞추기 위하여 교육을 받지 않은 임시 직원들을 많이 고용할 필요가 있을 수 있다. 이전에는 주로 사람

에 의해 이루어졌던 코딩, 에디팅과 자료입력저장 작업에서 컴퓨터가 인간의 의존도를 줄게 했다. 또한 컴퓨터는 이러한 일을 표준화되고 일치된 방식으로 실행하고 이것은 작업자가 범하는 가변오차들을 제거 시켰다(즉, 작업자분산). 그렇지만 컴퓨터 프로그램들은 오류를 낼 수 있으므로 자동화가 과연 자료 품질을 개선하는지 또는 얼마나 개선하는지 알기는 쉽지 않다.

- 6장에서 보았듯이 CASIC 새로운 기술과 기존 기술의 발전은 자료수집 활동과 자료처리 운영이 통합될 수 있게 하였다. 미래에는 조사 체계 전체를 더욱 효과적으로 하는 것 뿐만 아니라 자료수집과 자료처리 체계의 더 많은 통합을 예견된다.
- 자료처리 체계의 수작업 부분은 많은 다른 생산 환경에 흔한 조립 라인 작업을 닮았다(예를 들어, 자동차 제조). 그래서 조사설계자들과 조사관리자들이 산업용 품질관리에서 관례적으로 사용되는 합격판정추출법과 같은 품질보증안을 적용하는 것은 당연하다. 합격판정추출법은 한 작업에 대하여 미리 정한 평균 결과물 품질수준 (AOQL)을 보장한다. 예를 들어, 만일 어떤 제품의 최대 결함률이 1%라고 생각하면, 합격판정추출안은 이 수준의 결과물 품질이 보증되도록 만들어진다.

합격판정추출법(acceptance sampling) 산업 환경에 적용되기 위해 Dodge와 Romig(1944)에 의해 개발되었다. 1960년대와 1970년대에 이 방법의 사용은 여러 수작업 조사로 확대되었다. 조사 연구 문헌에서 이런 방법들은 품질관리 방법의 관리적용(administrative applications of quality control methods)이라고 불린다. 합격판정추출법은 어떤 과정의 오류가 품질에 큰 변동 없이 관리되고 미리 정해진 품질수준에 도달하는 것을 보증하는데 사용될 수 있다. 그렇지만 이 방법에 피드백 루프가 없는 결과로 Deming(1986)과 다른 사람들에 의해 이 방법들은 비난을 받아왔다. 즉, 작업의 오류는 확인되지만 오류의 원인들이 확인되지 않고 제거되지 않았기 때문에 작업을 계속해서 개선할 방법이 없다.

그렇지만 합격판정추출법의 주된 목적은 제품의 결함을 찾는 것이라는 점에 특히 주의해야한다. 전형적으로, 이 방법은 먼저 조사도구(즉, 설문지들)에서 묶음(batch)들 또는 로트(lot)들을 만들어 낸다. 각 로트에서 표본을 뽑고 정해진 기준에 순응하는지를 검사한다. 만일 표본의 검사를 근거로 그 로트에 너무 많은 결함이 있다고

확인되면 그 로트는 거부된다. 그 시점에서 그 로트를 재작업 할 수 있고 거부당한 로트를 작업한 기사는 더 높은 수준의 품질관리심사를 받을 수 있다. 일반적으로, 합격판정추출법의 적용에 관련된 개선의 요소가 지속적으로 있는 경우가 아니면 재작업의 양은 다소 안정적이다. 개선의 요소가 지속적으로 있더라도 시간이 지나면 결국 감소될 것이다.

캐나다 통계청과 다른 기관들은 여러 해 동안 합격판정추출법을 많은 과정에서 사용했지만 품질 개선의 목적으로 작업자들에게 응답루프를 편입시켰다. 이 기관들은 합격판정추출법을 이러한 방식으로 계속해서 사용하였고 직원 이동률이 높은 작업에서는 다른 품질관리 방법들보다 (예를 들어, 지속적인 품질 개선) 더 좋다는 것이 발견되었다(Mudryk 등, 1996, 2001a 참조). 합격판정추출법은 시간에 걸쳐 결과물 제품의 오류율을 계속해서 낮추는 수단으로는 결점이 있다. 지속적인 품질개선(CQI) 방법들은 제품을 가려내기 보다는 오히려 과정의 근본적인 개선을 강조한다. 7.7절에서 다시 거론 하겠지만, 분명한 것은, 양쪽 방법 다 가능성이 있다.

- 전에 언급했듯이 수작업 자료처리에서는 작업자들은 면접원들이 범하는 오류와 통계적으로 유사한 오류들을 범한다. 코더, 에디터, 자료 입력자들은 그들에게 배정된 요소의 기록값에 동일한 방법으로 영향을 미치는 경향이 있다. 그들은 상관오차를 일으키고 이런 오차들은 총조사오차를 특수한 실험을 하지 않으면 알 수 없는 방향으로 부풀린다. 어떤 조사에서 이러한 분산의 기여가 상당하다고 믿을만한 이유들이 있다. 상관오차에 추가하여 수작업자도 모든 작업자들에 공통인 가변오차와 계통오차를 일으킬 수 있다. 자료처리 오류의 결과로서 조사 자료에 미치는 총 효과를 측정에 대한 연구는 거의 없다. 이것은 주로 조사 예산의 감소와 문제에 대한 관심 부족 때문이다. Jabine과 Tepping(1973)과 미국 센서스국(1974)이 그러한 연구이다. 목적이 반드시 품질평가가 아니고 품질개선 일 수도 있다고 믿음은 오류의 원인에 대한 품질 평가의 필요성을 제거시킨다.

다음 몇 절에는 자료처리 작업의 오류의 결과와 오류를 줄이고 통제하는 방법들을 제시한다.

자료처리(*data processing*)는 조사 연구에서 경시되는 오류의 원인이다. 코더, 입력자, 에디터와 다른 작업자들은 면접원이 기여하는 것과 비슷하게 평균제곱오차에 상관오차를 기여한다. 기술 사용의 증가는 상관오차를 줄이지만 새로운 오류를 낼 수도 있다.

7.3 자료 입력저장 오류

7.3.1 자료 입력

자료 입력저장(*data capture*)은 용지나 설문지에 기록된 정보를 컴퓨터가 읽을 수 있는 형태로 변환시키는 조사의 단계이다. 입력, 표지 문자인식(*mark character recognition*; MCR), 지능형 문자인식(*intelligent character recognition*; ICR), 그리고 음성인식 입력(*voice recognition entry*; VRE)으로도 정보를 입력저장 시킬 수 있다. MCR은 스캔된 이미지에서 표지의 존재나 부재를 찾아내지만 표지의 모양은 찾아내지 못하기 때문에 손이나 기계로 인쇄된 문자에는 적절하지 않다. 응답자가 질문에 대한 그들의 응답을 표시하기 위해 버블(*bubbles*)이라고 하는 작은 동그라미들을 메우는 용지들이 흔히 사용된다. ICR은 손으로 적은 문자들(필기체가 아닌)의 이미지를 기계가 읽을 수 있는 문자들로 바꾼다. 광학 문자인식(*optical character recognition*; OCR)은 기계가 만든 문자들(예를 들어, 바코드)을 컴퓨터가 읽을 수 있는 문자들로 변환시킨다. VRE는 음성 패턴을 기계가 읽을 수 있는 문자들로 변환시키는 방법이다. 최근에는 팩스 전달, 전자자료교환(*electronic data interchange*; EDI), 인터넷을 통한 전자메일 전달을 포함한 새로운 형태의 자료 입력저장 기술이 개발되었다.

자료 입력(*data keying*)은 엄밀하게 관리하지 않으면 실수를 하기 쉬운 따분하고 노동 집약적인 작업이다. 면접원보조 방법에서는 CATI, CAPI와 다른 CAI 기술을 사용해서 입력하는 것을 피할 수 있다. 우편조사에서는 처음에 스캔하고 그 스캔된 이미지들을 OCR이나 ICR 기술을 사용해서 컴퓨터가 읽을 수 있는 문자들로 변환시키는 특수하게 설계된 종이 설문지를 사용해서 입력을 피할 수 있다. 다음에는

이러한 선택에 대한 이익과 불이익을 이야기한다.

자료처리에 대한 연구논문에 입력 오류율의 세 가지 정의가 나와 있다.

1.
$$\frac{\text{한 글자 이상 잘못 입력된 필드 수}}{\text{총 필드 수}}$$
2.
$$\frac{\text{잘 못 입력된 글자 수}}{\text{입력된 총 글자 수}}$$
3.
$$\frac{\text{한 개 이상의 입력 오류가 있는 기록 수}}{\text{총 기록 수}}$$

정의 1.은 잘못된 필드의 비율을 측정한다. 예를 들어, 소득에 대한 질문의 응답을 한 필드로 생각할 수 있다. 만일 100명이 질문에 답한다면 이 측정의 분모는 100이고 분자는 잘못 입력된 이 필드의 수이다. 정의 2.는 모든 입력된 글자 중에 잘못 입력된 글자의 비율을 측정한다. 그러므로 만일 10,000글자를 입력했는데 200개를 잘못 입력했으면 입력 오류율이 $200/10,000=0.02$ 또는 2%이다. 마지막으로 정의 3.은 최소한 한 개의 입력 오류가 있는 기록의 비율을 측정한다. 다음 예에서 설명하듯이 다른 정의도 가능하다.

예 7.3.1 소득에 대한 질문의 답이 17,400이라고 하자. 만일 입력된 값이 17,400과 다르면 입력 오류가 발생했다고 말한다. 다음 세 가지 입력된 것들은 입력 오류의 예이다: 1740, 17,500, 또는 17,599. 첫 번째 경우에는, 입력자가 마지막 “0”을 입력하지 않았고, 두 번째 경우에는 한 숫자가 잘못 입력되었고, 세 번째 경우에는, 세 개의 숫자가 잘못 입력되었다. 세 개의 오류 중에 어느 것을 가장 심각하다고 할 수 있는가? 첫 번째가 다른 두 개 보다 더 심각한 것은 명백하다. 위의 정의들이 이런 오류들을 적당히 반영하나?

대부분의 입력 작업에서 독립적인 재입력 검증이 기본적인 입력의 품질 관리 방법이다. 이 방법들은 독립적으로 입력자의 모든 또는 표본의 작업을 재입력하고 새로 입력된 것과 본래 입력된 것을 비교하는 것을 의미한다. 이 두 입력의 불일치를

컴퓨터로 검사하여 판정할 수 있다. 요즈음의 자료 입력 체계는 두 번째 입력 도중에 바로 이 검사들을 한다. 이렇게 하여 세 번째가 아닌 두 번째 입력자가 관정관의 역할을 하게 한다. 소규모 작업(예를 들어, 몇 백 개의 설문지)에서의 입력 품질 관리는 입력된 것을 하드카피 소스 문서들(설문지들)과 대조하는 종속 검사를 사용해서 쉽게 수행될 수 있다. 그러나 이 방법은 상당한 수의 입력 오류를 놓칠 수 있다. 이 장의 뒤에서 종속확인 효과와 독립확인 효과 비교하여 설명한다.

조사 자료에서 몇 개의 매우 심각한 입력 오류에 대한 걱정과 입력이 비교적 저렴한 작업이라는 사실 때문에 많은 기관에서의 입력 품질관리 활동이 광범하게 이루어지게 되었다. 다른 조사 작업에 비교하여 특히 선택형 질문의 응답은 입력 오류율이 상당히 작다는 사실에도 불구하고, 흔히 100% 독립적인 재입력 확인이 사용된다. 예를 들어, 미국 다시점연구의 제4차 추적조사(U.S. Fourth Follow-up Survey of the National Longitudinal Study)에서는 설문지의 품질관리 표본을 근거로 한 오류 수준이 품질관리 확인과 수정 전에 약 1.6%였다(Henderson과 Allen, 1981). 마찬가지로, 1988년 미국 SIPP(Survey of Income and Program Participation)에서의 오류 수준은 약 0.1%였다(Jabine 등, 1990). 1990년 미국 센서스에서 긴 형식의 설문지 입력 품질의 결정을 위해 설계된 연구에서 입력자들이 연출(? production) 입력의 초기 단계에서 필드의 0.62%에서 키를 치는 실수를(또는 생략을) 범했다(미국 센서스국, 1993). 이 비율은 각 변수값의 입력 오류와 관련이 있고 (1번 정의) 입력을 전문적으로 하는 작업과 작업자가 적용되었다.

컴퓨터 기술은 본래 서툰 자료 입력자들로 간주되는 CAPI 면접원들이 범하는 오류의 수에 대해 새로운 관심을 갖게 하였다. 그렇지만 Dielman과 Couper(1995)의 CAPI 환경에서 입력 오류에 집중한 연구에서 0.095%의 입력 오류율이 보고되었다. 따라서 CAPI 조사에서는 입력 오류가 중대한 문제가 아니라는 증거가 제공되었다. 오류율을 기록의 백분률 입장에서 보는 한 스웨덴 조사에서는 3번 정의에 의하면 기록당 평균 오류율이 1.2%였다(Lyberg 등, 1977 참조).

예 7.3.1에서 설명했듯이 입력 오류에 대한 관심은 매우 큰 오류가 가능하다는 것이다. 이런 매우 큰 오류는 에디팅 단계에서 찾아낼 수 있지만 입력의 품질을 관리하는데 에디팅에 의존하는 것은 현명하지 않다. 이것은 입력을 위하여 어떤 종류의 품질 관리가 필요하다는 것을 의미한다.

사실 위에서 인용한 오류율에 대한 연구는 모두 일부 종류의 입력 오류를 찾고

수정할 수 있는 에디팅 절차를 가진 품질관리 체계 하에서 실시되었다. 만일 품질 관리가 없다면, 오류율은 아마도 더 높을 것이다. 왜냐하면 입력자들은 그들의 일에 대한 검사가 없을 때 보다 확인된다는 것을 알고 있을 때에 더 나은 품질의 결과를 낸다. 이것은 특히 입력자 성과의 평가가 생산력과 정확도 모두를 근거로 할 때에 그러하다.

모든 입력 과정의 오류율은 몇 가지 요인에 따라 결정된다. 이 요인들에는 입력자의 경험, 입력자간의 오류의 변동, 입력자 이동률, 입력 과정에서 사용되는 품질 관리의 양과 종류, 입력할 자료의 읽기 쉬운 정도, 입력자가 입력해야 할 필드의 확인에 대한 용이함, 용지들을 입력하기 위한 준비에 필요한 스캔에디팅 과정의 양과 품질 등이 있다.

입력 오류율은 (*keying error rates*) 보통 작지만 오류의 효과는 상당할 수 있다. 다양한 오류율의 정의 때문에 조사와 기관들에 따른 입력 오류율은 비교하기 어렵다.

7.3.2 스캐닝

미국 센서스국이 컴퓨터입력용 광학필름 판독장치(*Film Optical Sensing Device for Input to Computer*; FOSDIC)의 사용을 선도했다. 이것은 스캔한 사본이 아닌 적절히 설계된 설문지의 마이크로필름 사본에서 표식을 감지한다는 것을 제외하고는 OMR과 비슷한 과정이다. FOSDIC은 응답 버블이 표시되었는지를 확인하는데 광선이 사용된다. 이 과정에는 약 0.02%의 평균 오류율이 있다(Brooks와 Bailar, 1978).

NCES(U.S. National Center for Education Statistics)는 "High School and Beyond Study"에서 OMR을 사용하였다(Jones 등, 1986). OMR은 다지 선택형 질문에서는 매우 효과적이지만 위에서 말했듯이 손으로 쓴 것을 감지할 수 없다. 자료가 손으로 적은 응답일 때에는 ICR이 자료 입력저장 문제에 대한 해결법으로 제안되었다. 미국 센서스국이 ICR의 새로운 연구를 독려하려고 이 주제에 대한 여러 회

의를 주최했다. 이 회의들에서 발견된 일반적인 것은 어떤 적용 대상은 다른 것들보다 정교함이 덜 요구되기 때문에 ICR의 실현 가능성은 적용 대상에 따라 결정된다고 제안한다. 그렇지만 지난 몇 년 동안 적은 응답을 입력 저장하기 위한 ICR 체계는 “단어와 문구를 읽는 기계의 성능은 결과의 정확도를 줄이지 않으면서 센서스 실시애 필요한 비용과 시간을 줄일 정도로 충분히 좋아졌다”고 할 정도까지 극적으로 개선되었다 (Geist 등, 1994).

ICR에서는 대입과 거부, 두 종류의 오류가 발생할 수 있다. 대입 오류는 (*substitution error*) ICR 소프트웨어가 문자를 잘 못 인식할 때 발생한다. 거부 오류는 (*rejection error*) ICR 소프트웨어가 문자를 인식할 수 없어서 거부할 때 발생한다. 거부된 문자들은 손으로 수정해야하고 시스템에 다시 입력해야 한다. 따라서 거부 오류는 처리에 비용이 많이 들지만 바르게 처리하기만 하면 자료 입력 과정에 아무런 오류가 생기지 않는다.

대입 오류율은 보통 작다. 1970년 스웨덴 센서스를 위하여 실시한 연구에서 손으로 쓴 응답을 읽는데 ICR이 사용되었고 숫자의 단지 0.14%만이 대입되었다. 그렇지만 전에 언급했듯이 ICR의 정확도는 적용 대상에 따라 좌우되기 때문에 이 비율은 다른 조사로 일반화 되어서는 안된다. 이 경우에는서는 숫자를 쓰는 과정이 많이 표준화되어 있었다. ICR 숫자를 쓰는 특수한 교육을 받은 직원들이 숫자를 썼다. 물론 이 상황은 일반 모집단에게 소득과 다른 변수들에 대한 자료를 ICR 처리가 가능하도록 입력하라는 것과는 상당히 다르다.

Tozer와 Jaensch(1994)는 호주 통계국이 실시한 소매 센서스의 대입률이 숫자 기준으로 1%였다고 보고한다. 영국 노동부에서도 비슷한 수준이 보고되었다 (Thomas, 1994). 1991년 캐나다 농업센서스에서 알파벳과 알파벳과 숫자로 이루어진 문자의 교체율은 각각 2%와 5%였다(Vezina, 1994).

많은 적용에서 설사 대입율이 작다고 할지라도 그 결과 생기는 오류는 상당히 문제가 있고 심각할 수 있다. 캐나다 통계청의 경험에서(Mudryk 등, 2002) 대입 오류는 사실상 계통적일 수 있고 그 때문에 심각한 문제들을 야기한다. 예를 들어, 작업 중에 광학 창에 얼룩이 있는 스캐너를 생각해 보자. 이 스캐너는 이 문제가 존재하는 한 아마도 많은 계통 대입 오류를 낼 것이다. 실제로 비슷한 계통적 대입을 야기할 수 있는 스캐너 전구의 낮은 휘도, 부품에 있는 먼지, 그리고 (?)문서 운반 문제들 같은 많은 다른 문제들도 있다.

대입 오류를 확실하게 최소화하기 위하여는 품질 관리 과정에서 잦은 도구의 재계수조정과 재배열이 필요하다. 대입 오류는 또한 읽거나 해석되는 자료의 종류에 관계가 있을 수 있다. 예를 들어, 바코드와 점검 표시 상자들(*tick boxes*)은 숫자 자료보다 훨씬 더 낮은 대입률을 가지고 있다. 숫자 자료는 알파벳이나 알파벳과 숫자로 이루어진 자료 보다는 대입률이 더 낮다. 또한, 일반적으로 잘 관리된 환경에서 준비된 서류들은 관리되지 않은 환경에서 준비된 것들보다 오류율이 더 낮기 때문에 서류들의 상태도 요인이 된다.

최근에 스웨덴 통계청과 몇 개의 다른 기관에서는 정보를 컴퓨터가 읽을 수 있는 형태로 변환시키기 위하여 조사 용지 전체를 스캔하고 ICR을 사용하기 시작했다. 따라서 일반적인 자료 입력은 필요하지 않다. 더구나 에디팅과정의 일부 기능들은 자료 입력저장 과정에서도 수행할 수 있다. 서류들을 에디팅과 수정 중에 저장하거나 검색 할 수 있다. Blom과 Lyberg(1998)는 스웨덴 통계청에서 실시한 초기 스캐닝 연구에서의 거부율은 7%에서 35%안에서 변동된다고 보고한다.

마지막으로, ICR을 조사 자료에 적용하는데 새로운 발전이 있었다. Mudryk 등(2001a)이 설명하듯이 ICR은 자동화 된 표시와 이미지 인식이 결합되고, 그 결합에 “빈틈없는 자료 입력저장 기법”을 사용해서 “이미지로 부터 입력하는” 작업자들에 의한 수작업의 입력저장이 부가되었다. 이것은 ICR 기술은 문자를 거부하지는 않지만, 그 필드에 대응하는 신뢰수준과 함께 항상 그것의 해석을 제시한다. 만족스러운 ICR 신뢰수준이 있는 필드는 ICR 소프트웨어가 인식할 수 있다고 간주되고 ICR의 해석을 받아들인다. 만족스러운 신뢰수준이 없는 필드는 작업자들이 광학적으로 획득한 이미지를 읽고 손으로 이미지의 원문을 입력하는 “이미지에서 입력하는” 작업으로 보내진다. 이러한 자료 입력은 간단하고 서류 처리가 필요하지 않아서 비용을 상당히 줄인다. 더구나 이런 이미지들은 이후에 서류를 쓰지 않고 에디팅에 사용될 수 있어서 비용을 한층 더 절감시킨다.

지능형 문자 인식에서는 대입 오류(*substitution error*)와 거부 오류(*rejection error*)가 있다.

7.4 자료 입력저장 후의 에디팅

7.4.1 용어의 정의

에디팅에는 여러 가지 정의 있다. 예를 들어, U.S. Federal Committee of Statistical Methodology(1990)에 따르면 에디팅(*editing*)은 “자료의 대체와 요약 절차를 수행하기 전에 되도록 많은 잘못된 자료(반드시 의심되었던 모든 자료는 아닌)를 수정할(수작업으로 또는 컴퓨터를 이용하는 방법으로) 목적으로 잘못되거나 의심스러운 조사 자료(조사응답자료 또는 신분증명 종류의 자료)를 찾기 위해 설계되고 사용되는 절차들”로 정의한다. IWSSE(International Work Session on Statistical Editing)은 “에디팅은 자료에 있는 오류들을 찾아내고 수정하는 활동이다”라고 간단히 명시하는 ECE(Economic Commission for Europe)의 정의(1995)를 뒷받침했다. 세 번째 정의(Granquist와 Kovar, 1997)는 후자를 약간 확대 해석한다. 즉 “에디팅은 통계 산출에 사용되는 각각의 자료에 존재하는 오류와 이상점을 확인하고 만일 필요하면 수정하는 것이다.”라고 하였다.

어떤 정의도 모든 오류들이 수정되거나 확인이라도 되어야 한다고 명시하지는 않는다는 것에 주목해라. 따라서 에디팅의 목적을 상세하게 말하면 다음과 같다.

1. 자료 품질에 대한 정보를 제공(*To provide information about data quality*). 조사 설문지들이 현장에서 들어올 때 그것들의 시각적 검사 또는 스캔에디팅 (*scan edit*)를 통하여 면접원들이 (PAPI 면접 방법에서) 그들의 역할을 어떻게 수행하는지 그리고 응답자들이 정확한 정보의 제공자로서 그들의 역할을 가지고 어디에 애를 썼는지가 많이 드러날 수 있다. 면접원의 메모와 논평으로도 자료수집 문제에 대한 중요한 정보를 모을 수 있거나 일부 면접원들이 그들의 용지를 바르게 완성하지 않는다는 증거가 찾을 수 있다. 예를 들어, 그들은 일부 관측자료를 입력하는 것을 잊을 수 있고 건너뛰어야 되는 모든 질문들을 건너뛰지 않을 수고 있다. 이런 문제들은 면접원이 이런 분야에 대해 다시 교육을 받아야된다는 것을 암시할 수 있다.

이러한 스캔 에디팅의 절차에 이어서 자료를 입력하거나 컴퓨터가 읽을 수 있는 형태로 변환시켜서 더 검사할 수 있다. 빠진 자료 값들, 범위 밖 값들과 믿기지 않

는 값들이 어느 정도인지를 결정하기 위하여 분석을 실시할 수 있다. 이 분석은 자료품질에 대한 추가 정보와 설문지나 면접 과정의 가능한 문제점들을 우리에게 줄 수 있다. 이 단계에서 자료를 개선하기 위하여 자료수정을 할 수 있다. 때로는 설문지에 제공된 다른 정보를 사용해서 이것을 정확하게 할 수도 있다. 그렇지만 많은 경우에는 이러한 수정들이 응답자와의 재접촉 또는 면접원이 한 가지 이상의 결정적인 자료를 매우거나 명백히 하는 것을 필요로 할 수 있다.

CAI 적용에서는 에디팅의 일부는 자료를 수집할 때 할 수도 있고 면접원들이 실행할 수도 있다. CAI 소프트웨어를 적용하여 면접 도중에 일치하지 않거나 또는 믿기지 않는 값들에 대하여 면접원이 통보를 받을 수 있고 응답자에게 분명히 하도록 부탁하게 할 수도 있다. PAPI 적용에서는 질문지를 기관으로 보내기 전에 면접원이 에디팅의 일부를 수행하게 하는 것이 가능하다. 특수 직원들이 용지들을 손으로 에디팅하는 것이 아직도 상당히 흔하지만 대형 조사에서는 에디팅은 최소한 부분적으로라도 특수한 소프트웨어를 사용해서 자동으로 한다.

2. 미래 조사의 개선에 대한 정보를 제공(*To provide information about future survey improvements*). 설문지들을 검사하고 컴퓨터 분석을 통하여 자료를 재검토하는데서 얻은 교훈으로 미래 조사에서 이런 종류의 오류를 제거할 자료수집 과정의 개선을 제안할 수 있다. 이런 이유로 조사 설계자들이 자료의 근본적인 오류 원인들을 이해하도록 노력하는 것이 중요하다. 이것은 자료수집이나 자료처리의 문제들을 수정하기 위해서이다.

3. 향상된 과정을 가능하게 하기 위한 자료의 “정화” (*To simply “clean up” the data so that further processing is possible*). 이것은 에디팅의 가장 중요한 동기 부여라고 전에 이야기했다. 오류를 제거하면 자료가 이상적인 수준의 품질에 더 가까이 가져간다고 생각하기 때문에 우리는 자료를 청소할 때 가능한 한 많은 오류들을 확인하고 수정하는 것이 당연하게 보인다. 그렇지만, 이 행동에 쓰이는 비용이 그 행동으로 얻을 수 있는 개선의 가치보다 일반적으로 훨씬 더 클 수도 있다. 즉, 비용과 오류의 최적화는 에디팅에 쓰인 비용의 많은 부분은 조사과정의 다른 요소들을 개선하는데 더 효과적으로 쓸 수 있다는 것을 말한다.

에디팅 방법의 개요는 국제 연합(United Nations)(1994b)에 나와 있다.

7.4.2 에디팅에서 사용되는 기본 개념

자료 에디팅은 조사 자료에 적용되는 규칙으로 구성되어 있다. 에디팅 규칙의 예로는 값은 파일의 특정한 위치에 항상 나타나야 변수의 어떤 값들은 허락되지 않고; 값의 어떤 결합들은 허락되지 않고; 합이 값은 그것의 요소들의 합과 똑같아야 하고; 그리고 특정한 값은 만일 그것이 미리 정의된 구간 안에 포함되어 있을 때에만 허락된다. 이런 에디팅들을 결정적 에디팅(*deterministic edits*)이라고 부르며, 만일 이것들을 어기면 틀림없이 오류가 나타난다. 에디팅 과정에서는 에디팅 규칙이 적용되고 수동적인 조정을 위하여 오류 메시지가 인쇄되거나 컴퓨터 화면에 표시된다. 물론, 이런 명세 사항들에 따라 오류를 찾아내고 수정할 수 있는 소프트웨어를 개발할 수 있다. 소형 조사에서는 이런 소프트웨어를 사용하는 것이 효율적이지 않을 수 있지만 대형 조사에서는 필수적이다. 에디팅 규칙을 근거로 자료를 수정하는 과정은 기록에서 입수할 수 있는 자료이거나 응답자와 면접원과의 재접촉을 통해서 수집될 수 있는 자료에 적용될 수 있다.

치명적 에디팅(*fatal or critical edits*)은 자료 기록을 사용 가능하기 위하여 수정되어야 하는 잘못된 변수값을 반드시 찾아내야 한다. 치명적으로 여길 수 있는 오류들의 예는:

- 신분 증명 번호의 오류 (예; 다당하지 않은 사업 ID 번호)
- 중요한 변수의 항목 무응답
- 중요한 항목에 대한 타당치 않은 값 (예; 범위 밖의 나이 값)
- 일치되지 않는 값 (예; 대리모의 나이가 아이의 나이보다 작다)
- 변수들 사이에 정의된 관계가 만족되지 않는 경우 (예; 세후순소득이 세전총소득보다 작다)
- 극단적이거나 터무니없는 값

의문 에디팅(*query edits*)은 의심스러운 값들을 확인하는 것이다. 의심스러운 값들은 잘못된 값일 수 있지만 추가 조사 없이는 그 결정을 할 수 없다. 예를 들어, 농장 기사가 농장을 둘러싼 땅 값의 10배로 그의 땅을 그 평가 할 수 있다. 이것은 농부의 보고에 “0”을 덧붙여진 입력 오류일 수 있다. 그렇지만 그것이 농장 기사가

제공한 실제 추정값일 수도 있다. 원래의 조사 용지 점검을 통해 어느 것이 사실인지가 드러날 것이다.

의문 에디팅은 그것들이 조사 추정값에 현저하게 영향을 미칠 가능성이 있을 때에만 실행되어야 한다. 오류가 발생했는지 명백하지 않기 때문에 각 의문 에디팅은 더 철저하게 조사해야한다. 이런 조사는 시간이 많이 걸리고 비용이 많이 들 수 있다. 그러므로 의문 에디팅의 규칙은 추정값에 영향을 미칠 상당한 확률이 있는 오류만을 확인하고 실행하도록 설계되어야하는 것이 중요하다. 이런 에디팅 규칙들은 확률적 에디팅(*stochastic edits*)이라고도 부른다(결정적 에디팅과는 대조적으로).

기록 또는 설문지 단계에서 수행하는 에디팅은 미시적 에디팅(*microediting*)이라고 한다. 최근 들어 총괄(*aggregate*) 자료(예를 들어, 평균이나 총계)에 수행되는 에디팅 또는 기록 전체에 적용하는 검사인 거시적 에디팅(*macroediting*)에 대한 관심이 높아졌다. 만일 거시적 에디팅 규칙에 근거하여 총괄된 값이 의심스럽다고 생각되면 총괄의 차이의 원인을 한 개나 몇 개의 잘못된 기록들에 돌릴 수 있는지를 결정하기 위하여 그 총괄을 구성하는 각각의 기록들을 검사한다. 그렇지만 만일 총괄 값이 의심스럽지 않으면, 그 총괄을 이루는 각각의 기록들은 에디팅을 통과한다. 물론, 설사 한 총괄 값이 에디팅을 통과하더라도 그 총괄 값을 이루는 각각의 기록값에 여전히 오류가 있을 수 있다. 그렇지만 이 오류들이 추정값에 인식될 수 있는 영향이 없기 때문에 (총괄 값이 입증했듯이), 그것들은 검사하고 수정하지 않아도 된다.

따라서 에디팅은 확인 단계(결정적 또는 확률적)와 수정 또는 조정 단계의 두 단계로 나눌 수 있다. 후자는 면접원과 응답자의 재접촉이 포함될 수 있고, 잘못된 값을 단순히 삭제하거나, 추론이나 대체한 값으로 그것을 대신하는 것일 수도 있다. 그러므로 대체는 항목 무응답 외에도 에디팅 실패에 사용되는 중요한 구제 방안일 수 있다. 에디팅에 적용한 대체는 Pierzchala(1990, 1995), Fellegi와 Holt(1976), 그리고 Legault와 Roumelis(1992)에 설명되어 있다.

자료는 아래와 같이 조사 과정의 여러 단계에서 에디팅할 수 있다.

- **면접중의 에디팅** (*editing during the interview*)

면접원 혼자 또는 응답자와 협조하여 면접원 지시서에 명확히 나와있는 에디팅 점검을 한다. 일치성 점검, 범위 점검, 변수 관계, 그리고 빠진 값들을 CAI 소프트웨어

트웨어의 일부로 짜 넣어서 면접을 진행하면서 이런 에디팅들을 실행하도록 하게 할 수 있다. 그렇지만 면접이 지나치게 길어지게 하거나 방해받지 않기 위하여 실시간의 에디팅 queries는 가장 중요한 에디팅 확인으로 제한해야한다. 또한, 우편조사 응답자들이 약간의 자기 에디팅을 실시한다는 증거(용지에서 명백한 교정 등)가 있다. 설문지 설계자들은 적당한 지시를 통해서 이런 자발적인 에디팅을 격려해야한다.

- **자료 입력저장 전의 에디팅(editing prior to data capture)**

이런 종류의 에디팅은 보통 수작업으로 하고 설문지가 조사 기관에 전달된 후에 수행된다. 이 단계에서의 에디팅은 설문지들을 계속해서 처리할 것인지에 관해서 받아들임, 거부함, 조치가 필요함 등으로 분류하는 상당히 일반적인 작업이다. 예를 들어, 설문지가 대부분을 채워지지 않았고 명백한 오류들이 있으면 그것은 거부되고 더 이상 처리하지 않을 수 있다.

- **자료 입력저장 중의 에디팅(editing during data capture)**

자료 입력저장에는 입력, ICR, OMR, 또는 다른 스캐닝 기술을 포함될 수 있다. 이 단계 중에 에디팅 소프트웨어를 대화형식의 과정으로 사용하는 것이 가능하다. 에디팅은 변수별로 또는 기록별로도 할 수 있다. 전형적으로 에디팅이 실패하면 소프트웨어가 그 과정을 멈추고 자료 입력저장은 작업자가 특정한 조치를 취한 다음에만 계속될 수 있다. 이런 조치들에는 확인된 값에 대한 작성자 쪽의 특별한 수용, 또는 작성자 변경한 것, 또는 미래 조치를 위한 조언의 역할을 하는 것들이 있다.

- **자료 입력저장 후의 에디팅(editing after data capture)**

대부분의 에디팅은 이 단계에서 일어나는데 결정적인 그리고 의문 (미시적-) 에디팅, 거시적 에디팅과 추론적인 그리고 모형을 근거로 한 대체가 포함된다. 이 에디팅은 제한된 수작업이 가미된 자동화가 되었다.

- **결과물의 에디팅(output editing)**

이것은 마지막 에디팅인데 사용자에게 제시되는 값들에 초점이 맞추어져 있다. 표의 칸들의 총괄 값이나 다른 통계적 추정값들이 논리적인지를 점검한다. 한 가지 널리 알려진 방법은 만일 조사가 지속적이면 결과들을 이전의 조사와 비교하고, 만일 조사가 일회성이면 외부 자료와 비교하는 것이다. 의심스러운 총괄 값은 거시적 에디팅 절차로 다룬다.

때때로 조사 기관들은 큰 오류가 있는 표나 출력한 데이터를 발표한다. 그 오류는 너무 늦게 발견되는데 더욱 나쁜 경우는 사용자가 이것을 발견한다. 발표된 결과에 있는 큰 오류들은 오류를 확인하는데 필요한 정보 부족 때문이거나 또는 그 오류는 이전의 추정값과의 비교가 모든 변수와 조사에서 적절한 방법이 아니기 때문에 관련된 사람들의 주의를 피할 수도 있다. 사용되어지는 있는 신뢰성 있는 그 래픽 시스템이 있기는 하지만(Houston과 Bruce, 1993 참조) 결과물 에디팅 방법들은 충분히 개발되지 않았다. 자료수집 기관들이 큰 오류들이 발표된 불행한 상황에 대처하기 위하여 정책과 조치 계획들을 적절하게 가지고 있는 것이 중요하다. 이런 정책과 계획들은 고객 염려에 대처하고, 오류의 원인을 입증하고 분석하고, 기관 전체가 그 실수를 교훈 삼도록 한다.

7.4.3 실제 에디팅

에디팅은 조사과정의 필수 단계이지만 에디팅에 의해 생기는 문제들은 꽤 심각할 수 있다. 새로운 기술에도 불구하고 에디팅은 비용이 많이 들고 많은 시간을 요한다. 전에 언급했듯이, 어떤 조사에서는 에디팅만이 전체 조사 예산의 40%를 소비할 수 있다(U.S. Federal Committee on Statistical Methodology, 1990). 광범한 에디팅은 조사의 공개를 지연시킬 수 있고 따라서 사용자들에게 자료 이용의 적절성을 줄인다. 조사 집단의 일치된 의견은 에디팅의 양은 비용과 오류의 최적화 전략을 근거로 해야 한다는 것이다. 특히 많은 에디팅 체계는 자료 품질을 감지할 수 있을 정도로 개선시키지는 못하기 때문에 에디팅에 드는 자원을 조사 오류를 줄이기 위해 선택할 수 있는 많은 대안들과 비교 평가해야 한다. 주목할 만한 품질 개선을 이루지 못하는 광범한 에디팅은 Granquist와 Kovar(1997)에 의해 과다 에디팅(*overediting*)이라고 부른다. 그들은 과다 에디팅에 대해 다음과 같은 설명을 하였다. 무응답이나 측정오차와 같은 자료수집 과정 중에 일어나는 대부분의 오류 보다 자료수집 이후의 문제들이 훨씬 다루기 간단하다고 지적한다. 자료를 과다 에디팅하는 조사 생산자들은 이 책에서 강조하는 총조사오차의 최적화 원칙을 지키지 않는 피실을 범하는 것이다. 그들은 과정의 모든 단계에서 품질을 염두에 두어야한다는 Deming의 아이디어에 충실하기보다(1장) 사후에 오류를 고치는 것을 더 선호한다

다.

에디팅에 대한 현대 연구에서의 중요한 교훈은 자료에 있는 모든 오류를 고치기 보다는 차라리 선택 에디팅(*selective editing*)을 수행하도록 해야 한다는 것이다. 다양한 연구들에서 선택 에디팅은 최종 추정값의 정확도의 감소 없이 시간과 돈이 상당히 절약된다고 하였다. 선택 에디팅에서는 모든 의심스러운 항목의 100% 미만을 재검토한다. 대신에 의문 에디팅에서는 표본 단위의 중요성, 연구 대상인 변수의 중요성, 오류의 심각성, 그리고 그 의심스러운 항목을 자세히 조사하는데 드는 비용을 근거로 항목이 선택된다. 표본 단위의 중요성은 보통 단위의 크기나 추정 과정에서 단위의 가중치를 근거로 한다. 농업 조사에서는 생산과 농업 통계의 추정값에 훨씬 적은 영향을 미치는 작은 농장보다 큰 농장에 대한 에디팅 과정에 더 높은 우선권을 준다. 가중치가 큰 단위들은 평균제곱오차에 영향을 미칠 더 큰 가능성이 있다. 가중치가 큰 의심스러운 항목들은 작은 가중치가 있는 항목들 보다 우선권을 준다.

의심스러운 항목을 자세히 조사하는 비용에는 노동과 컴퓨터 비용보다 더 많은 것이 포함될 수 있다. 항목이 잘못되었다거나 의심스럽다는 말을 듣는 응답자의 잠재적인 스트레스는 물론이고 각 접촉은 응답자에게 시간과 노력이 들기 때문에 항목을 수정하기 위하여 응답자에게 다시 연락하는 것은 응답자 부담을 늘리는 것이다. 이것은 패널 조사에서 차후의 파동이나 주기적인 조사의 미래 수행의 응답률을 감소시킬 수 있다. 예를 들어, 대규모 사업체나 농업 조사에서 매우 큰 단위들에 대해서는 표본 선택확률 1을 할당해서 해서 무조건 표본에 추출되는 선택 확실단위(*certainty units*)가 되게 한다(9장 참조). 그러므로 미래에 참여를 꺼리게 하거나 무응답을 막기 위하여 자료 항목의 수정을 위한 이런 단위들과의 접촉은 되도록 피해야 한다. 이러한 고려사항들은 에디팅 과정 설계자들이 참작해야 할 필요가 있다.

추천되는 또 하나의 방법은 과정 자료(*process data*)라고 불리는 에디팅 과정 자체에서의 자료 수집이다. 이것은 에디팅 운영의 품질과 효율을 모니터링하는 것이 중요하다. 만일 과정 자료가 과정이 설계된 대로 운영된다는 것을 보여준다면 과정에서 나오는 자료 또한 만족스러울 것이라는 가정이다. 운영이 미리 지정한 기준에 일치하는지를 모니터링하는 것은 그 운영의 중요한 과정 통계량의 측정과 모니터링을 필요로 한다. <표 7.2>에 이 목적으로 사용될 수 있는 Granquist 등(1997)이 제안한 과정 통계량이 제시되었다.

<표 7.2> 에디팅 성과를 모니터링하기 위한 중요한 과정 비율 통계량

과정 변수	비율의 정의		목적
	분자	분모	
1. 에디팅 실패율	에디팅에 실패한 대상의 수	에디팅된 대상의 수	검증된 양의 추정값
2. 재접촉률	재접촉의 수	에디팅된 대상의 수	재접촉의 수의 추정값
3. 결과물 에디팅률	결과물 에디팅에서 검증된 대상의 수	검증된 대상의 수	에디팅과 결과물 에디팅 사이에 균형이 잡혔는지를 확인
4. 수정율	수정된 대상의 수	에디팅된 대상의 수	에디팅과정 효과의 추정값 (변수 4가 변수 1보다 추정값이 작으면 에디팅과정이 비효과적이라는 것을 나타낸다)
5. 재접촉 생산성	1개이상의 대체된 변수가 있는 재접촉 대상의 수	재접촉 대상의 수	재접촉 효율의 추정값
6. 변수당 에디팅 실패율	변수 X에서 에디팅에 실패한 대상의 수	변수 X에서 에디팅된 대상의 수	변수당 검증된 양의 추정값
7. 에디팅 점검당 에디팅 실패율	에디팅 점검 K에 의해 에디팅에 실패한 대상의 수	에디팅 점검 K로 에디팅된 대상의 수	각 에디팅 점검으로 검증된 양의 추정값
8. 변수당 에디팅 변경율	변수 X에서 변경된 대상의 수	변수 X의 값이 있는 대상의 수	변수당 에디팅과정의 효과의 추정값
9. 변수당 에디팅 성공율	변수 X에서 변경된 대상의 수	변수 X에서 에디팅에 실패한 대상의 수	변수당 오류를 에디팅이 얼마나 성공적으로 확인하는지의 추정값

출처: Granquist 등(1997)

<표 7.2>에 있는 (6)-(9)의 특수한 변수들은 크기 순으로 배열되어야 한다. 과정의 어디가 부족한지 어디를 변경해야 하는지 확인하기 위하여, (1)-(9)의 모든 추정값들이 지속적인 방식으로 사용되어야 한다. 이런 과정 변화는 효과적이지 못한 에디팅을 생략하거나, 추정값의 품질에 가장 많이 기여하는 꼭 필요한 몇 개에 집중하거나, 설문지를 변화시킨다.

최근의 에디팅에 대한 논문은 전통적인 철저한 미시적 에디팅에 대한 많은 대안들이 있다고 제안한다. 우리는 다양한 형태의 선택 에디팅, 거시적 에디팅, 중요한 에디팅과정 변수들의 측정과 분석을 통한 개선에 대하여 이야기했다. Bethlehem과 van de Pol(1998)이 지적하였듯이 발표를 위해 자료를 준비하는 책임자들에게 이 대안들이 조사에서 통상적으로 실시하는 에디팅의 양을 감소 시키기 때문에 이 대안들이 품질에 해롭지 않다는 것을 납득시킬 필요가 있다. 역사적으로, 철저한 에디팅은 조사 품질을 위해서 절대적으로 없어서는 안되는 것이라고 보아왔다. 그렇지만 현재의 시각은 중요한 것은 조사 과정의 어떤 요소의 완벽보다는 오히려 총 조사 자료 품질이라는 것이다.

Granquist와 Kovar(1997)는 과다 에디팅이 상당히 흔하다는 사실을 설명하는 몇 개의 예를 제시한다. 그들은 에디팅은 매우 비싸고 많은 조사 예산의 많은 부분을 사용하고 추정값에 대한 효과가 흔히 의심스럽다고 한다. 또한 많은 비용을 들여서 비교적 중요하지 않은 수정을 많이 한다고 하였다. 비교적 중요하지 않은 수정은 추정값에 매우 작은 효과를 준다. Swedish Annual Survey of Manufacturing에 대한 Hedlin의(1993) 연구에서 에디팅 과정 중에 실시한 50%의 수정이 최종 추정값에 1%미만의 변화를 가져왔다고 보고된다. 이 결과는 추정값에 큰 효과를 초래하는 것은 비교적 작은 수의 중요한 변화들이라는 호주, 캐나다, 미국에서의 일반적인 경험과 일치한다.

또한, 에디팅이 오류를 증가시킬 위험이 있다. 예를 들어, 비논리적이거나 잘못된 에디팅 규칙의 결과로 그럴 수 있다. Granquist와 Kovar는 다음의 놀랄만한 예를 제시했다. 일반적으로 분석자들은 과다 에디팅된 조사 자료에서 지나칠 정도의 신뢰를 가지고 에디터의 모형을 재발견하게 된다. 예를 들어, 에디팅된 자료에서 아내들의 나이가 남편들보다 평균 2살 어리게 나타나는 것을 인구 통계학자가 '발견'했을 때 바로 그러한 대체를 수행하는 에디팅 규칙이 비로소 캐나다 센서스 체계에서 제거되었다.

에디팅은 조사 설계에서 없어서는 안 될 부분이지만 그것이 조사 추정값에 상당한 효과를 초래하는 변화가 온다는 것을 생각해야한다. 에디팅 규칙을 확인하는 가장 효율적인 유일한 방법은 에디팅된 자료와 에디팅되지 않은 자료를 비교하는 것이다. 따라서, 에디팅에 드는 자원은 다른 필요한 것들과 비교 평가되어야한다 (10장 참조).

에디팅(editing)에 총 조사 예산의 큰 부분이 소모된다. 에디팅의 이득을 최대한 활용하기 위하여 에디팅 단계의 결과들이 조사 과정을 개선하는데 사용되어야 한다. 예를 들어, 설문지, 면접원 교육, 에디팅 실패의 원인이 될 수 있는 다른 조사 과정들을 개선하기 위하여.(?) 또한 잠재적인 오류를 더 효과적으로 찾아내는 방법들을 사용해서 에디팅 과정 자체를 개선할 필요가 있다. 거시적 에디팅과 선택 에디팅이 이런 방법들의 예이다.

7.5 코딩

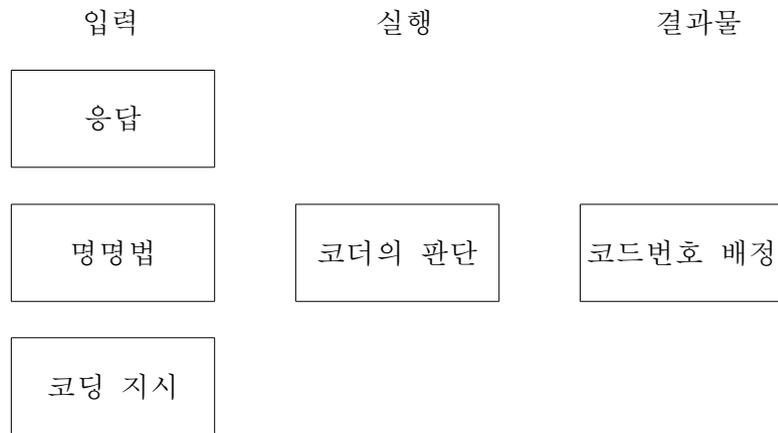
7.5.1 코딩 오류

코딩이 모든 조사에 필요한 것은 아니지만 잠재적으로 해로운 오류의 원인이면서 동시에 많은 조사에서 매우 중요한 작업이다. 코딩은 흔히 서술형 질문의 응답 형태로 되어있는 원래의 조사 자료에 추정, 도표화, 분석을 하기에 적절한 코드번호나 구분을 배정하는 분류 과정이다. 코딩은 작업자나 코더에 의해 수작업으로 실시되거나 특수하게 설계된 코딩 소프트웨어로 자동으로 실시할 수 있다. 때로는 쉬운 상황은 컴퓨터가 코딩하고 남은 것들은 코더가 하는 것이 선호된다.

<그림 7.1>에서 보여주듯이 코딩 작업에는 세 가지의 기본적인 입력 요소들이 있다. 이 입력을 근거로 코더는 판단을 하고 그 요소에 코드번호를 배정한다.

1. **응답(response)**. 표본이나 모집단에 있는 각 요소는 흔히 말로 나타낸 설명(변수와 관련된 질문들의 답)에 의해 특정 변수에 대해서 코딩된다.

2. **명명법(nomenclature)**. 이 변수에 대하여 미리 정해진 특정한 코드번호들이 있다. 이것은 변수의 특정한 범주에 대응하는 숫자들로 이루어져 있다(예를 들어, 00에서 99까지). 범주의 설명이 각 코드번호와 관련되어 있다. 코드번호들은 때로는 코딩 기준(*standard*) 또는 명명법(*nomenclature*)이라고 부른다.
3. **코딩 지시(coding instructions)**. 1.과 2.를 관련짓는 코딩 지시들이나 규칙들이 있다. 이 규칙들은 코더가 서술형 응답의 중요한 요소들을 명명법에 있는 코드번호들과 관련짓는 지시 형태로 되어있다.



<그림 7.1 일반적인 코딩 과정>

코딩 과정에서 일어날 수 있는 문제들은 많은데 조사 기관들은 이것들은 항상 자각하지는 못한다. 이들 중에는 다음과 같은 것들이 있다.

- 대부분의 코딩은 오류의 여지가 있다. 코딩 규칙들을 항상 바르게 적용하지 않고 코딩 규칙 자체가 불완전하기 때문에 오류가 발생한다. 예를 들어, 가능한 모든 서술형 응답이 포함되게 명명법을 구성할 수 없고 매우 노련한 코더들도 배정할 알맞은 코드번호에 대하여 의견이 일치하지 않는 경우가 많다. 어떤 작업에서는 코더 불일치율이 20%까지 될 수 있다.
- 코딩은 매우 주관적인 행동이기 때문에 우수한 코딩 작업을 개발하는 것은 어렵다. 때로는 서술형 응답들은 코드번호를 명확하게 배정하기에 적당하지 않다.

그래서 코더들은 응답을 코딩하기 위하여 그들의 판단을 사용해서 “숨은 뜻을 헤아려야” 한다. 코딩 기술은 개발하는데 시간이 많이 필요할 수 있다. 하지만, 불행히도 이런 작업에서 직원 이동률이 높을 수 있다.

- 큰 조사에서는 코딩 작업이 상당히 많아서 관리하기가 힘들 수 있다. 이런 작업에서 오류를 관리하는 것은 상당히 어렵다.

코딩 작업에는 몇 가지 형태가 있다. 코딩은 여러 위치에서 수동으로 (비집중식 수동 코딩(*manual decentralized coding*)) 또는 한 위치에서 수동으로(집중식 수동 코딩(*manual centralized coding*)) 실시할 수 있다. 또한 코딩이 잔여분은 수동으로 하는 자동화일 수 있고 또는 코딩이 컴퓨터가 보조하는 수동일 수 있다. 비집중식 코딩의 전형적인 형태에는 자료수집과 관련하여 면접원들이 코딩을 수행하는 것, 또는 응답자가 수행하는 코딩이 있다. 집중식 코딩은 조사기관 안에서 어느 정도 전문적인 직원들이 수행하는 코딩이다. 자동화 코딩(*automated coding*)은 컴퓨터 프로그램이 응답 기재 사항을 컴퓨터에 저장된 명명법 설명(사전)들과 대응시키도록 하고, 특정한 대응 기준에 대응이 만족되면 코드번호를 배정한다. 잔여분은 (즉, 대응기준에 만족되지 못한 경우들) 수동 코딩 작업으로 전환시킨다. 컴퓨터 보조 코딩에서는 코더들이 코딩 소프트웨어와 서로 작용한다. 소프트웨어에 의해 코더에게 응답 기재 사항과 잘 대응되는 코드번호의 목록이 제공된다. 코더가 그 목록에서 어울리는 코드번호를 선택하거나 적절한 코드번호에 대한 확신이 서지 않아서 소프트웨어의 도움을 청하면 소프트웨어는 요구에 따라 코딩을 할 수 있다.

전형적으로 코딩을 필요로 하는 변수들에는 산업, 직업, 대학의 전공 분야, 직장, 가정 구매가 있다. 이런 변수에 대해서는 명명법에 몇 백개의 코드번호나 범주가 있을 수 있다. 우리는 각 요소와 각 조사대상의 변수에 대하여 참의 코드번호가 존재한다고 가정한다. 코딩 오류는 요소에 올바른 코드번호가 아닌 다른 것을 배정했을 때 발생한다(*A coding error occurs if an element is assigned a code number other than the correct one*). 결보기에는 간단한 이 정의는 한층 더 상세해야 할 필요가 있다.

첫째, 응답이나 코드할 요소의 특징호합 때문에 올바른 코드번호를 결정하기 어렵다. 기본적인 가정은 각 요소는 단 하나의 범주에 속한다는 것이다. 실제로는 특정한 기재 사항의 해석에 따라서 다른 코드번호가 배정될 수 있는 어려운 상황들이

있다. 이것의 예는 직업에 대한 질문에서 “전화 오퍼레이터(operator)”이라는 응답이다. 이를테면, 직업의 명명법에 전화 오퍼레이터에 대한 두 개의 가능한 범주 있다. 하나는 전화 회사에서 일하는 전화를 다루는 기사이고 다른 것은 다른 종류의 회사에서 일하는 전화 교환원이다. 보조 정보 없이는 이 요소를 명명법의 더 미세한 수준으로 코딩할 수 없다.

둘째, 응답이 설사 자세하다 할지라도 정확한 코드번호를 배정하는데 문제가 발생할 수 있다. 코딩 전문가들 사이에 변동이 상당할 수 있고 그 결과 정확한 코드번호는 운영상의 규칙을 통해서 정의해야 한다. 한가지 이런 운영상의 규칙은 다수결의 원리이다. 예를 들면 세 명 이상의 코딩 전문가가 독립적으로 응답을 코딩한다고 하자. 그러면 정확한 코드번호는 전문가들의 다수가 배정한 것으로 정의한다. 과반수에 이르지 못한 경우에는 추가 규칙들이 사용되어야 한다.

셋째, 이를테면 직업에 대한 응답이 올바르게 코딩되었지만 아직 그 사람에 대한 잘못된 직업일 수 있다. 예를 들어, 설문지를 기입하는 항공 조종사가 “당신의 직업은 무엇입니까?”라는 질문에 어떤 이유로 “승무원”이라고 답할 수 있다. 코더는 이 직업의 올바른 코드번호인 코드712를 배정할 수 있다. 이것은 코딩 오류라기보다 오히려 조종사에 의한 측정오차에 원인을 돌려야 오류이다. 이 예는 직업 코딩의 오류의 일부는 응답자 오류이고 일부는 코딩 오류라는 것을 강조한다. 코딩 작업의 개선은 응답자 오류에는 효과가 조금 있거나 아니면 없을 수도 있다.

대부분의 기준이나 명명법은 숫자의 체계를 사용해서 체계적으로 구성되었다. 첫 자리 수는 주된 분류를 나타내고, 두 번째는 주된 것 내에 더 자세한 분류를 나타내고 등이다. 이 체계는 연속된 일련의 단계를 정하고 많은 통계 결과물(예를 들어, 표 또는 일람표)은 세부 코딩 단계보다는 높은 코딩 단계들을 더 자주 사용하기 때문에, 높은 단계에서 발생하는 오류가 낮은 단계에서 일어나는 것보다 더 심각한 것으로 여긴다. 특정한 단계에서 오류가 발생하자마자 그 이후의 모든 단계들도 마찬가지로 잘못 코딩된다. 예를 들어, 코드번호의 첫 자리의 오류는 모든 단계의 결과에 영향을 미친다. 반면에 낮은 단계들이 결과물의 일부가 아니면 낮은 단계에서만 발생하는 오류들은 결과물에 영향을 미치지 않을 것이다.

코딩 오류율은 다양한 방법으로 계산될 수 있다. 특정 조사 변수, 특정 코드번호나 코드번호 단계 (수의 자리 위치), 개별 코더들에 대한 오류율들이 계산될 수 있다. 이 오류율들은 잘못 분류된 것들의 총수를 측정하기 때문에 총 오류를 측정한다.

다. 하지만 요소들이 특정 코드번호에 잘 못 들어있거나 또는 빠져있으면 오분류된 것들은 상쇄될 수도 있다. 그 경우에는 총오차보다 훨씬 적을 수 있는 순 오차가 더 적절할 수 있다. 전형적으로, 일람표에 영향을 미치는 코딩 오류는 순 오차이다. 다음 두 예가 코딩 오류율을 줄이고 코딩된 자료를 개선하려고 시도하는 동기를 제공한다.

예 7.5.1 건강에 대한 조사에서 특정한 직업의 사람들을 자세히 조사하고 싶다고 가정하자. 우리는 임의로 사용할 수 있는 모든 개인들의 직업을 포함하는 센서스 파일이 있다. 이를테면 우리의 폐병 연구에서 우리의 관심은 광부, 석공, 주택 도장공에 제한되어있다. 이 세 가지 직업들과 관련된 코드번호를 써서 적절한 부모집단을 가려낼 수 있다. 가려낸 부모집단의 조사에서 우리는 그 부모집단에 속하지 않는 사람들을 찾아낼 수 있다. 이것은 코딩 오류 때문에 그들은 우리의 관심 대상인 세 구분 중 하나에 잘못 배정되었다. 이 사람들의 제거는 단지 재정상의 그리고 관리상의 문제이다. 보다 훨씬 나쁜 것은 알 수 없는 수의 광부, 석공, 주택 도장공들이 틀린 코드번호아래에 숨겨져 있고 우리가 그것들을 찾을 가망이 없다 것이다.

예 7.5.2 이 예는 노동 인구 조사와 총 변동에 대한 모수의 추정에 관한 것이다 (즉, 직업과 산업의 상이한 범주간의 총 전환의 수). 일부 연구(Lyberg, 1981)들은 산업 범주간의 변동들과 직업 범주간의 변동들 중 비교적 적은 수가 실제 변동이라는 것을 보여주었다. 대부분의 변동은 코딩 오류 때문이다. 이런 결과의 공식발표는 노동 시장의 이동성의 과장된 묘사를 야기할 것이다. 이런 문제들에 대한 해답은 물론 다양한 대책으로 코딩 오류율을 최소화하도록 하는 것이다.

매우 많은 연구들이 코딩 오류의 빈도가 상당할 수 있다는 것을 보여준다 (즉, 총 오류가 클 수 있다). 예를 들어, 1965년 스웨덴 Census of Population에서 산업 변수에 대한 오류율이 코딩 과정의 여러 순간에 8.2와 14.5% 사이에 변동했다. 높은 비율은 코딩 작업의 초기단계 중에 관측되었고, 높은 비율은 코더들이 더 경험을 쌓은 후반에 관측되었다. 1970년 스웨덴 센서스에서 직업 변수의 추정된 오류율이 13.5%이었고, 산업은 9.9%이었다. 그리고 세대주와의 관계나 일한 시간 수 같은 더 간단한 한자리 수의 변수에 대해서는 오류율이 3.7과 11.5%사이에 변동했다. 1975년

스웨덴 센서스에서 직업의 추정된 오류율은 7.8%이었고 산업의 추정된 오류율은 3.5%였다. 모든 한자리 수 변수들의 추정된 오류율은 0.5와 1.0% 사이였다. 1975년에 대단히 향상된 오류율은 우리가 나중에 검토할 오류 관리를 위한 새로운 절차들에 의해 설명된다.

1970년 미국 인구센서스의 평가에서 산업과 직업의 추정된 오류율은 각각 9.1과 13.3%이었다(미국 센서스국, 1974). 1980년 미국 센서스의 예비 검사에서 미국 Current Population Survey에서 온 경험이 있는 코더들의 산업과 직업의 추정된 오류율을 각각 6.9와 8.1%라고 했다(미국 센서스국, 1977).

지금은 대부분의 대형 기관들이 자동화, 부분적인 자동화, 그리고 수동 접근방법이 혼합된 방법을 사용하기 때문에 지난 몇 십 년 동안에는 비슷한 평가 조사를 찾기 어렵다. 수동 코딩을 사용하는 기관들은 그들의 코딩 작업에 대하여 평가 연구를 실시하지 않는 경향이 있다. 물론 예외도 있다. 하나는 Biemer와 Casper(1994)에 설명된 1991년의 RTI 연구인데 산업과 직업 코딩의 추정된 오류율이 각각 17%와 21%이었다. 또 하나는 Campanelli 등(1977b)에 보고된 영국에서의 직업 코딩의 품질에 대한 연구이다. 그들은 연구에서 상관코더분산의 추정값을 얻었다. ρ_{coders} 의 값(5장 참조)은 매우 작았고, 그들이 어느 중요한 직업군에 관계하나에 따라서 설계효과는 1과 1.79사이에서 변동했다. 설계효과의 대다수가 1과 1.3사이였는데, 300과 400사이의 작업량의 크기를 고려하면 상관코더분산에 기인하는변수의 분산팽창인자가 비교적 크지 않았다. 최근 몇 해 동안 연구가 없음에도 불구하고 문제의 상황은 명백하다. 만일 코딩이 관리되지 않은 상태로 방치되면 오류율은 높아지고 이것은 또한 늘어난 조사 오류와 결함이 있는 분석에 이르게 할 것이다.

7.5.2 수동 코딩 오류의 관리

수동 코딩의 관리에 사용되는 기본적인 두 개의 방법론들이 있다. 그것은 종속검증과 독립검증이다. 종속검증(*dependent verification*)에서는 입력 코더 A가 한 요소를 코딩한다. 이렇게 입력된 코드번호는 검증자인 B에 의해 재검토 된다. B는 코드번호를 면밀히 살피고 그것이 정확한지를 결정한다. 만일 그것이 맞다고 생각되면 그대로 두고 그렇지 않으면 그가 맞다고 생각하는 것으로 코드번호를 바꾼다. <그림 7.2>이 그러한 설계를 나타낸다.

코더 A가 코드번호 x 를 배정한다.

코더 B가 코드번호 x 를 검증하고
코드번호 y 를 배정한다.

$y=x$ 가
성립되니까?

아니오

최종 코드번호는
 y 이다.

예

최종 코드번호는
 y 이다.

<그림> 7.2 코딩의 종속검증

종속검증은 매우 비능률적이다. 경험에 의한 방법으로는 오류의 약 50% 정도만 수정 된다. 일부 연구들은 오류의 감소율이 더 낮을 수 있다는 것을 보여준다. 이러한 낮은 수정률을 야기하는 인식의 작용원리는 이미 배정된 코드번호가 검증자의 판단에 강하게 작용한다는 것이다. 이 방법은 명백하고 모호하지 않은 오류들만 수정하는 경향이 있다. 검증자는 흔히 본래의 코드번호가 전적으로 틀리지 않았기 때문에 바꾸지 말아야 한다고 논리적으로 생각할 수 있어서 덜 명백한 오류들은 원래대로 두는 경향이 있다. 바꿔 말하면, 최초의 코더가 한 판단에 따르는 경향이 있다.

독립검증(*independent verification*)에서는 검증자측의 그러한 불확실성의 근거가 제거된다(즉, 검증자는 처음 배정된 코드번호를 알 수 없다). 독립검증에서는 입력 코더, 이를테면 코더 A가 요소에 코드번호를 배정하고 이것을 x_A 라고 표시한다.

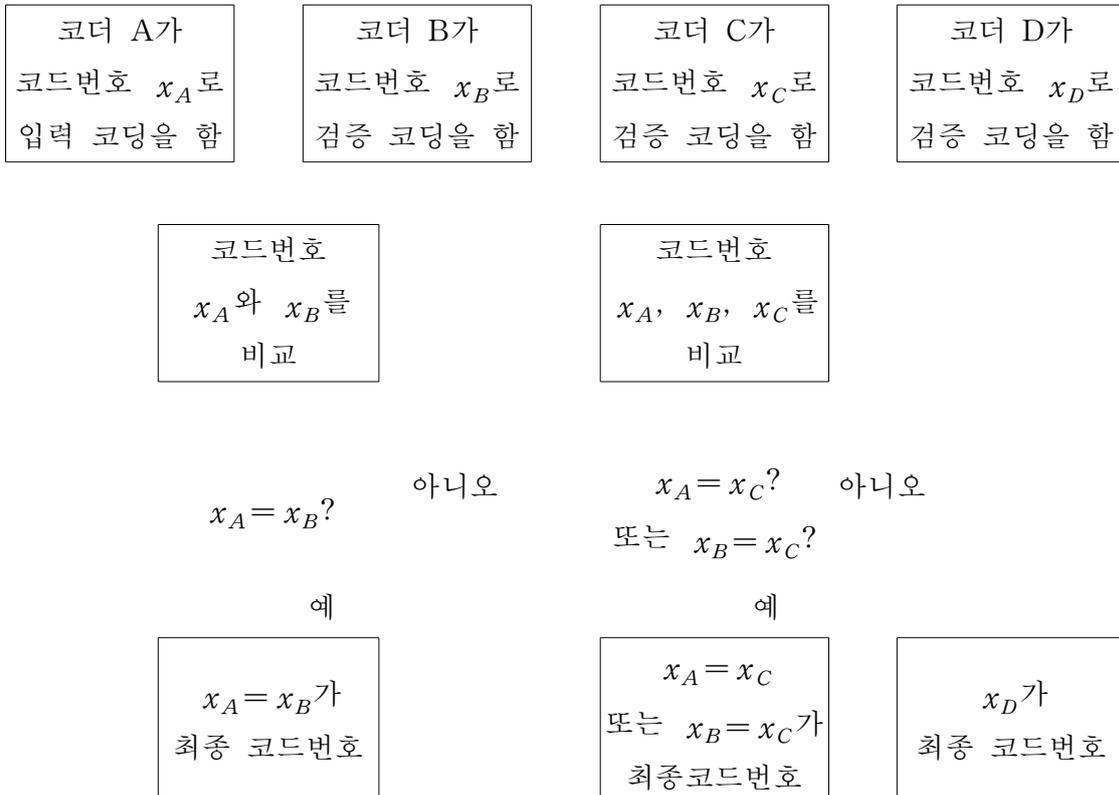
같은 요소를 두 번째 코더인 코더 B가 다시 코딩하는데, 코드번호 x_B 를 배정한다. 두 코더들이 독립적으로 코딩을 하기 때문에 서로 배정한 코드번호를 모른다. 두 코드번호들을 비교하고 최종적인 검증 코드번호를 결정하기 위하여 다음 결정 규칙이 적용된다

- 만일 $x_A = x_B$ 이면, x_A 가 최종 코드번호이다.
- 만일 $x_A \neq x_B$ 이면, 그 요소는 세 번째 코더, 이를테면 코더 C로 넘어가고 그는 코드번호 x_C 를 배정한다.
- 만일 $x_A = x_C$ 이면 x_A 가 최종 코드번호이다.
- 만일 $x_B = x_C$ 이면 x_B 가 최종 코드번호이다.
- 만일 x_A, x_B, x_C 가 모두 다르면 그 요소는 네 번째 코더인, 코더 D가 코딩하고 그가 x_D 를 최종 코드번호로 판결한다.

<그림 7.3>에 묘사된 이러한 체계를 최종 판결을 포함하는 이원독립검증(*two-way independent verification with adjunction*)이라고 한다. RTI 등에서 사용되는 이러한 접근 방법은 세 코더만 있고 필요할 경우 x_C 를 최종 코드번호로 정한다. 또 다른 형태는 세 방향으로 작용하는 삼원독립검증인데 세 명의 독립적인 코더들이 모든 요소들을 코딩하고 최종 코드번호를 코드번호 다수결 원칙에 의해 결정하고, 만일 처음 세 코더 모두의 의견이 다르면 네 번째 코더를 사용한다. 그렇지만 Lyberg(1981)은 최종 판결을 포함하는 이원독립검증이 삼원독립검증과 같은 결과를 내고 비용이 덜 든다는 것을 보여주었다.

독립검증과 종속검증을 비교해 보자. 독립검증 설계의 기본적인 가정은 독립적으로 배정한 두 개나 세 개의 코드번호들은 모두 일치할 것이고 잘못될 확률은 매우 작기 때문에 결정 규칙이 결정한 최종 코드번호 정확한 번호라는 것이다. 하지만 비교되는 코드번호들을 배정하는 코더들의 코딩 능력이 비슷해야한다. 그렇지 않다면, 코더들의 의견은 좀처럼 일치하지 않아서 너무 많은 경우를 판정원에게 보낼 수 있고 그것에 의하여 판정원의 작업량과 판정원오류를 잠재적으로 늘린다. 또한, 방법을 적절히 이해하지 못하는 서투른 두 코더들의 결정이 한 명의 유능한 코더를

결정을 무력화시키는 것이 가능하다. 의료 사망자수 코딩의 다음 예를 보자(Harris, 1974).



<그림 7.3> 최종 판결을 포함하는 이원독립검증.

예 7.5.3 코딩 지시는 다음과 같다. “만일 진단에 X가 존재하면 코드번호 111을 배정하고 다음 요소로 넘어간다.” 부주의한 코더는 만일 진단에 X가 존재하면 코드번호 111을 배정하고 다음 요소로 넘어갈 것이다. 이 코더는 “그렇지만 만일 진단에 Y도 존재하면 코드번호 112번을 배정해야한다.” 라고 명시된 지시의 두 번째 부분을 놓칠 것이다. 한명의 주의 깊은 코더와 같은 조에 있는 두 명의 부주의한 코더들은 코드번호 111을 배정할 것이고, 따라서 주의 깊은 코더의 결과가 오류로 여겨질 것이다.

다행히도 다양한 논문에서 증명되듯이 코딩 작업에서 이런 상황은 드물게 발생한다. 독립검증이 종속검증보다 비용이 더 많이 들기는 하나 더 올바른 코드번호를 배정한다. 최종 판결을 포함하는 이원독립검증 체계는 우리가 알고 있는 독립 체계 중에 비용이 가장 저렴하고 세계적으로 많은 기관들에서 사용되고 있다.

캐나다 통계청에서는 독립검증을 합격판정추출 품질관리 체계의 틀 안에서 실시한다. 그들의 체계에는 상이한 능력 수준의 검증자들이 있다. 첫 단계의 검증자는 입력 코더보다 더 경험이 많고 두 번째 단계의 검증자는 첫 단계 검증자보다 더 경험이 많다. 첫 단계 검증자는 입력 코더가 한 것의 표본을 독립적으로 코딩한다. 의견에 차이가 있으면 두 번째 단계의 검증자에게 보내져서 다수결의 원칙에 의해 판정된다. 두 검증자들이 입력 코더보다 항상 경험이 더 많으므로 위에 설명한 두명의 서투른 코더들이 한명의 유능한 코더를 무력화하는 상황은 아마도 일어나지 않을 것이다.

코딩의 검증 방법에는 종속검증(*dependent verification*)과 독립검증(*independent verification*)이 있다.

7.5.3 자동화 코딩

높은 코딩 오류율, 품질의 큰 편차, 고 비용으로 나타나는 코딩의 문제는 코딩 과정의 최소한 일부분을 컴퓨터를 사용해서 자동화하는 것을 고려하게 하였다. 이미 보았듯이, 컴퓨터는 에티팅에 광범하게 사용되고, 컴퓨터 코딩은 이런 노력의 연장으로 볼 수 있다. 최초의 자동화는 미국 센서스국이 자동화된 지리적 코딩 체계에 착수한 1963년까지 거슬러 올라간다. 과제는 산업과 직업(I & O), 교육, 말로 나타내는 자유 서술형 설명들같은 오류가 생기기 가장 쉽고 귀중한 변수들을 위한 체계를 발전시키는데 있다. 말로 나타내는 자유 서술형 설명은 많은 조사에 흔히 있는 “기타-명시해주시기 바랍니다” 응답항목에도 나온다. 모든 자동화 코딩 체계의 기본적인 특성은 다음과 같다.

1. 코드번호와 관련된 단어나 단어의 일부를 포함하는 컴퓨터에 저장된 사전이나 데이터베이스가 있어야 한다. 이것이 수동 코딩에서 사용되는 명명법에 상응하는 것이다.
2. 응답은 온라인으로 또는 스캐닝이나 컴퓨터에 입력하는 것 같은 다른 방법에 의하여 입력된다.
3. 응답들은 사전의 설명들과 대응시키고, 그 대응과 수반하는 결정 규칙을 근거로 코딩 번호를 배정하거나 응답들을 코딩을 하지 않은 채로 놓아둔다.
4. 중요한 과정 자료를 수집하고 분석해서 코딩 과정을 평가하고 지속적으로 개선한다.

사용되고 있는 체계에는 여러 가지 단계의 자동화가 있다. 한 방향의 전개는 컴퓨터보조코딩(*computer-assisted coding; CAC*)과 결합한 수동 코딩인데 이것은 코더가 어느 코드번호를 배정해야할지 확실하지 않을 때에 CAC 시스템에 도움을 요청할 수 있는 것이다. 예를 들어, 그들은 응답의 첫 세 글자를 입력할 수 있고 그러면 CAC 시스템은 코더가 선택할 수 있는 몇 개의 제안을 제시한다. CAC 소프트웨어가 그것의 최대한의 잠재능력까지 사용할 수 있게 응답들은 표준화된 방법으로 입력된다고 CAC 소프트웨어는 가정한다. CAC 소프트웨어는 그것의 최대한의 잠재능력까지 사용할 수 있도록 응답들이 표준화된 방법으로 입력된다는 것을 전제로 한다. Bushnell(1996)에서 보여주듯이 이 체계는 코더에게 지원을 사용하는 정도에 관하여 개인적인 자유를 주기 때문에 CAC는 상관코더분산을 늘릴 수 있다. 하지만 이 잠재적인 문제가 중요한 결점으로 생각되지는 않는다.

두 번째 방향의 전개는 일괄 방식의 자동화 코딩이다. 여기서는 코딩되어야 하는 모든 응답들을 컴퓨터에 입력하고 코드를 배정하던지 아니면 소프트웨어에 의해 코딩되지 않고 놓아둔다. 코딩되지 않은 경우들은 수동 코딩으로 보내는데 CAC 시스템과 결합될 수도 있다. 자동화 코딩에서는 데이터베이스 또는 코드 사전(*code dictionary*)이 수동 코딩에서 사용되는 명명법 설명들을 대신한다. 사전의 구조는 코딩 안내서의 내용을 근거로 할 수 있지만 경험에 의하면 응답의 경험적인 패턴을 사용하는 것이 훨씬 더 효율적이다. 사전의 실제 편집은 수동으로 하거나 컴퓨터 프로그램을 써서 할 수 있다.

자동화 코딩에는 두 가지 종류의 대응(*matching*)이 있다. 그것은 정확한 그리고

부정확한 대응이다. 정확한 대응(*exact matching*)에서는 코드번호를 배정하기 위하여 입력한 응답이 사전의 기입 사항과 똑같아야한다. 이런 간단한 알고리즘은 틀림 없이 응답의 많은 부분이 코드번호를 배정 받지 못하는 결과를 가져올 것으로 생각할 수 있지만 그것은 적용과 언어에 따라 결정된다. 스웨덴에서는 소비자 구매와 직업 자료는 흔히 간단한 한 단어의 구조라서 정확한 대응에 적격이다. 이런 적용에서의 코딩 정도는 (즉, 자동화 시스템으로 코딩된 응답의 비율) 60에서 80%의 범위 내에 있다(Lyberg와 Dean, 1992).

부정확한 대응을 허락하는 것은 코딩 정도를 늘릴 수 있지만 오류율이 보통 높아질 것이다. 부정확한 대응(*inexact matching*)에서는 만일 응답이 사전 기입 사항 중 하나와 충분히 비슷하면 응답을 대응로 여긴다. 부정확한 대응의 비슷한 종류의 규칙들에는 입력된 설명의 단어 순서를 무시하는 것, 입력된 설명의 계속적인 생략, 특정한 코드번호와 관련된 매우 유익한 단어나 표현을 확인하는 것, 단어에 발견을 돕는 각 단어의 정보 내용에 비례하는 가중치를 배정하는 것이 있다. 그렇지만 정교한 대응 알고리즘의 중요성은 다소 지나치게 높이 평가되었다. 코드번호들의 분포는 흔히 치우쳐 있기(즉, 어떤 코드번호들은 다른 코드번호들 보다 훨씬 더 높은 빈도로 일어난다) 때문에 비교적 소수의 범주와 이 범주에 관련된 설명에 집중하는 것이 매우 효과적이다. 다시 우리는 파레토의 법칙이 힘을 본다.

자동화 코딩 과정 중에 연구되어야 할 몇 개의 중요 과정변수들의 예가 다음에 나와 있다.

- 코딩 정도 (즉, 자동으로 코딩된 응답의 비율).
- 사전의 갱신 후 코딩 정도의 변화.
- 수동 대 자동화 코딩의 범주별 코딩 정도.
- 비용.
- 코딩 방법 (즉, 수동, CAC, 자동화), 범주, 사전 갱신에 따른 코딩 오류율.
- 코더가 시스템에 얼마나 자주 의견을 묻는지에 대한 CAC 자료.

이 모든 변수들은 표본별로 병렬 전문가 코딩(*parallel expert coding*)이 필요한 오류율을 제외하고는 쉽게 측정된다.

자동화 코딩의 목적은 코딩 정도를 최대화하는 사전을 개발해서 수동 코딩을 줄

이는 것이다. 동시에 사전은 자동화된 부분의 코딩 오류가 0에 매우 가까운 정도 이어야 한다. 후자는 사전에 애매하지 않은 표제어만 기재 되도록 하여 이를 수 있다. 만일 애매한 표제어를 허락하면, 코딩 정도와 오류율의 동시 평가는 필수적이 된다. 만족스러운 오류율을 우선 달성하고 그 다음에 오류율을 허용 수준까지 늘리면서 코딩 정도를 늘리는 것이 중요하다.

항상 확장되는 사전이 항상 코딩 정도를 늘리는 최고의 전략이 아니다. 예를 들어, 스웨덴 가구소비조사에서 구매 사전은 자료처리를 하는 한해 동안 17번이나 갱신되었다. 이 과정 중에 사전의 크기는 1459개의 설명에서 4230개로 늘어났다. 세 번째 갱신 후에 그 당시 1760개의 설명을 포함한 사전에 의해 코딩된 표본의 코딩 정도는 67%이었다. 지속적인 새로운 추가 설명에도 불구하고 사전의 나중판은 한번도 73%넘게 코딩하지 못했다. 1980년 스웨덴 센서스에서 사용된 직업사전은 11,000개의 설명으로 이루어졌는데 68%의 코딩 정도를 냈다. 사전을 두 배로 한 1985년 센서스에서도 사실상 같은 결과를 가져왔다. 배정된 코드번호의 치우친 분포로 이러한 상황을 설명할 수 있다.

대부분의 적용에서 수동코딩에 비교해서 비용절약을 보여준다. 센서스와 다른 대형 조사에서 많은 코더들을 임시직으로 고용하는 것이 점점 더 어려워져가고 있기 때문에 대부분의 대형 기관들에서는 어려운 변수의 코딩에서는 어떤 종류의 자동화 코딩을 사용한다. 미국 센서스국이 자동화 산업직업코딩체계(*Automated Industry and Occupation Coding System; AIOCS*)라고 하는 정교한 체계를 개발했다. AIOCS는 정보를 주는 유익한 단어들과 덜 유익한 단어들, 동의어들, 틀린 철자들, 약자들을 확인해서 수동 코딩을 가장하기 위한 설계이다. 정확한 대응이 일어나면, 코드번호 배정은 간단하다. 만일 대응이 정확하지 않으면, 확률적인 가중치들을 사용해서 코드번호들을 배정한다. AIOCS는 작업의 50%를 대략 10%의 오류율로 코딩할 수 있었다 (Chen 등, 1993 참조).

미국 센서스국은 PACE(Parallel Automated Coding Expert)라고 하는 또 하나의 체계를 개발했다. PACE 체계는 자료 병렬 컴퓨팅 기술을 사용하고 초대형 병렬 슈퍼컴퓨터에 수행한다. 대형 전문가-코딩의 데이터베이스는 그 체계에서 “훈련” 도구로 사용되는데 이것은 데이터베이스에 있는 가장 근접한 이웃자리들을 확인하기 위하여 기억력을 근거로 한 추론의 적용을 사용한다. PACE는 응답에 제공된 모든 단어들을 동시에 고려한다. 그리고 이 입력된 데이터의 흐름을 데이터베이스와 비

교하고 그 시스템은 가능한 근접한 이웃들을 내고, 최후의 코드번호는 점수 알고리즘을 근거로 한다. PACE는 산업과 직업 설명의 63%까지 코딩할 수 있었다. 체계의 자세한 설명은 Creecy 등(1992)과 Knaus(1987)에서 찾을 수 있다. 2000년 미국 센서스와 관련한 자동화 코딩의 최근 평가는 Kirk 등(2001)과 Gillman(2000)에서 찾을 수 있다.

캐나다 통계청은 ACTR(*Automated Coding by Text Recognition*)이라고 하는 체계를 개발했다. 이것은 단어 표준화 기술을 사용하여 입력된 원문 파일을 잘 표현된 참조 파일에 대응하고, 원문에 대한 코드번호를 만들어 낸다. 다른 체계와 같이, 이 체계는 사용자에게 표현된 참조 파일이나 원문들과 그것들과 관련된 코드번호들을 제공하도록 요구한다. 체계의 자세한 설명은 Wenzowski(1996)에서 찾을 수 있다. 호주 통계국은 ASCO(*Australian Standard Classification of Occupations*) CAC체계를 개발하였다. 이 체계는 비교적 미숙한 코더들이 숙달된 수동 코더들과 같은 방식으로 수행할 수 있게 하는 많은 특징들이 있기 때문에 매우 효율적이며 아마도 완전히 자동화된 체계보다 뛰어날 것이다. 숙달된 수동 코딩과의 일치성은 95%를 넘는다. 시스템은 가능성 있는 대응과 온라인 도움말 화면, 짧아진 자료 입력 설명, 빠른 찾기와 대응 절차들을 제공하는 사용자 친화적 인터페이스를 가지고 있다. CAC를 사용하는 다른 기관들에는 미국 National Center for Education Statistics, 뉴질랜드 통계청, 네덜란드 통계청, 영국 Office for National Statistics가 있다.

자동화 코딩(*automated coding*)은으로 또는 컴퓨터의 도움을 통하여 수행할 수 있다. 말로 나타내는 설명들을 컴퓨터에 저장된 사전과 대응시키고 받아들일 수 있는 대응이 일어나면 소프트웨어가 배정될 코드번호를 선택하거나 제시한다. 처리되지 못하고 남겨진 모든 코딩은 수동으로 한다. 오류율은 사용되는 대응 기준에 좌우된다.

7.6 파일 준비

자료처리의 마지막 단계는 자료 파일의 준비이다. 파일은 각 표본 단위당 한 개씩, 개별 자료 기록들로 이루어져있다. 파일은 통계 기관에서 표, 그림, 또는 다른 도표를 이용한 방법으로 나타내는 추정값을 얻는데 사용될 수 있다. 파일은 사용자들이 특정 부집단에 대한 그들 자신의 추정값을 만들어 낼 수 있는 데이터베이스의 일부로 볼 수도 있다. 파일이 바르게 기능을 다하기 위하여 다음의 두 가지가 필요하다: (1) 응답하는 각각의 표본단위는 가중치가 배정해야 하고, (2) 각각의 표본단위에 대한 정보가 공개되는 위험을 제한하는 조치들을 취해야 한다.

7.6.1 가중

확률표본조사의 모든 추정과정의 기본적인 원리는 각 표본 단위는 모집단의 단위들을 대표한다는 것이다. 추출틀의 각 구성원이 선택될 기회가 같도록 표본을 추출하면, 표본평균과 표본비율은 응답한 계층의 모집단 평균과 비율의 좋은 추정값이다(3장 참조). 그렇지만, 선택 확률이 같지 않다면 추정값을 얻기 위하여 각 단위에 가중값을 주어야 한다. 이것에 대한 근거는 9장에서 상세하게 다룬다. 또한, 가중은 무응답과 추출틀 미포함을 보완할 수 있다.

대개 표본의 각 단위에 선택 확률의 역으로 기본 가중치를 배정한다. 다음에 무응답과 미포함을 보완하도록 조정 요인들을 이 가중치에 포함시킨다. 그러나 최종 가중치는 언급한 것들만이 아니고 몇 개의 다른 요인들을 포함한다. 예를 들어, 다단계 표본추출을 사용해서 표본을 선택할 때에는 조정 요인들을 단계에 따라 상이하게 적용될 수 있을 수 있다(군 또는 관할 구역, 센서스 지역 등). 이 조정들은 특정 변수의 가중된 표본 분포를 센서스 모집단과 유사한 외면적인 기준에 일치시키려는 의도이다.

이러한 기본 가중치에 대한 사후 조정의 가장 중요한 목적은 무응답과 미포함의 원인이 되는 조사 추정값의 편향을 줄이기 위해서이다. 그렇지만, 기초 가중치의 큰 조정은 문제가 될 수 있다. 그 이유는 그러한 조정은 표본이 모집단의 일부 그룹들을 실제 보다 상당히 적게 대표한다는 것을 나타내기 때문이다. 이것은 응답하는 표본이 목표 모집단을 대표하지 않는다는 것을 의미한다. 조정은 대표되지 못하는 부분을 수정하는 역할을 하지만, 그것은 조정되는 집단의 표본 응답자들과 누락된 단위가 비슷한 특징을 가지고 있다는 강한 가정을 근거로 한다. 특히 무응답과 미

포함이 상당히 많은 조사에서는 사후조사 조정들이 편향을 줄일 수는 있지만 미관 측오차의 해결책은 결코 아니다.

가중 조정이 복잡하면 그것이 바르게 계산되지 않을 위험이 있고 추정값의 평균 제곱오차를 늘릴 수도 있다. 9장에 가중에 대한 간단한 설명이 나오지만 추정과 가중의 상세한 기술적인 문제들은 이 책의 범위를 벗어난다. 관심이 있는 독자들은 Horvitz와 Thompson(1952), Elliott와 Little(2000), Pfeffermann 등(1998), Hidiroglou 등(1995a) 을 참조하기 바란다.

7.6.2 공개의 기피

사실상 모든 국가적인 통계 기구들과 많은 조사 기관들은 거시적 자료와 미시적 자료의 외부 공개에 관한 정책들이 있다. 거시적 자료(*macrodata*)는 도표화, 총수, 빈도를 포함하는 파일들을 가리킨다. 미시적 자료(*microdata*)는 개별의 사람, 가구, 사업체, 또는 다른 단위들에 대한 자료들을 제공하는 기록들의 파일을 가리킨다. 공개 기피(*disclosure avoidance*)라는 용어는 모집단에 있는 특정한 단위가 표본에 있는 단위로 확인되어 일반적으로 알려지지 않은 단위에 대한 정보를 드러낼 수 있을 때 그러한 공개의 위험을 줄이려는 노력을 가리킨다. 그러므로 도표화나 미시적 자료의 제안된 모든 발표에서 공개 위험의 허용 수준을 평가해야 한다.

두 결합되는 필요 사이에 상당히 자주 상충이 생긴다. 한편으로는, 사회는 개인, 기업, 기관에 대하여 자세한 정보가 필요하다. 반면에, 이런 모집단의 표본에 있는 응답자들은 그들이 제공하는 정보가 그들의 기밀성이 보호되는 방식으로 사용되는 것이 보증되어야한다. 이것은 사용자들에게 발표되는 거시적 자료는 모집단에 있는 개인들을 확인할 수 있을 정도로 자세해서는 안 되고, 또한 미시적 자료 파일에는 이름, 주소, 또는 다른 독특하게 신원이 확인되게 하는 정보가 없어야한다.

직접적으로 증명되는 것들은 정의하기 쉽기 때문에 발표된 미시적 자료 파일에서 그것들을 제거하는 것은 간단하다. 하지만 직접적으로 증명되는 것들이 제거되더라도 단위의 특성들의 분석을 통해서 자료파일에 있는 특정한 단위를 재확인(*reidentify*)하는 (즉, 신원을 결정하는)것이 가능할 수도 있다. 이런 종류의 재 신분 증명은 때로는 의도되지 않은 직접 공개(*inadvertent direct disclosure; i.d.d.*)라고 부른다. I.d.d.는 두 가지가 일어날 때 발생한다.

1. 침입자가 (즉, 재 신분 증명을 하려고 시도하는 사람) 거시적 자료 또는 미시적 자료 파일에 포함된 모집단의 개인 구성원을 인지한다.
2. 침입자가 그 모집단 구성원에 대한 그가 알지 못했던 무언가를 다른 출처에서 알게 된다.

만일 위험에 대한 제한 조치를 취하지 않으면, 기업 조사에서 지리적 위치, 총매출액, 고용인 수, 그리고 개인 조사에서 성별, 나이, 직업 같은 몇 개의 흔한 변수들에 대한 정보의 결합을 통하여 i.d.d.들의 위험이 상당히 높아질 수 있다.

예 7.6.1 Sweeney(2000)는 지역 내의 모집단에서 어떤 드물게 생기는 인구통계학적인 결합값을 가지는 사람이 몇 명인지를 결정하기 위하여 1990년 미국 센서스 요약 자료로 실험했다. 그는 적은 수의 특징들의 결합으로 어떤 개인들의 신분 확인이 가능하다는 것을 발견했다. 예를 들어, 미국 인구의 87%가 (2억 1600만/2억 4800만) 단지 {5자리 수 우편번호, 성별, 생년월일}을 근거로 신분 확인이 가능하다. 미국 인구의 반 정도는 (1억 3200만/2억 4800만, 또는 53%) 단지 {위치, 성별, 생년월일}로 확인될 수 있다. 여기에서 위치는 기본적으로 그 사람이 거주하는 도시, 읍, 또는 지방자치체를 뜻한다. 군 단계에서 조차도 {군, 성별, 생년월일}은 미국 인구의 18%를 확인할 수 있다. 이런 이유로 공개적으로 발표된 자료는 군 단계나 그 아래의 지리적인 확인, 또는 나이는 허용할 수 있지만 생년월일을 포함해서는 안 된다.

기밀성 보호에 대한 많은 연구는 개인 단위의 독특한 신분 확인의 위험의 감소에 관한 것이다. 개인 자료 제공자들의 필요 사이의 균형을 유지하면서 이런 위험들을 줄이기 위하여 몇 개의 방법이 개발되었다. 공개 기피 기술 (*disclosure avoidance techniques*)이라고 부르는 이 방법들은 자료 파일 준비에 적용되어야 한다. 기밀성 보장은 응답율을 올리는 방법일 수 있으므로 이 기술들의 일부는 3장에서 언급했었다.

거시적 자료를 위한 방법들

도표화의 공개 기피를 위해 이용할 수 있는 기술들은 세 가지 일반적인 종류로

나뉜다. 셀 삭제, 자료 모으기, 자료 교란이 그것이다. 셀 삭제(*cell suppression*)에 서는 공개 가능성이 높은 표의 값은 생략하고 기밀성 염려 때문에 셀이 생략되었다 는 것을 나타내는 별표나 다른 부호로 대신한다.

자료 모으기(*Rolling-up the data*)의 예는 한 셀을 구성하는 수가 최소 셀의 한계 크기를 넘어서도록 표의 가로줄과 세로줄을 결합해서 더 큰 급 구간 또는 특징들의 새 배합을 만드는 것이다. 이것은 개별 항목의 삭제보다 더 간단한 해답일 수 있지만 표의 묘사적인 그리고 분석적인 가치를 줄이는 경향이 있다.

자료 교란(*disturbing the data*)은 도표의 값들을 어떤 체계적인 방식으로 바꾸어 서 값들이 개별 단위들에 대한 정보를 정확하게 공개하기에는 불충분하지만 도표의 유익한 가치를 악화시킬 만큼 왜곡되지 않는 결과를 낳는 것을 가리킨다. 보통의 어림수로 하는 것이 가장 간단한 예이다. 예를 들어, 표에 있는 값들을 5의 배수에 가장 근접하게 어림잡을 수 있다. 그렇지만 발표될 값들에 랜덤오차의 주어 공개를 기피하기 위한 일련의 기술이 시도되고 있다. 미시적 자료 파일에 “잡음”을 주어서 파일에서 나온 도표화의 어떤 공개의 가능성도 피한다. 이 방법은 자료 생산자를 위해서는 사태를 간단히 할 수 있지만 사용자에게는 문제를 일으킬 수 있다 (Dalenius, 1974).

대부분의 대형 통계 기관들에는 표를 언제 삭제하나에 대한 적절한 규칙이 있다. 예를 들어, n개의 가장 큰 공헌자들이 총 셀 값에 100k%이상을 기여하는 셀을 확인하는 규칙들이 있다. Flygare와 Block(2001)에 나온 다음 예는 이런 규칙의 필요 서의 동기가 된다.

예 7.6.2 특정 산업의 기업들의 총매출액을 보여주는 표에서 한 셀은 두 회사에 의 해 좌우된다. 그 셀은 다섯 회사들의 총매출액이 \$3,295,000이라는 것을 보여준다. 두 번째로 큰 회사의 회장은 자신의 회사의 매출액이 \$921,000이라는 것을 안다. 그는 작은 세 회사들을 합쳐서 많아야 자신의 회사의 매출액의 반이라는 것을 확신한다. 그는 이제 가장 큰 회사의 총매출액을 짐작할 수 있다. 만일 작은 세 회사들이 셀의 값에 아무런 기여를 하지 않으면, 가장 큰 회사의 총매출액은 $\$3,295,000 - 921,000 = 2,374,000$ 이다. 만약 그 회사의 반 정도만 기여한다면 $\$3,295,000 - 921,000 - 921,000/2 = 1,913,000$ 이다. 그러므로 만일 이 특정한 표의 셀이 발표되었다면 그 셀의 두 번째로 큰 회사의 회장은 그들의 가장 큰 경쟁 상대의 총매출액이

구간 (1,913,000, 2,374,000) 어딘가에 있다는 것을 짐작할 수 있었다.

미시적 자료를 위한 방법들

외부 자료 출처에 대응시키는 가능성 때문에 미시적 자료를 공개로부터 보호하기 어렵다는 것이 오랫동안 인식 되었다(Bethlehem 등, 1990). 더구나, 미시적 자료 파일에 대한 공개 위험의 일반적으로 인정된 척도가 없어서 보호가 충분하다는 것을 보증하는데 적용할 수 있는 기준이 없다.

공개 가능성을 줄이기 위하여 사실상 모든 공공 사용 미시적 자료 파일들은 (1)모집단의 표본에 있는 자료만 포함하고, (2)뚜렷한 증명하는 것들은 포함하지 않고, (3)지리상의 세부사항은 제한하고, (4)파일에 있는 변수의 수를 제한한다. 눈에 잘 띄는 변수들을 숨기는데 사용되는 추가 방법들은 다음과 같다.

- **천정 코딩(top coding)** (또는 바닥 코딩 (bottom coding)). 어떤 최대값, 이를테면 M 보다 큰 변수의 모든 값들은 M 으로 대신한다. 예를 들어, 일년에 100,000 이상의 모든 소득은 100,000으로 대신한다. 바닥 코딩은 값들을 bottom에서 자른다는 것 이외에는 같다.
- **구간으로 재코딩하기(recoding into intervals)**. 변수의 값을 의미 있는 구간으로 재코딩하고 변수의 정확한 값을 보고하는 대신 대응하는 구간만 보고한다. 예를 들어, 소득 변수를 $<40,000$, $40,000-100,000$, $>100,000$ 으로 재코딩할 수 있다.
- **난수를 더하거나 곱하기(잡음)(adding or multiplying by random numbers (noise))**. 파일에 있는 각 단위에 대하여 난수를 선택해서 변수값에 더하거나 곱한다.
- **교체 또는 순위 교체(swapping or rank swapping)** (전환(switching)이라고도 부름). 기록의 표본에 대하여 미리 결정된 변수들에 대한 대응을 데이터베이스에서 찾으면서 다른 모든 변수들을 교체한다.(???)
- **선택된 변수의 말소와 대체(blanking out selected variables and imputing for them)** (말소와 대체(blank and impute)라고도 부름). 미시적 자료 파일에 있는 몇 개의 기록에 대하여 선택된 변수들을 말소하고 대체된 값으로 대신한다.
- **응답자의 소그룹을 총괄하여 한 단위의 보고 된 값을 평균으로 대신하기**

(*aggregating across small groups of respondents and replacing one unit's reported value with the average*) (희미하게 하기(*blurring*) 라고도 부름). 희미하게 하는 것을 수행하는 많은 가능한 방법들이 있다. 평균을 내는 기록들의 집단들은 다른 변수들을 대응하거나 관심 대상인 변수를 가려내어 만들 수 있다. 집단에 있는 기록의 수는 (평균을 구할) 고정됐거나 랜덤일 수 있다. 특정 집단의 평균은 집단의 모든 구성원에게 또는 “중간” 구성원에게 배정될 수 있다. 변수별로 상이한 집단 분류가 있는 한 개 이상의 변수에 수행할 수 있다.(???)

Citteur와 Willenborg(1993)에서는 공개 기피 기술들이 재검토 되었다.

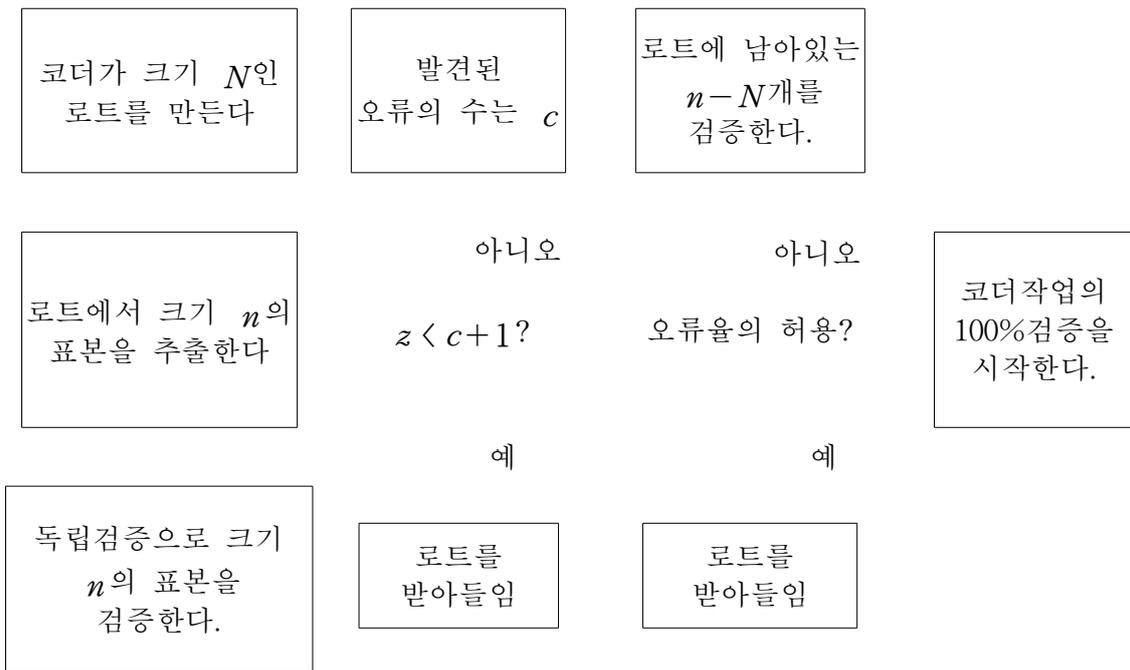
공개 기피 기술들로 연구 목적에 충분하지 않으면 이용할 수 있는 다른 선택들이 있다. 예를 들어, 미국에서 “기밀 부분이 삭제된” 공공 사용 미시적 자료 파일로 요구를 만족하지 못한 연구자들은 자료 제작자에게 보통 돈을 주고 원시 파일의 특수한 도표화를 만들게 할 수 있다. 이것은 연구자가 완전한 미시적 자료 파일에 접근할 수 있었더라면 그가 만들어 낼 수 있었던 것과 같은 표가 제공되는 것이다. 또한, 사용이 제한된 자료 파일들을 신중한 연구자들만 매우 엄격한 지침아래 이용할 수 있게 만들어낼 수 있다. 아마도 직접적으로 확인되는 부분의 삭제를 제외하고는 파일들은 기본적으로 가려진 것이 없다. 하지만, 파일의 이용권을 얻고 싶어하는 연구자들과 기관 대표자들은 그들이 기밀성을 위반하지 않겠다고 서약하는 계약에 서명해야한다. 계약의 위반은 연구자나 연구 기관 또는 모두 제재를 받을 수 있고, 벌금, 또는 징역형까지의 결과를 가져올 수 있다. 어떤 경우에는 연구자들이 제작하는 기관 안에서 일하면서 자료에 자유롭게 접근할 수 있기 위하여 연구자들을 그 기관의 임시 고용인으로 지명할 수 있다. 대부분의 다른 나라들도 비슷한 적절한 절차들이 있다.

공개 기피 기술(*disclosure avoidance techniques*)은 자료 제공자들의 기밀성을 보호하기 위하여 사용된다. 한 종류의 기술들은 거시적 자료에, 또 하나는 미시적 자료에 적용할 수 있다. 모든 조사기관들은 각 국가에서 결정한 자료 보호 규칙을 절대적으로 준수해야이다.

기밀성 보호와 통계 공개 방법론은 통계학자들 사이에 상당한 관심을 일으키는 분야이다. 이 주제에 대한 문헌은 엄청나다. 대표적인 것으로 Fienberg와 Willenborg(1998)와 Willenborg와 De Waal(1996)이 있다.

7.7 지속적인 품질 개선의 적용: 코딩의 경우

검증의 비용을 줄이는 한 방법은 통계품질관리 이론을 사용하여 검증을 표본별로 운영하는 것이다. 쓸모 있는 중요한 품질관리방법(*quality control methods*)들은 합격판정추출법, 과정관리, 그리고 두 가지를 결합한 방법이다. 일반적인 측면은 Ryan(2000)을 참조하라. 이것들은 자료처리 작업의 품질을 보증하기 위하여 몇 십년 동안 사용되고 있는 접근방법들이다. 전형적으로 합격판정추출법에 의해 관리될 항목들은 우선 작업단위(로트)로 나누는데, 작업단위(*work unit*)는 한 작업자가 (즉, 입력자, 편집자, 코더 등) 하루나 이틀의 짧은 기간 안에 할 수 있는 작업의 양이다. 그리고 검사를 위하여 각 작업단위에서 항목을 선택하기 위하여 표본추출 계획을 수행한다. 독립검증 같은 방법을 사용해서 항목들을 점검한다. 발견된 모든 오류들을 세고 어떤 경우에는 수정을 한다. 만일 작업단위에 있는 오류의 수가 미리 정한 허용 수 이하이면 그 작업단위는 통과한다. 그렇지 않다면, 그것은 탈락하고 작업단위 전체를 재작업한다. 그 시점에서 작업은 표본 검사에서 모든 항목들을 오류율이 만족스러운 수준으로 떨어질 때까지 검증하는 전체 검사로 전환할 수 있다. 이러한 검증 접근방법은 작업자들이 그들의 업무수행을 개선하는데 도움을 주기 위해 피드백을 제공한다. 전형적으로 그들에게는 검토의 결과만 주어진다. 만일 그들이 통과하지 못하면 그들이 미래에 검사에 탈락하는 것을 피하기 위하여 그들이 무엇을 해야 하는지에 관한 개선 교육이나 지시를 받을 수 있다. 코더의 작업에 대한 합격판정추출법의 계획을 <그림 7.4>에서 간략하게 보여준다.



<그림 7.4> 허용 수 c 를 사용한 크기 N 인 로트에서 크기 n 의 표본을 얻기 위한 합격판정추출법

캐나다 통계청을 포함한 일부 통계 기관들은 조사 자료처리 작업에서 좋은 품질을 얻기 위하여 합격판정추출법을 성공적으로 사용해왔다. 하지만 일부는 대량 검사 방법과 합격판정추출법의 사용에 이의를 제기했다 (예를 들어, Deming, 1986; Biemer와 Caspar, 1994). 합격판정추출법에 대한 몇몇 반대론은 다음과 같다.

- 대량 검사는 작업을 점검하는데 검증자 팀이 필요하기 때문에 그리고 그것은 일부 작업단위들을 재작업에 이르게 하기 때문에 비용이 많이 든다.
- 코딩에 관한 연구들은 종속검증은 오류를 확인하는데 흔히 효과가 없음을 보여주었다. 독립검증이 더 낫지만 만일 오류율이 낮으면 그것을 더욱 낮추기 위해 100% 검증이 필요할 수 있다.
- 품질 개선의 책임은 입력 코더보다는 오히려 검증자에게 주어진다.
- 검사에서는 검증자가 찾은 모든 오류에 대하여 작업자가 책임이 있다는 원칙이다. 사실 작업자들에게 피드백을 제공하는 이유는 그들에게 그들의 작업에서 오류를 찾았고 앞으로 오류를 피하려면 무엇인가를 해야한다는 것을 알리기 위해서이다.

그렇지만 오류의 원인은 작업자들이 아니라 오히려 작업 방법일 수 있다. 검사는 이러한 사실을 무시되는 경향이 있는데 이것이 품질 전문가들이 품질은 검사를 통해서만 이루어 질 수 있는 것이 아니라고 주장하는 주된 이유이다. 만일 절차들이 불명확하거나 작업자들이 작업 대상으로 할 자료들이 본래부터 불충분하면 검사와 검증은 통한 품질개선의 정도는 제한적일 것이다. 예를 들어 지시들을 정확하게 수행하는 두 코더들은 코딩 과정 자체의 근본적인 문제들 때문에 서로 다른 코드번호를 지정할 수 있다.

지속적인 품질 개선(CQI)에 대한 연구문헌들은 작업 결과물의 품질에 대한 변동을 두 가지로 구분한다. 그것은 특수 원인 변동과 일반 원인 변동이다. 특수 원인 변동은 개별 코더가 범한 오류들 때문에 일어날 수 있고, 일반 원인 변동은 과정 자체 일어날 수 있다. 초기의 품질 개선 노력은, 특수 원인 오류들을 제거하는데 집중해야 한다. 그 이유는 특수 원인 오류는 대개 과정에서 일어나는 오류들의 원인이 되기 때문이다. 하지만, 일단 특수 원인 오류들이 처리되면, 개선 노력은 일반 원인 오류들을 줄이는데 집중해야한다. 그렇지만 일반 원인 오류들을 줄이는 것은 오류율이 낮아지도록 과정을 바꾸는 것을 요구한다. 예를 들어, 그것은 수동에서 자동화 코딩 또는 명명법까지도 바꾸는 것을 필요로 할 수 있다.

일반 원인 오류들은 특수 원인 오류들로 자주 오인된다. 예를 들어, 작업자들은 실제로 시스템 오류 때문에 비난을 받는다. 예컨데, 모든 작업자들이 같은 오류를 반복적으로 범하는 것은 오류의 원인이 작업자 때문이 아니라 작업자들에게한 지시 때문이라는 증거이다. 그러므로 일반 원인 변동의 관측을 근거로 피드백을 작업자들에게 제공하는 것은 이 오류들을 줄이는데 소용이 없다. 그것은 원인이 작업자가 아니기 때문이다. 사실 이런 종류의 피드백은 품질관리 체계가 불공평하다고 느끼는 작업자들의 사기를 꺾을 수 있다.

합격판정추출법과 검사를 수반하는 유사한 품질개선 방법의 문제점들을 고려하면 자료처리 작업의 품질을 관리하는 대안의 전략들이 필요하다. CQI는 앞에서 언급한 결점들을 가지고 있지 않는 한가지 방법이다.

CQI는 검사 방법들과 근본적으로 다르다. 자료품질 개선이 반복되는 과정이라는 개념을 근거로 품질을 개선하는 팀 접근방법이 사용된다. CQI의 기본적인 개념은 <그림 7.5>에서 보여준다. 이 그림은 입력과 그에 이은 작업자들의 실행, 그리고

과정에서 나오는 결과물로 이루어진 전형적인 작업을 보여준다. 현재 수행되고 있는 실제 과정이 있고 이상적인 또는 선호되는 과정(*preferred process*)이 있다. 후자의 과정은 어떤 종류의 부조화 없는 것이다 (즉, 오류가 없거나 거의 없는 과정). 현재 과정에 있는 부조화들을 정의하는 한 방법은 그것의 입력, 실행, 결과물을 선호되는 과정의 그것들과 비교하는 것이다. 실제와 선호되는 과정 사이의 모든 차이는 다음 다섯 단계 접근방법으로 제거한다.

1. 운영의 한 사이클을 실시한다.
2. 이 사이클의 결과를 근거로 운영의 부조화를 확인한다.
3. 작업의 품질에 영향을 미칠 잠재적 가능성이 있는 모든 직원을 대표하는 구성원들로 이루어진 팀들이 관여하는 과정에서 부조화의 근본적인 원인들을 발견한다.
4. 부조화의 근본적인 원인들을 어떤 종류의 행동 또는 작업의 변화로 제거한다.
5. 운영의 새 사이클을 위하여 1 단계로 돌아가서 이 과정을 반복한다.



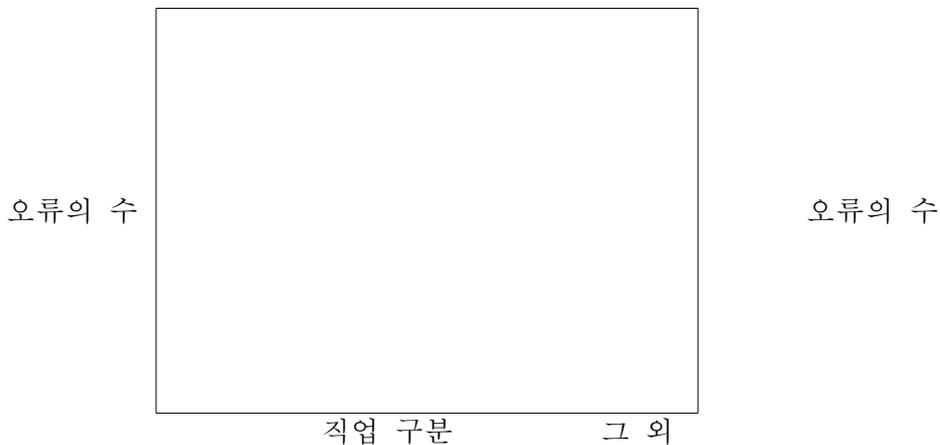
<그림 7.5> 지속적인 품질 개선. 작업의 입력, 실행, 결과물을 선호되는 작업의 입력, 실행, 결과물 관련지어 생각할 수 있다. 한 작업의 매 주기마다 실제 작업과 선호되는 작업 사이의 차이(불일치)의 수를 줄여서 실제 작업이 선호되는 작업에 더 가까워지도록 한다.

그러므로, CQI는 운영이 이상적인 운영에 자꾸 가까워지도록 작업과정의 각 사이클마다 운영의 부조화를 끊임없이 제거하는 과정으로 여길 수 있다. 이상적인 작업은 근본적으로 결함이 없는 것이기 때문에 (보통 이루기 힘든 목적), CQI는 지속적이고 끊임이 없다. 그런데도, 각 사이클은 새로운 개선을 가져오기 때문에 작업은

시간에 걸쳐 계속 개선될 것이다. 그러면 이론적으로 작업의 오류율은 끊임없이 줄어들겠지만 0에 이르지 않는 못할 것이다.

실제 경험에서는 작업에 많은 부조화가 있을 수 있고 이것들 모두를 한 사이클의 과정에서 확인하고 제거할 수 있다고 생각하는 것은 비현실적이다. 어떤 문제들은 제거하는데 여러 사이클이 걸릴 수 있고 직원의 작업투여면에서도 각 사이클에서 몇 개의 문제만에 전력을 쏟을 수 밖에 없다. 그러므로 각 사이클에서 어떤 부조화를 처리할지 결정하는 전략이 필요하다. 이 목적을 위한 매우 유용한 도구는 파레토 원칙이다.

우리가 이 책에서 몇 번 인용하는 파레토 원칙(*Pareto principle*)은 때로는 80/20 규칙(*80/20 rule*)이라고 부른다. 즉, 작업에 있는 문제들의 80%가 부조화의 (또는 오류의) 20%에서 발생한다. 파레토 원칙의 기본적인 개념은 작업의 부조화를 가장 중요한 것부터 가장 중요하지 않은 것까지 정렬시키고 이 목록의 맨 위에서 시작해서 가장 중요한 문제들을 먼저 해결하면서 아래로 내려가는 것이다. 순서는 보통 파레토 그림에 나와있다(간단한 예로 <그림 7.6>을 참조하라).



<그림 7.6> 파레토 그림

파레토 원칙은 작업에서 작업자들이 수행하는 활동의 20%가 그들이 범한 오류의 80%의 원인이 된다는 것을 제시한다. 물론, 80/20 규칙은 대체적인 것이지만 이 원칙이 많은 과정에서 놀랍게도 정확하게 들어맞는다. CQI 원칙은 작업자가 유일한 원인이라고 가정하기보다 오히려 오류의 근본적인 원인을 중점적으로 다룬다. CQI

는 작업자들을 감독자들과 이 과정에 관여할 수 있는 다른 요인을 포함하는 팀의 일부로 본다. 과정의 근본적인 오류의 원인들이 어디에 있던지 그 원인을 확인하고 제거하는 것이 그 팀의 목적이다. 때로는 근본적인 원인이 그들의 작업을 제대로 이해하지 못하는 작업자일 수도 있다. 하지만 그에 못지않게 근본적인 원인들은 과정 자체나 또는 개선하려고하는 작업보다 앞서 이루어지는 과정일 수도 있다. 한 예로, 코딩 오류의 한가지 근본적인 원인이 면접원들이 직업과 산업 자료를 수집하는 방법일 수 있다. 따라서 코딩 운영 한층 더 개선될 수 있도록 코딩 품질 팀에게 자료수집 과정의 변화들을 추천할 권한을 주어야 한다.

Biemer와 Caspar(1994)는 RTI가 사용되는 산업과 직업코딩을 위한 CQI의 적용을 설명한다. 그들의 접근 방법은 다음과 같이 요약할 수 있다:

1. 코딩 작업은 일주일동안 수행되고 최소한 한 개의 오류가 있는 코딩된 모든 코드번호들의 목록을 작성한다.
2. 가장 자주 발생하는 잘못 코딩된 코드번호들을 확인하기 위하여 파레토 분석을 수행한다.
3. 코딩 품질 팀이 만나서 이 코딩 번호들과 오류의 원인들에 관하여 더 상세하게 논의한다.
4. 품질 팀이 선택한 개선 대책을 수행하기 위하여 조치를 취한다.
5. 수행의 효과를 측정한다.

RTI는 코더, 코딩 감독자, 품질관리 전문가로 이루어진 팀들을 사용했다. 연구한 과정은 입력 코더와 검증자가 일치하지 않는 경우에 최종 판결을 포함하는 이원독립검증을 사용한 일반 산업과 직업의 코딩이었다. 이 체계에서는 판정원을 단지 세 번째 코더라기보다 오히려 전문가로 여기므로 판정원이 최종 코드번호의 판단에 최종의 권한을 가진다. 입력 코드번호와 판정원의 선택 사이의 모든 불일치는 부조화나 오류로 여겨졌다.

일주일 단위로 파레토 원칙을 잘못된 코드번호에 적용했다. 그래서 팀들은 가장 자주 잘못 코딩된 코드번호들이나 범주들을 확인했다. 목록의 제일 위에는 다섯 개의 산업과 다섯 개의 직업 범주가 있었다. 팀들은 이 범주들이 그렇게 자주 잘못 사용된 가능한 이유들을 논의하기 위하여 만났다. 많은 문제들은 자료수집 단계에

원인이 있었다. 흔히 산업과 직업 질문에 대한 응답들에는 코더들이 가능한 코드번호들을 구별하기 위해 필요한 중요한 정보가 들어있지 않았다. 면접원들은 코딩의 경험이 없어서 이러한 정보가 필요하다는 것을 알고 있지 못했다. 면접원이 이러한 중요한 정보를 수집하는 것이 그들에게 큰 변화는 아닐 것이다. 그것은 개선된 지시를 개발하고 그것들을 적용하도록 면접원들을 교육시키는 문제였다. 따라서 면접원들에게 새로운 지시들을 보냈고 코더에게 가장 유용한 종류의 정보를 명백하게 하도록 면접원 교육을 보완했다.

새로운 절차들의 수행을 통하여 응답들은 극적으로 향상되었고 또한 코힐씬 더 정확하게 코딩할 수 있었다. 팀워크의 결과 코딩 체계에 많은 다른 변화들이 생겼다. 매주의 회의는 파레토 분석에 전념했지만 개인의 오류 목록 작성도 병행하였다. 여기에서는 코더들은 그들이 올바르게 코딩하지 않은 다섯 개까지의 1목록을 받았다. 그 목록에는 코더가 처음에 본 응답 전체의 원문 그대들로 보여주었다. 목록은 또한 다른 코더가 배정한 코드 번호, 판정원이 배정한 코드 번호, 그리고 판정원의 의견들도 보여주었다. 회의 중에 코더들은 이 예들을 보고 그들이 어떻게 잘못된 코드번호에 이르게 되었나에 관하여 논의할 수 있었다. 감독자들은 이 코드번호들의 적용에 대한 해석 차이를 줄이고 이 코드번호들이 앞으로 올바르게 사용되게 하기 위하여 설명과 재교육을 할 수 있었다. 판정원들은 특정 코드번호를 배정한 그들의 이론적 설명을 제공할 수 있었다.

하지만, 매주의 회의들은 파레토 분석과 개인의 목록 작성에만 제한된 것은 아니었다. 코더들과 판정원들이 작업 환경, 그들이 작업 대상으로 하는 정보의 품질, 그리고 그들이 효율적으로 일할 수 있는 능력을 유발시키는 근무시간 외의 요구 사항 등의 해결하기 어려운 이슈들까지도 제기하도록 하였다. 필요할 경우, 품질 전문가는 결정과 작업 절차들이 코딩 작업에 영향을 줄 수 있는 RTI의 고위 경영자와의 교섭자로 또는 다른 RTI 부서의 직원으로 일하도록 하였다. 이러한 방식으로, 개인적인 코딩 능력을 향상시키고 코딩 작업의 기반 시설을 개선하여 코딩 작업도 개선할 수 있다.

이 모든 변화들은 다소 역동적인 품질 개선을 가져왔다. <그림 7.7>은 산업과 직업 코딩의 결과를 오류율의 각도에서 보여준다. Y축에 그려진 것은 코딩 오류율(CER)인데 이것은 결과 코드번호와 판정원이 배정한 코드 번호에 차이가 있는 경우 수를 검증된 총 경우 수로 나눈 값이다. 한 해의 첫 두 분기에는 산업의 CER은

약8% 정도에서 변동한 것을 볼 수 있다. 3분기 초반에 CQI를 수행했고 CER은 결국은 50% 감소된 4%로 줄었다.



(a)



(b)

<그림 7.7> 1년간의 지속적인 품질 개선 시도에서 (a)산업과 (b)직업의 코딩 오류율

직업에 대한 결과도 비슷하다. 첫 두 분기에는 어떤 주목할 만한 개선 패턴이 보이지 않는다. CQI는 3분기 이전에 수행했고 마지막 두 분기 동안 아주 주목할 만한 개선이 있었다. CER의 감소는 어떠한 비용의 증가 없이 계속되었다. CER이 감소하면 환경원에게 보내지는 경우가 적어지기 때문에, 비용 조차도 감소할 수 있다. 환경의 감소는 팀 회의의 결과이다. 또한 코딩 체계에 일어난 변화들은 생산성을 CQI의 수행 이전과 같은 수준으로 증가시켰다.

최근에, 미국 센서스국(2002)은 Biemer와 Casper(1994)에서 검토된 방법들에 근거하여 코딩 검증 절차와 대조표를 개발하기로 결정했다. 물론, 검토한 CQI 방법들은 다른 조사 단계에서도 사용할 수 있다(10장 참조).

합격판정추출에 매우 심하게 의존되었다고는 하지만 그런 종류의 검사가 최종 품

질을 보증하는 유일하게 합리적인 방식이다(Mudryk 등, 2001b, 2002). 만일 코딩 작업자의 높은 전직률 때문에 코딩 과정이 불안정하면 CQI를 근거로 새로운 작업 절차들을 개발할 충분한 시간이 없을 수 있다. CQI를 사용한다 할지라도 우리는 코딩할 때에 새 코더나 서투른 코더의 결과물이 진행과정에 들어가는 것을 관리할 수 있어야 한다. 이러한 상황에서 우리는 오류의 특수한 원인들을 제거하는데 필요한 피드백과 교육에 제공되는 오류 정보를 사용하면서 새 코더들과 서투른 코더들의 결과물을 관리하는데 합격판정추출을 사용할 수 있다. 그러면 우리는 일반적인 오류의 원인들도 처리할 이원 집단 피드백의 총괄된 품질 관리 결과들을 사용할 수 있다.

이런 접근방법은 품질 관리의 틀안에 있는 CQI로 볼 수 있다. 데밍(Deming)은 그가 스튜어트(Shewhart)의 관리도를 변동의 특수한 원인과 일반적인 원인을 구별하기 위한 통계 도구로 제안했을 때 이 시나리오에 관해 생각하였을 수도 있다. Juran도 품질 관리가 중심이 되는 삼중 품질 경영 모형(Trilogy Quality Management Model)을 품질 개선의 방법으로 개발했을 때 비슷한 생각하였을 수 있다(Mudryk 등, 2002).

7.8 과정의 통합 움직임

여러 저자들이 조사과정의 모든 단계에 뿐만 아니라 자료처리 단계에서 실시되는 작업들의 통합의 필요성을 논의했다. 그러한 논의에 대해서는 Bethel(1997), Shanks(1989), Keller(1994, 1995), Pierzchala(1990), Baker(1994), Weeks(1992)를 참조하라. 전통적으로, 자료처리는 각 작업마다 독립된 공정을 가진 집중식 시설에서 수행된다. 처리는 조립 라인과 비슷하게 순차적으로 한다. 각 작업에는 작업자들과 작업이 어떻게 기능되어야 하는지를 자세히 설명하는 안내서들이 있을 수 있다. 관련된 사람들의 여러 집단이 상당한 양의 자료를 방침에 따라서 과정에서 과정으로 전달한다.

새로운 기술의 출현은 자료처리를 비집중식 방식으로 수행하는 것을 쉽게 하고 지루한 수동 작업의 필요성을 줄일 수 있다. 하지만, 새로운 기술은 앞으로 기관들이 조사 자료를 처리하는 방법을 바꾸게 할 엄청난 잠재력이 있다. 우리는 일괄 처

리를 대신하여 대화식 처리가 표준 운영 기준이 될 것으로 믿는다. 작업은 일괄식으로 보다는 끊임없는 흐름으로 그리고 더 작은 처리 사이클로 과정을 지날 것이다. 또한, 우리는 자료 입력, 에디팅, 도표화, 추정 같은 처리 단계의 훨씬 더 많은 통합을 보게 될 것이다. 네덜란드 통계청은 관리 센터가 있는데 이곳은 조사 자료를 모든 필요한 처리 과정을 거치도록 하기 위한 사용자 친화적인 셸(shell)이다. 또한 그 관리 센터는 자료와 메타 자료 파일들(10장 참조)을 어떤 형태로도 만들 수 있고 따라서 과정 분석을 용이하게 할 수 있다. 보편화된 조사 처리 소프트웨어를 개발하는 유사한 일은 예를 들어, 캐나다 통계청(Turner, 1994)과 스웨덴 통계청(Blom and Lyberg, 1998) 같은 다른 기관에서도 실시되고 있다.

제 8 장 조사오차 평가방법의 개요

이 장에서는 조사 작업에서 비표본추출오차를 평가하는 방법과 기법을 중점적으로 다룬다. 비표본추출오차의 평가는 총오차에 영향을 주는 한 가지 이상의 조사오차 원인을 평가하는 과정이다. 이 책의 앞에 나온 여러 장을 읽은 독자에게는 조사오차 평가라는 주제가 생소하지는 않을 것이다. 이 이유는 조사 자료 수집 방법을 이야기할 때 다양한 종류의 평가 연구의 결과가 이 책 전체에 걸쳐 제공되었기 때문이다. 예를 들어, 3장에서는 조사 추정량의 포함 편향과 무응답 편향을 평가하는 연구의 결과가 검토되었다. 4장에서는 조사 면접 중에 발생할 수 있는 다양한 문제들을 설명하는데 응답 과정의 평가 결과를 사용했다. 5장에서는 면접원 분산 연구와 면접원 업무수행 조사의 결과가 검토되었다. 6장에서는 방식에 대한 비교 연구의 결과가 검토되었고, 마지막으로 7장에서는 자료 수집 후 조사 작업에서 발생하는 오류를 설명하는 조사 평가 연구에서 몇 가지 예들이 제시되었다. 앞의 여러 장에 있는 예들에서 조사오차 평가는 조사 방법론에서 절대 필요한 부분이라는 충분한 증거가 제공된다. 그 이유는 조사오차 평가가 한 방법과 다른 방법의 비교를 통하여 조사를 실시하는 최선의 방법에 도달하게 하는 수단을 제공하기 때문이다.

8.1 조사오차 평가의 목적

이미 보았듯이 조사자료의 수집은 조사비용과 조사오차 사이의 무수한 타협을 필요로 한다. 결국 모든 조사자료는 수많은 원인에 의해 오차가 생기게 된다. 조사오차평가의 목적은 이런 원인들에서 발생하는 오차들을 평가하고, 관리하고, 보완하는 것이다.

조사오차 평가의 목적은 평가를 실시하는 조사과정의 단계에 따라 많고 다양하다. 조사의 설계단계에서 주어진 조사예산으로 가장 높은 자료품질을 제공하는 설계 모수들의 조합을 선택하기 위하여 경합되는 설계 대안들을 비교하는 평가를 실시할 수 있다. 예를 들어, 계획하고 있는 중요한 조사에 선호되는 방법을 결정하려는 목적으로 응답률, 자료 완성률, 응답 신뢰도, 두 가지 자료 수집 방법의 비용비교를 하는 평가 연구를 실시할 수 있다. 이런 연구들은 시험 연구(*pilot studies*)라

고 부르고 많은 중요한 조사의 계획에서 중대한 요소가 된다.

많은 조사들은 실시되기 이전에 조사 절차의 사전 검사(*pretest*)를 위하여 몇 가지 평가 기법이 사용된다. 사전 검사 단계에서의 평가의 목적은 조사 설계의 잠재적인 문제들을 확인하고 그 문제들을 완화시키기에) 가장 좋은 방법을 결정짓기 위해서이다. 예를 들어, 목표 모집단에 있는 몇 명의 사람들을 면접해서 조사 설문지를 사전 검사할 수 있다. 이 장에서 우리는 하자가 있는 질문들을 확인하고 그것들을 어떻게 개선하는지를 판단하는 몇 가지 방법들을 이야기한다. 또한, 사전 검사는 면접의 주제에 대한 응답자의 태도, 면접을 실시했을 때 면접원의 경험, 조사 과제를 완성하는데 직면하는 어려움, 면접을 완성하는데 걸린 시간 등을 평가할 수 있다 (10장 참조).

일단 조사가 진행되면, 평가의 노력은 품질 관리와 품질 보장(*quality assurance*) 기법의 형태를 취한다. 여기서의 목적은 자료를 수집하면서 조사오차를 관리하고 생성 자료가 자료품질의 특정한 기준을 만족시키는 것을 보장하기 위해서이다. 우리는 면접 지침들을 따르고 있는지를 확인하기 위하여 면접원들을 모니터링하고 그들의 일을 검증할 것이다. 같은 설문지에 있는 다른 응답들과의 일치성을 확인하기 위하여 설문지 자료를 검사하고 응답들을 확인할 것이다. 컴퓨터화 된 조사 도구가 올바르게 *branching*하고 있는지 그렇지 않으면 면접원이 면접을 올바르게 인도하고 있는지를 확인하기 위하여 자료에 대한 다양한 검토를 할 수 있다.

마지막으로, 조사가 실시된 후에 몇 개의 상이한 목적을 실행하기 위하여 몇 개 사후조사평가(*postsurvey evaluations*)들을 실시할 것이다. 첫째, 우리는 미래의 조사 설계들을 최적화 하는데 사용할 수 있는 자료를 수집하고 싶을 수 있다. 이 목적을 위하여 조사 설계자는 총 조사오차의 어느 정도가 무응답, 추출틀 포함, 설문지, 면접원, 자료수집 방식, 응답자, 자료 에디팅과 같은 각각의 조사의 주요 오류의 원인에 의하여 영향을 받는지를 알고 싶어 한다. 그렇지만 비표본추출오차를 줄이기 위해 조사 자원의 최선의 배정을 결정하는 데는 오류 요소들에 대한 정보가 요구될 뿐만 아니라 설계자가 고려하는 예산 배정의 대안들에 의해 오류 요소가 어떻게 영향을 받았는지에 대한 정보도 요구된다. 이러한 목적을 가진 평가 연구들에서는 가장 적은 비용으로 가장 정확한 자료를 내는 설계 특징의 조합을 찾기 위해 여러 경합하는 설계 대안들의 자료 품질과 비용이 비교될 것이다.

사후조사 평가들의 두 번째 목적은 자료 사용자들에게 자료 품질에 대한 정보를 제공하는 것이다. 비표본추출오차의 측정에서 자료 품질이 뛰어나거나 매우 좋은 것으로 나타나면 자료의 품질에 대한 사용자의 높은 신뢰를 얻을 수 있으나 품질이 중간 이하로 나타나면 반대의 효과가 나는 경향이 있다. 조사 자료의 발표에 동반하여 공개되는 자료 품질에 대한 보고는 때로는 품질고지서(*quality declarations*)라고 부른다 (10장 참조).

세 번째로, 하지만 드물게 사용되는 사후조사 평가의 목적은 비표본추출편향을 가진 조사 추정값들을 조정하는데 사용되는 자료를 얻는 것이다. 무응답과 포함오차에 대한 사후조사 조정들은 조사 작업에서 상당히 흔하지만 (3장 참조) 측정오차를 보상하기 위한 사후조사 조정은 그리 흔하지 않다. 우리가 2장에서 이야기했듯이 조사 추정값의 측정 편향을 평가하는 것은 진실에 대한 지식을 필요로 하고 (즉, 조사과정 표적에서 정중앙이 어디인지) 따라서 조사 자료 자체만으로는 충분하지 않다. 측정오차의 추정에서는 평가 목적에서 진실로 여기는 절대기준 자료 (*gold standard data*)라고 불리는 추가 자료의 수집이 요구된다. 절대기준 자료를 수집하는 방법들과 기법들을 이 장에서 생각해 본다.

불행히도 조정된 어떤 추정값의 총 MSE가 대비되는 조정되지 않은 추정값의 MSE보다 실제로 더 클 것이다. 이것은 조정이 표본 분산을 늘리는 것보다 편향을 덜 감소시킬 때 일어난다. 또한, 사후조사 조정들은 평가를 실시하고 평가자료를 분석하고 추정값을 어떻게 조정해야 하는지를 판단하는 일정에서 훨씬 더 많은 시간을 요구한다. 이런 이유들로, 이런 조정들을 성공적으로 실천한 예들을 제시한 논문은 거의 없다. <표 8.1>에는 조사 과정의 각 단계에서 조사 평가의 일부 목적들을 요약되었다. 조사 방법론자가 이런 조사오차 평가의 목적을 검토하기 위하여 이용할 수 있는 몇 개의 방법들이 있다. <표 8.1>에는 가장 자주 사용되는 방법들과 우리가 이 장에서 알아볼 방법들의 목록이 나와 있다. 표에 있는 방법들은 그것들이 조사 과정에서 사용될 수 있는 순서로 정리되었다(즉, 조사 설계 단계, 사전 검사 단계, 조사 자료 수집 단계, 사후조사 단계).

<표 8.1> 조사 품질 평가의 방법과 기법

조사 과정 단계	평가 방법	목적
설계	전문가의 설문지 검토 • 비 구조적 • 구조적	설문지 지면 배정, 형태, 질문 표현법, 질문 순서, 지시들
설계/사전 검사	인식 방법 • 행동 코딩 • 인식 면접 • 다른 인식 실습 방법	응답 과정의 한 개 이상의 단계를 평가
사전 검사/조사/사후 조사	결과 보고 • 면접원 집단 토론 • 응답자 중심 토론	설문지와 자료 수집 절차들을 평가
사전 검사/조사	관측 • 감독자 관측 • 전화 모니터링 • 테이프 녹음/CARI	면접원 수행을 평가 설문지의 문제점을 확인
사후조사	사후 조사 분석 • 실험 • 비랜덤 관측 • 내부의 일치성 • 외부의 검증	자료 수집 대안의 비교 MSE의 요소들을 추정; 조사 추정값 검증
	사후 조사 자료 수집 • 절대기준의 방법 (예; 재면접 조사와 기록 점검 연구) • 무응답 추적 연구	MSE의 한 개 이상의 요소들을 추정

조사 평가를 수행하는 방법이나 과정의 단계에 관계없이 조사를 개선하기 위하여 적용하는 유용한 원칙은 품질 관리 논문에서 나오는 잘 알려진 계획-시행-점검-행동(*plan-do-check-act*; PDCA) 사이클이다(Scholtes 등, 1994 참조). PDCA 원칙에 따르면 평가 목적, 평가와 평가 결과의 분석에 사용되는 방법, 평가의 일정과 예산을 구체적으로 명시한 평가 계획을 우선 개발한다. 그런 다음에 평가를 실시하고 결과들을 분석한 평가 결과에 근거하여 조사 과정을 변경한다.

PDCA의 다음 단계인 변화의 평가는 실제로 자주 잊혀지기는 하지만 대단히 중요하다. 조사 과정의 강화를 마무리하기 전에, 수행한 변경들이 의도한 결과들이 달성되는지를 보증하기 위해 더 깊은 연구가 필요하다. 한 가지 문제에 대처하기 위

하여 수행한 변화들이 다른 문제들을 자주 일으킬 수 있는데 이것은 다음 번 과정의 평가를 실시할 때까지 발견되지 않고 지나갈 것이다. 예를 들어, 개념의 명료화를 강화시키기 위한 설문지의 변경은 응답자에게 추가적인 부담을 줄 수 있고 이것은 무응답을 늘릴 것이다. CAPI나 CATI로의 도구의 변경은 branching, 자료 입력저장, 자료 에디팅이 제대로 기능을 하는지 철저히 검사해야 한다.

마지막으로, 변화나 수정에 대한 평가를 근거로 변화들은 조사 전체에서 적용하여 수행하고 더 많은 개선을 위하여 PDCA 사이클을 다시 시작한다.

8.2 조사 설계와 사전검사 평가 방법

8.2.1 전문가 검토

조사 설문지를 개발하는 연구자들이 설문지 설계 문제에 대하여 그들의 동료의 조언을 구하는 것은 일반적인 관행이다. 이런 탁상 사전 검사(*desktop pretests*)에는 설문지 설계, 질문 표현법, 면접 중에 응답자들과 면접원들이 완성해야하는 다양한 일들의 비구조적인 검토가 전형적으로 포함된다. 탁상 사전 검사는 조사 “전문가”나 응답자의 관점에서 설문지에 대하여 의견을 달라고 부탁 받은 다소 비전문적인 동료 연구자를 통하여 개인 또는 집단 환경으로 수행될 수 있다. 비공식 검토도 오해할 수 있는 질문들, 혼동되는 레이아웃, 오해하게 만드는 또는 복잡한 지시, 오타, 다른 설문지 문제들을 확인하는데 효과적일 것이다. 그렇지만, 설문지 설계 전문가들이 이런 검토를 실시하더라도 일부 중요한 문제들은 그래도 확인되지 않을 위험이 있다.

더 포괄적인 검토는 Lessler 등(1992)이 개발한 구조적인 설문지 검토 (*structured questionnaire review*) 접근 방법을 사용해서 할 수 있다. 구조적인 설문지 검토에서는 설문지에 있는 각 질문은 우리가 4장에서 공부한 인지응답과정에 근거한 기준에 의하여 검토된다. 각 기준에는 질문이 제시하는 문제의 종류를 나타내는 코드를 배정한다. 코딩 체계는 조사 질문들의 체계적인 평가에서 지침으로 사용되고 질문 운영, 해석, 또는 인지 처리의 어려움에 이르게 할 수 있는 질문들의 표현이나 잠재적인 구성의 문제들을 검토자가 확인하는 것을 돕는다. 검토자는 각 질문을 단

계적인 방식으로 질문 특성의 특정 범주를 고려하면서 검토하고 각 단계에서 질문들이 읽는 것, 지시, 항목 명료성, 가정, 지식이나 기억, 민감성 응답 항목 등에 대한 문제들의 원인이 될 수 있는 특징들을 가지고 있나를 판단한다. 각 단계에서 질문 코딩 체계는 자주 발견되는 문제점의 종류를 나타내는 열련의 코드(약칭의 설명)를 제공한다. 검토자는 코딩 용지에 문제점에 대응되는 코드를 기록하고 그 코드를 배정한 이유를 써서 문제가 있는지를 나타내고 평가를 종료한다. 검토자는 코딩 범주에 해당되지 않는 항목에 대한 다른 의견도 기록할 수 있다. 면접원이 작업하는 설문지의 코딩 범주의 더 완전한 목록이 <그림 8.1>에 제공되었다

구조적인 전문가 검토(*structured expert review*)는 조사 설계나 사전 검사 단계 중에 수행할 수 있는 비교적 저렴한 방법이며 자료 품질을 크게 개선시킬 가능성이 있다. 이 방법의 한가지 단점은 특정 코드를 어떻게 그리고 언제 적용하는지를 아는 설문지 설계 전문가들을 필요로 한다는 것이다. 더욱이, 경험이 적은 검토자가 때로는 자료품질에 심각한 결과를 주지 않는 것을 설문지에 있는 문제점으로 제시할 것이다. 그렇지만 전문적인 검토자도 이러한 문제를 얼마나 잘 피하는지는 검토자에 따라 다를 것이다. 경험이 많은 설문지 설계자는 어떤 종류의 문제들이 중요하고 확인된 문제들을 어떻게 해결하는지를 판단할 수 있을 것이다.

설문지의 전문가 검토는 대개 설문지 사전 검사 과정의 초기 단계에서 실시한다. 평가에서 나타난 문제점에 대처하기 위하여 설문지가 일단 수정된 후 다른 구조적인 또는 비구조적인 검토를 받을 수 있다. 이러한 검토-수정-검토 과정은 전문가 검토로 나타날 수 있는 모든 중요한 문제점들이 해결될 때까지 여러 번의 반복할 수 있다. 이 단계에서 설문지는 다음절에서 설명할 한가지 이상의 방법이 사용된 인지 검사 와 같은 다른 형태의 검사를 받을 수 있다.

설문지의 전문가 검토는 구조적 또는 비구조적일 수 있다. 비구조적인 검토에서(*unstructured review*)는 검토자가 찾아낸 것들을 편리한 형태로 보고한다. 구조적인 검토(*structured review*)는 응답과정이 설문지 설계에 의해 불리하게 영향 받을 수 있을 모든 가능한 방법들에 대한 더 체계적인 과정이다.

1. **읽는 것에 관한 문제:** 면접원들이 질문을 모든 응답자들에게 일률적으로 읽어주는 것이 어려운지 판단한다.

- 1a - 무엇을 읽나: 면접원들은 질문의 어느 부분을 읽어야 하는지를 판단하는데 어려움이 있을 수 있다.
- 1b - 빠진 정보: 질문하는데 면접원에게 필요한 정보가 질문에 포함되어 있지 않다.
- 1c - 어떻게 읽나: 질문이 완전하게 쓰여있지 않아서 읽기가 어렵다.

2. **지시에 관한 문제:** 소개, 지시, 설명에 대한 문제를 응답자의 관점에서 찾아본다.

- 2a - 상반되는 또는 정확하지 않은 지시, 소개, 또는 설명.
- 2b - 복잡한 지시, 소개, 또는 설명.

3. **항목의 명료성에 관한 문제:** 질문의 의도나 의미를 응답자에게 전달하는 것에 관련된 문제를 확인한다.

- 3a - 표현: 질문이 길거나, 어색하거나, 문법에 어긋나거나, 복잡한 구문을 포함한다.
- 3b - 전문용어들이 정의를 내리지 않았거나, 명확하지 않거나, 또는 복잡하다.
- 3c - 애매한: 질문을 해석하거나 무엇을 포함하고 제외해야 하는지를 판단하는 다양한 방법이 있기 때문에 질문이 애매하다.
- 3d - 기준기간이 명확히 적혀있지 않거나 상충한다.

4. **가정에 관한 문제:** 가정한 것이나 근본으로 논리에 문제가 있는지를 판단한다.

- 4a - 응답자나 환경에 대하여 적당하지 않은 가정을 한다.
- 4b - 사실상 변하는 상황에서 일정한 행동이나 경험 패턴을 부적당하게 가정한다.
- 4c - 다수의 함축적인 질문들을 포함하는 애매한 (이중 목적에 적합한) 질문.

5. **지식/기억에 관한 문제:** 응답자들이 정보를 알지 못하거나 또는 기억하는데 문제가 있는지를 조사한다.

- 5a - 지식: 응답자가 답을 알 것 같지 않다.
- 5b - 질문한 의견이 존재하지 않을 수 있다.
- 5c - 생각해내지 못 한다.
- 5d - 계산 문제.

6. **민감도와 편향에 관한 문제:** 질문의 민감한 성질 또는 표현, 그리고 편향에 대해 평가한다.

- 6a - 민감한 내용: 질문이 사람들이 일반적으로 말하기 거부한 주제에 대한 것이다.
- 6b - 사회적으로 바람직한 응답을 암시한다.

7. **응답항목에 관한 문제:** 기록될 응답의 범위의 타당성을 평가한다.

- 7a - 적당하지 않거나 어려운 서술형 질문.
- 7b - 질문과 응답항목을 짝을 잘못 짓다.
- 7c - 전문용어가 정의를 내려지지 않았거나, 명확하지 않거나, 복잡하다.
- 7d - 명확하지 않은 응답항목.
- 7e - 겹치는 응답항목.
- 7f - 빠진 응답항목.
- 7g - 응답항목의 비논리적인 순서.

<그림 8.1> 구조적인 전문가 검토 체계의 코딩 범주.

8.2.2 인지 방법

지난 20년 동안 인지분야 방법들은 조사의 사전검사 활동의 중요한 요소가 되어 왔다 (Snijkers, 2002; Tucker, 1997 참조). 이런 방법들은 조사 응답을 하기 위하여 응답자들이 사용하는 과정에 대한 정보 그들로부터 얻는 것을 필요로 한다. 전문가 평가와 마찬가지로 방법들은 4장에서 우리가 알아본 응답 과정 모형을 사용한다. 면접을 하거나 설문지를 완성하는 응답자들에 대한 관측을 통하거나 그들이 어떻게 질문을 해석하고, 정보를 생각해내고, 어떤 정보가 적절한지 판단하고, 답을 구체화 하는지에 대하여 이야기를 들어서, 조사 방법론자들은 조사 질문들의 문제점에 대하여 많이 알 수 있다.

우리가 이 절에서 공부하는 방법들은 행동 코딩, 인지 면접, 몇 가지 특수 목적의 기법 즉, 바꾸어 말하기, 시나리오 분류, 응답 대기시간 측정법 등이다. 일부 방법들은 면접원 지원 방법에서 더 적절하지만 우리가 이야기하는 대부분의 방법들은 본 조사에서 사용되는 자료수집방법과 관계없이 응답과정의 다양한 상황을 평가하는데 사용할 수 있다.

행동 코딩

행동 코딩은 (*behavior coding*) 면접원들 및/또는 응답자들이 면접 중에 보이는 어떤 종류의 행동의 빈도를 기록하는 체계적인 방법이다. 특히 흥미로운 것은 조사 오차에 관계되는 것으로 생각되는 행동들이다. 행동 코딩은 전화나 대면조사 중에 응답자와 면접원사이의 상호작용에 대한 연구에 특히 유익한 기법이다. 한 예로, 연구자는 면접원이 조사 질문의 표현을 바꾸거나, 질문에 대한 응답을 분명하게 하거나, 더 상세하게 하는 것이 필요함에도 불구하고 탐색조사를 하지 않는 횟수에 관심이 있을 수 있다. 연구자는 또한 응답자가 설문지에 사용되는 용어들을 분명하게 해달라고 부탁하거나 질문의 의미에 대하여 혼동하는 횟수를 알고 싶어할 것이다. 많은 면접에 걸친 이런 행동에 대한 분석을 통하여 방법론자들은 특정한 조사 질문에 대한 문제점이나 면접 지침과 면접에 관련된 다양한 과제들을 수정할 필요성에 대하여 많이 알게 될 것이다.

행동 코딩은 사전검사 단계나 또는 입력 자료수집 단계에서 사용될 수 있는 중요

한 평가 도구이다. 하지만, 이 절에서는 설문지 설계와 사전 검사를 위한 도구로서의 행동 코딩에 중점을 두었다. 조사 과정의 단계와 관계없이 행동 코딩의 기법들은 근본적으로 모두 같다. 실시간이나 또는 테이프에 녹음된 면접으로부터 조사 질문을 행동 코딩을 할 수 있다. 하지만 후자의 방법이 보통 더 효율적이고 정확하다. 녹음된 면접들은 실시간 코딩과는 달리 면접원의 스케줄에 의존하지 않고 분석자가 편리한 때에 면접을 코딩 할 수 있다. 면접원과 응답자사이 사이의 대화를 분석자가 필요한 만큼 되풀이해서 재생할 수 있기 때문에 코딩의 신뢰도가 늘어난다. 또한, 실시간 코딩보다 녹음을 이용한 코딩을 했을 때 면접에 대해 훨씬 더 많은 정보를 기록할 수 있다. 이것은 실시간 코딩에서는 코더들이 면접원들과 보조를 맞추기 위해서 훨씬 더 단순한 코딩 체계를 사용해야하기 때문이다.

<표 8.2> 행동 코드의 예

코드	설명
면접원 행동 코드	
E	적혀있는 그대로 정확히 읽는다
W1	비교적 중요하지 않은 표현의 교체
W2	중요한 표현의 교체
P	적절한 탐색조사
F	탐색조사를 하지 않거나 충분하지 않은 탐색조사
응답자 행동 코드	
C	응답자가 분명하게 해 달라고 요구한다
I	질문을 처음 읽는 것을 응답자가 답으로 차단한다

행동 코더들은 면접의 원리와 무엇이 받아들일 수 있는 면접원 행동이고 무엇이 아닌지를 이해하기 위하여 기본적인 면접원 교육을 받아야 한다. 충분한 교육으로 코더들은 테이프에 녹음된 면접들을 고도의 일치성을 가지고 코딩할 수 있다. 그렇지만 분석자는 특정 행동을 어떻게 코딩하는지 판단하는데 면접원과 응답자 사이의 대화를 “재생”할 수 없기 때문에 실시간 면접에서 코딩하는 것에 믿음이 덜 간다.

전형적인 한 개씩 하는 질문에 대한 행동 코딩 계획의 예가 <표 8.2>에 나와 있다. 이 계획을 적용하기 위해서는 분석자가 질문을 읽는 면접원과 응답자의 답을

듣고 표에 있는 한 가지 이상의 코드를 배정한다. 예를 들어, 만일 면접원이 중요한 표현을 바꾸어서 질문을 읽고 응답자한테서 분명하게 해 달라는 요구가 있었으면 코더는 그 주고받은 것에 대하여 W2와 C를 코딩한다. 각 질문에 대하여 이런 방식으로 과정이 계속된다.

<표 8.3> 질문 A와 B에 대한 행동코드의 빈도의 비교 (퍼센트)

행동코드	질문 A	질문 B
E: 적혀있는 그대로 정확히	93	40
W1: 비교적 중요하지 않은 표현의 교체	2	23
W2: 중요한 표현의 교체	3	13
I: 질문 차단	0	17
C: 분명하게 해 달라는 요구	2	7
합계	100	100

상단에 코드 범주를 그리고 질문 번호를 왼쪽에 나열한 코딩 용지에 코드를 기록할 수 있다. 그러면 코딩은 적절한 코드 아래에 있는 상자에 표시하여 간단히 수행할 수 있다. 더 빠른 방법은 컴퓨터 데이터베이스에 직접 코딩하는 것이다. 이것은 코딩 작업 중 어느 시점에라도 결과의 일람표를 만들 수 있다. <표 8.3>은 164번의 면접으로 이루어진 행동 코딩 연구의 결과를 제시한다. 60개의 정도의 질문은 <표 8.2>에 있는 코딩 설계의 단순화한 형태를 사용해서 코딩했다. <표 8.3>은 이 질문들 중에 두 개의 코딩 패턴을 비교한다. 첫 번째 질문인 질문 A는:

당신의 [의료 제공자(health care provider)] 방문의 목적이 무엇이었습니까?

질문은 상당히 간단했고 표에서 볼 수 있듯이 면접원들의 93%는 질문을 나타난 대로 정확히 읽었고 단지 2%의 응답자들만 분명하게 해 달라고 요구했다.

두 번째 질문인 질문 B는 훨씬 더 복잡했다. 그것은 다음을 질문했다.

당신의 가장 최근의 방문에 당신 자신이 얼마나 지불했습니까 또는 지불해야합니까? 보험이 지불한 또는 지불할 액수는 포함하지 말시오. 만일 정확한 액수를 모르시면

당신의 최선의 추정값을 주세요.

질문을 바꾸어 말하는 것, 응답자 차단, 분명하게 해 달라는 요구의 높은 비율이 반영하듯이 이 질문에는 더 많은 문제점이 있었다. 이것은 질문이 응답자가 한 번 읽어서 이해하기에 너무 복잡할 수 있다는 것을 제안한다. 그 문제를 고치는 한 방법은 다음과 같이 질문을 두 개나 세 개의 질문들로 나누는 것이다.

당신의 가장 최근의 방문에 당신 자신이 얼마나 지불했습니까 또는 지불해야합니까?
만일 정확한 액수를 모르시면 당신의 최선의 추정값을 주세요.

당신의 보험이 이 지불 금액의 일부분이라도 부담할 것입니까?
(만일 그렇다면) 당신의 보험이 이 액수의 얼마를 부담할 것입니까?

이 방법이 더 많은 질문들을 요구하기는 하지만 응답자에게는 실제로는 쉬운 작업이 될 수 있다. 응답자가 질문 B가 묻는 것을 전부 이해하고 보험이 떠맡지 않는 자신이 부담하는 비용을 머릿속에서 계산하는 것을 요구하기보다 계산 자체가 아닌 이 계산을 위해 필요한 정보만 요구된다. 이러한 발견은 행동 코딩 연구에서 상당히 전형적이다. 특히 이전에 사전 검사하지 않은 질문들에 대해서 그렇다. 하지만 행동 코딩은 전문가 검토와 인지 면접과 같은 다른 사전 검사 기법들을 사용해서 사전 검사한 질문들에서도 문제점들을 찾을 수 있다. 그러한 이유로 모든 문제점들을 찾기 위하여 한가지 방법에 의존하는 것보다는 질문들을 사전 검사 하는데 두가지 이상의 방법을 사용하는 것이 보통 좋은 생각이다.

행동 코딩 조사의 결과로 흔히 응답자들이 질문이 가지고 있는 문제점을 제안될 수 있고, 따라서 질문을 어떻게 수정해야하는지를 제안할 수 있다. 예를 들어, 면접원이 자주 반복하거나 바꿔 말하는 질문들은 어색하게 표현이 되었거나 발음하기 어려운 단어들 포함 되었을 수 있다. 자주 차단되는 질문들은 면접원이 너무 쉽게 무시하거나 잊을 수 있는 예기치 않은 설명이나 수식어들이 질문의 끝에 포함되었을 수 있다. 분명하게 해 달라고 하는 요구를 받는 질문들도 어색하게 표현이 되거나, 애매하거나 불충분하게 정의된 용어들이 포함되거나, 응답자의 경험이나 기준틀과 맞지 않는 응답이 요구되었을 수 있다. 하지만 응답자에게 응답과정에서 부담치는 문제들에 대하여 질문하는 것을 허용하지 않기 때문에 문제들을 대처하는 적

당한 행동들을 찾는 것은 여전히 주관적이다. 따라서 개선을 위한 조치들이 바라던 효과가 있었나를 알아보기 위해서는 추가적인 시험과 평가가 필요하다. 그런 의미에서 행동 코딩은 설문지 평가와 수정 과정의 초기의 단계로 보아야한다.

마지막으로, 5장에서 보았듯이, 행동 코딩은 질문 전달, 탐색조사, 피드백, 설문지에 있는 지시들을 따르는 것, 조사질문들에 대한 응답자들의 질문들에 답하는 것, 공손하고 예의바름 등 같은 면접 과제에 대한 면접원의 업무수행을 평가하는 데에도 사용될 수 있다. 면접원 업무수행의 행동 코딩 검토에서 획득할 수 있는 정보의 종류는 단지 코딩 계획과 코드들을 정확하게 배정할 수 있는 코더들의 능력에 의하여 제한된다.

행동 코딩(*behavior coding*)은 면접 중에 일어나는 면접원과 응답자 사이의 다양한 종류의 상호작용을 몇가지 행동 범주로 분류하는 코딩 계획으로 이루어진다. 근본적인 목적은 응답 품질에 영향이 있다고 믿어지는 행동의 종류와 빈도를 확인하는 것이다.

인지 면접

인지 면접은 응답과정의 특정 단계에서 (즉, 기호화, 이해, 정보 검색, 응답의 구체화, 연락) 발생하는 오류들을 조사 설계자가 확인할 수 있도록 응답자를 면접하는 방법들을 가리킨다. 인지 면접들은 조사방법론자의 사무실이나 면접을 테이프에 녹음하거나 비디오테이프에 녹화할 수 있는 특별한 시설을 갖춘 인지 실험실과 같은 통제된 환경에서 실시된다. 하지만 응답자의 집에서 본인이 직접 또는 전화로 인지 면접을 실시할 수도 있다. 자기기입식 설문지가 사용될 수도 있지만, 면접은 대개 응답자와 1:1로 하는 면접원이 관리하는 설문지로 실시된다. 보통 면접 후에 더 자세하게 연구할 수 있도록 면접의 기록을 만든다.

한 예로, 미국 Behavior Risk Factors Surveillance System(BRFSS)를 위하여 실시한 인지 면접이 Beatty 등(1996)에 설명되어 있다. 이 면접에서는 응답자들의 질문에 대한 이해와 그들이 응답하는데 사용된 방법들을 유도하기 위하여 원고에 의한 병행 탐색조사(*concurrent scripted probes*)라고 하는 기법을 사용했다. 이 기법

로 면접원은 응답자가 특정 질문에 대한 답을 하자마자 생각 과정에 대한 정보를 탐색조사한다.

인지 면접에서 면접원이 물었다. “신체적 질병과 부상을 포함하는 당신의 신체적 건강을 생각하면 지난 30일동안 당신의 신체적 건강이 좋지 않았던 날이 몇일이었습니까?” 얼마간의 생각 후에 응답자는 “이틀 정도” 라고 답했다. 면접원은 응답자가 이 답에 이르는데 사용된 과정에 대하여 다음과 같은 질문들을 연속해서 하였다.

- 그러한 날짜 수를 어떻게 결정했습니까?
- 당신의 답에 포함시킨 질병과 부상을 설명하십시오.
- 어떤 날 “좋았나” 또는 “좋지 않았나”를 판단하는데 어려움이 있었습니까?

면접원들이 어떤 원고에 의한 탐색조사를 사용하는지 관해서는 그들 자신의 판단에 의존하도록 교육받았다. 원고가 없는 탐색조사도 면접원의 재량대로 허용되었다.

인지 면접의 다른 형식은 말하면서 생각하는 면접(*think-aloud interview*) 인데 여기서는 응답자들에게 그들이 질문을 해석하고, 응답하는데 필요한 정보를 검색하고, 응답을 구체화하고, 응답항목을 선택하는데 사용하는 과정에 관해서 소리내어 생각해 내라고 부탁한다. 말하면서 생각하는 병행 면접에서는 응답자들이 그들의 생각을 그들이 조사 질문에 답하는 것과 동시에 보고하라고 지시한다. 예를 들어, BFRSS 조사의 앞에 말한 질문에서 면접원은 질문을 같은 방식으로 묻고 다음과 바로 다음과 같은 지시를 한다. “자, 당신이 그 질문에 답을 하려고 하면서 정확히 무엇을 생각하고 있는지를 말씀해 주세요. 첫 번째로 생각하려고 하는 것이 무엇입니까?” ... 등이다.

말하면서 생각하는 병행 면접에서는 응답자가 과정을 수행하면서 동시에 면접원들은 응답 과정에 대한 정보를 얻는다. 이 방식으로, 인지 면접의 병행 방법들은 응답자가 조사 질문에 대해 답을 구성하려고 시도하면서 생각하는 과정에 대한 자세한 정보를 얻는데 효과적인 방법일 수 있다. 하지만 병행 방법들은 면접의 정상적인 흐름을 중단시키는 결점이 있다. 그러므로 병행 방법들에서 나온 결과들을 전형적인 조사 환경에 일반화 시키는 것은 위험하다. 또한, 어떤 응답자들에게는 질문에 답을 하면서 말하면서 생각하는 것이 어려울 수 있기 때문에 그 방법의 실행에 문

제가 있을 수도 있다.

회고적인 면접(*retrospective interviewing*)은 실제 조사와 더 비슷한 상황 아래 응답 과정을 연구하는데 사용될 수 있다. 회고적인 탐색조사와 말하면서 생각하는 회고적인 면접은 병행 탐색조사와 말하면서 생각하는 것에 대응하는 회고적인 면접의 두 가지 변형된 면접이다. 회고적인 탐색조사와 말하면서 생각하는 것(*retrospective probing and think-alouds*)에서는 응답자들은 우선 실제 조사와 비슷한 상황 아래 면접을 완성한다. 그 다음에는 조사 응답들을 검토하고 응답자들에게 그들이 어떻게 그들의 답에 이르렀나를 묻는다. 이것은 다음과 같은 탐색조사들을 사용해서 할 수 있다:

- 그 질문에 답했을 때 무엇을 생각하고 있었습니까?
- 어떻게 4일이라는 답에 도달했습니까?
- 좋지 않은 (*not good*) 이라는 용어는 당신에게 무슨 뜻이었습니까?
- 지난 30일동안 당신이 아팠었는지 생각해내는데 어려움이 있었습니까? 어떻게 당신의 질병을 기억했습니까?

말하면서 생각하는 면접과 탐색조사 방법(*think-aloud interviews and probing methods*)은 다음과 같은 응답 과정에서 발생하는 측정오차를 연구하는 도구로 널리 사용된다.

- 어려운 용어나 애매한 질문들에 기인하는 이해 오차
- 예를 들어, 건강에 문제가 생긴 총횡수가 요구될 때, 건강 문제에 대한 말로서의 설명을 탐구조사 하는 것 같은 질문 형식에 충실하는 문제
- 응답에 알맞은 답의 범주를 확인하는 문제
- 요구된 정보를 생각해내는 문제

이 기법을 사용하여 발견한 문제들은 설문지 수정, 자료수집 방법의 변경, 면접원 교육 등으로 대처될 수 있다. 인지 면접의 사용을 설명하는 몇가지 예들이 다음에 나온다.

인지 면접은 (*cognitive interviewing*) 응답자의 인지 과정을 확장된 또는 집중적인 면접 접근방법을 사용하여 알아보는 방법들이다.

예 8.2.1 이해 오차. 미국 CPS의 담배 사용에 대한 특별판에 나오는 질문들의 평가에서 (DeMaio와 Rothgeb, 1996) 회고적인 탐색조사 실험계획안(protocol)을 사용해서 다음 질문을 시험했다. “당신은 지난 12개월 동안 담배를 하루이상 끊은적이 몇 번 있습니까?” 이 질문의 의도는 담배를 끊으려는 시도를 측정하기 위한 것이었다. 그렇지만 응답자들은 담배를 끊으려는 의도 외에 다른 이유들, 예를 들어 그들이 아팠거나, 그 전날 과음했거나, 담배를 살 돈이 없었기 때문에 그들이 담배를 끊었던 경우들을 생각했다. 인지 면접 연구에서 나온 추천안은 단순히 “당신이 담배를 끊으려고 했기 때문에” 라는 표현을 원래의 질문의 마지막에 더하여 질문을 수정하는 것이었다.

예 8.2.2 질문 형식에 충실한 문제. 앞에서 설명한 Beatty 등(1996)의 연구에서 “신체적 질병과 부상을 포함하는 당신의 신체적 건강을 생각하면 지난 30일동안 당신의 신체적 건강이 좋지 않았던 날이 몇일이었습니까?”라는 질문은 인지적으로 평가되었다. 일부 응답자들은 숫자로 응답하거나 (예를 들어, “이틀이나 사흘”) 또는 “0”을 함축하는 방법으로 답을 했지만 대부분은 “제 건강은 좋습니다. 지난 30일 동안 의사를 찾아가지 않았습니다,” 또는 “저는 아픈 데가 많습니다.” 같은 답들을 했다. 이 결과들을 연구한 방법론자들은 숫자로 응답하지 않은 응답자의 수에 상당히 놀랐다. 그 결과는 응답자들로부터 적절한 답을 얻으려면 면접원들에게 광범한 탐색조사를 하도록 요구해야 한다고 제안했다. 따라서 어쩌면 높은 수준의 면접자의 주관성이 응답안에 들어있게 된다. 이러한 발견들은 요구되는 질문 형식으로 응답하지 않는 원인과 개선책을 다른 검사들에 의해 확인하는 것을 권유했다.

예 8.2.3 응답항목에 관한 문제. 의료 기관에 대한 조사에서 자기기입식 설문지를 말하면서 생각하는 회고적인 실험계획안을 사용해서 인지적으로 평가했다. 조사 질문 중에 하나는 “데스크탑 컴퓨터의 유지 보수는 한 부서에 맡아서 하는 집중식입니까 아니면 개별 부서가 유지 보수에 책임을 지는 비집중식입니까?”를 물었고 (1)

집중식 또는 (2)비집중식 두 개의 응답항목이 있었다. 많은 응답자들은 둘 중 어느 응답항목도 그들의 컴퓨터 서비스 체계의 조직을 정확하게 설명하지 못하고 “유지 보수”의 어떤 면에 관심이 있는지에 대하여 분명하게 하는 것이 필요하다고 했다. 특히 대형 의료 기관에서는 서비스의 어떤 양상은 집중식으로 되어 있고 다른 것들은 비집중식으로 되어 있다. 이 문제를 고치기 위하여 “집중식”과 “비집중식”을 “대부분 집중식”과 “대부분 비집중식”으로 대신해서 응답항목과 질문을 바꾸었다.

예 8.2.4 정보 검색에 관한 문제. 회고적이거나 병행의 말하면서 생각하는 탐색조사하는 기법은 응답자들이 정보 검색에 사용하는 절차를 알아보는데 사용되어왔다. 예를 들어, 1977년 미국 Census of Manufacturers에서 실시된 회고적인 탐색조사를 통해 많은 응답자들이 회사에 일하고 있는 고용인 수를 회사의 기록에서 그 수를 얻기보다는 기억을 근거로 한 추정값들을 제공하고 있음을 알았다. Sudman 등 (1996)은 인지 면접이 응답자들이 사건의 빈도를 추정하는지, 세는지, 또는 양쪽의 어떤 결합인지를 판단하는 것을 도울 수 있다는 추가적인 예를 제공한다. 그것들은 일어난 사건을 기억하고 사건의 날짜를 제공하기 위해 생각하는 과정이 응답자들에 의해 어떻게 구성되는지를 알아보기 위해 인지 면접이 사용되어진 예들을 제공하였다.

인지 면접방법은 자료의 중요한 측정오차들을 방지하는 강력한 기법들이긴 하지만 그것은 한계점도 가지고 있다. 인지 면접의 한 가지 뚜렷한 한계는 면접이 실시되는 환경이다. 인지 면접은 보통 사무실이나 “실험실”에서 실시된다. 하지만, 생산 환경이 잘 만들어져도 면접의 평가 성향은 실제 생산 면접과 매우 다르다. 인지 면접원은 면접원이라기보다 보통 연구자이고, 면접중에 사용되는 탐색 조사, 질문, 면접 방식은 일상 면접의 일부가 아닌 인지 과정들을 유도한다. 더구나 응답 과정 중에 일어나는 작은 성과로 만족하는 행동에 기인하는 오류들은 대체로 응답 과제에 높은 집중력을 가진 유급 지원자를 사용하는 인지 면접에서는 잘 일어나지 않는다. 이런 이유들로, 인지 면접을 한가지 이상의 다른 설계와 사전 검사 방법들과 겸하는 것이 최선의 방법이다.

인지 실험 방법

응답자들이 조사에서 질문의 응답을 하려고 할 때 부딪치는 어려움들을 알아보기 위한 몇가지 방법들이 있다. 인지 면접을 포함한 이러한 방법들은 인지 실험 방법(*cognitive laboratory methods*)이라고 부른다. 이 방법들은 과정의 한 개 이상의 단계를 연구하는데 사용되는데 이해, 기억, 의사전달 같은 응답과정의 특정한 면을 중점적으로 다룬다. Forsyth와 Lessler(1991)에는 이런 많은 방법들의 개요가 제공된다. 하지만 이 절에서 우리는 흔히 사용되는 응답 대기시간, 삽화, 정렬의 세 가지만 다룬다.

응답 대기시간

응답 대기시간은 응답자가 특정한 질문에 답하는데 걸리는 시간을 측정하고 분석하기 위한 기법들을 가리킨다. 이런 자료의 올바르게 해석은 응답 과정에 대한 통찰력을 제공하는데 유용한 것으로 밝혀졌다. CATI나 CAPI와 같은 컴퓨터화된 면접 방법은 응답 대기시간의 정확한 측정을 가능하게 하였다. 컴퓨터는 질문에 대한 응답을 입력한 정확한 시간 뿐만 아니라 컴퓨터 화면에 질문이 처음 제시된 정확한 시간을 기록할 수 있다. 이 두 시간의 차이가 응답 대기시간의 측정으로 사용될 수 있다. 이 결과, 응답 대기시간 방법들은 조사 연구자들이 더 널리 이용할 수 있게 되었다.

Bassili(1996)는 응답 대기시간 방법들이 질문의 해석, 기억 검색, 응답 선택의 문제들을 확인하는데 어떻게 사용되었는지의 몇 개의 예를 제공한다. 예를 들어, 응답자들이 간단한 질문들에 응답하는데 걸리는 시간을 기준으로 사용해서 Bassili는 응답의 지연된 시간이 긴 질문들을 확인할 수 있었고 그 질문은 이해하는데 어려움이 있을 것으로 생각했다. 그러한 질문들 중 하나는 “큰 회사에서 일정 퍼센트의 여성 고용을 보장하기 위하여 할당 인원수가 있어야한다고 생각합니까 아니면 여성들은 특별한 대우를 받지 말아야합니까?”이었다. 이 질문에서 응답은 평균 2.2초 걸렸고 가장 낮은 응답 대기시간은 1.4초였다.

길고 짧은 응답 대기시간의 해석은 연구되는 질문의 종류에 따라 판단된다. 예를 들어, 기억 검색이 요구되는 질문의 더 긴 응답 대기시간은 보통 더 광범한 인지 과정의 척도로 받아들이고 따라서 더 정확한 응답으로 여긴다. 그렇지만 기억이 요구되지 않는 질문에서의 더 긴 대기시간은 이해와 응답 구성을 위하여 과도한 시간

이 요구되는 것을 암시할 수 있고 따라서 더 어려운 질문으로 암시될 수 있다. 응답 대기시간 방법의 효과는 응답 시간 자료를 분석하는데 사용되는 기법에 크게 좌우된다는 것을 뒤에서 보게될 것이다.

한 예로, Bassili는 대기시간과 응답자가 선택한 응답항목과의 관계를 비교했다. 그가 발견한 것 중 하나는 질문에 “모름”으로 응답하는데 걸린 지나치게 긴 시간이었다. 긴 응답 대기시간 다음에 오는 “모름”은 요구된 정보를 제공하는 것을 거부하는 것(즉, 숨겨진 거부)을 대신한 “모름”이라기보다 오히려 진짜 “모름”을 암시할 수 있다. Bassili는 같은 종류의 논거와 분석적인 기법을 이용하여 의견 조사에 대한 응답에 내포된 마음가짐과 의견의 강도를 평가했다.

응답 대기시간 연구의 결과들은 설문지에 있는 잠재적인 문제점을 확인하는데 상당히 유용할 수 있다. 하지만 대기시간 결과가 제안한 문제들을 검증하고 적절한 개선책을 확인하기 위해서는 결과들의 포괄적인 사후 점검이 필요할 수 있다. 따라서 <표 8.1>의 목록에 있는 의 다른 방법들과 합쳐졌을 때 이 방법이 최고의 가치를 가지게 되는 것을 보게된다.

응답 대기시간(*response latency*) 측정법은 응답자가 특정 질문에 답하는데 걸리는 시간을 측정하고 분석하는 기법을 가리킨다.

삽화

가설적 삽화 기법에서는 응답자들에게 삽화나 이야기로 제시된 가설적 상황들의 짧은 설명들을 그들에게 읽으라고 부탁한다. 그리고 나서 그들에게 삽화에 설명된 가설적 상황이 특정 조사 질문이나 개념에 어떻게 관련이 있는지에 대한 질문들을 한다. 예를 들어, 가설적 삽화는 가공인물의 행동이나 활동을 설명할 수 있고, 그런 후에 응답자에게 다음과 같은 질문을 할 수 있다. “이 시나리오에서 묘사된 사람은 다음 질문에 어떻게 응답해야 할까요?”

미국 National Household Survey on Drug Abuse (NHSDA)를 위하여 RTI에서 실시한 연구는 처방이 필요한 약의 비의료적인 목적의 사용 가능성에 대하여 묻는 질문의 표현에 대한 대안들을 비교하기 위하여 이 기법을 사용했다. 이 실험에서는 주제들이 처방이 필요한 다양한 약의 비의료적인 사용을 확인하려는 질문과 같이

제시되었다. 그런 다음에 그들에게 연속된 삽화들을 읽고 각 삽화에 묘사된 인물에 대하여 질문에 답할 것을 부탁했다. 예를 들어, 질문의 한 형태는 다음과 같았다.

당신은 한번이라도 (1)당신에게 처방되지 않거나 (2)그것이 가져오는 경험이나 느낌을 위하여 의사의 처방이 필요한 약을 복용해 본 적이 있습니까?

실험에서 사용된 11개의 삽화 중에 세 개는 다음에 나온다.

삽화 1. Jim Gillman은 이를 몇 개 뽑았고 치과의사가 그에게 타이레놀과 코데인을 처방해 준다. 매 네시간마다 두 알씩 복용하는 것인데 그는 이가 너무 많이 아파서 매 네시간마다 세 알씩 복용한다. Jim에 대한 이 질문의 올바른 답은 무엇일까?

삽화 2. Greg Wagner는 코프 시럽에 중독되었다. 그는 근방의 여러명의 의사를 찾아가서 거짖으로 기침을 하면서 코프 시럽 처방을 여러 개 받는다. Greg에 대한 질문의 올바른 답은 무엇일까?

삽화 3. 멕시코에 휴가 중에 Elizabeth Clark는 설사를 하고 멕시코에서는 처방이 필요 없는 코데인을 약간 산다. 다음 날 그녀는 코데인의 처방이 필요한 미국에 돌아온다. Elizabeth는 증상들이 없어질 때까지 약을 계속 복용한다. Elizabeth Clark는 이 질문에 어떻게 답해야한다고 생각합니까?

각 경우에, 삽화 다음에 다음과 같은 지시가 뒤따랐다:

만일 [삽화에 있는 사람의 이름]이 [삽화에 있는 약의 이름]을 한번이라도 의사의 처방없이 복용한 적이 있으면 예 상자에 표시를 하시오. 만일 그가 한번도 자신을 위한 의사의 처방없이 복용한 적이 없으면 아니오 상자에 표시를 하시오.

예 아니오

처방이 필요한 약의 비의료적인 사용에 대한 실행적 정의에 따라 실험 대상자들이 11개의 삽화를 가장 바르게 분류를 하게 한 질문 표현법을 결정해서 연구자는 질문이 의도하는 의미를 가장 잘 전달하는 질문 표현법을 결정할 수 있었다(Caspar 등, 1993).

삽화 접근방법의 한가지 단점은 응답자에게 주어진 상황의 인위적인 면이다. 응답자들에게 다양한 가설적 상황에 그들 자신이 있는 것을 상상하도록 부탁했지만 그들은 그들 자신보다는 다른 사람들에 대하여 보고를 한다. 실제로 삽화에 제기된 상황에 있는 응답자는 상당히 다르게 반응할 수 있다. 예를 들어, 불법 약물을 사용하지 않는 응답자는 가설적인 약 사용에 대한 질문들에 응답할 때 공개와 사회적기대부응 편향의 두려움이 없겠지만 실제 약물 사용자는 확실히 그럴 수 있다.

삽화 기법(*vignette technique*)에서는 응답자에게 삽화나 이야기로 제시된 가설적 상황들의 짧은 설명들을 읽고 삽화에서 묘사된 가설적 상황에 대한 조사의 질문에 답해달라고 응답자에게 부탁한다.

정렬

정렬 기법은 다양한 설문지 설계를 판단하는데 적용할 수 있다. 정렬 방법은 비교적 간단하다. 응답자들에게 특별한 순서 없이 배열된 대상을 주고 어떤 기준에 따라 그것들을 정렬하라고 한다. <표 4.5>에 제시한 자료는 정렬 방법이 적용될 수 있다. 조사에서 어떤 주제에 대하여 이야기하는 것이 응답자들에게 얼마나 거부한지를 묻기보다 이야기하기에 “가장 호감이 가는” 주제에서부터 “가장 호감이 가지 않는” 주제까지 순서대로 정렬하라고 그들에게 부탁할 수 있다. 이 방법으로 제공되는 정보는 응답자들이 “민감하다”고 여기는 주제에 관한 정보와 매우 유사할 것이다. 이 정렬 방법에서는 대상을 어떤 기준이나 차원에 따라 분류하라고 응답자에게 부탁하기 때문에 때때로 “차원적인(*dimensional*)” 방법이라고 부른다.

자유 정렬은 응답자들이 대상을 그들에게 이치에 맞는 어떤 기준이나 방법으로 정렬하는 것도 허용한다. 예를 들어, 응답자에게 조사 설문지에 넣을 질문을 주고 그것들을 “비슷한” 질문의 그룹으로 정리하라고 부탁할 수 있다. 예를 들어, 응답자는 질문들을 주제별로(건강, 소득, 활동, 가족 구성 등) 또는 종류별로(행동, 태도, 인구통계학적 등) 그룹을 만드는 것을 선택할 수 있다. 응답자들의 대다수가 좋게 여기는 그룹화는 설문지에 있는 질문들을 모아 정리하는 자연스러운 방법을 제안할 수 있다.

자유 정렬 기법의 더 정교한 사용은 Brewer와 Lui(1996)에 보고 되었다. 한 연구

에서는 미국 NHIS (National Health Interview Survey)에서 사용되는 만성 건강상태의(chronic conditions) 목록에 관심이 집중되었다. 그 조사에서는 응답자들에게 30개 정도의 만성 건강상태가 포함 된 여섯 개의 목록 중에 한 개를 주고 자신이나 가족들이 현재 어떤 건강상태에 있는지 표시하도록 부탁했다. 연구에서 검토한 질문은 100개의 항목을 여섯 개의 목록으로 나누는 방법이 응답자에게 논리적이고 자연스러운지 이다. 그래서 NHIS 검토목록에서 68가지의 대표적인 만성 건강상태를 선택하였고, 색인 카드들에 각 건강상태에 대한 라벨을 타자기로 쳤다. 그 다음에는 70명의 대상자에게 자유 정렬 기법을 사용해서 건강 상태를 정렬하라고 부탁했다. 결과로서 생긴 그룹들이 여섯 개의 목록으로 된 항목들의 현재의 그룹과 얼마나 잘 일치하나를 판단하기 위하여 결과로서 생긴 그룹들의 분석을 수행했다. 저자는 응답자 자신의 자연스러운 그룹이 현재 NHIS 그룹과 일치한다는 결론을 내렸다.

정렬 기법(*sorting technique*)에서는 대상자에게 특별한 순서 없이 배열한 대상들을 주고 어떤 기준에 따라 그것들을 분류하라고 부탁한다.

8.2.3 결과보고와 사전 검사 관측

이 절에서는 조사 방법과 설문지를 설계하고 사전 검사하는 기법인 포커스 그룹을 포함한 결과보고 방법과 사전 관측 검사 방법이 설명된다. 결과보고에 의해 응답자와 면접원으로부터 조사 평가를 위한 유용한 정보를 얻을 수 있다. 면접원 결과보고는 조사 사전 검사 활동의 일부로 정기적으로 실시된다. 전형적인 사전 검사는 각자 12개 정도의 면접을 실시하는 6명에서 12명의 면접원들이 포함된다. 현장 조사가 끝나면 그들의 조사 작업의 경험, 설문지, 조사 요구에 대한 응답자들의 반응, 조사 설계의 다른 측면을 설명하고 논의하기 위하여 응답자들을 모이게 한다. 조사 자료를 수집하는 그들의 경험은 (즉, 설문지와 관련된 질문, 조사에 대한 응답자의 질문, 특정 종류의 표본 구성원으로서의 접근의 어려움) 앞으로의 조사 수행에서 가장 혼란 문제점들을 피하기 위해 자료 수집 절차를 어떻게 바꾸나를 판단하는데 큰 도움이 된다.

면접원 집단 결과보고의 더 형식적이고 구조적인 형태는 포커스 그룹(*focus group*)이다. 전형적인 포커스 그룹에서는 사회자가 미리 결정된 토론 지침에 따라서 토론을 이끌고 토론의 주제를 제기한다. 사회자는 참가자들이 자발적으로 이야기하게 하고, 집단이 토론하고 있는 조사의 특정한 측면에서 너무 멀리 벗어날 경우에만 조언을 하기 위하여 간섭할 수 있다. 면접원이 조사의 각 측면에서 그들이 직면한 문제들이나 성공사례를 밝히도록 하고, 또한 문제점에 대처하기 위한 제안을 제시하도록 유도한다.

면접원 결과 보고와 포커스 그룹은 현재의 설계에 의한 조사에서 어떤 문제에 직면할 수 있나에 대한 귀중한 정보를 제공할 수 있지만 오해를 야기시킬 수도 있다. 예를 들어, 조사의 어떤 측면에 대하여 상당히 부정적인 한 두명의 거리낌 없이 말하는 면접원이 조사에서 비교적 중요하지 않은 문제를 중요한 문제로 제시하도록 이끌 수 있다.

예를 들어, 이러한 한 포커스 그룹에서 다소 떠들석하게 주장하는 면접원이 그의 경험에서는 응답자들이 조사의 주제를 잘 이해하지 못했고 이것은 조사에서 목표 응답률을 달성하지 못할 것이라는 신호라고 불평했다. 포커스 그룹의 다른 면접원들은 거리낌 없이 말하는 면접원의 주제에 대한 강한 의견에 이의를 제기하기를 꺼려했다. 그런데 포커스 그룹을 근거로 한 예측과는 반대로, 실제 조사에서는 이러한 문제가 발생하지 않았다. 이것은 단지 한 면접원 결과 보고가 실제 조사 수행에서의 잠재적인 문제점들을 확인하는 유일한 방법이 되어서는 안 된다는 것을 제안한다.

다른 어려운 문제는 일부 면접원의 의견과 비판이 얼마나 중요한지를 판단하는 것이다. 면접원은 거의 일어나지 않는 응답자의 반응과 설문지의 업무수행에 대한 문제점을 자주 제기한다. 그러므로 문제가 광범하게 퍼진 것인지 아니면 단지 우연한 사건인지를 판단하기 위하여 면접원들이 특정 문제점에 얼마나 자주 직면하는지를 아는 것이 중요하다. 많은 경우에, 제기된 문제점은 빈번히 발생할 가망성이 전혀 없다고 생각될 수도 있다. 포커스 그룹에 의한 악영향의 가능성에도 불구하고, 면접원 결과보고는 조사설계의 장점과 단점들에 대한 훌륭한 정보를 제공한다.

응답자 집단 결과보고와 포커스 그룹들은 개선이 필요한 설문지 설계의 분야를 발견하는 방법으로 매우 효과적일 수 있다. 면접원 결과보고에서와 같이 응답자 포커스 그룹들은 보통 토론을 이끌 사회자와 6명에서 12명의 참가자를 필요로 한다.

응답자 포커스 그룹과 결과보고 회의에서는 이것이 없었다면 조사 설계자에 의해 무시되었을 수 있는 설문지 설계의 분야와 쟁점이 확인될 수 있다. 이런 쟁점들은 말하면서 생각하는 면접이나 관측을 사용해서 더욱 깊이 탐색할 수 있다. 하지만 면접원 결과보고에서는 각 응답자를 토론에 동등하게 참여시키려는 사회자의 노력에도 불구하고 응답자 포커스 그룹의 떠들석하게 주장하는 한 두명의 참가자들이 토론을 지배하고 집단이 설문지의 문제들에 관하여 잘못된 결론을 내도록 할 수 있다. 따라서, 집단에서 발견된 것을 확증하기 위해서는 포커스 그룹 편향을 방지하기 위한 수단으로 두 개 이상의 포커스 그룹이 운영되어야 한다. 또한, 이것은 독립적인 사전 검사 기법으로 보다는 다른 방법들과 함께 사용되어야 한다. 이상적으로, 면접원과 응답자의 제공한 자료는 더 양적인 자료로 보충되어야 한다.

면접을 실시하는 면접원을 관측하는 것은 또 하나의 자주 사용되는 사전 검사 방법이다. 여기에는 대면면접에서 현지 감독자나 조사 설계팀의 선임 구성원이 면접원이 면접을 실시할 때 그들을 동반하는 것도 포함될 수 있다. 관측자는 면접원의 업무 수행, 조사 절차, 설문지, 질문에 대한 응답자들의 반응, 면접 중에 면접원이나 응답자에게 요구되는 활동을 포함한 그들이 관측하는 모든 문제점에 대해 메모를 할 수 있다.

이 방법의 한가지 어려움은 면접 중에 관측자의 존재의 효과이다. 면접원이나 응답자는 보통의 조사 상황에서 그들이 반응하는 것과는 매우 다르게 조사 절차에 반응하게 할 수 있다. 따라서 전형적인 면접 상황에서 자주 일어날 수 있는 일부 문제점을 놓칠 수 있고, 오히려 면접원의 존재 때문에 다른 변칙적인 문제점이 발생할 수 있다. 그럼에도 불구하고, 이 방법은 조사 절차의 문제점과 중요한 논점을 확인하는데 상당히 유용할 수 있다.

덜 드러나는 관측 방법은 차후의 분석을 위해 면접들을 녹음하는 것이다. 이것은 아마 설문지나 면접 지침에 잠재해 있는 문제점의 징후를 요약하기 위하여 행동 코딩 기법이 사용될 것이다. 이 방법은 CAPI와 함께 작동하는 디지털 방식의 컴퓨터 음성녹음면접(*computer audio-recorded interviewing; CARI*) (Biemer 등, 2001) 이라고 부르는 방법을 사용해서 한층 덜 드러나게 만들어졌다. CARI에서는 면접원이 노트북 컴퓨터를 사용하여 면접을 실시하는데, 이 컴퓨터는 디지털 테이프 녹음기의 역할을 한다. CARI의 디지털 녹음 시스템은 노트북에 장착된 마이크를 사용하고 면접 전체 또는 조사 설계자들이 관심이 있을 수 있는 면접의 일부를 녹음하게

프로그램 할 수 있다.

오늘날까지 CARI는 주로 면접 검증 목적으로 사용되어 왔다 (즉, 면접 자료의 면접원 위조를 찾아내고 방지하는). 그렇지만, Biermer 등(2001)은 CARI가 미국 National Household Survey of Adolescent Well-being에서 면접원 업무수행의 평가 이외에도 설문지 평가를 위하여 성공적으로 사용되어 왔다고 보고한다. 여러 연구에 의하면 CARI의 음성 녹음의 품질은 현장에서 테이프 녹음기로 녹음할 때 발생하는 문제점의 걱정없이 적어도 테이프 녹음기에 의한 녹음 만큼 좋다고 보도되었다.

중앙집중식 전화 면접으로 실시된 조사의 사전 검사를 위하여, 전화 통화 모니터링이 관측 방법으로 자주 사용된다. 대면 면접 관측의 경우와 같이 사전 검사 단계 동안 전화 모니터링 관측은 설문지, 면접 지침, 조사 요구에 대한 응답자 반응, 조사 계획안의 전반적인 실현 가능성과 기능성에 초점이 맞추어져 있다. 행동 코딩 방법들이 발견들을 요약하기 위하여 사용될 수 있는데 코더들은 실시간 면접들을 코딩하거나 면접의 테이프와 디지털 녹음들을 코딩할 수 있다.

조사 설문지를 사전 검사를 하기 위한 결과보고 방법들 이외에도 면접원 관측 기법들은 현지에서 또는 전화 시설에서 자료 품질을 모니터링하기 위하여 조사중에 사용할 수도 있다. 기법을 어떻게 적용하나에 대한 유일한 차이는 관측 표본의 크기와 관측자의 관심일 수 있다. 조사품질 평가를 위한 이 기법들의 사용은 다음절에서 더 자세히 설명한다.

8.3 자료품질을 모니터링하고 관리하는 방법

8.3.1 품질관리 방법

품질관리 방법(*quality control methods*)은 자료품질을 개선하고, 자료품질이 허용 기준에 충족되는 것을 보증하고, 조사 자료품질에 대한 정보를 수집하기 위하여 조사의 다양한 단계에 적용하는 기법들을 가리킨다. 일부 품질관리 기법은 이 책의 다른 장(예를 들어, 5장과 7장에서)에서 다루었다.

조사과정의 각 단계에서 비표본추출오차의 위험을 고려하면, 품질관리 방법은 명

백히 필요하다. 조사오차는 그것을 제어하기 위해서 조사 중에 지속적으로 측정되어야 한다. 그 이유는 자주 발생하거나 중대한 오류가 조사 작업의 품질을 서서히 해치는 것을 막기 위한 조치를 취하기 위해서이다. 자료품질에 대한 정보를 실시간으로 주어지면, 조사 관리자는 오류가 관리되는 정도를 판단할 수 있고 필요하면 오류를 줄이기 위해 개입할 수 있다.

조사가 진행되면서 평균제공오차에 대한 비표본추출오차의 영향을 모니터링하는 것이 이상적이지만 거의 그렇게 하지는 않는다. 신뢰도, 측정 편향, 면접원 분산과 같은 평균제공오차의 요소들을 추정하는 것은 조사가 특수하게 설계되고 조사자료에 재면접이나 행정 기록 자료 같은 다른 자료가 보충되었을 때에만 가능하다. 그러므로, 과정의 진행 중에 특정 조사과정과 관련된 평균제공오차의 모든 요소를 (예를 들어, 면접원 오류) 직접적으로 그리고 지속적으로 모니터링하는 것은 좀처럼 가능하지 않다.

대신에 자료품질의 다양한 지표인 응답률, 자료 일치성 비율, 면접원 승낙율, 에디팅 실패율, 항목 무응답률을 측정하고 모니터링한다. 이 책의 여러 곳에서 중요한 과정 변수(*key process variables*)라고 부르는 이러한 품질 지표는 관심의 대상인 평균제공오차의 요소들의 대리 변수의 역할을 한다. 중요한 과정 변수들은 평균제공오차 요소들과 관련이 있기 때문에, 중요한 과정 변수들을 바람직한 수준으로 유지하는 것은 그것들이 반영하는 평균제공오차 요소들도 바람직한 수준에 있을 것이라는 것을 보증할 수 있다. 따라서 최선의 품질 지표는 조사 과정에 걸쳐 편리하게 그리고 지속적으로 관측할 수 있고 관리되어야 하는 오류 요소들과 높은 상관관계를 가진 과정 변수들이다.

예로, 현지에서 조사 자료를 받으면서, 조사 관리자는 중요한 설문지 항목들의 항목 무응답을 모니터링 하여 항목이 적절하게 응답되는지를 확인할 수 있다. 높은 항목 무응답률은 도구의 branching 오류, 면접 지시의 하자, 항목에 대한 예기치 않은 응답자 민감성, 또는 다른 체계 결점의 징후일 수 있고, 잘하면 더 많은 자료를 무응답으로 잃기 전에 수정될 수 있다. 항목 무응답을 줄이기 위해 조사 중에 취한 조치는 항목에서 얻어낸 조사 추정값의 편향과 분산을 줄여서 분석 단계에서 보답을 받을 것이다. 연구 질문들을 처리하기 위한 종속 변수나 독립 변수로 사용될 결정적인 조사 항목에 있어서 특히 중요하다.

자료 입력저장, 자료 에디팅, 코딩, 기록 연결 같은 자료 처리 단계 중에 수행한

품질 관리 방법은 품질 지표의 역할도 할 수 있는 몇 개의 통계량을 산출한다. 예를 들어, 품질 관리에서 사용된 검증 절차에서(7장 참조) 자료 입력과 코딩의 검증 과정의 부산물로 최초의 입력자와 재입력자 또는 최초의 코더와 검증 코더 사이의 일치율에 대한 통계량이 나온다. 이런 자료 처리 통계량이 작업에서 나오면서 자료 품질의 유용한 지표가 된다. 지나치게 높은 입력자 또는 코더의 불일치율은 자료 수집 과정에 대한 문제를 나타내거나 자료 처리 작업자들이 충분히 훈련을 받지 않았다는 것을 나타낸다. 하지만 매우 낮은 입력자 또는 코더 불일치율은 이 작업이 높은 품질의 결과물을 냈다는 보장을 한다.

다음절에서는 전화와 대면 조사에서 면접의 품질을 모니터링하는 몇 가지 방법들을 설명한다. 이 방법들은 면접원들이 일으킨 오류를 관리하기 위해서 뿐 아니라 조사 설문지 수행을 평가하기 위해서도 유용하다.

8.3.2 자료 수집 중의 감독자 관측

면접 중에 발생할 수 있는 오류는 완성된 설문지를 단순히 검사해서는 찾아내기 어렵다. 자료는 상당히 그럴 듯하고 좋은 자료로서 모든 가시의 기준을 만족시킬 수 있다. 그렇지만 면접원이 질문들을 표현된 대로 읽지 않았거나, 적사를 하지 않았거나, 응답자에게 적당한 피드백을 제공하지 않았거나, 응답자의 질문에 정확히 답하지 않았거나, 응답들을 올바르게 기록하지 않아서 생기는 숨은 오류들이 자료에 있을 수 있다. 비록 면접원들은 면접을 실시하고 면접 지침에 따르도록 교육을 잘 받았을지라도 그들은 시간이 지남에 따라 나쁜 습관에 빠져들거나 시간에 쫓겼을 때 손쉬운 방법으로 절차를 진행하기로 결정할 수 있다.

면접을 실시할 때 면접원을 직접 관측하는 것은 특정 오류를 확인하는데 상당히 효과적일 수 있고 필요하면 면접원에게 교정한 조치를 재교육 시킬 수 있다. 면접의 직접적인 관측은 면접원 원인이 아닌 다른 오류를 찾아낼 수도 있다. 예를 들어, 관측은 응답자들이 질문을 해석하고 요구된 정보를 생각해 내는데 문제에 직면하거나 특정한 질문들을 반복해 달라고 부탁하는 것을 드러낼 수 있다. 설문지 설계에서 예상하지 않았던 질문에 대한 응답자 반응이 면접의 직접적인 관측에서 나오는 또 다른 흔한 발견이다.

면접 중에 면접원과 응답자 사이의 상호작용을 모니터링하고 객관적으로 코딩하

는 것은 면접중에 발생하는 많은 “드러나지 않는” 오류를 찾아내는 체계적인 방법을 제공한다. 면접원과 응답자의 행동을 이 방식으로 요약하면, 과정에서 오류에 대처하는 방법들을 개발할 수 있고 면접의 품질이 개선될 수 있다. 자연스러운 면접 환경에서 면접원들과 응답자들을 관측하는 최고의 방법은 드러나지 않는 면접 모니터링의 통해서 가능하다.

전화 통화 모니터링

중앙 집중식 시설에서 실시하는 전화 조사에서는 통화 모니터링이 면접원 업무수행을 모니터링하고 면접원 오류를 관리하는데 널리 사용되는 방법이다. 이런 종류의 모니터링은 대개 면접장소에서 멀리 떨어진 곳에서 감독자나 감독자의 보조원 실시한다. 통화 모니터링 요원들은 면접원이 면접을 실시하는 동안 면접원의 말을 듣고 그들이 관측한 것에 대하여 메모한다. 보통 면접원은 면접을 시작할 때 응답자에게 감독자가 “내가 작업을 바르게 하나 확인하기 위하여” 도청될 수 있다고 알린다. 그렇지만, 정확히 언제 통화가 모니터링 되는지는 응답자와 면접원 모두에게 알려지지 않는다. 이것을 드러나지 않는 통화 모니터링 (*unobtrusive call monitoring*)이라고 부른다.

어떤 경우에는 면접원에게 그들이 언제 모니터링 되는지를 알리는 것이 오히려 더 나을 수 있다. 예를 들어, 면접원이 면접을 하기 싫어하는 응답자에게 전화를 다시 걸때 어떻게 처리하는지에 대하여 최근에 교육을 받았다고 하자. 교실에서 실시하는 교육의 연장으로 교육 후에 면접원의 업무수행을 관측하기 위하여 면접원이 집중적으로 모니터링 될 수 있다. 이러한 상황에서 모니터링 요원이 적당한 수의 이런 경우를 관측할 수 있도록 면접원에게 특정 기간동안 이런 종류의 경우에 집중하도록 부탁하는 것이 더 효율적이다. 그렇지 않으면, 모니터링 요원이 같은 수의 첫 번째 면접 요청을 거부하는 경우를 드러내지 않고 듣기 위해서는 더 긴 기간동안 관측해야 할 수 있다.

어떤 모니터링 체계들은 질문 전달, 탐색 조사, 피스백 등을 질문별로 체계적으로 코딩하기 위하여 절 8.2.2에서 설명한 것 같은 행동 코딩 계획을 사용한다. 다른 코딩 계획은 단순히 면접원의 친절, 지식, 프로 근성, 면접의 속도 이외에도 방식, 목소리 톤 등 면접원의 행동과 면접의 다양한 속성을 기록한다. 통화 모니터링의 중요한 양상은 면접 중에 관측된 지시에 순응하지 않는 행동이 즉시 수정될 수 있도록

록 면접원에게 모니터링 결과를 적시에 피드백을 하는 것이다. 우리가 5장에서 보았듯이 전화 면접원을 모니터링하고 그들의 업무수행에 대한 피드백을 그들에게 제공하는 일은 면접원 분산이 조사 추정값들에 미치는 영향을 줄이는 것으로 믿어진다.

통화 모니터링에서 질문별 행동 코딩의 사용의 한 가지 이점은 설문지에 있는 특정한 항목의 업무수행에 대한 정보를 얻는 것이다. 이 정보는 미래의 조사 수행을 위해 설문지에 필요한 개선을 알아 내는데 대단히 유용할 수 있다. 설문지의 문제점이 면접원에 의해 보고될 수도 있지만, 이런 문제들을 체계적으로 코딩하면 문제가 일어나는 빈도의 객관적인 측정이 제공된다. 어떤 경우에는 이런 문제들이 자료수집 중에 확인될 수 있고, 미래의 면접을 위하여 설문지를 고치게 할 수도 있다. 그래서 전화 통화 모니터링은 설문지를 사용하는데 직면하는 중요한 문제들 이외에도 면접원들의 행동을 수정하므로 자료품질에 즉시 효과를 가져올 수 있다.

사회 과학 연구에 종사하는 전화 조사 기관은 면접 운영에서 오류를 관리하는데 면접원들에게 적시의 피드백이 있는 드러나지 않는 통화 모니터링을 정기적으로 사용한다. 모니터링 대 면접의 전형적인 비율은 매 10시간의 면접당 약 1시간의 모니터링이다. 이 비율은 교육 직후에는 면접원을 더 집중적으로 모니터링하고(이를테면, 5에서 7시간의 면접 당 1시간의 모니터링) 모니터링의 결과가 안정되고 좋은 결과를 보이면 덜 집중적으로 모니터링하도록(이를테면, 1:10이나 1:15의 모니터링 대 면접 비율) 조사 중에 변화시킬 수 있다.

현지 면접원 모니터링과 관측

대면면접에서는 전통적으로 면접의 관측을 근본적으로 두 가지 방법으로 실시해왔다. 그것은 테이프 녹음과 감독자의 관측 (Lepkowski 등, 1998) 이다. 첫 번째 방법은 면접원이 면접의 일부나 전부를 테이프로 녹음하게 해서 나중에 그 녹음테이프를 감독자가 검토할 수 있게 하는 것이다. 테이프로 녹음된 것은 면접 과정 또는 설문지의 특수한 연구들을 위해 남겨둔다. 이것은 면접원 오류를 모니터링하고 관리하는 일상적인 방법으로는 거의 사용되지 않는다. 두 번째 방법은 감독자가 (또는 감독자 대리인이) 면접원이 그에게 배정된 경우의 모두 또는 일부를 완성하는데 면접원과 실제로 같이 있는 감독자 관측이다. 감독자는 각 면접 중에 특히 면접원의 면접 지침의 순응에 관련된 논점을 메모하면서 면접원의 업무수행을 기록하고 (코

당하고) 평가한다.

두 전통적인 방법들 모두 두 가지 제한점이 있다. 첫째, 면접원과 응답자는 면접의 전 기간이 관측하거나 테이프를 녹음한다는 것을 알기 때문에 그것들이 상당히 방해가 된다. 그러므로 그 방법들은 관측된 행동이 전형적인 면접원과 응답자의 상호작용을 대표되지 않을 수 있게 면접 환경을 바꾼다. 또한 감독자는 면접원을 관측하기 위하여 때로는 그들이 있는 위치로 상당한 거리를 여행해야하기 때문에 면접원의 현지 관측은 대부분의 나라의 국가적인 대형 조사에서 많은 비용과 시간을 요할 수 있다. 테이프 녹음의 문제점은 현지에서의 테이프 녹음기 사용에 수반하는 운반, 관리, 작동 등의 병참과 관련되었다. 이 방법은 기계적 실패와 기록의 유실 위험이 상당히 높다.

컴퓨터 보조 면접 기법의 최근 발달이 노트북 컴퓨터에 장착된 마이크만을 사용해서 CAPI 면접원의 노트북 컴퓨터의 하드디스크에 직접 드러나지 않으면서 면접들을 디지털 방식으로 녹음을 할 수 있게 만들었다. 시스템이 완전히 소프트웨어의 관리 아래 있어서 미리 결정된 도구의 어떤 부분에서도 모든 면접 또는 무작위로 선택된 면접에 대하여 녹음 스위치를 키거나 끌 수 있다. 이 기능은 정규의 면접 과정을 중단시키지 않고 면접원과 응답자가 말로한 대화를 음성녹음 하는 이전에는 현지 환경에서는 이용할 수 없었던 능력을 제공한다. 이 능력은 녹음 기록을 현지에서 집중식 모니터링을 실시하는 장소로 보내야하기 때문에 모니터링의 결과에 관한 피드백을 면접원에게 주는데 더 많은 시간을 필요로 하는 것을 제외하고는 전화 면접 통화 모니터링을 위한 전화 시설에서 사용되는 통화 모니터링과 매우 비슷하다. 이런 시스템으로 면접원의 집에서 실시하는 비집중식 전화 면접 뿐만 아니라 응답자의 집에서 하는 대면면접에서도 드러나지 않는 면접 모니터링이 가능하다.

8.2.3절에서 언급했듯이 CAPI 면접 중에 면접원과 응답자의 상호작용을 디지털 방식으로 녹음하는 시스템은 컴퓨터 음성 녹음 면접(CARI)이라고 한다. CARI는 현지 면접에 특유한 몇가지 결정적인 요구를 만족시킨다. 그것은 면접 중에 면접원의 행동과 조사 질문들에 대한 응답자의 반응과 함께 현지 면접의 품질을 모니터링하는 방법을 제공한다. CARI는 다음과 같은 적용 범위를 가지고 있다.

- 면접 위조와 면접 오류를 찾아내는 것.

- 면접원 업무수행을 평가하고 면접원들에게 피드백을 제공하는 것.
- 설문지 문제를 확인하는데 사용하기 위하여 그리고 면접원과 응답자의 상호작용을 코딩하기 위하여 음성을 바탕으로 한 정보를 수집하는 것.
- 서술형 질문에 답하여 정보를 기록하는 것.

표준 CAPI 노트북 컴퓨터의 외부에 장착된 고도의 유지보수를 필요로 하는 장비와 관련된 문제만 없다면, 컴퓨터 녹음의 음성 품질은 테이프에 의한 녹음과 최소한 같은 수준이라고 제안한 논문이 있다. 다른 이점은 면접 도중 특정 순간에 녹음을 시작하고 정지하도록 컴퓨터를 프로그래밍 하는 능력이다. 예를 들어, 설문지의 특정 질문이나 부분을 녹음하거나 컴퓨터가 단지 면접 도중 랜덤한 순간에 녹음하기 시작하게 하는 것이 중요할 수 있다. 또한, 컴퓨터는 어떤 면접원과 종류를 녹음할지 제어할 수 있다. 예를 들어, 미리 정한 기준에 의해 맞는 경험이 적은 면접원이나 응답자를 더 빈번히 녹음하게 할 수도 있다.

녹음의 단점은 응답률에 미치는 잠재적인 영향이다. 녹음을 시작되기 전에 “품질 관리 목적을 위하여” 또는 “내 감독자에게 내가 어떻게 수행하고 있는지 들려주려고” 면접의 일부분을 녹음한다고 응답자한테 허락을 얻는 것이 표준관행이다. 면접원과 응답자 결과보고에서 CARI를 이용한 녹음의 허락 요청이 응답률에 인지될 만큼의 효과가 있었다는 증거는 없었다(Biemer 등, 2001). 그러나 녹음이 응답률에 미치는 영향을 추정하기 위한 실험은 아직 실시되지 않았고 어떤 조사에서는 문제가 될 수도 있다.

하지만, 응답률과 응답 품질에 미치는 잠재적인 효과에도 불구하고, 집중식 전화 통화 모니터링에서 실현된 많은 이득은 디지털 방식의 녹음 면접의 사용이 더 널리 보급되면서 분산된 현지 면접에서도 일상적으로 쉽게 얻을 수 있게 될 것이다.

8.4 사후조사 평가

소규모 연구와 인지 방법은 조사에서 응답 과정과 오류의 원인에 소중한 통찰력을 제공할 수 있지만 때로는 다양한 비표본추출오차 편향, 면접원 분산, 그리고 다른 비표본추출오차의 변동 요소 같은 평균제곱오차 요소의 크기를 추정하는 것이

바람직하다. 이러한 목적에서 평균제곱오차의 요소를 추정할 수 있는 실험 설계나 측정 방법을 포함하는 더 큰 규모 조사들을 요구된다. 이런 조사들을 다음 몇 개의 절에서 논의한다.

8.4.1 실험

면접 방법의 대안에 대한 조사, 표현 방법의 대안에 대한 시험, 또는 면접원 효과의 추정 목적으로, 실험은 조사 자료 품질의 평가와 조사 방법들의 개선에 핵심 역할을 해왔다. 실험은 알려지지 않은 결과나 원리를 발견하거나, 가설을 시험이나 입증 하거나, 또는 알려진 원리를 설명하기 위하여 통제된 상태 아래에 실시하는 자료 수집 활동이다. 전형적으로, 조사 실험에서는 평가하거나 시험하거나 비교하기 위하여 표본 구성원이나 단위들이 다양한 처리(면접 방식, 자료 수집 방법 등)에 랜덤하게 배정된다.

조사에서 실험 설계가 처음으로 사용된 사건들 중의 하나는 1940년대 중반 인도에서 Mahalanobis(1946)가 면접원 편향의 증거를 얻기 위하여 농업 조사에서 면접원들을 랜덤하게 배정한 것이다. 그는 우리가 5장에서 논의한 이러한 랜덤화 방법을 상호관입이라고 불렀다.

전면적인 조사에서의 확률화 실험 설계의 사용은 보통 심층실험(*embedded experiments*)이라고 부른다(van den Brakel, 2001). 예를 들어, 앞에서 면접 방식, 사전 편지의 사용, 면접원 교육의 다른 방법 등의 조사 설계의 일부 요인들을 달리 하는 몇가지 실험들은 설명했다. 이런 실험들이 조사 방법 연구의 중심이 된다. 실험 설계는 소규모의 현지 시험이나 시험 조사에서도 사용되어 왔다. 예를 들어, 조사를 실시하기에 앞서 우리는 조사에서 인센티브가 어떻게 사용되어야 하는지를 결정하기 위하여 다양한 수준의 인센티브로 실험을 할 수 있다.

조사 상황에 대한 실험 조작이 더 수월한 실험실에서는 매우 복잡한 처리 구조가 있는 설계도 가능하다. 한 예로, O'Reilly 등(1994)에 지필면접(PAPI), 컴퓨터 보조 자기 면접(CASI), 음성 컴퓨터 보조 자기 면접(ACASI)의 세 가지 자기기입식 방식들을 조사하기 위한 실험 연구가 설명되었다. 그 설계에서는 응답자들이 세가지 조합인 PAPI-CASI, PAPI-ACASI, CASI-ACASI 중에서 한 가지를 택해 두 번 면접을 했다. 또한, 사실상 여섯 가지 처리 조합이 시험될 수 있도록 순서가 정해졌다.

현지에서는 처리 구조가 대개 간단하고한 두 가지 실험 상황들을 조작한다. 전통적인 분할 투표 실험은 두 가지 처리 상황인 현재의 자료 수집 방법 (즉, 설문지, 자료수집 방법 등)과 실험적 방법만을 시험한다. 하지만, 표본 단위들을 확률화하여 처리에 배정하는 것은 상당히 복잡할 수 있다. 한 예로, 두 종류의 면접원 교육과 두 방식의 자료 수집의 시험에서 면접원 교육의 종류는 면접원의 집단에게 랜덤하게 배정되었고 자료 수집 방식 처리는 면접원에게 배정된 가구들에 랜덤으로 배정되었다. 이런 다단계 확률화는 보통 통계적 효율성을 위해서 보다는 자료 수집 관점에서 필요하다.

Groves와 Couper(1998)는 조사 평가에서 분할 투표 설계와 같은 다중 요인 실험의 중요성에 대해 언급했다. 이 책 전체에 걸쳐 우리는 여러 조사 요인들의 상호작용으로 자료품질의 다양한 양상이 영향을 받는 많은 예를 보았다. 예를 들어, 면접원 특징은 응답자 특징과 서로 상호작용을 한다. 이러한 상호작용은 자료수집 방법의 변화에 의해 바뀔 수 있다. 한번에 한 가지 요인만 바꿀 수 있는 경우에는 이런 복잡한 삼원 상호작용을 연구할 수 없다. 또한, 요인들의 상호작용이 존재하는 경우에는 주효과를 서로 상호작용하는 다른 변수의 수준을 생각하지 않고 해석 (즉, 방식이나 또는 면접원 특징의 효과) 하는 것은 오해를 불러 일으킬 수 있다 (예를 들어, 방식 효과가 면접원 경험에 따라 바뀔 수 있다). 이런 이유로 조사 평가 실험에서 다중 요인에 대해 신중히 고려해야한다. 실험에서 주효과와 상호작용의 해석에 대한 자세한 설명이 Box 등(1978)에 나온다.

8.4.2 관측 연구

확률화 실험이 조사 작업에서 항상 가능하거나 바람직한 것은 아니다. 예를 들어, 사업체 크기가 응답률에 미치는 효과를 판단하기를 원한다고 가정하자. 특히 50인 미만의 고용인이 있는 사업체와 50인 이상인 사업체의 응답률을 비교하고 싶다고 하자. 순수한 확률화 실험에서는 각 처리 집단이 원래 표본의 확률화 표본이 되도록 실험의 표본 단위들은 실험에서 비교될 처리 집단들에 랜덤하게 배정되어야 한다. 그렇지만 사업체의 크기는 미리 정해져 있기 때문에 이런 종류의 랜덤 배정은 작은 사업체와 큰 사업체를 비교하는 데에는 가능하지 않다. 그럼에도 불구하고 우리는 사업체의 크기에 따른 응답률의 차이를 알고 싶은 경우가 있다. 이것은 관측

연구들을 위하여 개발된 방법을 사용하면 가능하다.

관측 연구(*observational studies*)에서는 처리 배정은 표본추출 이전에 미리 정해져 있고 실험자에 의해 관리되지 않는다. 즉, 우리의 관심은 아무런 실험적 조작이 없이 단위의 자연적인 특징에 의해 형성된 집단들을 비교 하는 것이다. 관측 연구에서 관심의 대상이 되는 처리는 나이, 인종, 성별 같은 단위들의 특징들일 수 있고 또는 면접 시간, 자료수집 방법, 면접의 길이와 같은 자료수집 운영의 특징일 수 있다. 일부 특징들은 실험에서 조작 될 수 있고 (예를 들어, 우리가 한 면접 방식에 단위들을 랜덤하게 배정할 수 있다) 다른 특징들은 (예를 들어, 인종) 미리 배정되어 있고 확률화되지 않았다. 사업체 조사 실험에서는 비록 크기가 랜덤하게 배정되지 않았을지라도 작은 사업체와 큰 사업체 사이의 응답률을 비교할 수 있다. 하지만, 우리가 관측한 차이가 순전히 크기 때문인지 또는 사업체의 크기와 관련될 수 있는 다른 요인들 때문인지는 절대 확신할 수 없다. 관측 연구에서는 관측된 차이의 원인을 실험의 요인들에 돌릴 수 있는 통계적 명분이 있는 것은 아니다.

관측 연구의 분석 방법은 확률화된 실험에서의 방법들과 매우 비슷하다. 회귀분석, ANOVA, 범주형 자료분석, 상관분석이 사용될 수 있다. 하지만, 설계된 실험에서 확률화를 통해서 제거될 수 있는 다양한 “교란(disturbance)” 변수의 효과는 관측 연구에서는 존재한다. 따라서, 분석자에게는 모형화와 추정 과정에서 이런 “교란(confounding)” 효과를 명확하게 관리하는 것이 필수적이다.

예 8.4.1 대리응답과 자기응답의 차이에 대한 많은 연구들은 확률화 되지 않은 관측을 근거로 한다. 연구자가 어떤 표본 구성원이 대리응답 처리를 받고 어떤 표본 구성원이 자기응답 처리를 받는지를 결정하기 보다는 조사가 정상적으로 진행되고 응답의 종류가 (대리 또는 자기) 기록된다. 조사에서 대리로 응답을 제공하는 사람의 특징은 자기응답을 제공하는 사람과 상당히 다를 수 있다. 하지만 그것들이 다른 이유를 말하기는 어렵다. 예를 들어, 대리응답자의 평균 가족 수는 보통 자기응답자의 가족 수 보다 많다. 또한, 자기응답자와 비교해서 대리응답자는 불균형하게 기혼이고, 나이가 더 많고, 여성이다. 그러므로, 대리와 자기응답자의 인구통계학적 차이로 인해 그들의 응답의 방식의 차이가 설명될 수 있다.

결혼 여부, 나이, 성별 같은 변수들은 때로는 교란 변수라고 부르고 응답자 규칙의 추정값 효과는 이 교란 변수들과 “교란” 되었다고 말한다. 이 경우에는 응답자

방식의 효과의 추정값은 편향될 것이고 이것을 때로는 선택 편향(selection bias) 이라고 부른다. 만일 연구 설계가 완전히 확률화 되었으면 (즉, 만일 응답자들이 연구자에 의해 각 종류의 응답자 방식에 랜덤하게 배정되었으면), 그 두 집단의 인구통계는 불완전한 랜덤 배정에 기인하는 변동성을 제외하고는 대략 같을 것이다.

관측 연구에서 연구자는 두 집단의 남성, 기혼자, 나이가 많은 사람 등의 비율이 같아지도록 관측들에 가중치를 가해서 두 응답 방식 표본의 특징을 같게 되도록 할 수 있다. 연구에서 관측된 교란 변수의 이러한 조절의 목적은 비교가 더 이상 교란되지 않고 대리 대 자기응답의 순수한 효과가 추정될 수 있도록 방해할 수 있는 모든 변수들을 같게 하는 것이다. 가중치 조절이 교란 변수의 모든 효과를 없애지는 않을 것이기 때문에 약간의 선택 편향이 남을 수 있다. 따라서, 비확률화에 기인하는 편향이 아직도 있을 수 있으므로 관측 연구의 결과를 신중히 해석할 필요가 있다. Moore (1988)에 대리대 자기응답자들의 차이를 평가하려고 시도하는 몇 개의 관측 연구들의 비판적인 검토가 제공한다.

8.4.3 내부 일치성 연구

내부 일치성 분석은 자료 품질을 평가하기 위하여 흔히 사용된다. 이 분석에서는 같은 조사에서 수집된 강하게 상호상관된 것으로 알려진 두 개 이상의 변수들이 기대된 정도로 상관되었는지를 판단하기 위하여 그 변수들을 비교한다. 같은 특징이 여러 변수들에 의해 측정된다는 점에서 변수들이 중복된다고 할 수 있다. 하나의 예로, 소득에 대한 조사에서 모든 출처에서 나오는 총 가구 수입에 대한 질문을 먼저 할 수 있고 그 다음에는 특정한 출처에서 나오는 수입에 대한 자세한 질문을 할 수 있다. 소득을 수집하는 이러한 이중 전략은 전체에 대한 질문에서 나온 소득 보고를 개인 소득 보고의 합과 비교할 기회를 제공한다. 두 액수들 사이의 큰 불일치는 소득 자료의 품질의 문제를 나타낸다. 물론, 다른 정보가 없으면 어느 것이 더 정확한지 아는 방법이 없다.

다른 종류의 일치성 분석에서는 매우 다른 특징을 측정하지만 변수 사이의 관계가 잘 알려진 조사에서 수집된 변수들에 대한 자료를 모은다. 예를 들어, 사업체 조사에서 고용인에게 지불된 총이익 대 총소득의 비율이 대부분의 기업에서 25%와 45%이어야 한다는 것이 알려졌을 수 있다. 총이익 대 총소득의 비율이 이 범위를

벗어나는 사업체는 추가 조사의 근거를 주는 불일치를 나타내는 것으로 볼 수 있다.

내부 일치성 연구들은 임시(ad hoc)적이거나 계획됐을 수 있다. 만일 그것이 임의적이면, 응답자 부담을 줄이기 위하여 설문지 설계에서 보통 중복되는 변수를 피하기 때문에 분석에서 사용될 수 있는 변수가 몇 개 안될 수도 있다. 만일 자료 품질의 측정이 조사의 우선 사항이면, 내부 일치성 검사가 계획되고 조사 도구에 들어 있는 것이 이상적이다. 하지만, 자료 품질의 측정이 설계 단계에서 예상하지 않았더라도 대부분의 조사에서 어느 정도의 일치성 분석이 가능하다.

다양한 인구통계학적 집단들의 성비가 예상대로인지 판단하기 위해서 최소한 그 비율을 비교할 수 있다 (응답자의 성별을 조사에서 얻을 수 있다고 가정하고). 인구통계학적 집단에 따른 성비의 표는 가장 최근의 센서스 자료에서 입수할 수 있다. 만일 자료의 성비가 공식 숫자와 주목할 만하게 다르면, 이것은 표본의 편향의 증거가 될 수 있다. <표 8.4>에 1985년에 미국의 제한된 지역에서 실시된 RDD 조사에서 구강암과 pharyngeal 암 연구를 위한 이러한 종류의 분석 결과가 제시 된다. 분석은 조사에 흑인 남성의 대표성이 적은 경향을 제시한다.

<표 8.4> 18세 이상 흑인의 나이에 따른 성별 비율

응답자의 성별	응답자의 나이			합계
	18-29	30-44	45+	
여성	426	438	440	1304
남성	338	321	339	998
성비	1.26	1.36	1.30	1.31

출처: Maklan와 Waksberg (1988), 표 1.

그렇지만 내부 일치성 연구는 계획되고 중요 항목의 반복되는 측정이 조사 설문지에 들어 있을 때 더 유익하다. 예를 들어, 1999년 이전에는 미국 NHSDA가 약물 사용의 보고의 일치성을 검사하기 위하여 심층 중복 측정을 사용했다. 대마초 사용에 대하여 이렇게 질문을 했다. “당신이 마지막으로 대마초나 헤시시를 사용한지 얼마나 되었습니까?” 그리고는 설문지의 나중에 다음과 같은 질문을 했다. “이제 당신의 12개월 기준(reference) 날짜에서 오늘까지인 지난 12개월에 대하여 생각해 봐

라. 지난 12개월 중 몇일동안 대마초나 헤시시를 사용했습니까?” 만일 응답자가 뒤에 말한 질문에 대해 하루나 그 이상이라고 답했으면, 앞의 질문의 답이 대마초의 마지막 사용한지 12개월을 넘어서는 안 된다. 하지만, NHSDA 연구에서는 응답자의 약 1.5%가 이 질문에 불일치한 응답했다는 것을 보여 주었다. 지난해에 대마초를 사용한 미국 인구의 비율이 대략 8%이므로 1.5/8.0, 또는 대략 19%의 지난 한 해의 대마초 사용자들이 이 두 질문들을 불일치하게 답했을 가능성이 있다.

조사 설계자에 의해 내부 일치성 연구에서 나온 결과는 자료수집 과정에서의 문제로서 질문의 해석이나 표현에 대한 문제를 확인하는데 사용할 수 있다. NHSDA의 경우에 통계학자들은 두가지 약물에 대한 질문 응답을 지난 한 해동안 약물 사용의 확산을 추정하는데 사용한다. 예를 들어, 만일 응답자가 두 질문 중 어느 한 질문의 응답에서 약물의 사용을 나타내면 그 응답자는 확산에 대한 추정의 목적에서 약물 사용자로 분류된다.

그렇지만, 내부 일치성 연구는 자료를 에디팅에도 사용될 수 있다. 예를 들어, the Subcommittee on Measurement of Quality in Establishment Surveys(U.S. Federal Committee on Statistical Methodology, 1988)는 대부분의 연방 조사는 자료를 에디팅하고 응답오차를 관리하기 위하여 내부 일치성 분석이 사용되었음을 발견했다. 응답오차를 측정하는 간접적인 기법으로서 이 조사들의 약 3/4은 에디팅 실패율을 계산하고 약 반은 면접원 오류율을 계산한다. 측정오차가 조사 결과에 영향을 미치는 가능성에 대한 정보는 응답자와 면접원과의 결과보고 이외에 내부 일치성 비율의 분석에서 꾸준히 모을 수 있다. 이 자료들은 이 장에서 설명한 방법을 사용하여 더 평가되어야 하는 분야를 정하는데 큰 도움이 될 수 있다.

패널 조사에서 전 기간에 얻은 자료가 현재 시점에서 나온 자료와 일치하는지를 검사하기 위하여 내부 일치성 분석이 패널 조사에서 나온 자료에 자주 적용된다. 예를 들어, 매달 실시되는 노동력 조사에서 이번 달의 보고된 산업과 직업을 전 달의 보고와 비교하는 것은 일반적이다. 특히 개인의 산업 분류에서 매달 변화가 너무 많은 것은 산업과 직업 자료에 응답오차가 있다는 것을 나타낼 수 있다. 사실 미국 센서스국은 CPS의 산업과 직업 자료 수집에서 이런 종류의 불일치를 관측했다. 이것은 자료 수집 방법을 종속 면접 접근방법으로 변화를 시켰다.

종속 면접 접근방법(*dependent interviewing approach*)에서는 전달에 면접한 모든 응답자의 산업 보고를 CAPI의 도구에 미리 저장해 둔다(Brown 등, 1998). 그

서 응답자가 이번 달의 보고를 제공하면, 그 전달의 보고와 대조하고 산업이 실제로 바뀌었다는 것을 확인하기 위하여 면접원이 두 보고들 사이의 불일치를 즉시 검증한다. 예를 들어, “지난 면접 이후로 직업을 바꿨습니까?”라고 면접원이 묻는다. 미국 농무부에서도 비슷한 접근방법이 사용되는데, 여기에서는 봄에 실시하는 농업 조사에서 농부가 보고한 경작면적을 그 농부의 직전 가을 조사의 응답과 대조한다.

8.4.4 외부 타당성 검사

외부 타당성 검사는 조사 추정값을 평가하는 과정인데 조사의 추정값을 더 정확하다고 생각되는 외부 추정값과 비교한다. 외부 추정값(*external estimates*)은 가장 최근의 센서스 또는 평가될 특징에 대한 절대기준(*gold standard*)으로 생각되는 조사에서 나온 추정값일 수 있다. 그것은 또한 매우 정확하다고 여겨지는 행정기록에서 나온 추정값일 수도 있다. 예를 들어, 소득 조사의 추정값을 평가하기 위한 세무 당국에서 나온 소득 추정값이다. 따라서, 외부 추정값이 정확하다면 비교가 되는 추정값을 얻는데 사용된 조사의 편향은 평가될 수 있다.

한 예로, 미국에서는 10년주기의 센서스와 CPS를 여러 조사 추정값의 편향의 평가에 기준이 되는 매우 정확한 조사로 여긴다. 예를 들어, 전국적인 조사에서 성별 비율, 나이와 인종 집단의 분포, 그리고 많은 다른 인구통계학적 특징의 조사 추정값에 대한 조사 표본의 왜곡을 평가하기 위하여 이 추정값들을 센서스나 CPS의 대응되는 추정값들과 비교할 수 있다. 성인의 흡연 행동에 대한 RDD 연구의 결과가 <표 8.5>에 나와 있다. 이 표에서는 같은 기간의 RDD 조사와 CPS의 모든 나이의 사람을 포함하는 가구의 크기를 비교한다. 자료는 RDD조사가 1인 가구와 4+인 가구를 과소 표시하고 중앙값 크기인 2.7명 부근의 가구를 과대 표시하는 경향이 있다는 것을 보여준다. 앞의 결과는 특히 전화 조사에서 1인 가구의 비접촉으로 인한 단위 무응답 때문이고 뒤에 말한 결과는 큰 가구에 대한 전화 가구 등록부 절차의 잠재적인 문제점 때문이라고 추측할 수 있다. 하지만 표의 근본 원인들을 이해하고 이 자료들이 나타낸 문제들에 대처하려면 자세한 검사가 필요할 것이다.

<표 8.5> RDD 조사와 1986년 3월 CPS의 가구 식구 수 비교

식구 수	RDD(%)	CPD(%)
1인	20.9	22.7
2인	34.1	31.7
3인	18.6	18.1
4+인	26.4	27.5

출처: Maklan과 Waksberg (1988), <표 5>.

8.4.5 행정 기록 검사

이 절에서는 조사응답을 어떤 종류의 행정기록에서 얻은 개인 단위의 값과 비교하여 조사에서 나온 개인 응답을 평가하는 방법을 논의한다. 세계적으로 많은 나라에서 기록과 인구 등록부에 있는 자료는 평가 목적에서 뿐만 아니라, 모집단 추정을 위해서 연구자들이 사용할 수 있다. 그 경우에 조사 프로그램의 일부 추정값은 행정기록 자료만을 근거로 했을 것이다. 만일 조사에서 관심대상인 특징을 행정기록에서 얻을 수 있으면 조사를 실시할 필요조차도 없을 것이다.

이 경우에는 자료를 새로 수집할 필요가 없기 때문에 행정기록의 사용은 상당히 경제적일 것이다. 불행히도, 조사에서 관심대상인 많은 특징은 쉽게 이용하거나 얻을 수 있지 않거나 존재하지 않을 수 있다. 그것이 존재할 때, 평가 목적으로 행정기록 자료를 얻는 비용이 상당할 수도 있는데 이것은 표본에서 특정 개인에 대한 기록을 찾는 어려움과 평가될 조사 변수와 비교될 변수를 기록에서 만들어내는 어려움 때문이다. 그 결과로, 기록 검사 평가 연구에 대한 조사 방법론 논문은 드물다. 그럼에도 불구하고, 그것은 조사오차를 평가하는데 매우 중요한 방법이다.

이 기법에서는 어떤 특징(예를 들어, 나이)에 대한 조사 단위의 조사 응답을 행정기록(예를 들어, 출생 증명서)에서 얻은 이 특징에 대응하는 값들과 비교한다. 행정기록의 값이 가장 정확한 또는 절대기준값(*gold standard value*)이라고 가정한다. 따라서 조사 응답과 행정기록 값의 차이는 조사 응답의 오류에 기인한다. 이렇게 하여, 특징에 대한 조사 추정값에 있는 편향을 추정할 수 있다. 조사 평가에 사용된 행정기록에는 연방/주 소득세 또는 판매세 보고, 출생 증명서, 면허 정보, 인구와 정부 복지 등록부, 경찰 기록, 그리고 다른 특수 목적의 기록이 있다.

기록 검사 연구의 한 가지 방법은 표본추출틀에서 조사 단위의 표본을 뽑고 그 다음에 이 단위에 대응하는 행정 기록을 찾아낸다. 예를 들어, 연구자가 먼저 가구의 표본을 뽑고 조사에서 관심대상인 특징에 대한 정보를 수집하고 조사와 기록의 값을 비교하기 위하여 적절한 행정기록 시스템에서 응답자에 대한 정보를 찾는다. 이런 종류의 연구를 전진 기록 점검 연구(*forward record check study*)라고 한다. 반대로, 연구자는 기록의 표본에서 시작하여 대응하는 가구를 면접하여 기록에 있는 정보를 얻을 수 있다. 역 기록 점검 연구(*reverse record check study*)라고 하는 이러한 연구 설계는 보통 더 효율적인데 드문 항목의 추정값의 정확도를 더 잘 관리할 수 있다. 예를 들어, 만일 연구 설계가 소득이 하위 5%인 500명을 요구하면 500명이 선택되는 것을 보장하기 위하여 매우 큰 가구 표본이 필요할 것이다. 하지만, 역 기록 점검 표본으로 정확한 숫자의 가구들을 소득 자료가 들어있는 행정 기록에서 선택할 수 있다.

이미 언급했듯이, 평가 목적을 위한 행정 기록의 사용에서의 중요한 가정은 그 기록은 관심대상인 조사 특징에 대하여 본질적으로 오류가 없는 정보(즉, 참 값들)를 포함한다는 것이다. 이러한 가정 아래에서 측정 편향의 추정값은 조사 자료에서 추정값과 기록 자료에서 대응하는 추정값의 차이이다. 하지만, 이러한 방법은 다음의 네 가지 문제점이 있을 수 있고 이는 기록 자료의 유용성을 제한한다.

- 기록 자료와 조사 자료의 기간이 일치하지 않을 수 있다.
- 기록 체계에서 보고되는 특징이 조사에서 측정되는 특징과 똑같지 않을 수 있다.
- 평가 조사비용을 절약하기 위하여 기록 연구는 매우 제한된 지리적 지역으로 한정되고, 이 제한된 모집단을 벗어나는 추론은 실효성이 없을 수 있다.
- 행정기록은 오류가 발생하기 쉽고, 때로는 오류가 상당히 많을 수도 있다.

사생활 제한으로 행정 기록에 대한 접근이 제한될 수 있다. 따라서, 평가를 위해 적절한 기록이 있어도 평가를 실행할 수 없을 수 있다. 행정 기록도 비표본추출오차가 있을 수 있다. 하지만 오차가 평가되는 자료에 있는 편향에 비교해서 작다면 기록은 평가 목적으로 여전히 유용하다. 즉, 기록은 “금본위”와 같은 절대기준이라기 보다는 “은본위” 정도의 표준을 제공한다. 그렇지만, 정확도가 평가 목적으로 사

용되기에 부적당한 공식 통계의 예들이 있다. 미국 Office of Management and Budget에서 실시한 연구는(U.S. Federal Committee on Statistical Methodology, 1980) 기록 점검을 실시하기 위하여 방법론적인 필요 조건에 대한 안내를 제공한다.

전화와 대면면접으로 소득을 수집하는 것과 관련된 편향을 비교하기 위한 행정기록의 사용의 예가 Kormendi(1988)에 제공되었다. 덴마크에서 실시된 조사에서 Kormendi는 두가지 면접 방법이 체계적으로 어느 정도로 다른 답을 가져오게 되는지 조사하기 위하여 세무 당국으로 부터 전화와 대면조사의 응답자들에 대한 총수입과 순수입에 대한 정보를 얻었다. <표 8.6>은 그러한 비교를 보여준다. 세무 당국의 액수를 절대기준이라고 간주하고, 조사 추정값과 세무 당국의 값의 차이는 조사 추정값에 있는 편향의 추정값이다. 표에 있는 자료는 전화조사 (-4000kr)에서의 여성과 대면조사(7000kr)에서의 남성의 소득보고에는 통계적으로 유의한 편향이 있다.

<표 8.6> 전화와 대면조사의 총수입의 추정값과 세무 당국의 소득 정보의 비교

면접 방식	조사추정값 (1000kr)	세무 당국의 값 (1000kr)	차이 (1000kr)
전화	106	107	-1
남성	138	137	2
여성	70	74	-4**
대면	108	107	1
남성	138	131	7*
여성	78	81	-3

출처: Kormendi (1988).

* 5%수준에서 유의함.

** 1%수준에서 유의함.

8.4.6 재면접 연구

조사의 측정 오류의 요소를 추정하기 위해 가장 널리 사용되는 방법 중 하나는 재면접 연구이다. 재면접 조사(reinterview survey)에서는 최초의 조사 표본의 응답자를 재방문하여 최초의 조사에서 물었던 같은 질문을 다시 묻는다. 재면접의 목

적은 최초의 조사에서와 같은 특징에 대한 정보를 얻는 것이므로 재면접 질문은 첫 면접에서 기준했던 같은 시점과 기간이 기준이 되도록 설계되었다. 재면접에서의 오류에 대한 다양한 가정하에서 첫 응답과 관련된 오류의 요소가 추정될 것이다.

<표 8.7> 절대기준과 반복 재면접 조사 설계의 비교

재면접 종류	설계의 목적	분석의 목적
절대기준		
조정된 재면접, 원인의 판단이 없음	매우 정확한 응답들을 얻기 위하여	측정 편향의 추정
조정된 재면접, 원인의 판단 탐색조사 또는 확대 재면접 방법들	오류의 원인을 판단하기 위하여 참값을 추론하는데 사용될 수 있는 자세한 정보를 얻기 위하여	면접원 업무수행; 조사 절차의 평가 측정 편향의 추정; 조사 절차의 평가
반복 재면접		
시험-재시험 재면접	최초의 면접 과정을 반복하기 위하여	단순응답분산과 신뢰도의 추정
반복 측정	다른 방법들을 사용해서 최초의 면접 특징의 추가 측정을 얻기 위하여	latent class 분석 같은 모형에 근거한 방법을 사용해서 측정 분산 및/또는 측정 편향의 추정

<표 8.7>에서 보여주듯이 두 가지 일반적인 재면접이 있다. 그것은 절대기준과 반복 재면접이다. 이름이 뜻하듯이 절대기준 재면접(*gold standard reinterview*)은 측정 편향을 추정하는 목적으로 또는 조사 방법을 평가하기 위하여 매우 정확하다고 여길 수 있는 정보의 수집을 시도한다. 즉, 진실을 측정하려고 노력한다. 반복 재면접(*replicated reinterview*)에는 두 종류가 있다. 그것은 시험-재시험 재면접 또는 단순히 되풀이된 측정 재면접이다. 시험-재시험 재면접(*test-retest type of reinterview*)은 최초의 조사 과정을 반복한다. 즉, 최초의 조사 측정에서와 같은 비표본추출오차의 조건아래에서 두 번째 측정을 얻는 것이다. 시험-재시험 재면접의 목적은 단순 응답분산과 흔히 시험-재시험 신뢰도(*test-retest reliability*) (2장 참조)이라고 부르는 응답 신뢰도를 추정하는 것이다. 반복 측정 재면접(*repeated measurement reinterview*) 설계의 유일한 필요조건은 두 번째 측정의 비표본추출 오차가 최초 조사 측정의 비표본추출오차와 상관되지 않아야 한다는 것이다. 이러

한 종류의 재면접이 모든 재면접 설계 중 가장 일반적인 것이다. 예를 들어, 첫 번째 측정은 우편조사에서 나오고, 두 번째 측정은 우편조사 응답자의 전화 재면접 조사에서 나올 수 있다. 상이한 자료수집 방식 때문에 두 번째 측정은 첫 번째의 반복도 아니고 절대기준 측정도 아니다. 이 재면접 설계를 이제 자세하게 논의한다.

절대기준 재면접 연구

행정기록의 부재, 고비용, 사생활 문제, 또는 다른 이유 때문에 행정기록에 접근하는 것이 실행 불가능하면 조사 변수의 참값을 얻으려는 목적의 재면접 접근방법이 실현 가능한 대안일 것이다. 보통 이런 종류의 재면접은 매우 정확한 자료를 얻기 위해서 면접 기간 중에 더 많은 시간을 들이고 보고 업무에 더 큰 주의를 가질 수 있게 하기 위하여 적은 수의 “중요한” 조사 특징에 한정된다.

절대기준 재면접(*gold standard reinterview*)은 측정오차 편향의 추정을 목적으로 조사 항목의 일부분에 대한 매우 정확한 (즉, 본질적으로 오류가 없는) 측정을 얻는 것을 겨냥하는 재면접이다.

절대기준 재면접의 한 형태는 조정 면접이다. 전형적인 조정된 재면접 조사(*reconciled reinterview survey*)에서는 몇 일이나 몇 주전에 처음 면접했던 응답자의 부표본을 감독자나 선배 면접원인 재면접원이 다시 연락한다. 재면접원은 응답자에게 최초의 질문의 일부분을 묻고 그 응답을 최초의 면접의 응답과 비교한다. 만일 응답이 모두 일치하면 재면접은 완료된다. 만일 한 개이상의 응답이 일치하지 않으면 재면접원이 응답자에게 불일치를 알리고 질문에 가장 올바른 답을 판단하도록 지시한다. 이 과정을 조정(*reconciliation*)이라고 한다. 조정된 응답은 최초의 응답, 재면접 응답, 또는 응답자가 (또는 재면접원이) 가장 정확하다고 생각하는 완전히 새 응답일 것이다. 조정 과정에 뒤이어, 재면접원에게 불일치의 주된 원인을 판단하고 직면한 각 불일치의 이유들을 기록하도록 지시할 수 있다. 이 정보는 측정오차의 원인을 이해하는데 상당히 유용할 것이다.

<표 8.8>은 이항 변수(즉, 0이나 1의 값을 취하는 변수)에 대한 전형적인 면접-재면접표를 보여준다. 예를 들어, “1”은 “노동 인구에 포함됨”을 나타내고 “0”은 “노

동 인구에 포함되지 않음”을 나타낸다. 표에서는 면접과 재면접 모두에서 답한 응답자를 면접 응답(행)과 조정된 재면접 응답(열)으로 분류한다. 그러므로, 표에 있는 a 는 면접과 재면접 분류가 모두 “1”인 사람 수를 나타낸다. 마찬가지로, 표에 있는 d 는 면접과 재면접 분류가 모두 “0”인 사람 수를 나타낸다. 면접과 재면접 응답이 다른 사람의 수는 $b+c$ 이다. 따라서, 재면접 분석에서 면접과 재면접 분류가 일치하지 않는 사람들의 비율은

$$g = \frac{b+c}{n}$$

으로 주어지는데 이것을 총 차이의 비율(*gross difference rate*) 또는 불일치율(*disagreement rate*)이라고 부른다. 이 통계량은 절대기준 재면접에서는 실제적으로 중요하지 않지만 아래에서 설명하는 바와 같이 조정된 면접에서는 중요한 측정이다.

<표 8.8> 이항 변수에 대한 기초적인 면접-재면접표

재면접 응답	면접 응답		
	1	0	
1	a	b	$a+b$
0	c	d	$c+d$
	$a+c$	$b+d$	$n = a+b+c+d$

절대기준 재면접들에서 중요한 통계량은 최초의 면접을 근거로 한 모집단 비율의 추정값과 조정된 재면접을 근거로 한 같은 비율의 추정값의 차이인 순 차이 비율(*net difference rate; ndr*)이다. <표 8.8>에서 면접에서 1번 범주로 분류된 표본의 비율이 $p_1 = (a+b)/n$ 이라는 것에 주목하라. 재면접에서는 이 비율이 $p_2 = (a+c)/n$ 이다. 만일 우리가 이 두 비율을 모집단 비율의 추정값으로 받아들이면, p_1 은 최초의 면접의 측정 편향을 가지고 있는 추정값인데 반하여 절대기준 가정에 의하면 p_2 는 측정오차에 의해 편향되지 않았다. 따라서, 차이 $p_1 - p_2$ 는 최초의 조사 추정의 편향된 추정값이다. 따라서

$$ndr = p_1 - p_2 = \frac{b-c}{n}$$

이라는 것이 쉽게 입증된다.

하나의 예로, 1981년-1990년 CPS 재면접 프로그램에서 나온 <표 8.9>에 있는 자료를 생각해보자. 이 표에서 $b=3604$, $c=1224$, $n=373,464$ 이다. 위의 ndr 공식을 사용하면 최초의 분류에서의 편향은 $ndr=-0.00637$ 또는 -0.637% 인 것을 알 수 있다. 노동 인구 응답의 편향은 또한 편향을 최초의 면접에서의 모집단 비율의 추정값으로 나누어서 상대편향으로 (2장 참조) 나타낼 수 있다. 즉,

$$\text{상대편향} = \frac{ndr}{p_1}$$

인데 <표 8.9>에서는 -0.0102 또는 -1.02% 이다.

<표 8.9> 1981년-1990년 CPS 조정된 재면접 자료의 면접-재면접 표

재면접 응답	면접 응답		
	노동 인구에 포함됨	노동 인구에 포함되지 않음	
노동 인구에 포함됨	230,559	3,604	234,163
노동 인구에 포함되지 않음	<u>1,224</u>	<u>138,077</u>	139,301
	231,783	141,681	

출처: Sinclair와 Gastwirth (1998).

일부 설계에서는 조정된 재면접이 최초의 조사 면접원의 업무수행을 평가하는데 사용되고 있다. 그 경우에는 재면접원에게 두 응답들 사이의 불일치가 최초의 응답자, 현재의 응답자, 최초의 면접원, 재면접원, 또는 어떤 다른 요인에 의해 만들어진 오류의 결과인지를 판단하라고 할 수 있다. 최초의 면접원의 잘못으로 결정된 오류는 그 다음에는 어떻게 그 오류가 발생했는지 그리고 미래 면접에서 이런 오류를 피하기 위하여 면접원이 취할 수 있는 조치가 무엇인지 알아보려고 면접원과 논의를 한다. 설문지나 응답자 중 한 명이 원인이 되는 오류는 무시할 수 있다.

미국 센서스국은 조정된 재면접을 사용해서 면접원 업무수행과 설문지 업무수행을 평가하고 응답 편향을 추정했다. 하지만, 몇 개의 문헌에서 조정된 재면접의 접근방법으로 충분히 정확하지 않은 추정값이 제공 될 수 있다는 증거를 제공했다. Biemer와 Forsman (1992)은 조정된 재면접의 자료가 최초의 면접의 자료처럼 틀릴 수 있다는 증거를 제공한다.

그들이 고려한 재면접 체계는 편향의 추정과 면접원 업무 수행의 평가 목적을 하나의 재면접으로 결합시킨 1980년대에 미국 센서스국에서 사용한 것이다. 분명히 재면접을 실시한 현지 감독자는 편향 추정을 위한 응답자 특징의 “참”값을 얻는 것보다는 현지 면접원의 업무 수행의 객관적인 평가를 제공하는데 더 관심이 있었다. 따라서, 최초의 면접원의 잘못이 아닌 것으로 결정된 일부 불일치들이 보고되지 않았다. 하지만, 설계에 대한 적절한 처리를 통하여 조정된 재면접 접근 방법은 어떤 특징에 대한 측정 편향의 좋은 측정을 제공할 수 있다. Forsman과 Schreiner(1991)는 조정된 재면접 접근 방법의 추가적인 장점과 단점을 다루었다.

1950년 센서스 이후로 미국 센서스국은 센서스 사후확인조사(PES)에서 한 종류의 조정된 재면접을 더 완전한 가구 명부를 얻는 수단으로 사용했다. PES는 최초의 센서스 총계에서 누락되거나 잘못 포함된 사람을 조정하기 위한 목적으로 센서스에 있는 표본 지역에 거주하는 사람의 또하나의 두 번째 총수를 제공한다. PES 명부를 얻기 위하여 가장 흔히 사용되는 절차는 첫째, 센서스 실시일(4월 1일)에 그 주소에 살았던 모든 사람의 새 명부를 얻고, 이 사람들의 목록을 센서스 명부에 보고된 사람들과 비교 조정하는 것이다.

Fecso와 Pafford(1988)는 농업 조사의 편향을 측정하는데 사용된 재면접 연구를 제시한다. 미국 NASS는 이전에 해왔던 대면 면접 대신 CATI를 사용해서 가축과 작물 목록에 대한 자료를 수집하기 시작했었다. 자료의 상세한 성질과 중앙집중식 주 전화 조사원이 농장 용어를 잘 알지 못하기 때문에 전화로 정확한 응답을 얻는 것이 문제점으로 생각되었다.

새로운 CATI 자료 수집 과정의 편향을 평가하기 위해서 NASS는 대면 면접 방법을 사용해서 CATI 응답자와 조정된 재면접들을 실시했다. 조사 절차와 내용에 대한 지식이 있고 조정 과정 동안에 최선의 응답을 얻기 위해 탐색 조사를 가장 잘 할 수 있는 경험이 있는 현지 감독 조사원들이 대면 재면접에 사용되었다. 연구에서 대략 1000개 농장의 운영을 재면접했고 그 결과인 곡물 저장 품목들(옥수수과 콩 저장량)은 다음 표들에 나와있다.

<표 8.10>은 참값의 퍼센트로 나타낸 상대 편향의 조정된 재면접 추정값들을 보여준다. 이 자료들의 편향은 한가지 항목(인디애나의 콩 저장물)을 제외하고는 모두 통계적으로 유의했다. 편향의 방향은 CATI 자료수집 방식이 옥수수와 콩의 저장량을 과소 평가하는 경향이 있다는 것을 나타낸다. 조정의 과정에서 차이의 원인이

수집됐다. <표 8.11>은 압도적인 41.1%의 차이의 원인이 단순한 응답 분산(랜덤 변동) 때문이 아닌 정의상의 문제들(편향과 관련된 불일치)과 관련이 있다는 것을 나타낸다. 이런 정의상 문제의 예에는 대여한 저장소의 불포함, 정부의 예비 곡물의 보고에 대한 혼동, 다른 사람의 소유물인 곡물을 포함하지 않는 것, 그리고 부속 농장에 있는 저장소를 잘못 포함하는 것이 있다.

<표 8.10> 대면 재면접을 절대기준으로 사용한 CATI 조사 추정값의 상대 편향 퍼센트의 추정값

주	옥수수 저장량	콩 저장량
미네소타	10.4*	14.9*
인디애나	17.9*	5.9
오하이오	12.0*	13.7*

출처: Fecso와 Pafford (1988).

* CATI와 최종 조정된 응답이 $\alpha=0.05$ 에서 유의적인 차이가 있음.

<표 8.11> 미네소타의 옥수수 저장량에 대한 CATI와 재면접 응답 차이의 원인

이유	수	비율(%)
추정/어림수	28	31.1
“정의상”	37	41.1
기타	25	27.8
합계	90	100.0

출처: Fecso와 Pafford (1988).

다른 형태의 절대기준 재면접에서는 최초의 조사 측정 보다 훨씬 더 정확한 측정을 얻기 위하여 조정보다는 면밀한 탐색조사 접근방법이 사용된다. 예를 들어, 1980년 미국 센서스에서, 센서스 설문지 항목에서 선택된 것의 편향을 평가하는데 이러한 집중적인 면접 접근방법을 사용한 절대기준 재면접 조사를 실시했다. 예를 들어, 응답자의 라틴 아메리카계 또는 스페인 출신 또는 더 정확한 가계 혈통을 판단하기 위하여, 양쪽 부모에 대하여 두 세대 위의 조상에 대한 자세한 질문들을 했다. 그러면, 정보처리 단계 중에 응답자를 아메리카계/스페인 출신으로 분류하는데 이 정보가 사용된다.

반복 재면접

이미 언급했듯이 반복 재면접에는 시험-재시험 재면접과 반복 측정의 두가지 종류가 있다. 이 두 설계의 차이는 다음과 같다. 시험-재시험 재면접 설계는 재면접에서 응답 과정을 반복하므로 재면접을 어떻게 실시하는지에 대한 엄격한 규칙을 지켜야 한다. 반복 측정 재면접 설계는 명칭이 뜻하듯이 최초의 조사 과정을 엄격히 반복하려고 노력하지 않고, 다만 응답자의 표본에 대하여 조사를 반복한다. 반복 측정 설계는 측정 편향 이외에도 측정 분산(즉, 신뢰도)을 추정하는데 유용할 것이다. 하지만, 이런 종류의 재면접에서 얻은 비표본추출오차 요소의 정값은 통계 모형과 잠재 변수(latent variable)의 가정에 따라 결정된다. 그렇지만, 이러한 복잡한 모형 접근방법에 대한 고려는 이 책의 범위 밖이다. 따라서, 우리는 반복 재면접에 대한 우리의 논의를 시험-재시험 재면접으로 제한한다.

시험-재시험 재면접(*test-retest reinterview*)은 최초의 조사 과정을 반복하는 재면접이다. 최초의 조사와 같은 조사 조건 아래에서 최초의 조사 응답자로 부터 새로운 응답을 얻는 것이 이러한 재면접의 목적이다.

재면접의 목적이 측정 편향보다 측정 신뢰도를 추정하는 것이면, 시험-재시험 재면접 설계가 필요하다. 신뢰도의 추정량이 타당하려면 응답자가 재면접에서 어떤 질문에 대한 조사 응답에 이르기 위해 사용하는 응답과정은 그가 그 질문에 대해 최초의 면접에서 사용했던 응답과정과 근본적으로 똑같아야 한다. 이것은 면접과정 중에 발생하였을 수 있었던 가변오차가 재면접에서 발생할 가망성이 같아야 한다는 것이다. 이렇게 하여, 최초의 조사 응답과 관련된 측정오차 분산은 최초와 재면접 응답들 사이의 변동으로 추정할 수 있다.

그러므로, 시험-재시험 재면접 설계의 목적은 최초와 재면접 면접에서 사용된 응답 과정이 같다고 가정할 수 있도록 최초의 조사 상황을 재현하려고 노력하는 것이다. 이런 이유로, 재면접은 보통 다음의 다섯 가지 설계 원칙들에 따라 설계한다

1. 재면접원은 최초의 면접원과 같은 노동 집단에서 선택되어야 한다.

2. 재면접 절차는 최초의 면접 절차와 가능한 거의 똑같아야 한다.
3. 가능하면 같은 조사 질문을 사용해야 한다. 재면접 기준기간이 최초의 면접 기준기간과 맞도록 조절하기 위해서 약간의 변경이 필요할 수 있다. 그러나 조사 질문의 변경은 되도록 적게 해야 한다.
4. 재면접 응답자는 최초의 면접 응답자와 같아야 한다.
5. 면접과 재면접 사이의 기간은 기억오차와 신뢰도 비율이 추정될 특징의 실제의 변화를 되도록 적게 할 수 있을 정도로 짧아야 한다. 반면에 응답자가 단지 최초의 면접 응답을 기억하여 재면접에서 최초의 조사 응답의 응답과정을 되풀이하지 않고 최초 응답을 반복하는 상황을 피할 수 있을 정도로 길어야 한다. 이 기간은 보통 5일에서 10일 사이로 예상된다.

이런 설계 원칙이 잘 지켜지더라도 재면접에서 최초 조사 상황이 정확히 반복되지 않을 수 있다. 예를 들어, 최초 면접이 응답자에게 어떤 영향을 미쳤을 수 있고 후에 다시 물어보면 질문에 대한 응답에 영향을 줄 것이다. 문헌에서는 이것을 조절응답효과(*conditioning effects*)라고 부른다. 예를 들어, 응답자에게 그가 지난주에 일한 시간을 물었다고 가정하자. 면접에서 이 숫자를 다소 정확하지 않게 보고했을 수 있다. 하지만, 면접 후에 응답자가 호기심에서 그 숫자를 확인하였고 재면접에서는 정정한 숫자를 보고한다. 분명히 응답 과정은 최초 면접에 의해 바뀌었다.

같은 예에서, 최초 면접의 약 3주 후에 재면접을 실시했다고 가정하자. 재면접에서도 기준 주가 같아야 하기 때문에 응답자는 최초 면접에서와 같은 지난 주가 아니라 4주 전에 그가 몇 시간 일을 했는지를 이제 생각해내야 한다. 그러므로, 재면접 응답은 최초 면접보다 더 많은 기억오차가 일어나기 쉬울 것이다. 다른 한편으로는, 만일 재면접원이 재면접을 최초 면접을 실시한 다음 날에 실시하면, 응답자가 그 전 주에 대해 생각을 하여 첫 면접 응답과는 관계없이 답을 찾으려고 노력하기 보다는 단지 전날의 답을 기억하여 그것을 되풀이해서 말할 가능성이 높다. 이런 현상이 일어나면 재면접을 근거로 한 응답은 사실보다 신뢰성이 더 높은 것으로 보인다.

많은 다른 요인들이 최초 면접의 응답 과정을 정확하게 반복해야 한다는 가정을 지키기 어렵게 한다. 예를 들어, 비용과 응답자 부담을 덜기 위하여 최초 설문지의

일부분만이 사용될 수 있는데 따라서 재면접은 최초 면접보다 훨씬 더 짧아진다. 만일 최초 조사를 대면 면접으로 실시했으면 재면접은 전화 면접으로 실시할 수 있는데 이것은 응답 과정을 바꿀 수 있다. 이러한 재면접과 다른 재면접의 논점들이 Forsman과 Schrenier(1991)에 더 상세하게 논의되었다. 그것의 한계에도 불구하고, 시험-재시험 재면접은 최초 조사 과정에 가까이 접근할 수 있고, 측정 신뢰도의 추정값은 평가 목적에서 상당히 유용할 것이다.

2장에서 설명했던 신뢰도비 R 을 시험-재시험 재면접의 자료를 사용해서 추정하는 몇 가지 방법이 있다. 연속형 자료에서 R 을 추정하는 한 가지 쉬운 방법은 간단히 최초 면접과 재면접 응답의 상관 관계를 계산하는 것이다. 이를테면 만일 소득과 같은 어떤 특징에 대한 최초와 재면접 조사 응답 사이의 상관계수가 0.80이면, 우리는 소득 변수의 신뢰도가 80%라고 말한다.

범주형 자료(즉, 이산형의 응답항목이 있는 질문에서 얻은 자료)에 대해서는 방법이 좀 더 복잡하다. 미국 센서스국에서 사용된 방법은 1에서 신뢰도비를 빼, 즉 $1-R$ 인 불일치 지수(*index of inconsistency*)를 계산하는 것이다. 신뢰도비는 모집단의 특징의 참값의 분산과 관측된 값의 총 분산의 비율인데 반해서 불일치 지수는 관측들의 총오차 분산과 총 분산의 비율이다. 그러므로, 불일치 지수는 총오차에 대한 측정오차 분산 비율의 측정이다.

불일치 지수의 공식은 다음과 같다.

$$I = \frac{g}{p_1(1-p_2) + p_2(1-p_1)}$$

한 예로, <표 8.12>의 CPS 자료를 생각해보자. 이 자료는 1996년 CPS 재면접 프로그램의 시험-재시험 재면접 부분에서 나온 것이다. 지수는 다음과 같이 계산된다.

$$g = \frac{157 + 213}{7647} = 0.0484$$

$$p_1 = \frac{4877}{7647} = 0.638$$

$$p_2 = \frac{4821}{7647} = 0.630$$

$$p_1(1-p_2) = 0.231$$

$$p_2(1-p_1) = 0.233$$

$$I = \frac{0.0484}{0.231 + 0.233} = 0.104$$

이 계산에서 I 의 값은 0.104 또는 10.4%이다. 이것은 $(1-0.104)=0.896$ 또는 89.6%의 신뢰도비 R 과 동등하다. 미국 센서스국은 불일치 지수나 신뢰도비의 수용성을 판단하는 경험에 바탕을 둔 방법을 개발했다. <그림 8.2>에서 보여주듯이 대략 10%(90% 신뢰도)의 불일치 지수는 좋다고 여겨진다.

<표 8.12> CPS 조정하지 않은 재면접 자료의 면접-재면접 표

재면접 응답	면접 응답		
	노동 인구에 포함됨	노동 인구에 포함되지 않음	
노동 인구에 포함됨	4664	157	4821
노동 인구에 포함되지 않음	213	2613	2826
	4877	2770	7647

<p>좋음(Good): $I \leq 0.20$ 또는 $R \geq 0.80$</p> <p>보통(Fair): $0.20 < I < 0.50$ 또는 $0.50 < R < 0.80$</p> <p>나쁨(Poor): $I \geq 0.50, R \leq 0.50$</p>
--

<그림 8.2> 응답 불일치와 신뢰도의 수용성을 판단하는 경험에 바탕을 둔 방법.

마지막으로, 시험-재시험 재면접 조사 분석에서 자주 보고되는 자료 품질의 또 다른 측정은 일치율(*agreement rate*) 또는 재면접 분류와 그들의 최초 면접 분류가 일치하는 사람의 비율이다. <표 8.8?>의 표기법에 의하면 일치율은 $A = (a + d)/n$ 또는 1에서 총 차이의 비율 또는 불일치율 g 를 뺀 것이다. <표 8.12?>의 자료에서 일치율은 $A = (4664 + 2613)/7647 = 0.952$ 또는 95.2%이다. 그러므로 재면접 응답자의 약 95%가 그들의 면접 분류와 일치했고 이것은 이 자료의 좋은 일치라고 볼 수 있다.

8.5 평가 방법의 요약

이 장에서 우리는 조사에서 오류를 평가하는데 흔히 사용되는 몇 가지 방법을 설명했고 그것이 어떻게 조사 과정의 다양한 단계에서 적용될 수 있는지 알아보았다. 설계와 사전검사 단계에 적용되는 방법을 통하여 설계자는 주어진 예산내에서 총조사오차를 줄이도록 조사 설계를 최적화 할 수 있다. 자료 수집과 정보처리에 동시에 적용되는 방법들은 자료의 품질을 모니터링하고 조사 과정에서 중대한 오류가 발생하면 경고를 울리는데 사용될 수 있다. 마지막으로, 조사가 완료된 후에 적용되는 방법은 사용자에게 오류를 설명하고, 조사 설계자에게 앞으로의 조사를 개선하기 위한 정보를 제공한다. 따라서, 품질 평가는 설계단계에서 부터 사후조사 분석단계까지 수행되는 지속적인 과정으로 볼 수 있다.

품질 평가가 없다면 조사 품질 개선 과정은 순전히 추측이기 때문에 품질 평가는 조사연구 분야의 중요한 부분이다. 조사방법론에 대해 우리가 알고 있는 많은 것들이 조사품질 평가의 결과를 근거로 한 것이기 때문에 이 책의 초반은 조사 평가 없이 쓰여질 수가 없었다.

대부분의 조사에서 자료품질 평가는 설문지와 자료수집 방법에 대한 사전 검사와 면접과 정보처리 실행을 위한 일부 품질 관리 대책으로 한정되어 있다. 평균제곱오차의 특정한 요소를 추정하기 위한 심층실험과 사후조사 평가는 거의 실시되지 않는다. 하지만, 조사에 대한 응답률이 계속 떨어지면서, 특히 전화 조사에서의 조사 추정값의 무응답 편향의 평가는 흔해지고 있다. 이러한 평가에는 응답을 하도록 설득하는 목적으로 무응답자를 추적하는 것이 포함될 수 있다. 더 일반적인 것은 무응답자에 대한 서술적인 연구이다. 이런 연구에서는 조사 추출틀이나 현지 작업에서 분석자가 사용할 수 있는 무응답자와 응답자에 대한 모든 변수에 대하여 무응답자가 응답자와 어떻게 다른지 설명한다.

지속적으로 실시되거나 정기적으로 반복되는 조사에서는 조사 평가가 결정적인 역할을 한다. 지속적인 품질 개선을 위해서 뿐 아니라 일련의 자료의 유용성과 한계를 전달하는 수단으로 중요하다. 이런 이유로 몇 개의 미국 연방조사가 조사의 표본추출오차와 비표본추출오차에 의한 결함 지적하는 것 이외에 알려진 모든 것들

을 요약하는 방법으로 품질분석서를 개발했다.

품질분석서는 잠재적 오류의 원인인 특정화, 무응답, 추출틀, 측정, 자료처리를 검토하면서 조사 품질의 종합적인 상황을 제공하는 보고서이다. 품질분석서에는 조사가 실시되었던 전기간에 걸쳐 모아진 조사에 대하여 존재하는 모든 정보가 검토되고 종합되었다. 조사 품질분석서의 목적은 다음과 같다.

- 조사의 설계, 추정, 자료수집 과정에 대한 자세한 설명.
- 조사에서 모든 오류의 원인인 비표본추출오차 뿐 아니라 표본추출오차에 대하여 알려진 것의 포괄적인 요약의 제공
- 비표본추출오차에 대한 지식이 부족한 조사과정의 분야의 확인.
- 조사오차를 줄이기 위한 개선이 필요한 조사과정의 분야를 추천.
- 중요한 조사 추정값과 일련의 자료의 총평균제곱오차에 대한 지식을 넓히고 강화하기 위하여 필요한 추가적인 평가와 방법론 연구가 필요한 분야를 제안.

품질분석서는 정식 조사 서류에 보충되는 것이고 조사방법보고서, 미시적 자료 파일 사용안내서, 특정한 것에 대한 세부사항을 제공하는 기술보고서 같은 다른 형태로 입수될 수 있는 정보를 근거로 해야한다. 지속적인 조사에서는 시간에 걸쳐 이런 종류의 정보를 축적할 수 있다. 따라서 품질분석서는 지속되는 조사에만 한정되어 있다.

미국에서는 CPS(Brooks와 Bailar, 1978), SIPP(Jabine 등, 1990), SASS(Kalton 등, 2000), AHS(Chakrabarty와 Torres, 1996), RECS(U.S. Energy Information Administration, 1996) 등을 위한 여러 품질분석서들이 개발되었다. Kasprzyk와 Kalton(2001)은 미국 통계기관의 여러 품질분석서를 검토하고 조사의 개선과 품질고지를 목적으로 그들의 장점과 단점들을 논의 하였다. 다음은 SASS와 AHS의 품질분석서에서 검토된 주제들의 예이다.

SASS 자료수집의 품질분석서 개요

- 절차
 - 방식
 - 일정
 - 무응답자의 전화 추적
 - 추적이 필요한 학교의 특징 분석
 - 방식의 효과
 - 감독과 품질 보증
 - 설문지를 완성하는데 걸리는 시간
- 인지 연구와 사전 검사
- 재면접과 응답 분산
 - 재면접 표본
 - 재면접 절차
 - 응답 분산의 측정
 - 한 시점당 평가되는 질문들
 - 재면접에서 방식 변화의 효과
- 무응답
 - 학교 무응답률
 - 가중된 학교 무응답률
 - 항목 무응답률

AHS 자료수집의 품질분석서 개요

- 에디팅
- 자료 처리 과정에서의 품질관리 운영
 - 조사원의 에디팅
 - 자료 입력
 - 사전 에디팅
 - 컴퓨터 에디팅
- 입력의 품질 보증 결과
 - 방법
 - 오류율과 거부율
 - 입력 검증 결과
 - 수정
- 지역 사무실의 사전 에디팅에 대한 연구 결과
 - 거부된 것의 상태
 - 오류의 종류
 - 컴퓨터 에디팅 조치

제 9 장 표본추출오차

2장에서 조사추정에서의 총오차에 대해 살펴보고, 총오차가 표본추출오차와 비표본추출오차라는 두 가지 형태의 오차로 구성된다는 것을 보았다. 3장에서 8장까지는 비표본추출오차, 비표본추출오차의 원인 그리고 비표본추출오차의 효과를 평가하고 이를 줄이기 위한 다양한 방법들을 포함한 논제들에 대해 살펴보았다. 이 장에서는 조사추정에서 표본추출오차에 대해 설명할 것이며, 특히 다음 물음에 대해 중점적으로 논의할 것이다.

- 우리는 왜 표본을 뽑고 표본추출에서 확률화의 중요성은 무엇인가?
- 표본추출오차는 무엇이고 그것은 어떻게 발생되는가?
- 조사에서 표본추출오차는 어떻게 측정되는가?
- 어떤 설계 요소들이 표본추출오차의 크기에 작용하는가?
- 표준오차, 설계효과, 층화추출과 같이 표본조사 관련 분야의 주요 개념의 의미는 무엇인가?

이 장에서는 표본추출오차 관련 주제를 광범위하게 다루려는 것은 아니다. 또한 표본을 어떻게 설계할 것인지, 추출을 어떻게 할 것인지, 표준오차를 어떻게 추정할 것인지 또는 추정량을 어떻게 선택할 것인지에 대해 논의하는데 중점을 두지 않는다. 상세하게 이러한 주제들을 다루는 표본추출이론과 활용에 관한 많은 훌륭한 책들이 있고, 이것들은 참고문헌에 포함되어 있다. 여기서는 표본추출방법의 연구에 관련된 개념들을 강조할 것이다. 우리의 목적은 표본추출의 중요 주제에 대한 간략한 개관을 제공하는 것이고, 주로 표본추출에 대한 개념, 용어 그리고 이론적 근거 등이 독자들에게 친숙해 지도록 하는 것이다. 또한 가장 기본적인 수학적 개념들과 용어만을 사용하여 각 주제에 대해 논의하기 위해 노력할 것이다. 그래서 수학과 통계에 대해 많은 지식을 갖고 있지 않은 사람들도 접근할 수 있고 이해할 수 있는 있도록 하고자 한다.

비록 수식과 수학적 기호의 사용을 완전히 피할 수는 없지만, 수학적 기호의 사용을 가능한 자제하였다. 조사방법론에 대한 문헌의 대부분이 통계적 이론을 근거로 하기 때문에 이런 문헌에 접근하기 위해서는 통계와 수학의 언어(수식과 수학적

기호들)의 사용이 필수적이다. 이 장은 이 책에서 가장 기술적인 내용을 담고 있다. 우리는 수학적으로 기울어지는 것을 최소화 했으며, 기술적인 묘사를 위해 사용한 수학적 표현을 설명하고 명백하게 하기 위해 가능하면 이와 관련된 개념들을 말로 표현하기 위해 노력했다.

9.1절에서 우리는 표본추출의 역사에 대해 간략히 소개한다. 여기서는 편의표본에서부터 시작하여 유의표본의 구성 그리고 마지막으로 현대 조사연구 분야에서 거의 배타적으로 사용되고 있는 확률표본추출 방법을 차례로 살펴봄으로써 표본추출법의 발전과정에 대해 알아본다. 이 절에서는 1장에서 제공된 역사적 측면에서의 논의를 보충한다. 그러나 전체적으로 조사방법론보다는 표본추출방법의 발전에 초점을 맞춘다. 9.2절에서는 확률화(*randomization*)에 대한 기본적인 동기와 왜 이것이 다른 비확률추출법 보다 더 선호되는지 논의할 것이다. 9.3절에서는 단순확률표본추출(*simple random sampling*)이라고 부르는 매우 기본적인 방법을 사용하여 표본추출의 기본 개념들을 설명하고, 이 방법이 오늘날 사용되고 있는 좀 더 복잡한 형태의 표본추출법을 구현하는 데 어떻게 반영되고 있는지 논의한다. 9.4절에서는 표본조사로부터 통계적 추론을 하는 방법에 대해 논의한다. 마지막으로 9.5절에서는 표본추출과 조사방법 문헌에서 간혹 인용되는 좀 더 발전된 개념과 용어에 대해 논의할 것이다.

9.1 표본추출에 대한 간략한 역사

센서스가 군사적 그리고 과세 목적으로 국가의 전체 인구를 파악하기 위해 수행되었다는 관점에서 보면 사람들로 구성된 모집단에 대한 조사는 적어도 성서시대에서 그 기원을 찾을 수 있다. 그러나 조사방법을 배우는 학생들은 확률추출법이 20세기 초에 시작된 상대적으로 새로운 개념이라는 것에 종종 놀라게 된다. 현대 표본추출이론은 표본추출의 본질적인 아이디어를 발전시킨 세 명의 핵심적인 통계학자인 Kiaer, Bowley 그리고 Neyman에 의해 시작되었다고 할 수 있다. 그러나 이 초기의 혁신자들을 따르는 다른 많은 통계학자들 또한 표본조사 분야에 크게 공헌하여 현재의 방법론의 형태를 이룰 수 있게 도움을 주었다.

1900년경 노르웨이 통계학자인 Kiaer는 모집단의 작은 표본이 큰 모집단의 모수

를 정확히 추정하기 위해 사용될 수 있다는 개념을 처음으로 내놓았다. 그때까지 대부분의 통계학자들은 모집단에 대한 완벽한 센서스만이 모집단 모수에 대한 정확한 추정결과를 제공할 수 있다고 생각했다. Kiaer는 그의 생각을 대표방법(*representative method*)이라고 불렀다. 왜냐하면 그 목적이 상대적으로 작고 세심하게 모집단으로부터 뽑힌 표본으로 전체 모집단을 대표하기 때문이다. Kiaer의 방법은 어떤 형태의 확률화 과정도 포함하지 않았고, 사실 확률화 개념은 그 당시 통계학자들이 잘 이해하고 있는 것이 아니었다. 오히려 그의 방법은 모집단을 대표할 수 있는 표본을 전적으로 표본선정자의 판단과 기술에 의존하여 구성한 것이다. 이 아이디어는 작아진 크기 때문에 전체 모집단보다는 연구하기 쉬운 모집단 축소판을 창출하기 위해 전문가의 판단과 모집단에 대한 지식을 사용한 것이다. 만약 적절하게 설계되었다면 이 축소판은 전체 모집단의 크기보다 훨씬 작음에도 불구하고 본래 모집단과 상당 부분 동일한 특성을 가질 수 있었다.

Kiaer는 가장 최근의 센서스로 얻은 자료를 그가 원하는 유의표본을 구성하기 위해 활용하였는데, 그는 이런 과정을 통해 조사목적상 가장 중요하다고 생각하는 특성에 있어서 모집단과 표본이 서로 대응이 되도록 표본을 선정했다. 예를 들어, 도시와 교외 지역이 혼합된 어떤 국가에 대한 표본을 설정하는 경우 만약 도시화 수준이란 특성이 조사목적 달성에 있어 가장 중요한 설명 변수가 될 수 있다고 판단하는 것이다.

Kiaer의 아이디어가 자리 잡고 확장됨에 따라 영국 출신의 Bowley는 표본추출 분야에 확률화 개념을 공식적으로 소개했다. Bowley는 Kiaer의 아이디어처럼 표본의 대표성 확보를 위해 특정인의 모집단에 대한 지식과 유의적인 선택에 의존하는 것보다는 확률 추출에 의해 보다 대표성 있는 표본 추출이 가능하다고 했다. Bowley는 확률 표본추출은 복잡한 이론과 모집단의 상황에 의존하지 않는다고 주장했다. 그래서 전문가 또는 모집단에 대한 특별한 지식을 가진 사람이 아닌 어떤 사람들도 얼마든지 대표성 있는 표본을 뽑을 수 있다는 것이었다. Bowley의 확률화 방법은 모집단의 모든 단위에 대해 동일한 추출확률을 할당한 것이다. 그는 모자에서 숫자를 뽑는 것과 같이 본질적으로 단순한 방식으로 어떤 크기의 대표성 있는 표본이라도 뽑을 수 있다고 하였다. 후에 이런 방법을 단순확률추출법(*simple random sampling*)이라고 부르게 되었다.

1934년에 Jerzy Neyman이라는 통계학자가 유명한 논문을 발표함으로써 조사 분

야에 일대 변혁이 발생했다. 그의 논문은 확률추출법에 대한 이론을 제공하였고, 등 확률과 불균등확률에 의한 표본추출을 모두 인정하였다. Neyman의 논문 이전까지는 조사연구 분야에서는 유의표본추출과 확률표본추출 중 어떤 방법이 우수한지에 대한 끈임 없는 논쟁이 있어 왔다. 의 든 조사분야에서 어떤 것이 더 선호되는지를 진행하는 토론이 있었다. 각각의 방법은 서로 장단점을 가지고 있었고, 어떤 이들은 유의표본추출이 확률추출법보다는 훨씬 더 효율적이라고 주장했다. Neyman의 논문은 확률화 원칙을 기초로 한 통계적 추론에 대한 통합 이론을 제공gkadmfhTJ 이런 논쟁을 종식시켰다. Neyman은 랜덤화를 통해 정확한 결과를 얻게 될 확률을 파악할 수 있는 모집단에 관한 추론이 가능하다는 것을 보여주었다. 그는 또한 유의추출법은 어떤 통계적 이론으로도 설명될 수 없다는 것을 보여주었다. 예를 들어 확률추출에서 사용하는 통계적 신뢰구간을 구성하는 것이 가능하지만 유의추출에서는 그렇게 하는 것이 불가능하다.

Neyman의 역사적인 논문은 조사통계 분야에 대단한 반향을 일으켰고, 1940년대에 이 아이디어를 확장시킨 많은 논문들이 발표되었다. Hansen, Hurwitz, Madow, Cochran, Kish, Deming과 같은 통계학자들은 현대 표본추출 이론의 기초가 되는 개념들을 체계화시켰다. Yates, Sukhatme, Murthy, Des Raj와 같은 다른 통계학자들은 표본추출 분야를 발전시키는 데 큰 도움이 된 관련 교과서를 저술했다. 또한 Dalenius, Godambe, Horvitz, Thompson, Jessen, McCarthy, Waksberg, Pritzker, Stephan, Tepping and Mahalanobis와 같은 통계학자들은 1940에서 1965년까지 이루어진 표본추출 이론 및 응용의 발전에 상당한 공헌을 했다.

역사는 표본추출의 발전에 명확한 동기를 부여했다. 전체 모집단(완벽한 센서스) 대신에 모집단의 일부분에서 조사 자료를 수집한다는 아이디어는 원하는 수준의 정확한 추정값을 얻으면서 조사비용을 줄이고자 하는 의도에서 원래 시작되었다. 그러나 초기통계학자들은 표본조사를 통해 센서스 보다 양질의 자료를 얻을 수 있다는 중요한 표본조사의 장점을 깨달았다. 이는 표본 자료는 모집단의 오직 일부분에서 수집되기 때문에 자료수집 작업의 관리가 용이하고 결과적으로 자료의 질을 충분히 관리할 수 있다고 생각한 것이다. 그러나 모집단의 일부에 해당하는 표본에서 얻은 자료를 기반으로 추론을 하기 때문에 발생하는 가장 큰 문제는 표본추출오차이다. 특히 시군구 또는 읍면동 등과 같이 소지역에 대한 추정을 목적으로 하는 경우 이런 문제는 더욱 심각해진다. 만약 소지역 추정에 대한 정확성 확보가 우선적

으로 고려되어야 한다면 표본조사는 최선의 선택이 아닐 수 있다.

오늘날 표본조사는 타당성 있는 결과들을 제공함으로써 신뢰 받게 되었지만, 50년 전까지만 해도 많은 통계기관에서 표본추출방법을 너무 급진적인 아이디어로 간주했다는 점은 매우 의외의 사실이다. 표본조사에 근거를 둔 추론은 타당하고 표본추출에 대한 오차는 표본 자료를 사용함으로써 측정할 수 있다는 것을 증명한 통계적 이론을 Neyman이 제공하면서 비로소 표본추출법이 폭넓게 수용되기 시작했다. Neyman은 확률화를 통해(즉, 알려진 확률로 표본 단위를 추출함) 추정값을 중심으로 한 신뢰구간을 구하는 것과 모집단 모수에 대한 추론을 위해 이 신뢰구간을 사용하는 것이 가능하다고 증명했다.

그러나 오늘날 전통 표본추출이론(classical sampling theory)이라고 불리는 Neyman시대의 이론은 충분히 고려되지 못한 결점을 갖고 있다. 전통적인 표본추출이론은 모든 표본 단위들은 오차 없이 관찰할 수 있다는 것을 전제로 하고 있다(즉, 자료에는 비표본추출오차가 전혀 없다). 만약 이 가정이 적절하면 전통적인 표본추출이론은 조사에 그대로 적용될 수 있다. 그러나 우리가 이 책을 통해 여러 번 설명했던 것처럼 비표본추출오차는 거의 모든 조사에서 항상 발생하고, 표본추출이론에 따르면 비편향성이 만족되는 추정량을 사용함에도 불구하고 비표본추출오차로 인해 편향이 발생할 수 있다는 것이다. 다시 말해 Neyman이론을 토대로 하는 모집단에 대한 추론 방법은 여기서 제안한 이론보다 상당히 더 큰 오차를 발생시킬 수 있다. 수년에 걸쳐 통계학자들은 조사 자료에 필연적으로 포함되는 비표본추출오차를 보정하기 위해 이론을 조정하고 보완하기 위한 방법들을 발전시켰다. 이 방법들에 대한 설명은 고전적인 표본추출이론의 기본개념을 몇 가지 배운 후에 이 장의 마지막 부분에서 다룬다.

현대 표본추출이론은 유의 표본추출에서 단순확률추출로 진화한 후, 궁극적으로 복합설계에서 사용되는 불균등 확률표본추출로 발전했다.

9.2 비확률 표본추출 방법들

조사연구에서 확률표본추출의 광범위한 사용에도 불구하고 비확률표본추출방법도 여전히 사용되고 있다. 비확률추출법은 특수한 상황에서는 확률추출법보다 더 선호될 수도 있고, 어떤 경우 확률표본추출법보다 효율적이고 효과적인 대안이 될 수도 있다. 그러나 부적절하게 비확률추출법이 사용되는 사례들을 흔히 볼 수 있다. 예를 들어 실제 표본은 비확률추출법으로 선정하였지만 통계적 추론을 목적으로 마치 확률추출법을 적용한 것처럼 간주하고 분석하는 경우를 흔히 볼 수 있다. 우리는 이 절의 마지막 부분에서 이런 방법들의 남용 사례들에 대해 논의할 것이다.

빈번하게 사용되는 비확률표본추출법은 편의추출(*convenience sampling*) 또는 무계획추출(*haphazard sampling*)이다. 다음과 같은 경우들이 이런 추출법에 의한 표본에 해당한다.

- 학교계시판의 공고를 보고 실험실 실험에 참여하기 위해 지원한 학생들
- 호텔방에 놓여진 고객만족 설문지에 대답하는 호텔 고객들
- 설문지를 조사하기 위해 연락한 친구들과 친지들
- 특정 주제에 대해 의견을 피력하기 위해 TV 투표에 참여하는 TV시청자들

비록 이런 표본들을 근거로 한 통계적 추론은 타당성이 없지만 8장과 10장에서 언급된 조사 질문 또는 절차를 평가하는 것과 같이 특수한 목적을 위해 편의표본이 유용할 수 있다. 불행하게도 편의표본들은 대규모 모집단에 대한 추론을 하기 위한 근거로는 부적절하다. 예를 들어 어떤 호텔체인에서는 고객들이 모든 호텔방에 놓여진 고객만족도 조사 설문지들에 대한 응답결과를 기초로 그 호텔의 서비스가 매우 만족스럽지 못하다고 추론할 수도 있다. 그러나 그러한 설문지들은 흔히 서비스에 불만을 갖는 고객들에 의해 자신들의 불만사항을 나타내기 위해 주로 작성될 수 있을 것이다. 따라서 이런 응답자들로 구성된 표본은 그 호텔의 모든 고객을 대표한다고 할 수 없고, 결과적으로 그것들로부터 얻은 추론결과는 굉장히 큰 오류를 범할 수 있다. 명확하게 말하면 편의표본에서 얻은 결과를 기초로 한 추론 결과를 정당화시킬 수 있는 통계 이론은 존재하지 않는다.

비확률표본추출의 또 다른 유형은 유의추출(*purposive sampling*) 또는 판단추출(*judgement sampling*)이라고 불리는 추출법이다. 그것은 본질적으로 9.1절에서 설명했던 오래전에 Kiaer에 의해 제안된 방법이다. 이는 표본추출 담당자가 목표 모

집단을 잘 대표할 수 있다고 생각하는 단위들을 전문가의 판단을 기준으로 선정하여 표본을 구성하는 유의표본추출법이다. 예를 들어, 큰 회사의 사장이 어떤 회사 정책에 대한 종업원들의 의견을 파악하기를 원한다고 하자. 이를 위해 사장은 10명의 종업원으로 구성된 유의표본을 선정했다. 이 선정과정에서 사장은 회사 종업원 인력현황을 참고로 모든 급여수준의 종업원을 대표할 수 있는 다양한 급여수준의 5명의 관리자와 5명의 일반 사원을 표본으로 선정했다. 동시에 사장은 회사정책에 대해 정직하고 솔직하게 대답해 줄 것으로 기대되는 사원들을 표본으로 선정했다면 이런 표본이 유의표본에 해당한다.

이런 표본추출 방식을 통해 회사정책과 관련된 타당하고 유용한 정보를 얻는 것이 가능하고, 사장은 그 회사 직원들의 일반적인 의견을 상당히 정확히 파악할 수 있다. 그러나 조사에 참여한 사람들은 직원들의 일반적인 시각을 반영하지 못할 가능성이 매우 높다. 여기서는 모든 것이 직원 현황에 대한 사장의 이해, 회사를 대표하는 사람들을 뽑을 수 있는 사장의 능력 그리고 청취된 의견을 종합 해석할 수 있는 사장의 판단력에 의존하게 된다. 한편 오직 10명의 매우 작은 표본을 고려하는 경우, 유의추출은 확률추출에 비해 더 정확한 정보를 제공할 수도 있다. 명확하게 말하면 표본크기가 아주 작은 경우 확률추출은 유의추출보다 편향을 어느 정도 줄일 수는 있지만 반면에 상당히 큰 표본추출오차를 갖게 될 위험이 높다. 예를 들어 회사에서 완전히 랜덤하게 10명을 뽑는다면 특히 회사정책에 대한 지식이 없거나 다소 특이한 종업원이 표본에 포함됨으로써 상당히 왜곡된 결과를 얻게 될 가능성이 있다는 것이다.

그러나 이 장의 마지막부분에서 보게 되겠지만, 표본크기가 증가함에 따라 표본추출오차는 감소하기 때문에 어느 정도 표본크기가 커지게 되면 상당히 정확한 결과를 얻게 된다. 예를 들어 만약 표본크기가 10명에서 50명으로 증가한다면 확률표본으로부터 얻은 추론 결과가 동일 크기의 유의표본에서 얻은 결과에 비해 훨씬 더 정확할 가능성이 높다. 50명으로 구성된 유의표본은 조사 설계자의 선택편향의 결과로 더 큰 총오차를 가질 가능성이 높다. 더욱이 확률추출과 달리 유의추출에서는 표본 추정량이 모집단 모수의 참값에 얼마나 근접하는지에 대한 정보가 전혀 없다.

미국 센서스국은 센서스 절차의 사전검사를 수행하기 위한 지역을 선정하기 위해 유의추출법을 사용해 오고 있다. 이러한 사전검사에는 센서스에 따른 높은 비용 때문에 오직 2개 또는 3개 지역(도시 또는 교외지역)만을 포함시킨다. 센서스를 위한

사전검사의 목적이 주로 국가 모집단에 대한 추론보다는 자료수집과 추정절차와 관련된 다양한 대안들을 테스트하고 비교하는 것이기 때문에 지역의 형편을 고려해 유의적인 지역 선정이 가능한 유의추출이 선호된다. 센서스국은 전수조사가 매우 까다로운 지역을 대표하는 도시, 계상에 있어서 매우 다양한 문제를 가지고 있는 교외지역을 대표하는 곳 그리고 인디안 원주민과 같은 소수 그룹이 밀집된 지역을 대표하는 곳을 선택한다. 만약 이런 지역들을 랜덤하게 선정한다면 이들 지역에서는 아마도 센서스 절차와 관련된 중요한 사항들에 대한 철저한 점검을 할 수 없을 것이다.

비확률추출법의 세 번째 형태는 할당표본추출(*quota sampling*)이다. 할당추출은 많은 표본추출 기법에서 사용되고 있는 개념으로 표본에 대한 사전에 설정된 특정 조건별 사람수를 채울 때까지 일정 조건을 만족하는 사람들을 계속 면접하는 표본추출 방법을 말한다. 예를 들면, 다양한 차종에 대한 고객 만족도를 평가하는 데 관심이 있는 자동차 제조업체에서 대형 할인점 방문 고객들을 가로막고 조사하기로 결정했다. 그러면 예를 들어 조사원에게 지나가는 고객 중 밴을 소유하고 있는 10명, 일반 승용차를 소유하고 있는 20명, 트럭을 소유하고 있는 15명을 대상으로 조사하라고 한다면 이는 할당추출에 해당한다.

할당추출은 마케팅조사에서 흔히 사용된다. 왜냐하면 이 방법이 확률추출법에 비해 관리하기 쉽고, 비용이 적게 들고, 더 빠르게 조사를 완료할 수 있기 때문이다. 그러나 할당추출은 편의, 판단, 유의추출과 마찬가지로 본질적인 한계를 갖고 있다 (통계적 추론을 위해 확률적인 기반을 갖고 있지 못하다). 할당표본추출에서 조사원들은 추출기준을 만족한다면 자유롭게 자신이 원하는 사람을 선택해 조사할 수 있다. 이런 대표자들은 흔히 목표 모집단을 대표하지 않고, 자료를 수집하는 조사원의 태도와 신념에 의해 구성된다. 아울러 그들은 대부분 자진해서 응답하기를 원하는 사람들이거나 또는 조사원에게 쉽게 의견을 표현할 수 있는 사람들이다.

확률추출에 해당하지는 않지만 확률추출의 요소를 일부 포함한 할당추출을 약간 변형한 추출법이 있을 수 있다. 예를 들어 할당추출에서 모집단에 대한 아주 큰 확률표본을 선택하여 조사원에게 주고, 조사원에게 이 표본에서 할당에 따른 조건을 만족할 때까지 계속 면접을 하도록 하는 것이다. 예를 들어 12-15세에는 x 명, 16-25세에는 y 명, 25세 이상은 z 명을 조사하라는 할당 조건이 주어졌다고 하자. 만약 처음에 시도한 표본 구성원이 거부하거나, 접촉이 안 되면 다음 표본 구성원을 조

사하는 방식으로 모든 할당이 채워질 때까지 계속 조사하라고 조사원들에게 지시하는 방식이다. 이런 방식을 사용하면 접근하기 어렵거나 참여를 꺼리는 사람들에 대한 대표성이 결여되는 문제를 가질 것이다. 왜냐하면 실제로 이 방법은 비협력적이고 접근 불가능한 표본 구성원들을 협조적이고 접근이 용이한 표본구성원들로 대체하는 효과가 있기 때문이다. 전통적인 확률추출에서는 조사원은 할당과 관계없이 표본의 모든 구성원들에 대해 조사를 수행해야 한다. 응답을 꺼리고, 일시적으로 접근 불가능하고, 접촉하기 힘든 사람들은 재방문하는 등의 모든 방법을 동원하여 모든 표본 구성원에 대한 조사를 완료해야 한다. 이런 방식의 자료수집 전략을 통해서만이 면접이 힘든 사람들도 여전히 조사결과에 적절하게 반영될 수 있다.

비확률추출법에는 편의추출법, 무계획추출법, 유의추출법, 할당추출법 등이 있다.

“확률” 할당표본추출의 예로는 최근 대학교 졸업생들로 구성된 확률표본에 대해 수행된 조사가 있다. 이 연구에서 조사 의뢰자는 면접이 1000명의 졸업생에 대해 수행되어야 한다고 규정했다. 조사 설계자는 이전 연구들로부터 조사의 응답률이 약 50%정도 된다고 판단했다. 그래서 목표 표본크기인 1000명에 대한 면접을 수행하기 위해 2000명으로 구성된 표본을 추출했다. 그러나 조사의 응답률이 예상보다 훨씬 높아서 1000명에 대한 면접은 모든 2000명의 표본 구성원들과의 접촉이 이루어지기 전에 종료되었다. 여기서 할당추출 전략이 적용되었기 때문에 표본추출은 1000명의 면접이 완성되었을 때 중단된 것이다. 그러나 이 전략은 협력적이고 쉽게 접근할 수 있는 사람들이 처음에 면접되어지기 때문에 이런 사람들이 과도하게 표본에 포함되는 문제를 갖고 있다. 확률표본추출에서의 규칙은 2000명 표본구성원들 모두에게 접촉을 하고 면접이 시도되어야 한다는 것이다. 더욱이 여기서 응답률은 본래 뽑힌 2000명의 표본을 기준으로 산출되어야 한다. 만약 본래 2000명 표본구성원들 중 1000명을 면접한다면, 비록 1000번째 면접이 완성된 후 실제 표본 구성원들 중 200명 또는 그 이상은 전혀 접촉이 이루어지지 않았더라도 이 연구의 응답률은 50%가 된다.

지금까지의 논의를 요약하면, 비확률추출은 조사연구 분야에서 그들 나름의 영역을 가지고는 있지만, 확률표본추출이 신뢰할 수 없는 결과를 산출할 때나 혹은 추

론의 통계적 타당성보다는 조사되어지는 사람이 누구인가 하는 것이 더 중요하게 통제되어야 할 때만 사용되어야 한다는 것이다. 그러나 대부분의 경우에 확률표본 추출이 비확률추출에 비해 다음과 같은 많은 장점들을 갖고 있다.

비확률추출법에 비해 확률표본추출법이 갖는 장점

- 전체모집단을 대표하는 표본을 생산한다: 표본선택편향으로부터 자유롭다.
- 통계적 추론에 대한 이론적인 근거가 있다.
- 상대적으로 적용하기 쉽다: 전문가의 판단이 필요 없다.
- 재수행이 가능한 방법을 사용하다: 표본추출법이 문서화될 수 있고, 다른 사람들에 의해서도 수행될 수 있다.
- 표본크기의 설정을 통해 연구자는 추정의 정확도를 사전에 상당히 정확하게 결정할 수 있다.

다음 절에서 확률표본을 구성하는 데 있어서 가장 간단한 방법 중 하나인 단순 확률추출법에 대해 알아볼 것이다. 이 방법은 실제 조사에서 적용한다는 측면보다는 다른 좀 더 복잡한 표본추출법의 기반을 제공한다는 측면에서 매우 중요하다.

9.3 단순 확률추출

9.3.1 표본추출에서 사용되는 용어

표본추출방법에 대한 논의를 용이하게 하기 위해 우선 관련 용어들을 정리하고자 한다. 비록 이 중 일부는 이미 앞에서 소개된 것들이지만 완성된 용어정리를 위해 여기서 다시 설명한다.

1장과 2장에서 논의한 것과 같이 조사 설계의 첫 번째 단계 중 하나는 연구 대상 모집단을 정의하는 것이다. 그러나 특정 조사와 관련해서도 몇 가지 형태의 모집단 개념이 정의될 수 있다. 조사의 목적과 대상물을 정의하기 위해서는 목표모집단 및

추출틀 모집단을 정의하는 것이 필요하다. 목표모집단은 조사에서 연구 대상이 되는 동시에 조사로부터 얻어진 추론의 대상이 되는 모집단이다. 그러나 표본추출을 목적으로 하는 경우 추출틀(*frame*) 또는 조사(*survey*) 모집단이란 개념이 더 유용하다. 이것은 표본을 추출하는 데 이용하는 표본추출틀로 대표되는 모집단이다. 2장에서 이미 두 모집단이 어떻게 다른지 설명하였고, 이들의 차이를 보여주는 사례들을 제시했다. 표본추출틀, 표본추출과정, 자료수집과정의 실제 제약조건 때문에 목표모집단은 추출틀모집단과 상당히 차이가 있을 수 있다. 3장의 내용을 상기해 보면 조사모집단과 목표모집단간의 차이로 인해 추출틀 포함오차와 편향이 발생하게 된다는 것을 알 수 있다.

일단 모집단을 정의하고 조사목적은 안다면, 조사 목적을 성취하기 위해 표본을 어떻게 설계하는지에 대해 초점을 두어야 한다. 표본추출의 목적을 위해 모집단은 조사단위와 모집단단위를 포함하는 것으로 생각되어질 수 있다. 모집단 조사단위(*population element*)는 조사 자료 수집대상이 되는 근본적인 실체이다. 예를 들어, 가구조사에서 만약 각 가구 구성원을 대상으로 자료가 수집된다면 조사단위는 가구원이 될 것이고, 가구를 대상으로 한 조사가 이루어진다면 가구 자체가 조사단위에 해당한다. 모집단 단위(*population unit*)는 자연적인 기준에 의해 구성된 조사단위들의 집합으로 정의될 수도 있다. 예를 들어 같이 집에 살고 있는 사람들(조사단위들)로 구성된 하나의 거주단위가 모집단 단위에 해당할 수 있다. 사업체조사에서 단위는 회사 내의 사업체나 지점을 모두 포함한 하나의 회사가 단위가 될 수 있다. 농업조사에서는 단위가 전체 농장과 이에 부속된 토지들이 될 수 있다. 그러나 전체 농장을 구성하는 토지 구획들이 모집단 조사단위이다.

확률표본추출을 하기 위해서는 모집단 조사단위가 선택되어질 수 있는 표본추출틀(*sampling frame*)이 반드시 존재해야 한다. 가장 단순한 경우 추출틀은 모든 모집단 조사단위의 목록이다; 예를 들어, 조사 대상이 되는 협회의 모든 구성원들의 목록이다. 그러나 매우 흔히 지역거주자 조사에서 가구의 주소 목록 또는 사업체조사에서 회사 목록처럼 모집단 단위들의 목록이 추출틀이 될 수도 있다. 어떤 경우에는 모집단 단위들에 대한 물리적인 목록이 아니라 암시적으로 정의되는 목록이 추출틀이 될 수 있다. 이에 대한 좋은 예는 랜덤번호전화걸기(RDD) 조사에서 사용되는 모든 거주자들의 전화번호 목록이다.

많은 대부분의 실제 상황에서 추출틀은 전체 목표모집단을 포함하지 않고 모집단

의 특정부분은 알고 있지만 제외된다. 예를 들어 RDD조사에서 전화가 없는 가구에 살고 있는 사람들은 자료수집 방법의 실행상의 제약 때문에 제외된다. 만약 전화가 없는 가구에 살고 있는 사람들을 포함한 조사대상 지역에 살고 있는 모든 사람으로 정의한 목표모집단을 추론 대상으로 한다면, 목표모집단은 조사모집단과 동일하지 않을 것이고 포함편향의 가능성을 고려해야만 한다. 그러나 만약 목표모집단이 전화가 있는 가구에 사는 사람만을 대상으로 한다면 목표모집단과 조사모집단은 일치하기 때문에 전화가 없는 가구의 배제 때문에 발생하는 편향의 위험이 없다.

이와 유사하게 지역추출틀 조사에서는 매우 작은 외딴 지역에 사는 사람들은 배제할 지도 모른다. 왜냐하면 오직 한두 명의 면접을 수행하기위해 이러한 지역으로 출장을 가는 비용 때문이다. 이러한 상황에서는 목표모집단에 포함된 작고 외딴 지역을 고의적으로 추출틀에서 제외할 수 있다. 그럼에도 불구하고 여전히 추출틀에서 제외된 지역에 사는 사람을 포함한 전체 지역의 모든 사람으로 목표모집단을 정의 할 지도 모른다. 만약 제외된 지역이 차지하는 비율이 전체 목표모집단의 1 또는 2%를 정도라고 한다면 포함편향에 따른 위험은 매우 작다(3장 참조). 표본추출과 자료수집상의 효율성을 높이기 위해 작은 포함편향을 수용하는 것이 많은 경우에 있어서는 실제적으로 바람직한 조사 수행방안일 수 있다.

모집단을 설명하는 요약된 특성을 모집단 모수(*parameter*)라 부른다. 조사의 목적은 모집단의 하나 또는 그 이상의 모수를 추정하는 것이다. 이 목적을 위해 사용되는 표본 통계량들을 추정량(*estimator*)이라 부른다. 예를 들어 모집단 모수가 모집단 구성원의 평균소득이라고 하면 추정량은 표본의 평균이 된다. 또는 모집단 모수가 모집단의 교육과 소득사이의 상관관계라고 한다면 추정량은 이 특성들에 대한 표본 상관관계이다. 조사 자료로부터 계산된 추정량의 실제 값을 추정값이라고 부른다. 요컨대 추정량은 본질적으로 표본 추정값을 계산하기 위한 공식이다.

모든 단순 추정값들은 표본추출오차(*sampling error*)의 영향을 받는다. 표본추출오차는 표본 추정값에 대한 오차(즉 추정값과 모집단 모수의 참값사이의 차이)를 나타내며, 이는 전체모집단 대신 오직 전체 모집단의 부분집합을 선택하기 때문에 발생한다. 표본추출오차를 나타내는 일반적인 척도는 표본분산(*sample variance*)이고, 이는 본질적으로 추정값과 모집단 모수의 편차 제곱을 모집단으로부터 뽑을 수 있는 모든 가능한 표본에 대해 평균한 것이다. 이 추정량의 분산에는 추정량이 모집단 모수와 얼마나 근접한지에 관한 정보가 포함되어 있다. 추정량의 분산에 대한

제곱근이 추정량의 표준오차(*standard error*)이다. 모집단 모수의 추정값과 추정량의 표준오차에 대한 추정값이 통계적 추론(즉, 신뢰구간과 가설검정)을 위해 필요한 기본적인 값들이다. 앞에서 언급한 것과 같이 표본의 확률추출을 통해 조사 추정량의 분산과 표준오차에 대한 추정값은 표본 자료로부터 직접 계산되어질 수 있다.

앞으로의 논의에서 중요한 추정량의 두 가지 특성은 비편향성(*unbiasedness*)과 효율성(*efficiency*)이다. 만약 모집단으로부터 추출 가능한 모든 표본으로부터 계산된 추정값의 평균이 모집단 모수와 일치하면 이 추정량은 비편향이다. 즉 비록 추출된 각각의 표본에서 계산된 추정값은 모수와 차이가 생기지만, 비편향추정량은 평균적으로는 모수의 참값을 산출한다는 것이다. 만약 그것이 작은 표준오차, 다시 말해 정도(*precision*)가 높다면 이 추정량은 효율적(*efficient*)이라고 한다. 어떤 경우에 특정 모집단 모수를 추정하기 위해 이용할 수 있는 몇 가지 추정량이 있고, 만약 모든 추정량이 비편향이라면, 가장 효율적인 것이 최량(*best*) 추정량이다.

9.3절의 나머지 부분에서는 확률적인 방법을 사용하여 표본추출틀로부터 모집단 조사단위를 선택하는 방법들에 대해 설명한다. 실제 상황에서 활용되는 경우는 드물지만 그럼에도 불구하고 이론적으로는 매우 중요한 단순확률추출법부터 우선 설명한다. 그런 다음 우리는 모든 모집단 단위가 같은 선택확률을 가지도록 표본을 추출하는 다른 몇 가지 방법들을 논의한다. 등확률추출법(*equal probability of selection methods; epsem*)이라고 불리는 이런 추출법은 표본크기(즉, 비용)와 정도라는 관점에서 매우 효율적이라고 알려져 있다.

표본추출에서 사용되는 기술적인 용어들의 요약

- 모집단 조사단위(*population element*): 측정 대상이 되는 기본적인 실체
- 모집단 단위(*population unit*): 한 개 또는 그 이상의 모집단 조사단위로 구성된 자연적으로 발생한 실체이며, 추출 대상이 되는 모든 모집단 조사단위를 포함하는 것
- 목표모집단(*target population*): 연구하기 원하는 모든 단위들의 집합
- 추출틀(*frame*): 표본추출을 위해 이용할 수 있는 모집단 단위 목록
- 추출틀모집단(*frame population*) 또는 조사모집단(*survey population*): 추출틀로 대표되는 단위들의 집합
- 모집단 모수(*population parameter*): 추정 대상 모집단을 특정 짓는 요약값
- 추정량(*estimator*): 모집단 모수를 추정하는 표본 관측값들의 함수
- 추정값(*estimate*): 실제 계산된 추정량의 값
- 표본추출분포(*sampling distribution*): 모집단으로부터 뽑힐 수 있는 각 표본에 대한 추정량의 값을 보여주는 빈도 히스토그램
- 비편향추정량(*unbiased estimator*): 추정량의 표본추출분포의 평균과 모수가 일치하는 추정량
- 효율적인 추정량(*efficient estimator*): 표본으로부터 구성될 수 있는 모수의 모든 추정량들 중 가장 작은 분산을 갖는 추정량
- 등확률표본추출설계(*epsem design*): 모집단의 모든 단위들이 선택될 확률이 동일한 등확률 선택방법에 의한 설계

9.3.2 단순확률표본추출의 정의

크기가 n 인 단순확률추출법은 표본추출틀로부터 추출될 수 있는 n 개의 단위들로 구성된 모든 가능한 표본이 추출될 확률이 동일한 추출법이다. 그 이름에 내포되어 있는 것처럼, 단순확률표본(*simple random sampling; SRS*)은 모집단으로부터 확률 표본을 뽑는 가장 단순한 표본추출방법이다. 그러나 보통 더 효율적인 표본추출방

법(즉, 더 나은 정도를 가진 추정량을 산출하는 표본추출방법)을 찾을 수 있기 때문에 대규모 사업체 조사나 사회 조사에서 널리 사용되는 방법은 아니다. 하지만 소규모 연구 또는 비공식적인 확률화를 목적으로 하는 조사에서는 SRS이 여전히 유용하다.

예를 들어, 복권추첨에서 당첨자를 뽑기 위해서 복권 소유자의 이름 쓰인 티켓을 큰 상자 안에 넣고 잘 섞는다. 그런 다음 상자 안에 손을 넣어 “랜덤하게” 하나의 티켓을 잡음으로써 당첨자의 이름이 선택되게 된다. 추가적으로 당첨자를 뽑기 위해서, 이미 뽑힌 티켓을 다시 넣지 않고 그 과정을 반복한다. 이런 방식으로 하면 모든 복권티켓이 선택될 확률이 같아지게 된다. 복권 당첨자, 출발선의 정렬순서, 게임이나 스포츠 경기에서의 순서를 뽑는 것들은 모두 이런 방법의 일종으로 볼 수 있다.

비록 추정량에 정도를 보다 향상시킬 수 있는 좀더 복잡한 표본설계를 할 수 있지만, SRS는 어떤 경우에는 유용한 방법일 수 있다. 그럼에도 불구하고 더욱 복잡한 표본추출방법에 대한 통계적인 기초를 제공하기 때문에 SRS는 여전히 표본추출방법의 연구에서 매우 중요하다. 더욱 복잡한 표본추출 방법의 효율성은 흔히 SRS와의 비교를 통해 설명되어진다. 예를 들어, 9.5.6에서 논의되는 추정량의 설계효과(design effect)는 SRS와의 상대적인 비교를 통해 표현된다. 아울러 SRS의 추정 공식과 분산 추정량은 이해가 쉽고 간단하기 때문에 이런 공식들은 종종 더 복잡한 표본추출방법들의 설명하는 수단으로 사용된다. 매우 복잡한 표본추출방법조차도 표본추출 과정 중에는 반드시 SRS를 사용하고 있으며, 그래서 SRS에서 사용되는 공식이 적용되게 된다.

다음 장에서 우리는 SRS 표본을 뽑는 방법을 논의하고, 그 방법을 이용하여 목표모집단에 관한 추론을 하는 기초적인 공식을 제시할 것이다. 또한 표본크기의 결정, 추론을 위한 “최량” 추정량의 선택, 통계적 추론에서 무응답오차의 영향을 포함한 조사 표본설계에 관련된 많은 문제들을 논의한다.

크기가 n 인 단순확률표본(SRS)은 표본추출틀로부터 추출될 수 있는 n 개의 단위들로 구성된 모든 가능한 표본이 추출될 확률이 동일하도록 추출한 표본이다.

9.3.3 단순확률표본추출의 기본 개념

SRS에 대한 기본적인 사항들을 설명하기 위해 우리는 가상적인 매우 작은 모집단으로부터 어떻게 표본을 뽑을 수 있는지 생각해 보도록 한다. 이를 위해 오직 10개의 단위를 포함하는 모집단을 생각해 보자. 물론, 이렇게 작은 모집단은 실제로 표본추출 대상이 되지 않을 것이다. 그러나 이렇게 작은 모집단을 사용하여 SRS의 개념을 설명하는 것이 더 쉽다. 모집단 내의 개인들의 평균 소득을 추정하기 위해서 모집단으로부터 표본을 뽑으려 한다고 가정해 보자. 이 설명을 위해서 그 모집단의 모든 10명에 대응되는 값들이 <표 9.1>에 제시되어 있다. 이 표의 첫째 줄은 모집단 단위의 라벨을 나타내고, 둘째 줄은 i 번째 단위의 소득을 나타내며 이는 Y_i 로 표시하기로 한다.

<표 9.1> 10명의 소득으로 구성된 소규모 모집단

모집단 구성원	실제 소득
P1	$Y_1 = 60,000$
P2	$Y_2 = 72,000$
P3	$Y_3 = 94,000$
P4	$Y_4 = 90,000$
P5	$Y_5 = 102,000$
P6	$Y_6 = 116,000$
P7	$Y_7 = 130,000$
P8	$Y_8 = 135,000$
P9	$Y_9 = 141,000$
P10	$Y_{10} = 160,000$
평균	$\bar{Y} = 110,000$

이 모집단으로부터 표본크기 $n=2$ 인 표본을 어떻게 뽑을지 생각해 보자. 몇몇 방법들을 통해 단순확률표본을 추출하는 것이 가능하다. 각 방법은 매우 일반적이고 모

집단 크기 N 또는 표본크기 n 에 관계 없이 어떤 모집단으로부터도 단순확률표본을 추출할 수 있을 것이다. 하지만 어떤 방법은 다른 방법들에 비해 실행이 용이할 수 있다. 우리가 논의하는 모든 방법들은 비복원(*without replacement*) SRS에 해당한다. 이는 모집단 구성원이 표본추출틀에서 한번 추출되면 이는 추출틀에서 본질적으로 제거되고 따라서 다시 추출될 수 없다는 것을 의미한다. 이 책에서 자세히 다루어지지 않는 복원(*with replacement*) SRS은 추출된 후에 선택된 조사단위가 추출틀에 그대로 남아 있어서 다음의 추출에서 다시 선택될 수 있다는 것을 제외하고는 비복원 단순확률표본과 매우 유사하다. 따라서 복원 추출을 사용하는 경우 하나의 표본에 동일한 모집단 단위를 여러 번 반복해서 뽑는 것이 가능하다. 그러나 매우 큰 모집단에서는 동일한 표본 단위를 두 번 뽑을 가능성은 매우 작다. 대부분의 실제 조사 상황에서의 표본추출은 비복원이거나 혹은 비복원을 가정한다.

제비뽑기 방법

이것은 앞에서 우리가 복권의 당첨자를 선택하기 위해 논의했던 방법이다. 제비뽑기 방법에서는 복권 당첨자를 선정하는 방식으로 10명의 이름이 적힌 것들 중 두 개의 표본 구성원을 뽑으면 된다. 10개 모집단 단위의 이름을 종이 조각에 쓰고, 그 종이 조각들을 상자에 넣어 잘 섞는다. 그리고 나서 이름을 보지 말고 두 개의 종이 조각을 상자에서 뽑으면 된다. 비록 손쉽게 이해되는 것은 아니겠지만, 이 방법은 크기가 2인 모든 가능한 표본이 선택될 확률이 같아진다는 SRS의 정의를 만족하게 된다.

난수를 이용한 방법

난수 활용법은 특히 크기가 크고 전산화된 표본추출틀에서 SRS을 선택하는 효과적인 방법이다. 이 방법으로 모집단으로부터 크기 2인 표본을 선택하기 위해서는 우선 모집단 구성요소에 1에서 10까지 번호를 매겨야 한다. 그리고 1에서 10까지 중에서 두 개의 숫자를 난수표를 이용해서 랜덤하게 뽑는다(난수표는 대부분의 기초 통계학 교과서에서 찾을 수 있다). 두 개의 난수와 일치하는 번호를 부여받은 모집단 내의 두 명의 사람들을 표본으로 뽑는다.

모든 가능한 표본 방법

이 방법은 단순확률추출법의 정의에 기초를 둔 것이다. 이 방법은 크기가 10인 모집단으로부터 구성될 수 있는 크기가 2인 모든 가능한 표본을 만드는 것으로부터 시작된다. 크기가 10인 모집단으로부터 비복원 추출할 수 있는 크기가 2인 표본은 모두 45개가 있다. 45개의 표본 중에서 한 개를 랜덤하게 선택한다. 이것은 1부터 45까지 표본들에 번호를 매기고, 1에서 45까지의 숫자 중 하나를 난수표를 이용해 선택함으로써 가능해진다. 이 방법에서는 모든 45개의 가능한 표본들이 표본으로 추출될 확률이 모두 (1/45)로 같다는 것이 명백하다.

방금 제시된 방법이 SRS를 근본적으로 이해하는 데 도움을 준다. 여기서 어떻게 크기가 2인 45개의 표본을 만들었는지 자세하게 생각해 보자. 모집단 단위 P1을 포함하는 표본들의 개수를 계산하는 것부터 시작해 보면, 이 경우 총 9개의 표본이 있다: {P1, P2}, {P1, P3}, ..., {P1, P10}. 다음으로 모집단 단위 P2를 포함하는 표본을 생각해 보자. 단, 이미 계산된 {P1, P2}는 제외한다. 이 경우 총 8개의 표본이 있다: {P2, P3}, {P2, P4}, ..., {P2, P10}. 이런 식으로 계속 계산해 가면, 전에 계산되지 않은 P9가 포함된 표본에 도달할 때까지 P3에 7개, P4에 6개, P5에 5개, 등등의 표본이 있음을 증명할 수 있다. 이미 계산되지 않은 P10을 포함하는 표본은 없다. 따라서 모든 가능한 표본의 총수는 이들의 합이다: $9+8+7+6+5+4+3+2+1$, 즉 45개의 표본이 가능하다. <표 9.2>의 첫 번째 두개 열에 크기가 2인 45개의 이런 표본들을 나열해 둔 것이다.

비록 이것이 명백하지 않을지도 모르지만, 설명된 모든 세 가지 방법들이 본질적으로 동일한 방법이고, 크기가 2인 SRS를 생성할 수 있다. 더욱이 각 방법은 크기가 3이나 4인 또는 다른 크기(크기가 10인 모집단에서는 9까지)의 SRS 표본을 선택하는 데 적용할 수 있도록 쉽게 일반화 할 수 있다. 미리 얘기한 것처럼, 난수 활용법은 대부분의 조사업무에서 SRS 표본을 뽑는 가장 일반적인 방법이다. 하지만 우리는 다음에서 확률표본추출의 개념을 설명하기 위해 모든 가능한 표본 방법을 사용할 것이다.

<표 9.2> 크기 10인 모집단에서 크기 2인 모든 가능한 표본 (단위: 1000)

표본	단위들	표본평균, \bar{y}	\bar{Y} 로부터 편차	제곱편차
1	P1,P2	66.00	-44.00	1936.00
2	P1,P3	77.00	-33.00	1089.00
3	P1,P4	75.00	-35.00	1225.00
4	P1,P5	81.00	-29.00	841.00
5	P1,P6	88.00	-22.00	484.00
6	P1,P7	95.00	-15.00	225.00
7	P1,P8	97.50	-12.50	156.25
8	P1,P9	100.50	-9.50	90.25
9	P1,P10	110.00	0.00	0.00
10	P2,P3	83.00	-27.00	729.00
11	P2,P4	81.00	-29.00	841.00
12	P2,P5	87.00	-23.00	529.00
13	P2,P6	94.00	-16.00	256.00
14	P2,P7	101.00	-9.00	81.00
15	P2,P8	103.50	-6.50	42.25
16	P2,P9	106.50	-3.50	12.25
17	P2,P10	116.00	6.00	36.00
18	P3,P4	92.00	-18.00	324.00
19	P3,P5	98.00	-12.00	144.00
20	P3,P6	105.00	-5.00	25.00
21	P3,P7	112.00	2.00	4.00
22	P3,P8	114.50	4.50	20.25
23	P3,P9	117.50	7.50	56.25
24	P3,P10	127.00	17.00	289.00
25	P4,P5	96.00	-14.00	196.00
26	P4,P6	103.00	-7.00	49.00
27	P4,P7	110.00	0.00	0.00
28	P4,P8	112.50	2.50	6.25
29	P4,P9	115.50	5.50	30.25
30	P4,P10	125.00	15.00	225.00
31	P5,P6	109.00	-1.00	1.00
32	P5,P7	116.00	6.00	36.00
33	P5,P8	118.50	8.50	72.25
34	P5,P9	121.50	11.50	132.25
35	P5,P10	131.00	21.00	441.00
36	P6,P7	123.00	13.00	169.00
37	P6,P8	125.50	15.50	240.25
38	P6,P9	128.50	18.50	342.25
39	P6,P10	138.00	28.00	784.00
40	P7,P8	132.50	22.50	506.25
41	P7,P9	135.50	25.50	650.25
42	P7,P10	145.00	35.00	1225.00
43	P8,P9	138.00	28.00	784.00
44	P8,P10	147.50	37.50	1406.25
45	P9,P10	150.50	40.50	1640.25
	평균	110.00	0.00	408.27

9.3.4 단순확률표본에서 모집단 평균 추정

크기가 n 인 단순확률표본을 선택함으로써 모집단의 평균 소득 같은 모집단 평균에 대한 추정량을 계산하는 방법을 생각할 수 있다. 논의를 단순화하기 위해서 표본으로 선택된 모든 사람의 소득을 관측했다고 가정해 보자(즉, 무응답 때문에 자료의 결측이 없다). 또한 표본에 있는 사람마다 실제 소득이 정확히 조사된다고 가정하자(즉, 측정오차가 없다). 이 장의 뒷부분에서 무응답의 보정 방법과 추정과정에서 발생하는 측정오차의 몇 가지 유형에 대해 논의한다.

표본이 SRS에 의해 선택되어질 때, 매우 좋은 통계적 성질을 갖는 모집단 평균의 추정량은 표본평균이다. 나중에 우리는 이러한 통계적 성질과 표본평균이 좋은 추정량인 이유를 논의할 것이다. 이제 표본평균과 모집단 평균을 계산하는 공식을 쓰는 방법에 대해 생각해 보자. 이를 위해서는 수학적 기호를 필요로 한다. 우리가 정의하는 기호들은 주로 표본추출 관련 문헌에서 흔히 사용되는 것들이므로, 그것들이 의미하는 것이 무엇인지 이해하는 것이 바람직하다.

i 를 표본 내의 i 번째 사람이라고 하고, j 를 모집단에서 j 번째 사람을 나타낸다고 하자. 예를 들어, $i=1,2,\dots,n$ 이 표본 내의 n 개의 조사단위(또는 단위)를 나타내고, $j=1,2,\dots,N$ 이 모집단 내의 N 개의 조사단위를 나타낸다고 하자. 그리스 문자 시그마(즉, Σ)는 합을 나타내는 기호로 그 다음에 있는 값들을 합하라는 의미이다. 우리는 표본자료를 나타내기 위해서 소문자를 사용하고, 모집단 조사단위들의 특성값을 나타내기 위해 대문자를 사용한다. 이런 표시를 사용함으로써, n 개의 표본단위의 y 라는 특성에 대한 표본에서의 관측값이 y_1, y_2, \dots, y_n 으로 나타내어지고, 모집단 단위들의 특성에 대한 값들은 Y_1, Y_2, \dots, Y_n 으로 나타내진다. 예를 들어 $\sum y_i$ 는 n 개의 표본 단위들의 합을, $\sum Y_i$ 는 N 개의 모집단 단위들의 합을 나타낸다. 표본평균은 \bar{y} 로 나타내고, 다음 공식과 같이 정의된다.

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{\sum y_i}{n} \quad (9.1)$$

이 공식은 평균이 표본의 n개 조사단위들의 합을 n으로 나눈 것임을 말한다. 유사하게 모집단 평균은 N개의 모집단 조사단위들의 합을 N으로 나눈 것을 말한다. 수학적으로는 다음과 같이 쓸 수 있다.

$$\bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_N}{N} = \frac{\sum Y_i}{N} \quad (9.2)$$

여기서 \bar{y} 를 \bar{Y} 의 추정량이라고 말한다.

예를 들어, <표 9.1>에서 10명으로 구성된 모집단의 평균소득은 다음과 같다.

$$\bar{Y} = \frac{60,000 + 72,000 + \dots + 160,000}{10} = \frac{1,100,000}{10} = 110,000$$

물론, 실제 조사에서 모집단 평균은 알려져 있지 않고 추정되어야 한다. 이것을 어떻게 할 수 있을지 생각해 보자.

이런 목적으로 이 모집단에서 크기가 2인 단순확률표본을 뽑았는데 선택된 두 조사단위가 j=4, j=8인 모집단 조사단위라고 가정하자. 그러면 모집단의 Y_4 의 값에 따라 첫 번째 표본 조사단위의 값 $y_1 = 90,000$ 이고 두 번째 표본 조사단위의 값은 모집단의 Y_8 의 값에 따라 $y_2 = 135,000$ 이다. 또한 \bar{Y} 의 추정값은 이 두 값들의 평균: $\bar{y} = (\$90,000 + \$135,000)/2$ 또는 $\$112,500$ 이다. 이것은 이 특정한 표본에서 구한 $\bar{Y} = \$110,000$ 에 대한 추정값이다.

\bar{Y} 에 대한 이 추정값의 오차가 $\bar{y} - \bar{Y}$ 또는 $\$2500$ 이다. 이 예에서 비표본추출오차는 전혀 발생하지 않는다고 가정하고 있기 때문에, 이 오차는 오직 표본추출오차에만 기인한 것이다. 그러나 우리가 얻은 표본추출오차는 선택된 표본에 의존한다. 예를 들어, j=4, j=8인 표본 대신에 j=1, j=5인 것을 뽑았다고 가정하자. 그러면 \bar{Y} 의 추정량은 $\bar{y} = (\$60,000 + \$102,000)/2 = \$81,000$ 이다. 이 경우 $\bar{y} - \bar{Y} = \$81,000 - \$110,000 = -\$29,000$ 이므로 표본추출오차는 더 커지고 반대 부호를 가진다. 이를 통해 표본들에 따라 모집단 평균으로부터 다양한 편차를 나타내게 되고, 결과적으로 발생하는 오차가 달라짐을 알 수 있다.

오차를 알기 위해서 우리는 모수의 참값을 필요로 하기 때문에 특정 표본과 관련된 오차의 크기는 실제 알 수 없다. 그러나 다음 장에서 설명될 통계적인 방법을 사용하면 실제 추출된 특정한 표본이 아니라 표본추출 과정(*process*)과 관련된 표

본추출오차를 나타내는 척도를 얻는 것이 가능하다. 이런 척도는 추정량의 표본추출분산(*sampling variance*)이라고 하고, 이는 모집단으로부터 추출할 수 있는 모든 가능한 표본에 대한 표본평균과 모집단 평균으로부터의 편차들을 제공한 것들을 평균한 것이다. 이 척도의 제곱근을 평균 추정량의 표준오차(*standard error*)라고 한다.

9.3.5. 단순확률표본추출에서 통계적 추론

이 절에서 우리는 표본추출오차의 본질과 그것이 추정량의 표준오차와 어떻게 관계되는지에 대해 논의한다. 또한 하나의 특정 표본으로부터 표준오차를 어떻게 측정할 수 있는지, 그 추정값을 어떻게 해석해야 하는지, 이 추정값을 사용해서 모수에 대한 신뢰구간을 어떻게 구할 수 있는지에 대해 논의한다. 아울러 비편향이라는 모수의 추정량에 대한 중요한 성질에 대해 논의한다.

앞에서 언급된 것처럼 비확률표본에 비해서 확률표본의 중요한 장점은 확률표본에서는 표본추정량의 오차 크기에 대한 평가가 가능하다는 것이다. 9.3.4절에서 설명된 것처럼 단일 표본으로부터 얻어진 모집단 모수의 추정값은 모수와 상당히 다를 수 있다는 것이다. 실제로 우리는 추정값이 모집단 모수와 얼마나 가까운지는 모르지만, 모수에 대한 신뢰구간(*confidence interval*)을 계산하는 것은 가능하다. 표본에서 구한 신뢰구간은 추정값이 모수와 얼마나 가까운지에 대한 아이디어를 제공해 준다. 표본에서 관찰한 단위들의 값으로부터 계산된 신뢰구간은 모수를 포함할 것으로 생각되어지는 값들의 범위를 나타낸다. 신뢰구간은 우리가 뽑은 특정한 표본과 관련되어 있다. 예를 들어, 특정 표본으로부터 모집단 평균에 대한 95% 신뢰구간을 계산할 때, 모집단으로부터 뽑을 수 있는 모든 가능한 표본들에서 얻어진 구간의 95%가 모집단 평균을 포함할 것이라는 구간을 계산한 것이다. 즉 우리는 전체 중 95%의 경우 모집단 평균을 포함하는 구간을 산출하는 추론 과정을 수행하는 것이다.

모집단으로부터 추출 가능한 모든 표본 중에서 미리 설정된 비율만큼 모수를 포함할 수 있도록 신뢰구간을 만드는 방법은 통계적인 이론을 통해 구현될 수 있다. 신뢰구간은 모집단에서 추출된 단일 표본으로부터 모수의 참값을 추론할 수 있는 수단을 제공한다.

단일 표본으로 모수를 어떻게 추론하는 지를 이해하는 핵심적인 개념은 표본추출과 추정을 하나의 과정(*process view of sampling and estimation*)으로 보는 데 있다. 이는 표본을 추출하고 추정값을 계산하는 행위 자체를 의미하며 따라서 이과정은 무한히 여러 번 반복될 수 있다는 것을 의미한다. 이 과정이 한번 반복되면 일정한 크기의 새로운 표본이 같은 표본추출 설계에 의해 생성되고, 관심 모수에 대한 추정값을 얻기 위해 주어진 추정량이 계산에 사용된다. 이런 관점으로 표본추출을 본다는 것은 단일 표본에서 추정값을 산출한다는 것이 아니라 표본추출 과정을 가상의 반복수행을 전제로 한 추정 행위로 본다는 것을 의미한다. 모집단에 대한 추론은 표본추출과 추론과정에 대해 통계학적인 이론을 기초로 한다. 추론을 위해 추출된 특정 표본은 과정과 모집단에 대한 정보를 제공해 준다.

표본추출과 추정을 하나의 과정으로 보는 것은 실제에서는 표본을 추출하는 확률을 염두에 둔 것이다. 실제 활용에 있어서는 조사를 위해 하나의 표본만이 추출되고 이 표본에서 자료가 수집된다. 모집단의 일부에 지나지 않는 작은 부분에서 관측된 자료를 기초로 모집단에 대해 추론을 어떻게 할 것인지 이해하기 위해서는 우리는 표본을 표본추출 과정에서 산출될 수 있는 무수히 많은 것 중의 하나로 보는 것이 필요하다. 이런 관점들에 대해서는 다음에 자세하게 다룰 것이다.

표본추출 과정의 고찰

<표 9.1>에 제시된 작은 모집단에 대해 표본추출을 하나의 과정으로 고찰해 보기로 한다. 앞에서 설명한 것과 같이 전체 모집단으로부터 크기가 2인 표본은 모두 45개가 가능하며, 이것들은 모두 <표 9.2>에 나열되어 있다. 이 표의 2번째 칸에는 각 표본에 포함된 모집단 단위들 나열되어 있고, 3번째 칸에는 이 표본에서 얻은 표본평균이다. 예를 들어, <표 9.2>의 첫 번째 표본은 P1과 P2로 이루어져 있다. 이 표본의 평균은 $(60,000+72,000)/2 = 66,000$ (3번째 칸으로부터)이다.

<표 9.2>의 45개 표본은 크기가 2인 모든 가능한 표본으로 구성되어 있기 때문에 만약 어떤 크기가 2인 표본이 추출되면 이 표본은 이것들 중 하나일 것이다. 여기서 우리가 볼 수 있는 것처럼 어떤 표본이 추출되는지에 따라 표본평균은 모수보다 상당히 작을 수도 있고(표본1의 경우 \$66,000), 어떤 표본의 경우 매우 클 수도 있다(표본45의 경우 가장 큰 \$150,500이다). 하지만 이 표본평균들의 평균값은 \$110,000(3열의 마지막 행에서 보듯이)로 모수의 값과 정확히 일치한다. 따라서 이

런 표본추출 과정을 통해 산출되는 평균은 평균적으로는 모집단 평균과 같아진다.

표본 평균의 평균값은 정확히 모평균이고(<표 9.1>의 마지막 줄 2번째 칸에서 보듯이), 이런 결과는 우연히 일어나는 것이 아니다. 이것은 비표본추출 오차가 없을 경우에 단순확률표본추출에서 항상 나타나는 결과이다. 즉, 단순확률추출에서는 모든 가능한 표본에서 얻은 표본평균들의 평균값은 모평균과 일치한다. 통계학적 용어로 이런 표본평균들의 평균을 기대값(*expected value*)이라고 한다. 표본평균의 기대값은 단순확률추출법을 사용하여 표본을 추출하였을 경우 모평균과 같다. 통계 문헌에서는 추정량의 기대값이 목표 모수와 같아지는 추정량의 성질을 비편향성(*unbiasedness*)이라고 한다.

단순확률표본의 평균이 갖는 비편향성은 표본의 크기나 모집단의 크기에 상관없이 항상 성립한다. 그러나 그 과정은 상당히 가변적이며, 이런 가변성 때문에 <표 9.2>의 4째 칸에 있듯이 -44,000에서 40,500 까지 빛나갈 수 있다. 하지만 표본크기가 자꾸 증가하면 할수록 표본평균의 오차는 점점 작아진다. 예를 들어, 표본크기가 10(완전한 센서스)이라면, 표본평균의 오차는 0일 것이다. 그러므로 조사 설계자는 표본크기의 선택을 통해 추정량의 표본추출오차를 조정할 수 있다.

표본추출을 과정으로 보는 것(*process view of sampling*)은 조사를 위해 선택된 표본을 어떤 표본추출과정을 통해 선택될 수 있는 많은 표본 중 하나로 보는 것을 의미한다. 그러면 통계적 추론은 이 단일 표본을 사용하여 목표모집단의 특성(모평균)과 표본추출과정(추정량의 표준오차)의 특징을 기술하는 것이다.

표본추출분포

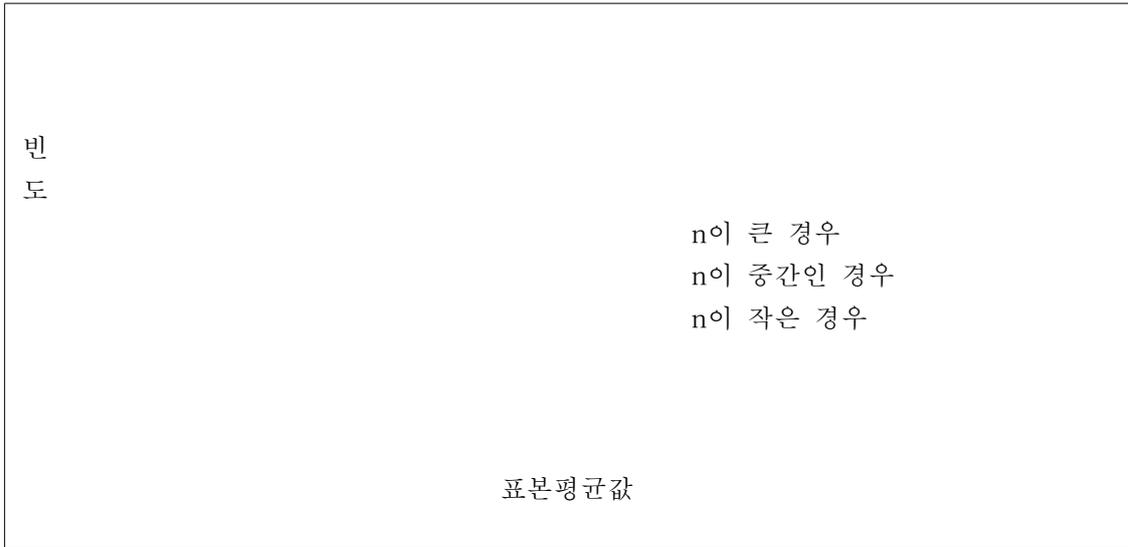
<그림 9.1>은 45개 표본평균에 대한 히스토그램이다(즉, \bar{y} 값의 빈도 그림). 이 히스토그램을 표본평균의 표본추출오차분포라고 한다. 이것은 대략적으로 모평균 110,000를 중심으로 대칭적이다. 이는 통해 표본평균은 평균적으로 모평균과 같다(단순확률표본에서 \bar{y} 는 비편향임)는 우리가 이미 알고 있는 표본평균의 분포를 확인할 수 있다. 이 히스토그램은 어느 정도 종모양이다(가운데는 둥글게 부풀어 올라 있고 양 옆으로 갈수록 점진적으로 꼬리가 내려간다). 표본평균의 표본추출분포가

대칭적인 종모양을 갖는다는 것이 단순확률추출의 또 다른 특징이다.



<그림 9.1> 가상적인 <표 9.1>의 모집단으로부터 크기가 2인 모든 가능한 표본의 표본평균에 대한 표본추출분포

\bar{y} 의 분포는 표본크기와 모집단의 크기와 상관없이 항상 이런 형태를 취한다. 만약 크기가 2인 표본 대신 3, 4 또는 그 이상의 표본크기에 대한 표본추출분포를 살펴보면 모양에 약간의 변화가 생긴다. 중앙은 110,000이 그대로 유지되지만 폭이 좁아진다. 즉, 분포의 꼬리부분이 히스토그램의 중앙으로 끌려 올라가기 시작한다. <그림 9.2>는 큰 모집단에서 표본크기(작음, 중간, 큼)가 증가함에 따라 발생하는 효과를 보여준다. 각 곡선에서 퍼짐의 정도는 표본크기에 따라 변화한다. 표본크기가 증가함에 따라 곡선의 퍼짐이나 산포도는 감소하는 것에 주목하라. 통계적 추론에서는 이런 효과 때문에 표본크기가 증가함에 따라 표본평균이 모평균과 큰 차이가 나는 표본이 추출될 가능성이 줄어들게 된다. 다시 말해서 표본크기가 작은 경우에 비해 표본크기가 큰 경우에는 상대적으로 나쁜 표본들(모평균과 표본의 평균이 크게 다른 경우)이 추출되는 빈도가 상당히 줄어든다는 것을 의미한다.



<그림 9.2> 표본크기가 표본평균의 표본추출분포에 미치는 효과

추정량의 표준오차

표본크기가 증가하는 것이 추정량에 미치는 효과는 추정량의 표준오차로 설명할 수 있다. 표본평균의 표준오차는 표본평균의 표본추출분포의 산포도(흩어짐)를 나타내는 척도이다. 그것은 표본평균의 표본추출분포에서 분산의 제곱근으로 정의된다. 표본평균의 표본추출분포의 분산은 일반적으로 $Var(\bar{y})$ 로 나타내고 표본평균의 분산(*variance of sample mean*)이라고 한다. 이는 표본평균과 모평균의 제곱편차를 모든 가능한 표본에 대해 평균한 것으로 정의된다.

<표 9.2>의 4번째 칸은 크기가 2인 45개 표본에서의 차이 $\bar{y} - \bar{Y}$ 를 나열한 것이다. 5번째 칸은 이들 편차의 제곱을 보여주고 있고, 이 칸의 마지막 줄에는 이들 편차의 제곱들의 평균값인 $Var(\bar{y})$ (즉, $Var(\bar{y}) = 4,082,700$)이다. 따라서 \bar{y} 의 표준오차는 $\sqrt{4,082,700}$ 또는 2021이다. 즉, $Var(\bar{y})$ 의 수학적 공식은 다음과 같다.

$$Var(\bar{y}) = \frac{\sum(\bar{y} - \bar{Y})^2}{M} \tag{9.3}$$

여기서는 \sum 는 모든 가능한 표본에 대한 제곱편차 $(\bar{y} - \bar{Y})^2$ 의 합을 나타내고, M은 가능한 모든 표본의 수이다. 따라서 식(9.3)은 표본평균과 모평균의 제곱편차를 모든 가능한 M개의 표본에서 구해 평균한 것으로 이는 표본평균의 분산과 같다.

이 공식은 모집단으로부터 추출이 가능한 모든 표본의 평균을 알아야하기 때문에 별로 유용한 것은 아니다. 이는 엄청난 수의 표본에 대한 것일 수도 있다. 예를 들어, 크기가 100인 모집단에서 추출이 가능한 크기가 10인 표본의 수는 10조이다! 그러나 약간의 대수학을 사용하면 좀 더 쓸모 있는 다음 공식을 얻을 수 있다.

$$Var(\bar{y}) = (1 - \frac{n}{N}) \frac{1}{n} \frac{\sum (Y_j - \bar{Y})^2}{N-1} \quad (9.4)$$

$$= (1 - \frac{n}{N}) \frac{1}{n} S^2 \quad (9.5)$$

여기서 분자에 있는 \sum 는 제곱편차 $(\bar{y} - \bar{Y})^2$ 의 합으로 모집단 전체 단위에 대한 것이고, 여기서

$$S^2 = \frac{\sum (Y_j - \bar{Y})^2}{N-1} \quad (9.6)$$

는 모집단에서 Y의 분산이다. 한편 <표 9.1>에 있는 자료에서 식(9.3)과 식(9.5)는 동일한 결과(408.270)를 산출한다는 것을 확인하기 바란다.

식(9.4)는 표본평균의 분산은 모집단 평균과 모집단 단위 값의 제곱편차의 합에 대한 “평균”을 구하고, 이를 다시 표본크기로 나눈 후 인자 $(1-n/N)$ 을 곱한 것과 같다. 여기서 따옴표 안의 “평균”은 평균을 구하기 위해 실제 N 대신에 N-1로 나눈 것을 나타내기 위한 것이다.

모든 가능한 표본을 구성하고 이들 표본들의 평균의 분산을 계산하는 것보다 모집단의 값들에 대한 분산을 계산하는 것이 훨씬 간단하기 때문에 식(9.5)의 $Var(\bar{y})$ 에 대한 공식을 식(9.3)보다 선호한다. 또한 n 이 커질수록 $Var(\bar{y})$ 가 작아진다는 사실을 식(9.5)에서 명백하게 볼 수 있다.

식(9.5)의 $(1-n/N)$ 은 유한모집단 수정인자(*finite population correction factor; fpc*)라고 부른다. 모집단이 수백만 명 또는 그 이상이 되는 경우처럼 모집단이 매우 크고, 표본 크기가 수천 명 또는 그 이하인 경우 *fpc*는 대략적으로 1이 되고, 계산상 그 효과가 매우 적기 때문에 이를 무시할 수 있다. 사실 표본추출률 $f=n/N$ 이 0.1보다 작다면 즉, 모집단의 10%이하가 표본이라면, *fpc*를 무시하는 것이 보통 수용된다. 이 경우 실제 계산결과에는 차이가 없는 간단하고 쉬운 공식이 유도된다. 이와 같이 *fpc*를 무시한다면 \bar{y} 의 분산은 간단하게 다음과 같이 표현된다.

$$Var(\bar{y}) = \frac{S^2}{n} \quad (9.7)$$

추정량의 표준오차는 간단하게 분산의 제곱근이고, 즉 \bar{y} 의 표준오차(SE)는

$$SE(\bar{y}) = \sqrt{Var(\bar{y})} \quad (9.8)$$

비록 $Var(\bar{y})$ 와 $SE(\bar{y})$ 가 모두 \bar{y} 의 표본추출분포의 산포도나 흩어짐의 척도이기는 하지만, 표준오차의 측정단위가 관심 특성에 대한 측정단위와 같기 때문에 통계 보고서에서는 표준오차가 주로 사용된다. 예를 들어, 관심 특성이 달러로 표시된 소득이라면, $Var(\bar{y})$ 의 단위는 달러제곱인 반면에 $SE(\bar{y})$ 의 단위는 달러가 된다. 더욱이 $SE(\bar{y})$ 는 신뢰구간을 구하거나 가설검증을 하는 데 사용되는 값이다. 예를 들어, 모평균 \bar{Y} 에 대한 95% 양측신뢰구간은 다음과 같다.

$$[\bar{y} - 2SE(\bar{y}), \bar{y} + 2SE(\bar{y})] \quad (9.9)$$

이와 같은 공식에서 식(9.9)의 2(좀더 정확히 1.96)에 $SE(\bar{y})$ 를 곱하는 것 대신에 정규분포의 $(1-\alpha/2) \times 100$ 백분위수를 사용하면 일반적인 $(1-\alpha) \times 100\%$ 신뢰구간을 구할 수 있다. 이런 백분위수 값들은 일반적인 통계학 관련 교재에서 찾을 수 있다.

이런 신뢰구간의 해석은 중심극한정리(*central limit theorem*)에 의존한다. 이 정리에 따르면, 만약 모집단으로부터 표본을 추출하고 식(9.9)에 의해 구간을 계산하면 그 구간 안에 모평균 \bar{Y} 가 포함된 가능성이 95%가 된다는 것이다. 다시 말하면, 모든 가능한 표본 중 95%의 표본에서 식(9.9)로 구한 신뢰구간에 \bar{Y} 가 포함된다는 의미이다. 본질적으로 표본의 크기가 충분히 크기만 하면, 즉 $n=30$ 이거나 그 이상, 단순확률추출에서 이 정리가 성립하게 된다.

표준오차(*standard error*)와 표준편차(*standard deviation*)는 매우 유사한 용어이고 쉽게 혼동된다. 표준편차는 모집단 분산의 제곱근이다. 예를 들어, 모집단에서 소득의 표준편차는 모집단 단위들의 소득에 대한 산포도 척도이다. 표준오차 역시 분산(즉, 추정량의 표본추출분포의 분산)의 제곱근이라고 정의할 수 있다. 이런 관점에서 보면 표준오차는 추정량의 표본추출분포에서의 표준편차를 말한다. 표준오차는 모집단이 아닌 항상 추정량에 대한 것이라는 점이 핵심적인 차이이다.

9.3.6 표본크기의 결정

원하는 폭을 갖는 신뢰구간을 얻는 데 필요한 표본크기를 결정하기 위해 (9.9)의 95% 신뢰구간 공식을 활용할 수 있다. 일반적인 실제 적용에 있어서는 보통 원하는 추정 오차의 한계를 설정하고 이를 만족하도록 표본크기를 정한다. 추정 오차의 한계(*margin of error*)는 간단히 95% 신뢰구간의 폭을 2로 나눈 것이라고 할 수 있다. 95% 신뢰구간의 폭은 상위 신뢰한계에서 하위 신뢰한계를 뺀 것이다. 만약 (9.9)식에서의 상위신뢰한계에서 하위신뢰한계를 빼면, 95% 신뢰구간의 폭($4 SE(\bar{y})$)을 나타내는 공식을 얻을 수 있다. 따라서 표본평균의 오차한계는 $4 SE(\bar{y})$ 를 2로 나눈 $2 SE(\bar{y})$ 이다.

오차한계를 나타내는 식을 우리가 원하는 오차한계 값으로 놓아 방정식을 만들고 이 방정식을 n 에 대해서 풀면, 원하는 오차한계를 만족시키는 표본크기를 구하는 식을 얻을 수 있다. $SE(\bar{y})$ 에 대한 식을 n 에 대해서 풀면 다음과 같은 표본크기를 결정해 주는 식을 얻을 수 있다.

$$n = \frac{4S^2}{d^2} \quad (9.10)$$

여기서 d 는 원하는 오차한계이다. 만약 fpc 가 1에 가깝다면(즉, 0.9나 그 이상으로 무시할 수 있다면), 식(9.10)은 상당히 정확하다. 그러나 만약 fpc 가 0.9보다 작다면 식(9.10)에서 얻어진 결과는 유한모집단 수정인자를 반영해 보정되어야 한다.

어떻게 보정하는 것인지 살펴보기 위해 우선 n_0 가 식(9.10)으로부터 얻어진 표본 크기라고 하자. 그러면 모집단에 대한 표본의 상대적 크기를 반영한 표본크기는 다음과 같이 구할 수 있다.

$$n = \frac{n_0}{1 + n_0/N} \quad (9.11)$$

즉, 앞에서 구한 표본크기로 표본추출률을 계산하고(이 추출률은 $f_0 = n_0/N$), 앞에서 구한 표본크기를 $1 + f_0$ 를 나눈 것이 보정된 표본크기가 된다.

식(9.10)은 모집단 분산 S^2 을 안다고 가정한 것이다. 실제 이런 경우는 거의 없기 때문에 S^2 의 근사값을 대신 사용해야만 한다. 어떤 경우에는 이전의 조사에서 이

모수에 관한 정보를 얻을 수도 있다. 하지만 많은 경우 S^2 의 값을 잘 파악할 수 없다. 이런 경우에는 S^2 의 구체적인 값을 추측하여 n 을 구하거나 또는 S^2 의 범위를 추측해서 이를 근거로 조사를 위한 n 의 범위를 산출해야 한다. 그러면 이 범위 안에서 모든 상황을 고려했을 때 가장 최상이라고 여겨지는 n 의 값을 선택할 수 있다.

표본크기를 가장 간단히 구할 수 있는 경우는 모집단에서 하나의 비율에 대한 추정에 관심이 있을 때(예를 들어, 자기 집을 소유한 사람들의 비율)이다. 비록 이 비율에 대해 전혀 정보가 없을지라도, 만약 우리가 약간은 보수적인 것(산출된 표본크기가 실제 정확한 모집단 비율을 아는 경우 산출되는 표본크기보다 다소 커지는 것)을 수용한다면 필요로 하는 표본크기를 산출할 수 있다. 이를 위해 우리는 모집단 비율을 대략적으로 0.5나 50%로 가정한다. 만약 이 값을 비율로 사용하면, 산출되는 표본크기는 정확한 경우에 비해 항상 커진다. 이와 같이 연구를 위해 표본크기를 보수적으로 계산할 수 있는 데, 이런 접근법은 특히 모비율에 대한 정보가 전혀 없는 경우 또는 크고 작은 여러 가지 모비율들을 한 조사에서 동시에 추정하고자 하는 경우 매우 유용하다.

식(9.10)을 비율 추정문제에 적절하게 전환하기 위해, 추정 대상 비율은 P 로 이에 반대 비율을 Q (즉 $1-P$)로 나타내면. 다음은 조사 이론에서 잘 알려진 사실이다.

$$S^2 = PQ \quad (9.12)$$

즉, 모집단 분산은 간단하게 $P(1-P)$ 가 된다. 이를 식(9.10)에 넣으면 비율에 대한 다음 공식을 얻게 된다.

$$n = \frac{4PQ}{d^2} \quad (9.13)$$

예를 들어, 만약 $P=0.5$ 라면 $Q=0.5$ 이고 $PQ=0.25$ 이다. 이를 식(9.13)에 대입하면, 원하는 비율 추정에 대한 오차한계가 d 인 경우 필요로 하는 표본크기를 보수적으로 계산할 수 있는 다음과 같은 단순한 식으로 얻을 수 있다.

$$n = \frac{1}{d^2} \quad (9.14)$$

예 9.3.1 알려지지 않은 크기의 모비율을 오차한계가 5%포인트 이하가 되도록 추정하기를 원한다고 가정하자. 즉, P 가 어떤 값일지라도 추정된 비율 p 로 구한 신뢰

구간이 $\pm 5\%$ 포인트가 되기를 원한다. 식(9.14)을 사용하면, $n_0 = 1/(0.05)^2 = 400$ 이다. 모집단 크기가 2000이고 유한모집단수정인자가 $1 - n_0/N = 1 - 400/2000$ 또는 0.80 이라고 가정하자. 이 값이 0.9보다 작으면, 유한모집단 크기를 반영하기 위해 n 의 초기 값을 식(9.11)을 이용해서 조정해야 한다. 이에 따라 최종 표본크기는 $400/(1+0.2) = 333.33$ 이다. 계산된 표본크기의 값이 정수가 아닐 때, 가장 가까운 정수로 올린다. 그러므로 5%포인트 오차한계를 만족하기 위한 표본크기는 334이다. 즉, 만약 334명에 대한 조사를 완료하고 식(9.9)을 사용하여 P 에 대한 95%신뢰구간을 구하면, 신뢰구간의 폭은 5%포인트를 초과하지 않을 것이다.

9.3.7 표본평균의 표준오차 추정

9.3.5절에 표본조사에서의 통계적 추론을 위한 기초적인 몇몇의 공식을 제시했다. 그러나 제시된 추정에 대한 표준오차 공식을 상기해 보면, 이를 사용하기 위해서는 본질적으로 모집단의 모든 단위에 대한 관심 특성에 대한 값이 필요하다. 이 절에서는 하나의 단일 표본으로부터 표준오차를 추정하는 방법을 배운다. 표본추출 이론에서 경이로운 사실은 어떤 다른 모집단에 관련된 지식 없이 단지 단일 표본에서 얻은 정보를 사용하여 통계적 추론을 위해 특정 표본의 유용성을 평가 할 수 있다는 것이다. 이러한 확률표본추출의 중요한 특징은 비확률표본추출법에 비해 커다란 장점이 될 수 있다.

\bar{y} 의 분산과 표준오차의 추정공식은 추정 대상이 되는 모집단의 모수에 대한 공식과 매우 유사하게 보인다(즉, $Var(\bar{y})$ 와 $SE(\bar{y})$ 의 추정 공식은 모집단 공식의 표본 버전에 해당한다). 공식의 주된 차이점은 모집단 값 대신에 표본 값을 사용하는 것이다. 우리는 $Var(\bar{y})$ 의 추정공식을 제시한 후, $SE(\bar{y})$ 에 대한 공식을, 그리고 최종적으로 그것의 적용을 설명하는 예를 제시할 것이다.

추정량의 비편향성에 대한 개념이 모든 가능한 표본에서의 추정값들의 평균이 추정대상 모수와 일치한다는 것을 의미한다는 것을 상기하기 바란다. $Var(\bar{y})$ 의 비편향 추정량은

$$Var(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{s^2}{n} \quad (9.15)$$

여기서

$$s^2 = \frac{\sum (y_j - \bar{y})^2}{n - 1} \quad (9.16)$$

따라서 $SE(\bar{y})$ 의 추정량은 다음과 같다.

$$se(\bar{y}) = \sqrt{var(\bar{y})} \quad (9.17)$$

예 9.3.2 분산과 표준오차의 추정 공식을 어떻게 사용하는지 설명하기 위해 <표 9.1>의 모집단을 다시 이용한다. 그러나 여기서는 크기가 4인 표본(P3, P5, P6, P10의 단위)을 추출한 것으로 가정한다. 식(9.16)을 적용하여 모분산 S^2 의 추정값을 구하기 위해 우선 표본의 평균 \bar{y} 를 계산한다. 4개 표본단위의 값은 각각 94,000, 102,000, 116,000, 160,000이다. \bar{y} 는 이들 4개 값의 평균이므로 118,000이다.

다음으로 4개 표본 값에 대해 각각 차이 $(y_j - \bar{y})$ 의 값을 계산한다. 이 차이를 나타내는 값은 각각 -24,000, -16,000, -2000, 42,000이다. 이 4개 값들의 합이 0인지 확인해 보아라. 다음으로 그 차이 값들을 제곱하고 그것을 합하면 2.6×10^8 이다. 그러면 이 값을 4-1 또는 3으로 나누어 식(9.16)에 의해 추정값 s^2 을 구하면 0.8667×10^8 이다. 또한 (9.15)을 계산하면, $(1-4/10)=0.6$ 을 곱하고 4로 나누어서 $var(\bar{y})$ 의 추정값인 0.13×10^8 을 얻는다. 마지막으로 (9.17)를 계산하면, 이 값의 제곱근인 0.36×10^4 또는 3600을 얻는다. 따라서 $SE(\bar{y})$ 의 추정값은 \$3600이고 \bar{Y} 의 95% 신뢰구간은 $\bar{y} \pm 2 \times 3600$ 또는 [110,800, 125,200]이다.

예제 9.3.2는 공식을 어떻게 사용하는지를 보이기 위한 아주 간단한 예라는 점에 유의하기 바란다. 실제 적용에 있어서는 추정의 정확성이 너무 떨어지기 때문에 단지 표본크기가 4밖에 안되는 표본에서 모평균이나 표준오차를 추정하는 경우는 없을 것이다. 또한 신뢰구간의 타당성은 중심극한정리에 의존하게 된다. 중심극한정리는 표본크기가 4보다 훨씬 큰 경우에 비로소 적용이 될 수 있고, 따라서 이와 같이 표본크기가 작은 경우에는 계산된 신뢰구간의 타당성에 문제가 있다.

9.3.8 단순확률추출에서 \bar{Y} 의 다른 추정량

표본추출오차는 표본설계와 표본크기에 의해 통제될 수 있다. 그러나 표본추출오차를 결정하는 또 다른 요인은 추정량의 선택이다. 단순확률추출에서는 모평균에 대한 자연스러운 추정량이 \bar{y} 이고, 이를 단순확대추정량(*simple expansion estimator*)이라고 한다. 그러나 이 추정량이 단순확률추출에서 \bar{Y} 에 대한 유일한 추정량은 아니다. 사실 가장 좋은 추정량이 아닐 수도 있다. 비추정량, 회귀추정량, 표본 중앙값과 같은 다른 추정량들이 있다. 보통 평균제곱오차(*mean squared error*)를 추정량의 선택 기준으로 사용한다. 이 절에서는 평균, 비율, 총계 등을 추정하는데 있어 단순확대추정량의 대안으로 사용할 수 있는 추정량(예를 들어, 비추정량)에 대해 논의한다. 비추정량(*ratio estimator*)은 표본추출이 적용되는 실제 활용에서 매우 중요하며, 특히 복합 표본설계가 적용되는 경우 가장 중요한 유형의 추정량이다. 여기서 단순확률추출에서 이 추정량의 사용에 대해 논의한다. 복잡한 표본추출에서 이를 확대 적용하는 것은 단순 확장에 지나지 않는다.

여기서는 평균제곱오차에서 비표본추출오차에 의한 영향은 무시한다고 가정하자. 그러면 모집단에 대한 완벽한 센서스를 제외하면, 추정량의 오차는 오직 표본추출에만 기인하게 된다. 추정량의 표본추출분산 그리고 표본추출분산과 표본크기의 관계에 대해 이미 논의했다. 약간 새로운 개념은 추정량의 편향이고, 때때로 이를 추정량의 기술적 편향(*technical bias*)이라고 한다. 이 편향은 비표본추출오차에 의해 발생하는 것이 아니고, 비편향성을 갖지 않는 추정량을 선택함으로써 발생하는 편향을 나타낸다. 비추정량은 작은 기술적 편향을 가진다. 하지만 어떤 경우 단순확대추정량보다 분산이 훨씬 작아지기 때문에 결과적으로 그것의 평균제곱오차가 더 작아지게 된다. 비추정량의 편향에 대한 더 이상의 논의는 이 책의 범위를 벗어나기 때문에 생략한다. 하지만 독자는 표본추출이론에 관한 책들(예를 들어, Lohr, 1999)을 보면, 이런 개념에 대한 논의와 여러 가지 표본추출에서 최량 추정량을 선택하는 방법에 대한 내용을 참고할 수 있다. 이 절에서는 특히 복합 표본설계를 포함한 일반적인 표본추출에서 비추정량이 매우 중요하기 때문에 이에 대해 간단히 논의한다.

비추정량은 표본 및 모집단 단위들에 대한 어떤 보조정보(*auxiliary information*)

가 존재하는 상황에서 사용될 수 있다. 보조정보는 본질적으로 관심을 갖는 특성 y 와 상관관계가 있는 표본 단위들에 대한 또 다른 특성이다.

예를 들어 모집단 모든 단위들에 대한 5년 전 소득을 알고 있다고 가정하고, 이 모집단의 현재 평균소득을 추정하기를 원한다고 하자. 크기가 n 인 SRS를 추출하고 각 표본 구성원의 현재 소득을 조사했다. 5년 전 자료를 기초로 구한 표본의 평균 소득을 \bar{x} 로, 현행 조사에서 얻은 현재 평균 소득을 \bar{y} 로 표시하자. 마지막으로 5년 전 자료를 기준으로 한 전체 모집단의 평균소득을 \bar{X} 로 나타낸다. 여기서 현재 조사 이외의 다른 출처로부터 얻어진 이전의 개인 소득 자료가 보조자료(*auxiliary data*)에 해당한다. 그러면 비추정량은 다음과 같이 정의된다.

$$\bar{y}_R = \frac{\bar{y}}{\bar{x}} \bar{X} \quad (9.18)$$

이전의 소득 X 는 현재 소득 Y 와 매우 높은 상관관계를 갖고 있을 것으로 예상되므로 비추정량 \bar{y}_R 은 단순확대추정량 \bar{y} 보다 훨씬 더 정확한 \bar{Y} 의 추정량일 것이다. 이것의 증명과 단순확대추정량과 비추정량의 비교 결과에 대해서는 표본추출이론 교재(예를 들어, Cochran, 1977)에서 찾을 수 있다.

다음 절에서처럼 표본추출오차는 조사 설계의 선택뿐만 아니라 추정량의 선택에 의해 영향을 받는다. 예를 들어, 단순확률추출처럼 어떤 방법에 의해 모집단으로부터 표본을 선택했다고 가정하자. 자료가 수집되면 통계학자는 관심모수를 어떤 추정량을 사용하여 추정할 지에 대해 결정해야 한다. 예를 들어, 통계학자가 모집단에서 평균소득 추정을 고려한다고 가정하자. 그는 보조자료로 무엇이 사용 가능할지에 따라 여러 가지의 추정량을 생각할 수 있다. 보조자료가 가능하지 않다면 모평균의 추정량으로 일반적인 단순확대추정량 또는 표본평균을 사용할 것이다. 하지만 만약 이전의 센서스나 조사와 관련된 외부 자료를 활용해 총소득 또는 평균소득에 대한 정보를 갖고 있다면 좀더 정확한 추정량이 구현될 수 있을 것이다.

9.4 비표본추출오차가 존재하는 경우 통계적 추론

9.3절에 제시된 통계적인 결과들은 비표본추출오차에 발생하지 않는다는 가정에서 성립한다. 고전적인 표본추출 이론에서는 완벽한 표본추출틀에서 표본을 추출하

고, 모든 표본 단위에서 정확한 응답을 얻을 수 있으며, 자료처리과정에서 오차가 생기지 않는다는 것을 전제로 한다. 비론 이런 가정이 실제 응용에서는 비현실적일 지라도 통계적인 이론들을 단순화 하는데 도움이 된다. 많은 경우 이런 “이상적인 세계”에서 얻어지는 통계적인 결과들이 개략적으로 “현실 세계”에서도 자주 적용되고 있다. 하지만 이상적인 가정을 전제로 한 표본추출이론이 실전에서는 거의 성립하지 않는다는 것을 인식하는 것이 매우 중요하다. 더욱이 실제 조사에서 피할 수 없는 비표본추출오차는 고전적인 표본추출이론을 위한 가정하에서 유도된 결과들이 더 이상 유효하지 않도록 표본추출분포를 왜곡시킬 수 있다. 이 절에서는 비표본추출오차가 통계학적인 추론에 미치는 영향을 예시하고, 비표본추출오차가 고전적인 조사 추론에서 성립하던 결과들을 어떻게 타당하지 않게 만드는지 보여줄 것이다.

비표본추출오차가 존재할 때 조사에서 통계적 추론의 기초적인 아이디어에 대해 살펴보기 위해 크기가 $N=10$ 인 <표 9.1>의 모집단을 다시 사용하자. 비표본추출오차를 도입하기 위해 10개의 단위 중 오직 8개의 단위만 추출틀에 있다고 가정하자: 예를 들어, 추출틀에서 P1과 P2가 손실되었다고 가정한다. 비록 P1과 P2가 손실되었다고 하더라도 여전히 이것들은 목표모집단에 속해 있고 적합한 연구 대상에 해당한다. 이는 3장에서 논의한 개념으로 보면 추출틀 누락에 해당한다. 아울러 소득 보고 과정에서 측정오차가 발생한다고 가정한다. 즉, 만약 단위 j 가 표본으로 추출되면, 모집단 j 단위의 경우 실제 소득 Y_j 대신에 약간 오류가 있는 보고 소득 X_j 가 기록된다고 가정하자.

<표 9.3>은 이 모집단에서 추출틀과 보고 정확성에 대한 가정에 따른 상황을 요약한 것이다. 이 표에서 두 번째 열은 <표 9.1>의 것을 그대로 옮긴 것이고, 세 번째 열은 만약 표본에 추출되는 경우 해당 구성원이 보고할 소득을 나열한 것이다. P1과 P2가 누락되었기 때문에 추출틀은 오직 P3에서 P10까지 $N=8$ 단위로 구성되었다는 점에 유의하기 바란다. 이 자료를 이용하여 우리는 $N=8$ 단위들로 구성된 추출틀에서 표본크기가 2인 표본들을 추출했을 때, 이들 표본의 표본평균에 대한 표본추출분포에 대해 살펴볼 것이다.

<표 9.3> 10명의 소득으로 구성된 소규모 모집단

(단위: 1000)

모집단 구성원	실제 소득	보고 소득
P1	$Y_1 = 60$	missing
P2	$Y_2 = 72$	missing
P3	$Y_3 = 94$	$X_3 = 80$
P4	$Y_4 = 90$	$X_4 = 76$
P5	$Y_5 = 102$	$X_5 = 90$
P6	$Y_6 = 116$	$X_6 = 100$
P7	$Y_7 = 130$	$X_7 = 132$
P8	$Y_8 = 135$	$X_8 = 142$
P9	$Y_9 = 141$	$X_9 = 157$
P10	$Y_{10} = 160$	$X_{10} = 135$
평균	110	114

크기가 8인 모집단으로부터 추출이 가능한 크기가 2인 표본의 수는 모두 28개이다. 이를 살펴보면 P3가 포함된 것은 7개({P3, P4}, {P3, P5}, ... , {P3, P10}), P4가 포함된 것 중 이미 고려하지 않은 표본의 개수는 6개, P5가 포함된 것은 이전에 고려하지 않는 것이 5개 등등이다. 따라서 전체 표본의 개수는 $7+6+5+4+3+2+1=28$ 이다. 이 28개의 각 표본에서 <표 9.2>의 Y 값을 사용하는 것 대신 <표 9.3>의 X 값을 사용하여 표본평균, $\bar{x} = (x_1 + x_2)/2$ 을 계산할 수 있다. 그 다음에 28개 \bar{x} 값들의 빈도분포를 구하고 이를 45개 \bar{y} 값들에 대한 빈도분포와 비교해 볼 수 있다. 이런 비교를 통해 실제 소득과 보고 소득에 대한 표본평균들의 표본추출분포를 비교 할 수 있다. 이 비교 결과는 <그림 9.3>에 제시되어 있다.

실제 소득을 사용할 때, 45개 모든 가능한 표본에서 얻은 표본평균들의 평균(즉, \bar{y} 의 기대값)은 모평균 \bar{Y} 와 같은 110,000이었다는 점에 유의하라. 단순확률추출에서 이런 표본평균의 성질을 비편향성(*unbiasedness*)이라고 한다. 하지만 비편향성은 비표본추출오차가 있으면 더 이상 성립하지 않는다. 만약 모든 가능한 표본의 평균으로 <표 9.3>의 X 값을 사용한다면, 이들 28개 표본평균 \bar{x} 의 평균은 114,000이고, 이는 <표 9.3>에서 \bar{X} 로 표시되는 X 값들에 대한 모집단 평균과 같다. 이들 두 기댓값의 차이가 비표본추출오차 때문에 발생하는 편향이다. 즉,

$$Bias(\bar{x}) = \bar{X} - \bar{Y} \quad (9.19)$$

또는 114,000-110,000=4000



<그림 9.3> <표 9.3>의 가상적인 모집단에서 크기 2인 표본의 실제소득과 보고소득의 표본평균에 대한 표본추출분포의 비교. 보고 소득에 대한 표본평균의 분포가 약간 오른쪽으로 이동한 것으로 보아 모평균 추정에 있어 편향의 발생 가능성을 보여준다.

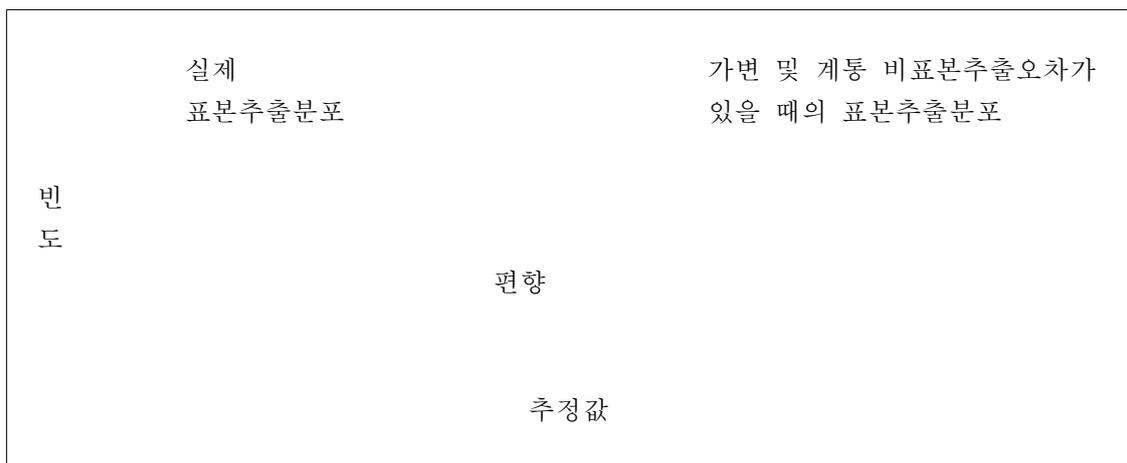
4000이라는 편향은 매우 크게 보이지만, 실제적으로 추정대상 모수의 값에 대한 상대비율($4000/114,000=0.035$ 또는 \bar{X} 의 3.5%)을 생각하면 그리 크지 않다. 상대편향(*relative bias*)이라고 부르는 3.5%라는 값은 모수에 대한 편향의 비를 나타낸다. 편향이 큰 것인지 또는 작은 것인지를 결정하는 일반적인 규칙은 없다. 일반적으로 추정량의 표준오차보다 훨씬 큰 편향이 발생하면 일반적으로 편향이 크다고 볼 수 있고, 표준오차에 비해 작으면 편향은 작다고 본다. 상대편향과 마찬가지로 표준오차도 모수에 대한 상대비율로 나타낼 수 있다. 이 값을 상대표준오차(*relative standard error*) 또는 변동계수(*coefficient of variation; CV*)라고 흔히 부른다. 따라서 만약 \bar{y} 의 CV가 10%라고 하면, 3.5%의 상대편향은 작은 편향이라고 볼 수 있다.

\bar{y} 의 편향은 <그림 9.3>에 잘 나타나 있다. \bar{x} 의 표본추출분포는 \bar{y} 의 표본추출분포

에 비해 약간 오른쪽으로 이동해 있다. 이는 \bar{x} 의 분포는 이것의 평균인 114,000를 중심으로 하고, \bar{y} 의 분포는 이것의 평균인 110,000을 중심으로 하기 때문이다. 이와 관련된 상황은 <그림 9.4>에서 명확히 볼 수 있으며, 이에 대해서는 차후에 좀 더 자세한 논의 한다. 두 그림 모두 비표본추출오차가 표본평균에 미치는 영향을 보여주고 있다. 비표본추출오차의 존재는 분포를 더 평평하게(더 흩어짐) 만드는 데, 이는 자료에 가변오차가 추가됨으로써 분산이 증가된다는 것을 보여준다. 오른쪽으로 분포가 움직이는 것은 또한 계통오차를 보여주고 있는 데, 이 예에서는 이 때문에 표본평균의 추정에 있어서 양의 편향이 발생하고 있다.

실제 상황에서 편향의 정도를 알고 있다면, 간단하게 \bar{x} 로부터 편향을 뺌으로써 \bar{x} 의 편향을 보정할 수 있다. 여기서 $(\bar{x} - 4000)$ 의 표본추출분포는 분포를 왼쪽을 이동시켜 중심이 110,000 또는 \bar{Y} 와 일치하게 해 준다. 불행히도 실제 상황에서는 $Bias(\bar{x})$ 는 거의 알 수 없기 때문에 리너 편향의 보정은 불가능하다.

일반적으로, 비표본추출오차는 모수의 추정값을 편향시킬 수 있기 때문에 모수에 대한 추론 결과는 더 이상 유효하지 않다. 비표본추출오차는 표본추출분포의 모양을 변화시킬 수도 있으며, 이에 따라 표본추출분포의 위치를 이동시켜 편향에 영향을 줄 뿐만 아니라 분산과 분산추정에도 영향을 줄 수 있다. 일반적으로 이런 영향은 조사 분석자가 파악하기 힘들기 때문에 분석과정에서 고려되지 못 한다.



<그림 9.4> 비표본추출오차가 있을 때, 추정량의 표본추출분포에서 발생하는 편향과 분산의 증가

<그림 9.4>에서 비표본추출오차가 표본평균의 분산과 편향에 미치는 잠재적인 영향을 볼 수 있다. 왼쪽의 종모양의 분포는 완벽하고 정확한 모집단 자료를 기초로 한 올바른 표본추출분포이고, 반면에 오른쪽의 것은 비표본추출오차로 왜곡된 자료에 기초한 표본추출분포이다. 올바른 표본추출분포의 경우, 추정량의 기댓값은 모수의 참값과 일치한다. 하지만 왜곡된 분포에서는 분포의 중심이 실제 올바른 분포의 중심에서 벗어나게 된다(즉, 비표본추출오차가 있기 때문에 추정량의 기댓값이 모수의 참값과 다르다). 이 두개의 기댓값의 차이가 비표본추출오차에 의한 편향이다.

<그림 9.4>는 표본추출분포의 퍼짐이나 산포도도 비표본추출오차에 의해 영향을 받는다는 것을 보여준다. 여기서 비표본추출오차가 있는 표본추출분포의 산포도는 비표본추출오차가 없는 경우보다 크다. 표본평균의 표준오차는 표본추출분포의 산포도와 직접 연관되어 있기 때문에 비표본추출오차가 있는 표본평균의 표준오차는 비표본추출오차가 없는 경우보다 커진다. 이와 같은 왜곡된 표본추출 분산추정과 추정량의 편향이 통계적 추론에 영향을 줄 수 있다. 예를 들어, (9.9)에 따라 95% 신뢰구간을 구하면 실제 포함 비율은 95%보다 훨씬 적어진다. 즉, 비표본추출오차 때문에 생기는 표본추출분포의 왜곡의 결과로 정상적인 방법으로 구한 95% 신뢰구간은 실제적으로는 85% 또는 70% 신뢰구간이 될 수가 있다. 불행히도 조사 자료에 대한 통계적 추론에 있어서 이런 사실은 거의 반영되지 않고 있다.

비표본추출오차에 의한 표본추출분포의 왜곡에 따라 통계적 추론에서 발생하는 문제를 보정하는 방법은 별로 없다. 예를 들어, 이미 논의 되었던 것처럼 편향에 대한 좋은 추정이 가능하다면, 편향을 보정하는 데 사용될 수 있다. 하지만, 편향 추정량 자체도 비표본추출오차와 표본추출오차의 영향을 받기 때문에 이런 보정방법을 적용할 때에는 상당한 주의를 요한다. 그래서 평균제곱오차를 기준으로 보았을 때 이와 같이 보정된 추정량이 원래의 보정 안 된 추정량에 비해 좋지 않거나 심지어 더 나쁠 수도 있다.

몇 가지 표준오차 추정방법은 표본추출분포에 포함된 표본추출 및 비표본추출 분산을 성공적으로 반영할 수 있다. 예를 들어, (9.17)의 표준오차의 추정량은 응답자에 의해 생기는 측정오차와 같이 서로 연관이 없는 비표본추출오차는 잘 반영할 수 있다. 하지만 면접원에 의한 분산과 같은 서로 연관된 오차는 잘 반영할 수 없기

때문에 결과적으로 표준오차 추정값은 실제 분산을 과소추정하게 된다. Wolter(1985)는 서로 연관 안 된 비표본추출오차뿐만 아니라 연관된 비표본추출오차를 반영할 수 있는 다수의 분산 추정방법에 대해 논의하고 있다.

관련된 오차를 완전히 성공적으로 보정할 수 있는 비표본추출오차 처리 방법은 없다. 각 조정 방법은 하나의 오차 발생 원인만을 주로 염두에 두고 있고, 이에 따라 어떤 보정 방법을 적용함에 따라 다른 유형의 오차가 새로 발생할 수도 있다. 비표본추출오차를 처리하는 가장 좋은 방법은 사후에 오차를 보정하는 것 보다 오차의 발생을 사전에 방지하는 것이다. 물론, 이것이 바로 이 책을 저술하는 동기이다! 아직 비표본추출오차를 완전히 피하는 것은 불가능하지만, 이런 오차들이 통계적 추론에 미치는 영향과 결과를 이해하는 것이 매우 가치 있는 일이다.

예제 9.4.1 많은 RDD 조사에서 우선 층화단순확률추출로 가구들의 표본을 1차적으로 추출하고, 각 표본 가구에서 조사대상 적격 가구구성원 중 랜덤하게 한 명을 추출하여 면접을 한다. 여기서 랜덤하게 한 명의 가구원을 선정하는 가장 정확한 방법은 Kish의 명부방법(*Kish roster method*)이다. 이 방법은 모든 적격자의 명부를 만든 후 난수생성 프로그램을 사용하여 명부에서 랜덤하게 한 명을 추출하는 것이다. 어떤 연구자들은 많은 경우 낮은 사람에게 함께 거주하는 모든 가구구성원에 관한 정보를 전화로 가르쳐주기 것을 꺼리기 때문에 이 방법을 적용하면 무응답이 많이 발생할 수 있다고 주장한다. 이들 연구자들은 응답자 선택을 위해 최근 생일 방법(*last birthday method*)의 사용을 권장하고 있다.

최근 생일 방법을 실행하기 위해서는 면접자가 가족구성원 중 적격자의 명부를 파악할 필요가 없다. 대신에 면접원은 단지 “18세 이상의 가구구성원 중 가장 최근에 생일이 지난 사람과 통화할 수 있습니까?” 라고 물으면 된다. 그러므로 이 방법은 주제넘지 않으면서 상당히 간결하고, 시간을 저축하면서 응답 부담을 줄일 수 있다. 어떤 연구에서는 이 방법이 Kish의 보다 과학적인 방법과 거의 차이가 없는 좋은 방법이라고 주장하고 있다. 그러나 다른 연구에서는 이 방법이 여성과 젊은 성인쪽으로 표본을 치우치게 하는 방법이라는 결론을 제시하고 있다. 한 연구(Lavrakas 등, 2000)에서는 이 방법을 적용하면 20%의 가구에서 잘못된 사람을 추출하게 된다고 것을 보이고 있다. 특히 저학력 가구나 대규모 가구에서 이런 현상이 많이 나타난다. 더욱이 이 방법에 대한 응답자들의 반응은 일정치 않지만 많은

경우 이 방법이 신뢰할만하고 과학적이라고 생각하지 않는다. 이런 이유로 우리는 엄격한 과학적인 조사에서는 이 방법을 사용하지 말 것을 권한다.

비표본추출오차(*nonsampling error*)는 추정량의 표본추출분포를 왜곡하기 때문에 고전적인 추론 방법은 더 이상 완벽하게 타당한 것이 아니다. 예를 들어, 계통오차는 추정에 편향이 생기게 하고, 결과적으로, 명목상의 95% 신뢰구간은 실제로는 훨씬 더 낮은 유효 신뢰수준을 가질 수 있다. 또한 가변오차는 회귀계수, 연관관계수, 그리고 다른 비선형추정량에 대해 편향이 생기게 하여 결과적으로 표본추출분포를 왜곡시키는 동시에 신뢰성을 떨어뜨린다.

9.5 다른 확률표본추출방법

단순확률표본추출은 완벽하고 정확한 추출틀의 사용이 가능하고, 표본 단위들에 대한 자료 수집을 위해 많은 이동을 필요로 하지 않는 경우 적합한 추출방법이다. 예를 들어, 전문가 단체의 회원을 대상으로 한 우편조사 또는 RDD 전화조사 같은 경우 표본 단위들간의 거리가 면접 비용에 아무런 영향을 주기 않기 때문에 단순확률표본추출의 사용이 바람직할 수 있다. 하지만 표본 단위들이 있는 곳을 돌아 다니면서 조사해 해야 하는 경우 이 방법은 적용상 문제가 있다. 예를 들어, 대면 면접조사(face-to-face)에서는 면접원들을 표본 단위가 있는 곳으로 보내 자료를 수집하는 비용이 전체 자료 수집 비용의 대부분을 차지하게 된다. 미국이나 캐나다처럼 커다란 국가에서 모든 가구에 대해 단순확률표본추출을 적용하면 하나의 가구를 면접하기 위해 면접원이 수백마일이 넘는 먼 거리를 여행해야 하는 경우도 발생할 수 있기 때문에 막대한 여행비용을 지출해야 할 것이다. 더욱이 조사의 표본추출을 위해 국가의 전체 인구를 포함하는 완벽한 추출틀을 만들어야 한다. 어떤 국가(특히, 스칸디나비아 국가들)에서는 전국 인구등록부를 사용하여 상당히 경제적으로 이런 작업을 할 수 있다. 하지만, 미국을 비롯한 여러 다른 나라에서는 국가의 모든 가구를 포함한 완벽한 명단을 모으기 위해서는 국가차원의 센서스 수준의 노력이 요구되기 때문에 이런 작업은 엄청난 비용을 필요로 한다.

모집단 조사단위에 대한 완벽한 목록이 없는 경우, 단위들의 집락을 추출함으로써 현장 조사비용을 줄일 수 있는 표본을 추출하는 방법이 보다 효율적이다. 또한 단순확률추출이 가능한 경우라 하더라도 보다 적용이 간단하거나 보다 정확한 추정이 가능한 다른 표본추출방법이 있다면 이를 이용하는 것이 바람직할 것이다.

이 절에서는 조사 연구에서 일반적으로 사용되는 표본추출방법들에 대한 간단한 개요를 제시하고자 한다. 우리의 목적은 이런 방법들이 무엇이고 또한 이런 방법들을 어떤 경우에 활용하는 것이 바람직한지에 대한 지침을 제시하고자 한다. 여기서는 이런 방법들에 대한 자세한 사항들을 다루는 것을 원하지 않는다. 이런 내용들을 심층적으로 다루고 있는 표본추출에 관한 교재들이 많이 있으니 이들을 참고하기 바란다(예를 들어, Cochran, 1977).

9.5.1 계통 표본추출

모집단 목록에서 물리적으로 표본으로 추출해야 하는 단순확률추출은 상당히 번잡한 작업을 필요로 할 수 있다. 예를 들어, 난수를 이용한 방법을 사용하여 5000명의 명단에서 500명을 추출한다면, 500개의 난수를 생성하는 것이 필요하며 그리고 나서 명단에서 500명의 위치를 일일이 파악해서 추출해야 하므로 상당히 귀찮은 손이 많이 가는 작업을 수행해야 한다. 훨씬 간단한 방법으로 명단에서 모든 10번째 사람을 선택하는 방법을 사용할 수 있다. 일정한 간격을 두고 명단에서 추출하는 것이기 때문에 이런 방식으로 추출하는 것이 훨씬 손쉽다. 만약 단위들이 명단에 상당 수준 랜덤하게 정렬되어 있다면 이런 추출방식은 본질적으로 단순확률추출법과 거의 동일하다.

5000으로부터 500개의 단위를 표본으로 선택하기 위해서는 우선 “매 몇 번째” 또는 뛰는 간격을 먼저 계산해야 하는데, 이는 본질적으로 추출률의 역수에 해당한다. 여기서 추출률은 $500/5000$ 또는 0.1 이므로 명단에서 매 10번째 단위를 선택하면 500명의 이름으로 구성된 계통표본을 얻게 된다. 다음 단계는 “시작하는” 번호 또는 랜덤 출발점을 정하는 것이다. 이것은 1과 “매 몇 번째”(이 예에서는 10이다)에 해당하는 숫자 사이의 난수이다. 예를 들어 난수가 5라고 가정하자. 그러면 계통표본은 5번째, $5+10=15$ 번째, $5+20=25$ 번째, ..., $5+499 \times 10=4995$ 번째까지에 해당하는 이름들로 구성된다. 이 명단에서 크기가 500인 추출 가능한 표본은 오직 10개라는 점에 유의

하기 바란다. 이런 방식으로 하면 랜덤 출발점의 선택에 따라 10개의 가능한 표본 중 어떤 특정 표본이 추출될 것이지가 결정되게 된다.

엄격하게 말하면 계통표본추출(systematic sampling)은 목록이 랜덤한 순서로 정렬되어 있는 특별한 경우를 제외하고는 단순확률추출과 동등하지는 않다. 그런 경우는 실전에서 별로 흔하지는 않지만, 많은 경우 목록이 랜덤한 순서로 정렬된 것으로 취급되는 경우가 많다. 예를 들어, 이름이 알파벳순서대로 정렬된 목록이라면, 조사에서 관심 대상이 되는 특성과 목록의 순서는 전혀 상관관계가 없다고 볼 수 있다. 또는 조사 관심 특성과는 전혀 관계가 없는 어떤 신분을 나타내는 번호에 따라 정렬된 목록도 마찬가지일 것이다. 이런 경우, 계통추출을 단순확률추출처럼 취급하는 것이 정당화 될 수 있다.

그러나 계통추출의 주된 이점 중 하나는 인위적인 순서에 따라 정렬된 목록에서 표본을 뽑는 것에서 찾을 수 있다. 관심 특성과 관련이 있는 변수로 우선 목록을 정렬한 후에 계통추출을 하면 추정의 정도를 향상시킬 수 있는 아주 강력한 기법이 될 수 있다. 이런 방법은 다음 절에서 논의되는 층화확률추출법과도 관련이 있다. 실제로 목록을 잘 설계된 방식으로 순서대로 정렬한 후에 뽑는 계통표본추출은 비록 동일한 추정량을 사용하더라도 단순확률추출에 비해 통계적 정확도라는 측면에서 훨씬 더 효과적인 표본을 추출할 수 있다.

관심 특성에 있어서 거의 동일한 값을 갖는 단위들이 목록상에 인접하도록 하는 것이 정렬을 하는 목적이다. 이를 통해 목록상에서 먼저 일종의 그룹화 또는 층화를 하고 목록에서 골고루 표본을 추출함으로써 가능하면 서로 다른 단위들이 표본으로 선정되도록 하는 것이다. 이를 다시 표현하면 하나의 계통표본에서 얻어지는 추정값은 동일한 표본크기의 다른 계통표본들에서 얻어지는 추정값들과 거의 같아지도록 하기 위해 목록을 정렬한다고 볼 수 있다.

정렬이 되어 있는 않은 목록에서 계통표본추출함으로써 발생하는 위험 중 하나는 그 목록이 랜덤하게 정렬되어 있기 않기 때문에 발생하는 추정의 정확도 감소 문제이다. 예를 들어, 여성에게는 홀수 주민등록번호가 남성에게는 짝수 주민등록번호가 배정되어 있는 상황에서 목록이 주민등록번호에 의해 나열되어 있는 경우를 가정해 보자. 이런 경우 제대로 정렬하지 않고 추출된 계통표본은 모두 남성이나 모두 여성만을 포함한 표본일 수 있다. 이를 방지하기 위해서는 주민등록번호 이외의 파일상의 다른 변수를 이용한 정렬이 필요하다.

계통표본추출(systematic sampling)은 가장 간단한 표본추출방법 중 하나이다. 이것은 관심 특성과 관련 있는 기준에 따라 모집단 추출틀상의 모든 단위들을 정렬하는 것에서 시작된다. 그런 다음 목록에 있는 N 개의 단위 중 근사적으로 n 개의 단위를 표본으로 선택하기 위해 먼저 목록에서 r 번째 단위를 선택한 후 그 다음에 k 번째에 해당하는 단위를 표본으로 선정한다. 여기서 r 은 1과 k 사이의 난수이고 k 는 N/n 에 가장 가까운 자연수이다.

9.5.2 층화표본추출

흔히 표본 추출틀에는 표본추출에 유용하게 활용될 수 있는 모집단 단위들에 대한 특성이 포함되어 있다. 9.5.1절에서 보았듯이 추출틀의 정보는 조사 관심 특성 측면에서 성격이 유사한 단위들을 그룹화하여 계통추출하는데 활용될 수 있다. 그러나 층화추출은 이 정보를 본질적으로는 이와 유사한 작업을 수행하는데 사용하는 또 다른 추출법이다. 층화추출의 목적은 성격이 유사한 단위들로 하나의 그룹 또는 층을 구성하는 것이다. 그런 다음 조사를 위한 표본을 뽑기 위해 각층에서 독립적으로 표본을 추출한다.

층화 추출에서는 층화추출하지 않을 때보다 훨씬 더 정확한 전체 모집단에 대한 추정량을 산출할 수 있다. 예를 들어, 추출 대상 모집단이 흑인 25%, 백인 60%, 기타 15%로 구성되어있다고 가정하자. 이 모집단에서 단순확률추출로 표본을 추출하면 모집단에서의 인종 구성비율과 매우 다른 표본이 나올 수도 있다. 예를 들어, 표본에서 5%만이 기타 인종 범주에 해당할 수도 있고, 혹은 표본에 백인 75%, 흑인 15%이 포함될 수도 있다. 인종에 상관없이 모집단으로부터 랜덤하게 표본추출 했기 때문에 이런 표본이 가능하다. 단순확률추출법을 적용함에 따른 표본들간의 표본 구성비율에 있어서의 변동 때문에 표본추출분포의 변동이 매우 커질 수 있으며, 결과적으로 추정의 정확성을 감소시키게 된다.

그러나, 만약 모집단에서 인종의 구성비율이 표본에 반영된다면, 각 표본에서 다르게 나타나는 이런 인종 구성비율의 차이에 따른 문제를 해결할 수 있다. 이를 위

해 3개의 층을 구성하고, 흑인 층, 백인 층, 기타 인종 층. 그런 다음 흑인 층에서 25%를 뽑고, 백인 층에서 60%, 기타 인종 층에서 15%뽑자. 이렇게 하면 우리가 추출하는 모든 표본들의 인종 구성비는 정확하게 모집단의 인종 구성비와 일치할 것이다. 이를 통해 표본들간의 변동을 줄이고, 결과적으로 추정량의 표본추출분산을 줄일 수 있게 된다. 이런 형식의 표본추출방법을 비례배분에 의한 층화확률추출(*stratified random sampling with proportional allocation*)이라고 한다. 각 층에서 추출(배분)되는 표본크기는 표본추출틀에서 층의 크기에 비례한다. 비율배분에 의한 층화확률추출에서 1000개의 표본을 뽑는다고 하면, 흑인 층에서 250명, 백인 층에서 600명, 기타 인종 층에서 150명을 각각 단순확률추출로 뽑는 것이다.

그러므로 단순확률 표본설계에 내재되어 있는 표본추출에 따른 변동이 비례배분에 의한 층화추출을 적용함으로써 줄어들 수 있다. 이것은 층화 변수와 상관관계가 있는 특성에 대한 추정량의 정확도에 상당한 향상이 있을 수 있다는 것을 의미한다. 더욱이 단순확률추출에서 사용된 것과 동일한 추정량이 비례배분에 의한 층화추출에서도 사용되기 때문에 평균, 총계, 비율을 추정하는 데 있어 더 복잡해지지 않는다. 하지만 추정량의 표준오차에 대한 추정에 있어서는 단순확률추출보다 좀 더 복잡해진다. 한편 만약 각 층에 대한 표본 배분을 비례적으로 하지 않으면 모수 추정식과 표준오차 추정식이 모두 단순확률추출보다 훨씬 복잡하다.

층화가 함축된 계통표본추출(*systematic sampling with implicit stratification*)이라고 알려진 방법을 사용하면 비례배분 층화확률추출을 통해 계통추출에서 얻을 수 있는 정확성 향상과 거의 동일한 효과를 얻는 것이 가능하다. 이 방법은 각 층에서 독립적인 표본추출을 하는 대신, 목록을 층에 따라 정렬한 후에 전체 목록을 대상으로 하나의 랜덤 출발점을 사용하여 계통추출법을 적용하는 것이다. 그 결과로 얻어지는 표본은 비례배분에 의한 결과와 거의 같다. 위 예에서 이런 방법으로 얻어진 표본은 정확하게 흑인 250명, 백인 600명, 기타 인종 150명을 포함할 것이다. 다른 예에서는 비례배분의 경우와 많아야 1 또는 2개 차이가 날 수 있다. 비례배분에 의한 층화확률추출의 경우와 마찬가지로 이 경우에도 단순확률추출에서의 추정식을 그대로 사용할 수 있다.

층화의 또 다른 유용한 점은 가구의 크기가 큰 가구들과 같이 혼하지 않은 특성을 갖는 단위들에 대해서 충분한 표본크기를 확보할 수 있다는 것이다. 만약 목록에 있는 모든 사람에 대한 가구의 크기가 알려져 있다면 가구 크기가 큰 가구에 살

고 있는 사람들을 모아서 하나의 층을 구성하여 추출틀상의 다른 단위들과 분리하여 추출할 수 있을 것이다. 이와 같은 방법을 통해 각 유형에 속하는 정확한 수의 사람들을 표본으로 추출할 수 있다.

순수한 랜덤 추출과정으로 뽑히게 되는 수보다 많은 수의 단위를 특정 유형에서 추출하는 것을 과대추출(*oversampling*)이라고 한다. 관심을 갖고 있는 특정 유형에 속하는 단위가 전체 모집단에서 5% 정도만 존재한다고 하자. 이런 상황에서 표본으로 이런 특정 유형에 속하는 단위들을 충분히 추출하기를 원한다면, 우선 추출틀을 층화하고 희귀한 특성에 해당하는 층으로부터 보다 높은 비율로(예를 들어, 30% 이상) 표본을 추출한다.

층화표본추출은 또한 다음에 논의할 다른 표본추출방법과 함께 사용될 수 있다. 일단 층이 구성되면 각 층에서 어떠한 추출 방법을 적용해도 되고, 각 층에서 다른 추출법을 사용해도 된다. 예를 들어, 어떤 하나의 층에서는 전혀 표본을 추출하지 않을 수도 있고, 이런 층을 비표본층(*nonsampled stratum*)이라고 부른다. 이 층에 속하는 단위들은 전체 모집단에서 매우 작은 비율을 차지할 것이고, 자료수집 비용을 절감하면서 표본추출의 효율성을 향상시키기 위해 이들 단위들에는 0의 선택확률을 할당한 것이다. 또 다른 유형의 층은 선택확률이 1인 단위들로만 구성될 수도 있다. 즉, 이런 단위들은 전수 다 표본으로 선택되게 되는데, 이런 층을 전수층(*certainty stratum*)이라고 한다. 이런 단위들은 매우 크거나 매우 중요하기 때문에 선택의 여지없이 표본으로 전수 선택하는 것이다. 마지막으로 세 번째 유형의 층은 표본층(*noncertainty stratum*)으로 앞의 두 가지 유형에 해당하지 않는 단위들이 포함되 층이다. 이 단위들 중에서는 가능한 가장 효율적인 방법으로 표본을 추출한다.

층화(*stratification*)는 본질적으로 조사에서 측정되는 특성에 대해서 내부적으로 동질적이고, 서로 겹치지 않으면서 전체를 망라하는 그룹들로 추출틀 모집단을 분할하는 것이다. 층화확률추출(*stratified random sampling*)은 각 층에서 단위들을 단순확률추출하는 것이다.

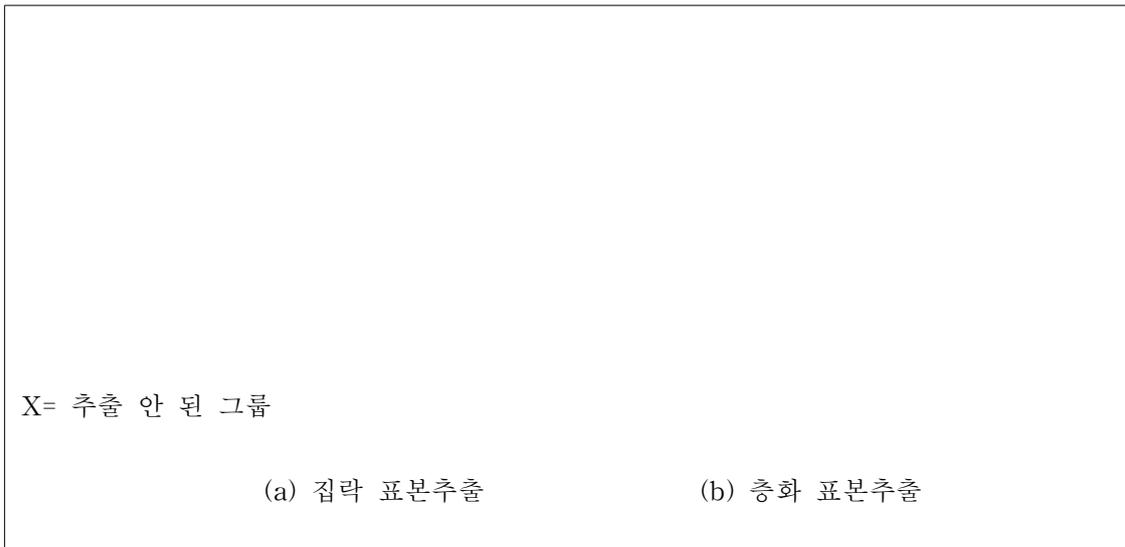
9.5.3 다단계 표본추출과 집락 표본추출

지금까지 우리는 추출틀에서 활용 가능한 보조정보를 확보할 수 있는 경우 추정량의 정확도를 향상시키는데 사용할 수 있는 세 가지 방법에 대해 살펴보았다: 비추정법, 계통표본추출, 층화표본추출. 비추정법(혹은 보조정보를 사용한 다른 추정법들)은 여기서 논의하는 다른 추출방법들과 함께 사용하는 것도 가능하다. 다단계표본추출과 집락표본추출은 추출과정에서 보조정보를 사용하는 또 다른 방법들이다.

다단계표본추출(*multistage sampling*)과 집락표본추출(*cluster sampling*)은 층화추출과도 관련이 있는데, 세 가지 방법 모두 인구학적 특성이나 지리학적, 혹은 다른 특성에 따라 추출틀의 단위들을 나누어 그룹들을 만드는 것에서 출발한다. 하지만 이들 사이에는 다음과 같은 몇 가지 중요한 차이가 있다.

- 층화추출에서는 각 그룹(층)에서 표본이 추출된다. 다단계추출에서는 전체 그룹 중 오직 몇 개만이 표본으로 확률추출된다(<그림 9.5> 참조).
- 층화추출에서는 그룹들은 관심 특성에 대해 그룹내에서 동질적이 되도록 표본추출자에 의해 구성된다. 예를 들어, 만약 조사에서 측정될 특성이 인종에 따라 상당한 차이가 있다면 표본추출자는 인종에 따른 층을 구성한다. 만약 저소득자와 고소득자의 응답에 차이가 있을 것으로 예상되면 표본추출자는 저소득층과 고소득층을 구성하려 할 것이다.
- 집락추출에서 그룹은 자연적으로 발생하는 구분에 따라 분할된다. 예를 들어, 인구조사의 경우 읍면동 또는 부락; 농업조사의 경우 지역 또는 토지구획; 학교조사의 경우 학군 또는 학교와 같은 것들이다.
- 일반적으로 층은 수는 집락의 수에 비해 상대적으로 작다. 예를 들어, 미국의 CPS(Current Population Survey)에서는 카운티와 카운티내의 센서스 블록으로 구성된 수천 개의 집락이 있는 반면 층은 오직 수백 개가 존재한다.

층화추출과 집락추출은 종종 혼돈된다. 가장 중요한 차이점은 층화표본추출(*stratified sampling*)에서는 모든 그룹(층)이 추출되고, 반면에 집락표본추출(*cluster sampling*)에서는 오직 그룹(집락)의 확률 표본이 추출된다는 것이다.



<그림9.5> 집락표본추출과 층화표본추출. 집락추출(a)의 경우, 전체 조사 대상 지역(사각형에 해당) 내의 단위들은 많은 조그마한 단위들의 서로 겹치지 않으며 전체를 망라하는(즉, 조사 지역의 모든 모집단 단위들을 포함하는) 집락으로 그룹(작은 원들)화 한다. 그런 다음이 집락들의 표본을 추출하고, 계속해서 집락내의 단위(원안의 별 표시)를 뽑는다. 추출되지 않은 집락들은 X로 표시되어 있다.

층화추출(b)의 경우, 비슷한 과정이 수행된다. 조사 대상 지역은 마찬가지로 서로 겹치지 않으며 전체를 망라하는 그룹으로 분할된다. 그림에서 볼 수 있듯이 그룹 혹은 층은 집락추출에서와는 다소 다르게 정의된다. 집락추출과 층화추출의 핵심적인 차이점은 집락추출처럼 단지 몇 개의 그룹에서 추출하는 것이 아니라 층화추출에서는 모든 층에서 추출한다는 것이다. 따라서 집락추출에서 발생하는 그룹화 작업에 따른 변동이 층화추출에서는 전혀 발생하지 않는다.

1단계 및 다단계 집락표본추출이 모두 가능하다. 만약 확률 추출된 그룹내의 모든 단위들이 표본에 포함되어 있다면, 이를 1단계 집락표본추출(*one-stage cluster sampling*)이라고 한다. 다단계 표본추출(*multistage sampling*)은 표본추출이 계속해서 집락 내에서 수행되고, 집락 내에서 보다 높은 단계의 집락들에 대한 새로운 집락화 과정과 추출이 이루어 질 수 있다. 그룹 내에서 추출된 단위들이 실제 관측 대상이 되는 조사단위인 경우 이 방법을 2단계 집락표본추출(*two-stage cluster sampling*)이라고 한다. 1단계(*primary stage*)는 집락들의 추출이고 2단계는 집락내의 조사단위들의 선택이다. 하지만 3단계 또는 4단계 추출 설계는 흔하지 않다.

2단계 표본설계의 예로, 어떤 주에서 학교들에 대한 조사를 고려해 보면 우선 학

군을 표본으로 뽑은 후, 뽑힌 학군 내에서 학교들의 표본을 뽑는다. 학군들에 대한 표본추출이 1단계 추출이고, 학교들의 표본추출은 2단계 추출이다. 만약 추출된 학교에서 학생들의 표본을 뽑고자 한다면, 학생들의 표본추출이 3단계 추출이다. 혹은 우리가 3단계로 학교 내에서 학급들을 뽑아 면접한다면, 학생들은 4단계 표본에 해당한다. 그러나 만약 표본추출 대신 우리가 추출된 반의 모든 학생들을 조사한다면 이 설계는 3단계 집락표본(*three-stage cluster sample*)에 해당하며 여기서 학교는 집락이다. 다단계와 집락 표본추출은 흔히 다단계집락표본추출(*multistage cluster sampling*) 또는 그냥 집락표본추출(*cluster sampling*)을 나타내기도 한다.

층화 추출과 집락 추출이 모두 추출에 앞서 모집단 단위들을 그룹화 하는 것을 필요로 하지만, 두 방법은 완전히 다른 추출 목적을 갖고 있다. 층화표본추출(*stratified sampling*)은 추정의 정확도를 향상시키기 위해 사용되는 방법이다. 이것은 조사에 있어서 관심 특성에 대해 동질적인 단위들의 그룹을 형성함으로써, 각각의 동질적인 그룹에 대해 상대적으로 작은 표본으로 그룹을 적절하게 묘사할 수 있기 때문이다. 이것은 층화추출이 사용될 때 전체 표본크기가 더 작아질 수도 있다는 것을 의미한다.

반면에 집락 표본에서는 단지 집락들 중 몇 개만이 추출되게 된다. 그러므로 각 집락이 가능한 한 모집단에 대한 많은 정보를 포함하기를 원한다. 다시 말하면 집락들은 내부적으로 이질적인 것이 바람직하고, 이는 층화추출에서 원하는 것과 매우 다르다. 하지만 집락들은 보통 모집단에서 카운티, 부락, 학교, 학군, 토지 구획 등과 같이 자연스러운 구분에 따라 형성되기 때문에 이런 인위적인 집락의 구성은 집락추출에서 일반적으로 가능하지 않다. 이는 불행하게도 집락추출에서는 보통 집락이 조사특성에 대해 내부적으로 동질적으로 구성된다는 것을 의미한다. 결과적으로 집락추출은 단순확률추출이나 계통추출과 같이 집락을 이용하지 않는 추출방법들보다 항상 정확성이 떨어지게 된다.

집락추출을 하는 가장 중요한 이유는 표본추출과 자료수집에 있어 특히 면접조사에 있어서 경제적이라는 것이다. 단위들을 집락화 함에 따라 표본 단위들 간의 이동비용은 급격하게 줄어들 수 있다. 예를 들어, 미국 전 지역에서 단순확률추출로 뽑힌 50,000 가구의 표본은 미국 전역의 모든 3,141개 카운티를 돌아다니면 면접을 해야 한다는 문제가 생기게 한다. 더욱이 가구들이 수마일 이상 떨어져 있을 수 있어 이에 따른 이동 비용은 거의 감당할 수 없을 정도로 클 것이다. 하지만 가구를

집락화 함으로써, 예를 들어 400에서 500개 정도의 카운티에 있는 가구들을 조사함으로써 이동비용을 급격하게 줄일 수 있다.

그러나 집락추출의 단점은 추정량의 분산이 같은 크기의 단순확률추출보다 상당히 커질 수 있다는 것이다. 예를 들어, 미 전역에서 추출된 표본 가구수가 50,000인 단순확률표본에서의 분산을 V_{SRS} 라고 하자. 동일한 가구수를 포함하는 400개 카운티를 추출하는 집락추출에서의 분산을 V_{CL} 이라고 하면, V_{CL} 은 $D \times V_{SRS}$ 으로 표현될 수 있다. 여기서 D 는 1보다 큰 어떤 수로 이를 설계효과(*design effect*)라고 한다(이에 대해서는 나중에 좀더 자세히 논의할 것이다). 집락추출의 분산은 집락내 상관(*intracluster correlation*) 때문에 보통 단순확률추출의 분산보다 크다. 즉, 어떤 집락에서 랜덤하게 추출된 두 단위들은 미국 전체에서 랜덤하게 추출된 두 단위들보다 훨씬 더 유사할 가능성이 높다.

예를 들어, 2000년 미국 대통령 선거에서 미국은 본질적으로 거의 같은 비율로 부시(Bush)편과 고어(Gore)편으로 나누어 졌다. 각 후보에 대한 유권자들의 지지율을 추정한다고 가정하자. 10개의 선거구로부터 1000명(선거구마다 100명)의 유권자를 추출하는 표본과 미국 전역으로부터 단순확률추출로 1000명을 추출하는 표본 중 어느 것이 더 정확할까?

이런 경우 특정 선거구 내에서는 한 후보를 전폭적으로 불균등하게 선호하는 경우가 흔히 있을 수 있다(즉, 많은 선거구에서 특정 후보 지지율이 90% 이상일 수 있다). 그러므로 수천 개의 선거구 중에서 단지 10개의 선거구를 확률추출하면 이들 모든 10개 선거구에서 두 후보 중 한명을 큰 차이로 선호하는 상황이 벌어질 수도 있다. 그러나 1000명의 유권자로 구성된 단순확률표본에서는 이렇게 큰 차이로 한명의 후보자를 선호하는 결과가 나타날 가능성이 매우 낮다.

달리 설명해 보면 전 지역에서 지지율 차이가 1% 미만일 때, 같은 선거구에서 추출된 100명의 유권자는 미국 전역으로부터 추출된 100명의 투표자들 보다 큰 차이로 특정 후보자를 선호할 가능성이 높다. 그러므로 선거구내에서의 투표자들의 유사성(즉, 선거구 집락효과)은 두 후보에 대한 지지율 차이를 추정하는 데 있어 추정의 정확성을 떨어뜨린다.

집락효과에 따른 정확성 손실에도 불구하고, 주어진 조사예산으로 조사를 수행해야 한다는 점을 고려하면 집락화 하지 않은 훨씬 더 작은 표본보다 매우 큰 규모의 집락표본을 선택함으로써 최고의 정확성을 갖는 추정결과를 얻을 수도 있다. 예를

들어 주어진 조사 예산으로 표본이 지리적으로 집락화 되었을 때 50,000 가구를 방문 조사할 수 있다고 가정하자. 그러나 막대한 여행비용 때문에 만약 표본이 전혀 집락화 되지 않았다면 동일한 예산으로 10,000 가구에 대한 조사가 가능할 수 있을 것이다. 결과적으로 설계효과인 D 가 집락표본에서 3.0이라 가정하자. 즉, $V_{CL} = 3 \times V_{SRS} = 3 \times \sigma^2 / 50000$ 이고, 여기서 σ^2 은 모집단 분산이다. 결국 분산 $\sigma^2 / 16667$ 을 단순확률추출에서의 분산은 $\sigma^2 / 10000$ 과 비교해 보면, 주어진 예산을 고려할 때 집락추출이 더 바람직할 수 있다는 것을 볼 수 있다. 따라서 이 경우에 같은 예산으로 좀더 작은 분산을 가진 추정을 할 수 있는 집락추출이 좋은 선택이다.

다양한 형태의 표본추출에 대한 장단점은 많은 표본추출 교재에 잘 수록되어 있다. 또한, 많은 표본추출 책들에는 주어진 조사 예산으로 보다 작은 표본추출분산을 얻을 수 있는 추출방법을 어떻게 결정할 수 있는지 설명해 줄 수 있는 비용과 오차를 동시에 고려한 최적화 공식들이 제시되어 있다.

2단계 집락표본추출(*two-stage cluster sampling*)은 1차추출단위(*primary sampling units; PSU*)의 표본으로 시작한다. 여기서 PSU는 균등 또는 불균등 추출확률로 선택될 수 있다. 각 PSU내에서 2차추출단위(*secondary stage units; SSU*)의 표본이 추출된다. 만약 각 PSU내의 모든 단위들이 뽑힌다면(즉, 2단계 추출이 아님), 이 방법을 단순집락표본추출(*single-stage cluster sampling*)이라고 한다. 다단계집락표본추출(*multistage cluster sampling*)에서는 각 단계에서 부차적인 추출이 이루어져 최종단계단위(*ultimate stage unit; USU*) 또는 모집단 조사단위에 이를 때까지 3단계 또는 그 이상의 단계에 걸친 추출이 수행된다.

9.5.4 불균등 확률 표본추출

단순확률추출, 계통추출, 비례 배분에 의한 층화확률추출은 모두 등확률추출법(*equal probability selection method; epsem*)에 해당한다. 즉 이런 설계들에서는 모집단의 모든 조사단위 또는 단위가 동일한 확률로 추출된다. 마찬가지로 집락표본

에서 집락들이 SRS로 선택되면 이 역시 epsem 설계이다. 다단계표본들도 각 단계에서 선택 확률을 적절하게 할당함으로써 epsem으로 구성될 수 있다. 예를 들어, PSU와 SSU를 모두 SRS으로 추출하는 2단계 설계에서 각 PSU 내에서의 추출율이 모든 PSU에서 같으면 이 또한 epsem 설계이다. epsem 설계의 좋은 성질은 표본이 자체가중(*self-weighting*)이 된다는 점이다: 즉, SRS에서처럼 표본평균이 모집단 평균에 대한 비편향성 추정량이다.

그러나, 많은 집락표본들, 다단계표본들, 그리고 비례배분이 아닌 층화표본들은 epsem에 해당하지 않는다. 예를 들어, 규모가 큰 가구 구성원의 수가 모집단에서는 단지 5%이지만 표본에는 50%가 이런 가구에 해당하는 비례배분을 적용하지 않은 1000명으로 구성된 표본을 추출한다고 가정하자. 이를 위해 2개의 층을 구성했는데 하나의 층은 규모가 큰 가구들에 거주하는 사람들로 구성된 것이고, 다른 층은 다른 사람들로 구성된 것이다. 그런 다음 각 층에서 500명의 사람을 추출했다. 앞에서 언급한 것과 같이 규모가 큰 가구에 거주하는 사람들은 모집단에 비해 표본에 과다하게 포함되어 있기 때문에 이런 비례배분이 아닌 층화추출을 과대표본추출(*oversampling*)이라고 한다. 규모가 큰 가구의 사람들을 과대표본추출한 결과 다른 층에 있는 사람들에 비해 이들이 표본으로 추출될 확률이 커지게 된다(즉, epsem이 아니다). 이에 따라 표본평균은 더 이상 모집단 평균에 대한 비편향 추정량이 아니다.

다단계추출에서 PSU는 흔히 크기비례확률(*probability proportional to size*; PPS)표본추출법으로 선택된다. 이러한 추출법은 PSU 크기가 클수록 높은 추출 확률을 할당하는 방법이다. 예를 들어, PSU가 학교인 다단계표본에서, 5000명의 학생을 갖고 있는 학교는 500명을 갖고 있는 학교보다 10배 더 큰 추출 확률을 할당 받는 것이다. 비록 PSU가 불균등 확률로 추출되었다 할지라도 종합적인 학생들에 대한 추출확률은 여전히 epsem을 만족할 수 있다. 만약 각 학교에 대한 선택확률과 무관하게 각 학교에서 같은 수의 학생이 추출된다면, 그 표본은 epsem에 해당한다. 그러나 만약 각 학교내의 표본크기가 학교에 따라 차이가 있으면 학생들에 대한 종합적인 추출확률은 epsem에 해당하지 않는다.

결과적으로 집락추출에서는 불균등 확률추출이 많이 이루어진다. 만약 단순집락 표본이 PPS로 추출되었다면, 크기가 큰 집락에 있는 조사단위들은 크기가 작은 집락에 있는 조사단위들보다 높은 선택확률을 갖게 되기 때문에 epsem이 아니다.

9.5.5 표본 가중과 추정

앞에서 서술했듯이 표본설계가 epsem이 아니면(즉 자체가중이 아니면), 표본평균은 모집단 평균의 비편향추정량이 아니다. 이러한 경우에 불균등 선택확률을 보정해 주기 위해 표본 단위에 대한 특성 값들의 평균을 구하기 전에 가중값을 부여해 주는 것이 필요하다. i 번째 표본 단위에 대한 가중값, ω_i 는 이 단위가 표본으로 선택될 확률인 π_i 의 역수로 정의된다.

$$\omega_i = \frac{1}{\pi_i} \quad (9.20)$$

그러면 모집단 평균 \bar{Y} 의 추정량은 다음과 같다.

$$\bar{y}_w = \frac{\sum \omega_i y_i}{\sum \omega_i} \quad (9.21)$$

불균등 선택확률에 따른 문제를 수정하는 것뿐만 아니라 무응답 또는 추출틀 오차를 조정하기 위한 조사후 보정에서도 가중값을 이용한다. 이런 가중값 사용에 대해서는 2, 3, 그리고 7장에서 어느 정도 논의 되었다. 따라서 본질적으로 두 가지 유형의 가중값이 있는데, 하나는 기초 가중값(*base weight*)이라고 하는 표본 선택 가중값(ω_{Bi} 로 표시)이고, 다른 하나는 조사후 보정 가중값(*postsurvey adjustment weights*)이다. 추정을 위한 최종 가중값, ω_{Fi} 은 두 가중값의 곱(즉, $\omega_{Fi} = \omega_{Bi} \omega_{Ai}$)이다. 그러면 식(9.21)에서 ω_i 를 사용하는 대신에 \bar{Y} 의 추정량을 얻기 위해 ω_{Fi} 를 사용하면 된다.

무응답과 추출틀 포함문제에 대한 보정이 가중값에 포함될 때, 이런 비표본오차 발생 원인에 따른 추정량의 편향을 감소시킬 수 있다. 그러나 가중값의 적용에 따른 단점은 이것이 추정량의 분산을 증가시킬 수 있다는 것이다. 일반적으로 표본에서 가중값들이 단위들에 따라 변동이 심하면 심할수록 가중값이 분산을 증가시킬 가능성이 점점 더 커진다. 경우에 따라 가중값이 분산에 미치는 영향을 줄이기 위해 가장 큰 가중값을 줄이거나 절사하는 방법을 쓴다.

가중값(weight)은 불균등 추출확률을 보정하기 위해서 뿐만 아니라 무응답과 추출
 틀 포함 오차를 보정하기 위해 필요하다.

9.5.6 표준오차 추정과 설계효과

앞에서 설명했듯이 단순확률추출은 다른 표본설계들을 비교하는 유용한 기준으로
 활용될 수 있다. 예 비해 유용한 기준점으로서의 역할을 한다. 이런 비교를 위해 흔히
 사용되는 척도는 설계효과(*design effect; Deff*)인데, 이는 복합설계에서의 추정량의
 분산을 같은 표본크기의 SRS에서 얻어지는 추정량의 분산을 나눈 상대 비에 해당
 한다. 예를 들어, 어떤 복합표본설계(δ)로부터 얻은 \bar{Y} 의 추정량을 \bar{y}_w 라 하자. 그러
 면 추정량 \bar{y}_w 의 설계효과는 다음과 같다.

$$Deff(\bar{y}_w) = \frac{Var_{\delta}(\bar{y}_w)}{Var_{SRS}(\bar{y})} \quad (9.22)$$

여기서 $Var_{SRS}(\bar{y})$ 는 (9.5)에 주어져 있다. Deff의 제곱근 Deft는 다음과 같다.

$$Deft(\bar{y}_w) = \sqrt{Deff(\bar{y}_w)} \quad (9.23)$$

복합조사설계에서 Deft는 표준오차가 설계에 적합한 추정 공식으로 계산되는 대신
 에 SRS 공식을 사용해서 계산되었을 때, 신뢰구간 추정을 위한 표준오차의 추정을
 하기 위해 직접 활용될 수 있다.

예 9.5.1 노동력조사 결과 모집단 실업률 추정값이 7%인 것으로 가정하자. 노동력
 조사를 위한 표본설계는 복합 층화 다단계 설계이기 때문에 추정량의 분산 계산을
 하기 위해서는 상당한 계산과정을 필요로 한다. 그러나 우리가 이전 조사로부터 실
 업률 추정값의 Deff가 약 1.83임을 알고 있다고 가정하자. 만약에 추정값이 10,000
 명의 응답자에 기반한 것이라면 95% 신뢰구간은 어떻게 될 것인가?

식(9.9)에 근사적 95%신뢰구간 공식을 제시했다. 큰 모집단으로부터 뽑은 단순확
 률표본에서 비율의 분산 추정값은 간단히 pq/n 이며, 여기서 p 는 표본 비율이고 q 는

1-p이고, n 은 표본크기이다. 따라서 노동력조사에서 추정량의 분산은 $Deff \times pq/n$ 로 추정될 수 있다. 이것을 SRS에 대한 신뢰구간인 식(9.9)에 넣으면 비율 p에 다음과 같은 95% 신뢰구간을 얻을 수 있다.

$$\left(p - 2 \times Deft \times \sqrt{\frac{pq}{n}}, p + 2 \times Deft \times \sqrt{\frac{pq}{n}} \right) \quad (9.24)$$

노동력조사 예에서 실업률에 대한 95% 신뢰구간을 계산하기 위해 이 공식을 사용할 수 있다. Deft는 $\sqrt{1.83} = 1.35$ 이고, $\sqrt{pq/n} = \sqrt{(0.07)(0.93)/10000} = 0.0026$ 이다. 그러므로 신뢰구간의 하한은 $0.07 - 2 \times 1.35 \times 0.0026 = 0.063$ 이고, 상한은 $0.07 + 0.0069 = 0.077$ 이다. 결과적으로 근사적인 95% 신뢰구간은 퍼센트로 (6.3, 7.7)로 나타낼 수 있다.

설계 효과인 Deff는 두 분산의 비이다. 분자에 해당하는 분산은 사용된 실제 표본 설계를 위한 공식을 사용해 구한 추정량의 분산이다. 분모에 해당하는 분산은 표본이 SRS를 사용해서 추출된 것으로 가정한 공식을 사용해 구한 추정량의 분산이다. 만약 Deff를 알고 있다면, 우리는 SRS 공식을 사용해 분산을 계산하고 이 결과에 Deff를 곱해 원하는 결과를 얻을 수 있기 때문에 Deff는 분산 추정을 위한 지름길을 제공한다.

마지막으로 설계효과는 부차모집단과 모집단의 부차그룹들에 대한 추정에도 활용할 수 있다. 예를 들어, 여성 혹은 18세에서 25세까지의 사람들에게 대한 실업률을 추정하기를 원할 수 있다. 일반적으로 총 모집단 추정을 위해 계산된 Deff는 부차모집단 추정을 위한 Deff보다 클 것이다. 이것은 Deff는 어느 정도 표본에서의 집락내 상관관계를 반영하고 이 상관관계는 각 집락내에서 추출하는 표본크기가 작아지면 추정에 미치는 효과가 줄어들기 때문이다. 따라서 모집단의 부차그룹에 대한 표준오차의 추정을 위해 모집단 Deff를 사용하게 되면 설계효과를 다소 과대 반영하는 표준오차의 추정을 하게 된다. 그럼에도 불구하고 이런 방법이 많은 유형의 통계적 추론에서 큰 문제없이 받아들여질 수 있다. 부차모집단 영역에 대한 추정과 관련된 많은 사항들은 Cochran(1977) 등을 참고하기 바란다.

9.6 결론

이 장에서 설명했듯이 조사 표본추출은 추출과정과 측정과정에 대해 잘 설정된 가정들을 전제로 하고 있다. 이러한 가정 아래서 제시된 이론은 표본크기에 상관없이 표본의 정확성을 매우 정확하게 예측할 수 있고, 매우 좋은 추론 결과를 가져올 수 있다. 대부분의 실제 상황에서는 통계적 추론을 위한 가정들이 정확하게 성립되지 않고, 이러한 가정 아래서 유도된 공식들은 현실적으로는 근사적인 것들이다. 예를 들어, 표본은 계통추출로 뽑혔지만 단순확률추출로 뽑은 것으로 간주하고 통계적 추론을 할 수 있을 것이다. 이런 경우 표준오차는 단순확률추출에서의 공식에 근사적인 설계효과를 반영하여 계산될 수 있다. 이런 근사적인 접근법을 통해 시간과 조사비용을 절감할 수 있고 보통 상당히 좋은 결과를 얻을 수 있기 때문에 이런 방법을 흔히 사용한다. 좋은 표본추출 통계 전문가가 갖추어야 할 중요한 자질 중 하나는 허용될 수 있는 수준의 결과를 얻을 수 있으면서 가장 간단하고 가장 비용효율이 높은 표본추출방법을 선택하는 능력을 갖추는 것이다. 사실 표본추출 이론을 활용하는 것은 정확한 과학이라기보다 근사적인 과학이라고 말할 수 있다.

고전적인 표본추출 이론을 위한 중요한 가정 중 실제 상황에서 거의 성립하지 않는 것은 조사 자료에 비표본추출오차가 없다는 가정이다. 결과적으로 고전적 이론을 적용해서 얻어진 결과들은 비표본추출오차에 따른 편향에 노출될 수밖에 없다. 이런 이유 때문에 무응답과 추출틀 미포함 오차 등과 같은 비표본추출오차를 조정하기 위해 일상적으로 조사 후 보정 작업을 하게 된다. 비록 이 책에서 간단하게 다루었지만 단위 무응답에 대한 표본 가중값 조정, 항목 무응답에 대한 대체, 추출틀의 과소포함을 조정하기 위한 사후층화 등과 같은 조사후에 이루어지는 보정작업들이 많은 대규모 조사의 추정과정에서 빠져서는 안 될 매우 중요한 부분을 차지하고 있다. 이와 관련해서는 Cox 등(1995), Rubin(1987), Little과 Rubin(1987), 그리고 Lundstrom과 Sarndal(2001) 등을 참고하기 바란다.

조사에서 표본추출은 고도로 전문화되고 매우 발달된 조사과정의 구성 요소인데도 불구하고, 이 장에서는 몇 개의 기초적인 개념만을 논의해 볼 수 있었다. 이런 분야에 대한 기초, 중급, 또는 고급 수준의 수많은 훌륭한 교재들이 있다. 이것들 중 일부는 참고문헌 목록에 있다. 하지만 가장 경험이 많은 표본추출 통계 전문가들이 증언하는 바와 같이 실제 조사 설계는 표본추출의 정확도를 최대로 하면서 비

용은 최소로 하는 것을 목적으로 하는 추정방법들과 표본추출기법들에 대한 복잡한 조합과정을 흔히 필요로 한다. 표본추출 과정의 복잡성과 추정과정에서의 비용절감의 필요성 때문에 일반 교재에 제시된 표준 공식들은 실제 상황에서는 직접 적용될 수 있는 경우가 드물다. 그럼에도 불구하고 고전적인 표본추출 이론을 철저히 이해하는 것이 실제 표본추출 설계를 하기 위한 전제 조건이다. 그러나 일종의 “예술”에 해당하는 표본추출 기술은 다년간의 경험을 통해서만 얻어질 수 있다.

제 10 장 총조사오차를 최소화하는 조사 설계

앞에서 우리는 총조사오차의 구성요소, 발생원인, 평가방법에 대해 다루었다. 우리가 총조사오차에 관해 그린 그림은 어떤 조사 설계자들에게 다소 궁색해 보일지도 모른다. 우리는 대부분의 조사업무에서 오차가 생기기 쉽고, 조사오차는 누적적이기 때문에 많은 작은 오차들이 모여져 받아들이기 어려울 만큼 큰 총오차로 귀결된다. 또한 오차 크기와 오차가 추정에 미치는 영향에 대한 정보를 모으는데 비용이 많이 든다. 이에 따라 조사 품질에 관한 책에서 반드시 다루어져야만 하는 한 가지 질문이 있다: 실제적인 조사 설계 업무에서 총조사오차를 최소화하기 위해 조사 계획과 실행 단계에서 사용되어야하는 전략이 과연 무엇인가? 이 장에서 우리는 이 질문을 다룰 것이다. 그 논의에 있어 앞에서 다루어진 많은 개념과 방법들을 통합할 것이며, 조사 설계와 실행에 이런 정보를 어떻게 적용할 것인지에 대한 지침을 제공할 것이다.

조사 업무에서 “진행 규칙”으로 표현될 수 있는 MSE를 최소화하는 많은 일반적인 전략들이 있다.

1. 가능하다면 언제든지 앞장에서 설명한 방법들 같은 신뢰할 수 있다고 알려진 방법들을 사용하라.
2. 조사 설계 과정의 일부로써, 각 조사 진행단계로의 자원 배분에 대한 계획을 개발하라. 이 계획에는 각 과정에서 중요한 업무에 투여해야 할 총 조사자원의 배분을 명확히 해야 한다. 사전 연구와 관련 문헌에서 얻은 조사오차에 대한 활용 가능한 모든 정보를 조사 업무에 최적으로 자원을 배분하는 데 사용하라.
3. 조사 설계에 조사 품질 정보를 수집하는 계획을 하나의 조사 과정으로서 집어넣어라. 이 정보는 향후 조사 설계에 정보를 제공하고, 현재 조사 과정을 모니터 하는 데에 사용될 것이다.
4. 예상하지 못한 문제들과 오차 원인에 대한 답이 필요하므로 품질 기준에 적합한지 모든 과정들을 모니터하라. 그리고 자원을 재분배하고 조사 설계를 수정하라.
5. 자료 생산자뿐만 아니라 자료 이용자들에게 자료 품질에 대한 정보를 널리 알려라. 이를 통해 자료 사용자들은 그들의 결정에 영향을 줄 수 있는 자료가 갖

고 있는 한계를 알 수 있고, 자료 생산자들은 자신들이 사용한 방법들이 자료 품질에 어떤 영향을 주었는지에 대한 정보를 얻을 수 있다. 이 마지막 단계는 향후 조사를 계획하고 실행하는데 매우 중요하다.

이 장의 나머지 부분에서는 어떻게 이 전략들을 개인 조사 연구자뿐만 아니라 조사 기관들의 행동 계획으로 전환할 수 있는지에 대해 논의한다. 조사 설계와 실행을 하나의 과정으로 보고 있는 2장의 아이디어들에 대해 다시 생각해보는 것으로 이 장을 시작한다.

10.1 비용, 조사오차, 다른 품질 특징들 간의 균형

<그림2.1>에서 우리는 조사 설계와 실행을 요구 수준을 충족하거나 초과하는 통계적인 생산물을 산출하는 것을 목적으로 하는 많은 단계를 포함하고 있는 하나의 과정으로 보았다. 조사 과정 각 단계들 사이의 상호관계와 어떻게 한 가지 구성요소에 관한 결정이 다른 구성요소들에 관한 결정에 영향을 미치는지에 대해 2장에서 논의하였다. 자원과 시간이 실제 조사 업무에서 항상 제한적이기 때문에 이런 논의에서 핵심적인 사항은 모든 단계에서 가장 좋다고 알려진 방법들을 단순히 적용하는 것은 불가능하다는 것이었다. 다음은 이와 관련된 다른 예들이다.

1. 조사 주제가 매우 민감한 것이면, 자기기입식 조사방법이 면접원의 도움을 받는 방법보다 더 선호될 수 있다는 것을 암시할 수 있다. 하지만 중요한 조사 목적은 조사 시작 후 바로 그 조사 결과를 생산해 내는 것이다. 이에 따라 불가피하게 전화조사의 사용이 필요할 수 있다. 더욱이 조사 설계자는 적시성과 자료 품질이 서로 상충하는 문제를 해결해야만 한다.
2. 표본추출틀이 거의 100%의 포함률을 만족하는 경우에도, 목표 모집단을 완전히 조사에 포함시키기에는 조사 예산이 부족할 수 있다. 예를 들어, 사업체 조사에서처럼 크기가 작은 모집단 단위들의 수가 현저히 많아 모집단이 이 쪽으로 매우 치우쳐 있지만, 작은 단위들이 표본에서 추정하려는 총계에 미치는 영향은 매우 작을 수 있다. 이런 경우 조사 설계자들은 절단표본추출(*cutoff sampling*)

법을 실행해야 할지 여부를 결정해야만 한다(즉, 절단기준 크기보다 작은 모든 단위들에 0의 선택확률이 주어져야만 하는지를 결정해야만 한다). 하지만 이러한 결정은 비용을 감소시키지만 통계적 추론과 조사 결과의 일반화 가능성을 제한하는 포함편차를 발생시킬 것이다.

3. 조사에서 무응답률이 매우 높을 것으로 예상되는 상황에서 비용에 대한 제약조건을 만족하면서 허용될 수 있는 무응답률을 얻기 위해서는 몇 개의 자료수집 방법을 동시에 사용하는 것이 필요할 수 있다. 하지만 조사 주제와 질문 형식에 있어 하나의 특정한 자료수집방법이 측정오차를 줄일 수 있어 특별히 선호될 수 있다.

이러한 것들은 조사업무에서 고려되는 일반적인 고려사항들 중 서로 상충되는 것들에 대한 예이다. 모든 조사에서 다른 품질 요소들과 비용이 수용할만한 수준으로 만족될 수 있도록 품질 요소들이 적절히 절충이 되어야 한다. 조사가 진행됨에 따라 추가적으로 절충이 이루어지는 것이 필요할 수도 있다. 예를 들어, (1) 추출단계에서, 완전한 추출틀을 개발하기에 너무 비용이 많이 들어서 다른 한 개 이상의 불완전한 추출틀들을 대신 사용해야만 할 수도 있다; (2) CAPI에 사용된 한 가지 질문이 이상하고 잠재적으로 응답자를 혼란시킬지도 모른다고 파악되었지만 설계 단계의 일정 내에서 수정될 수 없을 수 있다. 그런 경우 그대로 수행될 수밖에 없다; (3) 기관에서 두 가지 중요한 조사를 동시에 수행해야 할 수 있으며, 이런 경우 그 기관의 면접수행 수용 능력에 한계가 있기 때문에 상대적으로 많은 수의 무경험 면접원들을 고용해야 될 수 있다.

1장에서 품질 요소에 대한 논의에서 우리는 적절성, 정확성, 접근성, 비교가능성, 일관성, 완전성 사이의 상충이 조사를 설계하고 실행하는데 불가피하게 나타나게 된다는 점을 알았다. 그것들은 상호 의존성뿐만 아니라 비용에 있어서도 균형을 이루어야 한다. 비록 오차를 줄이고 품질을 최대로 향상시킬 수 있는 방법들과 기술들이 있지만, 비용이나 다른 품질 요소들이 희생될 수 있기 때문에 우리는 이러한 방법들을 마음껏 충분히 활용할 수 없다. 우리는 정해진 시간과 예산 내에서 조사를 마무리하면서 자료에 오차가 발생하도록 절충해야 한다. 그러므로 다른 오차들에 비해 어떤 오차를 줄이기 위해서는 보다 많은 비용이 들어야 한다는 점과 같은 자원을 갖고도 적절한 자원 배분을 통해 훨씬 더 많은 자료 품질의 향상을 얻을

수 있다는 점을 인식하는 것이 필요하다. 오차원인과 오차축소를 비용 함수로 설명할 수 있는 방안이 조사 설계뿐만 아니라 조사 전 과정에 걸쳐 반드시 고려되어야 한다.

오차를 줄이기 위해서는 추가적인 시간을 필요로 하기 때문에 품질 기준 중 정확성(*accuracy*)과 적시성(*timeliness*)은 보통 상충된다. 예를 들어, 무응답률을 줄이기 위해서는 추적조사 또는 금전적인 인센티브 사용 승인을 위한 시간 등이 필요하며, 이런 것들은 신속한 조사결과를 얻는 것과 상충된다. 하나는 양적(무응답오차의 축소) 요소이고 다른 하나는 질적(가장 최근의 것인지 여부) 요소로 볼 수 있는데, 이 문제를 풀 수 있는 방법은 두 가지 요소를 적절히 절충하는 것이다; 예를 들어, 자료 산출이 너무 지연되는 것을 방지하기 위해 무응답 추적조사를 축소하라. 정확성과 적시성에 대한 수준을 모두 높게 유지하면서 이런 상충문제를 해결하는 다른 방법은 우선 요구되는 일정에 따라 예비 추정 결과를 발표하고 그 후에 좀 더 정확히 수정된 추정 결과를 제공하는 것이다. 자료 사용 고객들에게 적시에 추정결과를 제공하는 것이 매우 중요한 사업체 조사와 같은 유형의 조사에서는 이런 전략이 흔히 사용된다. 예를 들어, 미국의 무역통상조사(*Foreign Trade Survey*)에서는 이런 방식을 택하고 있다.

다른 일반적인 상충은 정확성(*accuracy*)과 적절성(*relevance*) 사이에서 발생한다. 많은 통계 사용자들은 매우 상세한 수준에 관한 정보를 요구하는데, 이런 요구를 충족시키기 위해서는 정확성이 희생되어야 한다. 어떤 사람들은 정책결정을 위해 아주 상세한 정보를 필요로 하기 때문에, 4자리 또는 5자리 수준의 산업분류보다 6자리 수준의 산업분류 코딩을 원할 수 있다. 그러나 이런 세분화된 수준의 산업분류를 할 수 있는 자세한 사항에 대한 응답을 얻는 것은 힘들기 때문에 이런 세분화된 수준의 산업분류 코딩은 정확성이 떨어지게 된다. 이런 경우 수용 가능한 정확성을 응답자로부터 확보할 수 있는 세부 산업분류에 대해서만 6자리 코딩을 하는 것이 하나의 절충안이 될 수 있다.

조사가 전국 수준의 추정값을 구하려고 계획되었음에도 불구하고 사용자가 소지역의 자료를 원할 때 적절성과 정확성 사이에 또 다른 상충이 흔히 발생한다. 대부분의 소지역들에 대한 추정은 소수의 표본을 근거로 산출되기 때문에 정확성이 수용할 수 없을 수준으로 떨어지게 있다. 이 딜레마에서 벗어나는 방법은 적절한 수준의 정확성을 얻을 수 있는 표본크기의 확보가 가능한 대도시 지역에 대해서만

직접 추정값을 제공하는 것이다. 한편 소지역에 대해서는 유사한 속성을 갖는 많은 인접 소지역들의 자료를 합성하는 소지역 추정모형 기법을 근거로 산출된 간접 추정값을 제시한다.

흔히 또 다른 상충은 비교가능성(*comparability*)과 정확성(*accuracy*) 사이에서 발생하는데, 이는 주로 계속되는 조사에 개선된 새로운 기법을 채택할 것인가 여부와 관련된 문제이다. 사용자들은 정확한 자료를 필요로 하지만 또한 자료의 연속성을 계속 유지하는 것도 매우 중요하기 때문이다. 개선된 새로운 자료수집방법이 개발되면 조사 기관들은 자료의 정확성을 높은 수준으로 유지하기 위해서 이러한 개선을 채택하도록 내·외적으로 압력을 받게 된다. 그러나 이러한 새로운 방법들의 실행은 자료의 연속성에 영향을 미치게 된다. 이런 경우 시간의 흐름에 따른 변동을 추정하는 데 있어서, 모집단의 변동에 기인한 시간에 따른 변화와 비표본추출오차를 새로운 기법으로 감소시킴으로써 발생하는 변화와 구분할 수 없게 된다. 결과적으로, 조사 기관은 그들이 그 동안 사용하던 덜 정확한 추정방법을 계속 사용하도록 강요받을 수도 있다. 조사 분야에서 시대에 뒤떨어지고 수준이하라고 판정된 방법들을 계속 사용하면 그 기관의 명성은 떨어지기 시작할 것이다. 이런 경우 해결 방법은 새로운 방법을 채택하되 사용자들로 하여금 새로운 조사로부터 얻은 추정 결과에서 비표본추출오차 변화에 따른 영향을 조정할 수 있는 방안을 채택하는 것이다. 이를 통해 기존 설계와 새로운 설계를 연계하여 변화에 대한 추정을 가능하게 할 수 있다. 새로운 통계 시리즈와 기존의 시리즈를 연계하는 이러한 접합(*splicing*) 과정을 위해서는 보통 얼마 동안은 두 가지 조사를 동시에 수행하는 것이 필수적으로 필요하게 된다: 여기서 기존의 조사는 지금까지 해오던 그대로의 방식을, 새로운 조사는 개선된 설계를 사용하는 것을 말한다. 이런 중복설계(*overlapping design*)를 통해 조사기관은 과거 자료와의 비교를 위해 새로운 추정값을 예전의 추정값으로 변환할 수 있는 새로운 자료 시리즈에 대한 조정 인자를 계산할 수 있다. 그러나 중복설계에 따른 자료수집 비용의 증가로 이런 해결방법을 도입하기 위해서는 많은 비용을 지불해야 한다.

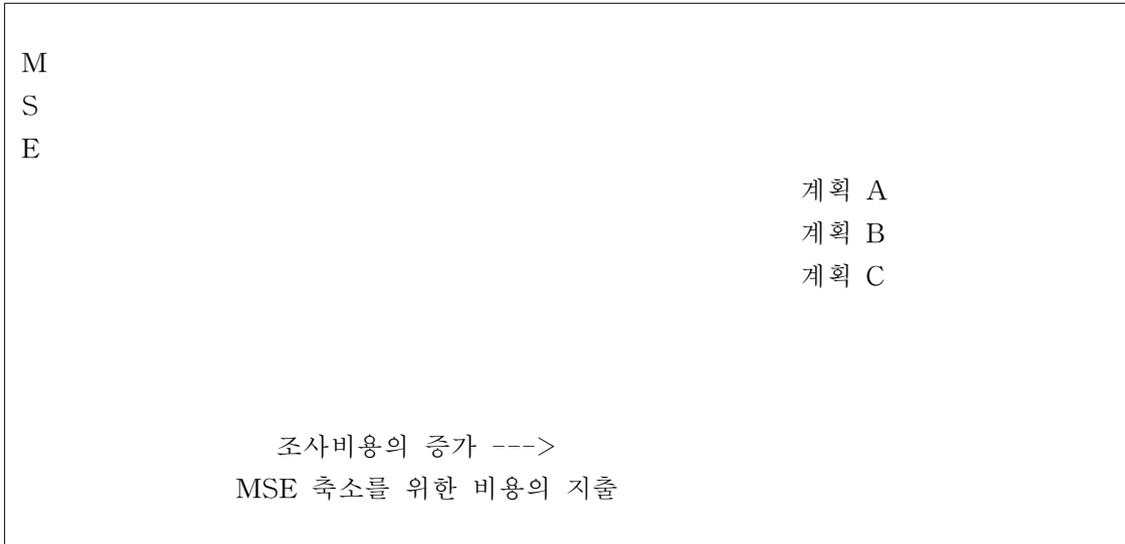
물론, 우리가 이 책에서 보아온 모든 조사에서 필연적으로 발생하는 중요한 상충들은 조사오차와 조사비용 사이에서 발생한다. 예를 들어, 다음과 같은 두 가지 사항 중 하나를 선택해야 할 것이다; 무응답 축소에 매우 효과적인 프로그램 개발과 추출틀 포함률 개선; 어떤 모집단 부그룹에 대한 충분한 표본크기의 확보와 면접원

들에 대한 보다 더 철저한 훈련 프로그램; 금전적 인센티브의 사용과 설문지에 대한 더 광범위한 사전검사; 이외에도 표본설계와 관련된 너무도 많은 사항들에 대해 조사 설계자는 전 조사과정에 걸쳐 선택을 해야 한다.

품질 요소들 간의 상충은 자료 생산자와 사용자 간의 상호작용을 통해서 가장 잘 해결된다. 대부분의 경우에 사용자들(또는 의뢰자들)은 자료에서 요구하는 것이 무엇인지, 그리고 자료의 적절성, 접근성, 비교가능성, 일관성, 완전성에 관한 상충되는 요구사항 중에 어떻게 우선순위를 정할지 가장 많은 지식을 가지고 있다. 한편 생산자들은 비용, 정확성, 적시성 그리고 이러한 분야에서 품질 목표를 어떻게 달성할지에 관해 가장 많은 지식을 가지고 있다. 그러므로 품질 요소들과 설계 대안들 사이의 상충을 해결하기 위해서 생산자와 사용자가 함께 일하는 것이 중요하다 (Holt and Jones, 1998). 자료 정확성의 궁극적인 수준은 오로지 조사기관의 생산자의 지식과 전문성에 의해 결정되기 때문에, 조사 자료의 품질과 관련된 조사 기관의 명성은 그 과정에 대한 의뢰자들이 느끼는 신뢰감과 안정감에 의해 결정된다.

경험 있는 설계자는 조사의 중요한 오차 원인에 관한 충분한 정보를 갖고 그 오차 원인들을 해결하기 위한 가용 예산을 참고로 하여 수용할 있는 자원의 배분을 가능케 하는 많은 전략을 사용할 줄 안다. 불행하게도 조사의 자원 배분에 관한 문헌은 다소 빈약한 실정이다. 이 문제에 대한 논의는 Linacre와 Trewin(1993)의 연구를 참고하면 된다. 이 저자들은 조사 문제들을 다루는 가장 비용 대비 효율성이 높은 접근법을 결정하는 평가 연구에 관한 광범위한 사용을 촉진하고 있다. 그들은 조사 추정량의 평균제곱오차가 오차를 줄이는 데 사용되는 자원의 양에 대해 감소 함수임을 제시하고 있다. 그러나 <그림 10.1>에서 보는 것처럼 동일한 자원을 지출하더라도 총오차에 미치는 영향은 작아질 수도 있고 더 커질 수도 있다. 그림에서는 A, B, C로 표시된 세 가지 전략 또는 계획에 의한 MSE의 감소를 나타내는 세 개의 곡선을 보여주고 있다. 세 가지 모두 감소하는 결과를 보이지만, 계획A를 사용한 경우의 감소량이 계획B 또는 C를 사용한 감소량에 비해 가장 적다. 더 많은 자원을 사용할 때, 계획C가 점점 더 다른 두 개의 MSE 감소 계획보다 효과적이다. 계획C가 다른 두 개의 계획보다 더욱 효과적인 방법, 더 나은 자원의 분배, 더욱 적절한 일정관리 등의 사용을 나타내고 있다. 어떤 지출 수준에서도 MSE를 최소화하기 위해 자원을 사용한다는 관점에서 설계는 최적화될 수 있음에 유념하라. 조사를 개선하는 데 투여할 수 있는 자원이 부족하더라도 여전히 그 수준에서 MSE를 최

소화시키는 분배 계획이 존재한다.



<그림 10.1> MSE 축소를 위한 지출과 3개 조사 계획들의 MSE 간의 관계

최상의 전략을 결정하는 것(즉, 비용과 시간에 대한 제약 조건하에서 조사오차를 최적화하는 것)은 더 많은 기술이나 정보 또는 그 둘 다의 이용을 필요로 한다. Dalenius(1971), Linacre와 Trewin(1993), Groves(1989), Fellegi와 Sunter(1974)는 조사를 최적화시키는 능력을 저해하는 많은 문제들에 대해 언급했다:

- 조사 기관은 조사오차를 최소화할 수 있는 비용 대비 효율성이 높은 방법을 판별할 수 있는 조사 방법에 대한 전문적인 지식을 가지고 있지 않을 수도 있다.
- 비용과 오차 사이의 관계는 <그림 10.1>에 나타난 것보다 더 복잡하다. 오차의 행태 특성상 얼마나 더 많은 자원을 그 오차를 해결하기 위해 투여해야 하는지를 알 수 없다. 예를 들어, 설문지에 대해 사전검사를 더 많이 하는 것이 측정오차를 좀더 감소시킬 것이라는 것을 알지만, 특정한 수준의 정확성을 얻기 위해 얼마나 많은 사전검사가 필요한지는 모른다.
- 조사 오차는 여러 오차 원인들에 걸쳐 동시에 상호작용을 한다. 예를 들어, 설문지의 개념을 명확히 하는 부가 질문들은 측정오차를 줄일 수도 있지만 응답자의 부담을 증가시킬 수도 있고 협조율을 저하시킬 수도 있다.
- 중요한 조사는 많은 항목에 대한 정보를 수집하도록 설계되므로 어떤 자원 배

분 모형이 모든 조사 항목에 최적일 수는 없다. 따라서 가장 중요한 항목들에 초점을 맞추어야 하고, 여기서도 절충이 필요할 수 있다.

- 정확성과 적시성이 아닌 모든 품질 요소와 그것들에 대한 제약 조건들이 사용 가능한 설계 대안들을 제한하게 된다.
- 얼마나 많은 조사 자원이 오차를 줄이는데 투입되어야 하는지, 오차와 자료 품질이 얼마나 자주 측정되어야 하는지를 결정하는 것은 어렵다. 만일 모든 자원이 오차를 줄이는데 투입된다고 하면, 비용, 오차, 방법들에 대한 정보의 부족으로 향후 관련 조사들을 최적화하는 것이 불가능하게 될 것이다.
- 협상 불가능한 것처럼 제약 조건들을 다루는 것은 좋지 못하다. 매우 자주 상대적으로 적은 예산의 증가로 상당한 개선들이 이루어질 수도 있다. 그런 경우에는 더 좋은 접근방법은 설계의 어떤 영역에 조금 더 많은 자원을 이용하는 것이 가져다주는 이익을 설명하면서, 자원의 추가적인 투입의 필요성과 여러 가지 옵션들에 대해 조사 의뢰자와 논의하는 것이다. Deming(1986)이 그의 14가지 요점에서 경고한 것처럼, 항상 가장 저렴하게 입찰하는 자와 계약하는 것은 좋지 않은 결정 전략이라는 사실을 조사 의뢰자들이 깨닫는 것이 중요하다.

비용의 함수로써 조사 오차를 모형화하는 대부분의 연구는 표본오차를 줄이기 위해서 표본 배분을 최적화하는 것에 초점을 두어 왔다; 그러나, 한 가지 예외적인 경우가 비용 모형을 표본오차와 비표본추출오차의 구성요소로 이루어진 것으로 생각하는 Groves(1989)의 연구결과이다. 그의 연구의 요점은 조사비용과 오차의 관계가 극도로 복잡해서 이러한 복잡성의 많은 부분이 알려져 있지 않다는 것이다. 그럼에도 불구하고, 비용과 오차 사이의 관계를 지나치게 간소화하는 아주 기초적인 모형들이 그래도 자원 배분에 관한 매우 훌륭한 길잡이 역할을 할 수 있다는 것이다. 이런 부분적인 이유는 자원의 실제적 배분과 최적 배분이 상당히 다를 수도 있지만, 그 결과는 여전히 꽤 유사하다는 것이다. 이것을 최적화 관련 문헌에서는 흔히 균일 최적(*flat optimum*)이라고 한다.

품질 요소들은 항상 조사 업무에서 서로 상충된다. 그러므로 조사 설계(*survey design*)는 어떤 요소에서 얻기 위해 다른 요소에서 희생하며 맞교환하는 절충의 과정이다. 이런 설계에서 어떻게 선택을 해야 하는지가 조사 방법론의 본질이다. 조사 설계자와 의뢰자 사이 또는 주요 자료 사용자들 사이의 협동작업을 통해 설계상의 여러 상충되는 문제들을 가장 잘 해결할 수 있다.

10.2 최적 품질을 위한 조사 계획

조사 계획(*survey planning*)은 조사의 규모, 자료수집 기관의 유형, 조사가 제안 요구사항(*request for proposal; RFP*)에 맞게 계획되었는지 또는 하나의 기관으로부터 재정 지원을 받는지 여부 등에 따라 많은 형식을 취할 수 있다. 그 상황이 어떻든지, 도로 지도가 어떤 나라를 가로지르는 장기 여행에 대한 계획의 중요한 부분인 것처럼, 계획은 조사 과정에서 중요한 요소이다. 특히 복잡한 조사 프로젝트에서, 조사 계획은 가용 자원을 효율적으로 사용하기 위해 그리고 총조사오차 최소화 전략에 따라 자원을 배분하기 위해 필수적이다. 또한 계획은 조사의 목적, 목적을 달성하기 위해 사용될 기술적인 접근법, 어떻게 조사가 관리될 것인지, 실행가능한 일(즉, 최종 조사 자료와 분석 보고서뿐만 아니라 조사 실행하는 동안의 중간 부산물), 그리고 인도 일정 등을 담고 있어야 한다. 그래서 그것은 향후 참고를 위한 조사 설계의 문서화뿐만 아니라 조사팀 구성원들과 조사 계획에 관해 서로 의사소통할 수단으로 사용될 수 있어야 한다. 조사 계획에 대한 개요에 대한 예시는 <그림 10.2>에 있다.

- 작업 설명서
 - 목표 모집단
 - 조사 목적
- 기술적인 접근법
 - 표본추출설계와 과정
 - 표본추출틀의 생성
 - 자료 수집과 과정
 - 자료 처리
 - 데이터베이스 준비
 - 분석과 보고
- 관리 계획
 - 지도력, 직원, 기관의 구조
 - 품질 관리 계획
- 활동과 인도 예정표
- 예산

<그림 10.2> 조사 계획의 전형적인 개요의 예시

계획은 충분히 구체적으로 작성되어야 하며, 그래서 조사를 실행하는데 드는 모든 유의적인 비용들이 추정될 수 있어야 한다. 이것은 조사 계획이 비용 대비 실행 가능한지 여부를 점검하는 역할을 한다. 만일 그렇지 않다면, 전체 자료 품질과 명시된 목적들을 달성하기 위한 조사의 능력은 유지하면서 전체 비용에 실질적으로 많은 영향을 주는 요소들을 주의 깊게 재설계함으로써 계획은 수정되어야 한다. 중요한 조사 변수들에 대한 정확성을 최대화하면서 주어진 시간과 예산 내에서 조사를 실행할 수 있도록 하는 조사 설계자들의 능력에 전체 조사 품질이 달려있기 때문에 명확하게 이것이 계획의 중요한 목적이다.

주의 깊게 조사를 계획하는 노력에도 불구하고, 좀처럼 계획한대로 정확하게 조사가 수행되지 않는다. 그러나 주의 깊은 품질관리시스템의 도입으로 많은 피해를 입기 전에 문제점들을 발견할 수 있다. 예를 들어, 어떤 조사에서 설문지 전반에 걸쳐 CAPI 시스템이 잘못 프로그램 되어 있었다. 이런 문제가 설문지의 각 질문에 대한 응답 빈도에 대한 계산을 통해 품질관리팀에 의해 조사 초기에 발견되어졌다. 많은 경우들이 영향 받기 전에 문제가 수정되어지고, 예전의 결함 있는 버전이 새로운 CAPI 모듈로 대체되었다. 조사 실행하는 중에 발생하는 예상치 못한 문제들을 성공적으로 해결하기 위해서는 초기 탐지와 빠른 반응이 열쇠임을 많은 반복

적인 경험들이 입증하고 있다.

10.2.1 내부와 외부의 전문적인 기술과 자원의 활용

조사 설계에서 고려되어야 하는 많은 요인들이 주어졌을 때, 조사오차를 최소화하는 최근의 기술에 대한 교육과 경험이 조사기관의 매우 귀중한 자원이 된다. 조사전략을 계획하는 동안에 내부 조사방법론 전문가들에게 의견을 구해야 하며, 특히 계획된 조사의 형식에 대한 전문적 지식을 가진 사람의 의견을 들어야 한다. 명백히 전략을 계획하는 단계에서 조사 설계상의 결함이나 비효율성을 고치는 것이 조사가 현장에서 실행되고, 자료가 일부 수집된 후에 수정하는 것보다 더 쉽다. 보통 일단 조사가 현장에서 실행되고 나면 중요한 설계의 결함을 고치는 것은 매우 어렵다. 또한 문제를 갖고 조사가 시작되면(CAI 프로그램이 제대로 작동되지 않거나, 거부율이 예상보다 훨씬 높은 경우 등) 현장 조사원들의 신뢰감과 충성심을 잃게 되어 생산성이 떨어지고 무응답률이 높아질 수 있다. 이런 문제는 극복하기 매우 어려운 품질의 악순환을 만들면서 현존하는 문제들에 더해지게 된다. 또한 자료가 그런 혼란 속에서 계속 수집되기 때문에 자료 품질에 회복할 수 없는 피해가 생길 수 있다. 중요한 문제들이 사전에 검토되고 자료 수집 시작 전에 조절될 수 있다면 이러한 문제들은 피할 수 있을 것이다.

조사방법에 대한 지식기반을 확장하는 한 가지 방법은 다른 조사 연구자들과 함께 전문가 활동에 참가하는 것이다. 예를 들어, 미국통계학회(ASA), 미국 여론조사 연구학회(AAPOR), 그리고 국제통계학회(ISI)에 의해 매년 또는 격년으로 통계 학술대회가 개최된다. 이 기관들의 일부에서 조사방법 또는 공식통계를 전문적으로 다루고 있다. 또한 조사 관련 특정 주제에 관한 많은 일회성 학술회의들이 매년 열린다. 이런 것들의 목록은 다른 통계학회의 회보뿐만 아니라 미국 통계학회의 회보인 “*AmStat News*”에서 발견할 수 있다.

또한 무응답, 에디팅, 데이터베이스 관리, 설문지 설계 같은 특정 주제를 위한 국제적인 연구 그룹이 조직되어져 왔다. 이것들은 그 분야에서 가장 최근의 연구개발에 관한 정보를 교환하는데 매우 유용하다. 어떤 조사기관들과 개인들은 특정 주제에 관심 있는 전문가 동료들로 구성된 그들만의 비공식적인 네트워크를 만들어 오고 있다. 정보교환과 공동연구를 이끌어 가는데 그런 네트워크들은 매우 유용할 수

있다. 조사기관들이 서로의 시스템과 방법론을 연구하기 위한 벤치마킹 활동을 포함한 공식적인 활동들을 함으로써 모든 실제 조사 수행방법들에 있어서 향상을 가져올 수 있을 것이다.

10.2.2 최상의 실행방법에 관한 문서의 활용

많은 조사기관들이 표본추출, 분산 추정, 무응답 추적조사, 에디팅 같은 다양한 조사 과정들에 대한 모범 적용사례에 대한 보고서를 작성해 오고 있다. 이러한 문서들은 조사 설계와 실행의 표준이나 지침서로써 조사기관들을 통해 유포된다. 기관에서 조사를 계획될 때, 모든 조사가 일관되게 설계되고 높은 품질 기준을 달성할 수 있도록 이러한 문서들을 참고로 한다. 이러한 문서들에는 보통 현행 최상 방법(*current best methods*; CBMs)들이 제시되어 있으며, 이런 방법들은 어떤 조사 업무나 과정을 위한 현행 최상의 방법들을 반영하기 위한 개선 의지가 확실하게 담긴 것들이다. 어떤 기관에서는 CBMs를 최상 실행(*best practices*) 또는 표준 운영과정(*standard operating procedure*; SOP)이라고도 부른다. 면접원을 고용하고 훈련하는 과정, 현장 직원의 급여 조정, 조사과정의 문서화와 배포, 예산 관련 문서의 준비 등까지도 CBMs로 목표를 설정할 수 있다.

CBMs의 목적은 그 기관 내부적으로 다른 조사에서 또는 다른 기관에 의해 외부적으로 개발되었는지 여부와 상관없이 최상의 방법들이 그 기관의 모든 조사에 사용되고 있는지를 확인하는 데 있다. 이것을 통해 일관되게 조사 산출물의 높은 품질을 확보 할 수 있다. 조사과정의 관리에 상당한 변동이 있었다면, 산출물의 품질은 일관되게 높은 품질을 유지하지 못할 것이다. 물론 CBMs이 주어진 활동에 대한 현행 최상의 방법이라는 것에 대한 기관내의 합치된 의견이 될 수 있다는 것이 중요하다. 기관에서 CBMs의 개발과정에 대해서는 10.3.2에서 논의된다. 한번 자리를 잡게 되면 CBMs는 기관내의 모든 부서에서 일관된 품질을 확보할 수 있는 매우 중요한 수단이 될 수 있다.

10.2.3 조사방법 관련 문헌의 활용

앞의 10.2.2절에서 설명한 것처럼, 조사방법 문헌은 조사 설계에 대한 최적 방법

을 확인하는 매우 귀중한 자원이다. 이런 관련 문헌들은 두 가지로 분류될 수 있다. : 교재와 학술지 논문, 단일 주제 연구논문집 및 기타 보고서. 먼저 중요한 교재들에 대해서 살펴보자. 표본추출방법론에 대해서 핵심 교재들로는 Kish(1965), Cochran(1977), Wolter(1985), Sarndal 등(1991)이 있다. 조사 표본추출에 관한 대부분의 다른 교재들처럼 이 교재들은 비표본추출오차에 관한 장을 포함하고 있지만, 표본추출에 관한 어떤 교재들도 비표본추출오차에 대해 내용을 상세하게 다루고 있지 않다.

조사에서의 비표본추출오차에 대한 논의를 위해서 참고할 수 있는 좋은 교재들로는 Groves(1989)와 Lessler와 Kalsbeek(1992)가 있다. 또한 조사방법의 특정 주제를 다룬 책들이 있는데, 조사 자료수집에 대한 Dillman(1978, 2000), 가구 조사에서의 무응답에 대한 Groves와 Couper(1998), 조사결과의 그래프 표현에 대한 Willgren 등(1999), 그리고 인지 과정과 설문지 설계에 관한 Payne(1951), Schuman과 Presser(1981), Sudman 등(1996), Tourangeau 등(2000)이 있다.

학술지 논문, 단일 주제 연구논문집, 보고서 형식의 문헌들도 많이 있다. 조사방법론의 다양한 특정 주제를 다루기 위해 발행된 연구논문집 시리즈가 있는데, 모두 미국통계학회와 다른 전문가 학회가 후원한 것들이다. 이런 전문 연구논문집에서는 패널조사(Kasprzyk 등, 1989), 전화조사방법론(Groves 등, 1988), 조사의 측정오차(Biemer 등, 1991), 사업체 조사방법(Cox 등, 1995), 측정오차와 과정의 품질(Lyberg 등, 1987), 컴퓨터 보조 조사 정보수집(Couper 등, 1998), 조사 무응답(Groves 등, 2002) 등의 단일 주제들을 다루고 있다. 다른 중요한 전문 연구논문집에는 표본조사에서 불완전한 자료에 대한 Madow 등(1983), 복합표본조사의 분석에 대한 Skinner 등(1989), 조사에서 인지의 기초에 대한 Tanur(1992), Schwarz와 Sudman(1992, 1994, 1996)이 있다. 많은 통계기관들은 전문논문집이나 일반적 관심사에 관한 방법론에 대한 시리즈 책자를 출판해왔다. 이런 기관들에는 ISI, Eurostat, UN, UN의 FAO, 미국의 연방통계기구, OECD, 그리고 세계 각국의 국가통계기관들이 있다. 이러한 기관들의 중요한 공헌에 대해서는 이 책에 많이 언급되어 있다. 좀더 최신의 정보는 항상 그 기관들의 웹사이트에서 찾을 수 있다.

많은 통계 학술지에는 조사방법론에 관한 논문들을 실려 있다. 조사연구 관련 주제를 전적으로 다루고 있는 네 가지 학술지는 스웨덴 통계청의 *Journal of Official Statistics*, 캐나다 통계청의 *Survey Methodology*, Eurostat의 *Research on*

Official Statistics, 그리고 미국통계학회(Survey Research Section, Social Statistics Section, Government Statistics Section)의 *Proceedings of the American Statistical Association* 등이 있다. 다른 학술지들도 조사연구 주제에 관해 지면을 할애하고 있다. 이런 대표적인 학술지들로는 *Journal of the American Statistical Association*, *Public Opinion Quarterly*, *Sociological Methodology*, *Journal of the Royal Statistical Society(series A와 B)*, *Sankhya*, *Journal of Applied Psychology*, *International Statistical Review*, *Bulletin of the International Statistical Institute*, *Journal of the Market Research Society*, *Biometrika*, *American Journal of Public Health* 등이 있다. 미국 센서스국과 캐나다 통계청에서는 끊임없이 개최되는 각종 학술 대회 및 심포지엄에 대한 회보들을 출판하고 있다.

특정한 조사연구 주제를 다룬 논문을 확인하는 좋은 방법은 키워드로 검색이 가능한 연간 발간물인 *Current Index to Statistics(CIS)*를 활용하는 것이다. 예를 들어, 무응답이라는 키워드로 검색해 보면 연구논문 제목에 이 단어를 포함하고 있는 대략 100개 학술지에 관한 CIS 데이터베이스에 수록된 모든 출판물을 확인할 수 있다. 또한 무응답에 관한 논문은 자료수집, 인센티브, 응답 부담 등과 같은 제목에서도 찾아질 수 있다는 것에 유의해야 한다. 그리고 각종 통계기관이나 다른 기관의 직원 등에 의해 작성된 출판되지 않은 연구논문에서도 많은 연구 결과들을 찾을 수 있다. 이런 연구결과가 매우 중요한 것이라면 보통 출판된 연구논문의 참고 문헌 목록에서 찾을 수 있을 것이다. 일반적으로 많은 연구논문에서 인용되면 될수록 그 연구논문이 그 분야에서 매우 중요한 연구결과를 담고 있다고 보면 된다.

많은 조사기관들은 그들의 직원들, 특히 조사방법론 분야의 전문가들에게 관련 연구 논문을 꾸준히 읽고, 가능한 이런 연구 논문 작성에 적극 참여하도록 독려한다. 그러나 심사가 있는 학술지들의 경우 대부분 모든 게재 신청 논문 중 25% 이하만이 출판되기 때문에 이런 학술지 연구논문 게재는 쉬운 것은 아니다. 하지만 기관에 있는 직원들이 이런 문헌에 실린 최근 연구동향을 파악하기 위해 꾸준히 시간과 노력을 기울이지 않으면 이런 출판된 문헌에 수록된 아주 가치 있는 정보들을 그 기관에서는 전혀 활용할 수 없게 된다.

예를 들어, 조사 참여를 유도하기 위한 사전 편지통지문이 무응답률을 감소시키는 것이 아니라 오히려 증가시킨다는 일화와 같은 증언이 이에 해당한다. 어떤 경

우 사전 편지통지문이 응답자가 조사에 참여하지 않을 구실을 준비할 수 있는 시간을 줄 수 있다고 주장할 수도 있을 것이다. 비록 이것이 몇몇 응답자들에게는 맞을 수 있지만 관련 문헌에 수록된 많은 연구 결과들에 위하면 이런 주장은 사전 편지통지문의 주요 효과로 받아들일 수 없다는 것이 입증되어 있다. 이런 사례를 통해 직관이나 일화와 같은 경험에 의해 잘못된 방향으로 인도될 수 있지만, 이런 잘못된 문헌에 제시된 연구결과들을 참고함으로써 올바르게 교정될 수 있다는 것이다. 다시 말해 관련 연구논문에 대한 꾸준한 검토가 상당히 가치 있는 작업이라는 것을 인식하기 바란다. 비록 관심 있는 특정 모집단이나 조사 주제를 정확히 다루는 연구결과를 문헌에서 찾을 수 없다면, 아직 연구되지 않은 상황에 관한 직관을 얻기 위해 기존의 연구 및 실험에서 얻을 수 있는 일부 교훈들을 적극 활용하는 것은 가능할 것이다.

몇 년간의 조사업무 경험에서 얻어진 직관은 조사업무에 큰 도움이 될 수 있다. 많은 설계, 사전조사, 자료 수집, 자료과정 활동 등에 대해 조사예산을 어떻게 할당하는지를 결정하는데 있어서는 경험이 필요하기 때문에 조사업무 경험은 특히 자원 배분 결정에 있어서 매우 중요하다.

비록 조사연구자들이 관련 문헌을 검토하는 것이 매우 도움이 되지만, 문헌들에는 많은 대립되는 결과들이 존재하기 때문에 이런 작업도 상당히 난해한 것일 수 있다. 특히 심사가 필요 없는 학술회의 논문집이나 미출간 보고서 같은 곳에 실린 방법론적인 연구결과들은 형편없이 설계되어지고, 다수의 설계요소들이 혼란스럽게 다루고 있을 수도 있다. 또한 잘 설계된 연구에서도 특정 방법의 효율성이나 유효성이 특정 조사 상황에 의존하는 것이기 때문이 결과를 잘못 해석하는 경우가 종종 발생한다. 예를 들어, 같은 조사에 대해 동일한 방법을 어떤 조사기관이 적용하느냐에 따라 그 결과에 차이가 생긴다는 것은 조사 연구에서 일반적인 결론이다. 이런 차이를 어떤 경우 기관 효과(*house effects*)라고 하는 데 여기서 “기관”은 조사를 수행한 조사기관을 나타낸다. 기관 효과는 조사기관의 문화, 정책, 고용 방법, 구성원 등의 측정할 수 없는 요인들 때문에 발생하는 결과일 수 있다. 문헌에 수록된 연구결과를 활용하는 데 있어서 얼마나 이런 요인들이 조사 결과에 영향을 줄 것인지에 대한 평가는 전적으로 조사 설계자에게 달려 있다.

조사계획(*survey planning*)은 기관 내부와 외부의 전문기술, 최상의 실행경험, 연구문헌에서의 권고, 근거가 충분한 품질 지침들을 활용하는 자원 지향적인 활동이다.

10.2.4 품질 지침의 적용

품질 지침은 통계 생산물에 대해 일반적으로 수용되는 원칙을 나타낸다. 이것들은 그렇게 하지 않아야 할 매우 뚜렷한 이유들이 없다면 따라야 하는 실행방법들이다. 지침들은 생산물 품질의 효과 면에서 덜 중요하다고 생각되어지는 것들과 중요하다고 생각되어지는 설계의 여러 가지 요소들을 고려해 개발된다. 품질 지침은 캐나다 통계청(1998), 영국의 ONS(2001), 미국의 NCES(2002) 등을 포함한 많은 기관들에 의해 제시되었다. 후자의 기관은 통계적인 기준과 안내지침을 단일 문서로 통합했다. 다음은 “수용할만한 응답률 얻기”라는 주제에 대한 사례이다.

예 10.2.1

기준 1: 자료 수집, 수집방법의 독립성(예를 들어, 우편이나 인터넷으로 설문지가 발송되었는지 또는 면접원에 의한 개별 또는 전화 조사인지)은 응답자의 참여율을 높인다는 관점에서 설계되고 운영되어야 한다.

지침 A. 자료 수집 방법(예를 들어, 우편, 전화, 인터넷)은 목표모집단과 자료수집 목적에 맞게 선택되어야 한다.

지침 B. 자료는 일년 중 가장 적절한 시점에 수집되어야 한다.

지침 C. 자료수집 시기는 보다 나은 응답률을 얻기 위해 적당하고 합리적인 기간 동안 이루어져야 한다.

지침 D. 적당한 시점에 응답자에 대한 인센티브가 고려되어야 한다.

미국여론조사협회인 AAPOR(1997)은 전체 조사과정의 다양한 측면을 고려한 조사업무를 위한 12가지 지침을 발표했다. 이것들은 <그림 10.3>에 제시되어 있다.

AAPOR에서 발표한 품질 지침

1. 조사에 대한 구체적인 목표를 갖자. 목적은 구체적이고 명확해야 한다.
2. 정보를 수집하기 위해 조사를 하는 것의 대안에 대해 고려해 보아라. 어떤 정보는 이미 현존하는 자료를 통해 완벽하게 얻을 수 있고, 이미 시행된 조사들이 여기에 해당할 수 있다.
3. 연구 대상 모집단을 잘 대표하는 표본을 추출하라. 올바른 모집단에서의 확률표본추출이 표본의 대표성에 관련된 문제를 해결해 준다.
4. 오차와 비용에 균형이 맞는 설계를 사용하라. 예산을 세울 때 모든 오차 요소들을 고려해야 한다.
5. 질문의 표현이 연구 모집단과 측정되어야 할 개념과 일치하도록 모든 주의를 기울여라. 아마도 이것이 조사계획과정 중 가장 중요한 부분일 것이다. 그것 때문에 전체 조사를 위태롭게 한 실패한 측정 과정들에 대한 많은 사례들이 있다.
6. 조사에 앞서 문제점을 확인하기 위해 설문지와 절차를 검토하라. 조사과정 중보다 조사 시작 전에 문제점을 확인하는 것이 항상 좋다.
7. 면접 기술과 조사의 주된 문제에 대해 면접원들을 주의 깊게 훈련시켜라. 응답자를 찾고, 동기유발하고, 의미 있는 자료를 얻는데 능력이 없는 면접원들은 조사의 품질을 상당히 떨어뜨린다.
8. 조사과정마다 체크 항목을 구성하라. 매 조사 과정은 전체 조사 오차에 영향을 줄 수 있는 잠재적인 요인들이고, 재작업에는 많은 비용이 든다. 적절한 품질확인 시스템은 확실한 품질 기준을 보장해 준다.
9. 사람들에게 대한 윤리적인 문제가 생기지 않는 한계 내에서 협조나 응답률을 극대화하라. 협조를 극대화 하는 것은 무응답으로부터 발생하는 오차를 줄여주는데 보다 중요한 역할을 하기 때문에 매우 중요하다.
10. 수집된 자료에 맞는 통계 분석법과 보고서 작성 기술을 사용하라. 발견된 사실과 한계와 해석에 대한 정직한 보고뿐만 아니라 조사의 모든 국면들을 문서화 하는 것이 조사기관의 성실성과 신뢰성에 영향을 주는 매우 중요한 사항이다.
11. 응답자들에게 한 비밀을 보장 서약을 주의 깊게 개발하고 이행하라. 그러한 서약이 안 지켜 질 경우는 기관이 큰 피해를 입을 수 있을 뿐만 아니라 명성과 사업 자체가 곤경에 빠질 수 있다. 보통 국가 지정 통계관련 법령에 의해 비밀성이 규정된다. 개개의 표본 단위들에 대한 정보가 노출되는 것이 불가능하도록 하는 방안을 개발하는 것이 통계전문가가 해야 할 일이다.
12. 평가와 반복이 허용되도록 조사의 모든 방법들을 공개해라. 전문 연구팀이 공식적인 문서를 기초로 연구를 다시 수행할 수 있을 정도로 문서화 작업이 상세하게 이루어져야 한다.

<그림 10.3> AAPOR에서 발표한 조사설계를 위한 지침에서 인용

10.2.5 사전검사와 시험조사

8장에서 설명했듯이, 조사를 설계하는 것은 종종 엄청난 양의 사전정보를 필요로 한다. 사전정보에 의해 조사를 설계하는 사람은 연구 대상 모집단에 대한 어느 정도의 지식이 있어야하고, 심지어 연구 주제가 되는 특성에 관해서도 지식이 있어야 한다. 이것은 조사 설계자는 조사가 끝날 때까지 알 수 없는 정보를 필요로 하기 때문에 원칙적으로는 역설적이다. 분명한 것은 이런 정보는 조사 자체 보다는 다른 방법 또는 좀더 작은 규모의 조사를 통해 수집될 수 있다는 것이다. 사전검사를 위한 조사와 여러 가지 설계 대안들 중의 선택과 같은 방안들을 포함한 조사를 평가하기 위한 많은 방법들이 8장에서 논의 되었다. 이 절에서는 조사 계획에 있어 사용될 수 있는 정보를 수집하는 이러한 방법들, 특히 시험 연구의 사용에 대해 논의할 것이다.

효율적인 조사 설계는 모집단 특성의 변동에 대한 정보와 이런 변동을 설명할 수 있는 자료를 필요로 한다. 우리는 MSE의 구성 성분에 대한 논의에서 이미 이런 필요성을 설명하였다. 또한 오차, 비용, 자료수집방법과 자료처리 과정의 실행 가능성에 대한 아이디어를 반드시 갖고 있어야 한다. 예를 들어, 여러 가지 자료수집방법들과 자료처리방법들 중에서 적절한 방법을 선택하기 위해서는 면접조사와 전화조사, 광학적 문자 인식 입력과 직접 자료 입력, 수동 코딩과 자동 코딩, 종속검증과 독립검증 등의 상호 장단점에 대한 폭넓은 지식을 필요로 한다. 앞에서 언급했듯이, 표본추출 단위, 표본추출 시스템, 추정 시스템에 대한 적절한 선택을 할 수 있어야 한다. 이러한 방법들과 절차에 대한 폭넓은 정보나 지식을 갖고 하지 않은 선택은 연구자의 전문지식, 경험, 일반적인 조사조건만을 기초로 가능한 최상의 결과가 얻어지는 바라면서 도박을 하는 것과 마찬가지로 일 것이다. 이런 정보에 대한 요구를 달성하기 위해서는 본조사의 설계를 향상시키기 위해 사용될 수 있는 정보를 생성하는 것을 목적으로 설계된 사전검사와 시험연구를 수행하는 것이 필요하다.

시험연구의 설계와 활용에 대해서는 불행히도 조사 관련 문헌에서 다루고 있지 않다. 시험연구는 본조사의 특별한 경우에 해당하고 본조사와 반드시 동일한 방식으로 설계 되어져야 한다고 설명할 수 있을 것이다. 그러나 시험연구를 위한 설계에서 직면하는 문제들은 보통의 정식 표본조사 설계에 있어 직면하는 문제들과는

다르다. 정식 조사나 전수조사의 목적은 표본 추정 결과나 센서스 계상 결과를 생산하는 것이지만, 시험조사는 조금 다른 목적을 갖고 있다. 정식 본조사에 대한 설계 원칙이 어떤 경우의 시험연구 목적에 대해서는 효율적 일수 있지만 다른 경우에 있어서는 문제가 있을 수 있다. 예를 들어, 시험연구에서는 정식 본조사와는 달리 필요에 따라 확률표본 대신 주관적인 표본을 사용할 수도 있다.

조사 관련 문헌에서 시험조사 설계에 대한 논의가 무시되고 있는 것처럼 실제 조사에서도 이런 상황이 벌어지고 있다. 추론은 종종 엄격한 통계적 원칙 보다는 직관에 의존하고, 조사 목표는 종종 약하게 정의되거나 희미하며, 그리고 비용 대비 효율성은 좀처럼 중요하게 고려되지 않고 있다. 이것들이 모두 조사설계자들의 시험연구 설계에서 주의를 기울일 필요가 있는 중요한 사항들이다. 시험조사 설계 분야에서는 일반적으로 수용되는 용어들도 정리되어 있지 않다. 이런 조사는 사전검사, 정식 무대연습, 가능성 검사, 실험, 심층 실험, 공식적인 검사, 비공식적인 검사, 방법연구 등으로 불리고 있다. 그러나 이것들 모두는 약간 다른 의미와 강조점을 갖고 있지만(8장 참고), 아직까지는 이것들을 모두를 시험연구(pilot study)라고 한다. 다음에 제시된 용어들은 문헌에서 발췌되거나 개인적인 의견교환을 통해 정리된 것들이다.

- 본조사(main survey): 시험조사 활동이 수행된 후에 실시되는 공식적인 조사
- 사전검사(pretest): 보통 연구 주제와 자료수집 도구를 탐색하기 위한 비공식적인 질적 기법을 사용한 소규모 연구. 일반적으로 요구된 정보를 얻기 위해 몇 개의 사전조사가 연속적으로 필요하다.
- 시험조사(pilot survey): 본조사를 향상시킬 수 있는 정보를 얻기 위해 설계되고 수행된 조사. 다양한 목적을 갖는 하나의 조사일 수도 있고, 각각이 제한된 수의 목적을 갖는 연속된 조사들일 수도 있다. 설계는 조사 목적에 주로 의존하나, 보통 신뢰할 수 있는 양적 정보를 얻게 되고, 본조사를 위한 예비 설계를 아직 조정할 수 있거나 심지어 상당 수준 변경할 수 있을 때 수행된다.
- 가능성연구(feasibility study) 또는 가능성검사(feasibility test): 실용성에 대한 사안이나 의문이 있을 때 하는 방법들이나 절차들에 대한 공식적이거나 비공식적인 연구. 공식적인 연구인지 비공식적인 연구인지에 대한 구분은 명확치 않다. 기본적으로는 공식적인 검사는 실험설계에 입각한 것이고, 반면에 비공식적

인 검사는 본질적으로 정보를 얻을 수는 있지만 질적인 측면이 상당히 강한 것이다.

- 심층실험(embedded experiments), 공식검사(formal test), 방법론연구(methodological study): 의견이 나뉘지는 경우 자료수집방법, 자료 또는 처리시스템, 다양한 설문지 등을 검토하기 위해 시험조사 또는 본조사의 일부로 심층 실험이 실시될 수 있다. 이런 실험들은 엄격하게 설계되어야 하고 보통 큰 표본 크기를 필요로 한다.
- 정식무대연습(dress rehearsal): 조사설계 또는 조사기관이 갖고 있는 약점을 파악하기 위해 본조사에 임박하여 실시되는 정식 조사의 축소판이다. 이를 통해 조사 방법의 개선을 위한 기초 자료를 얻고, 조사 종사자들에 대한 훈련과 경험을 제공하며, 조사의 시작부터 끝까지 전반적인 작업의 실행 가능성을 “증명” 하는 동시에 조사 업무를 검사할 수 있는 실제적인 자료를 얻을 수 있다.

<그림 10.4>는 다양한 시험연구를 통해 평가되어 왔던 설계와 관련된 사안들이다. 물론, 좋은 설계를 하기위해 모든 요소들에 대해 실제 자료를 확보하는 것은 필요하지도 않을 뿐더러 가능하지도 않다. 논의되었던 것처럼 재정상, 운영상, 방법상, 기술상의 제약에 직면하는 것이 일반적이고, 이것은 자동적으로 선택할 수 있는 설계 방안들을 제한한다. 더욱이 하나의 요소에 관한 구체적인 설계상의 해결책은 다른 설계 요소들에 대한 선택의 여지를 줄어줄게 한다.

<ul style="list-style-type: none"> • 재통보 기간의 길이 • 수집방법의 선택과 방법들의 조합 • 주제의 민감성 • 응답 부담감 • 개념과 정의의 명확성 • 비밀보장 서약의 효과 • 질문의 표현과 문맥 • 설문지 레이아웃 • 대체 응답 규칙 • 소요 시간 추정 	<ul style="list-style-type: none"> • 새로운 장비의 도입 가능성 • 설계효과 크기 • 모집단 변동성 척도 • 면접원 수행 보고 • 응답자 탐색 대체 과정 • 필요한 에디팅 범위 • 단위 및 항목 무응답률 • 무응답률 감축 수단의 효과 • 추출틀, 응답자, 면접원, 자료처리로 인한 비표본오차의 기대 비율 • 비용 요인
--	---

<그림 10.4> 시험연구를 통해 검토되어야 할 사안들

모집단, 검사일정, 조사 상황, 필요로 하는 추론 등을 종합하여 가능한 검사와 시험연구를 결정해야 한다. 정확한 추정을 위해서 크기가 큰 확률표본을 필요로 한다. 예를 들어, 공식적인 검사를 하기 위해 면접원의 선택하는 경우 전체 면접원 중에서 랜덤하게 뽑아야 한다. 공식적인 연구에서는 스스로 지원하는 면접원들을 선택하거나 현재 업무가 별로 없어 추가적인 업무가 가능하기 때문에 면접원을 선택해서는 안 된다.

비확률표본추출, 작은 표본크기, 제한된 지역 또는 PSU, 계절적 변동, 검사된 횟수 등에 의한 효과가 설명되기 전에는 시험연구의 결과를 너무 신뢰하지 말아야 한다. 시험연구의 표본크기와 비표본추출오차들에 의한 복합 효과는 심각하게 추론에 영향을 줄 수 있다. 만약 시험조사가 많은 목적을 갖고 있다면, 몇 가지 옵션들은 동시에 비교할 수 있는 실험이 보통 효과적이다. 시험연구에서는 유의수준 검정보다는 추정결과로부터 결론을 도출하는 것이 더 안정적이다. 시험연구는 정식 본조사의 최종 설계를 변경할 수 있는 충분한 시간을 확보할 수 있는 적절한 시점에 수행되어야 한다. 시험연구와 관련된 문제들은 Brackstone(1976), Jabine(1981), Hunt 등(1982), Nelson(1985), Lyberg와 Dean(1989)에 논의되어 있다.

시험연구(*pilot study*)는 계획을 위한 정보가 부족할 때 필요하다. 그것들은 목적에 따라 다양한 방법으로 수행될 수 있다. 만약 본조사의 설계자가 시험연구가 필요하다고 생각되면, 본조사의 설계를 보완할 수 있는 충분한 시점에 결과를 얻을 수 있도록 이런 연구를 수행하는 데 필요한 자원들을 별도로 비축해 놓는 것이 중요하다.

10.3 조사품질의 문서화

조사 실행과정과 경험을 문서화하는 것은 조사방법론 분야를 위해서 뿐만 아니라 통계기관에 있어서도 중요한 활동이다. 문서화의 주요한 목적은 과정들, 절차들, 그리고 조사들에서 얻은 결과들을 그 분야에 종사하는 다른 실무자들 뿐만 아니라 자

료의 사용자에게 전달하는 데 있다. 자료 수집을 위해 사용된 방법들과 자료의 한계점은 자료를 잘못 해석하는 것을 막아줄 수 있기 때문에 자료 사용자에게 특히 중요하다. 더구나 문서화는 우리가 확보하고 있는 조사 방법론에 대한 지식을 더 보완해 줄 수 있고, 향후 조사들의 질을 향상시키는데 일조 할 수 있다. 문서화는 여러 형태로 할 수 있다: (1) 조사 운영 과정의 문서화, (2) 추천할 만한 또는 최상의 실행방법에 대한 문서화 (3) 품질 보고서. 이 세 가지 형태의 문서에 대해 이 절에서 논의할 것이다.

10.3.1 조사 운영 과정의 문서화

이장에서 일찍이 조사 과정상의 세부사항들과 관련된 많은 것들을 전달하기 위해 조사계획이 중요함을 설명했다. 사실상 조사 계획은 조사과정의 각 단계에 투여할 활동과 노력의 수준을 나타내기 때문에 조사의 품질 수준을 정의한다. 이런 과정에서 자원은 조사 설계자에 의해 의도된 방법에 따라 조사에 포함되어 있는 다양한 작업에 균형적으로 배분되어야 한다. 그렇지 않으면 좀더 많은 자원이 조사의 시작 단계에서 소비되어 의도한 조사품질에 도달하기 위해 필요한 마지막 단계들에 있어서는 적은 양의 자원만이 사용될 수 있게 된다. 또한 계획은 조사 작업에 속한 직원들의 책임을 설명하고, 어떻게 프로젝트 팀이 함께 작업들을 잘 수행할 것인가를 그려준다. 또한 문서화는 프로젝트에 새로 참여하는 직원에게 조사 목적과 조사 설계에 대한 정보를 제공하는 중요한 역할을 한다.

거의 항상 그러는 것처럼 실행 단계에서 조사설계가 변경되면, 조사계획도 이에 따라 수정되어야 한다. 즉 조사계획은 생산과정의 모든 단계들에 대한 문서로 진화되어야 한다. 어떻게 두드러진 문제점이나 성공을 포함한 핵심적인 모든 과정이 어떻게 설계되었고 수행되었는지에 대한 문서화가 완료되기 전에는 조사 노력에 대한 완전한 평가가 이루어지면 안 된다. 만약 조사가 일정 간격으로 반복된다면, 토대가 되는 문서화 작업 없이는 작업 향상을 꾀하는 것이 어렵기 때문에 문서화는 굉장히 중요한 역할을 한다. 이것은 생산자와 사용자 모두 과정들을 이해할 수 있도록 충분히 상세하게 작성되어야 한다.

흔히 그 기관에서 수행된 모든 조사에 대해 적절히 사용되는 문서화 시스템이 있다. 만약 그런 시스템이 없다 하더라도, 문서화의 논리는 상대적으로 매우 간단하

다. 예를 들어, 추출틀과 추출틀 개발에 관한 것을 설명하면 다음과 같다.

- 목표 모집단과 추출틀 모집단, 그 차이를 명시
- 추출틀 개발과 추출틀 단위에 대한 서술
- 추출틀에서 접근 가능한 정보
- 추출틀 작성 과정
- 포함률과 포함률 제고 방안

표본추출에 대한 문서화에서는 다음을 포함한다.

- 층화와 그 목적
- 사용된 표본추출 설계
- 다단계 추출을 위해 사용된 크기 척도와 이것을 구성한 방법
- 층별 표본크기와 각 추출단계별 표본크기
- 사용된 보조정보
- 추정량의 예상 정도
- 표본추출상의 문제점과 표본추출 실행과정상 계획에서 벗어난 사항들

자료수집에 대해서는 다음을 포함한다.

- 사용된 방법에 대한 서술
- 무응답자 추적조사를 포함한 표본 단위 접촉 절차
- 면접 또는 자료수집 과정
- 면접원의 고용, 훈련, 감독, 모니터링, 관찰
- 인센티브를 포함한 무응답을 감축을 위한 특별한 방안
- 거부율, 접촉 불가능 비율, 기타 무응답률을 포함한 주요 응답그룹에 대한 무응답률
- 설문지와 면접 또는 자료수집 지시서(목록으로 첨부 가능)

자세한 내용을 담은 문서는 기관내의 모든 부서에서 접근이 가능해야 하고, 주요

설계 특징에 중점을 둔 덜 상세한 형식의 문서는 모든 자료 사용자가 (가능하면 인터넷으로) 접근이 가능하도록 해야 한다.

10.3.2 추천할 만한 또는 최상의 실행방법에 대한 문서

10.2.2절에서 우리는 조사 계획 과정에서 CBMs의 사용의 활용에 대해 논의 했다. 거기서 논의 되었듯이, 많은 조사 기관들은 더욱 높은 기준의 통계 생산 과정을 달성하기 위한 노력으로 그런 문서들을 만들고 있다. 물론 그들의 목표는 가장 성공적이고 입증 되었다는 의미에서 “최상” 이라고 여겨지는 기준을 적용하는 것이다. 이 절에서 최상의 실행을 정의하고 기관에서 CBMs를 개발하기 위한 핵심적인 개념을 정의할 것이다.

이름이 의미하듯이 CBMs에서는 조사방법론 분야에서 새로 개발된 방법들로 기존의 것들을 주기적으로 갱신되어야 한다. 이런 갱신의 빈도는 그 분야의 연구발전 및 기술 개발 속도에 달렸고, 기관 내에서 CBMs를 유지하기 위한 노력에 우선권을 두어야 한다.

Morganstein과 Market(1997)는 조사품질의 향상에 있어서 CBMs의 역할에 대해 자세하게 설명하고 있다. 그들은 가장 빈번하게 확인되는 변천의 원인들 중 하나는 동일한 업무를 수행하는 사람들 간의 수행방법 또는 접근방법상의 차이라고 설명하였다.

스웨덴 통계청은 무응답률 감소(Japoc 등, 1997), 에디팅(Granquist 등, 1997), 프로젝트 작업(Statistics Sweden, 1999), 공개 관리(Flygare와 Block, 2001), 설문지 검사(Lindstrom 등, 2001), 추정과 무응답 보정(Lundstrom과 Sarndal, 2001)에 대한 CBMs들을 개발했다. 한 예로, 다음에 우리는 스웨덴의 무응답률 감소를 위한 CBM의 개발에 대해 살펴보고자 한다.

조사에서 무응답을 제어하는 책임은 조사 관리자에게 달려있다. 최근에 스웨덴 통계청에서 실시된 대부분의 조사에서 무응답률은 떨어지지 않고 있으며, 많은 경우 오히려 증가하고 있다. 이러한 추세를 설명하기 위해 스웨덴 통계청은 조사에서 무응답률을 줄이기 위한 여러 접근방법들에 의한 변동을 축소하는 것을 목적으로 하는 CBM를 개발하기로 결정했다.

1986년 이후로 스웨덴 통계는 자신들이 수행한 조사들에서의 무응답률에 관한 정

보를 수집하고 이를 무응답 바로미터(*Nonresponse Barometer*)라고 불리는 책자에 수록해 왔다. 수년에 걸쳐 이 책자에 포함된 조사들의 수는 증가해왔고, 지금은 약 50개 정도를 포함하고 있다. 그러나 여기에는 무응답률을 감소하기 위해 사용된 방법에 대해서는 많은 정보가 포함되어 있지 않다. 따라서 사전 편지통지서, 자료수집 전략, 추적조사, 설문지 설계, 면접원 훈련, 응답 부담의 경감, 인센티브의 사용과 같은 극히 중대한 과정들이 이 책자에 포함된 조사들에서 어떻게 관리되었는지 파악하기 위해 추가적인 그런 자료들의 수집이 필수적이다. 처음에는 이러한 기술은 주요 과정 단계들을 확인하고 이들에 대한 최상의 수행방법을 CBM에 서술하는 파레토분석으로 가능할 것으로 가정되었다.

그들의 개선 작업에 있어서 조사 관리자를 안내할 수 있는 무응답과 무응답자들에 대한 자료에 일반적인 결함이 있다는 것이 연구에서 나타났다. 또한, 그 연구는 일반적인 조사 여건이 거의 동일한 유사한 조사에 있어서도 절차와 방법들이 상당히 다르다는 것을 보여주었다. 결과적으로, 그 연구는 지침이 필요한 무응답 감축 과정에 있어서 많은 중요하고 어려운 단계들을 확인하는 데 도움을 주었다.

이런 결과들을 바탕으로 CBM을 개발하는 작업이 시작되었다. 무응답 감소를 위한 단계별 처방에 따른 “요리책” 형식의 접근보다, 핵심적인 과정 변수를 정의하며 관련 자료 수집에 대한 지침을 제시하고 알려진 믿을만한 방법들의 사용을 강조하는 것을 통해서 체계적인 개선 작업을 위한 체계를 구축하는 형식으로 CBM을 작성했다. 이러한 일반적인 접근은 무응답률 감소를 위한 전략에 관한 어떤 전문성을 제공할 뿐만 아니라 많은 조사에 걸쳐 다양한 설계 모수들에 적용하는 데 필요한 융통성도 제공해 주었다.

CBM은 6명으로 구성된 팀에 의해 개발되었는데, 연구 개발 부서에서 3명의 전문가, 과제별 부서에서 2명의 통계학자, 면접방법론 전문가인 행동 과학자로 구성되었다. 팀은 앞에서 언급된 바로미터 책자에 포함된 50개 조사의 연구 자료를 분석하는 작업으로 연구를 시작했다. 이 작업이 완성 되었을 때, CBM에 대한 윤곽과 일반적인 내용에 대한 의견이 모아졌다. 이미 알려진 조사방법 관련 문헌과 믿을만한 방법들에 대한 검토와 다른 조사 기관들에 대한 벤치마킹을 위한 연구에 많은 노력을 기울였다. 각 장에 대한 원고가 작성되고, 기관의 다양한 부서에서 선정된 15명으로 구성된 그룹이 이에 대한 검토 작업을 벌였다. 이런 작업의 최대 성공 비결은 이 기관의 최고 관리자가 이것에 매우 높은 우선순위를 두었다는 것이다.

이런 작업의 결실로 얻어진 CBM은 4개의 장으로 구성된 책이다. 1장에서는 무응답률의 정의와 계산방법, 원인과 무응답 유형, 그리고 조사 참여자 이론과 같은 무응답에 관한 기본적인 개념을 다루었다. 2장에서는 주요 과정이라고 불리는 사실상 모든 조사에서 나타나는 과정들에 초점을 맞추었다. 여기서는 설문지 설계, 사전 편지통지문 설계, 추적조사 절차, 프라이버시와 기밀성 보장, 자료수집, 응답률 제고를 위해 어떻게 다양한 수단들을 조합할 것인지 등에 대한 것을 다루었다. 3번째 장에서는 민감한 질문의 처리, 응답자 부담, 면접원 관련 사항, 인센티브의 사용과 관리, 대리 응답자의 활용, 패널조사의 운영관리와 같은 항상 조사에서 나타나는 것은 아닌 과정들을 중점적으로 다루었다. 마지막 4장에서는 각 조사 관리자가 자신의 업무 개선에 도움이 될 수 있는 핵심 과정 변수를 확인하고 측정할 수 있도록 하기 위한 체계를 제시하였다. 여기서 핵심 과정 변수에 해당하는 것들을 예로 들면 표본 파손에 따른 무응답률, 수집방법에 따른 무응답률, 흔적 탐색에 의한 추적 성공률, 평균 접촉 시도 횟수, 시간 경과에 따른 접촉시도 상황, 독촉에 따른 유입량, 거부 전환율, 마지막 10% 응답자에 대한 자료수집비용, 변수별 항목 무응답률 등이다.

10.3.3 품질 보고서와 품질 고지서

이 장에서는 Eurostat에 의해 제안된 적절성, 정확성, 적시성, 접근성, 비교가능성, 일관성, 완전성과 같은 조사 품질 요소들에 대해 논의해 왔다. 1장에서 지적했듯이, 품질을 정의하기 위한 이런 구조는 조사 통계뿐만 아니라 어떠한 통계 산출에도 적용될 수 있다. 이 절에서 우리는 통계, 특히 조사통계 사용자들에게 이런 품질 요소들에 대한 정보를 제공하는 의도를 갖고 작성되는 문서들에 대해 논의할 것이다. 앞의 8장에서 설명했듯이, 품질고지서(quality declarations), 품질보고서(quality reports), 품질분석서(quality profiles) 등이 이런 문서에 해당한다(여기서 품질분석서는 좀더 광범위한 목적을 갖고 있다)

품질고지서의 주목적은 생산물의 적절한 사용을 유도하기 위해 통계 생산물의 품질 특성에 대한 정보를 제공하는 것이다. 그런데 품질분석서처럼 품질고지서도 역시 개선이 필요한 조사 영역을 확인하고 향후 조사를 개선하는데 유용할 수 있다. 품질고지서를 위한 또 다른 체계는 스웨덴 통계청에서 공식적인 통계를 위해 사용한 것이다. 이것은 7개 요소 대신에 5개 요소만을 포함하기 때문에 Eurostat의 형식

보다 다소 간단하다. 여기에 포함된 5가지 요소는 내용, 정확성, 적시성, 비교가능성과 일관성, 그리고 가용성과 명확성이다. 이것들은 다음과 같이 정의된다.

- 내용(*content*): 목표 모집단 특성, 측정, 영역, 기준 시점 등을 포함한 조사에 의해 추정될 모집단 모수들을 고려한다. 이 요소는 또한 포괄성에 대한 정보를 포함한다(즉, 통계적인 내용들이 얼마나 완벽하게 당면 과제의 주요 문제들에 대해 설명하는지 고려).
- 정확성(*accuracy*): 평균제곱오차의 다양한 성분을 포함한 총 조사오차에 대한 전반적인 평가를 고려한다.
- 적시성(*timeliness*): 조사주기, 생산 시점, 주요 생산물의 배포일정을 고려한다.
- 비교가능성(*comparability*)과 일관성(*coherence*): 다른 통계들을 얼마나 잘 함께 활용할 수 있는지를 고려한다(즉, 시간경과에 따른 비교가능성, 영역간의 비교가능성, 다른 통계들과의 일관성).
- 가용성(*availability*)과 명확성(*clarity*): 통계의 물리적 가용성과 지적인 명확성을 고려한다. 배포형식, 발표형식, 추가 정보 취득 방법, 상세 자료 접근 방법, 정보 서비스에 대한 사항들이 문서에 포함된다.

일단 이러한 요소들이 설정되고 적절하게 정의되었다면, 품질이 보고 될 수 있다. Eurostat은 품질보고서 모형(*model quality reports*)이라고 불리는 것에 대한 작업을 시작했다. 품질보고서가 통계 생산물의 모든 측면에 대한 품질을 충분히 설명할 수 있는 정보를 포함하는 것은 가능하지 않다. 그 이유는 조사가 수행 될 때마다 매번 총 조사오차를 평가하기 위해서는 너무 많은 비용이 들기 때문이다. 더욱이 조사 중 발생하는 현상의 특징으로 인해 오차의 모든 성분을 추정하는 것은 가능하지도 않다. 그러므로 보고 모형에서는 재정적인 자원 및 방법론적 측면에서 현실적인 고지서를 산출하는 것을 지향한다(Davies와 Smith, 1998 참조). 품질보고서에 담아야 할 내용에 대해 다음과 같이 추천하고 있다.

- MSE의 구성 성분들에 대한 실제 추정 결과가 없어도 조사품질 지표들을 산출하라. 비록 그것들이 품질에 대한 직접적인 척도는 아니라 할지라도, 이런 지표들은 조사과정의 부차적 결과인 동시에 보통 품질 척도와 강한 상관관계를 갖

고 있다. 가중된 경우와 가중되지 않은 경우의 응답률, 추출률 포함 오차율, 그리고 자료 에디팅 실패율 등이 이런 예에 해당한다.

- 품질 척도들은 주기적으로 산출되어야 한다. 표본추출 오차들, 무응답 편향의 추정값, 질문 항목 신뢰성 등이 이런 예에 해당한다.
- 하나 또는 몇 개의 오차원인의 효과를 매년 조사하는 롤링 평가 체계(rolling evaluation system)를 수행하라. 예를 들어, 1차 년도에는 포함 오차를 조사 보고하고, 2차 년도에는 응답 오차를 평가하며, 3차 년도에는 무응답 오차를 평가하는 방식이다.
- 사용된 방법들을 문서화하라.

그러므로 품질보고는 실제적으로 품질 추정값과 품질지표(무응답률, 에디팅 실패율 등), 사전조사 결과, 메타데이터(설문지, 정의 등)와 같은 다른 유형의 정보를 혼합한 것이다. 예를 들어, 처리 오차의 효과에 관한 정보를 제공하는 경우, 다음 사항들 전체보다는 오직 일부에 대해서만 추정값을 제시하는 것이 가능할 것이다: 처리상의 오차에 기인한 분산과 편향; 처리 오차들의 비율들과 이들의 추정에 대한 방법론상의 참고사항; 에디팅, 입력, 코딩 시스템에 대한 설명과 특정 유형의 경우에 발생한 에디팅 실패율과 같은 것들.

정확성 요소는 표본추출, 특정화, 포함, 측정, 무응답, 그리고 자료처리를 포함한 너무 많은 오차 요인이 있기 때문에 확실히 가장 평가하기 어렵다. 다른 요소들은 메타데이터(*metadata*) 특성이 있기 때문에 상대적으로 처리하기가 쉽다. 예를 들어, 날짜와 지연(적시성), 배포 일정, 출판물과 데이터베이스(접근성), 잠정 추정값과 최종 추정값의 차이, 연간과 단기 결과의 차이, 다른 자료로부터 얻어진 결과와의 차이(일관성), 그리고 미완성의 이유들(포괄성)에 대한 정보를 제공하는 것은 쉽다.

8장에서 논의한 것처럼 품질분석서(*quality profile*)는 품질보고서에서 한 단계 더 나아가는 것이다. 품질분석서는 품질보고서 보다 보통 훨씬 더 종합적이고, 총 조사 오차를 이해하기 위해서는 무엇에 대한 연구가 더 필요한지 보여주거나 또는 조사 설계에서 어떤 변화가 필요한지를 지적해 주는 것을 주요 목적으로 한다. 따라서 품질분석서는 보통 어떤 조사에 큰 변화를 가져오는 중요한 재설계 작업 이전에 작성된다. 이와 관련된 보다 세부적인 사항을 원한다면 8장과 Doyle과 Clark(2001), Kasprzyk와 Kalton(2001)을 참고하기 바란다.

모든 문서화 노력들은 장기간에 걸쳐 계획되고 수행되어야 한다. 그렇지 않으면, 이 작업은 손대기에 너무 방대한 작업이 되어 버린다. 사실 문서화를 꺼리는 가장 큰 이유는 자원과 시간이 부족한 상황에서 보통 조사가 끝날 때까지 이 작업을 미루어 두기 때문이다. 효율적이고 포괄적인 문서화를 위해서는 조사를 계획하고 수행하는 과정 중에 지속적으로 손쉽게 문서로 옮겨 질수 있는 형태로 정보를 수집하는 것이 필요하다. 또한 조사에 대한 많은 일반적인 정보와 노하우를 데이터베이스에 저장하여, 제안서를 작성하는 경우처럼 이런 것들을 발췌하고 재사용함으로써 유사한 중복적인 문서화 작업을 손쉽게 할 수 있도록 하는 것이 중요하다.

문서화(*documentation*)는 처리의 문서화, 추천할 만 하거나 최상의 실행방법에 대한 문서화, 품질 보고서와 같이 여러 형태로 나타날 수 있다. 문서화는 사용자와 생산자 모두에게 매우 중요하다.

10.4 조사 품질과 관련된 조사기관의 문제

10.4.1 작업 환경

조사품질에 대한 문헌에서는 어떤 기관들에 의해 생산되는 통계 생산물의 품질과 강한 연관성이 있는 것으로 보이는 그 기관들의 몇 가지 특성들을 제시하고 있다. 우선 첫 번째는 기관이 자신들이 수행하는 조사들에 적용되는 표준화된 절차 혹은 과정을 갖고 있는지 여부이다. 이것은 앞에서 논의 했듯이 좋은 품질을 촉진할 뿐만 아니라 비용 대비 효율성을 향상시킬 수 있다. 표준화를 통해 중요한 과정에 대한 개선이 가능한 새로운 것이 발견되면 이런 개선이 전체 기관에 걸쳐 효과적이고 한결같은 방식으로 확실하게 파급될 수 있도록 할 수 있다. 성공적인 기관들은 과정들을 일괄적으로 향상시키는 것을 항상 장려하고 지원하며 이를 매우 중요한 것이라고 생각한다.

성공적인 기관은 중요한 운영상의 결함, 큰 오차, 다른 실패들을 분석하고 그것들의 발생원인을 이해하고자 하며, 이를 통해 문제점의 근본적인 원인을 파악하는 것

을 확실하게 한다. 더욱이 이에 대한 예방책을 전체 기관에 알림으로써 이런 실수로부터 기관 전체가 무언가를 배우게 된다. 이를 통해 오차가 발생한 업무를 책임지고 있는 특정인을 곤경에 빠뜨리는 것이 아니라, 오차를 발생시킨 시스템상의 문제점을 강조해야만 한다. 그렇게 하지 않으면 갖고 있는 문제점을 기관 전체에 공개하는 것을 꺼리게 된다. 만약 그런 일이 발생하면 기관의 다른 부서에서도 유사한 오차가 계속 발생할 가능성이 높아진다. 같은 이유로 성공사례를 축하하고 함께 공유하는 것도 역시 중요하다. 예를 들어, 제안서가 공모에서 선발되는 것, 뛰어난게 높은 응답률의 달성, 의뢰자를 만족시킨 성공적인 자료의 전달 등이 이런 것들에 해당할 수 있을 것이다.

팀원들의 협력과 전임자로부터 전수 받은 교훈과 자료를 활용하는 것으로 특정지을 수 있는 작업환경을 갖는 것이 중요하다(Batcher와 Scheuren, 1997). 이것은 지속적인 과정에 대한 자료 수집과 그 분야의 진보된 지식을 위한 심층 실험의 이용을 포함한다. 국내에서 또는 국제적으로 다른 통계 기관들과 지식을 공유하고 협력하는 것이 매우 유용하다는 것이 강조되어야한다. 협력에는 연구 학술회의 참여하고 네트워크를 구축하는 것이 포함된다. 이런 접근들을 통해 지속적인 품질 개선이 가능한 일터를 만들 수 있을 것이다(Morganstein과 Hansen, 1990; Lyberg, 2000; Lyberg 등, 2001).

Martin과 Starf(1992)는 미국 연방통계기관들을 위해 “무엇이 효과적인 기관을 만드는가?”라는 질문을 제기했다. 품질과 전문적인 기준에 관해서 그들은 정부기관뿐만 아니라 모든 통계기관에 적용될 수 있는 6가지 활동을 정리했다. 이 저자들에게 의하면 기관들은 다음을 행해야만 한다.

- 통계학 이론과 응용뿐만 아니라 임무와 관련된 학문 분야에 철저한 전문성을 갖는 직원을 양성하라.
- 자료의 타당성과 정확성에 대한 이해를 증진하고 불확실성의 측정 결과를 사용자에게 전달해라.
- 자료의 타당성과 신뢰성을 증진하고, 자료를 모으고 컴파일하며 에디팅하고 분석하는 과정을 개선하기 위한 품질보증 프로그램의 수행을 계속하라.
- 모든 전문적인 작업에서 현대 통계이론과 바른 통계적인 실행 방법을 사용하라.

- 적절한 전문적인 조직들과 강하고 지속적인 관계를 개발하라
- 보고서 또는 정의, 문서, 자료수집방법의 서술, 불확실성의 척도, 가능한 오차발생 요인에 대한 발표에 있어서 인정된 기준을 지켜라.

10.4.2 윤리 지침과 원칙들의 준수

모든 실제적인 조사작업은 윤리적 지침과 원칙에 따라 수행되어야 한다. 조사 참여자들 혹은 조사기관들조차도 그러한 지침이 있다는 것을 인식하지 못하는 경우가 많이 있다. 많은 전문분야들은 종종 수행 규약을 갖고 있다. 이런 규약의 목적은 넓은 의미의 전문적인 가치들을 정리하고, 이런 가치를 유지하는 결과로 발생하는 기술적인 측면과 윤리적인 측면의 충돌을 어떻게 해결할 것인지에 대한 지침을 주는 것이다.

의학 분야의 윤리를 다룬 1947년에 제정된 뉘른베르크 규약(Nuremberg Code)가 아마도 가장 유명한 전문분야의 규약일 것이다. 하지만 다른 분야에서도 이런 윤리 규약이 자리를 잡아가고 있다. 전문적인 수행에 대한 규약들은 인류학, 심리학, 사회조사, 엔지니어링, 경영, 사회사업, 마케팅 조사 등에 있어서도 발견할 수 있다. 한편 통계조사 전문직을 위한 업무 수행 규약에 대해서는 미국 통계학회(ASA, 1983)와 국제통계기구(ISI, 1985)를 비롯한 다른 조사 관련 기관들에 의해 때때로 논의 되어왔다. 이에 대한 자세한 사항은 Jowell(1986)에 논의 되어 있다. 또한 공식 통계의 생산으로 범위를 한정된 규약 또는 규칙의 제정에 대한 논의도 있었다. 예를 들어, UN(1994a)에서는 공식통계의 기본원칙들을 제시했다.

전문직 윤리에 대한 국제통계기구(ISI)의 선언문은 통계 업무와 관련된 가장 일반적인 것이다(<그림 10.5>). 여기에 우리는 규약의 본질을 제시하고, 전 세계 통계학자와 조사전문가들이 기대하는 것이 무엇인지 예를 들기 위해 몇 가지 원칙들만을 나열했다. 규약의 전문을 확인하기를 원하면 규약 원본을 참고하기 바란다. “통계학자(*statistician*)”라는 용어는 조사를 계획하고 수행하고 분석하는 데 책임이 있는 모든 사람들을 포함하는 포괄적인 의미로 사용된다는 점에 유의하기 바란다.

조사연구 종사자들을 위한 국제 윤리 규약

1. 통계학자들은 수집된 자료에 대한 예상되는 오용과 오해를 방지해야 한다.
2. 통계학자들은 가능한 폭넓은 공동체의 이익을 위하여, 통계적인 탐구 범위의 확장과 발견된 결과들의 공유 가능성을 고려해야 한다.
3. 통계학자들은 잘못된 결과를 산출하도록 설계된 방법의 선택에 간여치 말아야 한다.
4. 통계학자들은 사전에 고용주나 의뢰자와 통계학자의 책임과 의무를 명확히 해야 한다.
5. 통계학자들은 방법상의 대안들을 공정하게 평가해야 한다.
6. 통계학자는 특정 연구 결과를 조건으로 하는 계약상의 조건을 받아들여서는 안 된다.
7. 사용된 통계적인 방법들과 절차들은 공개되어야 한다.
8. 통계학자들은 자신들이 사용한 방법들이 평가되는 것을 허락해야 한다.
9. 지식을 발전시키고 정보를 추구하는 것 자체가 다른 사회적이거나 문화적인 가치보다 우선한다는 것은 충분히 정당화 될 수 없다.
10. 사람들의 활발한 참여를 필요로 하는 통계적 탐구에서는 정보에 입각한 자유로운 참여자의 동의를 받아야 한다.
11. 상황에 따라 기술적이거나 실행상의 문제로 이전의 정보에 입각한 동의를 지킬 수 없게 되면, 이런 경우 다른 방법으로 연구 대상자의 권익을 보호해야 한다.
12. 통계학자들은 연구 대상자 자체와 그것들의 주의환경과의 관계를 방해하는 것을 최소화하도록 노력해야 한다.
13. 협력 여부와 상관없이 조사 대상자의 신분과 기록은 항상 비밀에 부쳐져야 한다.
14. 통계학자들은 어떠한 조사 대상자들의 신분이 폭로되거나 추론될 수 있는 형태로 자료가 공표되는 것을 방지해야 한다.

참고: 여기서 “통계학자“라는 용어는 넓은 의미에서 모든 조사연구 종사자들을 포함한다.

<그림 10.5> 국제통계기구(ISI)의 전문가 윤리 선언에서 발췌 (출처: ISI, 1985)

조사 연구자들 또는 통계학자들의 윤리적 책임감을 가볍게 여기면 안 된다. 자료의 사용 범위에 관한 약속과 참여를 꺼리는 사람들을 조사에 실제 참여토록 하기 위해 적용된 다양한 방법들과 관련된 사항들이 특히 중요하다. 또한 많은 지침들을 지키기 위해서는 적절한 문서화 작업이 많은 도움이 된다는 것은 분명하다.

국제 통계 공동체는 모든 조사연구자들이 지켜야 할 업무수행 윤리를 담고 있는

지침서를 1985년에 채택했다. 정부 공무원들과 공식통계 관련 종사자들에게는 추가적인 규칙들이 적용되는 데, 이런 공식통계에 추가적으로 적용되는 윤리와 관련된 사항은 Gardenier(2000)에서 찾을 수 있다.