310.6
통142ㅈ

조사에서의 측정 문제와 과정별 품질

2003. 10.

2005. 1. 07
B41660

통계청

조사에서의 측정 문제와 과정별 품질

편집자:

Lars Lyberg
Paul Biemer
Martin Collins
Edith De Leeuw
Cathryn Dippo
Norbert Schwarz
Dennis Trewin

Willey-Interscience 출판사
John Wiley & Sons 주식회사
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

목 차

기고자 1

서 문 3

조사에서의 측정문제와 과정별 품질 9

도입말 11

조사평가의 유래 11

제 1 장: 설문지 설계

제 1 절 조사표설계 : 개념에서 응답에 이르는 험난한 길 35
(Nobert Schwarz, University of Michigan, Ann Arbor)

제 2 절 이론적 개념에서 조사 질문까지 48
(Joop J. Hox, University of Amsterdam)

제 3 절 질문 순서의 효과 줄이기: 보조 문항의 작용 71
(Michaela Wänke, university Heidelberg ; Norbert Schwarz, University of Michigan, Ann Arbor)

제 4 절 조사에 있어 효과적인 측정을 위한 평정 척도의 설 96
(Jon A. Krosnick, The Ohio State University ; Leandre R. FabrigarQueen's University at Kingston)

제 2 장 자료 수집(Data Collection)

제 5 절 자료 수집방법과 조사품질: 개요 119
(Edith de Leeuw, Vrije Universiteit, Amsterdam ; Martin Collins, City University Business School, London)

제 6 절 조사자료 품질에 대한 새로운 자료수집기술의 효과 140
(Willian L. Nicholls II, 美 통계국; Martin Collins, 시카고 대학, 여론조사센터; Jean Martin, 영국 통계청)

제 7 절 면접원의 CAPI 기술 활용평가 166
(Mick P. Couper, Sue Ellen Hansen, Sally A. Sadosky, 미시건대학 Survey Research Center)

제 3 장 조사 처리 및 작업
제 8 절 조사후 처리의 몇 가지 측면 189
(Lars Lybe 스웨덴 통계청 ; Daniel Kasprzyk,미국 국립교육통계연구소)

제 9 절 조사처리를 위한 통합관리시스템 205
(Jelke Bethlehem, 네덜란드 통계청)

제 4 장 품질 평가 및 관리
제 10 절 조사측정 및 프로세스개선: 개념과 통합 231
(Cathryn S. Dippo, 美노동 통계청)

제 11 절 통계기관들의 지속적인 품질개선 249
(David Morganstein and David A. Marker, Westat, Inc., Rockville Maryland)

제 12 절 국가 통계기관들의 품질정책, 표준, 지침 및 권고관행 275
(Michael Colledge, 호주 통계청 ; Mary March, 캐나다 통계청)

제 13 절 기업경기조사 추정치의 비교 가능성 개선 296
(Bill Gross and Susan Linacre, 호주 통계청)

제 14 절 조사자료 평가: 사전조사에서 품질평가까지의 변천 과정 315
(James L. Esposito, 미국 노동통계 ; Jennifer M. Rothegb, 미국 인구조사국)

제 5 장 추정, 분석 및 해석에 대한 오차의 영향
제 15 절 측정오차의 조사자료분석에 대한 영향 검토 345
(Paul P. Biemer, Research Triangle Institute, Research Triangle Park ; Dennis Trewin, 호주통계청(Australian Bureau of Statistics)

제 16 절 상호 관입 할당량으로 감독되는 면접 - 재면접 조사에서 비표집 오차의 추정량 376
(Francesca Bassi and Luigi Fabbris, University of Padua)

부록 393

기고자

Reginald P. Baker, 미국 일리노이 주, 시카고, 국립 여론 조사 센터
Alison K. Baldwin, 미국 일리노이 주, 시카고, 국립 여론 조사 센터
Francesca Bassi, 이탈리아, 파두아, 파두아 대학
Mary K. Batcher, 미국 워싱턴 D. C. 연방국세청(IRS)
Jelke G. Bethlehem, 네덜란드, Voorburg, 네덜란드 통계청
Paul P. Biemer, 미국 노스캐롤라이나 주, Research Triangle Park, Research Triangle 연구소
Steven Blixt, 미국 델라웨어 주, Wilmington, MBNA-America
Bill Blyth, 영국 런던 Tayler Nelson AGB
Pamela Campanelli, 영국 런던, 사회 공동체 계획 연구소
Noel Chavez, 미국 일리노이 주, 시카고, 일리노이 대학
Michael Colledge, 호주, Belconnen, 호주 통계청
Martin Collins, 영국 런던, City 대학 경영대
Frederick Conrad, 미국 워싱턴 D. C., 노동통계국
Mick P. Couper, 미국 미시간 주, Ann Arbor, 미시간 대학
Edith de Leeuw, 네덜란드 암스테르담, Vrije 대학
Don A. Dillman, 미국 워싱턴 주, Pullman, 워싱턴 주립 대학
Cathryn S. Dippo, 미국 워싱턴 D. C., 노동통계국
Jennifer Dykema, 미국 위스콘신 주, 매디슨, 위스콘신 대학
James L. Esposito, 미국 워싱턴 D. C., 노동통계국
Luigi Fabbris, 이탈리아, 파두아, 파두아 대학
Leandre R. Fabrigar, 캐나다 온타리오 주, Kingston, Queen's 대학
Wayne Fuller, 미국 아이오와 주, Ames, 아이오와 주립 대학
Leopold Granquist, 스웨덴 스톡홀름, 스웨덴 통계청
Bill Gross, 호주, Belconnen, 호주 통계청
Patricia M. Guenther, 미국 워싱턴 D. C., 농무부
Sue Ellen Hansen, 미국 미시간 주, Ann Arbor, 미시간 대학
K. Piyasena Hapuarachchi, 캐나다 오타와 주, 캐나다 통계청
Anthony Heath, 영국 옥스포드, Nuffield 대학
John Horm, 미국 메릴랜드 주, Hyattsville, 국립 보건 통계 센터
Joop Hox, 네덜란드 암스테르담, 암스테르담 대학
Cleo R. Jenkins, 미국 워싱턴 D. C., 국세조사국
Jared B. Jobe, 미국 메릴랜드 주, Hyattsville, 국립 보건 통계 센터
Timothy Johnson, 미국 일리노이 주, 시카고, 일리노이 대학

Daniel Kasprzyk, 미국 워싱턴 D. C., 교육부
John G. Kovar, 캐나다 오타와 주, 캐나다 통계청
Jon A. Krosnick, 미국 오하이오 주, Columbus, 오하이오 주립 대학
Jouni Kuha, 영국 옥스포드, Nuffield 대학
Loretta Lacey, 미국 일리노이 주, 시카고, 일리노이 대학
Rolf Langeheine, 독일 Kiel, Kiel 대학
James M. Lepkowski, 미국 미시간 주, Ann Arbor, 미시간 대학
Susan Linacre, 호주, Belconnen, 호주 통계청
Lars E. Lyberg, 스웨덴, 스톡홀름, 스웨덴 통계청
Mary March, 캐나다, 오타와 주, 캐나다 통계청
David A. Marker, 미국 메릴랜드 주, Rockville, Westat 주식회사
Jean S. Martin, 영국 런던, 국립 통계국
Nick Moon, 영국 런던, NOP 사회 및 정치부
David R. Morganstein, 미국 메릴랜드 주, Rockville, Westat 주식회사
Colm O'Muircheartaigh, 영국 런던, 런던 정경 대학
William L. Nicholls II, 미국 워싱턴 D. C., 국세조사국
Diane O'Rourke, 미국 일리노이 주, Urbana, 일리노이 대학
Sarah M. Nusser, 미국 아이오와 주, Ames, 아이오와 주립 대학
William F. Pratt, 미국 메릴랜드 주, Hyattsville, 국립 보건 통계 센터
Thomas W. Pullum, 미국 텍사스 주, Austin, 텍사스 대학
J. N. K. Rao, 캐나다 오타와 주, Carleton 대학
Kenneth A. Rasinski, 미국 일리노이 주, 시카고, 국립 여론 조사 센터
Jennifer M. Rothgeb, 미국 워싱턴 D. C., 국세조사국
Sally Sadosky, 미국 미시간 주, Ann Arbor, 미시간 대학
Fritz J. Scheuren, 미국 워싱턴 D. C., 조지 워싱턴 대학
Norbert Schwarz, 미국 미시간 주, Ann Arbor, 미시간 대학
Jacqueline Scott, 영국 캠브리지, Queen's 대학
Randy R. Sitter, 캐나다 오타와 주, Carleton 대학
Chris Skinner, 영국 Southampton, Southampton 대학
Tessa Staples, 영국 런던, 국립통계국
S. Lynne Stokes, 미국 텍사스 주, Austin, 텍사스 대학
Seymour Sudman, 미국 일리노이 주, Urbana, 일리노이 대학
Katarina Thomson, 영국 런던, 사회 공동체 계획 연구소
Roger Tourangeau, 미국 일리노이 주, 시카고, 국립 여론 조사 센터
Dennis Trewin, 호주, Belconnen, 호주 통계청
Frank van de Pol, 네덜란드, Voorburg, 네덜란드 통계국
Richard Warnecke, 미국 일리노이 주, 시카고, 일리노이 대학
Adam Wronski, 캐나다 오타와 주, 캐나다 통계청
Michaela Wänke, 독일 하이델베르크, 하이델베르크 대학

서 문

조사의 품질은 조사 오차와 직접적으로 관련되어 있다. 조사 오차는 크게 표본 오차와 비표본 오차의 두 가지 종류로 분류할 수 있다. 표본 오차를 다루는 종합적인 이론들은 있지만 비표본 오차의 경우에는 이런 이론이 없다. 사실, 지난 50년 동안 비표본 오차에 대한 연구가 지대한 관심을 끌고 있어, 여러 특정 오차 원천들의 처리와 이 원천들을 동시에 모델링 하거나 통합적으로 다루려는 시도들을 설명하는 많은 문헌들이 출현하게 되었다. 이 연구에서는 몇 가지 측면에서 대단한 성공을 거두어 몇몇 오차의 근원들을 처리하는 효율적인 방법이 개발되었다. 한편, 그 외의 많은 유형의 오차들을 처리하기 위해 제안한 해결책들은 전혀 효과가 없었다. 물론, 한 가지 이유는 이 오차들이 수많은 근원에서 발생했기 때문인데, 이것은 몇몇 복잡한 조사 설계들로 인해 더욱 혼란이 가중되었다.

또 다른 이유는 비표본 오차 처리는 보통 누적적인 것이 아니기 때문이다. 즉, 한 유형의 오차의 수준을 감소시키면 다른 유형의 오차가 증가할 가능성이 높아진다. 실제로, 몇몇 조사의 질에 관한 목표들은 상충되어 이로 인해 이에 관한 의사결정 시 불합리한 요소가 포함된다. 예를 들면, 강력한 추가[1]권유(follow-up)를 통해 무응답율을 감소시키려는 시도는 적시성(timeliness)을 훼손할 수 있고, 조사 자료에 관한 많은 세부적인 정보를 입수하는 과정에서 비밀 유지 및 보안 규정을 위반할 수 있으며, 처리상의 오차를 줄이려면 응답자에게 더 불편한 설문지를 사용해야 할지도 모른다. 몇몇 오차 원천들은 처리할 수 있지만, 다른 일부 원천들은 포착하기가 너무 어려워 추정은 고사하고 표현조차 못할 정도이다.

비표본 오차들을 통제하지 않은 채 그대로 방치하는 것은 이것들이 영향을 받은 많은 중요한 조사 대상들의 조사결과 자료를 쓸모없게 하는 원인이 될 수 있다. 신뢰도, 타당도 및 편향 추정치 등을 측정하는 사후조사를 통한 품질 평가 척도들은 자료의 정확도를 나타내는 매우 중요한 지표들이지만, 반복 조사의 경우를 제외하고 조사 자료를 향상시키는 데 거의 아무런 소용이 없을지도 모른다. 따라서, 사후조사를 통한 품질 평가 및 이에 따른 수정이 아닌 설문지 구성, 면접 및 기타 자료 수집 활동, 코드 부여, 자료 포착, 편집 및 분석과 같은 조사 과정을 통제하는 데에 관심을 기울여야 한다. 조사 과정의 품질이 조사 결과물의 품질을 결정하는 것이다.

---

1) 설문지를 우송한 후 아직 응답하지 않은 사람들에게 편지, 전화 등으로 응답하도록 권유하는 것

전 세계의 많은 조사 기관들은 현재 전사적 품질 경영(TQM)이라는 개념을 조사 설계, 자료 수집 및 자료 처리에 적용하고 있다. 산업계에서 개발하고 있는 프로세스 통제, 품질 관리팀, 고객 중심, 과학적 방법에 근거한 의사결정 등과 같은 과정의 품질을 감시하고 확보하기 위한 방법들을 조사 작업에 적용하여 성공을 거두고 있다. 이 작업에 관한 사항들이 문헌상으로 보고된 바는 거의 없지만, 이 방법들이 갖고 있는 잠재력은 대단하다.

1992년에 미국 통계 협회(ASA)의 조사 연구 방법 담당 부서는 조사의 측정 및 제반 과정의 품질이라는 주제의 중요성을 감안하여, 조사의 측정 및 제반 과정의 품질이 이 조사 연구 방법 담당 부서가 후원하는 회의의 주제가 되어야 한다고 결정하고, 이 아이디어를 발전시키기 위해 학술 논문들을 편집하고 Lars Lyberg를 만났다. 이 부서는 이 회의에는 전 세계의 연구가들이 참석해야 하고, 지리적으로 이 주제에 관한 관심의 중심지인 미국 외의 나라들에서 연구 결과를 보다 널리 전파하고 이 나라들에서 일고 있는 대단한 관심을 활용하기 위해 한 유럽 국가에서 이 회의를 개최해야 한다고 결정하였다. 또한, 조사 연구 방법 담당 부서는 학제간 연구 및 문화권 간의 연구 필요성을 역설하였다.

1993년 초에 구성된 조직/편집 위원회의 위원들은 Lars Lyberg(위원장), Paul P. Biemer, Martin Collins, Edith De Leeuw, Cathryn Dippo, Norbert Schwarz 및 Dennis Trewin이었다. Lee Decker는 회의용 물자 조달 및 대부분의 회의 관련 사무를 담당하도록 임용되었다. Patricia Dean은 학술 논문 고문 편집자로 임용되었고, Lilli Japec는 기고 논문 및 기타 기획 업무를 담당하도록 임용되었다.

이 위원회는 수많은 연구 기관들을 만나 재정적 기부를 요청하였다. 또한 이 위원회는 학술논문집의 대요를 작성하여 전 세계에서 여기에 기고할 연구가들을 잠재적인 저자로서 물색하고 섭외 하였다. 그들에게 논문 개요를 제출할 것을 요청하였고 이 학술 논문집에 기고하는 데 관심이 있는 연구가들로부터 133개의 논문 개요들을 제출 받았다. 위원회는 이 중에서 본 학술논문집의 각 절을 구성하는 34개를 선정하여 회의 프로그램을 수립하였다. Steve Quigley는 John Wiley & Sons사가 본 학술논문집을 출간하는 업무를 총지휘하였다.

미국통계협회(ASA), 국제조사통계사협회(IASS), 시장조사협회(MRS), 왕립통계협회(RSS) 및 세계여론조사협회(WAPOR)의 5개 전문 기관들에 본 회의를 후원해 줄 것을 요청하였다. 이 5개 기관들 모두가 열성적으로 호응하고 본 프로젝트를 지원하기 위한 자금도 기부하였다. 이와 아울러 아래의 연구 기관들도 자금을 기부하였다.

호주통계청
중앙 통계국(더블린)
경제 & 사회 연구 위원회(런던)
국제 노동국(제네바)
국립 여론 조사 센터(시카고)
NSS(헤이그)
국립 과학 재단(워싱턴 D. C.)
국립 통계국(런던)
Research Triangle 연구소(Research Triangle Park)
SPSS(런던)
덴마크 통계청
핀란드 통계청
네덜란드 통계청
뉴질랜드 통계청
스웨덴 통계청
미시간 대학, 사회 조사 연구소, 조사 연구 센터
미 노동 통계국
미 국세조사국
미 농무부/농업통계청(NASS)
미 국립 보건 통계 센터
Westat 주식회사(Rockville)
조사 연구 방법 센터(ZUMA)(Mannheim)

이 기관들의 재정 지원이 없었다면 회의 및 논문 편집이 불가능했을지도 모른다.

'조사에서의 측정 문제와 과정별 품질에 관한 국제 회의'는 1995년 4월 1일부터 4일까지 영국 Bristol에서 개최되어 27개국에서 291명이 참석하였다. 이 프로그램은 기고한 논문들 중 본 학술논문집에 싣기 위해 선정된 34개 논문들과 113개의 기고 논문들로 구성하였다. 기고 논문의 저작자들에게는 미국 통계협회가 출간하는 회의논문집(proceedings volume)에 수록하도록 자신의 논문을 제출할 기회를 부여하였다. 이 회의논문집은 이 기고 논문들 중 60개의 논문들을 담고 있다. 이와 함께 2회의 Short Course 과정이 있었는데, 하나는 조사 연구를 위한 다각적 수준 분석(multilevel analysis)에 관한 강연이었고 나머지 하나는 통계 조사 기관들의 총체적 품질 경영(TQM)에 관한 것이었다. 또한 조직위원회는 '대영(大英) Ronald Fisher 경 추모 위원회'가 우리의 회의가 제19회 Fisher 추모 강연을 맡도록 선정하였다는 사실에 매우 흡족해 하였다. 이 강연 내용은 본 회의 논문집에 수록되어 있다.

이 책을 설계하는 데 있어서, 조사의 측정 및 제반 과정의 품질 분야에서 가장 중요한 문제들을 논의하고, 가능하다면 여러 견해들을 통합하는 것을 목표로 삼았기 때문에, 각 장마다 상당히 많은 편집, 검토 및 수정을 거쳤다. 이 책은 5개 장들로 구성되었고, 각 장의 제목 및 편집자는 다음과 같다.

제 1 장: 설문지 설계(Norbert Schwarz)
제 2 장: 자료 수집(Edith de Leeuw 및 Martin Collins)
제 3 장: 조사처리 및 작업(Lars Lyberg)
제 4 장: 품질 평가 및 관리(Cathryn Dippo)
제 5 장: 추정, 분석 및 해석에 대한 오차의 영향 (Paul Biemer 및 Dennis Trewin)

학술 논문 저작자들의 성향이 다양하였기 때문에 모든 장들에서 용어 및 표기법의 통일을 기할 수 없었다. 제 5 장을 제외하고 본 학술논문집의 통계 수준은 사회학, 심리학, 언어학, 교육학이나 시장조사를 전공하는 대학원생들이라면 무난하게 읽을 수 있는 수준이다. 하지만 제 5장은 조사 표본 및 수학적 통계에 관한 심오한 기초지식이 필요하다.

이 책은 교과서로 사용할 수 있지만, 그 주요 대상은 조사 연구에 관한 약간의 사전 지식을 갖춘 연구자들이다. 이 책은 여러 학문 분야들과 관련된, 조사의 측정 및 제반 과정의 품질에 관한 몇몇 검토 논문들을 수록하고 있으므로, 이 조사의 측정 및 제반과정의 품질에 적극적으로 관여하고 있고 다양한 이론적 관점들에서 이것을 논의하기를 원하는 연구자들에게 유용할 것이다. 또한 이 책은 조사 과정에 주안점을 두고 더 나은 설계, 자료 수집 및 분석 기법을 통해 조사의 품질을 개선하는 방법에 관하여 더 많은 것들을 알기 원하는 방법론자들에게도 유용할 것이다. 우리가 편집자로서 주의 깊게 판단해 보건대, 이 책은 1995년 현재의 최신 지식을 반영하고 있다. 우리는 모두 이 책의 출간을 통해 앞으로 이 흥미진진한 분야에 관한 연구가 촉진될 것으로 기대한다.

각 장의 편집자들 중 대부분은 다른 장에도 보조 편집자로 책임을 맡았다. 각 장의 저작자들은 이것을 심도 있게 저술하고 수정하는 활동 이외에도 다른 장들의 검토 과정에도 참여하였다. 그들은 자신이 저술한 장에 대해 외부인의 검토를 받아보라는 권유를 받았다. 따라서, 각 논문은 다수의 저작자, 편집자 및 검토자의 수고와 기여가 결집되어 나타난 결실이다. 본 위원회는 모든 원고를 최종적으로 편집한 Patricia Dean에게 진심으로 감사를 표한다. 그녀는 고문 편집자로 일하면서 통상적인 문구 편집은 물론 저작자들에게 모든 장들의 스타일 및 구성에 관한 제안을 하는 수고도 아끼지 않았다.

우리는 본 회의와 회의 논문집 작성에 필요한 물자의 조달과 관련된 세세한 일들을 정성을 다해 효율적으로 처리한 미국 통계 협회의 Lee Decker에게도 진심으로 감사를 표한다. 또한 회의와 논문집을 위해 아주 능숙하게 일을 잘 처리한 데 대해 Lilli Japec에게도 감사를 표한다.

회의에서 사용한 "조사에서의 측정 문제와 과정별 품질(SMPQ), 발췌 문헌"이라는 소책자를 편집한 Rotterdam 소재 에라스무스 대학의 조사 연구 방법 담당 부서 (SRM) 문서 센터에서 근무하는 Joke Hoogenboom과 Truus Kantebeen에게도 심심한 사의를 표한다. 그들은 본 논문집의 모든 장들의 문헌들을 찾기도 했다.

미국 통계 협회에서 여러 직위를 겸임하면서 본 주제에 관한 국제 회의들을 주로 개최하는 미국 외의 지역에서 이 회의를 개최하는 아이디어를 홍보한 Barbara Bailer, Seymour Sudman과 Judith Lessler에게 감사를 표한다. 본 위원회가 여러 장들을 검토하는 일을 도와 준 Dan Kasprzyk, Jun Liu, Joop Hox, John Bosley, Fred Conrad, Sylvia Fisher, Roberta Sangster, Linda Stinson과 Clyde Tucker의 노고에 감사를 표한다. 색인에 넣을 용어들의 리스트를 처음 편집한 Eva von Brömssen에게 심심한 사의를 표한다. 우리가 속해 있는 아래의 기관들에게도 우리가 회의를 진행하고 논문집을 편집하는 일을 지원해 준 데 대해 진심으로 감사를 표한다.

스웨덴 통계청(Lyberg), 미시간 대학(Schwarz), 암스테르담 Vrije 대학(de Leeuw), City 대학(Collins), 미 노동통계국(Dippo), Research Triangle 연구소(Biemer), 뉴질랜드 통계청 및 호주 통계청(Trewin)

Lars Lyberg
Paul Biemer
Martin Collins
Edith De Leeuw
Cathryn Dippo
Norbert Schwarz
Dennis Trewin

조사에서의 측정문제와 과정별 품질

도입말

조사의 측정오차: 역사 개관

Colm O'Muircheartaigh
London School of Economics and Political Science

1. 조사평가의 유래

초기 조사실시자들이 작업에 착수하게된 다양한 맥락들과 그들이 그들의 작업에 활용했던 다양한 견해들을 고려해볼 때, 그들의 시도에 대한 성공여부를 측정하는 공통의 기준을 찾기란 어렵다. 현대적 의미에서 조사의 역사는 단지 100년을 거슬러 올라가지만, 이에 관련된 환경, 주제, 철학 그리고 수행기관들은 매우 다양하였다. 초기단계에서는 조사설계 및 조사실시에 관련된 주제와 조사오차의 주제들 간에 특별한 구분이 없었다.

품질의 개념, 그리고 실제로 오차라는 개념도 그 작업이 수행되었던 동일한 맥락에서만 만족스럽게 정의될 수 있었다. 맥락이 다양해지고, 목표도 다양해지는 만큼 오차의 의미도 변하게 된다. 필자는 오차의 정의를 *하지 않았는데도 했다고 나타나는 일*이라고 할 것을 제안한다.

이것은 임의의(객관적으로 보이는) 기준을 규정하는 것이기보다는 목표에 관련되고 연구자가 참조할 만한 문제를 재정의하는 것이다. 절대적인 의미에서의 *참값*을 고려할 필요가 없고 자료는 수집되어야 한다는 것이다. 일반적으로 모든 조사작업에는 목표, 결과 그리고 결과에 대한 기술이 있다. 오차(품질불량)는 이러한 요소들이 일치하지 않는 것을 찾는 것이다.

조사연구는 정부/공식 통계, 학술/사회조사연구, 상업/광고/시장조사연구의 세 가지 뚜렷한 분야에서 역사적인 발전을 해오고 있다. 이들 각 분야는 그들만의 고유한 지식체계, 학문적 관점 그리고 성패를 평가하는 기준을 가지고 그들 나름대로 조사에 활용하고 있다.

국제통계기구(ISI: the International Statistical Institute)는 노르웨이 통계국장(Norwegian Bureau of Statistics) Kiaer가 "대표적 조사연구"에 대해서 자신의 견해를 발표하고 조사현장에 대한 심층연구를 주장했던 19세기 후반 무렵 정부

통계학자들의 논쟁의 중심이었다. 조사의 평가는 주로 통계적인 것이었고 조사(survey)란 모집단의 전수조사를 대신하는 수단으로만 간주되었다. 런던 정경대학 최초의 교수였던 Bowley는 표본(1906 그리고 1926)과 측정(1915)에 관한 자신의 저서를 통해 과학적인 표본조사에 기여한 주요 인물 중의 한 사람이었다. 이 후 표본조사는 정부의 자료수집에 주로 사용하는 방법이 되었고, 표본조사를 실시하는 정부기관은 정치인과 정치적인물 모두에게 자료제공의 중요10한 출처가 되었다. 그들의 기원을 나타내기라도 하듯, 이러한 부서는 주로 국가(정부)통계기관에 있었고, 그리고 그들의 전문 부서에서는 수학이나 통계학 교육을 받는 경향이 있었다. 여기에서 오차의 개념은 추정량의 분산(Variance of the estimator)이다. 특히, 1934년 Neyman의 주요 논문집에 의하면 표본 분포의 분산(variation of sampling distribution)이라고 한다. 품질과 분산을 동일시하고 얼마간의 편향을 인정하는 반복표본에 의한 측정은 1940년대 중반 인도의 Mahalanobis의 연구에서 확립되었고, 특히 변동이나 변이를 추정하기위한 상호관입표본(interpenetrating sample)에 대한 그의 설계는 현장조사자 등에 의해 도입되었다. 오차의 개념화와 그것의 측정에 대한 통계학자(통계전문가)들의 영향은 이런 전통을 따라 계속되고 있고, 조사표본의 모든 고전적인 문헌들에서도 찾을 수 있다. 이런 전통에서 "오차"의 용어는 한가지 이상의 의미를 갖게 되었는데, 이 용어는 막연하게 조사로부터의 결과나 산출물 또는 추정치에서의 분산의 원인을 기술하는데 사용되고 있다.

사회조사분야 내에서는 과학적인 표본조사의 강력한 입지가 인정되고 있지만, 1895년 ISI에 Kiaer가 제안한 사항은 널리 인정되지 못하였고, 하나 혹은 소수의 사례들에 대한 세밀한 검증(오늘날의 *사례 연구case study* 접근 방법)을 하는 학술논문집필자(monographers)들의 지지가 없었다면 거부되었을 것이 거의 확실하다.

ISI의 논의에서 학술논문집필자들의 활발한 참여는 조사의 발달에서 두 번째로 중요한 분야에 대한 연결점을 제공했다는 점에서 특히 흥미롭다. 이것은 *사회정책* 과 *사회조사* 운동이었는데, 그들의 시작은 아마도 1988년에서부터 1903년까지 런던의 빈민정책에 관한 Booth'의 연구와 1892년 미국의 Hull House [2]논문에서 가장 잘 드러난다. 공식적이거나 조직적인 움직임은 아니었지만, 접근에서 대한 공통점과 광범위한 활동분야에서 어떤 목표들이 있었다. 이 움직임의 목표는 사회개혁이었고 이를 위한 수단은 공동기술(community description)이었다. 활동의 성패는 정책 결정자와 정치인에 대한 조사결과물에 미치는 영향으로 좌우되었다.

2 헐하우스 ((미국의 사회 봉사가 Jane Addams가 1889년 시카고에 세운 복지 시설))

이 집단에 중요한 영향을 미친 것은 사회개혁운동과 새로운 학문분야로 태동하고 있었던 사회학이었다. 표본조사의 몇몇 선구자들은 공식통계와 사회조사분야 모두를 섭렵하였다. 특히 조사표본의 발달에 상당한 공헌을 한 Bowley는 1915년 사회조사방법에 관한 독창적인 연구를 했는데 그것은 사실정보나 행동정보에 관한 자료의 품질 및 오차의 추정치들(parameters)을 정의하는데 도움을 주었다. Bogardus(1925), Thurstone(1928), Likert(1932)는 태도의 측정에 대한 과학적인 접근방법을 제공했다. 이 분야는 사회학과 사회심리학을 기본 방향으로 삼았고, 사회통계와 정신측정학으로부터 어느 정도 영향을 받았다. 1946년 미시간대학에서 사회조사연구소의 창립자 겸 소장이었던 Likert는 조직에 관한 이후의 연구에서 사회조사운동의 초기 선구자들과 동일한 실무중심의 경향을 나타냈다.

세 번째 분야는 의사소통 수단의 확장과 시장의 성장을 통해 발생하였다. 1890대부터 광고대상의 꾸준한 증가 그리고 동반자인 시장조사연구의 발달과 형성으로 미약한 시작이 일었다. 시장조사연구는 원래는 재화와 서비스의 생산자 입장에서 상업적인 정보에 주안점을 두었고 이것을 비공식적으로 수집하고 평가하였지만 점차적으로 독자적인 영역을 확보한 전문적인 활동으로 발전하였다.

이 과정에서 특히 심리학자들의 영향이 컸다. Link 등이 Psychological corporation 에서 행한 연구는 시장조사연구 분야에서 측정을 위한 명백한 과학적 기초를 제공하는 데에 있어서 그 영향력이 상당했다. 이런 심리학자들에게 실험심리학은 사회심리학보다 우위로 평가받았다. 그들이 사용한 용어와 접근법은 과학과 기술에 관련된 것들이었다. "오차"라는 용어도 명시적으로 사용되지 않았고 그 보다는 도구에 있어서의 *신뢰도*와 *타당도*를 기술하였다. 이것은 통계학자들의 "오차"와 기본방향에 있어 특히 대조를 이룬다.

이렇게 해서 1940년대와 1950년대 기초가 확립되기 시작한 조사연구분야는 정부, 학술단체 그리고 경영분야의 세 가지의 다른 부문에 관련되었고, 통계학, 사회학, 그리고 실험심리학의 세 가지 다른 학문적 기초를 두게 되었다. 그리고 각 영역에 있어서 서로 다른 틀작업(framework)과 전문용어들이 발달하게 되었다.

2. 틀(framework)

일반적으로 조사에서 자료의 품질과 오차를 기술하데 있어서, 모형들은 조사수행 그 자체, 특히 자료수집과정을 집중적으로 다룬다. 모형들은 수학적이거나 (대수방정식에서 이상치로부터 벗어나는 수치를 오차항(disturbance terms)으로 표현하는) 도식적(자료수집과정의 수행요소를 묘사하는 개념적 모형)일 것이다. 개념

적 모형은 자료수집과정의 핵심으로서 면접을 강조한다. Hyman(1954), Kahn 그리고 Cannell(1957), scheuch(1967) 등의 연구를 토대로 하여, Sudman 과 Bradbun은 조사상의 응답효과에 관한 그들의 저서에서 보다 유용한 것을 제시한다. 이 도식적 모형은 조사원과 응답자 그리고 면접조사의 결과를 결정하는 조사수행(task) 간의 관계를 나타내주고 있다. 이 모형에 의하면 조사응답자의 전반적 품질을 좌우하는 주요요소들의 영향을 다음과 같이 파악할 수 있다.

- 조사연구자의 대리인으로써 조사원은 자료수집과정의 결과에 대하여 일차적인 책임을 가진다. Sudman과 Bradburn은 조사원의 공식적 지침, 조사원의 실제 행동, 그리고 조사원의 부가적인 역할 등 세 가지의 요소로 구분한다
- 응답자는 일반적으로 오차의 원인으로 간주되지 않았다.(불성실한 조사수행에 대한 일반적인 불만은 제외) 조사연구는 초기에 조사를 연구한 많은 학자들이 응답자에게 동기부여를 할 필요성을 언급했을지라도 당연하게 응답을 받아내도록 응답자를 다루는 경향이 있다. 하지만 종합적 접근방법에서는 응답자를 조사과정의 적극적인 참여자보다는 극복해야할 장애물로 보고 있다.
- 일반적으로 조사수행(task)에 주안점을 둔 조사에서 응답오차의 모형들은 응답이 조사의 최종목표-특히 분석에 필요한 자료를 제공하는-라고 규정되고 맞춰진다. 응답은 면접시간, 면접장소, 조사문항, 설문지작성, 수집할 자료유형, 그리고 개인의 기억이나 사회적 바람직성과 같은 면을 포함한다.
- 응답은 자료수집 활동의 결과이고 자료분석을 위한 원시자료이다. 대부분의 조사분석은 원시자료를 오차가 없는 것으로 처리하는데, 측정오차에 대한 통계적 접근방법에서는 응답을 개인적 자료에 대한 *참값에 응답편차* 또는 *응답효과* 라는 오차가 더해진 것으로 간주한다.

모든 조사과정과 조사오차에 관한 일반적인 모형이 이러한 요소들을 포함해야 한다는 것은 확실하다. 그러나 조사의 맥락이나 조사의 목표가 상이하기 때문에 이런 요소들만 고려하는 것은 충분하지 않다. 조사연구에 대한 다른 접근을 비교하는 것은 세 가지 주요한 부문들을 포괄하는 종합적인 틀작업을 제공할 필요가 있는 1장에서 다뤄질 것이다.

한 가지 가능한 틀작업은 조사를 위한 통계학적 설계에 관한 Kish의 저서에서 몇 가지 아이디어를 이용하는 것이다. Kish는 연구자가 조사설계 과정에서 다루어야 할 필요가 있는 세 가지 사항을 제시하였다. 필자는 이와 유사한 분류법이 오차의 원천 대부분을 포괄하는 차원을 분류하는 데 사용될 수 있다고 제안한다. 이들 각각의 "차원"은 그 자체가 다차원적인 개념이며 대표성, 무선화, 현실성이 해당된다.

조사연구가 응용사회과학을 다루기 때문에, 우리는 모집단 요인들에 대한 실제 조사에 근거해서 조사를 이해해야 한다. 사회이론이나 통계이론에서는 이런 요구사항들을 가지고 있지 않다. 하지만 이런 경험적 차원에서는 변수들간의 정도와 관계의 방향은 항상 모집단 요인들의 상황에 따라 다르다. 그래서 조사에서 얻어진 어떠한 추론도 모집단의 요인들이 적절하게 설정되어 있음에 근거해야 하고, 요인들의 하위 계급간 비교는 요인들 간의 비교 가능한 부분집합에 근거해야 한다. *대표성*은 확률 표본의 사용, 층화, 선택 편향의 방지 그리고 무응답의 고려와 같은 것들을 포함한다. 일반적으로 우리는 사회과학 내에서 발견되는 것들은 모집단의 가능한 모든 요인에 대한 상황과 환경에서 동일하게 적용될 수 있다고 생각하지 않는다. 실제로 사회 과학을 올바르게 다루는 것은 모집단의 부분집합이나 모집단내에서 발생하는 차이점 방법을 이해하는 것이다.

대표성은 조사에 포함되는 모집단 요인에 대한 관심설정을 반영한다. 특히 목표모집단이 요인들의 표본에 적절히 반영되는 것까지 포함한다. 완벽하게 구체화된 모형에서 모집단의 어느 요인을 표본에 포함시킬 것인가를 고려하는 것은 필요가 없다. 완전하고 완벽하게 구체화된 모형(이미 고려하고 있는 관계성이나 변수들에 영향을 주는 잠재성을 지닌 모든 변수를 가진)이 없는 경우, 대표성의 개념은 구체적으로 영역들(또는 하위 계급들)의 적절한 대표성, 선택적 편향의 방지, 그리고 차별적 무응답의 최소화를 뜻하기도 한다. 대표 표본이라는 용어는 통계학에서 파란만장한 역사를 가진다(예를 들어, Kruskal & Mosteller(1980), OMuircheartaigh & Soon(1981) 참조). 이 용어는 일반적으로 가치가 있는 것으로 인정되고 있다. 또한 (선택적 편향을 초래하는)선택력(selective forces)의 부재를 의미하는 것으로 받아들여지기도 한다. 이것이 원래 함축한 의미는 모집단의 축소판이나 거울이다, 이것은 때때로 전형적인 혹은 이상적인 사례로 간주되고, 모집단(cf. 층화)의 적절한 범위라고 할 수 있다. 이것은 확률 표본으로 가장 잘 명시되며, 확률 표본은 학계 조사 그리고 (대부분) 정부조사가 채택하는 방법이다.

무선화(그리고 이것의 반대개념은 *통제*)는 실험 그리고 복잡한 변수들의 통제와 관련이 있다. 조사에서 중요한 목적으로는 아주 드물게 정식 실험을 하기도 하지만, 측정오차(자료 내 왜곡)의 원천에 대한 정의나 이런 "오차"에 대한 크기를 추정할 때는 빈번하게 행해진다. 무선화는 허위의 상관관계와 효과를 잘못 파악하는 것을 피하거나 그럴 가능성을 줄이는데 사용된다.(무선화(또는 최소한 무작위 추출)는 또한 확률표본에서 대표성을 확보하는 데에도 사용된다)

*현실성(Realism)*은 이러한 맥락에서 두 가지 측면에서 문제가 되고 있다. *변수의 현실성*은 측정되거나 드러난 변수들이 기술하고자 하는 구성개념에 관련되는 것이다. *환경의 현실성*은 자료 수집이나 실험에 대한 설계가 연구자가 처해있는 실제환경의 상황과 유사한 것을 의미한다. 조사의 상황은 변수들과 환경들이 우리가 측정하기를 원하는 현실성에 더 밀접한 관찰 연구들과는 대조적일 지도 모른다. 이러한 차원들은 Campbell & Stanley(1963) 등이 사회조사연구의 평가를 기술하는데 사용했던 *내적타당도*와 *외적타당도*의 의미와 관련이 있다. 특정 조사 상황에서의 비교 타당도는 내적 타당도이고, 내적으로 타당한 추론을 특정 상황을 벗어나 일반화할 수 있는 것은 외적 타당도이다.

응답과정의 여러 구성요소들은 다음 장에서 소개된다. 그들 각각은 기본 모형의 다양한 요인들에 관련된다. 제 3장은 공식(정부)통계의 관점을 제시하고 *응답*에 대한 집중적인 논의가 있을 것이다, 이 전통은 조사연구의 *자연과학* 학파에서 여전히 따르고 있다. 제 4장은 *조사수행*의 요인들을 고찰한다. 제 5장에서는 우선 *응답자*에 초점을 둔 후, *조사원*과 그들간의 관계에 대해서 논할 것이다. 4장과 5장에서는 조사연구과정에 대해서 다루게 되는데 이에 대한 대부분의 내용은 시장조사와 여론조사에 관여하는 심리학자들과 사회조사연구와 정책연구에 관여하는 사회학자들과 사회심리학자들에 의해 이루어졌다. 제 6장에서는 최근 개발된 몇 가지 사항들을 사용하여 조사상의 측정오차를 재고찰하는 방법을 설명한다. 제 7장에서는 몇 가지 잠정적인 결론들을 제시한다.

3. 표본조사 관점의 영향-응답자료의 통계적 분석 : 공식 통계 및 조사통계

Kiaer(1897)를 비롯한 정부 기관의 창시자들에 의해 사용된 표본 조사는 전수조사의 완벽한 대안으로써 좀 더 상세하고 주의 깊은 측정에 대한 요구로 인해 필요하게 되었다. 1897년 Kiaer는 "사회현상을 좀 더 깊이 이해하기 위해…이것은 필수적이고…전체에 대한 일련의 특별한 설문을 구성하고…한 나라의 모집단에 대한 완전한 조사를 수행하는 것을 금하고, 심지어는 대도시의 모든 거주자에 대한 어느 것도"(p.38) 조사라는 것과 통상적인 정부 여론조사라는 것의 차이점을 보여주기 위해서 *표본*의 개념은 필수적인 것이다. 그 차이점이라는 것은 결과를 오염시킬지도 모르는 오차를 말하는데, 정부조사에 관여하고 있는 통계학자 1세대 등에 의해서 관심의 초점이 되어왔기 때문에 표본은 필수적이다. Kiaer는 반복조사(replication)-비교가능한 하위표본들의 집합을 간단하게 구성(본질적으로는 표본과정을 반복하는 것)-을 조사결과를 평가하는 수단으로 제안하였는데(p.51), 이것은 최초의 *총분산모형(total variance model)*이다.

Bowley와 다른 통계학자들에 의해 널리 알려진 이 접근법은 Neyman이 1934년 왕립통계학회에 제출한 고전적인 논문 "대표성 방법의 두 가지 측면에 대하여"에서 절정을 이루었는데, 이 논문은 *표본분포-표본설계와 표본과정의 모든 가능한 결과들의 집합-*에 대한 그의 개념을 구체화 한 것이다. 표본설계와 그렇게 해서 이루어진 표본조사에 대한 품질은 추정의 *표본오차*에서 요약될 수 있다, 일반적으로 "오차"라는 용어가 사용된다고 해도 이것은 분산이나 변동성을 순수하게 측정하는 것이지, 상식적인 의미의 오차를 측정하는 것은 아니다. 특히, 통계적 추정치들 중에서 선택을 할 경우 기술적인 문제로 삼는 것을 제외하고는 *편향* (또는 체계적 오차)은 일차적인 고려대상이 아니다.

Kiaer-Bowley-Neyman 접근방법은 사회조사분야에서 정의되었던 표본에 대해서 이후의 통계학자들을 위한 연속적인 문헌들을 만들어냈다. 고전적인 표본에 관한 문헌들의 연속물은 Yates(1949, UN Sub-Commission에서 1950년 세계인구총조사를 집행하는데 도움을 주는 통계적표본에 관한 매뉴얼로 사용하기 위해 집필), Deming(1950)과 그를 따르는 Hansen(1953)등, Cochran(1953), Sukhatme(1953), 그리고 Kish(1965)에서 완결되었다. 이러한 문헌들을 가지고 조사통계학자들은 그들의 분야를 조사추정에서의 부정확성에 대한 표본설계의 효과를 측정하는 것이라고 정의했다. 다른 고려사항들은 보조부문으로 분류하거나 이 책의 말미에 있는 장에서 다뤄질 것이다. 문헌들은 수리 통계학에서 부여한 상대적 가중치 면에서 상당한 차이가 있다, 가장 성공적인 교재는 Cochran(1963년 제 2판, 1977년 제 3판)으로, 그는 가장 수학적이고 비표본과 비통계적인 사항들의 영향을 가장 적게 받았다.

두 번째 분야는 인도의 Mahalanobis의 연구에서 찾아볼 수 있다. Kiaer처럼 그는 조사에서 파생된 추정의 정확성을 평가하기 위해 그가 상호관입표본이라고 불렀던 반복조사를 사용할 것을 주창하였다. 그는 이것을 "독립적으로 반복조사된 표본단위들의 네트워크"라고 정의하였다. 게다가 그는 조사에서 *인간의 매개 (human agency)*를 강조한 최초의 통계학자였다(1946, p.329), 그는 오차를 표본오차, 기록오차, 그리고 물리적 변동(불안정성) 오차로 구분하였다. 분산을 추정하기 위해 다양한 반복성이 "현장 연구자의 다양한 분야"에 의해 다뤄지고 그래서 표본오차뿐만 아니라 인간에 의한 오차도 정확성을 평가하기 위해서 포함되어져야 한다고 했다, 그는 또한 반복조사들 간의 중요성을 테스트하기도 했다. 또한 부분연구에 대한 또 다른 장점이(그의 상호관입표본을 사용하는)그것들이 정부기관의 보고서에서도 볼 수가 없었던 오차의 추정을 용이하게 한다는 것인데, Mahalanobis는 이것에 대해서 확실히 입증했다. 그의 초기 평가들 중 하나는, 1923에서 1925년 간 인도총독부(Indian Civil Service)가 행한 표본에 근거를 둔 것으로, 농작물 산출량에 관한 추정에서 편향을 보여준 것이다(1946, p.337).

반복조사는 통계학자들이 오차를 다룰 때 가장 기본이 되는 도구이다. 분산-어떻게 정의하든 반복조사시의 변동성-그리고 오차-정확한 값이나 참값으로부터의 체계적 편차)는 여전히 오차에 대한 통계학자의 접근방법에 대한 정보를 준다. 반복조사(또는 Mahalanobis의 상호관입표본)은 조사과정 그 자체에서 얻은 추정치의 변이를 추정하기 때문에 측정가능성을 얻을 수 있는 방법이다. 표본은 표본단위를 독립적으로 선택하고 그들간의 차이점을 전체 추정치의 내적 안정성이나 불안정성을 판단하는 수단으로 사용하는 것이다. 단순 응답 분산의 경우, 통계학자들은 응답자의 부분집합에 대한 관찰을 반복했고 단순히 그 결과를 비교했다, 이것은 대개 재면접 프로그램이라고 하는데 이것은 되풀이되었다는 의미에서 반복조사라고 한다. 조사원 효과의 맥락에서 반복조사는 조사수행 내에 있는 것이고, 비교가능한 사례들의 부분집합을 구성하고 그것들을 비교하는 것이다. 조사원 효과를 측정하기 위해서 응답자들은 여러 조사원들에게 무작위로 배정되고 얻어진 응답자료들은 비교된다. 통계이론을 통해 만약에 응답에 체계적 조사원 효과가 없었다면 조사원들의 업무량에서 우리가 예측할 수 있는 얼마나 많은 변이들이 있었는지 알 수 있다. 귀무모형보다 더 많은 분산이 있었다면 우리는 조사원 효과가 발생한 것으로 간주한다.

1960년대 초반에는 통계학자들이 조사오차를 고찰하는 관점이 진일보하였다. Hansen(1961) 등은 조사오차에서 "U.S. 미국 센서스국"의 모형로 알려지게 된 논문을 발표했다. 그들은 조사과정과 이 조사를 실시하는 조사기관의 안정적 특성으로 *필수적인 조사의 조건*들이라고 정의했다, 분산은 그러한 필수적인 조사의 조건들에 관련지어 정의하였다. Hansen과 그의 동료들은 "참값"이라는 개념에 문제가 많다는 것을 잘 알고 있었지만, 그들은 참값을 오차의 정의를 내리고 그에 따라 오차를 추정하는 가장 유용한 기초 개념으로 제안했다. 그들의 모형은 본질적으로 분산-공분산 모형이고, 상당한 일반화를 허용하며(예를 들어 Fellegi, 1964, 1974 참조), 조사 통계학자들에게 가장 많은 영향을 주고 있다. 특히 이것은 조사원과 감독자의 효과와 가계내에서 상관오차(correlated error)가 발생할 가능성을 인정하고 있다.

이와 거의 동시에 Kish(1962)는 조사원 오차를 다루는데 있어서 분산분석(ANOVA)모형이라는 또 하나의 대안적인 기술적 접근방법을 사용했다, 이것은 Yates 등이 가장 선호한 접근방법이었고 농업 통계학자들의 실험설계관점으로부터 유래한 것이다. 통계학자들은 다시 그것을 사실성있게 단순화시켰고 ANOVA 모형은 그들의 모형구조 내에 수용될 수 있었다, 조사원 효과는 그가 면접한 각 응답자들의 응답내용에 미치는 부가적인 효과로 보고 있다. 이러한 접근방법은 조사에서 다른 매개자들(코드부여자, 감독자, 편집자 등)의 효과에 쉽게 일반화될

수 있다(예를 들어 Hartley & Rao, 1978), ANOVA모형의 한가지 결점은 범주형 자료분석에는 그다지 도움이 되지 않는다는 것이다.

이러한 두 가지 접근방법은 일반적으로 단일변수(또는 비율)를 사용하여 분산을 추정하는 목적을 가진다. 조사의 핵심은 모집단의 기술적(記述的) 특징을 추정하는 것이다. 총분산이란 이 추정치에 영향을 주는 다양한 오차 원천들의 총합이다. 단순 랜덤 표본추출(SRS)에 근거를 두고 오차 없이 측정된 변수의 표본평균의 분산으로부터 시작하여, 또 다른 요인들을 구성하는 것은 각각 별개의 오차 원천으로 대표되는 분산이 추가되는 것이다. 이러한 진행과정, 무응답, 비포함 그리고 측정의 오차들은 모두 이 접근방법에 포함된다. 일반화되어지는 모든 편향들의 역시 추정치의 총오차로 평균제곱오차에 주어져 그 안에 추가될지도 모른다. 단일 기술 모수 추정에 초점을 두는 것은 부분적으로는 정부통계에서 온 것이고(이것은 모집단의 중요한 단일특성, 예를 들면 실업률과 같은 것을 추정하는 것과 밀접한 관련이 있다), 부분적으로는 분포의 모수에 대한 구간추정에 대한 일반적인 통계학의 전통에서 온 것이다.

조사 통계학의 접근은 20여 년 동안 그런 기술적 모수에 관심을 두고 있다. 하지만 1980년대 자료분석에서 통계 모형의 사용과 설계중심과 모형중심 요소들의 추론 간의 조사표본관리는 조사에서 통계모형로 직접 연결되는 응답오차의 결합을 초래했다. O'Muircheartaigh & Wiggins(1981)은 항공기 소음이 불쾌함에 미치는 영향에 관한 로그선형분석에서 조사원오차를 모형링하였다, 그리고 Anderson & Aitkin(1985)은 조사원 변이를 조사하기 위해 분산성분모형(다중수준모형으로 알려져있는)을 사용하였다, Hox, de Leeuw & Kreft(1991), Pannekoda(1991)그리고 Wiggins 등(1992)은 이러한 방법들의 응용하고 발전시킨 사람들이다.

이와 같이 최근에 전개된 사항은 통계적인 경향에 변화를 가져다 주었다. 즉 측정오차에 대한 추정치의 "정확성"요소는 종래에는 조사/통계적 관점이었다가 측정오류가 조사자료에 대한 중요한 분석에 영향을 미친다는 광범위한 관점으로 변하게 된 것이다. 이와 같은 후자의 관점은 다른 조사에도 전통적으로 내재되어 있었던 것이다.

4. 조사수행(TASK)

조사수행은 조사 질의와 조사가 이루어지는 환경의 핵심이다. 이 작업에서 가장 중요한 것은 조사설문지 또는 조사일정이다. 조사표의 정확한 본질 및 기능에

대해서 조사연구의 초기에는 보편적인 의견의 일치가 있지 않았다(지금도 그렇지만). 초기의 연구자들 중에, Galton-사회조사에 대한 그의 견해는 잘 알려져 있지 않다-은 *인간의 정신능력에 대한 연구*(1983)에서 조사표에 대한 네 가지 필수요건을 제시했다(Ruckmick, 1930), 조사설문지는 "(a) 빠르고 쉽게 이해되고 (b) 간단한 응답이 가능하며 (c) 조사의 주제를 적절히 반영하며 (d) 응답자들이 그들의 응답과 이와 유사한 주제에 대한 설명을 빠짐없이 자유롭게 쓰도록 해야한다. …… 이러한 사항들은 주어진 설문에 대한 응답보다 더 유용하고 흥미로운 것으로 입증되었다". 사회개혁가 Booth(1889)는 자료수집에 대한 문제를 도구적인 관점으로 보았다. "가장 밑바탕에 있는 생각은 … 나에게 필요했던 모든 사실들은 누군가에게도 알려져 있다는 것이고 그러한 정보는 간단하게 수집되고 함께 다루어져야 된다는 것이다."

그러나 얼마 되지 않아, 사회조사 연구자들은 그들의 자료에 대해 신뢰성과 타당성을 의문시하기 시작했다. 그리고 실험하거나 그들의 방법을 테스트하고 평가하기 위한 다른 조사로 수행되는 사례들이 생겨나기 시작했다. 심리학자들은 물론 설문지의 문구를 달리해보면서 그 때마다 발생 가능한 효과들에 관심을 갖게 되었다, 질문지형식에 대한 실험에 대해 설명한 Muscio(1917)는 Lipmann이 1907년도에 출판했던 조사에 대해서 언급했다. 사회조사와 직접적인 관련이 있는 초창기의 것 중 하나는 Hobson이 1916년에 *미국통계학회지*에서 발표한 것으로 우편조사에 있어서 다양한 조사표 유형과 질문지의 문구에 관하여 비교한 것이다. 이러한 영역에서 조사의 대부분은 Link & Gallup이 선두로 활약하던 1930년대와 방법론에 관한 연구들이 출간되기 시작한 1940년대에 이르러서야 등장하기 시작했다.

이 분야에도 조사연구자들간에 다양한 견해들이 있었다. 특히 1920년에서 1950년까지의 많은 기간동안에 외적태도측정척도(Thurstone, Likert, Guttman)의 개발로 심리학자들로부터 척도법이 사용되었다. 그들은 자신들의 측정에 대한 특징점들을 묘사하기 위해 신뢰성과 타당성이라는 용어를 사용하면서 자신들의 전문용어들을 도입하였다. 이러한 용어들은 과학적인 의미를 내포하고 있으며 측정에 대한 부정적인 측면보다는 긍정적인 측면을 강조한다. 그 용어들은 통계학자들의 분산과 편이-추정의 강점보다는 결점을 강조하지만 광범위하게는 유사점을 가진 용어-와는 대조된다. 심리학적인 전통은 태도척도의 구성과 같은 그들의 도구를 시험하기 위해 내적일치도, 반분실험 그리고 검사-재검사 비교들을 사용하면서 또한 실험의 가능성과 실험적인 비교점들을 강조했다. Cattell은 응용심리학을 상업적으로 활용하기 위해 Psychlogical Corporation을 창립했다, 1932년에는 미국심리학회(ASA: American Psychological Association)의 절반 가량은 이 회사

의 주주가 되었다(Converse, 1986, p.107). 조사방법론의(또는 응용의) 조사에서 가장 활발한 활동을 보인 건 Henry Link였다. 그는 1930년에 Psychlogical Corporation에 합류했는데 이후 20여 년 이상 조사와 척도의 방법론에 관한 통찰력있는 논문을 작성하였다.

대체로 시장조사 사업을 수행하는 기업들은 자신들의 독자적인 기법에 대해 진지한 연구를 했다. Cantril은 갤럽-여론조사의 아버지라고 일컬어지는-및 영미 여론조사기관과 귀중한 제휴를 맺었다. 1936년에서 1949년까지 갤럽은 약 400회의 분할투표방식(split ballot experiment)의 조사를 실시하였다. 이들 중 대다수는 Cantril에 의해 설계된 것으로 그는 "*규제나 제한 없이 결과를 공표 할 수 있도록*"허용해주었다. 이러한 노력의 주요 결과는 *여론조사평가기준(gauging public opinion)*(1944)인데 이것은 조사설계, 실사, 분석의 방법에 대한 연구들의 개론서이다. 이후 50여 년 동안 이와 같은 각 분야별 공동연구의 사례들이 보여지지 않은 것이 아쉽다.

이 기간 중 다른 주요한 연구는 또 하나의 이례적인(다소 강제적인)분야별 제휴였다. 2차 세계대전 중 1941~1945년 동안 Stouffer의 지시로 U.S. Army의 사기진작사단(Division of Morale)에 조사부(Research Branch)가 설치되었다(Converse, p.165. 우수논문에 대한 *Passim).* 이 전쟁을 계기로 군대 설치에 관한 흥미로운 사항들에 대한 지속적인 여론조사 프로그램이 생겨났다. 이 조사는 사회과학을 발전시킬 목적으로 수행된 것이 아니었다, 군인들의 태도와 견해에 대한 근거가 되는 정보를 제공하는 실용적인 업무를 위해 부서가 설치된 것이다. 조사부의 직원과 자문관들은 전쟁 중 수집한 상당한 자료를 상당히 심도 있게 분석 처리하여 전쟁 후 *2차 대전 중(1945-50)의 사회심리학 연구*라는 4권의 책으로 출간하였다. 1권과 2권은 *미국 군인(the American Soldier)*라는 공동 제목으로 루틴적용연구[3]라는 적절한 분석을 통하여 얻게된 방법론적이고 이론적인 흥미로운 결과점들을 찾게된 빛나는 사례들이다.

1950년대와 1960년에는 조사도구의 여러 장단점들에 대한 연구들이 발표되었다(Parry & Crossley, 1950; Cannell & fowler, 1963의 타당성검증연구). Sudman & Bradburn(1974)은 당시까지의 조사에 대한 개관을 소개하고 조사응답과정의 명시적인 모형을 정립하였다. Schuman & Presser(1981)는 자신들과 다른 연구자들의 조사표 문구를 기술한 저서를 발간하였다.

---

3 Routine applied work : 컴퓨터 루틴(프로그램에 의한 컴퓨터의 일련작업)

조사수행 관련 연구들은 *타당성 정보를 사용하는지 내적 일관성 분석*을 사용하는지에 따라 구분이 된다. 타당성 정보의 경우-덜 빈번하지만-정보는 조사표에 대한 개인적(또는 집단의) 응답에 대한 타당성을 체크하는 것으로 조사의 외부에서도 얻을 수 있다.

그런 정보들은 다량의 조사에서는 얻기 어렵고, 행동정보에 국한된다. 방법론적 타당성은 두 가지 범주에서 연구된다-*식별연구(identification)*는 특정오차를 식별하고 정량화는 것을 목적으로 하며, 진단연구*(diagnostics)*는 특정오차를 만드는 요인을 파악하고 그것을 제거하거나 감소시키기 위한 절차를 고안해내는 것을 목적으로 한다.

수년간 측정오차의 정도에 잠재적인 영향을 미치는 것으로 알려진 조사수행의 가장 중요한 특징은 다음과 같다, 그것들은 면접장소와 조사실시방식, 응답자 선정, 조사표의 길이와 질문의 순서 그리고 질문지구성, 조사표의 문구, 질문의 길이와 난이도, 질문의 중요성, 그리고 사회적 바람직성에 대한 위협의 수준과 가능성이다. 역시 영향을 주는 것으로 조사되거나 가정된 요인들은 응답자부담, 기억효과, 그리고 개방형 대 폐쇄형 질문에 대한 고전적인 논쟁, 자료수집방식, "모름"에 대한 명시적인 지문, 척도점의 개수 등이다.

조사수행 변수에 대한 문헌들에서의 어려움은 존재하는 것으로 알려져 있는 많고 다양한 효과들을 설명하기에 적합한 명확한 패턴들을 찾을 수 없는 것이다. 명백한 이론적 틀이 없기 때문에 효과들을 분명하게 분류하는 것도 어렵고 조사모형에 대한 일반적인 이론을 정립하기도 어렵다. 통계학자들의 경우에는 오차를 정량화하는 단순한 내용이지만(총분산이나 평균제곱오차의 효과를 더해서), 관찰한 여러 효과들을 뒷받침해주는 심리학적, 사회학적, 또는 실용적인 만족스런 원리들이 발견되지 못하고 있다. 응답과 그 응답효과가 상황에 따라 달라질 수 있기 때문에 어떤 결과든 다양한 맥락에서 반복되어야 한다는 주장이 있었다 (Cantril, 1944, p.49). *"단 하나의 의견자료는 오직 개인적이고 사회적인 맥락에서만 의미가 있기 때문에, 여론조사에 의해 이루어진 많은 단일 질문에 대한 응답은 다른 상황에서 같은 주제를 가지고 수행되어진 응답과 비교를 하는 것은 가장 중요하다."* 조사응답의 맥락-중심적인 본성에 대한 연구들은 질문지-관련효과(question-related effect)에 대한 우리의 이해를 강조하는 이후의 연구의 효시가 되었다.

5. 조사원과 응답자

사회조사연구에서 조사원의 임무와 행동을 응답자의 임무과 행동으로부터 완전하게 구분하는 것은 어렵다. 이 장에서는 특히, 우리의 사고방식이 변함에 따라 사회조사연구에서 그들의 역할이 어떻게 발전하고 변화되었는지를 다룰 것이다.

사회연구의 초창기에는 조사 상 자료수집의 본질적인 면을 표준화하는 도구가 없었다. 고전적인 빈민층 연구의 수많은 사례들의 경우, 자료를 수집하고자 하는 사람들과 직접적인 의사소통을 하지 않았다. 예를 들어, Booth의 경우 "*사실이 존재했다*", 그리고 남은 일은 사실을 수집하는 방법을 찾는 것이었다. 그는 학생과 그들의 부모 그리고 그들의 가정형편에 대해서 세세한 자료를 제공하기 위해 학교의 출석 담당자와 같은 "전문"정보 제공자를 이용했다, Beatrice Webb는 이러한 절차를 "대량의 면접 방법(method of wholesale interviewing)"이라고 칭하였는데, 이 당시에는 조사원 수행에 대한 관심이 없었다. 조사원이 응답자를 만나지 않고 정보를 적는 일은 계속되었고, 몇 가지 정보들은 제한된 맥락에서 여전히 이렇게 수집되고 있다. Yates(1949년 판과 그 후속판)는 "*집안에 습기가 많은지의 여부를 집주인에게 묻기보다는 직접 그 집을 조사하는 것이 더 낫다.*"고 말하였다. 1896년 Philadelphia에서 DuBois는 도시의 흑인 인구에 대한 자료를 얻기 위해 가가호호 호별 방문을 포함한 다양한 방법을 사용했다(DuBios, 1899). Booth 연구 10년 후, Rowntree(1902)는 조사원을 이용하여 가족으로부터 직접 정보를 얻어냈다.

조사원을 어떻게 이용하는 지에 대해서는 다양한 변수가 있다. 면접역할에는 두 가지 극단적인 유형이 있다. 그 하나는 응답자와 "대화"를 나눔으로써, 그리고 동시에 자료를 적지 않으면서 정보를 얻는 전문 조사원이다. 다른 하나는 빈 종이와 준비된 질문지를 가지고 질문을 하는 "질문지형" 조사원이다. 사회정책개혁자들 은 모든 경우에 있어서 공식적인 도구를 사용하고 있다, 공식적인 도구들이란 조사표의 순서표(조사표에 대한 *면접 순서*라는 용어가 생겨나고 면접도구들을 가지고 수행하게 되는)또는 점수표나 목록카드(후에 조사표가 되는)와 같은 것을 말한다. 시장조사와 여론조사의 연구자들도 이와 유사한 이분법적인 방법을 사용했다, 전문 조사원들은 기업의 경영진이나 상인들을 만날 때 이용되며, "질문지형"조사원은 일반인을 만날 때 이용되었다. 질문지형 조사원들은 인쇄된 질문지 세트를 가지고 응답자가 보는 앞에서 해당 공란에 응답을 기입했다.

사회조사의 정의가 모호하고 일반적으로 인정된 표준화된 도구 세트가 없었기 때문에, 사회조사는 초기에 그 뛰어난 가치를 인정받지 못했다. 미국 사회학회지(AJS)의 Thomas(1912)는 " … *주 작업인 면접은 향후 관찰에서도 비교할 목적으로 사용할 되어지는 오류의 본체로 취급된다.*" 라고 하였다. 미국통계학회지(JASA)의 Gillin(1915)은 *"조사는 실업자가 되었을지 모르는 사람들의 수다거리가 된다거나 오락거리가 되는 위험에 처해있다"*고 하면서 품질 관리의 부재에 대한 경향을 비난하였다. 조사원과 그들의 행동은 언제나 조사의 품질에 있어 핵심이다. 1920년대 초기의 연구자들은 조사원이 미치는 악영향에 대한 연구가 수행되었다(Rice, 1929). 1930년대에는 직장면접 같은 맥락에서의 면접에 대한 다양한 비판점들이 등장하면서 조사원의 행동에 대한 표준화를 시도했다(Hovlans & Wonderlic, 1939; Bingham & Moore, 1934). 1940년에는 조사원의 효과에 대한 좀 더 체계적인 조사가 수행되었다.

Katz(1942)는 사무직과 노동자계층 출신의 조사원들을 비교하고 사무직 조사원의 면접결과가 더 보수적인 경향을 보이는 것을 발견했다, Cantril과 그의 동료들은 조사원 편향, 조사원의 척도평가에 대한 신뢰성, 조사원의 훈련, 그리고 친밀감형성(Rapport)에 관한 주제들을 고찰하였다(Cantril, 1944). Wechsler(1940)은 조사원들이 설문을 설명하는 방식에 관심을 기울였다, *"조사원은 자신이 no라는 대답을 받아들이지 않을거라는 생각이 들도록, 도전적이고 긍정적인 말투로 질문을 할지도 모른다. 그는 너무 친절하고 부드러워서 피면접자가 내는 목소리가 이의를 제기하는 것처럼 보일까봐 응답을 주저할지도 모르는 억양을 사용할 수도 있다. … 그러한 사례들은 개인적으로는 사소하게 보일지 몰라도, 티끌 모아 태산이 될 수 있다."*

점차적으로 조사원의 행동을 표준화시키도록 압력이 가해지자 선두의 조사기관들은 이에 반응을 보였다. 1942년 여론조사계간지(誌)(POQ)에서 Williams는 국립 여론 조사 센터(NORC)를 위한 *조사원에 대한 기본적인 지침*을 기술했다. 이런 지침들은 그들의 취지를 명백하게 밝힌 하나의 모형이고 매우 주의 깊게 표준화된 면접실무를 나타낸 것이다. 1946년 Central City Conference는 면접에 대한 토론이 행해졌다. Molone가 가장 이상적인 조사원으로 묘사한(37세의 기혼여성, 정책에 대해 적대적이거나 열렬한 추종자도 아니며, 지침을 이해하고 따를 수 있는) 이러한 전형은 오랫동안 유지된다. 같은 섹션에서 Tamulonis는 그녀가 특정 질문에 대해 극단적인 견해를 가진 조사원보다는 중도적인 조사원이 자신의 응답자들이 편견을 갖도록 유도할 가능성이 더 적다는 것을 보여주려고 실시했던 실험에 관해 보고했다. Sheatsley(1947-48)는 1980년대 Cannell에 의해 만들어진 대화식 코드 부여 방식의 효시가 된 조사원의 보고서양식의 사용에 관한 조언도 했다.

1950년대는 Hyman(1954)과 Kahn & Cannell(1957)이 면접에 관한 중요한 두 권의 책들을 출간하였다. 이 당시에는 면접조사의 기초사항들이 확고하게 정립되어 왔, 조사원들은 중립적인 수단으로써 행동하도록 기대되었다, 중립적인 수단이란 어떤 특정한 대답에 대한 선호나 놀라움을 표현하는 것 또는 특정응답에 대해 기뻐하는 등의 어떠한 것도 암시하지 않고 글자 그대로 그들의 질문지를 읽어주는 전문가를 의미한다. 이상적인 조사원은 엄격한 훈련으로 보강된 가장 구조적인 역할을 해야한다. 그리고 묻는 질문에 악영향을 줄 수 있는 별도의 역할을 해서는 안 된다.

이 당시 조사대상(응답자)의 역할은 중요치 않거나 의미가 없었다. 언제나 어떤 방법으로 정보를 얻을 것인가에만 관심이 있었다. Booth는 그의 조사방법인 "대량의 면접 방법(method of wholesale interviewing)"을 가지고 그가 정보를 얻기를 바라는 대상으로부터 개별적인 상담을 하지 않았다. Rowntree & DuBis가 응답자에게 직접 질문을 했다고 해도 응답자는 효과적인 조사를 방해하는 장애물과 같다는 생각이 일반적이었다. 자료를 수집하는 것에 대한 문제점이 거론되어도 주로 응답자를 *중심*에 두지 않는 것이었다. 때로 조사작업이 너무 오랫동안 이루어지거나 지루하다는 우려가 제기되었다. 1936년 White는 소비자구매 연방조사(Federal survey of Consumer Purchase)에서 *"응답자는 일년동안 그가 몇 리터의 석유를 구입했는지 알고 있고 다리통행료나 나루터통행료로 지불한 금액에 대한 계산서를 가지고 있는 것으로 여겨진다."*White는 그런 자료를 보관하고 있는 누군가에 대해서 *"그들은 너무 비정상적이어서 그들의 보고가 일반적인 대중을 대표할 수 없다"*라고 했다. 하지만 대체로 응답자 또는 응답자 피로는 아무 문제가 없는 것으로 보인다. 1946년 Central City 회의과정에서 "그들은 설거지보다 자신들의 의견을 제시하고 싶어한다."라는 표현에서 응답자(아마도 여성차별)에 대한 관심이 부족했던 것으로 보인다.

심지어는 사려 깊은 방법론자들이 갖고있는 응답자에 대한 일반적인 견해조차도 그들이 조사과정에서 상대적으로 수동적인 역할을 하고 있다는 것이다, 예를 들어 Sudman & Bradburn(1974)은 "응답자의 가장 기본적인 역할은 조사원의 질문에 응답하는 것이다."라고 하였다. 주 관심사는 응답자가 응답하도록 동기부여를 하는 것이다. Gates & Rissland(1923)는 이러한 관심사를 공식적으로 다룬 최초의 인물들이었고 Cannell(1975)등도 동기부여의 중요성을 강조하였다.

상기한 조사원과 응답자 역할에 대한 견해들에 영향을 미친 것은 대규모 조사에서 통계담당자들의 기본적인 요구사항들이다. 조사방법론자들은 방법상 발생가능한 결점들을 우려했다. Cannell & Kahn(1968)은 한 비평 논문에서 응답자와

조사원의 *상호작용*이 면접의 중심이라는 지적을 했다. 그 이후로 수행된 일련의 연구들에서는 면접시 발생하는 상호작용에 대한 더 복잡한 견해들이 제시되고 있다.

Cannell과 그의 동료들이 행한 연구(1981년에 Cannell 등에 의해 요약)는 면접(예를 들어, 약속절차)에서 응답자의 관여를 드러내게 하고 응답자에 대해 적절한 행동을 하도록 강화된 조사원으로부터 피드백을 얻는 방식을 고안하게 되었다. 그들은 또한 면접에서 상호작용을 위해 행동의 코드부여체계(behavior coding scheme)를 개발했다. 이것은 조사의 질문과 설문지의 사전조사(pretesting)를 위한 정교한 도구가 되었다(Cannell 등, 1989).

친밀감형성(Rapport)은 언제나 성공적인 면접의 열쇠이다, 그 용어는 응답을 위한 만족스런 동기를 이끌어내는 조사원의 역할의 일부로 성공을 의미하는 품질의 다양성을 일컫는다. 친밀감형성에 대한 최근의 문헌들은 공식적인 또는 전문적인 스타일의 한 면과 비공식적, 대인관계, 사회 감정적(socio-emotional) 또는 개인적인 스타일의 다른 면 사이의 구분을 가진 *"면접 스타일"*이라는 용어로 대체해서 사용하고 있다(Schaeffer, 1991). Dijkstra & van der Zouwen(1982)은 *"연구자들은 응답자 행동에 영향을 미치는 과정에 대한 이론을 가지고 있다."*고 주장했다, 면접동안에 이루어지는 상호작용의 세부적인 연구는 그런 이론을 개발하는 하나의 방법이다.

물론 구조화된 면접의 비판도 끊이지 않았는데, 자계식질문지(self-completion questionaire)에 대해 제기된 비판들과 유사하였다. 사회연구에 조사(survey)를 더 많이 사용하게 되면서 이러한 비판점들은 다시 나타나기 시작했다. Cicourel(1964)은 연구자의 과제와 응답자의 인생에는 필연적인 관계가 없다고 했다. 이러한 관점은 응답자의 관점에서 면접 조사를 고찰하는 문헌들에서 다시 나타나고 있다.

면접조사의 분석에 대한 더욱 질적인 접근은 응답자의 역할에 초점을 두거나 이 문제를 고려할 때 응답자에게 최소한 동일한 지위를 부여하는 경향이 있다. 면접에 대한 몇 가지 분석에서 Suchman & Jordan(1990)은 응답자들의 면접조사에 대한 인상이 심문이나 시험에서 대화 심지어는 치료에 이르기까지 다양하다는 점을 기술했다. 인지심리학에 공헌을 한 Tourangeau & Rasinski(1988)는 응답과정의 모형을 제공했다. 그것은 초기 조사문헌, 특히 Cannell의 연구에서 기술한 것을 발전시킨 것이다. 이러한 절차모형은 만약 믿을 수 있고 타당한 정보를 얻기 위해 왜 응답자가 적극적으로 참여해야 하는지 몇 가지 이유를 명확하게 설명한다. "대화분석"이라는 좀 난해한 분야도 우리가 조사응답과정을 이해하는데 도움이 되는 가능성을 보여준다(Schaeffer, 1991).

이런 모든 연구들이 기여하는 한가지 공통점이 있다. 대체적으로 유일한 관심의 초점이었던 조사원을, 중립의 매개자로 활동하도록, 특정응답에 대해 누구나 관계없이 지침을 따르도록 하는 것으로부터, 각자의 책임과 기대가 있는 의사소통인 "연설사건"에 참가하는 조사원과 응답자 양쪽 모두를 고려하는 것으로 전환된 것이다. 현재 이에 대한 한 쪽의 생각은 응답자 권한부여(그것의 가장 극단적인 것은 응답자를 연구자의 과제로 삼는 것이다)이고 다른 쪽의 생각은 허용 가능한 상호작용을 세세하게 설명하는 것이다, 그래서 응답자행동의 중요성을 알고 있다면 그것을 자유롭게 하기보다는 통제하는 것이다. 이런 양쪽의 생각들은 전문 조사원과 질문지형 조사원이 자료수집의 양극을 이루었던 100년 전 상황과 유사하다.

6. 최근의 발전

이 장에서는 오차와 오차의 처리에 대한 우리의 관점을 변화시켜주었던 현대의 조사방법론의 네 가지 예를 제시한다. 그것들은 전체를 망라하는 것이 아니라 이와 관련된 문제들의 범위를 예시하기 위해 선정한 것이다.

6.1 통계적 모형

1955년 조사 표본에 관한 비교적 이론적인 한 논문은 통계학자의 자신들의 역할에 대한 견해를 근본적으로 변화시키고 조사자료의 통계분석으로 오차를 수용하는 가장 적절한 방법을 모색하는데 (최소한 일부라도) 도움이 되었다. Godambe(1955)는 *모든 유한 모집단*을 가장 적합한 유일한 추정량(모집단 평균)은 없다는 사실(그 자체는 놀랍거나 특이하지 않은)을 증명하였다. 이것은 1960년대 중반에서 시작하여 이십여 년 동안 지속되어 모형중심추정(model-based estimation)으로 알려진, 즉 표본설계(요소선택의 결합확률)에 근거한 분석이 아닌 표본의 구성여하에 관계없이 분석 내에서 다양한 변수들이 서로 관계되어 있는 공식적인 모형에 근거한 관심이 폭발적으로 증가하게 되었다.

이런 경향은 경제학, 심리학 그리고 사회학에서 비교적 복잡한 분석을 상당히 많이 필요로 하였기 때문에 더 강화되었다. 이러한 분석에 사용되는 통계방법은 대개 일반선형모형(generalized linear model)에 근거를 둔 것인데, 거의 언제나 독립항등분포(i.i.d.; independent identically distributed)관찰을 가정하고 때때로 오차분포의 정규성을 가정한다. 이 모형은 구체성과 검증을 강조했다.

이와는 대조적으로 조사 통계학자들의 접근방법은 전통적으로 설계중심이었다. 3장 언급한 바와 같이, 일차적인 관심은 표본이 추출된 유한모집단을 추정하는 것이고, 응답오차는 전수조사의 또 다른 왜곡으로 개념화된다. 가중치는 표본설계와 무응답을 고려하여 추정치를 조정하기 위해 사용된다. 그리고 표준화된 오차는 표본군집의 효과와 조사원과 다른 매개자들의 상관효과(correlated effect)를 고려하여 산출한다. 이렇게 수정된 표준오차는 기술통계량의 신뢰구간을 나타내주거나 그런 통계량에 대한 가설검증에 주로 사용된다.

이런 접근방법들의 차이점은 조사통계학자들이 각각 무선화와 통제에 서로 다른 강조점을 두고 있다는 것이다. 조사통계학자들의 경우 최우선적인 기준은 표본설계를 통한 *대표성*을 위한 무선화이지만, 실험통계학자들과 실체적 분석가들은 외생변수의 *통제*를 위한 무선화와 모형에 강조점을 두었다.

최근 컴퓨터의 성능과 사용가능성의 획기적인 증가로 통계적 방법론의 발달은 서로 상충하는 접근방법들이 조화될 수 있다는 희망을 안겨주었다. 다수준모형(multilevel modelling)(분산분석모형을 발전시킨 것, Bryk & Raudenbush, 1992; Goldstein, 1995; Longford, 1993)은 본래는 학교, 교사, 학급, 학생 등과 같은 교육시스템에서 계층적(위계적) 자료구조를 분석하기 위해 설계된 것이다. 이 틀은 조사관찰치에 강조를 둔 군집화 *그 중에서도 특히,* 표본설계, 응답오차, 그리고 동일 개체에 대한 반복관찰을 다루기 위해 응용된 것이다. 이 분석에서는 그것들을 무시하거나-표준통계분석의 경우- 조정치로 가산하기보다는-표준오차에 대한 설계효과나 조사원효과를 산출한 결과- 분석 자체에 명백하게 이 요소들을 분석에 포함시킨다. *잠재변수모형(latent variable modelling)*(Joreskog, 1973; Joreskog & Sorbom, 1989; Bartholomew, 1987)은 정신측정학의 척도법에 관련되는 문헌에서 소개된 것이다. 이 분석에서는 자료와 오차의 분산-공분산 구조를 동시에 명백하게 모형화한다. 각각의 맥락에 가장 적절한 가정을 정립하는 것은 어렵지만, 이러한 접근방법들이 상당한 잠재력을 가지고 있다는 것은 의심의 여지가 없다.

6.2 인지적 접근방법

조사응답과 이와 관련된 오차의 맥락중심의 본질은 조사수행관련, 조사원관련 그리고 응답자관련 효과들에 대한 일관성있는 접근방법에 있어 상당한 장애물이 된다. 조사에 대한 초기의 많은 연구들이 심리학자들에 의해 수행되었다 해도 조사자료수집에 대한 공식적인 이론적 고찰은 없었다. 1978년 왕립통계협회와 (영국)사회과학연구 위원회의 통계위원회에 의해 영국에서 개최한 회의에서

Baddeley에 의해 조사에 있어 최초로 기억과 관련된 주제에 대한 심리학자의 논의가 이루어졌다.

그 후 미국 사회과학연구위원회에 의해 개최된 회의는 조사연구자들, 통계학자들, 심리학자들이 함께 모여 여러 학문분야들의 기법을 결합함으로써, 특히 인지심리학과 사회심리학의 어떤 아이디어들이 조사 방법을 향상시킬 수 있는지 살펴보면서 조사방법의 발전 가능성을 모색했다(Jabine 등 1984). 이로 인해 직접적 혹은 간접적으로 질문지설계의 방법론, 특히 도구설계를 고려한 좀더 확장된 틀의 개발이 상당히 진전되었다. Tanur(1992)에서 상당히 가치 있는 논문을 볼 수 있다.

*조사표에 대한 인지적 접근*의 목표는 응답생성과정을 이해하고 효과에 대해 "안정적인" 예측을 하는 지표들, 질문지설계에서 실제적인 질문에 대한 해결점, 그리고 질문지형식에 대한 "규칙들"을 만드는 것이다. 이러한 방향으로 진전한 한 가지 예는 Schwarz와 그의 동료들이 동화와 대조효과에 대한 연구를 한 것이다(Schwarz 등, 1991)

6.3 컴퓨터 보조 방법

조사자료수집에서 컴퓨터보조방법의 개발은 두 가지 다른 동기에 의해 이루어졌다. 그 중 하나는 특히 북미의 전화면접의 방향에 대한 움직임인데, 이것은 특정조사에서 많은 조사원들이 동시에 작업을 하는 중앙집중식 전화시설을 설치하는 계기가 되었다. 이것은 자료입력과 자료분석을 위해 동일장소에서 메인프레임 컴퓨터의 성능을 유용하게 사용하는 기회를 제공해주었다. 그리고 설문 자체에 대한 컴퓨터보조면접도구개발, 조사표 기입경로, 그리고 시스템으로 통합된 자료에 대한 다량의 체크를 가능하게 해주었다. 또 다른 하나는, 조사과정에서의 난점들과 병목현상으로 인해 특히 유럽에서 컴퓨터보조자료입력시스템이 개발되었는데, 자료의 입력과 동시에 자료의 편집이 이루어졌다. 이런 두 가지 동기로 인해 면대면, 전화, 그리고 자계식 조사를 위한 가장 강력하고 유연성있는 컴퓨터보조 시스템이 개발되었다.

컴퓨터보조방법에는 세 가지 특징점이 있다. 첫째, 시스템은 이미 *유연하다*, 그리고 앞으로 더 많이 유연하게 될 것이다. 조사도구는 필요하다면 불러올 수 있는 긴 자료들로 체계적인 구조화를 할 수 있다. 둘째, 시스템은 또한 본질적으로 *휴대가 가능하다*. 실제로 도구의 크기는 거의 제약이 없고, 이것을 운반하는 조사원의 짐이 되지 않을 것이다. 셋째, 시스템은 *상호작용이 가능하다*. 구체적인

면접 일정이 응답자의 요구로 조정가능하고, 응답자가 더 선호하는 것으로(또는 면접과정 중에 응답자에 의해 수정된 것으로)조정될 수 있다.

이 시스템의 강점을 설명하는 두 가지 사례가 있다. 소비자지출조사(consumer expenditure survey)에서는 긴 코드구조를 이 도구에서 사용할 수 있을 뿐 아니라 조사표의 작성순서(사람별 또는 활동별)가 면접 중인 응답자에 의해 결정될 수 있고 다양한 주제에 따라 다르게 설정할 수 있다. 이로 인해 조사표가 하나의 구조일 필요가 없게 되었는데, 최소한 일부 응답자에게는 "응답자 비친화적"인 구조가 될 지도 모른다. 패널조사에는 항상 독립적이거나 비독립적인 재면접이 이뤄져야 하는지 결정하는 문제가 있다. 가장 적절한 컴퓨터보조방법을 사용하면 이전 정보를 시스템에 저장할 수 있지만, 새로운 응답을 입력할 때에만 조사원에게 알려주기 때문에 두 접근방법 모두의 장점을 갖게 된다.

컴퓨터보조방법은 측정오차의 맥락에서 커다란 장점을 갖는다. 즉, 이것은 응답자에게 처치를 할당할 때 무선화를 도입하는 능력이다. 우리는 다음 장에서 이 장점을 다룰 것이다

6.4 오차 개요

전통적으로 오차 지표들은 특정 관점으로 제한되는 경향이 있다. 그래서 표본오차 또는 무응답의 계산이 수록된 부록을 첨부하거나, 또는 본문에서 질문지세트를 해석하거나 작성하는 난점들을 논할지도 모른다. 최근에는 조사에 대한 오*차개요*에 대한 방향으로 관심을 두는 움직임이 있다. 이것은 조사전반에 대한 다차원평가가 될 것이고 양적 및 질적인 척도를 포함할 것이다. 이탈리아 노동력조사 와 미국의 소득 및 프로그램 참여조사는 이러한 시도를 보여주는 사례이다.

7. 결론

이장의 취지는 사회조사를 더 넓은 맥락에서 고려해야 된다는 것을 강조하기 위한 것이다. 2장에서 소개된 틀작업은 사회조사자료에 대한 어떤 평가를 의미한다. 이 때의 사회조사자료란 것은 조사원, 응답자, 그리고 조사수행이라는 전통적인 요소들뿐 아니라 *대표성, 무선화 그리고 현실성*이라는 세 가지 차원에 대한 관련성에서 조사의 위치를 고려해야 한다. 만약 3차원 공간의 원점을 이 세 가지 차원들을 최적화시킨 지점이라고 한다면 한 차원에서 원점을 향해 움직이는 것은 다른 차원에서 원점으로부터 멀어지는 움직임이 되기 쉽다는 것이다. 이것을 보여주는 세 가지 사례가 있다.

1. 시장조사응답자 중 소비자패널에게 어떤 일정 기간동안 또는 수주간에 걸쳐서 그들이 행했던 특정 종류의 거래에 대해 빠짐없이 적도록 요구한다. 충분한 수의 응답자를 얻기 위해 훨씬 더 많은 수의 응답자를 접할 필요가 있다, 아마도 한 명의 응답자를 성공하려면 열 명은 접해야 할 것이다. 조사수행에 기꺼이 응하겠다는 응답자의 특성은 아마도 연구하는 변수의 특성을 가진 모집단의 나머지 사람들과는 다를 것이다. 5장에서 소개된 이와 관련된 1936년 White의 연구를 기억해보자 이 경우 *대표성* 차원은 조사수행을 가능하게 하기 위해 희생되는 사례가 될 것이다.

2. 미국의 노동력조사(CPS; Current Population Survey)의 재설계는 미국 노동통계국 과 미국 인구조사국에 의해 합동으로 수행된 중요한 방법론적 실습이었다. 다수의 현장 및 실험실 실험의 재설계 과정은 새로운 절차에 대한 다양한 요소들을 테스트하기 위해 수행되었다. 최종버전이 공식적으로 제자리를 잡았을 때 시계열 자료를 함께 비교할 수 있도록 일련의 시간동안 신버전의 CPS는 구버전의 CPS와 병행되었다. 이렇게 병행해서 수행한 결과 신버전에서 보고된 실업률이 0.5% 더 높은 수치를 보였다. 공식적으로 교체할 때, 이후에도 두 개의 시계열자료를 비교하기 위해 구버전은 계속 병행되었다. 놀랍게도, 교체된 후 구버전의 CPS에서 더 높은 실업률 수치가 나왔다.

신버전의 실시 그리고 두 개의 버전을 병행해서 실시하기로 한 결정은 방법론의 실습에서 좋은 예가 된다. 이러한 놀라운 결과를 가능하게 하는 설명은 사실 이러한 절차는 최종 실험에서 *현실성*의 문제를 고려하기보다는 *대표성과 통제(무선화)*에 중심을 두었다는 것이다. 이것은 *실제의* CPS와 두 버전을 병행한 관계자들에게는 잘 알려져 있다. 조사의 절차는 다를 수밖에 없었고 조사에 대한 인식도 달랐다. 이런 주제에 접근하는 대안적인 전략은 어떤 주(states)에서는 구버전의 CPS를 하고 있는 동안 대부분의 주에서는 신버전의 CPS를 하는 것이다., 이것은 실제 정책관련조사에 대한 처치에서 측정되지 않는 요인들이 함축하는 것에 대한 중요성을 밝혀줄 것이다.

3. 세 번째 예는 어떤 의미에서는 반증이라고 할 수 있다. 사회조사에서 방법론적 연구의 가장 큰 어려움은 실험연구가 현장조사에서는 어렵다는 것이다. 만약에 조사원들이 조사일정에서 다양한 선택사항들을 적용시킬 수 있다면 조사의 방대함이 하나의 문제가 된다. 현장에서 무선적 처치(또는 조사표의 버전)를 하기 위한 방법을 고안해 내야 하는 것도 또 다른 어려움이다. 그들의 지침을 수행하기 위해 조사원을 교육시키는 것도 세 번째의 문제가 된다. 실험이 아닌 것처럼 실험에 수행되는 응답자를 속이는 것도 문제이다.

자료수집에서 컴퓨터보조방법은 조사수행에 대한 최소의 방해로 조사현장에서의 실험실시를 가능하게 해준다. 시스템은 조사원이나 응답자의 직접적인 개입없이 무선적으로 수행하고, 가장 적절한 일정을 찾아서 할 수 있게 해주며, 결과를 기록하는 것을 가능하게 해준다. *현실적*이고 *대표성있*는 조사에서 *통제된(무선화된)* 실험을 가능하게 해준다.

Bowley는 1915년 측정에 관한 그의 저서에서 이 주제들에 관한 적합하고 폭넓은 견해를 피력했다.

" 주 작업은 … 정확하게 무엇이 측정에 결정적인 것인가를 밝혀내는 것이다, 그리고 측정에서 그 노력을 최소화시키는 가장 완벽한 도구를 개발하는 것이다. … 결론적으로, 우리는 측정이란 것은 목표를 달성하기 위한 도구라는 것을 깨달아야 한다, 단지 자신을 위한 수치에 만족하는 것은 어리석은 일이다. 한편으로 측정은 복잡한 현상에 대한 시각화가 가능하도록 정확하고 이해하기 쉬운 기술로 결론지어야 다. 다른 한편으로, 실제적인 개혁자가 필요하다, 그는 문제의 중대성을 알고 척도에 관한 계획을 세울 것이다."

측정의 정의에 무엇을 측정할 것인가에 대한 실체적인 결정과 어떻게 측정할 것인가에 대한 기술적인 문제 모두를 포함시키는 것은 자신의 학문분야가 가지는 기술들에 매료되어 측정절차에 자신의 관점을 지나치게 제한시키는 것이 얼마나 위험한 것인지를 상기시키는 데 유용할 것이다. 우리는 이러한 한계를 벗어나서 다른 학문분야들도 뭔가 유용한 것을 제공할 수 있다는 것을 인정할 필요가 있다. 조사방법론자들은 외적인 시각(externalist)을 더 가져야 하고 내적인 시각(internalist)을 덜 가져야 한다(Converse, 1986). 그리고 자신의 기준 뿐 아니라 다른 분야의 기준도 존중해야 한다. 이 책의 나머지는 이것을 행하는 이상적인 기회를 제공해 줄 것이다. 자신의 관심분야와 비슷한 특정 부분만 읽지 말고 전문적인 관심과 직접적인 관련이 없는 부분도 선택하기를 바란다. 조사연구가 점점 전문화되어가고 있다는 것은 조사의 발전에 가장 적극적으로 몰두하고 있는 연구자들로부터 조사의 존재이유가 망각되는 것이라고 할 수 있다.

제 1 장

설 문 지 설 계

제 1 장: 설문지 설계

제 1 절 조사표설계 : 개념에서 응답에 이르는 험난한 길

Nobert Schwarz
University of Michigan, Ann Arbor

1.1 머리말

일반 교재들에 따르면, 연구는 연구자들-또는 그들의 고객이나 자금을 지원하는 기관들-이 흥미롭게 생각하는 이론적인 혹은 응용의 질문에 대한 대답을 얻기 위해 실시된다. 더 나아가, 조사설계의 구체적인 사항들은 이론적인 혹은 응용의 목표들을 위해 조정된다. 조사환경에 따라, 이러한 가정들은 "자명한 이치"(어느 한 학자의 표현에 따르면) 또는 "학계 이론가들의 거만한 망상"(또 다른 학자의 표현에 따르면)을 반영한다. 이 절에서는 그리고 "질문지 설계"에 관한 본 절의 기고 논문은 개념적인 주제에서 조사 응답자에 의해 제공된 응답까지 길고도 험한 길의 다양한 측면을 다룬다. 이 절의 첫 대목은 조사설계의 열쇠가 되는 요소들을 검토하는데, 이에 관한 좀 더 자세한 사항들은 다음 절에서 다루었다.

1.2 조사설계의 구성요소들

Schuman & Kalton(1985)는 조사의 주요 요소들에 대해 설명하였다. 연구자들은 일단 조사목적을 설정한 후, 목적에 맞는 모집단을 정하고 적절한 표본을 추출한다. 조사목적은 조사하고자 하는 개념을 정의하는 것인데, 이 것은 적절한 질문으로 표현되야 한다. Schuman & Kalton(1985, p.40)은 "조사할 문제에 주안점을 두고 표본설계와 질문구성을 병행하여 진행하는 것이 가장 이상적이다. 이러한 단계들이 잘 통합되지 않으면(일반적으로 실패하기가 쉽다), 표본의 일부와 잘 맞지 않는 질문을 하게 되거나 주요 분석에 대한 사례를 거의 제공하지 못하는 표본을 얻게 된다"고 지적했다. 표본이나 특정 질문은 모두 연구자의 일차적 관심의 대상은 아니다. 그보다 "연구자들은 하나(표본, 질문)를 사용해서 다른 것(모집단, 개념)을 추정하는데, 후자에 좀 더 관심을 갖는다." 모집단을 표본하고 개념에 대해 조작적 정의를 내리는 것은 모두 관찰대상자(observed)에서 비관찰대상자(unobserved)로 우리를 이끄는 것이다(Schuman & Kalton, 1985, pp. 640-641)

사용 가능한 예산 뿐 아니라 표본과 질문의 특성에 따라 실시할 조사방식을

결정하고 이에 따라 질문 문구를 조정할 필요가 있을지도 모른다. 사전조사를 거치면서, 조사는 관련 자료를 수집하기 위해 실시된다. 이 단계에서, 자계식조사가 아닌 경우를 제외하고 연구자에 의해 만들어진 질문들은 조사원에 의해 적절하게 전달이 되야 한다. 질문에 대답하기 위해서 응답자는 그것의 의미를 이해해야하는데, 연구자가 의도한 의미와 일치할 수도 있고 아닐 수도 있다. 다음으로 응답자는 기억하고 있는 관련정보를 회상해서 판단을 내리고, 이것을 주어진 응답선택문항에 맞게 적절한 형식으로 구성한다. 그리고 나서 응답자들은 조사원에게 그것을 전달하기 전에 사회적 바람직성과 자기체면유지에 관련되는 질문에 대해서는 그들의 대답을 편집할 것이다. 마지막으로 조사원들은 응답자의 대답을 이해하고, 나중에 처리할 수 있도록 그 대답을 기록한다.

그 조사의 사후단계에서, 자료처리 전에 조사원의 기록(ptotocol)은 편집되고 부호화된다. 마지막으로 자료분석, 해석 그리고 최종결과를 발표하면서 조사과정은 마무리된다.

1.3 조사목적, 개념 그리고 조작화

2장에서 Hox가 지적한 바와 같이, 조사방법론자들은 표본과 조사표구성에 많은 관심을 보이고 있다. 그와 반대로 조사표 구성의 이전 단계들, 즉 명확한 조사목적의 설정과 구체적인 질문으로 표현되는 이론적 개념의 정립에는 상당히 적은 노력을 들인다. 사실 조사방법론에 관한 대부분의 교재에는 이런 주제들이 없고, 방법은 조사목적에 맞게 조정되어야 한다는 일반적인 조언만 제공할 뿐이다.

조사방법에 관한 자문을 하는 많은 사람들이 *밝히듯이*, 많은 연구에 있어서 조사의 목적은 놀랍게도 잘못 정의되어 있다. 질문을 통해 정확히 무엇을 측정하고자 하는 것인지 연구자에게 물어보면 그들은 애매모호한 대답을 하는데 이것은 물론 동일 프로젝트에 참여한 여러 연구자들이 제각기 다른 대답을 할 정도는 아니지만. 이러한 문제는 다양한 그룹의 연구자들, 이질적인 감독자들, 그리고 종종 자료를 사용하는 외부의 이질적인 고객들이 참여하는 대규모조사에서 더 복잡해진다(cf. Davis 등, 1994). 이러한 경우에, 최종 합의 하에 만들어진 질문지 세트는 개념이 잘못 정의되었을지 모르는 위험을 안고있는 복잡한 협상 결과라는 것이다. 확실히, 모호하게 정의된 개념은 다양한 그룹의 연구자들이 다양한 연구목적을 갖게 한다. 그것은 이 개념이 타협에 의해 수정될 가능성이 높고 달성해야 할 구체적인 목표 중 특정 목표에 대해서는 엉성하게 구성될 수도 있다. 게다가 연구자들은 종종 시계열로 이루어지는 계속적인 관심사에서 이미 물어본 질문을 바꾸길 주저한다. 이미 물어본 질문이 현재 조사의 목적에 있어 결점이 될지라도 말이다. 질문프로그램들을 결정하는 것에 관한 협의들에 관한 체계적인 분석은

과학적 사회학에 대한 흥미로운 주제가 될 것이다. 그리고 대규모의 사회과학연구기관에 대해서는 잠재적으로 중요한 함축성을 가지게 될 것이다.

조사목표에 대한 협의를 하면서, 연구자들은 조사표로 표현되는 적절한 개념을 구성하게 된다. 방법론에 관한 교과서에서는 이러한 개념의 개발에 대한 주제를 다루는 경우가 드물다. 이론적 개념의 개발은 대개 검증의 맥락에서보다는 발견의 맥락에서 이루어진다는 것이다(Popper, 1968). Hox(2장에서)는 "개념은 진위여부를 밝혀내는 것이 아니라 조사과정에서 만족스러운 결과를 얻기 위한 기본이라고 할 수 있다"고 강조했다. 어쨌든 이론적 발견에 대한 창조적 활동(구체적인 규칙은 지나치게 제한적일 수 있지만 이에 대한 조언은 얻을 수 있음, Root-Bernstein)과 이론적 개념의 논리적 발달을 구분 짓는 것은 유용하다.

후자는 개념에 대한 법칙망의 정교함, 그 의미의 하위영역에 대한 정의, 그리고 적절한 경험적 지표의 검증에 관련되어 있다. 조사의 경우, 이러한 경험적 지표들은 구체적인 질문으로 표현되거나, 물어보게 될 질문지 세트를 구성하는 것이다. Hox는 개념개발을 위한 다양한 전략들을 검토하고 과학철학에서의 이론적 근거들을 강조했다.

1.4 질문작성과 조사표설계

원하는 경험적 지표들에 합의하면서, 연구자들은 숨은 정보를 찾아내는 질문을 만드는 작업에 착수한다. 조사방법론의 역사에서, 질문작성은 전형적으로 경험에 의해 얻어진 "기술(art)"이라고 생각되어왔다. 단지 최근에야 연구자들이 조사응답 하에서 인지적 및 의사소통의 과정을 탐구하기 시작했다. 언어이해, 기억 그리고 판단에 대한 심리학적 이론을 도출하면서, 심리학자들과 조사방법론자들은 질문응답과정 모델을 만들고 이러한 모델을 엄격하게 통제된 실험실 실험과 분할투표방식의 조사를 통해서 검증했다. 몇 몇 편집한 책들(Jabine 등, 1984; Jobe & Loftus, 1991; Hipper 등, 1987; Schwarz & Sudman, 1992, 1994, 1996; Tanur, 1992)과 포괄적인 학술연구보고서(Sudman 등 1996)는 1980년 초까지 급격한 진보를 요약했다.

이러한 발전양상은 또한 유럽뿐만 아니라 미국의 주요 조사센터와 정부기관들에서의 급속한 인지실험실 제도화를 반영하는 것이다(Dippo & Norwood, 1992). 이러한 실험실은 조사응답과정을 연구하고 개별질문들이 가진 문제를 파악하기 위해 광범위하고 다양한 기법들을 도출해냈다. 이러한 방법의 응용을 통해, 광범위한 현장의 사전조사보다는 더 알차고 비용이 덜 드는 절차를 포함해서 조사 질

문지의 사전조사에 중대한 변화가 일어났다. 이러한 접근방법의 유용한 결과는 Johnson 등(4장)이 다양한 문화에 혼재되어 있는 표본에서 사용된 질문응답전략을 탐구하기 위해 생각했건 것을 말해버리는(think-aloud) 절차를 사용한 방법으로 알 수 있다.

Johnson 등에 더해서, 이 책의 다른 몇 개의 절에서도 인지적 이론을 보여주는 질문작성과 질문지설계에 대한 주제를 다룬다. 아래에서는 이 절의 논의에 도움이 되도록 질문응답과정의 주요소들을 검토한다.

1.4.1 질문지 제시 및 응답 : 인지적과정과 의사소통과정

설문조사에 대해 응답하려면 응답자는 여러 작업을 수행해야 한다(Cannell 등, 1981; Strack & Martin, 1987; Tourangeau, 1984, 동일한 기본가정에 밀접하게 연관된 견해). 응답자의 첫 번째 과제는 물론, 그것이 무엇을 의미하는지 이해하기 위해 질문을 해석하는 것이다. 만약에 질문이 의견을 묻는 것이라면, 응답자는 이미 형성된 의견을 기억해 내거나 그 순간 의견을 "계산"할 수도 있다. 그렇게 하기 위해서 응답자는 그들이 평가해야 하는 대상에 대한 정신적 표상(mental representation)을 형성하기 위해 기억하고 있는 관련정보를 이끌어내야 한다. 대부분의 경우, 응답자는 대상을 평가하기 위한 몇 가지 기분도 기억해 내거나 구성할 필요가 있다.

만약 질문이 행동에 관련된 것이라면, 응답자는 기억에서 이러한 행동에 대한 관련 사례들을 회상하거나 재구성할 필요가 있다. 만약에 질문이 대상기간("지난 주"또는 "지난 달")을 명시하는 경우라면, 그들은 이러한 사례들이 대상기간에 발생한 것인지, 아닌지의 여부도 결정해야 한다. 이와 비슷하게 질문이 그들의 "일상적인"행동에 대해 언급된 것이라면, 응답자는 이러한 사례들이 그들의 일상적인 행동을 대표하는지, 일상을 벗어난 것인지를 회상하거나 재구성해야 한다. 만약 응답자가 행동에 대한 구체적인 사례들을 회상하지 못하거나 재구성하지 못한다면, 또는 이런 노력을 할 정도로 충분한 동기가 없다면, 그들은 응답수행에 관련된 추정을 계산하기 위한 근거로 자신들이 가진 일반적인 지식이나 더 잘 알려진 정보에 의존할지도 모른다.

응답자의 마음에 "개인적인"판단이 형성되면, 응답자는 이것을 연구자에 전달해야 한다. 이렇게 하기 위해서, 응답자는 질문에서 제공되는 응답선택문항에 맞게 이 판단을 형성할 필요가 있다. 게다가 응답자는 자신의 판단을 전달하기 전에 사회적 바람직성이나 상황적합성(situation adequacy)에 영향을 받을 지도 모르기 때문에 그들의 응답을 편집하고 싶어할 것이다.

따라서, 질문을 해석하고, 의견을 형성하거나 관련행동을 표상 짓고, 응답형식을 정하고, 대답을 편집하는 것은 주로 과정의 심리학적 요인들이다. 이것은 응답자가 설문조사에 응하는 것으로 시작해서 그들의 공개적 보고로 끝난다.

1.4.1.1 질문의 이해

질문의 이해 단계에서 주요 문제는 응답자가 질문을 이해하는 점이 연구가 의도한 것과 일치하는 지의 여부이다. 심리학적 관점에서, 질문이해력은 두 개의 서로 얽힌 과정을 나타낸다(Clark & Schober, 1992; Strack & Schwarz, 1992; Sudman 등, 1996, 3장). 첫 번째는 씌여진 말에 대한 의미를 이해하는 것이다. 문장에 대한 *문자적 의미*를 이해한다는 것은 단어를 식별하고, 의미적기억(semantic memory)에서 어휘정보를 회상하는 것, 그리고 씌여진 말의 의미를 문맥에 따라 구성하는 것을 포함한다. 조사 관련 교재들이 연구자에게 간단한 질문으로 쓰되 생소하거나 애매모호한 문장을 피하라는 것을 강조하는 것은 당연하다(cf. Sudman & Bradman, 1983).

Johnson 등(4장)이 지적하는 바와 같이, 의미이해의 문제는 표본이 문화적으로 혼재되어 있는 경우 매우 복잡해진다. 다양한 응답자들은 같은 문장일지라도 다양한 이해를 할 수 있기 때문에, 같은 의미로 전달될 수 있는 다양한 질문을 하는 것이 필요하다. 이런 점에서 응답자는 그들이 이해하는 대로 대답한다는 것을 명심해야 한다. 만약 그렇다면 표준화에 대한 노력은 묻는 질문의 표면적인 특성을 표준화하는 것보다는 전달되는 의미의 표준화에 초점을 맞춰야 한다. 이것은 특히 질문이 다양한 언어로 번역될 때, 질문어구의 표면적인 특성을 유지하려고 하면 전달되는 의미가 심하게 왜곡될 수 있다(Alwin 등, 1994, 다국적/다언어 조사에서 측정 주제에 관한 논의). 이러한 주제들은 미래의 인지적 조사를 위해 해결해야할 과제이며, 표준화된 질문어구(standardized question wording)에 대해 다시 고찰할 필요가 있다.

그러나, 표준화된 질문어구는 Heath & Martin(3장)에 의해 도입된 표준화된 측정(standardized measures)의 주제와는 다르다. Heath & Martin은 왜 사회과학자들은 이론적 구성의 열쇠가 되는 표준화된 측정을 사용하지 않는지 의문을 가졌으며 여러 연구들간의 비교가 되는 표준화된 측정은 유용하다고 강조하였다. 표준화된 측정의 사용에 대한 결정적인 요구사항은 주어진 척도로 이루어진 질문은 모든(또는 최소한 대부분) 응답자에게 같은 의미로 전달되어야 한다는 것이다. 다양한 응답자에게 같은 의미를 전달하는 (다소)다양한 질문을 할 필요가 있다고 해도, 특히 척도가 국가간 또는 문화간 연구에서 사용될 때 척도의 의미를 표준화하는 것은 구체적인 문구를 사용하는 변수가 될지도 모른다.

더 복잡한 것은, 단어를 이해하는 것만으로는 질문에 응답을 하기에 충분하지 않다는 것인데, 오히려 조사표 설계에 관한 책 읽기에서 짐작할지도 모르겠다. 예를 들어, 만약에 응답자에게 "오늘 무슨 일을 하셨습니까?"라고 묻는다면 그들은 단어의 의미를 쉽게 이해할 수도 있다. 단어들은 연구자가 관심을 가지고 있는 어떤 종류의 활동을 측정해야 한다. 예를 들어 응답자들이 샤워를 했는지 안 했는지의 여부를 보고해야 하나? 적절한 대답을 할 수 있는 방법으로 질문을 이해하는 것은 질문의 *문자적* 의미를 이해하는 것뿐 아니라 질문의 *실제적인 의미*를 측정하기 위해 질문자의 의도를 추측하는 것도 포함된다(Clark & Schober, 1992; Strack & Schwarz, 1992; Sudman 등, 1996, 3장에서 좀 더 자세한 논의). 질문자의 의도를 파악하기 위해, 응답자들은 관련된 질문의 내용과 그들에게 제공되는 응답선택문항을 고려해야한다. Wänke & Schwarz(5장)는 실제적인 과정들이 질문이해에 미치는 영향과 좀 더 자세하게 맥락효과의 발생을 다루고 있다. 그리고 Schwarz & Hippler(1991)는 질문이해에서 응답선택문항들의 역할을 검토했다.

1.4.1.2 회상 또는 판단 계산하기

연구자가 관심을 갖는 것이 무엇인지 판단한 후, 응답자는 관련정보를 기억에서 회상해낼 필요가 있다. 어떤 경우, 응답자는 이미 형성되어 있는 자신의 태도나 행동에 관련된 판단에 직접 접근해서 응답을 할지도 모른다. 그러나 대부분의 경우, 기억에 저장되어 있는 적절한 응답을 찾지 못해서 그 순간에는 판단을 계산해낼지도 모른다. 이런 판단은 해당시점에서 가장 접근가능성이 큰 정보에 근거를 두게 되며, 그런 정보들은 앞선 질문들에 관련된 것을 사용하는 경우가 많다(Bodenhausen & Wyer, 1987; Higgins, 1989, 정보접근성에 대한 논의).

앞선 질문들이 관련 가능한 정보의 접근성에 미치는 영향은 태도측정에서 맥락효과를 가져온다(Sudman 등 1996, 4장과 5장, Tourangeau& Rasinski, 1988ckarah, 참고). 이런 과정의 구체적인 결과물은 *무엇이* 마음속에 떠오르는가뿐 아니라 어떻게 이런 정보가 *사용되는가*도 포함된다. Wänke & Schwarz(5장)는 정보의 접근성에 대한 이론적 주제를 논하고 맥락효과의 출현과 완충기억장치 문항(buffer items)에 초점을 두면서 좀더 구체적으로 태도측정을 사용한다.

만약에 질문이 행동에 관련된 것이라면, 얻어진 응답은 응답자가 사용했던 회상과 추정의 전략에 좌우된다. Bradburn 등, 1987; Pearson 등, 1992; Schwarz, 1990; Sudman 등, 1996, 7장에서 9장까지 개관, 최근의 연구는 Schwarz & Sudman, 1994). Johnson 등(4장)은 이러한 전략에 대해 세세하게 논했다. 그들이 발견한 점은 이러한 전략들의 사용이 응답자의 문화적 배경에 따라 다르다는 것

이다. 다양한 문화의 집단이 보고하는 행동에 대한 응답의 차이점은 실제 행동에 있어서의 차이점이기보다는 그들이 선택한 회상전략에서 있어서의 차이점을 나타내는 것이다.

1.4.1.4 응답형식결정

응답자는 일단 판단을 내린 후, 일반적으로 그들 자신의 단어로 보고할 수 없다. 그보다, 응답자는 연구자가 제공하는 응답선택문항 중 하나를 지정하는 방식으로 이를 보고한다. 그래서 응답자들은 주어진 선택에서 벗어나지 않도록 그들의 응답에 대한 형식을 정할 필요가 있다. 따라서 연구자들이 응답선택문항을 선택하는 것은 조사의 결과에 상당한 영향을 줄 수 있다(Schwarz & Hippler, 1991). 응답선택문항의 영향은 응답형식결정단계에서만 국한되지 않는다. 그보다 응답선택문항들은 질문이해, 회상전략 그리고 공개적인 응답으로 편집하는 것을 포함하는 일련의 응답과정의 모든 단계에 영향을 준다.

응답형식결정단계에서 확실히 드러나는 한 가지 효과는 평정척도의 양극점을 정하는 것과 관련이 있다(Ostrom & Upshaw, 1968; Parducci, 1983). Krosnick & Fabrigar(6장)은 평정척도를 사용하는 것에 관련되는 이것과 많은 다른 주제들을 다룬다. 그들은 평정척도의 설계와 사용에 대한 광범위한 문헌들에 대해 포괄적인 메타분석의 방법으로 개관하고 응용할 수 있는 중요성에 대한 많은 해답을 제공한다.

1.4.1.4 응답편집

마지막으로 응답자는 그들의 의사를 전하기 전에 사회적 바람직성과 자기체면유지를 고려하여 응답을 편집하려할 것이다. 이러한 것을 고려하는 효과는 전화면접보다 면대면조사에서 더 두드러지고 자계식(self-administered) 질문지에서는 많이 감소한다. DeMaio(1984)는 사회적 바람직성에 대한 조사문헌을 개관하고, Hox 등(1991) 뿐 아니라 van der Zouwen 등(1991)은 이러한 관점에서 조사원의 역할을 탐구하였다. 이러한 주제들은 본 책의 2장에서 더 자세히 다룬다

1.4.2 조사표 사전조사: 최근 발전 현황

인지적 조사의 결과로써 빠르게 향상된 조사실시의 한 분야는 조사표의 사전조사이다. 인지적 방법-원래는 응답자의 사고과정에 대한 통찰을 얻기 위해 사용했던-은 전통적인 현장 사전조사를 보충하기 위해 사용하고 있다(2장에서

Sudman 등, 1996, Schwarz & Sudman, 1996의 기고논문). 이러한 방법들은 현장 사전조사에서 놓치기 쉬운 많은 문제들을 밝혀내기 위한 잠재력을 가지고 있다. 질문이해의 문제에서는 특히 그렇다, 이 문제들은 응답자가 명확한 설명을 요구하거나 아주 무의미한 대답을 줄 때 현장에서는 발견하기만 할 뿐이다. 이와 반대로, 사전조사의 응답자들에게 질문을 다른 말로 표현하거나 이해의 문제에 대한 통찰을 제공하도록 대답하는 동안에 생각하는 것을 큰소리로 말하도록 요구하는 것은 응답자의 명확한 요구나 무의미한 응답에 대한 결과를 초래하지 않을 것이다. 게다가, 이러한 절차들은 사전조사에 필요한 응답자의 수를 줄여주어 비용을 절감할 수 있다, 그러나 아 주장은 심리학자들이 잠재적으로 잘못 판단하는 가정을 반영할지도 모른다. 그 가정은 인지적 과정은 사회인구학적 특성을 가진 변수를 반영할 가능성이 거의 없다는 것이다(Groves, 1996, 비평). 현재, 주요 조사기관은 다양한 절차를 돌아가며 사용하고 있다.

가장 많이 사용되는 방법은 생각하는 것을 동시적 또는 회고적 소리내어 말하기(think-aloud) 절차에서 *언어 기록(protocol)*을 모으는 것이다(Groves 등, 1992; Willis 등, 1991). 동시적 소리내어 말하기(concurrent think-aloud)는 응답자가 그 질문에 대답하는 것처럼 그들의 생각을 명료하게 발음하는 것이고, 그 반면에 회고적 소리내어 말하기(retrospective think-aloud))는 응답자들이 응답을 제공하고 나서 다시 어떻게 그 대답에 이르게 되었는지 과정을 묘사하는 것이다. 후자의 절차는 응답자가 덜 부담스러워 하지만, 응답자가 어떻게 대답에 이르게 되었는지에 대해서 응답자의 주관적 이론에 근거를 둔 자료를 얻게 되는 굉장히 큰 위험을 가지고 있다. 반대로 동시적 소리내어 말하기는 실제상황에서 얘기하는 것처럼 응답자의 실제의 사고과정을 반영하기가 쉽다(Ericsson & Simon, 1980, 1984; Nisbett & Wilson, 1977). 이와 같이 좀 더 정교한 기법들에 관련해서, 응답자는 종종 질문을 다른 말로 표현하도록 요청받는데, 이렇게 해서 질문의 의미에 대한 그들의 해석에 관하여 통찰을 얻을 수 있다. Johnson 등(4장)은 이러한 절차들의 결과를 기술했다.

소리내어 말하기 절차의 사용이 실험실 세팅에 지나치게 제한되어 최근의 발전을 이루고 있지만, 응답자에게 그들이 응답을 한 후에 다시 그들의 사고과정에 대해 보고하도록 요구하는 것은 조사연구에서 오랜 전통을 가지고 있으며 여전히 주목할만하다. 예를 들어 Cantrill(1944)은 응답자에게 그들이 응답하고 난 후에 응답자가 실제 생각하는 것이 무엇인지 요청했고 Nuckols(1953)는 응답자의 질문이해를 체크하기 위해 응답자에게 질문내용을 그들의 언어로 다시 표현하도록 했다. 이런 연구를 발전시켜 Schuman(1966)은 폐쇄형 질문지를 제시한 후 "무작위 탐침(random probe)(무작위 하위표본에 실시)"를 해서 응답자에게 그들의 대답을

자세하게 설명하도록 요구했다. Schuman의 무작위 탐침이 응답자의 사고과정을 해석하거나 탐험하려는 질문은 아니었지만, 이러한 주제에 접근해서 탐침은 실험실 세팅에서 사용되었던 소규모 응답자집단과는 달리 질문지이해에 관련된 좀 더 대규모 조사를 위한 유용한 방법이다.

*행동코드부여(behavior coding)*는 실험실 세팅보다 조사실행에서 자주 적용되었던 절차이다(Cannell 등, 1981; Fowler & Cannell. 1996). 원래는 조사원의 행동을 평가하기 위해 개발되었지만, 이 절차는 조사표에 대한 문제들을 밝혀내는데 효율적이다. 이것은 응답자의 명료한 설명요구, 응답자의 끼어들기, 응답자의 적절한 응답과 같은 면접내용을 테이프에 녹음하고 행동에 코드부여를 하는 것이다. 그러나 보통의 현장 사전조사처럼 이 방법은 질문의 의미가 무엇인지 잘 못 이해하는 것처럼 언어기록에서 드러난 문제를 밝혀내는 데에 실패할 지도 모른다. 최근에 이 접근방법이 확대 적용된 예에는 조사면접대본(interview transcript)의 자동화된 코드부여가 포함되는데, 이를 위한 표준화된 코드부여체계(coding scheme)가 개발되었다(Bolton & Bronkhorst, 1996).

언어기록과 행동코드부여에서 얻은 통찰에 근거해서, Forsyth 등(1992; Lessler & Forsyth, 1996)은 좀 더 세부적인 코드부여체계를 개발했는데, 이것은 인지전문가들이 미리 질문의 문제점을 쉽게 밝혀내도록 해주었다. 현재 이 것은 인지적 접근으로부터 축적한 지식을 가장 많이 응용할 수 있는 방법의 하나이다. 분류절차의 사용(Brewer 등, 1989; Brewer & Lui, 1996) 또는 응잡잠재시간측정(Bassili, 1996)과 같은 다른 기법들은 적용 범위 면에서 더욱 제한되어 있고, 사전조사에서는 아직 사용되지 않고 있다.

1.5 자료수집의 방식: 응답자의 응답수행과 조사표 구성을 위한 함축

자료수집방식은 많은 요인에 의해 결정되지만, 비용과 표본을 고려하는 것이 가장 우선이다(Groves, 1989). 면대면 면접조사, 전화조사 또는 자계식 질문지에 의존하는 것은 조사표 구성에 중요한 의미를 갖고 있다. 다음에는 심리학적 관점에서 면대면 면접조사와 전화조사 또는 자계식 질문지 간의 차이점에 대한 몇 가지를 요약했다(Schwarz, 1991). 이 차이점에 관련되는 많은 주제들은 자료수집에 관한 2장에서 다룬다

1.5.1 인지적 변수

이러한 자료수집방식 간의 가장 명백한 차이점은 자료를 인식하는 감각기관이다. 자계식 질문지에서 문항은 자료를 읽어야 할 응답자에게 시각적으로 제시되

므로 Jenkins & Dillman(7장)이 언급한 것처럼, 시각적 지각과 읽기에 대한 조사가 질문지 설계에 밀접하게 관련되어 있다. 이와는 정반대인 전화조사에서, 문항과 반응선택문항은 응답자에게 읽어줘야 한다. 면접조사에서는 이 두 가지 방법 모두가 사용될 수 있다.

이러한 구분에 밀접하게 관련되어 있는 것은 자료가 제시되는 시간적 순서이다. 전화조사와 면접조사는 엄격한 순차적 구조를 갖고 있다. 따라서, 응답자는 조사원이 제시하는 시간적 순서와 속도에 맞춰서 정보를 처리해야 한다. 조사원은 일반적으로 문항들을 되돌아가거나 앞서갈 수 없고 어떤 특정 문항에 대해서 상대적으로 더 많거나 적은 시간을 사용할 수도 없다. 응답자가 그들의 응답을 수정하기 위해서 이전의 문항으로 되돌아갈 수 있다고 해도, 거의 그렇게 하지 않는다. 전화조사나 면접조사의 상태 하에서 이전 응답을 추적하기 위해 기억을 더듬는 것은 어려운 일이기 때문이다. 반대로, 응답자의 응답을 추적하는 것 그리고 문항들 간에 되돌아가거나 앞서가는 것은 자계식 설문조사의 상태 하에서는 어려움이 없다. 그 결과, 자계식 조사표는 질문의 순차적 구성이 미치는 영향이 적고, 연달아 제시되는 설문은 바로 *앞 질문의* 응답에 영향을 미치는 것으로 나타났다(Bishop 등, 1988; Schwarz & Hippler, 1995). 따라서, 맥락효과의 발생은 면접조사와 전화조사 상에서는 순서-의존적(order-dependent)이지만, 자계식 질문지 상에서는 그렇지 않다.

게다가, 다양한 조사방식은 응답자에게 주는 시간압박감(time pressure)에 따라서 다르다. 시간적 압박은 좀 더 광범위한 회상과정은 방해를 받게되며 간단한 처리전략을 채택할 가능성이 높아진다(Krosnick, 1991; Kruglanski, 1980). 전화조사 시에 시간압박감이 가장 큰데, 침묵의 순간을 나타내는 것이 응답자가 응답수행에 여전히 집중하고 있다는 것을 암시하는 비언어적 의사소통으로써의 제 역할을 하지 못하기 때문이다(Ball, 1968; Groves & Kahn, 1979). 만약 질문이 회상을 하거나 응답수행에 대한 판단을 하도록 하는 것이라면, 전화조사에서는 응답자가 필요로 하는 시간을 갖도록 배려하는 것이 특히 중요하다. 자계식 질문지의 경우 응답자들이 자신만의 작성속도로 응답할 수 있기 때문에 응답자들이 느끼는 시간압박감이 가장 적다. 면접조사는 일시 중단하는 경우 비언어적 의사소통으로 의미가 전달되므로 중간정도의 시간압박감을 갖는다.

1.5.2 사회적 상호작용 변수

사회적 상호작용이 모든 표준화된 조사면접에 심각하게 제한되어 있는데, 자료수집방식도 그들이 비언어적 의사소통을 제한하는 정도에 따라 다르다. 면접조

사는 참가자들이 비언어적 단서에 완벽하게 접근할 수 있는 반면, 전화조사의 참가자들은 준언어적(paraverbal)단서에 제한되어 있다. 반면 자계식 설문조사에서는 사회적 상호작용이 거의 없다. 심리학적 연구는 면접조사동안에 비언어적 단서의 다양한 기능을 발견하였다(Argyle, 1969). 가장 중요한 것은 비언어적 단서가 상대방에 대한 관심과 반응을 나타내고, 무엇이 말하는지에 대한 묘사(제스츄어의 형태로)뿐 아니라 피드백을 제공해준다. 자계식 질문조사 상태에서는 이렇게 도움이 될만한 단서들이 없기 때문에 Jenkins & Dillman(7장)이 강조한 바와 같이 조사원 개입이 없는 상태에서 응답동기를 유지시키고 질문지의 내용을 자명하게 만들기 위해 각별한 주의가 필요하다. 7장에서는 이에 관한 유용한 지침들이 제공된다.

게다가 자명한 질문지를 만들 필요가 더 절실한 이유는 응답자가 자계식 질문지 상에서 조사원으로부터의 부가적인 설명을 들을 수 있는 기회가 없다는 것이다. 이와 반대로 면접조사에서는 조사원이 응답자의 비언어적 표현을 살펴볼 수 있는 상황이기 때문에 쉽게 조사원의 도움을 받을 수 있다. 전화조사는 이들 양극단의 사이에 있는데, 조사원은 응답자의 언어적 표현만 얻을 수 있다. 조사원이 추가정보를 제공한다 해도 그 정보는 대개 규정된 피드백에 제한되어 있다, 이것은 응답자가 질문의 의미를 알 수 있도록 도와주는 것이다. 비정보적-특별하지 않은-명료성, "당신이 이해하고 있는 바대로"는 응답자가 질문을 적절히 해석하기 위해 그 문맥을 검토하는 과정을 단축시킬 수 있다. 반면에 자계식 설문조사에서는 질문의 숨은 의도를 추론할 수 있는 조사표에 명시된 문맥에 의존해야 한다. 그리고 그들은 그렇게 할 수 있는 시간과 기회가 있다. 이것은 질문지해석에 따른 질문과 선택반응문항에 관련된 효과를 증가시킨다(Strack & Schwarz, 1992).

자계식 조사표는 문맥정보에 대한 의존도를 높일 가능성이 큰데(제시되는 질문순서의 영향에는 독립적일 수 있지만)조사원 효과에서는 자유롭다. 일반적으로 조사원특성은 조사원이 보이지 않을 때보다 면대면 접촉을 할 때 응답자에게 노출되기 쉽다. 전화조사에서 조사원특성은 비언어적 단서와 말하는 스타일로부터 추론되는 특성으로 제한되어 있다(성별, 연령 또는 인종과 같은). 물론 자계식 설문조사에서는 조사원이 없기 때문에, 응답자가 표지글이나 질문지를 주고 간 사람 등에서 연구자의 특성을 알아낼 수 있다. 응답자가 지각하는 조사원의 특성이 사회적 바람직성을 나타내는 응답들에서 증가할 지라도 조사원과의 친밀감형성(Rapport)도 증가시킬 수 있다. 이처럼 응답행동에 미치는 조사원특성의 효과는 양면성을 갖는다.

1.5.3 기타 차이점들

이 외에, 자료수집방식은 비밀이라고 인식하는 정도와 예를 들어 다른 가족구성원같은 외부에서 주의를 산만하게 하는 요소들을 연구자가 통제 할 수 있는 정도에 따라 차이가 있다. 게다가, 조사방식별로 서로 다른 특성을 가진 응답자를 구별하여 차별적 자기선택(differential self-selection)을 하게 된다. 일반적으로, 낮은 교육수준의 응답자는 면접조사나 전화조사보다는 상대적으로 우편조사에서는 대표성이 낮은 것으로 추정된다(Dillman, 1978). 더 중요한 것은 우편조사의 응답자는 그들이 조사에 응하기로 결정하기 전에 질문지를 미리 볼 수 있는 기회를 갖는다는 것이다(Dillman, 1978). 이와 반대로 면접조사나 전화조사의 응답자들은 각 조사를 소개하는 도입부에서 제공되는 정보에 근거하여 이런 결정을 내려야 한다(조사가 진행되는 동안에는 그들의 결정이 바뀌기가 쉽지 않다.). 그 결과, 우편응답자들은 면접조사나 전화조사보다는 주제관련 자기선택을 할 위험성이 높다.

우편조사에서 응답률을 증가시키는데에 있어서 주제관련 자기선택이 덜 문제가 되고 있는 것은 사실이지만, 응답률이 높더라도 자기선택문제는 여전히 존재한다. 우편조사와 전화조사가 모두 80%의 응답률을 얻고 있다고 가정해보자. 전화조사에서, 20%의 무응답은 응답자가 휴가나 또는 다른 이유 등의 불편한 시간에 전화를 걸었기 때문에 참여자들이 거절을 한 경우다. 이 경우는 조사의 주제에 대하여 심사숙고하여 주제가 그들의 시간을 낭비할 만큼의 가치가 없는 것이라고 생각하는 응답자는 포함시키지 않았다. 반대로, 우편조사의 응답자는 그들이 선택하는 시간에 질문지를 작성할 수 있다, 그래서 시간문제로 생길 수 있는 무응답률은 감소될 수 있다. 다른 면으로, 그들은 조사표를 훑어보고 그들이 관심을 갖는 분야의 주제라면 조사에 기꺼이 응해줄 가능성이 높아진다. 그 결과 두 가지 방식에서 모두 무응답률이 20%로 동일하더라도 면접상황에서는 주제에 무관하나, 우편조사에서는 그렇지가 않다. 다양한 방식에서 유사한 응답률을 보이는 것은 반드시 공통점이 있는 표본을 나타내는 것은 아니다.

게다가, 우편조사에서 자기선택을 좌우하는 변수는 주제에 대한 응답자의 관심인데, 이것은 사회인구통계학적 변수들과는 관련성이 적다. 따라서 추출된 표본이 연령, 성별, 수입 등과 같은 인구통계학적 변수를 고려하여 대표성을 가진 것일지라도 주제에 따른 자기선택은 존재할 수 있다. 주제관련 자기선택의 문제는 존재하는 것으로 평가되고 있으며, 미래의 연구는 주제에 관한 응답자의 관심과 다른 조사방식에서 실질적인 응답과의 관련성에 대해서 평가할 필요가 있을 것이다.

1.6 결론

본 절에서는 조사표 설계에 관하여 응답과정의 인지적이고 의사소통적인 면에 일차적으로 초점을 두고 조사과정의 주 요소를 개관하였다. 조사과정의 다른 요소들은 이 책의 다음절에서 다룰 것이다. 앞서 몇 가지 연구들을 선정하여 검토한 바대로 응답자는 어떻게 질문에 반응을 보이며 대답하게 되는가 하는 우리의 이론적 이해는 최근간 발전해 오고 있다. 게다가 연구자들은 최근의 이론적 사항들을 조사실시에 관련하여 활용하기 시작했는데, 이 발견사항들은 조사표 구성에 대한 인지적 접근의 잠재적 유용성이다. 그럼에도 불구하고, 탐구해야 할 사항들이 아직 많이 남아있고 향후 10년 간의 인지적 연구가 표준화된 조사실무에 미치는 영향력은 낙관론자들이 희망하는 것 이상으로 여전히 그 한계가 남아있을 것이다. 어떤 측면에서 볼 때 이 점은 실망스러운 것이지만, 지침이 되는 이론적 원리들을 체계적으로 개발하고 검증하지 않으면, 조사측정에 대한 우리의 지식은 Groves & Lyberg(1988, p. 20)가 관련논의에서 지적했듯 "결과에 대한 반복실험이 계속된 실패로 인하여 흩어진 발견점으로 남아있기 쉽다."

제 2 절 이론적 개념에서 조사 질문까지

Joop J. Hox
University of Amsterdam

2.1. 머리말

조사방법론은 조사표에 포함시킬 질문작성에 대한 문제에 많은 관심을 두고 있다. 질문의 문구, 조사표 흐름, 질문의 문맥, 응답범주의 선택과 같은 문제들은 많은 연구자들의 관심대상이다. 게다가 경험이 많은 연구자들은 질문을 만들고 질문지를 구성하는 것에 대한 지침을 만들어왔고 인지적 실험방법에서는 조사질문의 품질을 검증하는 수단을 개발하고 있다.

이와 대조해볼 때, 조사의 첫 번째 질문을 종이에 기록하기 *전에* 발생하는 문제들을 파악하는 노력은 상대적으로 소홀하다. 질문이 만들어지기 전에 연구자들은 자신들이 측정하고자 하는 개념들을 결정해야 한다. 그들은 개념의 명칭을 정하고, 그 특성 및 범위를 기술하며 이것이 다른 관련 개념들과 어떻게 다른가를 기술함으로써 측정하고자 하는 것이 무엇인지를 정의해야 한다. (전부는 아니더라도)대부분의 방법론자들과 철학자들은 검증의 차원이 아닌 발견의 차원에서 사회조사를 위한 개념을 개발한다. 그러므로 연구자의 상상을 제한하는 고정된 규칙은 없고, 엄정하게 개념들의 진위를 검증하지 않고 단지 조사과정에 유용한지의 여부만으로 그 가치를 판단한다.

개념형성과정은 개념을 정교하게 구성하고 의미상 중요한 하위영역들을 정의하는 것이다. 다음 단계는 개념이나 하위영역 각각의 경험적 지표들을 찾는 것이다. 이 단계에서는 아직 구체적인 질문을 작성하거나 측정(척도)모형을 제시하지 않았으므로 이 경험적 지표들은 준변수(not-quite-variables)로 볼 수 있다. 마지막 단계는 실제로 질문을 작성하는 것인데, 이것은 각 지표의 개념적 내용을 완전하고 정확하게 조사질문으로 변환하는 것이다.

본 절의 서두에서는 과학철학의 몇 가지 주제들을 논하지만, 주요 논점은 이론적 개념으로부터 원형조사질문(prototype survey question)에 이르는 과정을 체계화하는 데 사용하는 일반적인 전략들과 구체적인 기법들을 제시하는 것이다. 이 단계를 거친 다음에야 비로소 조사표 설계에 있어서 조사하려는 모집단, 질문문구 등의 *기타여러가지* 통상적인 관심거리들이 중요한 역할을 하는데, 이 문제들은 여기에서 다루지 않는다.

본 절의 서두에서는 과학적 개념에 대한 이론적 본질을 논한다. 최근의 과학철학은 과학적 개념들이 모두 이론과 결부되어(laden with theory)있다고 결론을 내렸는데, 본 절에서는 그 이유와 후에 이것이 개념화와 조작화에 미치는 영향을 논한다. 아래 항목에서는 질문개발을 위하여 제안된 몇 가지 접근방법들을 이론중심접근방법과 자료중심접근방법으로 구분하여 논하고, 마지막 항에서는 간단한 요약과 논의를 제시한다.

2.2 과학적 이론 및 과학적 개념의 모호성

과학적 연구의 두 가지 주요 목표들은 특정분야에 관한 우리의 지식을 확대하고 실제적인 문제를 해결하는 것이다. 사회과학에서 조사연구는 두 가지 목적으로 사용되는데, 첫째, 조사가 제공하는 정보는 이론을 탐구하거나 검증하는 데 사용할 수 있고, 둘째 정책을 형성하거나 결정하는 데에도 사용할 수 있다. 두 가지 경우 모두에서 과학적 이론은 중요한 역할을 담당한다. 우리는 일반적으로 가장 실제적인 사례에도 응답 자체에는 관심을 갖지 않고, 좀 더 추상적인 수준으로 함축하는 의미들에만 관심을 갖는다. 예를 들어, 응답자에게 어느 후보에게 투표할 것인가 질문한다고 하자. 이것은 직설적인 질문이지만, 조사 목적은 도수분포를 산출하기 위한 것이 아니라, 예를 들면, 선거결과를 예측하기 위해 일반적인 대중의 "투표의도" 같은 것에 대해 뭔가를 추론하는 것이다. 이런 면에서 질문은 전혀 직설적이지 않다.

우리는 응답자들에게 어느 정도 그들이 원하는 대로 투표결과가 나타날 것이라고 확신하는가와 다른 후보 중 누구에게도 투표할 것을 고려하고 있는지에 대해 질문함으로써 "투표의도"를 더 잘 파악할 수 있다. 다른 예는 응답자의 직업을 묻는 전통적인 질문이다. 질문의 목적은 보통 그가 근무시간에 무엇을 하는지 알고자 하는 것이 아니라 응답자를 사회경제적 지위(SES; Social- Economic Status) 중 어느 한 위치에 배정하기 위한 것이다. 여기에서도 이 질문은 직설적이지만 "SES"라는 개념은 그렇지 않다. "SES"를 다르게 개념화하면 직업을 묻는 질문도 다르게 작성될 것이다.

조사의 설계나 해석을 더 추상적인 수준으로 고찰하는 이유는 과학적 개념들이 더 일반적인 이론을 나타내기 때문이고, 일반적인 이론이란 것은 우리가 연구하는 현상을 나타내는 체계적인 관점이기 때문이다. 우리는 특정 개인을 표본으로 추출하여 그들이 지지하는 후보들의 인기도를 표로 작성하는 대신에 정치적 성향을 나타내는 다른 여러 개념들과 관련지어 "투표의도"를 설명(예측)하려고 한다. 우리는 "목수"라는 값을 "직업"이라는 변수에 부여하는 대신에 "SES" 변수 중 "숙련된 노동자(skilled labor)"라고 배치한다.

*개념(concept), 구성개념(construct) 그리고 변수(variables)*라는 용어들은 종종 혼용되고 있다. 이 용어들의 차이점을 이해하기 위해 Kerlinger(1986, p.9)의 정의를 소개한다. "이론이란 개념/구성개념들, 정의들 및 명제들이다. 이들은 현상을 설명하고 예측할 목적으로 변수간의 관계를 규정함으로써 이 현상을 체계적으로 나타낸다." 따라서 이론은 개념과 구성개념의 관계에 관한 명제의 체계이다. 변수라는 것은 논란의 여지가 없이 규칙에 따라 경험적 관찰에 근거하여 값을 부여한 하나의 용어나 기호이다. 개념과 구성개념의 차이는 미미하다. 개념은 유사한 현상들이나 유사한 속성들을 일반화하여 형성한 추상적 관념(abstraction)이고, 구성개념은 과학적 이론에서 사용하기 위해 체계적으로 정의한 개념이다(Fiske, 1971; Kerlinger, 1986). 예를 들어, "범법"은 범법으로 분류된 행동을 드러내는 개인적인 성향을 일컫는 개념이고, "일탈"은 일탈로 분류된 행동들을 나타내는 개인적인 성향을 일컫는 구성개념이다. 그 차이를 보면 범법은 우리 사회가 범법으로 분류한 행동들을 지칭함으로써 정의할 수 있는 반면에, 일탈은 어떤 행동을 일탈이라고 할 것인가에 대한 ("규범" 및 "사회준거집단"과 같은 기타 구성개념들을 포함하는) 특정 이론의 틀 내에서만 정의할 수 있다는 것이다.

구성개념 및 개념은 모두 이론적인 추상적 관념이기 때문에, 필자는 이 용어들을 폭 넓게 사용하여 개념은 응용 연구와 이론화 초기 단계에서 사용하는 추상적 용어들을 지칭하고, 구성개념은 이론 연구에서 사용하는 더욱 공식적인 과학적 개념을 지칭하는 것으로 다룬다. 이러한 용어정의방식은 사회 조사에서 사용하는 개념이 일상적으로 사용하는 단어일지라도(예를 들면, 사회 계층, 장애, 실업 등), 연구자들은 자신의 연구를 위한 목적으로 응용하거나 생성한 것이다. 이 개념은 이론적인 추가적 의미(surplus meaning)를 갖고 있는데 이 의미의 경계는 불분명하다. 이론적 구성개념은 다양한 추가적 의미를 가지고 있다. 응용을 목적으로 연구를 하는 경우에는 추가적 의미가 거의 없는 개념으로 연구할 것이다. 그러나 경험적 측정치에 가까운 "성별"이나 "나이"와 같은 개념을 측정하는 경우를 제외하면, 개념들에는 *약간씩은* 추가적 의미가 있을 수 있다. 따라서, 개념 및 구성개념은 조작적 정의(operational definition)에 의해 관찰된 변수들에 반드시 연결된다. 조작적 정의에 의해 이들이 어느 변수와 연결되어 있고 어떻게 값을 주는지를 구체화시키는 것이다. 이 과정은 종종 변환과정이라고 하는데, 여기에서 이론적 구성개념은 구성개념을 대표하기에 적절하게 보이는 관찰가능한 변수로 변환된다.

2.2.1 조작주의(Operationism)

조작주의(Bridgman, 1927)라고 하는 과학적/철학적 접근방법은 모든 개념이 단 하나의 관찰 변수와 완벽하게 일치할 것을 요구한다. 조작은 하나의 변수에 하나의 값을 매기도록 하는데 이렇게 해서 관련 구성개념이 완전히 정의된다. 사회학에서 Lundberg(1939)가 Bridgman의 아이디어를 지지하였고, 심리학에서는 행동주의의 발달에 영향을 주었다(Watson, 1924). 조작주의의 주요 원칙은 구성개념을 정의한 후에 이것을 관찰하기 위한 과정이 아니라 조작적 정의가 바로 구성개념*이다*라는 것이다(Lundberg, 1939). 조작주의 접근방법의 논리적 귀결은 우리가 조작적 정의에 대해 한 측면만 바꾸더라도 전혀 새로운 이론적 구성개념을 가지게 된다는 것이다. 따라서, 수은 온도계와 알코올 온도계로 각각 열을 측정하는 경우 두 개의 다른 대상을 측정하는 것이 된다. 마찬가지로 질문 내에 있는 한 단어를 변경하게 되면 다른 구성개념을 측정하게 된다.

조작주의가 사회과학에 미친 영향은 명확한 정의 및 측정절차를 강조하는 것이다. 하지만 조작주의를 극단적으로 채택하여 결과들을 일반화하려고 시도하는 경우 해결할 수 없는 문제들이 발생할 것이다. 조작주의와 관련된 문제는 두 개의 조작적 정의를 가지고 어느 것이 더 나은가를 비교할 수 없다는 것이다; 그 둘은 정말 다르다.

2.2.2 논리실증주의

조작주의 접근방법의 여러 문제들과 모순들로 인해 이론적 구성개념과 관찰변수를 구별하는 것이 유용하다는 결론이 나오게 되었다. *논리실증주의*로 알려진 철학적 접근방법이 그 중심사상인데, 비엔나 학파라는 철학자 집단에서 발생한 것이다(cf. Hempel과 Oppenheim, 1948). 논리실증주의는 사회 과학자들이 이론적 구성개념과 관찰 변수의 관계에 대해 인식하는 방법을 형성하는 데에 많은 영향을 주었다. Carnap(1956)은 논리실증주의의 견해를 명쾌하게 설명하였다. Carnap은 *이론언어(theoretical language)와 관찰언어(observation language)*를 구별하였는데, 이론언어는 이론적 구성개념과 이들의 상호관계에 대한 가정들을 내포한다. 관찰언어는 단지 조작적 정의 혹은 형식적 논리나 수학적 조작에 의해 조작적으로 정의된 개념들과 연결되는 것만을 포함한다. 또한 관찰언어는 관계 있는 모든 과학자들이 완벽하게 이해할 수 있어야 하고, 관찰의 정당성에 대해서는 논란의 여지가 없어야 한다. 이론언어와 관찰언어의 연결은 가정과 추론의 대응규칙(correspondence rule)에 의해 이루어진다. Carnap의 견해에 따르면, 이론언어는 관찰언어보다 훨씬 더 풍부하다. 따라서 과학적 이론은 다른 구성개념들과의 관

계로만 정의되는 구성개념, 즉 조작적 정의가 없는 구성개념들을 포함할 수도 있다. 이론언어가 관찰언어보다 더 풍부하다는 견해는 이론언어가 관찰언어에 관한 추가적 의미를 갖고 있다는 말로 요약할 수 있다. 우리는 연구 과정에서 우리의 지식을 확대함으로써 이 추가적 의미를 줄이고 싶어하지만 약간의 추가적 의미는 항상 남기 마련이다.

사회학에서 논리실증주의는 Northrop(1947), Lazarsfeld(1972) 그리고 Blalock(1982)의 연구에 큰 영향을 미쳤다. Northrop은 관찰 즉시 이해할 수 있는 직관에 의한 개념들(concepts-by-intuition)과 이론에서 의미의 일부분을 최소한이라도 도출하는 가정에 의한 개념들(concepts-by-postulation)을 구별하였다. 이 두 가지 종류의 개념들은 Carnap의 대응규칙과 유사한 인식적 상관관계(epistemic correlation)로 연결된다. Blalock은 Northrop의 직관에 의한 개념들을 관찰언어의 용어로, 가정에 의한 개념들은 이론언어의 용어로 정의했다. 이 두 종류의 용어들은 "보조측정이론(auxiliary measurement theory)"으로 연결되는데, 이것은 Carnap의 대응 규칙을 구성하는 가정 그리고 추론을 기술하는 이론이라고 할 수 있다. 심리학에서 논리실증주의는 법칙적 연결망(nomological network)의 일부분을 구성하고 있는 이론적 구성개념에 대한 것과 이 이론적 구성개념을 어떻게 측정하는 지 구체적으로 기술하는 조작적 정의를 구분하는 데에 영향을 주었다. 정신측정학 이론에서의 주요 개념은 측정의 *구성 타당도(construct validity)*이다. 구성 타당도는 측정 측정이 측정하기로 되어있는 이론적 구성개념을 얼마나 잘 측정하는 지를 나타내는 것이다(Cronbach와 Meehl, 1955; Campbell과 Fiske, 1959). 인식적 상관관계와 마찬가지로 구성 타당도는 구체적인 값으로 평가될 수 있는 어떤 모수(parameter)가 아니다. 이것은 대응 규칙의 일반적인 품질을 지칭한다.

2.2.3 세련된(고급) 반증주의(Sophisticated Falsificationism)

논리실증주의자들의 견해들은 Popper, Kuhn과 Feyerabend와 같은 철학자들로부터 신랄한 비판을 받았다. Popper는 반증 원리를 도입한 것으로 가장 잘 알려진 인물일 것이다. 과학은 이론적 추측을 정립하고, 반증을 통해 이 이론들의 진위를 가릴 수 있게 함으로써 발전한다(Popper, 1934). 하지만, Popper(그리고 기타 철학자들)는 논리실증주의에 대해서 직접적이고 편견이 없는 관찰과 같은 것도 있다는 그들의 믿음을 비판하였다. Popper가 여러 가지 사례를 들어 밝힌 바와 같이(Popper, 1934; Chalmers(1976)도 참조), 관찰에 대한 진술은 항상 이론언어로 만들어진다. 정확한 관찰은 명확하고 명료하게 정립한 이론을 요구한다. 이론에 오류가 있으면 이론의 문맥에서 기술한 관찰의 진술도 반증의 대상이 된다.

관찰이 이론과 결부되어 있다면, 직접적인 반증은 불가능한데, 이는 이론과 관찰이 모두 거짓일 수 있기 때문이다. Popper는 이론뿐만 아니라 관찰의 진술도 비판과 검증을 받게 함으로써 이 문제를 해결하였다. 과학계는 비판적인 검증 후에 오류가 없는 것으로 판명된 관찰은 당분간 받아들인다(Popper, 1934). 이 견해는 대부분의 사회과학 방법론자들이 채택하였다. 예를 들면, Blalock(1968)은 Northrop을 언급하면서 직관에 의한 개념은 사회과학자들이 합의를 하여 받아들인다고 주장하였다. Cronbach(1971)는 구성개념들을 부정적인 결과들에 의해 발생 가능한 효과 중의 하나라고 열거하는데, 하나의 구성개념은 폐기되거나 분할되거나 다른 구성개념들과 결합하게 될지 모른다. De Groot(1969)는 "과학적 포럼", 즉 과학계에서의 지속적인 논의를 통해 이론적 구성개념과 조작적 정의에 관한 결정을 내려야 한다고 주장하였다.

Kuhn(1970)과 Lakatos(1978)는 모두 과학적 이론을 전개하는 방식을 고려하여 관찰진술이 이론에 종속되는 문제를 약간 다른 방식으로 다루었다. Chalmers(1976, pp. 63, 71)가 증명한 바와 같이, 과학자들이 Popper의 반증 원리를 엄격하게 따랐더라면, Newton의 만유인력의 법칙과 같은 가장 강력한 이론들 중 몇 가지는 태동할 시부터 폐기되었을 것이다. Kuhn(1970)은 과학의 연구는 거의 언제나 일반적으로 인정되는 가정, 법칙 그리고 기술을 포함해서 일반적으로 인정되는 패러다임 내의 정상적 과학의 상태에서 수행되었다고 주장하였다. 외관상의 반증 사례들(apparent falsifications)은 추후 연구와 측정기술 및 실험기술의 향상에 의해 해결될 수 있는 변칙적인 결과이다. 변칙사례들(anomalies)이 증가하면, 과학혁명이 발생하는데 여기에서 변칙사례들로 가득 찬 패러다임은 새 패러다임으로 대체된다. Lakatos(1978)는 과학을 서로 경쟁하는 연구 프로그램들의 세계로 본다. 각 연구 프로그램은 주 이론적 가정은 기각되지 않는 다는 견고한 핵(hard core)을 갖고 있다. 이론적 핵의 보호대는 보조이론이다(수용한 연구 방법에 대해 진술되지 않은 가정들을 포함하는). 반증사례에 직면할 때, 이론적 핵은 변하지 않고 남아있고 보조이론만이 조정된다. 지속적으로 흥미롭고 새로운 결과를 이끌어 내는 연구프로그램은 *진보*라하고 그렇지 않은 연구프로그램은 *퇴보*라고 한다. 한 연구프로그램이 퇴보하는 징후를 강하게 나타내면 과학자들은 이것에 대한 경쟁 프로그램을 찾는다(옛 이론은 죽지 않는다. 다만 사라질 뿐이다). Lakatos의 견해에 대한, 조사연구에서 연구프로그램의 한 예는 응답행위를 연구하는 인지적 접근방법이다. 여기에서 이론적 핵은 일단의 인지적 이론들과 정보처리과정들이다. 보호대는 이론들을 조사연구에 적용하는 데 필요한 여러 조정사항들로 구성되어 있다. 일반적으로 한 반증사례가 발견되었다고 해서 인지적 접근방법이 전부 폐기되는 것이 아니라, 보조이론이나 연구방법을 조정한다. 인지적 연구프로그램이 지속적으로 흥미로운 결과를 낳는 한, 연구가들은 이것을 고수하고, 그렇지 못하는 경우에는 다른 프로그램(있는 경우)에 관심을 갖게 된다.

Kuhn과 Lakatos의 견해에 따르면, 대부분의 연구는 정상이거나 진보적인 연구프로그램 내에서 수행한다. 따라서 이론적 구성개념과 관찰에 대한 논의는 기술적인 사항을 다루는 경향이 있다. 이론적 구성개념에 관한 비판적 검증은 아마도 이 개념의 차원성(dimensionality)(이것을 분할하거나 다른 구성개념과 결합해야 하는가?) 혹은 법칙망(다른 구성개념들과 관련되어 있는 방식)에 주안점을 둘 것이다. 새로운 이론적 구성개념들은 일반적으로 인정되는 틀에 맞춰지고, 이것들을 관찰 가능한 척도로 변환하는 지침도 있다. 우리는 이론적 위기 상태에서만 Feyerabend가 말한 무엇이든지 허용되는(anything goes) 위치에 있게 된다.

조작주의와 논리실증주의가 이론적 구성개념의 모호성과 이론에 대한 관찰의 종속성을 제거해야 할 문제들로 본 반면에 Popper, Kuhn과 Lakatos는 이 문제들을 과학적 진보의 핵심으로 간주하였다는 것은 흥미로운 것이다. 특히 일반적으로 "세련된 반증주의"(Chalmers, 1976)라고 불리는 견해를 가진 Lakatos는 개념의 이론적 의미와 발생에 대한 관찰의 비판적 논의를 계기로 새로운 이론들이나 새로운 연구문제의 형태로 과학적 진보가 일어나는 것이라고 강력하게 주장하였다.

2.3 개념화 및 조작화

전 항에서 설명한 대로, 철학자와 방법론자들은 순수한 관찰은 없다는 결론을 내렸다. 조사측정의 맥락에서 완전히 객관적인 조사질문은 없다. 대신질문은 이론에서 관찰에 이르는 연속선에 있다. 일부 질문은 관찰의 끄트머리에 가까워서 대부분의 경우 이것들을 구성개념을 완벽하게 측정하는 척도들로 생각할 수 있는데, 나이나 성별에 관한 질문이 그 예가 된다. 그러나 이것들도 이론과 강하게 결부되어 있을 수 있다. 성역할 조사(gender study)에 있어서의 "성"이나 월력상 "나이" 및 생물학적 나이에 관한 조사나 생활주기조사(life cycle research)에 있어서의 나이를 생각해 보라.

실제로 이론적 구성개념들을 정교하게 만들고 조사표 문항을 구성하는 문제에 직면할 때, 방법론자들은 이론언어와 관찰언어간에 분명한 차이가 있다고 하였는데, 이로 인해 개념화와 조작화를 명확하게 구별하게 되었다. 개념화는 개념형성으로, 정교하게 개념의 법칙망을 구성하고 그 의미에 대한 중요한 하위영역들을 정의하는 것들을 의미한다. 조작화는 개념과 그 하위영역들의 경험적 지표를 규정하여 이론적 구성개념을 관찰 가능한 변수로 변환하는 것이다. 이러한 이론과 측정의 틈을 잇기 위해 두 가지 다른 연구 방법이 소개되었다, 하나는 이론중심 또는 "하향식" 방법인데, 이것은 구성개념으로 시작하여 관찰 가능한 변수로 향하는 것이다. 다른 하나는 자료중심 또는 "상향식" 방법으로 관찰에서 시작하여 이론적 개념으로 향하는 것이다(cf. Hox와 De Jong-Gierveld, 1990).

2.3.1 개념화에서 조작화로: 이론중심의 접근방법

본 항목에서 설명하는 이론중심의 방법들은 차원/지표 분석, 어의 분석 및 다면 설계이다.

2.3.1.1 차원/지표 분석

Lazarsfeld(1958, 1972)는 일반적이면서 약간 모호한 개념으로 접근을 시작한다. 다음으로 *개념구체화(concept specification)* 과정이 이어진다; 개념은 개개의 구성요소나 차원들로 구분된다. 구분된 하위의 차원들은 이론적 구성개념의 의미를 더욱 자세하게 구체화한다. 이론적 구성개념의 하위차원들은 논리적으로 도출되거나 선행연구의 경험적 결과에 근거하여 얻을 수 있다. 다음에는 각 하위차원에 대한 일단의 경험적 지표들을 정의한다. 실용적인 연구를 위해 지표의 부분집합 하나를 선택하고 보통 이것들을 합하여 각 하위차원의 총점수를 산출한다. Lazarsfeld의 견해에 의하면 특정 하위차원의 지표들은 고도의 호환성을 갖추고 있어 지표들을 정확하게 선택하는 것이 그렇게 중요한 것은 아니다.

Lazarsfeld는 우리가 실제로 어떻게 개념을 구체화해야 하는지를 설명하지 않았다. Fiske(1971)는 Lazarsfeld의 접근방법과 유사한 절차를 설명하였다. Fiske의 견해에 따르면, 개념을 명확하게 규정하면 조작화할 대상을 명확히 제시하고 여러 구성요소들의 조작화를 허용하도록 잘 정의된 구성개념을 반드시 얻을 수 있다. Lazarsfeld처럼 Fiske도 우리가 일반적이면서 모호한 개념으로 시작한다고 가정한다. 그 첫번째 단계는 이 개념에 대한 이론적 맥락을 파악하는 것이다. 종종 개념들은 이것들의 이론적 배경을 제공하는 기존의 일반 이론에서 추출된다. Fiske(1971, p. 97)는 이 접근방법을 선호하는데, 이미 정립한 개념들과 과거의 경험적 연구를 기반으로 하여 구축되기 때문이다. 여기에서 연구자가 할 일은 이 이론적 배경뿐만 아니라 특정 연구 문제에 맞추기 위해 개념을 확장하는 방법을 명확하게 설명하는 것이다. 또한 연구자들은 자신만의 개념을 정립할 수도 있는데 이 경우 그들은 그 이론적 맥락을 정교하게 구성하고 이 개념의 유용성과 이것이 다른 개념들보다 나은 장점들을 증명하여야 한다.

Fiske의 접근방법의 두 번째 단계는 구성개념의 독특한 특성인 구성개념의 핵심을 설명하는 것이다. 이론적 핵심은 개념의 본질이고 개념이 지칭하는 모든 현상들을 다루어야 한다. 이 요구사항을 충족하는 경우, 이론적 개념은 너무 광범위해질 가능성이 있다. 이러한 광범위한 이론적 개념은 하위개념들 전체를 지칭하는 일반적 명칭으로 생각하는 것이 더 적절하다. 실제로 하위구성개념들의 이론

적 의미를 검토하고 각 의미별로 별개의 척도를 구성하는 것이 바람직할 것이다. Fiske는 핵심적 의미를 정의하는 몇 가지 방법들을 기술하였는데, 이것들은 *차별화(differentiation)*라는 표제 하에 포함시킬 수 있다. 따라서 구성개념에 포함되지 않는 것들과 그 구성개념에 관련되어 있는 다른 구성개념들 간의 차이점을 명확하게 기술하는 것이 중요하다. 또한 우리는 구성개념의 양극단을 기술하고, 이것이 척도의 하한을 정의하는 것인지 아니면 새로운 특성을 도입하는 것인지 여부를 결정할 수 있다. 예를 들면, 우리가 "삶의 만족도"를 측정하기를 원하는 경우 이 척도의 하한을 만족의 부재로 정의할지, 많은 불만의 존재로 정의할지를 결정해야 한다.

Fiske의 접근방법의 세 번째 단계는 측정도구를 구성하는 것이다. 구성과정을 시작하기 전에 여러 사항들을 구별하는 것이 도움이 된다. 그 중 하나는 척도의 일반화를 가정하는 양식(modalities)이다. 자계식(self-description), 타계식(대리 측정) 및 관찰에 일관성이 있다고 기대해도 좋은가? Guttman의 다면 설계를 사용하여(후에 더 자세히 설명할), 예를 들면, 여러 상황이나 행동유형을 구분함으로써 양식들을 체계적으로 구별할 수 있다. Fiske도 측정 척도와 측정설계에 관하여 명확한 결정들을 내릴 것을 권고한다.

Fiske의 접근방법이 Lazarsfeld의 접근방법과 공통점이 많다는 점은 분명하다. Lazarsfeld의 하위차원은 Fiske의 하위구성개념과 유사하다. Lazarsfeld는 왜 사람들이 하위차원들을 구분하기를 원하는지 분명히 밝히지 않았다. Fiske는 이것을 연구문제와 관련지었다. 응용 연구에서, 우리는 아주 구체적인 목표를 가질 수 있는데, 이것에는 한정된 의미의 구성개념이 적절하다. 기초 연구에서, 특히 초기 단계에서 우리는 더 일반적인 구성개념들을 사용할 것이다. 나중에 우리는 더 높은 정확성을 얻기 위해 하위차원들을 규정할 것이다. 이에 관한 결정은 주어진 분야의 연구 문제와 지식 상태에 따라 좌우되고, 실용적인 이유로 인해 이러한 결정을 내릴 수도 있다.

차원/지표 분석 접근방법이 복잡해 보이지만 조사연구에서 이에 관한 사례를 찾는 것은 어렵지 않다. 사회조사에서 몇몇 주요 개념들의 구체화에 관한 통찰력 있는 논의는 Burgess(1986)에서 찾아볼 수 있다. 개인적인 복지에 관한 Andrews와 Withey(1976, pp. 11-14)의 연구는 좋은 사례가 된다. 일반적으로 우리는 복지가 무엇인지 알고 있다. 몇 가지 조사 목적에 따라 응답자들에게 그들의 복지에 관한 질문을 한 가지만 해서 복지를 측정할 수도 있지만, 우리는 대개 이보다 더 정확하게 측정하고 싶어한다. Andrews와 Withey는 복지에 대해서 다양한 영역들과 다양한 기준들을 구분하는 복지의 개념모형을 논하였다.

Andrews와 Withey의 영역은 사람들의 관심사에 관한 시험조사에 근거한 것으로, 사람들의 관심사를 구분할 수 있고 포괄할 수 있는 삶의 영역을 찾는 것을 목표로 한다. 그 결과로 나온 것이 주택, 직업, 가족 생활 및 이웃과 같은 생활 영역들의 목록이다. 여기에서 기준은 가치, 표준 또는 목표인데, 개인이 이런 다양한 영역들에 대해 어떻게 생각하고 있는지를 판단한다. 더 나아가 이 기준들은 절대적 기준과 상대적 기준, 일반적 기준 및 구체적 기준으로 구분된다. 이 기준들의 예를 들면 생활 수준, 독립성, 재미와 안전이다. 그림 2.1은 구체적인 개념화이다(cf. Andrews & Withey, 1976, p. 234).

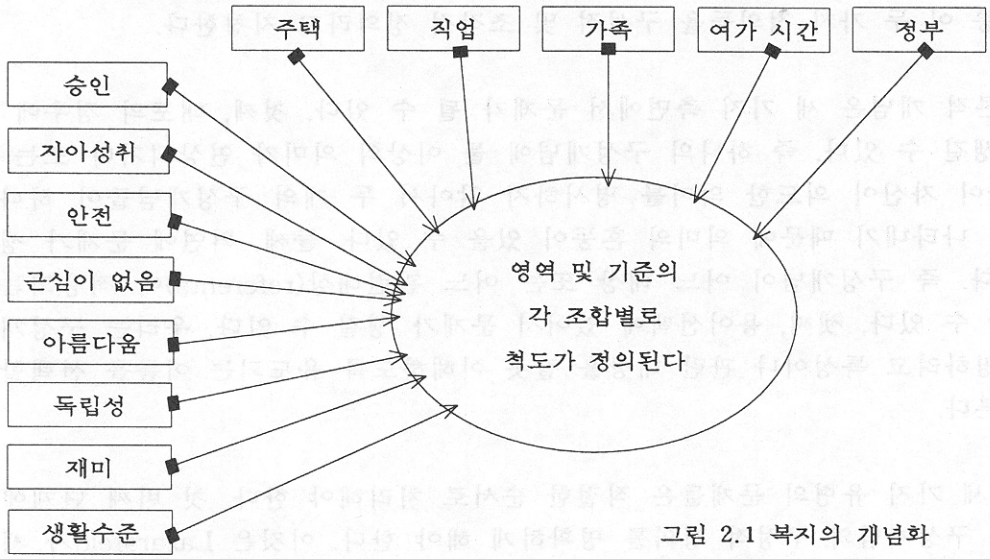


그림 2.1 복지의 개념화

최종 도구설계 또한 사용된 응답 척도 유형에 대한 고려사항들을 포함한다. 이런 구분을 감안하여, Andrews와 Withey는 개인적인 복지의 다양한 면을 알아낼 수 있는 많은 조사 질문들을 작성할 수 있었다. 그리고, 이것들을 이론적 또는 경험적으로 이론적 구성개념에 관련지을 수 있었다. 이 설계에 따라 작성한 질문의 한 가지 예는 "귀하의 주택이 가진 아름다움과 매력에 대해 어떻게 생각하고 계십니까?"라는 것이다(Andrews와 Withey, 1976, p. 13)이다.

2.3.1.2 어의 분석(semantic analysis)

Sartori(1984)는 개념화는 언어에 의해 매개되기 때문에 진술에 대한 우리의 어의적 구조를 연구하는 것은 중요하다고 지적했다. 언어이론들은 언어시스템이란 당연하게 표현하는 것이라고 한다. 언어시스템의 중요한 면은 어휘다. 이론에서는 기본 어휘를 사용하는 데 이러한 어휘는 세상에 대한 광범위한 인식을 반영하는 것이다. 이 점은 중요한데, 왜냐하면 이름이 부여되지 않은 것은 종종 인식

되지 않으며, 이름을 선택하는 것은 때로 광범위한 해석을 수반하기 때문이다. 따라서 우리의 이론적 구성개념을 둘러싸는 어의망을 측정하는 것은 중요하다.

이론적 구성개념의 의미는 두 가지 일반적 방식들로 정의할 수 있다. 첫째는 구성개념의 어의 내포(connotation)인데, Sartori(1984, p. 78)는 이것을 사용자가 마음속에 가지고 있는 연상(associations)이나 구성개념에 포함된 모든 특성들의 목록이라고 기술하고 있다. 둘째는 구성개념의 어의적 외연(denotation)인데, 이것은 예를 들면 이 개념이 지칭하는 일단의 사물의 목록을 말한다. 사회 과학자들은 보통 이 두 가지 정의들을 구성적 및 조작적 정의라고 지칭한다.

이론적 개념은 세 가지 측면에서 문제가 될 수 있다. 첫째, 내포의 경우에 문제가 생길 수 있다. 즉 하나의 구성개념에 둘 이상의 의미가 연상되거나 또는 연구자들이 자신이 의도한 의미를 명시하지 않아서 두 개의 구성개념들이 하나의 의미를 나타내기 때문에 의미의 혼동이 있을 수 있다. 둘째, 외연에 문제가 생길 수 있다. 즉 구성개념이 어느 대상 또는 어느 관련대상(referents)에 적용되는지 모호할 수 있다. 셋째, 용어선택에 있어서 문제가 생길 수 있다. 우리는 구성개념을 명명하려고 특성이나 관련 대상을 잘못 이해하도록 유도되는 이름을 선택할지도 모른다.

이 세 가지 유형의 문제들은 적절한 순서로 처리해야 한다. 첫 번째 단계에서 우리는 구성개념의 구성적 정의를 명확하게 해야 한다. 이것은 Lazarsfeld가 제안한 실체적(이론적) 분석인 차원분석과 동일한 것이 아니고, Fiske가 제안한(하위) 구성개념들의 핵심적 의미의 분석에 더 가까운데, 이것도 강력한 어의적 요소를 갖고 있다. Sartori(1984, p. 32)는 이것을 *부수(附隨)적인(변동하는)* 특성에 대립되는, 구성개념의 *정의적(본질적)* 특성을 평가하는 것이라고 말한다. 구성적 정의는 적절하고 간명해야 한다. 즉, 이것은 경험적 경계 및 관련대상을 만족스럽게 기술(외연적 정의)할 수 있을 정도로 충분한 특성들을 포함하는 면에서 적절해야 하고, 반면, 불필요한 부수적인 특성들을 전혀 포함하지 않는 면에서 간명해야 한다. Sartori가 제안한 한 방법은 기존의 내포적 정의의 특성들을(사전이나 선행연구에서) 수집하여 이들로부터 공통적인 핵심을 추출하는 것이다. 하지만, 구성적 정의가 다양하게 보이지만, 정의를 하는 특성들의 공통적 핵심은 다루기 쉬울 정도로 작을 수 있다. 중요점검사항은 동음이의어 및 동의어의 발생이다. 한 가지 이상의 의미를 지닌 구성개념은 둘로 나누어 각 의미별로 하나의 단어를 만들어야 한다. 동일한 의미를 가진 하나 이상의 구성개념은 앞의 경우보다는 더 낫지만, 모호할 수 있으므로 삼가야 한다.

다음 단계는 구성개념의 경험적 관련대상을 결정하는 것이다. 이것은 조작적 정의를 제공하는 것과는 다르다. 왜냐하면 우리가 (아직) 측정의 실시를 하지 않은 경험적 관련대상을 포함시킬지도 모르기 때문이다. 여기에서 중요한 일은 경계를 설정하는 것과 경험적 관련대상이 구성개념에 속하는지 속하지 않는지를 결정하는 것이다. 다시 말하지만, 문제가 되는 것은 이 분야에서 수행할 측정의 실시나 분류가 아니다. 이 문제는 조작적 정의를 규정하는 것으로 해결할 수 있으며, 잘못 분류한 것은 측정 오차로 다루어진다. 문제는 무엇이 정확하게 이론적 구성개념에 속하는지를 원칙적으로 정하는 것이다. 경험적 경계나 경험적 관련대상에 관한 어려운 점들은 구성개념의 구성적 정의에 있어서 결함으로 될 수 있다. 문제는 구성개념의 핵심을 충분히 정확하게 정의하지 않는다는 것이다. 이에 대한 명확한 해결책은 구성개념에 대한 정의할 수 있는 특성의 수를 늘림으로써 구성개념의 의미를 제한하는 것이다. 그러나 Lakatos(1976)가 증명한 바와 같이 *개념확장(concept stretching)*은 대개의 경우 이론적 구성개념의 내용에 관한 문제를 다루는 혁신적인 방법이지만, 구성개념을 변경함으로써 나타나는 결과들은 주의 깊게 검토해야 한다.

세 번째 단계는 구성개념에 대한 언어의 명칭을 명확하게 이해할 수 있도록 확실하게 하는 것이다. 한 가지 고려할 점은 선택한 용어의 추상도(level of abstractness)이다. 굉장히 추상적인 용어들은 대부분 뚜렷한 특징들이 거의 없고 많은 경험적 관련대상을 지칭한다. Sartori는 어의적으로 다른 구성개념과 혼동하지 않을 용어들을 선택하도록 권한다. 용어를 평가하는 편리한 방법은 대체검사인데, 이는 구성적 정의에서 한 단어를 다른 단어로 대체할 시 명료성이나 정확도가 증가하면 단어를 잘못 사용하고 있다는 것이다.

어의 분석의 한 예는 "발전"이라는 구성개념에 관한 Riggs(1984)의 분석이다. Riggs는 어휘분석으로 시작하였는데, 이 분석에서 "발전하는 것(to develop)"이란 어휘는 어원적으로 감싸다(envelop)의 반대말임을 증명하였다. 현대적 의미로는 "어떤 것이 잠재적으로 가지고 있는 모든 것을 끄집어내는 것이다"(옥스포드 영어 사전). 사회과학에서 이 의미는 더 구체적인데, 인간 시스템이 더 복잡하거나 더 성숙하거나 더 고차원적인 형태로 진화하는 것을 뜻하거나 정치학의 경우에는 주로 거시적인 차원에서 정치적 체제가 상기한 형태로 변화하는 것을 뜻한다. 정치학에서 "발전"에 관한 초기 정의들(23개의 다양한 정의들)을 분석한 한 연구는 정치학에서 "발전"이라는 개념은 서로 중복되는 의미들로 과도하게 정의되어 있어 사실상 무용하다는 결론을 내렸다. 여러 정의들에서 "발전"은 "변화", "성장", "현대화" 등과 같은 구성개념들로 변화된다. 몇 가지 경우에는 이러한 변화를 통해 원래의 의미가 더 향상된다.

어의 분석은 조사 질문과는 직접적으로 연관되어 있지 않지만, 이것은 이 구성개념들에 있는 여러 의미들을 정리하고 그 모호한 의미들을 구별하는 데에 도움이 된다. 어의 분석은 언어를 다루기 때문에 언어에 종속적이다. 이것은 문화간 조사에서 사용하기 위해 번역된 척도들의 의미를 검토하는 경우에는 장점이 될 수 있다. 예를 들어, 네덜란드어의 "행복"을 어의적으로 분석하면 결국에는 "행운"이라는 개념을 측정하게 되는데, 이것은 *geluk*이라는 네덜란드어 단어가 "행복한(happy)"과 "운이 좋은(lucky)"이라는 영어 단어의 의미들을 결합한 것이기 때문이다. 따라서 "geluk"에 관한 네덜란드 질문지는 "운이 좋은"에 관한 믿음을 묻는 질문을 포함할 가능성이 높지만, 미국인을 위한 질문지에서 "행복"에 대해 질문을 할 때는 "운이 좋은" 과 같은 질문이 포함되지 않을 것이다.

2.3.1.3 다면 설계(facet design)

구성개념을 어의적으로 분석하기에 유용한 기법은 Guttman의 다면 설계이다 (Guttman, 1954). 다면 설계는 다면 내에 포함되는 요소들을 가진 다면체계로 관찰치들을 분류함으로써 관찰치들의 총체(universe of observation)를 정의하는 것이다. 여기에서 다면은 관찰치들을 분류하는 여러 방법들이고, 요소들은 각 다면 내에 있는 서로 구별되는 종류(class)들을 말한다. 관찰치들의 총체는 (1) 모집단을 분류하는 모집단 다면, (2) 변수들을 분류하는 내용 다면과 (3) 변수들을 위한 응답 범주들의 공통 범위라는 세 가지 기준들을 사용하여 분류된다.

우리는 현재 우리가 추구하는 목표를 위해 변수들의 다면 구조를 집중적으로 다룬다. 다양한 내용 다면은 분산분석(ANOVA)설계와 유사한 교차 분류(cross-classification)로 볼 수 있는데, 이것은 질문지 문항들 간의 유사점들과 차이점들을 구분하는 것이다. 각 *면(facet)*은 존재 가능한 관찰치들을 정의하는 일단의 요소들로 구성된 특정개념 분류체계를 나타낸다. 공통 *응답 범위*는 필요한 정보의 유형을 규정한다. 예를 들면, 태도 질문은 일반적으로 매우 부정적…매우 긍정적으로의 응답 범위를 갖고 있고, 가치 질문은 일반적으로 매우 중요하지 않은…매우 중요한이라는 응답 범위를 갖는다.

표 2.1 대응문장의 예

"(X)라는 사람이 느끼는 것이 무엇인지 만들어주세요

출처		반응
(그녀 자신의 경험은)	그녀는 확신한다	(더 건강하다고)
(그녀의 남편은)	믿고 있다	(더 날씬하다고)
(그녀의 의사는)	그녀는 될 것이다	(육체적으로 훨씬 더 매력적으로)
(언론은)		(옷 문제가 덜 생길것이다)
		(사회적 치욕으로 덜 고통받을 것이다)
		(사회적 지위에서 덜 불안할 것이다)
		(덜 의기소침해질 것이다)

	반응
만약에 그녀가 몸무게를 줄인 다면,	(실제로는 전혀 줄지 않을 것)
	(많이 줄지는 않을 것)
	(아주 약간 줄어들 것)
	(상당히 줄어들 것)
	(꽤 많이 줄어들 것)
	(아주 많이 줄어들 것)
	(정말로 아주 많이 줄어들 것)

여기에서 (X)는 체중 감량 프로그램에 참여하는 결혼한 여성이다.

(출처: Gough, 1985, p.247)

내용 다면은 이것들이 정의하는 구성개념에 적합해야 한다. 가장 적절한 다면과 요소들을 선택하는 데 있어서, 내용 영역에서 중요한 모든 측면들을 명확하고 분명하게 기술하는 것을 목표로 삼아야 한다. 예를 들어, 대부분의 구성개념들의 경우, 관련 행동들을 정의하는 행동 면과 이 행동들이 발생하는 상황들을 정의하는 상황 면을 구별하는 것이 유용할지도 모른다. 다면 설계에서 다면 구조는 종종 관찰치들을 하나 이상의 통상적인 문장들로 기술한 대응문장에 의해 언어로 표현된다. 대응문장에 의한 다면 설계의 예는 표 2.1에서 제시하였는데, 이것은 Gough(1985, p. 247)의 체중감량 프로그램 참가이유에 관한 대응문장이다.

이 다면 설계에서 두번째면(출처)은 확신의 출처를 가리키고, 세번째 면(이유)은 체중 감량의 구체적인 결과를 가리킨다. 앞에서 볼 수 있는 다면 설계를 사용하여 질문지 문항들을 만들 수 있다. (X)면은 개체들의 목표 모집단을 나타낸다.

출처 면은 네 가지 요소들을 갖고 있고 이유 면은 일곱 가지 요소들을 갖고 있어 4 X 7 = 28가지의 질문들을 정의할 수 있다. 예를 들어, 출처 면과 이유 면의 첫 번째 요소들을 결합하면 "귀하께서는 자신의 경험을 통해 몸무게가 줄어들면 건강이 증진될 것이라고 확신하게 되었습니까?"라는 조사 질문을 얻을 수 있다 (Gough, 1985, p. 257)

다면 설계는 다면 이론이라는 더 일반적인 접근방법 중 일부인데, 이것은 다면 구조를 사용하여 항목들 간의 유사점들에 관한 가정을 만들고, 이 구조를 분석 및 해석하는 데 기하학적 표현을 주로 사용하는 것이다(Shye 등, 1994; Borg 와 Shye, 1995). 다면 이론은 특정 다면을 도입할 필요 여부를 결정하기 위한 일반 지침들은 포함하지 않는다. 대신에 다면 설계가 하나의 정의임을 강조하는데, 이것은 설계를 진위 여부로 판단해서는 안 되고 생산적인 조사 결과를 낼 수 있는지 여부로 판단해야 한다는 것을 의미한다. 또한 다면 설계에서는 조사 문제를 항상 염두에 두고 구성해야 한다. 따라서, 구성개념의 의미를 정하는 한 방법인 다면 설계는 우리가 조사하기를 원하는 경험적 영역을 이미 잘 알고 있다고 가정한다. 예를 들면, Gough의 다면 정의는 문헌 및 시험 조사(pilot research)에서 얻은 것이다.

보통, 다면 설계는 구성개념의 외연적 의미를 정하는 데 사용한다. 이 설계에서는 경험적 관련대상들을 정하고, 대응 문장에서 실제 조사 질문들을 얻는다. 하지만, 차원/지표 분석이나 어의 분석에서 구별한 사항들을 다면 설계에서 별개의 면들로 도입할 수 있는데, 이 경우에 다면 설계는 내포적 의미와도 관련이 있다. 이 두 가지 사례에서 다면 설계를 사용하면 체계적이고 포괄적인 개념 분석이 촉진된다. 또한 기존 도구 내에 있는 질문들을 분류하는 데에도 다면 설계를 사용할 수 있다.

2.3.2 조작화에서 개념화로: 자료중심 접근방법

본 절에서 설명하는 자료중심 방법들은 내용 표본, 상징적 상호주의 및 개념도(concept mapping)이다.

2.3.2.1 내용 표본

이론적 구성개념과 실제 조사질문 간의 간격을 좁히는 대안적 접근방법은 조사 문제와 관련되어 보이는 대규모 질문세트를 구성하고, 자료를 수집하고, 어떤 질문들이 공변(co-vary)하는지 따라서 동일한 구성개념을 측정하는 것으로 보이는지를 판단하는 것이다. 이것은 철저히 귀납적인 접근방법이다. 우리는 하나 또

는 둘 이상의 개념적 핵심을 공유하는 질문들을 구성할 수 있을 정도로 피험자를 충분히 알고 있다고 가정하고, 이 공통 핵심들이 무엇이고 어떤 질문으로 측정할 수 있는지를 판단하기 위해 요인분석이나 이와 유사한 기법들을 사용한다. 우리는 자료로부터 시작하여 이론적 구성개념을 추론하는데, 이 개념들은 후속 경험적 연구에서 변경할 수도 있다.

측정 및 개념화에 관한 자료중심 접근방법은 피상적이고 *임시적*인 것으로 보인다. 그럼에도 불구하고 Schuessler의 복지척도분석에서 그 예를 찾을 수 있는 바와 같이, 이로부터 유용한 표준척도를 얻을 수 있다(Schuessler, 1982). Schuessler는 1920년부터 1970년까지 미국 사회학 연구에서 사용하였으며 9500개 이상의 질문들을 보유하고 있는 140개 이상의 척도들을 분석하였다. 그는 이 척도들 중 다수가 *임시적*이라는 것을 알았다. 이 척도들은 한 번이나 두 번 일회성으로 사용되었다. 여기에는 중복되는 문항들이 많이 있다. 혼란스러운 한 가지 발견점은 개념적으로 다양한 척도들이 여러 문항들을 공통적으로 갖고 있는 것인데, 중복되는 것을 제외하면 질문은 통틀어서 약 500개였다. Schuessler는 복지를 측정하기 위한 표준척도세트를 구성하기 위해 237개 문항들을 선정하였다. 1,522명의 응답자 표본의 응답들에 관한 요인분석에서는 17개 요인들이 발견되었고, 이로부터 최종적으로 95개 문항들을 가진 12개 척도를 얻었다.

Fiske(1971)가 지적하듯이, 하위요인을 찾기 위한 질문지 하위세트의 요인을 포함하여 구성개념들과 구성개념 간 관계를 추론하기 위해 요인분석 결과를 사용하는 경우, 우리는 출발점만 다를 뿐 차원/지표 접근방법과 유사한 작업을 하게 된다. 이런 유형의 자료중심 접근방법에서 피해야 할 위험은 이미 찾은 요인들을 *구체화(reification)*하는 데에 있다. 즉 요인들은 고정된 인과적 실체들(causal entities)보다는 응답일관성(response consistencies)을 요약하는 편리한 기술범주로 해석해야 한다(Anastasi, 1985). 이론적인 관점에서 볼 때, 이런 일관성이 나타나는 이유는 흥미롭다. 요인에 부여하는 명칭과 그에 대한 해석은 이러한 응답사건들(response contingencies)을 초래하는 과정을 어떻게 설명하느냐에 따라 다를 수 있다.

내용 표본에서는 조사 영역의 중요한 측면들을 빠뜨리지 않는 것이 중요하므로, 통상적인 접근방법은 *대규모* 문항세트를 구성하는 것이다. 예를 들어, Cattell(1973)은 사전에서 기분(mood)을 기술하는 데 사용할 수 있는 모든 단어들을 취하여 기분척도를 개발하였다. 단어들을 무작위로 추출하여 질문지 문항들을 구성하였고, 후속 조사를 통해 "평가(evaluation)"와 "각성(arousal)"과 같은 여러 하위척도를 가진 기분척도를 만들었다. Cattell 등(1970)은 16PF(Personality

factor) 테스트를 구성할 때 이와 유사한 접근방법을 사용했다. 연구자는 대규모 질문지세트를 가지고 시작하기 때문에, 기본차원을 찾기 위해 요인분석을 자주 사용한다. 이 분석의 목적은 동일 구성개념을 측정하는 질문세트들을 찾는 것이다. Fiske(1982)와 Marradi(1981)는 한 질문세트가 동일 구성개념을 측정하는 경우 이 세트를 요인분석하면 단일요인으로 묶여져야 한다고 지적하였다. 그리고 나서 2단계로 진행된다. 1단계에서는 설명요인의 분석이나 유사한 수단을 사용하여 예비 질문세트들을 찾아낸다. 제2단계에서는 실제로 단일요인만으로 충분한지 여부를 검증하기 위해 이 세트들을 개별적으로 분석한다.

앞 문장에서 "요인분석"은 자료중심의 구성개념형성의 첫 단계로서 차원이나 척도가능성(scalability)이라는 측면에서 질문지 문항들 간의 유사점들을 검토하기 위한 모든 기법들을 가리키는 일반적인 용어로 사용되었다. 하나의 기법으로서 요인분석은 차원(dimensionality)을 엄격하게 검증하지 않고 실제로 측정치를 산출하지 않기 때문에 신랄한 비판을 받아(예를 들면 Duncan, 1984), 대신 Rasch모형과 같은 문항응답모형이 선호되고 있다. 그러나 여전히 이 모형도 잠재변수들의 구체화와 잘못된 해석을 하지 말라는 상기한 경고들을 준수해야 한다. Mokken모형이나 Rasch모형과 같은 일부 척도모형에 따라 잘 측정할 수 있는 문항 세트들을 찾더라도 이 세트들은 단지 관찰된 응답 일관성들을 찾기 위한 것뿐이다. 척도들의 이론적 해석은 우리가 이 응답 사건들이 발생하는 이유를 어떻게 설명하는가에 좌우된다.

2.3.2.2 상징적 상호주의

사회과학에서는 우리가 흔히 사용하는 언어에서 추출한 개념들을 사용한다. 이 개념들은 우리의 사회생활에서 분명한 기능을 한다. 연구자들은 과학적 이론에 더 유용하게 사용하려고 이런 개념들에 정확한 의미를 부여한다(Fiske, 1971; Kerlinger, 1986). 하지만, 사회에서 개인들의 경험과 행동에서 기술하거나 설명하려면 사회생활에서 이 개념들이 어떻게 사용되고 이해되고 있는지를 알아야 한다. 따라서, 구성개념형성에 관한 한 가지 접근방법은 일상생활에서 보통 사람들이 흔히 사용하는 개념들로부터 시작하여 이에 대한 과학적 논의로 발전해나가는 것이다.

이 접근방법은 상징적 상호주의라는 사회조사 패러다임과 부합된다. 상징적 상호주의는 사회를 개인들이 서로 상호작용하는 세계로 보는데, 이들은 상호작용하면서 서로 공감할 수 있는 정의를 찾기 위해 부단히 협상한다(cf. Blumer, 1969). 상징적 상호주의 조사에서는 개인들이 사회적 상호작용을 어떻게 해석하

는지를 알아내서 이 해석을 기술할 수 있는 개념들을 개발한다. 이를 통해 상호작용 및 해석과정에 주안점을 두는 질적조사전략을 수립한다. 개념화는 *민감화 개념들(sensitizing concepts)*로 시작하는데, 이것은 조사과정이 희망적인 방향으로 진행되는 것 이상의 의미가 없는 모호한 견해이다. 질적 현장조사에서는 민감화 개념들에 대해 자세한 경험적 준거(empirical reference)를 부여하여 여러 개념들 간의 관계를 밝힌다.

이것은 Sartori가 이론적 구성개념의 외연적 의미 및 내포적 의미를 명확하게 정했던 것과 유사하다(Sartori, 1984). 차이점이 있다면, Sartori는 주로 어의분석에 의한 이론적 개념의 명료화에 관심을 두었던 반면에, 상징적 상호주의자들은 개인들과 사회에 가능한 더 가까이 접근하려 했다는 것이다. 이것은 Glaser 및 Strauss(1967)가 명명한 현장기반이론(grounded theory)이라는 용어로 잘 요약할 수 있다.

상징적 상호주의 조사에서는 종종 비구조적 면접 및 자연주의적 관찰을 사용한다. 데이터들은 현장 메모, 조사지침서나 관찰요령 등의 기록자료들이다. 민감화 개념들에서 시작하여 개념을 개발하고 경험적으로 규정하는 것은 자료를 검토하고 어떤 개념과 기초과정들이 포함되었는가를 조사하는 방식으로 진행한다. 분석과정 중에 새로운 개념의 가치에 대한 증거를 수집할 수도 있다. 조사의 결과물은 "신개념"들의 파일인데, 이것은 새로운 사항을 발견하거나 새로운 자료를 입수할 시 지속적으로 갱신된다. 이 과정은 개념들이 *포화*된 경우, 즉 개념들의 새로운 요소가 발견되지 않는 경우에는 중단된다. 일단의 개념들을 더 작은 개수의 주개념으로 축소함으로써 분석을 종결할 수 있는데, 그 다음에 주개념들은 이들간의 관계를 정립함으로써 하나의 종합적인 이론으로 통합시킬 수 있다.

상징적 상호주의 및 기타 질적 접근방법들은 하나의 통합된 사회조사방법으로 발전하지 않고 있다. 하지만 이들로부터 조사질문을 조작화할 수 있는 개념의 개발에 유용한 방법들이 나타났다. 이 방법들의 일반적 과정은 개방형 면접을 사용하여 응답자로부터 정보를 수집한 후에 상기한 대로 분석을 하는 것이다. 코드부여(coding) 과정에서 정밀도와 신뢰도를 높이기 위한 방법들을 사용할 수 있고(Miles와 Huberman, 1994), 이 분석을 지원하기 위해 컴퓨터 프로그램을 사용할 수 있다. Foddy(1993)도 조사 질문을 작성하기 위해 상징적 상호주의 방법을 사용하는 것을 논하였다.

상징적 상호주의 접근방법의 한 예는 Schaeffer와 Thomson(1992)의 근거 있는 불확실함(grounded uncertainty)에 관한 설명이다. 그들은 조사 응답자들이

"모르겠다", "나에게는 해당이 안 된다"나 "상황에 따라 다르다" 등 여러 가지 방식으로 불확실함을 표현한다는 관찰 결과로부터 연구를 시작하였다. 이 것은 두 가지 유형의 불확실함인데, 하나는 상태 불확실함(응답자 자신의 상태에 대해 불확실함)이고 다른 하나는 응답작업 불확실함(응답자 자신의 상태에 대한 표현방법이 불확실함)이다. Schaeffer와 Thomson은 출산 동기에 관한 조사에서 준구조적인(semi-structured) 전화면접을 사용해서 표준화된 개방형 질문을 하였다. 그들의 분석은 대체로 상기한 상징적 상호주의 방법을 따른다. 그들은 중립, 명확성 결여, 양면가치(ambivalence), 우유부단 그리고 뒤섞인 표현(mixed expression) 등의 핵심개념들을 얻었다. 마지막으로, 그들은 폐쇄형 질문지에서 모르겠음 범주를 확대하고 상태 불확실함의 여러 유형들을 직접 측정했다.

다른 분야에서 볼 수 있는 또 다른 예는 "지능"의 의미에 관한 Sternberg의 조사이다(Sternberg, 1985). Sternberg는 응답자들에게 전형적인 "지능적" 행동들과 "비지능적" 행동들을 열거하도록 했다. 그는 이 리스트에서 250가지의 행동들을 선정하고, 두 번째 표본에 대한 요인분석에서 "실제 문제해결능력, 언어능력 그리고 사교능력"이라는 세 요인들을 산출하였다(Sternberg, 1985, p. 60). 재미있는 관찰결과는 일반인과 전문가의 요인구조들이 매우 유사하다는 것이었는데, 이 사실을 통해 지능에 관한 질문을 일반 응답자들에게 제시할 수 있다는 결론에 도달하였다.

2.3.2.3 개념도

상징적 상호주의 접근방법의 단점은 시간과 노력이 많이 소요된다는 것이다. 다른 약점은 구조(structure)의 취약으로 응답자의 기여와 연구자의 기여를 구분하기가 어렵다는 것이다. 응답자가 특정 주제에 관한 자신의 사고를 정립하는 기본적인 방법은 직접적으로 적용할 수 있다. 질문지 문항 작성을 위해 상징적 상호주의 조사는 진술의 생성과 구조화라는 두 개의 주 원리로 요약할 수 있다. 진술생성단계는 탐색단계로서 관심영역 전부를 다루는 포괄적인 진술목록을 만드는 것을 이상적인 목표로 삼는다. 구조화단계는 진술목록을 이것을 포함하는 주개념의 목록으로 좁히는 것이다.

표적집단(focus groups)은 응답자가 자신의 경험이나 반응을 기술하기 위해 어떤 개념을 사용하는지 알아내기 위해 이용한다. 표적집단은 세심하게 선정한 응답자 집단들이 중심 주제를 논의하고 이것을 여러 가지 다른 관점으로 검토한다. 일반적으로 표적집단은 목표 모집단으로부터 추출한 개인으로 구성된다. 표본의 기준은 *분산의 대표성*은 아니지만 광범위하게 다양한 사람들이 포함된다. 때

로 특정 상황이나 경험을 공유하는 사람들과 같은 특정 집단을 사용한다. 표적집단에서의 논의는 기록하고, 나중에 관심 영역에 대한 진술들을 추출하기 위해 분석한다.

연구자가 스스로 이 진술 목록을 구조화하거나 응답자들로부터 더 많은 자료를 수집하여 이것을 구조화할 수 있다. 진술들의 수가 작은 경우 모든 짝의 조합들(pairwise combinations)을 응답자에게 제시하고 각 쌍이 유사한 만큼 척도값을 매기도록 요청할 수 있다. 이로부터 나오는 유사행렬은 이 진술들을 어떻게 더 추상적인 개념들로 묶을 것인가를 결정하는 다차원 척도법(multidimensional scaling)이나 기타 유사성 분석의 대상이 된다. 진술의 수가 k인 경우, 짝 비교의 수는 $(k \times (k-1))/2$이기 때문에 이 짝 비교는 진술들의 수가 10개를 넘는 경우에는 바람직하지 않다. 진술들의 수가 큰 경우, 진술의 경험적 구조를 검사하기 위해 분류작업을 자주 수행한다. 분류작업에서 응답자는 유사한 진술들을 한 군집으로 묶는다. 이 분류 작업은 완전히 무제한적이다, 다시 말하자면 여러 군집들이 있을 때 진술을 각 군집으로 분류하는 것은 간단하지만, 군집의 수와 각 군집에 해당하는 진술의 수를 임의로 정할 수 있다(보통 모든 진술을 하나의 군집에 놓거나 진술의 수만큼 많은 군집들을 만드는 것은 허용되지 않는다). 예를 들어 자연집단분석(natural grouping analysis)의 경우에서는 분류 작업을 제한할 수 있다. 자연집단분석은 일단 모든 진술들을 하나의 큰 군집에 포함시킨다. 응답자는 이 군집을 둘로 나누고 왜 이 두 군집이 서로 다른지 설명해야 한다. 다음으로 두 개의 군집 중 하나의 군집을 더 나누고 역시 응답자는 왜 이 두 군집이 서로 다른지 설명해야 한다. 이 과정은 응답자가 더 이상 구분할 수 없다고 생각할 때까지 반복한다. 진술들이 얼마나 빈번하게 군집화되는지를 표로 작성함으로써 유사성 행렬을 구성할 수 있고, 그 다음에 이것에 대한 유사성 분석을 실시할 수 있다.

개념도를 체계화하는 여러 가지 접근방법들이 제안되었다(이에 관한 개관은 Jonassen 등(1993) 참조). Trochim(1989)은 표적집단들에서의 개념도를 위한 포괄적 시스템을 개발하였는데, 이것은 (1)관심개발 (2)진술생성 (3)진술구조화 (4)진술표현 (5)개념도 해석 및 (6)개념도 활용의 여섯 단계이다. 1단계 "관심개발"에서는 주요 관심, 즉 관심영역을 규정하고 표적집단에 참여할 자들을 선정한다. 2단계 "진술생성"은 관심영역에 관련된 모든 측면에 대한 진술들을 생성하기 위해 표적집단과 함께 하는 브레인스토밍 회의이다. Trochim(1989)은 "노인들이 갖고 있는 문제들, 관심사나 요구사항들…."을 집중적으로 다루는 브레인스토밍 회의의 예를 제시하였다(Trochim 1989, p. 4). 제3단계 "구조화"에서 개개의 참여자들은 인쇄된 진술문들을 "자신에게 의미가 있는" 여러 묶음으로 분류하라는 지시를 받

는다. 또한 참여자들에게 중요성에 따라 진술에 등급을 매길 것을 요청한다. 개개의 군집들(individual sorts)을 군집 유사성 행렬에서 조합한다. 제4단계 "진술 표현"에서는 이 유사성 행렬에 다차원 척도법을 적용하고, 해석을 용이하게 하기 위해 이 개념도 위에 군집해(cluster solution)를 겹쳐 놓는다. 사실 군집은 "새로운 개념"이라고 할 수 있다. 중요성 등급을 개념도 위에 겹쳐 놓을 수도 있다. 노인들에 관한 Trochim(1989)의 연구로부터 얻을 수 있는 개념도의 예는 그림 2.2에서 보는 바와 같다. 제5단계 "개념도 해석"은 다시 집단활동이다. 참여자들은 각 진술 군집이 가질 수 있는 의미들과 이 군집에 적절한 명칭들에 대해 논의한다. 아울러 그들에게 명칭을 부여할 수 있는 군집 집단들(groups of clusters)이나 군집 공백 내의 영역들(regions in the cluster space)이 있는지도 확인한다. 기본적으로 이 단계에서는 참여자 집단이 승인한 개념도의 형태로 개념들 간의 관계를 파악하려고 시도한다. 제 6단계 "개념도 활용"에서는 계획 평가, 개입(intervention), 그리고 *기타*에 최종 개념도를 적용한다. 이 경우 활용은 개개의 진술들을 사용하여 개념들을 조사 질문으로 변환하는 것과 개념도를 사용하여 질문지의 경험적 구조에 관한 첫 번째 가정을 제시하는 것을 의미한다.

그림 2.2 노년층에 관한 관심사들의 개념도. (영국 Kidlington OX5 1GB, Langford Lane, The Boulevard, Elsevier Science 주식회사의 허가를 얻어 W.M.K. Trochim의 평가 및 프로그램 기획 12, pp. 1-16, ©(1989)의 평가 및 기획을 위한 개념도의 머리말에서 전재함)

2.4 요약 및 논의

사회 과학에서 사용하는 이론적 구성개념들은 복잡하고 설명하려고 하는 경험적 관찰치들과 간접적으로 관련되어 있다. 이론적 개념과 관찰 가능한 현상간의 차이로 인해 연구자들은 자신들이 이론화하려는 대상을 어떻게 측정할 것인가를 명시해야 한다. 조사연구에서 이론과 데이터는 당시의 이론에 내재되어 있는 구성개념과 관련된 조사질문을 작성함으로써 연결된다.

본 절에서는 개념화와 질문작성을 위한 두 가지 일반적인 방법들을 제시하였다. 첫 번째 방법은 이론과 이 안에 포함된 구성개념들에 논리적인 우선순위를 주는 것이다. 연구자는 이론적 구성개념과 이것에 추가적 의미를 부여하는 것으로 작업을 시작한다. 구체적인 조사 질문을 얻기 위해서는 구성개념의 추가적 의미를 명확하게 할 필요가 있다. 이 과정에 지침이 될 세 가지 접근방법들을 설명하였다. 첫째는 차원/지표 분석인데 이 접근방법은 Lazarsfeld와 Fiske 등의 많은 방법론자들이 주창하였고, Blalock, Cronbach와 Meehl 등의 많은 다른 학자들의

저서에도 은연중에 표현되어 있다. 많은 연구자들은 이것을 *유일한* 접근방법으로 보는 것 같다. 차원/지표 분석에서 우리는 중심개념을 하위개념들로 구분하고 각 하위개념의 경험적 지표들을 규정한다. 두 번째 접근방법은 구성개념에 관련된 의미들을 규정하고 각 하위구성개념의 경험적 관련대상을 찾기 위해 어의분석을 사용한다. 세 번째 접근방법은 다면설계인데, 여기에서는 다면구조와 공통응답범위를 규정함으로써 구성개념내용의 모든 요소들을 설계한다.

두 번째 방법은 경험적 관찰에 논리적 우선순위를 부여한다. 물론 중심 사건이라는 관념이 반드시 존재하지만, 이것은 모호한 것일 수 있다. 연구자는 관찰치들로부터 이론적 구성개념을 추출해야 한다. 이 과정에 지침이 되는 세 가지 접근방법들을 설명하였다. 첫 번째는 내용표본인데, 이 경우에는 중심사건과 관련있어 보이는 대규모 질문지세트가 이미 존재하거나, 이 세트를 쉽게 구성할 수 있다. 다음에 요인분석이나 관련 방법들을 사용하여 대규모 질문지세트를 소규모 이론적 구성개념 집단으로 축소시킨다. 나머지 두 가지 접근방법들에서는 흔히 사용하는 언어를 활용한다. 그 중 한 접근방법은 상징적 상호주의에서 사용하는 연구방법들을 구성개념형성문제에 맞게 변형시킨다. 다른 한 접근방법은 인지적 연구에서 가져온 개념도 방법을 사용한다.

이론중심 대 자료중심 방법들의 장단점은 Hox와 De Jong-Gierveld (1990)에서 자세히 논의하고 있다. 탐구해야 할 질문세트로 시작하는 경우를 제외하고 이 접근방법들은 조사질문을 직접적으로 산출하지 않는다. 이론중심 방법에서 우리는 조사질문으로 해석하기가 충분히 명백한(명백하기를 바람) 개념들의 목록을 얻게 되었는데, 이것은 비교적 기계적인 작업이다. 자료중심방법에서 우리는 상당히 구체적인 진술 그리고 이 진술과 추상적인 구성개념과의 관계에 대한 목록을 얻었다. 이 방법은 이론중심 접근방법과 비교할 때, 진술 내에 간략하게 요약된 원자료도 보유할 수 있다는 장점이 있다. 이 원자료는 조사질문으로 변환할 수 있는, 응답자의 언어적 표현을 그대로 기록한 자료이다. 자료중심방법의 또 다른 장점은 응답자가 특정개념을 이해하는 것으로 시작한다는 것이다. 하지만 자료중심방법에서는 우리가 사용하려는 이론에 맞지 않는 이론적 구성개념들을 얻을 위험성이 있다.

타당성 있는 척도를 구성해야 하는 연구자들은 구체적인 방법을 선택하는 것이 항상 중요한 것은 아니라는 Fiske(1971)의 주장에서 위안을 얻을 수 있다. 타당성 있는 척도를 구성하는 것은 일회성이 아니라, 척도를 고안하고, 이론적 및 경험적 의미를 연구하며, 더 나은 척도로 개발하는 지속적인 과정이다. 척도구성에 대한 체계적 접근방법의 중요성에 관한 논의는 Heath와 Martin이 저술한 제3

장을 참조하기 바란다. 이 과정이 종료할 시점에서 실제 출발점은 중요하지 않다는 것을 알게 될 것이다(cf. Oosterveld와 Vorst, 1995). 중요한 것은 이론적 분석 및 경험적 검증 모두가 적당한 시점에서 시작하여 발전하는 지침이 될 수 있도록 해야 한다는 것이다. 이 과정에서 우리는 이론적 극단 및 경험적 극단의 연속선 사이를 왔다갔다하게 된다.

개념 형성의 최종 산출물은 *원형조사질문(prototype survey question)*이다. 이것은 조사대상인 모집단, 조사문항연구 등의 통상적인 관심사들을 처리해야 한다는 것을 의미한다. 본 절에서는 질문지 작성과 이것을 인지적 실험실과 현장에서 검증하는 것에 관련된 문제들을 다뤘다.

감사의 글

필자는 본 절의 초안에 대해 논평을 해 준데 대해 Edith de Leeuw, Anthony Heath, Jean Martin, Norbert Schwarz와 Arie van Peet에게 감사를 표한다. 또한 본 절에서 제시한 주제들에 관하여 격려가 되는 논의를 한 데 대해 개념화 및 조사 설계에 관한 네덜란드 연구 위원회 위원들 모두에게 감사를 표한다.

제 3 절 질문 순서의 효과 줄이기: 보조 문항의 작용

Michaela Wänke
university Heidelberg
Norbert Schwarz
University of Michigan, Ann Arbor

3.1 머리말

선행 질문이 다음 질문의 응답에 영향을 준다는 것은 조사연구자들 간에 잘 알려진 사실이다. Sayre(1939)가 질문순서의 효과에 관한 최초의 체계적 연구결과를 보고한 후로 조사연구자들은 광범위한 문맥효과에 대하여 보고하였다(Schuman & Presser(1981), Schwarz & Sudman(1992), Sudman(1996) 제3장에서 제6장까지 Tourangeau & Rasinski(1988) 참조). 이들 연구에서 알 수 있듯이, 질문순서의 효과들은 매우 다양하다. 선행 질문이 다음 질문의 응답에 영향을 주는 것은 질문맥락이 함축하는 의미와 동일한 방향(동화효과 발생)으로나 반대 방향(대조효과 발생)으로 발생할 수 있다(예를 들면, Schwarz & Strack, 1991; Tourangeau & Rasinski, 1988). 게다가 선행 질문은 응답의 생략을 야기하거나(예를 들면, Martin 등(1990)) 응답의 질적 특성을 변화시킬 수 있다(예를 들면, Smith, 1992).

연구자들은 선행 질문의 효과를 줄이기 위해 일반적으로 관련 질문들 사이에 여러 개의 무관 문항들을 끼워 넣는데 이것을 "보조 문항(buffer item)" 이라고 한다. 보조 문항에 관하여 최초로 논한 학자는 Cantril(1944)이었다. 그는 미국인들에게 프랑스나 영국 군대에서 복무하는 것에 대한 찬성여부를 묻는 질문을 할 때, 그에 앞서 독일 군대에서 복무하는 것에 대한 찬성여부를 묻는 질문을 먼저 제시하면 전자의 질문에 찬성할 가능성이 낮아진다는 것을 발견하였는데, 이 효과에 대해서는 아래에서 더 자세히 논한다. 게다가 그는 순서 효과는 질문 사이에 여러 보조 문항들을 삽입하는 것으로 감소시킬 수 있다는 것을 관찰하였다. 하지만 보조 문항을 다음 문장에 삽입시키는 것만이 최선은 아니다. 오늘날까지 보조 문항들이 효과를 나타내는 조건들에 대해서는 거의 알려진 바가 없다. 일부 연구에서는 보조 문항을 사용하였을 때 질문 순서의 효과가 감소하지 않았다고 밝힌 것도 있고(예를 들면 Schuman 등, 1983; Bishop 등, 1984, 실험. 2; Bishop, 1987), 다른 학자들은 전혀 유의하지 않은 경향을 발견하였다(예를 들어, Tourangeau 등, 1989a). 게다가 보조 문항을 사용하는 것이 문맥의 효과를 유의

하게 감소시키는 경우에서도, 보조 문항을 사용하지 않은 경우와 비교해 볼 때 효과가 단지 약화되었을 뿐이지 제거된 것은 아니었다(예를 들면, Bishop 등, 1984, 실험. 1; Tourangeau 등, 1989b). 이보다 더 혼란스러운 것은 문맥 효과를 제거하기보다는 오히려 역효과가 발생할 수 있는 관찰 결과이다. 예를 들어, 대조 효과가 동화 효과로 바뀌어 얻어질 수도 있다(Ottati 등, 1989).

이 예들이 설명하는 바와 같이 보조 문항의 작용에 대해서는 아직 밝혀지지 않은 사항들이 많이 있다. 보조 문항은 선행 질문들이 다음 질문의 응답에 미치는 영향을 줄이기 위해 사용하는 것이므로, 이들의 작용에 관한 모든 이론적 분석은 우선 문맥 효과 발생의 근본 원인이 되는 인지적 그리고 의사소통 과정에 대해 우리가 알고 있는 사항들을 가지고 수행해야 한다. 현재, 보조 문항들은 일반적으로 인지적 접근가능성(cognitive accessibility)에 관한 다소 단순화한 가정들에 근거하여 사용하고 있다. 선행 질문은 머릿속에 질문에 관련된 대답에 뒤이어서 사용되는 정보를 불러오는데 이것은 질문순서효과의 결과이다.

뒤이어서 관련 질문들에 응답하는 데 사용하는 정보를 머리 속에 떠올리게 하여 질문 순서의 효과를 일으킬 수 있는 것으로 추정된다. 관련 없는 여러 개의 보조 문항들을 삽입하면 문맥 질문(context question)과 대상(목표) 질문(target question) 간의 시간적 간격뿐만 아니라, 그 사이에 머리 속에 떠오르는 정보의 양이 증가하게 된다. 이 두 가지 요인들로 인해 문맥 질문이 미리 제공한 정보가 응답자가 대상 질문에 응답할 시 기억 속에 아직 남아 있을 가능성이 감소되고 따라서 순서 효과가 줄어들 것으로 예상된다. 이 이론적 가정은 일반적으로 옳고 인지적 심리학상의 연구들은 이를 강력하게 뒷받침하고 있다(Anderson(1983), Wyer와 Srull(1989) 참조). 하지만, 이 요인들을 조사 측정에 응용한 결과는 예상했던 만큼 좋은 편은 아니다. 한 편으로, 순서 효과의 구체적인 특성은 어떤 정보가 머리 속에 떠오를 것인가에 좌우될 뿐만 아니라 응답자들이 이 정보를 어떻게 *사용하는가*에 좌우된다. 하지만 접근가능 정보(accessible information)의 사용 그 자체도 보조 문항의 존재여부에 의해 영향을 받아, 조사 문헌들이 보고한 바대로 많은 놀라운 결과들을 초래할 수 있다. 다른 한 편으로는, 일부 순서 효과들은 정보 접근 가능성 그 자체와는 *무관하여*, 이전에사용한 정보의 접근 가능성을 감소시키는 조건들의 영향을 비교적 받지 않는다.

본 절에서 우리는 질문 순서 효과를 끌어내는 인지적 과정 및 의사소통 과정에 대해 우리가 현재 갖고 있는 지식을 검토하고, 보조 문항이 이들 각 과정에 미칠 수 있는 효과를 탐구한다. 우리의 논의는 질문이해, 정보회상 그리고 판단형성과 응답형식구성 및 편집을 포함하는 응답자 수행의 구성으로 진행한다. 우리

는 *정보 접근 가능성* 및 *정보 사용*간의 차이점에 대해 각별한 주의를 기울일 것인데, 이 차이점은 보조 문항들이 선행 질문들의 영향을 약화시키거나 보강하거나 반전시키는 조건들을 파악하기 위한 개념틀을 제공한다. 입수 가능한 연구들의 대다수와 일관성을 유지하기 위해 우리가 든 예들의 대부분을 태도 측정 분야에서 추출하였다. 현재로서는 우리가 논하는 사항들의 대다수가 추론적일 수밖에 없는데, 철저하게 통제하여 실시한 실험들이 거의 없기 때문이다. 하지만, 우리의 개념 분석이 보조 문항들의 작용에 관한 향후 연구의 과제를 설정하는 데 도움이 되기를 바란다.

3.2 질문 순서 효과의 인지적 원천

선행 질문들이 다음 질문들의 응답에 어떻게 영향을 미치는가? 응답 과정을 더 면밀히 검토하면, 문맥이 미치는 영향을 이해하는 데 도움을 얻을 수 있다. 응답자는 조사 질문에 응답하기 위해 몇 가지 작업들을 수행한다(이에 관한 요약은 Schwarz, 제1장 참조; 더 자세한 사항은 Strack과 Martin(1987), Tourangeau(1984) 참조). 먼저, 응답자는 무엇을 의미하는 것인지 이해하기 위해서 질문을 해석한다. 이해의 수행은 문자 그대로를 번역하는 것과 어의적 이해에 도달하는 것이 포함된다. 하지만 질문의 어의적 의미(semantic meaning)를 이해하는 것은 충분치 않다. 유용한 정보가 될 수 있는 응답을 제공하기 위해, 응답자들은 면접자가 알고 싶은 것이 정확하게 무엇인지 판단해야 한다. 즉 그들은 질문의 실제적 의미(pragmatic meaning)를 추론해야 한다(예를 들면, Schwarz, 1994; Strack, 1992). 다음으로, 응답자는 질문에 응답할 수 있는 정보를 회상한다. 만약에 의견을 묻는 질문을 한다면 응답자는 이전의 경우에 형성했던 판단을 회상할 것이다. 하지만, 이전에 형성된 판단 중에 이번 질문과 직접적인 관련이 있는 것을 기억해내는 경우는 드물고, 대부분의 경우 응답자는 면접하는 바로 그 시점에서 즉각적인 판단이 가능한 정보를 회상한다. 질문이 행동에 관한 것도 마찬가지로 응답자가 기억에서 적절한 응답을 찾는 경우는 드물고, 이 경우에도 적당한 추정치를 얻기 위해 관련 정보를 회상해야 한다(Sudman 등(1996) 제7장부터 제9장까지 참조). 일단 판단이 형성되면, 응답자는 이것을 보고해야 한다. 대다수의 경우, 응답자는 면접자가 제공하는 응답선택문항에 맞게 판단의 형식을 구성하게 된다. 또한 일부 응답자들은 자신의 "사적인" 판단을 드러내는 것을 꺼려 구체적인 상황에 따라 자신들이 바람직하다고 인식하는 대로 이 응답을 편집할 수도 있다.

선행 질문들은 응답자의 작업수행에 영향을 미쳐 질문 이해, 회상, 판단 형성 및 응답 형식 구성이나 편집 단계들에서 질문 순서의 효과를 발생시킬 수 있다.

다음에 우리는 각 단계별로 관련 과정들을 검토하여 보조 문항들이 미칠 수 있는 효과를 파악한다. 앞으로 밝혀지게 되겠지만 질문 순서 효과의 원인이 되는 과정들은 각 단계별로 다르기 때문에 보조 문항들이 나타내는 효과도 구체적인 과정에 따라 다르다.

3.2.1 질문 이해

질문 파악 단계에서, 선행 질문들은 응답자가 질문의 *어의적* 의미뿐만 아니라, *실제적* 의미를 해석하는 데에도 영향을 줄 수 있다(이를 검토하려면 Strack(1992), Strack과 Schwarz(1992) 참조). 응답자가 자신의 어의적 지식을 활용하고 대화 규범을 준수했기 때문에 그런 영향이 발생할 수 있고, 뒤따르는 판단에 대해서 동화 효과나 대조 효과를 나타낼 수 있다. 우리는 먼저 접근가능 개념이 애매모호한 정보의 어의적 해석에 어떻게 영향을 주는지를 다루고, 사람들이 이 영향을 피할 수 있는 조건들을 찾는다. 우리는 인지적 사회 심리학상의 연구 결과를 활용하여 이 과정들을 소개하고, 그 다음에 이것들이 조사 면접에 대해 갖는 의미들에 대해 논한다. 마지막으로 우리는 응답자가 질문의 실제적 의미를 판단하는 방식을 다룬다.

3.2.1.1 어의적 의미

인지 심리학의 많은 연구들이 밝힌 바와 같이, 개념이 적절하게 들어맞도록 제공될 때, 애매모호한 정보는 가장 접근가능성이 높은 개념으로 해석한다(이를 검토하려면 Higgins(1989), Wyer와 Srull(1989) 참조). 예를 들면, "돛단배를 타고 대서양을 횡단하는 것"은 모험적인 것으로 인식하거나 무모한 것으로 인식할 수 있다. Higgins 등(1977)은 응답자들이 앞선 작업에서 "무모한"보다 "모험적"이라는 개념을 사용한 경우는 이 행동을 모험적이라고 해석할 가능성이 높다는 것을 관찰하였다. Strack 등(1991, 실험. 1)은 이 관찰 결과를 조사 분야에 확대 적용하여 선행 질문에서 사용한 개념들이 뒤이은 애매모호한 질문의 해석에 어떻게 영향을 미칠 수 있는지를 증명하였다. 그들은 응답자들에게 "교육 기부금 제도"의 가상(假想)도입에 관한 의견을 제시할 것을 요청하였다. 이 질문의 앞에는 수업료나 학생들을 위한 학자금 지원에 관한 질문을 두었다. 예상대로 이 조사에서 학생 응답자들은 이 애매모호한 교육 기부금 제도에 대해서 선행 질문이 수업료에 관한 것인 경우는 돈을 더 내도록 하는 것과 관련지었고, 선행 질문이 학자금 지원에 관한 질문이었던 경우에는 돈을 받는 것과 관련 있는 것으로 추론하였다. 그 결과 그들은 전자보다는 후자의 문맥에서 더 찬성하는 태도를 나타내었다.

이 사례가 보여주는 바와 같이, 응답자들은 질문이 애매모호할 때 이를 현재 질문에 가장 잘 들어맞는, 접근 가능성이 높은 개념으로 해석한다. 이것은 동화 효과의 결과로, 선행 질문에 의해 접근 가능하게 된 개념인 경우이다. 기억 속에 있는 관련 개념들의 접근 가능성만을 고려할 때, 관련 개념의 접근 가능성은 시간이 경과함에 따라 점점 줄어들기 때문에, 우리는 중간에 삽입된 보조 문항들이 문맥 효과의 강도를 줄일 것이라고 기대할 수 있다. (예를 들면, Higgins, 1989). 기억 속에 있는 정보의 접근 가능성은 본 절에서 여러 차례 언급하는 것처럼 고려해야 할 요소들 중 일부에 불과하다. 접근가능 정보의 효과를 이해하려면, 우리는 이 정보를 사용하거나 사용하지 않는 조건들도 고려할 필요가 있다.

중요한 것은, 인지 심리학에 관한 문헌들에서는 이런 유형의 동화 효과는 응답자들이 인식하지 *않을 때에만* 얻을 수 있다고 주장한다는 것이다. 관련 개념은 단지 최근에 다른 목적으로 사용한 적이 있기 때문에 지금 머리 속에 떠올랐다는 것이다. 이 사례가 왜 그런지 알기 위해 Higgins 등(1977)의 실험을 수행했던 피험자처럼 "도날드가 돛단배를 타고 대서양을 횡단했다." 라는 글을 읽었다고 가정하자. "무모한"이라는 개념이 떠오르면 당신은 그 행동에 대한 자신의 반응이라고 생각할 것이다. 그러나, 앞선 작업으로 인해 "무모한"이라는 말이 떠올랐을 것이라고 인식하게 되면, 당신은 이 반응을 무시하고 문장속의 행동을 해석하는 데 있어서 "무모한"이라는 것을 고려하지 않을 것이다. 그 결과 동화 효과는 나타나지 않을 것이고, 이것은 접근 가능성이 높은 정보를 *사용하지* 않은 것이다. 게다가 당신은 "부적절한" 생각에 미치는 영향을 감안하여 정확하게 반응하려고 노력할 것이다. 그 결과, 그 전에 생각했던 것보다 덜 무모한 것으로 해석할지 모른다. 사실, 몇몇 조사들에서는 이러한 정확성이 대조 효과를 나타낸다는 것을 관찰하였는데, 이는 사람들이 부적절하다고 생각하는 영향을 감안하여 "overcorrect"하였음을 나타내는 것이다. 이 점과 관련하여 여러 이론들이 제시되었는데, 이것은 본 절의 범위를 벗어나는 것이다(Martin, 1986; Martin 등, 1990; Higgins, 1989; Schwarz와 Bless, 1992a; Strack 등, 1993 참조).

안타깝게도, 이 실험 연구 결과들을 조사 연구 환경으로 일반화하기는 어렵다. 응답자의 인식이 미치는 효과를 다루는 실험들은 첫 번째 작업을 강조하는데, 이는 관련 개념을 활성화시키는데 사용되는 것이다. 두 번째 작업은 이 활성화된 개념의 효과를 평가하는 데 사용하는데 첫 번째 작업과는 서로 완전하게 독립적이다. 이 두 작업의 독립성을 감안할 때, 피험자가 그런 영향력을 인식하고 있는 경우, 그는 두 번째 작업에 미치는 첫 번째 작업의 부담을 피하려고 할 것이다. 하지만 조사에서 응답자는 일련의 질문들이 서로 관련이 있는 것으로 인식할 수 있다. 만약 그럴 경우, 그들은 관련된 선행 질문에 의해 머리에 떠오르게 된 정보

를 이용하는 것이, 왜 머릿속에 그런 생각이 떠올랐는지 완전하게 인식하고 있더라도, 잘못된 것이 아니라고 생각할 것이다. 예를 들어 Strack 등(1991)의 교육 기부금 제도에 관한 조사에서도 응답자들이 교육 기부금이라는 애매모호한 용어가 의미하는 바를 판단하기 위해 선행 질문들의 내용을 일부러 활용했을지도 모른다. 사실, 문맥 정보를 일부러 사용하는 것은 대화수행 규범에 의해 허용되는데, 이 규범은 일련의 대화가 중심 주제와 의미 있게 관련되어 있다는 것을 암시한다.

현재, 우리는 조사 응답자가 일련의 질문들을 서로 관련이 있는 것으로 인식하는 조건들에 대해 거의 아는 바가 없으므로, 다음에 뒤따라오는 질문의 해석에서 관련 정보를 사용하는 것은 타당하다. 우리는 사람들이 기본적으로 인접한 질문들을 서로 관련지어 생각한다고 가정하는데, 이런 경우 응답자는 선행 질문에 의해 떠오른 정보를 탐탁치 않은 외부 영향으로 생각하지는 않을 것이다. 이러한 기본 생각에 대해서는 다른 이슈를 담고 있는 다음 질문들을 삽입하는 것과 같은 적절한 변화를 주어 질문하는 것이 가능하다. 마찬가지로 관련 없는 다음 질문을 삽입해서 주제를 변화시킬 수 있다. 이런 경우 응답자는 실험 보고에서 관찰된 것처럼 명백하게 무관한 질문에 대한 과도한 영향을 피하려고 노력할 것이다.

이러한 고려사항들은 보조 문항들의 효과는 응답자가 질문 세트에 대해서 서로 관련되어 있는지 여부를 인식하는 것에 좌우된다는 것을 암시한다. 질문들을 관련된 것으로 인식하는 것은 하나의 질문에 의해 머리 속에 떠올리게 된 정보가 다음 관련 질문을 해석하는 데 있어서 부적절한 것이 아니라는 것이다. 따라서 응답자는 접근가능 정보를 이용할 것이고 이로써 동화 효과가 발생하게 된다. 동화 효과의 강도는 보조 문항의 수가 증가하면 감소하는데, 이것은 삽입된 보조 문항들로 인해 정보의 접근 가능성이 낮아진다는 것을 나타낸다.

응답자가 질문들이 서로 관련이 없는 것으로 인식한 이유가 예를 들면 보조 문항에 의해 분리되었기 때문이라면 꼭 그렇지도 않다. 접근가능 정보가 머릿속에 떠오른 이유는 선행 질문에서 이것을 다루었기 때문이라고 인식한 응답자들은 이것을 사용하지 않을 것이다. 이로 인해 동화 효과는 나타나지 않고 대조 효과가 나타날 수 있다. 게다가 응답자가 선행 질문이 미칠 수 있는 영향을 인식하는 그 자체는 삽입되는 보조 문항의 수에 좌우될 수 있다. 보조 문항의 수가 많으면 많을수록 응답자는 이 영향을 덜 인식하게 될지도 모른다. 따라서 우리는 보조 문항 수가 접근 가능성이 증가된 정보를 완전히 제거할 만큼 충분하지 않은 경우가 아니라 응답자가 가지는 선행 질문이 미치는 영향에 대한 인식을 감소시킬 수 있을 만큼 충분히 클 때 명백히 무관한 문항 간에 가장 강력한 동화 효과를 기대할 수 있다.

앞서 설명한 대로, 질문 이해 단계뿐만 아니라, 아래에서 살펴볼 판단 형성 단계에서도 발생하는 문맥 효과를 이해하려면 어의적 정보의 *접근 가능성*뿐만 아니라 *사용*도 고려할 필요가 있다. 이런 효과가 기억 속에 있는 정보의 접근 가능성에 의해서만 나타난다는 가정은 응답자가 선행 질문의 영향을 인식하지 못하거나 타당하지 않게 생각하는 경우에만 이루어진다. 만약 응답자가 발생 가능한 영향력을 인식하고 있지만 이를 타당하지 않게 생각하는 경우, 다음 질문을 해석하는 데 있어서 접근가능 정보를 사용하지 않게 되어 동화 효과가 나타나지 않을 것이다. 아이러니하게도, 보조 문항은 발생 가능한 영향력에 대한 응답자의 인식을 감소시켜 동화 효과가 더 쉽게 나타나도록 할 수도 있다. 향후 연구에서는 질문지의 어떤 특성이 응답자가 인식하는 질문 간의 관련성과 발생 가능한 영향력이 타당한 지 아닌지에 대한 응답자의 인식에 영향을 주는 지 규명할 필요가 있다. 다음에는 응답자가 질문의 실제적 의미를 해석하는 데 영향을 미치는 문맥 효과를 고찰한다.

3.2.1.2 실제적 의미

질문의 어의적 의미를 이해하는 것만으로는 이 질문에 응답하는 데 충분치 않다. 그 보다 응답자는 연구자에게 유용한 정보를 제공하기 위해 그가 어떤 정보에 관심을 갖고 있는지를 판단해야 한다. 어의적 해석과 마찬가지로 실제적 해석도 문맥 효과의 영향을 받는데, 대화 과정의 작용에서 흔적을 찾을 수 있다. 일상생활에서와 마찬가지로 조사 면접에서도 우리의 대화 행위는 언어 철학자 Grice의 저서(1975)에서 설명한 암묵적 지식에 의해 규율된다(이를 검토하고 조사 연구에 응용하려면 Clark & Schober(1992), Schwarz(1995), Strack & Schwarz(1992) 참조). Grice의 대화 논리 분석의 핵심은 대화자들이 협동 원리(co-operativeness principle)에 따라 대화를 진행한다는 가정이다. 이 원리는 대화에 기여한 모든 사항들이 현재 진행 중인 대화의 목적과 관련 있어야 하고 화자(話者)는 현재의 대화와 무관한 정보를 제공해서는 안 된다는 것이다. 게다가 화자가 제공하는 정보는 유용한 것이어야 하는데, 청자(聽者)가 이미 갖고 있거나 당연한 것으로 여기는 정보가 아닌 그가 필요로 하는 정보를 제공해야 한다. 이 대화 규범에 부응하려면 주어진 구체적인 문맥에서 어느 정보가 "유익한 것"인지를 판단하기 위한 추론이 필요하다. 조사 면접에서는, 선행 질문, 그리고 함께 제시된 응답 선택 문항들(이것은 본 절에서는 다루지 않는데, 이것을 검토하려면 Schwarz와 Hippler(1991) 참조)이 하나의 질문이 가지는 문맥의 일부를 구성한다.

대화에 기여한 모든 사항들은 진행 중인 대화의 목적과 관련이 있다는 이러한 가정은 다음 질문을 이해하기 위해 선행 질문을 사용하는 것을 허용한다. 앞서

논한 바와 같이, 설문지의 특성 또는 대화 상황처럼 선행 질문에 의해 떠오른 정보를 일부러 사용하는 것이 부적절하다는 것을 제안하지 않고, 우리는 응답자가 이 정보를 일부러 사용할 것이라고 가정한다. 따라서 대화 규범의 이런 면은 질문 파악 단계에서 동화 효과가 발생하는 데 일조를 하는데, 이것은 관련된 질문들이 서로 유사하다는 것을 가정하는 것이다.

하지만, 이미 제공한 정보를 다시 제공하지 말고 유익한 새로운 정보를 제공해야 한다는 요구는 공통점보다는 차이점을 강조하는 질문이라는 해석을 초래할 수 있다. 예를 들면, Strack 등(1991, 실험. 2)은 응답자에게 일반적인 행복과 삶의 만족도를 보고하도록 하고, 이 질문들에 대해서 응답자가 인식하는 대화의 관련성을 다양하게 제공하였다. 한 조건에서 행복에 관한 질문은 첫 번째 질문지에서 마지막 질문으로 제시한 한편, 삶의 만족도에 관한 질문은 다른 연구자가 구성한 것처럼 표시하여 외관상으로는 관련 없는 것처럼 보이도록 한 후 두 번째 질문지의 첫 질문으로 제시하였다. 이 경우에 응답자가 보고한 평균 행복도와 평균 만족도는 거의 동일하였고 이 두 척도들의 상관계수는 r=0.96이었다. 하지만, 다른 조건 하에서 두 질문들은 같은 질문지의 말미에 제시하였고, 이 질문들의 서두에는 "당신의 생활에 대해 두 가지 질문을 하겠습니다."라는 공통의 소개문을 넣었다. 이 경우 응답자들이 보고한 행복도와 만족도의 평균은 차이가 있었고, 이 두 척도의 상관계수는 r=0.75로 유의하게 감소하였다. 두 조건을 비교해 볼 때, 행복 및 삶의 만족에 관한 질문에 있어서 서로 다른 연구자들이 별개의 질문지에 제시할 시에는 응답자는 두 개의 의미를 동일하게 해석한 것으로 보인다. 하지만 동일한 질문지에서 공통의 소개문과 함께 제시한 경우, 응답자는 행복과 삶의 만족에 관한 질문의 의미를 차별화시켰다. 두 질문 모두에게 같은 의미를 부여하면 두 번째 질문의 의미가 중복되어 무익한 정보가 되기 때문이다.

이와 같이, 응답자가 "유익한" 정보가 될 응답을 제공하려면 어떻게 할 것인가를 결정할 때 중복 금지 규범(norm of non-redundancy)은 유사한 질문으로 보이는 것 간의 차이점을 이끌어내도록 만든다. 보조 문항들은 이 과정에 두 가지 방식으로 영향을 줄 수 있다. 첫째, 이 문항들은 응답자에게 유사한 질문 간에 대화적 관련성이 있다는 인식을 약화시켜 중복 응답을 덜 생각하도록 만든다. 짐작하건데, 행복과 삶의 만족에 관한 질문을 동일 질문지에 제시하는 것은 두 질문 사이에 무관한 몇 가지 질문을 삽입하더라도 두 질문이 서로 다르다는 해석을 유도하는 데 충분치 않을 것이다. 둘째, 다음의 사례처럼, 보조 문항의 구체적인 내용은 무엇이 "유익한" 정보이고 유익한 정보가 아닌지를 판단하는 데 영향을 줄 수 있다.

Schwarz 등(1991)은 응답자들에게 결혼생활 만족도와 일반적인 삶의 만족도에 대해서 순서를 달리하여 질문했다. 한 조건에서는 자계식 질문지에 결혼생활 만족도에 관한 질문은 페이지 하단에 제시하고 일반적인 삶의 만족에 관한 질문은 다음 페이지 상단에 제시하였다. 이 경우 양 질문들의 상관계수 r=0.67로 나타났는데, 이것은 응답자들이 일반적인 삶을 평가하는 데 있어서 자신의 결혼에 관한 정보를 고려했음을 의미한다. 하지만 다른 조건에서는 양 질문들을 동일한 페이지 상에 두고 그 서두에 "당신의 삶에 대한 두 가지 질문을 드리겠습니다. 첫 번째 질문은 당신의 결혼생활 만족도에 관한 것이고, 두 번째는 일반적인 삶의 만족도에 관한 것입니다."라는 공통의 소개문을 넣었다. 이 경우 상관계수는 r = 0.18로 낮고 유의하지 점수로 하락하였다. 응답자는 일반적인 삶의 만족도에 관한 질문을 자신의 결혼생활에 관한 정보를 제공한 후 아직 보고하지 않은 다른 측면을 가리킨 것으로 해석한 듯하다. 이런 해석과 일치하는 결과로서, 응답자들이 "결혼 생활을 제외한" 삶의 "다른 측면"들에 대해 얼마나 만족하고 있는가에 대한 명시적인 질문을 받은 조건에서 r = 0.20으로 거의 동일한 상관계수가 나타났다. 따라서 두 질문들을 동일한 대화 맥락에 두면 응답자는 중복을 피하기 위해 일반적인 삶의 만족도에 대한 질문을 삶의 다른 측면들에 관한 것이라고 실제적인 해석을 하게 된 것이다. 여러 문헌(Schuman과 Presser, 1981; Schuman 등, 1981; Kalton 등, 1978; Bishop 등, 1985)에서 볼 수 있는 바와 같이, 일반적인 질문 앞에 구체적인 질문을 제시한 경우 응답자는 일반적 질문을 해석할 시 구체적인 정보를 배제할지도 모른다는 사실은 이러한 부분-전체 대조 현상(part-whole-contrast)을 설명할 수 있다.

하지만, (두 개 이상의) 여러 구체적인 질문들을 일반적인 질문 앞에 두었다고 가정하자. 예를 들면 응답자에게 그의 직업 만족도, 여가 활동 만족도 및 결혼생활 만족도에 대한 질문을 제시한 후에 일반적인 삶의 만족에 대한 질문을 할 수 있다. 이 경우 그들은 마치 일반적 질문에 대해서 "이런 측면들을 종합해 볼 때 삶 전체에 대해 얼마나 만족하고 계십니까?"라는 의미로 질문 어구를 구성한 것처럼 생각하고, 이미 보고한 삶의 다양한 측면을 종합하여 판단하도록 요청받았다고 해석할 수도 있다. 구체적인 질문을 먼저 했다면 이런 방식으로 일반적 질문을 해석하는 것은 대화적 관점에서 볼 때 당연하다. 이 경우 일반적 판단은 삶의 각 영역의 상대적 중요성에 대한 "새로운" 정보를 제공하는데, 이것은 대화의 핵심이다. 하지만, 선행질문에서 하나의 구체적인 질문만 한 경우는 종합할 것이 없으므로 종합적인 질문이라는 해석이 이치에 맞지 않는다.

이러한 고려 사항들을 검증하기 위해 Schwarz 등(1991)은 응답자들의 결혼생활 만족도, 여가 활동 만족도 및 직업 만족도에 대한 세 가지 구체적인 질문을

한 다음에 일반적인 질문을 했다. 이 사례에서는 모든 질문들의 앞에 공통의 소개문을 두어 이들 모두를 명시적으로 동일한 대화 맥락에 두었지만 결혼 만족도와 일반적인 삶의 만족도 간의 상관계수는 감소하지 *않았다*. 이것은 응답자들이 하나가 아닌 세 가지의 구체적인 질문 다음에 일반적 질문이 위치한 경우 이 질문을 "모든 측면들을 종합하는" 질문으로 해석했다는 것을 나타낸다. 일반적인 질문을 종합적인 판단으로 요청하는 어구로 재구성한 경우에도 이와 유사한 상관계수를 보였고, 다른 측면들에 대한 질문으로 어구를 재구성한 경우에는 상관계수가 훨씬 낮게 나타났다는 사실에 의해 더욱 강하게 입증되었다(다른 예들을 보려면 Schul과 Schiff(1993), Benton과 Daly(1991) 참조).

이 발견점들은 응답자가 질문의 실제적 의미를 파악할 때 보조문항이 미치는 효과에 대한 중요한 의미를 갖는다. 우리가 앞서 살펴본 바와 같이, 하나의 구체적인 질문과 하나의 일반적인 질문을 동일한 대화 문맥에 두면, 응답자들은 중복을 피하기 위해 일반적 질문을 구체적인 질문이 다루지 *않은* 측면에 관한 것으로 해석하게 된다. 하지만, 일반적인 질문 앞에 몇몇 구체적인 질문을 제시하면, 일반적 질문의 실제적 의미가 추가 정보에 대한 것 부터 *종합적 요약* 판단을 요청하는 것까지 바뀌게 된다. 하지만 일반적 질문을 후자로 해석하는 것은 구체적인 질문들이 상호 관련되어 있고 하나의 판단으로 종합할 수 있는 경우에만 타당하다. 이 질문들이 응답자의 삶의 질과 관련 없는 문제들에 관한 것인 경우에는 이러한 해석은 타당성이 없었을 것이다. 따라서 질문의 실제적 의미에 관한 응답자의 해석에 보조 문항들이 미치는 영향은 보조 문항의 *구체적인 내용*에 따라 다르다. 우리가 이 절 전체에서 살펴보는 바와 같이, 삽입되는 문항들의 수는 이 문항들의 구체적인 내용보다는 상기한 영향의 정도가 훨씬 더 작다.

지금까지 검토한 연구들에서는 공통적인 소개문으로 관련 질문들을 소개함으로써 대화 문맥에 관한 응답자들의 인식을 조종하였다. 명시적인 소개문 없이 질문들을 단순하게 나열할 때에도 이런 과정이 작용할 것인가? 이에 대한 대답은 질문들의 유사성에 좌우되는 것으로 추정되는데 유사성이란 응답자들이 관련있다고 인식하는 것을 결정하는 것이다. 표 5.1의 Ottati 등(1989)에 의한 연구와 같이, 최소한 이 사례에서는 밀접하게 관련된 질문의 경우 그 대답은 "그렇다 "이다.

이 연구에서는 응답자에게 "사람은 자신의 의견을 공개적으로 표명할 권리가 있다."와 같은 일반적인 명제들에 동의하는지 여부를 물었다. 이러한 각 일반 명제 앞에는 예를 들면, 긍정적으로(예: 학부모회(PTA: Parents teacher Association))나 부정적으로(예: KKK단: Ku Klux Klan)) 평가된 집단의 자유로운 언론에 관한 권리를 언급하는 더욱 구체적인 한 문항을 두었다. 이외에도 각

구체적 문항 및 이와 관련된 일반적 문항 사이에 여덟 개의 보조 문항들을 두거나, 보조 문항을 두지 않고 구체적 문항을 일반적 문항 바로 앞에 두었다. 그 결과, 문항 간 간격의 효과가 아주 큰 것으로 나타났다. 보조 문항 없이 구체적인 문항과 일반적인 문항을 제시할 시, 바로 앞에 있는 구체적인 문항이 긍정적인 집단보다는 부정적인 집단에 관한 것인 경우 응답자들은 일반 명제에 더 높은 찬성률을 보였다. 예를 들면, 그들은 KKK단에게 언론권을 부정해야 한다고 응답한 직후에 일반 질문을 "KKK단을 제외하고 일반적으로 언론의 자유에 대해 어떻게 생각하는가?"로 해석하여 이 일반 명제에 더 높은 찬성률을 나타낸 것으로 추측된다. 하지만, 이 구체적인 질문과 일반적인 질문 사이에 여덟 개의 보조 문항들을 두었을 시에는 이러한 해석을 하지 않았다. 이 경우 응답자들은 구체적인 문항이 긍정적인 집단보다는 부정적인 집단에 관한 것이었을 때, 일반 명제(예를 들면, 언론의 자유)에 대해 더 낮은 찬성률을 보였다. 이것은 예를 들면 KKK단에 대한 구체적인 문항이 응답자들이 일반적 문항에 관한 판단을 내릴 때에 고려한 사항들을 머리 속에 떠올리게 하여 이 일반 명제에 대한 찬성률이 더 낮아졌다는 것을 나타낸다.

표 5.1 보조 문항들은 문맥 효과의 방향을 반전시킬 수 있다.
우호적인 집단이나 비우호적인 집단과 관련된 구체적인 명제들 및 보조 문항의 함수로 나타낸, 일반 명제에 대한 평균 찬성률

이 사례가 설명하는 바와 같이, 보조 문항들은 실제로 순서 효과의 방향을 *반전*시킬 수 있다. 이 사례에서 반전은 보조 문항의 수가 구체적인 질문과 일반적인 질문의 대화적 관련성을 줄일 수 있고, 선행 질문으로 인해 머리 속에 떠오르는 것들에 의해 영향을 받을 수 있다고 응답자가 인식하는 것을 감소시키는 데 충분하다는 것을 나타낸다. 문항의 수가 더 많았더라면, 이전에사용한 정보의 접근 가능성을 감소시키는 데 충분하여 순서 효과를 완전히 제거할 수 있었을 것이다. 하지만, 보조 문항이 없는 경우에 응답자는 이 질문을 중복 금지 규범에 따라 해석하여 대조 효과를 초래하였다.

3.2.1.3 요약

결론적으로, 앞에서 검토한 연구들은 질문 파악 단계에서의 문맥 효과가 정보 접근가능성 *그 자체*와 유일한 상관관계에 있는 것은 아니라는 것을 나타낸다. 그보다는, 문맥 효과의 구체적인 특성은 응답자가 질문 해석 단계에서 접근가능 정보를 어떻게 사용하는가에 좌우되고 대화 규범과 상관이 있다. 그 결과, 보조 문항의 작용은 단순 접근가능성 모형이 제시하는 것보다는 더 복잡한데, 그 이유는

	근접한 명제들	분리된 명제들
	일반적인 명제	
	사례1	
호의적 집단	2.19	2.58
비호의적 집단	3.78	2.06
	사례2	
호의적 집단	1.96	2.54
비호의적 집단	3.06	1.40
	사례3	
호의적 집단	1.86	3.50
비호의적 집단	2.71	1.57

참고 : 평점의 범위는 5(찬성 안 함)에서 +5(찬성함)까지이다.

Ottati 등(1989)으로부터 발췌하여 변형함. 미국 심리학 협회 저작권 보유. 허락을 받은 후 전재함.

보조 문항이 적절한 정보의 접근가능성 뿐만 아니라, 선행 질문이 미칠 수 있는 영향을 응답자가 인식하는 것, 그리고 질문의 대화적 관련성에 대한 인식에도 역시 영향을 주기 때문이다. 이러한 상호 작용을 더 자세히 파악하는 것은 중요한 향후 연구 과제이다.

3.2.2 판단 내리기

응답자는 일단 연구자가 의도한 질문의 의미를 파악한 후에, 과거에 내렸던 판단을 기억 속에서 찾거나 그 시점에 접근가능 정보들에 근거하여 즉석에서 판단을 내려야 한다. 선행 질문이 이 두 가지 행위 모두에 영향을 줄 수 있는 방식은 세 가지 정도가 있는데, 같은 질문에 대해 과거에 내렸던 판단의 접근 가능성을 더욱 높이거나, 응답자들이 나중에 활용할 수 있도록 관련된 판단을 내리도록 유도하거나, 새로운 판단을 내리는 데 사용할 수 있도록 정보를 접근 가능하게 만드는 것 등이 있다. 후자의 경우에 이 정보는 판단의 대상과 직접적인 관계가 있거나, 더 일반적인 성격을 가진(예를 들면, 일반적 규범과 관련이 있는) 정보일 수도 있다. 마지막으로 응답자는 이미 어떤 선행 질문에 쉽게 또는 어렵게 응답해 본 주관적인 경험을 갖고 있을지도 모른다. 이러한 경험도 특정 조건 하에서는 후속 판단에 영향을 줄 수도 있다. 앞으로 이러한 발생 가능한 경우들 각각에

대해 더 자세히 논하고, 보조 문항들이 그 기저를 형성하는 과정들에 어떻게 영향을 줄 수 있는지를 중점적으로 다룬다.

3.2.2.1 일반적 규범들

조사 연구에서 상당한 관심을 모았던 일반 규범은 공정성 규범(norm of even-handedness)이다. 이 규범은 "논쟁 중 일방적인 이익(또는 불이익)을 주는 경우 상대에게도 이와 동일한 것을 주어야 한다."고 규정한다(Schuman와 Ludwig, 1983, p.112). 예를 들어, "공산국가의 신문기자들이 미국에 들어와 그들이 본 뉴스를 자신의 신문사에 전송하는 것을 허용해야 한다고 생각하십니까?"라는 질문을 생각해보자. Hyman과 Sheatsley(1950)가 관찰한 바와 같이, 이 질문에 동의한 사람들은 반대로 공산국가에서도 미국기자들이 입국하여 그들이 본 대로 뉴스를 보고하도록 허용해야 하는지의 여부를 묻는 유사한 질문이 뒤따라오는 경우, 같은 질문을 앞에 제시할 경우보다 더 적었다. 대부분의 미국인 응답자들이 미국기자들이 공산국가에서 언론의 자유를 누려야 한다고 응답한 것을 고려해 볼 때, 이것은 선행질문에서 미국 기자들에 관한 질문을 했던 것이 공정성 규범을 환기시켜 공산국가의 기자들에게도 마찬가지로 언론의 자유를 인정하도록 하게 된 것이다(다른 사례들에 대해서는 Schuman과 Ludwig(1983) 참조). 이 효과는 30년 후에도 반복해서 관찰되었는데(Schuman 등, 1983), 공산국가의 기자들에게 언론의 자유를 인정하는 것에 대해 먼저 질문한 경우 미국인 표본의 44.4%가 이것을 허용하였고 미국기자에 관한 질문을 먼저 제시한 경우 70.1%가 이것을 허용하였다. 더욱 중요한 것은 관련이 없는 17개의 보조 문항들을 그 사이에 삽입한 경우 유의하지 않은 정도로 이 효과가 완화되어 응답자 중 66.4%가 공산국가의 기자들에게 언론의 자유를 허용하도록 지지했다는 것이다.

안타깝게도 앞의 조사에서 보조 문항이 비효율적으로 발견된 이유는 아직 불분명하다. 혹자는 사용한 일반적 규범의 인지적 접근 가능성이 기억에서 사라지는 시간이 길기 때문에 17개의 보조 문항들로는 불충분하다고 추측할 수도 있다. 건망성 실어증(Martin, 1980)이나 걱정(Hayes, 1964)과 같은 다른 일반 개념들을 평가하는 질문지에 보조 문항들을 삽입한 조사들도 보조 문항을 통해 순서 효과를 감소시키는 데 실패하였다는 사실은 이 추측을 뒷받침하는 것이다. 다른 한편으로, 미국 및 공산국가의 기자에 관한 질문의 문구는 거의 동일하여 앞서 유사한 질문이 제시되었다는 것을 모를 가능성은 거의 없었다. 따라서, 두 번째 질문은 응답자들에게 첫 번째 질문을 상기시켜주어 보조 문항의 효과를 잠식했을지도 모른다. 그러므로, 두 번째 질문의 문구를 구성할 때 첫 번째 질문이 덜 생각나는 방법을 사용했다면 공정성 규범이 여전히 작용하였을지 여부를 알 수 없는

것이다. 건망성 실어증 (Martin, 1980) 및 걱정(Hayes, 1964)에 관한 연구에서도 마찬가지로 관련 문항들이 아주 유사하였다. 향후 연구에서는 문맥과 대상 질문의 표면적 유사성이 질문의 순서효과를 발생시킬 수 있다는 것을 명심하여 이러한 가능성을 감소시킬 필요가 있다(예를 들면, Duncan과 Schuman, 1980; Feldman과 Lynch, 1988).

하지만, 선행 질문을 떠오르게 하는 효과는 먼저 문맥 효과를 좌우하는 과정에 달려 있다는 점에 주목하자. 문맥 효과가 어떤 일반적 규범에 기인한다면, 응답자들에게 이 규범을 상기시킬 경우 보조 문항을 삽입해도 문맥 효과가 여전히 나타날 것이다. 그러나 대상 질문의 의미를 해석하기 위해 이전에사용했던 정보를 다시 사용하면서 문맥 효과가 발생한 경우에는 응답자에게 문맥 질문을 상기시키는 것은 선행 상황의 조건에 따라 동화 효과나 대조 효과를 발생시킬 것이다. 이런 발생 가능한 다양한 경우에 관한 체계적인 탐구는 향후 유망한 연구 과제가 될 것이다.

3.2.2.2 판단의 대상에 관한 정보의 의미

조사의 문헌에서 논하는 많은 질문 순서의 효과들은 선행 질문들이 다음 판단 형성에 사용하게 될 정보의 인지적 접근 가능성을 증가시킨다. 기본적인 과정들은 인지 심리학 및 사회 심리학의 현대 이론들을 구성하는 여러 모형에서 개념화되고 있다(Feldman과 Lynch, 1988; Feldman, 1992; Schwarz와 Bless, 1992a; Schwarz와 Strack, 1991; Strack와 Martin, 1987; Strack, 1994a; Tourangeau와 Rasinski, 1988; Tourangeau, 1992). 다음에 논하는 것들은 Schwarz와 Bless(1992a)의 포함/제외 모형(Sudman 등(1996) 제 5장도 참조)에 따른 것이다. 이 모형은 동화 효과의 개념화에 있어서는 Tourangeau(1992)의 확신 표본 모형(belief-sampling model)과 매우 흡사하지만, 다른 모형에서는 제시하지 않았던 대조효과의 명확한 개념화를 제시한다는 장점을 갖고 있다.

이 모형의 기본 가정은 평가적 판단에는 대상(즉, 판단의 객체)의 심상(representation)뿐만 아니라, 이 대상을 평가할 기준에 대한 심상도 필요하다는 것이다. 이 두 가지 심상들은 문맥에 따라 다르며 장시간 접근가능 정보뿐만 아니라, 예를 들면 선행 질문에 응답하기 위해 사용하던 *일시적인* 접근가능 정보도 포함한다. 접근가능 정보가 판단에 어떤 영향을 미치는지는 사용 방법에 따라 다르고, 이 주제는 질문 파악에 관한 항목에서 이미 집중적으로 다룬 바 있다.

동화 효과

응답자가 형성한 대상의 일시적 심상에 *포함된* 정보는 *동화 효과*를 낳을 것이다. 이것은 사용했던 심상에 포함시킨 정보에 근거하여 판단을 내린다는 것을 나타낸다. 따라서, 긍정적인 의미를 갖는 정보를 추가하면 더욱 긍정적인 판단을 낳을 수 있는 반면에 부정적인 의미를 갖는 정보를 추가하면 더 부정적인 판단을 낳을 수 있다. 접근가능성이 높은 정보를 포함시킨 효과는 이 정보에 대한 평가의 극단성(evaluative extremity)에 따라 증가한다. 하지만, 동화 효과의 크기는 다른 정보들의 양과 극단성이 장시간이나 일시적으로 접근가능지 그리고 이미 형성되어 있는 심상에 포함되는지에 의해 제한된다. 따라서 오로지 하나의 정보만 포함하는 심상에 또 다른 정보를 추가하는 것은 이미 많은 정보를 포함하고 있는 심상에 하나를 추가하는 것보다 더 많은 효과를 낼 수 있다.

따라서, 보조 문항들은 동화 효과의 발생 및 크기에 두 가지 방식으로 영향을 줄 수 있다. 첫째, 보조 문항은 이전에사용한 정보 X의 접근 가능성을 감소시켜, 대상의 심상을 형성할 시 X가 머리 속에 떠오를 가능성을 감소시킬 수 있다. 이것은 조사 연구에서 흔히 가정하는데, 보조 문항과 대상 문항의 관련성이 없을 때 가장 명확하게 나타난다. 둘째, 보조 문항은 대상의 심상을 형성할 시, 사용가능한 다른 정보(Y와 Z)를 머리 속에 떠오르게 할 수도 있다. 이미 형성되어 있는 대상의 심상에 X가 포함되어 있을지라도 그 효과를 줄일 수 있는데, 이는 심상이 더 큰 정보의 집합(X, Y와 Z)을 포함하기 때문이다. 안타깝게도 대상에 영향을 주거나 영향을 주지 않는 보조 문항의 상대적 효과를 체계적으로 비교한 연구는 아직 없다.

실질적으로 관련 있는 보조 문항의 효과에 대한 예들에는 Smith(1981, 1988)의 연구가 포함되는데, 그는 선행 질문이 미치는 영향은 하나의 질문 세트 중 첫 질문에 대한 응답으로 제한된다는 것을 관찰하였다. 아마도 첫 질문은 진단 정보(diagnostic information)를 추가로 머리 속에 떠올리게 하고, 다음 판단에 대한 문맥 질문의 효과를 제한하는 것 같다(Feldman과 Lynch(1988)도 참조). 또 다른 예로서 앞서 논한 결혼생활 만족도 및 삶의 만족도에 관한 Schwarz 등(1991)의 연구를 고찰해 보자. 응답자가 자신의 결혼생활 만족도를 평정하기 전에 일반적인 삶의 만족도를 평정하였을 경우에는, 두 척도 간 상관계수r = 0.32였다. 이것은 어떤 응답자들은 삶에 대한 전반적인 심상을 형성하면서 자신의 결혼에 대한 장시간 접근가능 정보를 사용하였다는 것으로 추정된다. 하지만, 일반적 삶의 만족도에 관한 질문을 하기 전에 결혼생활 만족도에 관한 질문을 제시한 경우, 상관계수r = 0.67로 증가하였다. 이것은 결혼생활 만족도에 관한 질문이 결혼에 관한 정보를 일시적으로 접근가능하게 만들었다는 것을 나타낸다. 하지만 앞서 보았던 한 조사에서 응답자들이 자신의 여가활동 만족도, 직업 만족도 및 결혼생활

만족도에 대해 평정한 후에 일반적인 삶에 관한 질문에 답을 했을 때, 결혼생활 만족도와 삶의 만족도에 관한 질문 간 상관계수r = 0.46으로 상관이 더 적어진 것으로 관찰되었다. 이것은 응답자가 직업 및 여가 활동에 관한 질문들로 다양한 정보를 기억해야만 했다는 것이다. 이 모형이 예측하는 바와 같이, 응답자가 삶에 대한 전반적인 심상을 형성할 때 결혼에 관한 정보와 이런 정보들을 포함시킴으로써 결혼 정보의 효과가 약화되었다.

하지만, 대부분의 조사 연구에서 사용하는 보조 문항들은 대상 질문과 관련이 없다. 포함/제외 모형에 따르면, 보조 문항들은 문맥 질문이 미리 제공한 정보의 일시적 접근 가능성을 줄이는 만큼 동화 효과를 약화시킬 수 있는데, 이는 조사 연구자들이 일반적으로 채택하는 가정과 일치한다. 한편, 실제로 관련이 있는 보조 문항들은 대상에 대한 정신적 심상을 포함하는 정보를 추가로 접근가능하게 해서 동화 효과를 약화시킬 수도 있다. 아울러 관련된 보조 문항들은 소급적 간섭(retroactive interference)으로 인해 앞서 얻은 자료들의 접근 가능성을 감소시킬지도 모른다(Feldman과 Lynch, 1988). 따라서, 여러 가지의 문맥 효과들은 서로를 약화시킬 수도 있는데, 이것은 보조 문항을 적용한 사례의 어떤 부분이 성공한 것처럼 보이는 근본적인 이유일지도 모른다.

대조 효과

포함/제외 모형에 따르면 동화 효과를 일으키는 동일한 정보가 *대조 효과*도 일으킬 수 있다. 이것은 이미 형성되어 있는 대상의 인지적 심상에 이 정보가 포함된 것이 아니라 *제외된* 경우이다.

첫 번째 발생 가능한 경우로, 긍정적(부정적) 의미를 갖는 주어진 한 정보가 대상 범주의 심상으로부터 제외되었다고 가정하자. 이 경우, 심상은 덜 긍정적(부정적)인 정보를 포함하게 되고, 결과적으로 덜 긍정적(부정적)인 판단을 내리게 한다. 이것을 차감에 근거한 대조 효과(subtraction-based contrast effect)라고 한다(Bradburn(1983), Schuman과 Presser(1981)참조). 차감형 대조 효과의 크기는 대상의 심상으로부터 제외된 일시적 접근가능 정보의 양과 극단성에 따라 증가하고, 대상의 심상에 남아 있는 정보의 양과 극단성에 따라 감소한다. 따라서, 많은 정보들을 포함하는 심상에서 제외하는 것은 이것을 다른 몇 가지 정보만 포함하는 심상에서 제외하는 것보다 효과가 더 작다. 동화 효과와 마찬가지로 보조 문항은 문맥 정보의 접근 가능성을 감소시키거나 오히려 대상의 심상에 포함시킬 수 있는 다른 정보를 접근 가능하게 함으로써 영향을 미칠 수 있다.

두 번째 발생 가능한 경우로서, 응답자들은 이미 형성되어 있는 대상의 심상에서 접근가능 정보를 제외시킬 뿐 아니라, 비교 기준이나 척도 양극점(scale anchor)을 구성하는 정보를 사용할 수도 있다. 기준이나 척도 양극점을 구성하는 데 사용한 장시간 접근가능 정보보다 일시적으로 접근가능 정보가 더 극단적이라면, 일시적인 정보는 더 극단적인 기준을 생성하게 되고 대조 효과가 나타난다. *비교*에 근거한 대조 효과(comparison-based contrast effects)의 크기는 기준이나 척도 양극점을 구성하는 데 사용한 일시적 접근가능 정보의 양과 극단성에 따라 증가하고, 이것을 구성하는 데 사용한 장시간 접근가능 정보의 양과 극단성에 따라 감소한다. 따라서 많은 정보들을 포함하고 있는 기준의 심상에 하나의 정보를 추가하는 것은 다른 몇 가지 정보들을 포함하는 심상에 추가하는 것보다 기준을 덜 변화시킨다. 마찬가지로, 보조 문항은 문맥 정보의 접근가능성을 감소시키거나 또는 관련 기준의 심상을 형성에 사용할지 모르는 어떤 정보의 접근가능성을 증가시킴으로써 영향을 줄 수 있다.

포함 및 제외의 결정인자들

포함/제외 모형(Schwarz와 Bless, 1992a)에 따르면, *초기 과정은(default operation)*은 머리 속에 떠오른 정보를 대상의 심상에 포함시키는 것이다. 이것은 조사 연구에서 대조 효과보다는 동화 효과를 발견할 가능성이 높다는 것으로, 메타분석으로 처리해야 한다. 이와는 대조적으로, 대상의 심상에서 정보를 제외시키는 것은 질문응답과정의 두드러진 특징들에 따라 행할 필요가 있다. 원칙적으로 정보의 범주화에 영향을 주는 모든 변수는 동화 효과 및 대조 효과의 발생에 영향을 주기 쉽기 때문에, 이 모형은 대체로 범주화 과정에 관한 인지적 연구와 관련이 있다. Schwarz와 Bless(1992a)는 사회적 판단을 내릴 때 문맥 효과에 영향을 주는 이질적 변수들을 검토하였다. 이 변수들은 머리 속에 떠오른 정보와 관련하여 응답자들이 내려야 할 세 가지 암묵적 결정들(implicit decisions)과 관련있는 것으로 개념화할 수 있다.

머리 속에 떠오른 일부 정보는 질문과 무관한 문제이기 때문에 전혀 고려할 가치가 없을지도 모른다. 다른 정보는 현재의 응답수행과 잠재적인 관련성이 있을지 모르므로, 응답자들은 이것을 어떻게 처리할 것인가를 결정해야 한다. 첫 번째로 결정할 사항은 왜 이 정보가 머리 속에 떠올랐는가에 관한 것이다. "엉뚱한 원인"으로 인해, 예를 들면, 응답자들 스스로 선행 질문이 영향을 미친다는 것을 인식하고 있기 때문에 머리 속에 떠오른 것으로 보이는 정보는 제외될 가능성이 높다(예를 들면, Lombardi 등, 1987; Ottati 등, 1989; Strack 등, 1993). 이것은 우리가 질문파악단계에서 논했던 내용 중 응답자 스스로 선행 질문이 영향을 미칠 수 있다는 것을 인식하는 경우 접근가능성이 높은 정보를 사용하지 않는 것과 유사하다. 앞서 설명한 대로, 보조 문항은 인식의 정도를 줄일 수 있다. 이런 이유로, Ottati 등(1989)의 연구에서 밝힌 바와 같이 보조문항은 대조 효과의 발생 가능성은 낮추고 동화 효과의 발생 가능성은 높일지도 모른다.

두 번째 결정은 머리 속에 떠오른 정보가 대상 범주(target category)에 "속하는지 여부"를 정하는 것이다. 문맥 질문의 내용(예를 들면, Schwarz와 Bless, 1992a), 대상 범주의 범위(예를 들면, Schwarz와 Bless, 1992b), 정보의 극단성(예를 들면, Herr, 1986)또는 대상 범주의 대표성(예를 들면, Strack 등, 1985)이 관련이 있다. 하지만, 보조 문항이 문맥 정보를 접근 불가하게 만들만큼 충분치 않은 경우(충분한 경우 응답자는 이 정보를 어떻게 사용할 지 결정할 필요가 없다), 변수들 중 어느 것도 보조 문항 삽입의 영향을 받지 않을 것이다.

마지막으로, 질문 파악에 관한 항목에서 살펴 본 바와 같이 대화 규범들은 응답자들이 접근가능성이 높은 정보를 가지고 할 수 있는 일에 대한 그들의 인식을 결정할 것이다(예를 들면, Schwarz 등, 1991; Strack 등, 1988). 이 맥락에서 논한 바와 같이, 응답자는 중복을 피하려고 하기 때문에 자신이 이미 제공한 정보를 또 다시 제공할 가능성은 낮다. 따라서 그들은 고려 대상에서 이미 제공한 정보를 제외시킨다. 그러나 앞서 언급한 바와 같이 관련 없는 보조 문항들을 삽입하면 질문의 대화적 관련성을 인식하는 정도를 저하시킬 수 있다. 그리고 반대의 경우는 중복 정보를 제외하지 않을 수도 있다. 게다가 앞서 논한 바와 같이, 실질적으로 관련된 보조 문항들을 삽입하면(예: Schwarz 등, 1991), 대화 과정의 본질이 변할 수 있다.

이러한 결정 *중 어느 하나라도* 이미 형성되어 있는 대상의 심상에서 정보를 제외시키는 경우에는 언제나 대조 효과가 발생한다. 대조 효과가 대상에만 국한되는지 아니면 관련 대상 전체에 일반화되는지 여부는 제외된 정보가 단지 대상의 심상에서만 제거되었는지 아니면 기준이나 척도 양극점의 구성에 사용되었는지의 여부에 따라 결정된다. 한편 머리 속에 떠오른 정보가 이미 형성되어 있는 대상의 심상에 포함될 때에는 언제나 동화 효과가 발생한다. 따라서 이 모형은 태도 측정에 있어서 문맥 효과의 발생, 방향, 정도 및 일반화를 예측한다. 이외에도 이 모형은 앞서 설명한 바와 같이 접근가능 정보의 사용에 미치는 보조 문항의 효과를 구체적으로 예측한다. 이러한 예측 사항들은 향후 유망한 실험 대상으로 가치가 있다.

3.2.2.3 회상 경험

선행 상황에서는 머리 속에 떠오른 정보에 주안점을 두고 응답자가 이것을 어떻게 사용하는가에 따라 정보의 효과가 좌우된다는 것을 강조하였다. 하지만, 응답자의 판단은 그가 기억한 내용에 좌우되는 것이 아니다. 오히려 내용의 함축적 의미는 머리 속에 떠올리는 것이 쉬운지 어려운지에 따라 다를 수 있다(이를 검

토하려면 Clore(1992) 참조). 한 예로서 Bishop(1987)이 검토한 일련의 조사들을 고찰해 보자. Bishop과 그의 동료들은 응답자에게 그들이 "정치적이고 공적인 사건에 얼마나 많은 관심을 갖고 있는지" 질문하였다. 그리고 질문의 앞이나 뒤에 응답자 지역의 미하원의원이 지역구를 위해 어떤 일을 했는지의 대답을 요청하는 정치적 지식에 관한 질문을 두었는데, 이 질문은 많은 응답자들이 응답할 수 없었던 것이었다. 응답자들은 정치적 관심을 묻는 질문의 뒤에 정치적 지식에 관한 질문을 두었을 때보다 앞부분에 두었을 때, 정치문제들에 관심을 덜 갖고 있다고 응답하였다. 응답자들은 정치 지식에 관한 질문에 응답할 수 없었다는 사실로부터 자신들이 정치 문제들에 많은 관심을 갖지 않았다는 결론을 냈거나 또는 관심을 가졌더라면 자신의 지역구 의원이 어떤 일을 했는지를 알았을 것이라는 결론을 내린 것으로 추측된다.

Bishop(1987)은 일련의 흥미 있는 조사들을 통해 보조 문항이 질문 순서의 효과에 미치는 영향을 탐구하였다. 세 번에 걸친 조사에서 관련 없는 보조 문항의 수는 각각 33개, 40개 그리고 101개를 사용했고, 조사 시간은 7분에서 17분까지였다. 보조 문항들의 수가 이례적으로 많았음에도 불구하고, 대부분 연구자들의 예상과는 반대로 문맥 효과는 사실상 변하지 않았다.

이것을 어떻게 설명해야 하는가? 응답자들은 자신이 얼마나 정치에 관심을 갖고 있는가에 대해서 가장 접근가능성이 높은 관련 정보의 의미에 근거하여 판단한 것으로 추정된다. 보조 문항의 대부분이 지역사회 문제나 케이블 TV에 관련된 내용으로 대상 질문과는 무관하다는 점을 감안할 때, 이 판단과 관련된 가장 접근가능성이 높은 정보는 응답자가 자신의 지역구 의원의 활동에 대해 거의 아는 바가 없다는 경험이었다. 따라서, Bishop(1987, p. 182)은 "응답자가 면접 중이나 그 이후에 자기 지각(self-perception)에 변화를 일으키는 경험을 할 때까지 이 효과는 지속될 것"이라고 결론지었다"

자기 지각에 의한 설명은 암시적 태도를 전제로 한다는 것을 주목해야 한다. 즉, 응답자들은 하원의원에 대한 본인의 지식이 부족한 것은 자신의 행동, 다시 말하면 그들이 공적인 문제들에 관심을 갖지 않았기 때문이라고 생각해야 한다. 하지만, 이와는 달리 그들은 하원의원이 자신의 활동 결과를 응답자게 알리는 일을 제대로 하지 않았다고 생각할 수도 있다. 그렇다면 101개의 관련 없는 문항들에서는 얻을 수 없었던 결과를 이런 가능성이 있는 하나의 보조 문항에서 얻을 수 있을지 모른다. 이러한 가능성을 검증하기 위해 Schwarz와 Schuman(1995)은 Bishop(1987)의 조사를 반복하고, 응답자가 자신의 하원의원이 지역 활동 결과를 얼마나 잘 알리고 있는가를 평가하는 보조 문항을 삽입하였다. Bishop의 연구와

마찬가지로 의원 활동에 관한 지식을 묻는 질문은 응답자의 정치적 관심에 대한 자기보고(self-reported) 를 18.5% 감소시켰다. 하지만, 하원의원의 지역 활동 보고에 관한 보조 문항 하나를 삽입했을 때는 이 효과가 약 8.8%로 약화되었다. 이러한 감소는 위에서 사용한 101개의 보조 문항들이 아무런 효과가 없었다는 사실처럼, 응답자들이 지식에 관한 질문에 응답하는 데 어려움을 겪었다는 것은 인지적 접근 가능성이 감소하였기 때문은 아니다. 그보다 지역 활동 보고에 관한 질문에서 응답자들의 지식 부족은 자신들이 정치 문제에 관심이 없기 때문이 아니라 그들의 하원의원이 지역 활동 보고를 제대로 수행하지 않았기 때문이라고 암시함으로써 응답자 자신이 인식한 어려운 경험의 의미들을 변경하였다.

이처럼 응답자 자신의 지식 부족에 대한 원인을 엉뚱한 곳에서 찾도록 유도하는 것은 응답자가 자기 지각을 변경하는 유일한 방법이다. Martin과 Harlow(1992)는 응답자가 정치적 지식에 관한 질문의 응답에 실패하는 것은 자신의 자아 개념(self-concept)에 손상을 가져온다고 주장하였다. 그들은 이 주장을 검증하기 위해 Bishop(1987)의 조사를 확대 반복하여 서로 다른 보조 질문들을 삽입하였다. 한 질문 세트에서는 응답자가 정치적 태도를 표명함으로써 자신의 자아 개념을 확고히 할 수 있게 한 반면(cf. Steels, 1988), 다른 질문 세트에서는 비정치적인 문제들을 다루었다. 예상했던 대로, 두 개의 정치적 태도에 관한 질문의 응답은 정치적 지식에 관한 질문이 갖는 효과를 감소시켰고 반면에 비정치적 질문들은 아무 영향을 미치지 않았다. 하지만 이것으로 응답자가 정치적 태도에 관한 보조 질문을 통해 자신의 자아 개념을 확고하게 할 수 있었던 것인지 또는 질문에 응답하는 도중 최소한 몇 가지 사항에 관하여는 정치에 관심을 둔 적이 있다는 정보를 떠올린 것인지 분명하게 알 수는 없다.

사례들을 종합해 보면, 여러 심리적 과정들이 질문 순서 효과의 발생 및 제거에 영향을 미칠 수 있다는 것을 알 수 있다. Bishop(1987)의 101개 보조 문항들, Martin & Harlow(1992)의 두 개의 보조 문항 그리고 Schwarz & Schuman(1995)의 한 개 보조 문항이 각기 다른 효과를 낸다는 사실은 보조 문항의 효율성이 이들의 개수와 상관이 있는 것이 아니라 주로 이들의 구체적인 내용에 좌우된다는 것을 명백하게 나타내는 것이다.

3.2.3 형식 구성

응답자는 일단 판단을 내린 후 연구자가 제시한 응답 선택 문항들에 맞게 판단의 형식을 구성해야 한다. 한 응답자가 긍정적인 판단을 내렸다고 가정하자. 그가 이것을 표시하기 위해 7점 평정 척도에서 7을 선택할 것인가, 아니면 6이나 5

를 선택할 것인가? 평정 척도의 눈금을 어떤 방식으로든 정하지 않으면, 이 수치들은 무의미하다. 정신 물리학(psychophysics)(이를 검토하려면 Poulton(1989) 참조)과 태도 측정(이를 검토하려면 Strack(1994b) 참조) 분야의 수많은 연구들이 증명하는 바와 같이, 응답자들은 현재 진행 중인 연구자와의 대화 맥락에서 척도의 눈금을 정하기 위해, 보통 머리 속에 떠오른 가장 극단적인 자극을 척도의 양극점들로 정한다. 예를 들어, 1파운드의 무게는 더 무거운 무게들이 주어진 맥락보다 더 가벼운 무게들이 주어진 맥락에서 제시되었을 때 더 무겁다고 평가된다.

이러한 "관점"(Ostrom & Upshaw, 1968)이나 "범위"효과(Parducci, 1982, 1983)는 맥락 자극을 준거틀로 사용하고 이 자극 중 가장 극단적인 자극으로 척도의 양극점을 정하는 것을 말한다. 하지만, 응답자들은 이 극단적 자극이 무엇인지 알고 있을 때에만 제시된 자극 연속의 양극단치들로 척도의 양극점들을 정할 수 있다. 한 번에 하나의 자극만 제시하는 경우, 응답자들은 기억을 더듬어 척도의 양극점을 정하므로, 판단 시에 어떤 관점이든지 사용할 수 있는데, 이것은 부분적으로 선행 질문들의 내용과 상관이 있다. 하지만, 아래의 사례에서 볼 수 있는 바와 같이, 극단적 자극이 이 관점에 속하기 위해서는 현재의 응답수행과 관련이 있어야 한다. 정신 물리학의 고전적인 한 연구에서, Brown(1953)은 피험자에게 여러 물체를 들어올리고 그 무게를 평정하도록 했다. 역시, 주어진 물체의 무게는 먼저 더 무거운 물체를 들었을 때보다 더 가벼운 물체를 들었을 때에 더 무거운 것으로 평가되었다. 하지만 이것은 무거운 물체가 평가할 물체의 일부로 인식된 경우에만 이런 결과가 나타났다. 이와 대조적으로, 피험자에게 실험자 옆에 있는 동일 무게의 쟁반을 들어올리라고 요청한 경우에, 대상 물체에 대한 그들의 판단은 변하지 않았다. 분명히 이 쟁반은 자극 물체의 일부로 인식되지 않았기 때문에, 피험자가 형성한 관련 "관심"의 심상에 포함되지 않았고, 따라서 평정 척도의 양극점을 정하는 데 사용되지 않았다.

다시 말하지만, 선행 질문의 내용은 어떤 정보가 머리 속에 떠오를지 여부에 영향을 줄 수도 있으므로 척도의 양극점을 정하는 데 사용된다. 보조 문항들은 이전에사용한 정보의 인지적 접근가능성을 감소시키거나 정보가 현재 응답수행에 관련이 있다는 인식을 줄이는 것으로 영향을 줄 수 있다. 하지만, 많은 경우 관찰된 대조 효과가 단지 평정척도를 다르게 사용해서 발생하였는지 아니면 이미 형성되어 있는 심상의 변화로 나타났는지 여부를 판단하기는 어렵다. 해석이 거의 필요하지 않은 자극의 경우(무게의 경우와 같이), 발생하는 모든 대조 효과는 응답 언어의 변화가 원인일 가능성이 높다. 반대로, 더 많은 해석을 요하는 자극일 경우 평정척도를 다르게 사용하는 것뿐만 아니라 이미 형성되어 있는 심상의 변화도 문맥 효과 발생에 기여할 수 있다(예를 들면, Lynch 등, 1991; 자세한 설명은 Strack(1994b) 참조)

3.2.4 편집

예를 들어, 응답자의 종교나 도덕적 가치에 관한 선행 질문은 다음 질문에서 자기 표상의 유지에 관한 우려를 증가시키거나 응답자들이 사회적 바람직성에 대한 응답에 영향을 줄 수 있는 정보가 머리 속에 떠오르게 할 수 있다(DeMaio, 1984). 다시 말하지만, 보조 문항이 관련 정보의 인지적 접근 가능성을 감소시키는 경우에는 이런 영향들을 감소시킬 수 있다. 하지만, 현재 우리가 아는 바로는 편집 단계에서 보조 문항이 미치는 영향을 다룬 연구는 없다.

3.3 결론

아마도 지금까지의 검토내용에서 도출할 수 있는 가장 명백한 결론은 우리가 보조 문항의 작용에 대해서 거의 아는 것이 없다는 것이다. 대부분의 연구자들이 질문지에서 질문을 배치하는 방식에 세심한 주의를 기울였지만, 이에 대한 결정은 주로 직관이나 상식에 근거한 것이었다. 보조 문항의 작용에 관한 체계적인 실험은 드물다. 게다가 조사 연구 문헌에서 보고한 실험들 중 다수는 해석하기가 어려운데, 이것은 이 실험들이 문맥 효과의 기저를 형성하는 인지적 및 의사소통 과정들에 관한 구체적인 가정들을 검증하기 위해 설계된 것이 아니기 때문이다. 그 결과, 일관성 없는 증거 자료들로부터 문항들의 배치에 관해서 이론적으로 잘 구성된 지침들을 도출하는 것은 어렵다. 앞서 언급한 바와 같이, 일부 연구들은 보조 문항과 질문 순서 효과의 감소 간의 상관을 규명하지 못한 반면(예를 들면, Schuman 등, 1983; Bishop 등, 1984, 실험. 2; Bishop, 1987), 다른 일부 연구들은 유의하지 않은 경향을 발견하였다(예를 들면, Tourangeau 등, 1989a). 심지어 가장 성공적으로 보조 문항을 적용한 사례에서도 문맥 효과의 제거를 기대했지만 단순히 약화시키기만 하였다(예를 들면, Bishop 등, 1984, 실험. 1; Tourangeau 등, 1989b). 마지막으로, 다른 실험에서는 보조 문항으로 대조 효과를 얻을 수 있었던 경우들이 오히려 동화 효과로 변하여 문맥 효과를 제거하기보다는 반전시키기도 했다(예를 들면, Ottati 등, 1989). 이러한 일관성 없는 연구 결과들은 여러 문맥 효과들이 각기 다른 인지적 및 의사소통 과정들에 기인한 것이기 때문에 동일한 메커니즘에 의해 약화될 가능성은 낮다는 점을 고려하면 당연하다. 일을 더욱 복잡하게 하는 것은 보조 문항 자체는 이전에사용한 정보의 접근 가능성에 영향을 줄 뿐만 아니라, 나중에 이 정보가 머리 속에 떠올랐을 때 응답자가 이것을 어떻게 처리할 것인가를 결정하는 데에도 영향을 줄 수 있다는 것이다.

일반적으로 정보의 접근 가능성은 이를 마지막으로 사용한 시간이 지남에 따라 감소한다(이를 검토하려면 Wyer와 Srull(1989) 참조). 정보의 접근 가능성이

감소하는 속도는 정보 자체의 특성(독특하거나 기억에 선명한 정도)뿐만 아니라, 중간에 삽입된 질문에 의해 활성화된 정보의 특성에 좌우된다. 예를 들면, 정치 문제들에 관한 두 가지 관련 질문 사이에 소비재에 관한 20개의 질문들을 삽입한 경우 정치에 관한 선행 질문은 중간에 삽입된 질문도 정치 문제와 관련되어 문맥 질문과 대상 질문 간의 관계가 불분명한 경우보다 머리 속에 떠오를 가능성이 더 높다. 또한 두 질문이 내용 및 어구 구성 면에서 매우 유사하여 대상 질문이 문맥 질문을 기억해 내는 데 아주 좋은 힌트가 된 경우에는 문맥 질문과 대상 질문 사이에 보조 문항을 삽입하는 것도 아무 소용이 없을지도 모른다.

보조 문항은 문맥 정보의 접근 가능성을 감소시켜 문맥 효과를 약화시킬 뿐만 아니라, 대상 질문과 관련 있을지도 모르는 정보가 머리 속에 떠오르게 함으로써 초기의 문맥 질문의 효과를 약화시킬 수도 있다. 이 경우 "본래의" 문맥 효과는 보조 문항 그 자체에 의해 야기된 다음 문맥 효과들에 의해 약화될 수 있다. 이런 가능성이 있기 때문에 실질적으로 관련 있는 보조 문항들은 바람직하지 않은데, 보조 문항이 초기의 문맥 질문 효과를 약화시키는 데 가장 효율적이더라도 마찬가지다.

일을 더욱 복잡하게 만드는 것은 응답자들이 문맥 질문에 의해 접근가능하게 된 정보를 어떻게 사용할 것인지 결정하는 데에 보조 문항이 영향을 줄 수 있다는 것인데, 이런 가능성은 이제까지 조사 문헌에서 다룬 바가 없었다. 보조 문항이 이러한 영향을 미치는 예를 보면, 특정 정보가 응답자에게 기억이 나는 이유를 이해하는 데에 영향을 주거나 대화 규범의 적용에 영향을 미치거나 인식한 접근가능 정보의 유용성 진단에 영향을 주는 것 등이 있다. 우리는 위에서 이 예들을 하나씩 좀 자세하게 다루고 요약하였다.

우리는 최대 101개의 무관 보조 문항들도 응답자 자신이 보고한 정치적 관심에 대해서 하원 의원에 관한 지식을 묻는 질문에 대한 응답 실패가 미치는 효과를 제거하는 데 불충분하였다는 것을 증명한 Bishop (1987)의 연구 결과를 가지고 보조 문항이 인식한 접근가능 정보의 진단에 미치는 효과를 설명하였다. 대조적으로, 하원의원의 지역 활동 보고의 질에 관한 한 가지 질문을 사용한 경우, 이 질문이 없었을 때는 관찰되었던 문맥 효과가 유의하지 않은 수준으로 감소되었다 (Schwarz와 Schuman, 1995). 이것은 하나의 보조 문항이 당시의 판단을 위해 응답자의 하원 의원에 대한 지식 부족에 대한 사실을 고려하지 못하게 만들었음을 나타낸다. 따라서, 보조 문항은 문맥 정보의 접근 가능성을 감소시켜 문맥 효과를 제거할 수 있을 뿐만 아니라, 인식한 문맥 정보의 함축적 의미를 변경함으로써 효과를 제거할 수도 있다.

이외에도 보조 문항들을 삽입하는 것은 중복 금지에 관한 대화 규범의 적용에 영향을 미칠 수 있다. 예를 들면, KKK단의 언론의 자유에 관한 질문을 한 후 언론의 자유에 관한 다음 질문을 일반적인 것으로 하게 되면 응답자는 이것을 다른 화자에 관한 것으로 해석한다. 구체적인 질문과 일반적 질문 사이에 몇 개의 무관한 보조 문항들을 삽입하여 질문 간의 대화적 관련성을 인식하지 못하게 한 경우에는 이런 해석을 하지 않는다. 그 결과, Ottati 등(1989)의 조사에서 KKK단에 대한 질문은 보조 문항이 없는 경우에는 대조 효과를 일으켰지만 보조 문항이 있는 경우에는 동화 효과를 일으켰다. 따라서 *무관한* 보조 문항들은 일반적 질문 및 구체적인 선행 질문이 동일한 대화 문맥에 속하는 것으로 인식할 가능성을 줄임으로써 일반적 질문을 다르게 해석하게 할 수 있다.

마찬가지로 실질적으로 *관련된* 보조 문항들은 예를 들면, 공통 소개문에 의해 모든 질문들을 명시적으로 동일한 대화 문맥에 배치한 조건 하에서도 질문을 다르게 해석하게 할 수 있다. 따라서, Schwarz 등(1991)은 응답자들이 결혼생활 만족도에 관한 하나의 구체적인 질문 다음에 일반적 삶의 만족도에 관한 질문을 제시하는 경우 이것을 새로운 정보를 요청하는 것으로 해석하였지만, 이 질문의 앞에 여러 삶의 영역들에 대한 만족에 관한 몇 개의 구체적인 질문들을 둔 경우에는 종합적 요약 판단을 요청하는 것으로 해석한 사실을 관찰하였다. 이런 다양한 해석은 하나의 구체적인 질문만 제시한 경우에는 부분-전체 대조 효과를, 몇 개의 구체적인 질문들을 제시한 경우에는 부분-전체 동화 효과를 이끌어 냈다. 보조 문항들이 실질적으로 무관한 경우에는 이처럼 특별히 질문을 다양하게 해석할 것이라 예상하지 않으므로 일반적 질문을 "종합하는 것"으로 해석하지 않게 된다. 이런 예들이 설명하는 바와 같이, 보조 문항들의 개수와 내용은 부분-전체의 순서로 구성된 질문들에서 일반적 질문의 해석에 영향을 주어 응답자들이 접근가능 정보를 사용하는 방식을 결정하게 한다.

마지막으로, 보조 문항은 응답자들이 어떤 정보에 대해서 선행 질문에서 이미 다루었기 때문에 머리 속에 떠올랐을 것이라 생각한 경우에는 어떤 영향을 미칠 수 있다. 일반적으로 사람들은 "엉뚱한"원인에 의해 머리 속에 떠올랐다고 생각되는 정보는 고려 대상에서 제외하기 때문에 대조 효과가 발생한다. 보조 문항들의 개수는 문맥 정보가 미치는 영향에 대한 응답자의 인식을 감소시키기에는 충분하지만 증가된 문맥 정보의 접근 가능성을 제거하기에 불충분한 경우에는, 대조효과를 얻을 수 있었던 것을 동화 효과로 반전시킬지도 모른다. 선행 질문으로 인해 머리 속에 떠오른 정보를 의도적으로 사용하지 않는 결과는 응답자 자신이 선행질문에 대해서 인식한 효과를 바람직하지 않게 생각하는 여부에 좌우된다.

질문 파악에 관한 5.2.1항에서 더 자세히 논한 바와 같이, 현재 이러한 인식을 결정하는 조건들은 잘못 이해되고 있다.

요약하면 보조 문항은 접근 가능성에 관한 연구들에서 주장하는 것보다 훨씬 복잡한 질문지의 한 요소이다. 보조 문항은 이전에 사용한 정보의 접근 가능성을 감소시킬 뿐만 아니라, 응답자가 머리 속에 떠오른 정보를 어떻게 처리할 것인가를 결정하는 데에도 영향을 준다. 그 결과 보조 문항은 동화 효과와 대조 효과를 약화시킬 수 있고, 야기할 수도 있다. 향후 연구들에서는 본 장에서 제시한 가설들에 대해 체계적인 실험을 통해 검증함으로써 궁극적으로는 문맥 효과 및 보조 문항들의 작용의 원리들을 보다 더 잘 이해하는 데 기여하기를 바란다.

제 4 절 조사에 있어 효과적인 측정을 위한 평정 척도의 설계

Jon A. Krosnick
The Ohio State University
Leandre R. Fabrigar
Queen's University at Kingston

4.1 머리말

태도나 신념과 같은 주관적인 현상들을 측정하는 현대 조사 연구에서 평정 척도는 필수적이다. "척도"라는 용어는 종종 하나의 개념을 측정하는 일단의 문항들을 지칭하는 데 사용하지만, 본 절에서 우리는 "평정 척도"라는 용어를 평정 형식(rating format)을 사용하는 하나의 질문 상에 있는 응답 선택 문항들을 지칭하는 데 사용한다. Thurstone과 Chave(1929), Likert(1932)와 이들과 동시대의 학자들의 선구자적인 연구로부터 시작하여 일련의 많은 연구들이 이 목적을 위한 평정 척도의 효용성을 입증하였다. 평정 척도는 그 자체만으로도 심리학자, 사회학자, 정치학자, 경제학자 및 많은 다른 사회과학자들의 경험적 연구에서 중추적인 역할을 하고 있다. 하지만, 놀랍게도 조사 방법 및 질문지 설계에 관한 교재들은 특정 목적에 대한 평정 척도를 구성하는 가장 좋은 방법에 관한 구체적이고 실용적인 지침들을 제시하는 일이 드물다.

평정 척도를 설계하는 데에는 여러 가지 많은 결정들을 내릴 필요가 있으므로 이러한 지침들은 유용할 수 있다. 예를 들면, 척도의 길이가 어느 정도여야 하며 이것이 중간점을 포함해야 하는지 여부를 결정해야 하고, 모든 점에 문자명칭을 쓸 것인지 일부는 숫자명칭을 쓸 것인지 여부도 결정해야 한다. 문자명칭을 사용하기로 선택한 경우, 정확하게 어떤 단어를 사용할 것인지를 결정해야 한다. 숫자명칭을 사용하는 경우, 의견없음(no-opinion) 응답선택문항을 포함시킬 것인지 생략할 것인지도 고려해야 한다. 대안적 접근방법이 없기 때문에 대부분의 질문지 설계자들은 자신의 직관에 의존하여 이런 결정들을 내리므로, 때로 조사자마다 아주 다른 절차들을 사용하게 된다.

하지만, 대부분의 조사방법론자들은 여러 사회 과학 분야들에서 80년간에 걸쳐 산발적으로 축적된, 이러한 설계에 관한 결정을 내리는 데 있어서 과학적 기초를 제공하는 매우 많은 경험적 연구 문헌들이 있다는 것을 알지 못한다. 현재, 이 문헌들을 종합하는 많은 작업과 질문지 설계자들이 관심을 갖는 많은 관련 주제들에 대한 연구들이 진행되고 있다(예를 들면, 곧 출간될 Krosnick와 Fabrigar

참조). 본 절에서 우리는 이 책에서 밝히고 있는 사항들의 일부를 미리 살펴본다. 구체적으로 말하면, 우리는 척도점들의 개수, 문자 명칭이나 숫자 명칭 및 의견없음 선택 문항의 포함과 같은 평정 척도 설계의 세 가지 주요 관심사들을 중점적으로 다룬다. 우리는 먼저 평정 척도 설계의 질을 평가하는 데 사용하는 기준인 신뢰도와 타당도에 관하여 간단히 설명하고, 그 다음에 상기한 세 가지 기본 주제들 각각에 관한 문헌들 중 일부를 검토한다.

4.2 자료의 질 평가

사회 과학에서 조사 방법에 관한 교재들은 보통 척도의 질에 관한 두 가지 기본 요소인 신뢰도와 타당도를 다룬다. 본 절 이외에도 이 문제들을 상세하게 논하는 우수한 책들이 많이 출간되었기 때문에, 여기에서 이 기준들을 정의하거나 정교하게 구성할 필요가 없다. 하지만, 모순처럼 보일 수 있는 증거를 이해하기 위해 조사 문항들의 신뢰도와 타당도의 평가 방법에 대한 내재적인 장단점들을 언급할 필요가 있다.

4.2.1 신뢰도

태도를 측정하는 조사를 할 때 우리는 태도를 개인이 마음에 두고 있는 잠재적 구성개념이라고 가정한다. 하나의 질문으로 이러한 구성개념을 완전히 알아낼 수 없는데, 어떤 질문에 대한 응답들은 태도 그 자체라기보다는 이를 포함한 다양한 외생적 요인들과 상관이 있기 때문이다. 그러나 구성개념은 개인의 마음속에 있기 때문에, 질문은 다른 요인들이 영향을 미치는 가운데에서도 이것을 측정할 수 있는 한 수단이 된다. 한 질문의 신뢰도는 동일한 사람에게 수회에 걸쳐 동일한 질문을 하여 얻은 결과들의 일관성(종단 신뢰도)이나 한 사람의 동일한 태도를 알아내기 위해 한 번에 여러 질문들을 함으로써 얻은 결과들의 일관성(횡단 신뢰도)으로 그 측정치의 일관성을 평가함으로써 판단할 수 있다.

이 방법들은 각기 결점을 갖고 있다. 시간을 두고 얻은 한 문항에 대한 응답들이 일관성이 없는 것은 두 가지 원인에 의해 발생한 것으로 관찰할 수 있는데, 즉 척도가 신뢰성이 없을 수도 있고, 태도 자체가 조사 시점들 사이에 변화했을 수도 있다. 이 조사 시점들 사이의 시간이 짧을수록 태도가 변화할 가능성이 더 낮아지지만, 응답자는 첫 번째 조사 시의 응답 내용을 회고하고 이를 단순히 되풀이하여 응답함으로써 명목상으로만 인위적인 신뢰도를 높일 가능성이 높다.

한 번의 조사에서 동일한 태도를 측정하려고 제시하게 되는 일련의 질문들에 대해서 응답자는 그 의도를 인식하고 문항들 전체에 걸쳐 일관성이 있는 것처럼 응답할지도 모른다. 또한 연구자가 의도적으로 완전히 동일한 문제에 대한 일련의 질문들을 할 리가 없다고 생각하여, 질문들 간의 세세한 차이를 추론하고 그 차이를 과장해서 해석할 수도 있다. 이런 결점들을 감안하여 각 방법에 의한 조사가 완전한 결과를 산출할 것이라고 기대해서는 안 된다. 오히려 여러 가지 방법들을 사용하여 얻은 조사 결과들에서 어떤 경향을 찾아야 한다.

4.2.2 타당도

척도의 타당도란 관심을 두고 있는 구성개념을 알아내는 정확도(accuracy)를 말한다. 신뢰성이 전혀 없는 척도는 결코 타당하지 않다는 점에서 신뢰도와 타당도는 서로 인과관계가 있다. 하지만, 신뢰도가 높은 척도도 상당히 부정확할 수 있다. 이 경우 이 척도의 일관성은 정확성 대신에 근본적인 무감각(insensitivity)이나 관심을 두고 있는 구성개념이 아닌 다른 개념을 찾는 것에 기인할 수 있다.

조사 척도의 타당도는 여러 가지 방법으로 평가할 수 있다. 예를 들어, 소위 상관타당도(correlational validity)는 주어진 척도가 이것과 관련이 있는 다른 변수들을 예측할 수 있는 정도이다. 더 높은 관련성은 더 높은 타당도를 나타낸다. 예를 들어, 하나의 태도 척도가 다른 척도보다 태도 관련 행위를 더 잘 예측하는 경우, 전자가 후자보다 더 타당도가 높은 것으로 간주할 수 있다. 판별 타당도(discriminant validity)는 하나의 측정 접근방법으로 서로 구별되는 것으로 추정되는 개념들을 구별할 수 있는 정도를 말한다. 따라서 평정 척도가 서로 다른 객체들이 갖는 인식의 차이를 더 잘 감지할 수 있는 경우 타당도는 높다고 할 수 있다. 물론 타당도 평가를 위한 접근방법의 유용성은 척도에 의해 예측되는 기준들을 정확하게 식별하고 서로 다르게 평가해야 할 사물들을 식별할 수 있는 연구자의 능력에 좌우된다. 여기에서도 타당도는 설계 측면에서 서로 이질적인 많은 조사 결과의 경향을 찾는데 매우 유용하다.

4.3 척도점의 수

평정 척도를 구성할 때 결정해야할 한 가지 중요한 것은 포함시킬 척도점의 수에 대한 것인데, 여기에는 고려해야 할 몇 가지 이론적 문제들이 있다. 양극 척도(bipolar scale)(즉, 명확한 개념적 중간점을 가지면서 두 개의 대립되는 선택사항들을 나타내는 척도)와 단극 척도(unipolar scale)(즉, 개념적 중간점이 없고 한쪽 끝이 0점인 어떤 개념의 다양한 수준을 나타내는 척도)를 구별하는 것이 중요하다. 태도는 그 범위가 가장 긍정적인 것에서 가장 부정적인 것까지 이르고

긍정도 부정도 아닌 중립으로 특정의 중간점을 갖기 때문에 양극 개념(bipolar construct)이라고 생각할 수 있다. 어떤 사람이 자신이 갖고 있는 특정 태도에 부여하는 중요도는 단극 개념의 예이다. 이것은 0의 중요도에서 최대 수준까지의 범위를 가지며 정확한 중간점이 없다.

일반적으로 척도점이 많은 것이 적은 경우보다 더 효과적으로 생각되는 몇 가지 이유들이 있다. 사람들 스스로 자신의 태도에 대해 갖는 인식이 극단적인 긍정에서 극단적인 부정까지 하나의 연속선상에 있는 것으로 생각하는 것이다. 이 연속선상의 한 점을 범주형 응답 척도상의 한 점으로 변환하려면 일련의 점이 연속선 전체를 나타내야 한다. 세 가지 선택 사항(예를 들어, 찬성, 반대 그리고 보통)만 갖고 있는 척도에서 사람들은 약간만 찬성한다고 말할 수 없다. 약간만 찬성하는 사람이 "찬성"을 선택(실제보다 더 긍정적인 것을 선택)해야 할지, "찬성도 반대도 아님"을 선택해야 할지 불명확하다. 더 많은 점들을 가진 정교한 척도에서는 온건한 성향을 가진 사람들이 자신의 입장을 정확하고 자신 있게 표현할 수 있을 것이다. 게다가 척도점이 많을수록 하나의 사물에 대한 자신의 태도와 다른 사물에 대한 태도를 더욱 세밀하게 구별할 수 있다. 연구자도 마찬가지로 척도점이 많을수록 동일 사물에 대한 사람들의 태도들을 더욱 세밀하게 구별할 수 있다. 따라서 척도점이 많은 척도는 더 유용한 정보를 전달할 수 있다.

반면에, 너무 많은 척도점을 사용하면 응답 선택 문항들의 의미에 대한 명확성이 떨어질 수 있다. 몇 가지 선택 문항들만 갖는 척도(예를 들면, 찬성, 반대 그리고 보통의 3점 척도)의 경우, 이 선택 문항의 의미는 매우 명확하다. 그러나 선택 문항의 수가 매우 많은 경우(예를 들면 101점 온도계 척도), 이들 각 점의 의미는 덜 명확하다. 이 경우 응답자들이 특정 응답 선택 문항에 부여한 의미들에 관하여 한 개인의 경우나 개인 간에도 일관성이 감소되는 결과를 낳을 수 있다. 이외에도 너무 많은 응답 선택 문항을 포함시키면 응답자 자신이 그 중 어디에 해당하는지 결정하는 수고를 덜기위해 만족을 줄만한(via satisficing)쪽으로 선택할 수도 있다(Krosnick, 1991).

따라서 최적의 척도 길이는 중간 길이인 것 같다. 그러나 중간 길이는 정확히 어느 정도인가? 양극 개념의 경우에는 척도란 사람들이 자연스럽게 구별하는 차이점들을 나타내는 인지적 범주를 표시해야 하는 것으로 생각할 수 있다. 예를 들면, 태도에 있어서 사람들이 특정 사물을 좋아하는 경우 약간, 중간 정도 또는 상당히 좋아할 수 있다. 태도 척도의 중간 좌측에 세 가지 점들을 생각하고 우측에도 세 가지 점을 생각한다면, 7점 척도가 최적인 것 같다. 척도의 한 쪽이나 반대쪽으로 약간 또는 상당한 정도로 구분하는 것이 자연스럽다면 5점 척도가 최적인 경우도 있다.

단극 척도의 경우 너무 적은 척도점을 사용하면 수집한 정보의 질이 떨어질 수 있고, 너무 많은 척도점을 사용하면 의미의 명확함이 떨어질 수 있다. 그러나 최적의 길이가 어느 정도인지 예상하는 것은 쉽지 않다. 사람들은 단극 연속선상의 0, 약간의, 보통의 그리고 아주 많은 같은 것을 쉽게 떠올리는 듯하다. 아마 이보다 약간 더 세밀하게 구분을 해도 편하게 느껴져, 최적의 척도 길이는 4-7점 정도라고 예상된다.

4.3.1 신뢰도

척도점의 수가 측정의 질에 미치는 효과에 대한 경험적 연구들의 대부분은 이것이 신뢰도에 미치는 영향을 연구하였다. 이 연구들에서는 기존 자료의 이차 분석 및 직접 실험 비교를 포함하여 다양한 접근법들을 사용하였다.

일부 연구들은 양극 척도상의 척도점의 수와 횡단 신뢰도 간의 관계를 발견하지 못했지만(Bendig, 1954a), 대부분의 연구들은 대략 7개의 점들을 가진 척도의 경우 횡단 신뢰도가 가장 높았다는 것을 밝혀내었다. 예를 들면, Masters(1974)는 2, 3, 4, 5, 6 및 7점 척도들을 실험적으로 비교하여 신뢰도가 4점까지 증가하다가 그 다음부터는 일정한 수준을 유지하였음을 발견하였다. Birkett(1986)는 2, 6 및 14점 척도를 실험적으로 비교하여 6점 척도에서 가장 신뢰도가 높다는 것을 알아냈다. 마찬가지로 Komorita & Graham(1965)은 6점 척도가 2점 척도보다 신뢰도가 더 높다는 것을 입증하였다. 종단 신뢰도를 검증한 한 연구에서는, 7점-9점 척도가 이보다 더 짧은 척도보다 신뢰도가 더 높은 것으로 나타났다(Alwin과 Krosnick, 1991).

단극 척도의 횡단 신뢰도에 관한 연구들은 최적의 척도점 수가 5점과 7점 사이라고 주장하였다. 일부 연구들은 척도점의 수와 신뢰도 간의 체계적인 관계를 찾는 데 실패하였지만(Bendig, 1953; Matell과 Jacoby, 1971; Peterson, 1985), 다른 연구들은 2점 척도보다는 4점 척도(Watson, 1988), 7점이나 11점 척도보다는 5점 척도(McKelvie, 1978), 3점이나 9점 척도보다는 5점-7점 척도의 횡단 신뢰도가 더 높은 것을 알아내었다. 종단 신뢰도에 관한 연구(Jacoby와 Matell, 1971; Matell과 Jacoby, 1971)에서는 7점 및 8점 척도가 3점-6점 척도나 9점-19점 척도보다 신뢰도가 더 높았다.

4.3.2 타당도

많은 연구들은 척도점 수가 측정의 타당도에 미치는 영향을 검증하였다. 이런 연구들의 대부분은 자료를 관계의 연속적인 표시로부터 불연속적인 척도점 표시로 변경하는 것이 이미 알려진 자료형태를 어떻게 왜곡시키는가를 검증하기 위해 컴퓨터 시뮬레이션들을 사용하였다. 몇 가지 예외가 있지만(Martin, 1973, 1978), 이 시뮬레이션들은 척도점의 수가 증가함에 따라 자료 왜곡이 감소하지만 이러한 자료 왜곡의 감소폭은 5점-7점을 넘는 경우에는 비교적 작다는 것을 나타내었다(Green과 Rao, 1970; Lehmann과 Hulbert, 1972; Ramsay, 1973).

상관타당도에 관한 연구의 결과들은 신뢰도에 관한 결과와 대체로 일치하지만, 일부 연구들에서는 이와 상반되는 증거를 얻었다(예를 들면, Smith와 Peterson, 1985). 예를 들면, Matell과 Jacoby(1971)는 7점 및 8점 단극 척도들이 이보다 더 길거나 더 짧은 척도보다도 상관타당도가 더 높다는 것을 발견하였다. 이와 마찬가지로, Rosenstone 등(1986)은 5점 양극 척도가 3점 양극 척도보다 개념적으로 관련 있는 변수들을 더 잘 예측한다는 것을 발견하였다.

척도점의 수가 타당도에 미치는 효과를 평가하는 것과 관련 있는 다른 연구들에서는 평정 척도가 문맥 효과의 영향을 받는 정도를 검증하였다(Wedell과 Parducci, 1988; Wedell 등, 1990). 예상대로, 다양한 판단 차원(예를 들면, 얼굴에 대해 느끼는 행복, 기하학적 물체의 크기, 인생사(life events)에 대해 느끼는 행복)상의 여러 자극들을 평정하는 데 문맥이 미치는 효과는 척도점 수와 상관관계에 있었다. 물론 결과는 판단을 내리는 것에 좌우되지만, 문맥 효과는 척도점의 수가 증가함에 따라 감소하였고, 7점 척도를 넘는 경우에는 비교적 완화되었다.

척도를 평가하는 다른 기준은 응답자 내의 태도와 응답자 간의 태도 사이에 차이점이 얼마나 많은지에 대한 정보의 양이다. 물론, 너무 짧은 척도는 응답자가 구별하는 다양한 사물들의 차이점을 충분히 제공할 수 없다. 이 명제와 관련하여, 몇몇 연구들에서는 척도점의 수를 7-9점까지 높일 경우 더 유용한 정보를 제공하고(Bendig, 1954b; Bendig와 Hughes, 1953; Garner, 1960), 척도점의 수가 12점 이상일 경우는 제공정보량이 감소함을(McCrae, 1970a, 1970b) 입증하였다. 사용한 척도점은 4점과 19점 사이의 걸쳐 그 비율이 일정하였지만, 7점보다 더 긴 척도들에 추가된 점들은 실질적인 정보를 추가적으로 제공하지 않은 것 같다(Matell과 Jacoby, 1972).

4.3.3 크기측정 척도

"크기측정척도"라는 척도구성에 대해 본질적으로 다른 접근방법을 제안하는 문헌들이 출간되고 있다(예를 들면, Lodge, 1981). 이 접근방법에서는 길이가 무한한 척도들을 사용하여 응답자가 사물 간의 차이를 구분하는 정도를 제한하지 않는다. 예를 들면, 응답자에게 길이가 4인치인 선을 보여주고 이것이 그들이 오렌지를 좋아하는 정도를 나타낸다고 말한다. 다음에 그들에게 사과를 좋아하는 정도를 선으로 그리되 오렌지보다 덜 좋아하면 짧은 선을, 많이 좋아하면 긴 선을 그리라고 요청한다. 이러한 시각적 방법 외에도 응답자에게 그들이 듣는 소리의 음높이를 조절하게 하거나 악력계를 눌러 자신들이 가하는 압력의 양을 비교하여 판단을 내리게 하는 방식으로 이런 척도를 구성할 수 있다.

원리로 볼 때 이 것은 매력적인 것 같지만, 특히 전형적인 조사 연구 환경에서 실제로 실시하는 것은 쉽지 않다. 크기측정척도는 한 평가 차원에서 자극들이 놓인 위치를 가지고 이것들 간의 비율만을 나타내는 것이고 응답자가 이것을 위치시킨 지점의 절대적 수준을 나타내는 것은 아니다. 즉, 응답자가 오렌지보다 사과를 두 배나 더 좋아한다는 것을 알 수 있지만 사과나 오렌지가 그의 태도 연속선상에서 긍정적인 쪽이나 부정적인 쪽에 있는지는 알 수 없다. 게다가 크기추정 자료의 신뢰도 및 타당도와 전통적인 범주형 평정척도를 사용한 자료의 신뢰도 및 타당도를 비교한 연구들은 후자가 더 우수하다는 것을 증명하였다(예를 들면, Kaplan 등, 1979; Miethe, 1985). 따라서 현재로는 전통적 접근방법이 더 나은 것으로 보인다.

4.3.4 중간점

설계상의 또 다른 중요한 문제는 중간선택문항을 포함시킬지 여부를 결정하는 것이다. 양극 차원에서 중간 선택사항은 찬성/반대, 동의한다/동의하지 않는다 또는 좋아한다/싫어한다 척도에서 "보통"이나, 또는 대규모 증가에서 대규모 감소에 이르는 변화를 나타내는 척도에서(예를 들면, "미 국방 예산 지출의") "현상 유지"를 포함하는 여러 형태가 있다. 단극 척도에서 중간점은 온건한 입장을 나타내지만, 양극 차원의 경우처럼 그 의미가 정확하지 않다.

실제 중립적인 입장을 갖고 있는 개인에게 찬성(기호)이나 반대(혐오)의 응답을 요구 할 경우 측정 오차가 증가된다면 양극 차원에서 중간선택문항을 사용하는 것은 적절한 방법이 될 수 있다. 하지만, 이쪽이나 저쪽 방향으로 약간 편향된 입장을 갖고 있지만, 제시된 중간점이 적은 수고만으로 쉽게 선택할 수 있고 정

당하다고 하기 쉽기 때문에 많은 사람들이 중간선택문항을 선택할 것이다. Krosnick(1991)은 이런 일이 발생할 수도 있다는 한 시나리오를 제시하였다.

Krosnick(1991)에 따르면, 조사 응답자들이 진실로 성실하게 응답하려고 한다면(조사 *최적화(optimizing)*라고 명명한 행위), 인지적 과제는 매우 부담이 된다. 따라서 개인은 때때로 응답의 책임감은 보여주면서 이런 노력을 더 들이지 않는 방법을 찾을 것이다(조사 *만족화(satisficing)*라고 명명한 응답 행위). Krosnick(1991)에 따르면, 만족을 원하는 응답자는 질문에서 이것을 할 수 있는 계기(cue)를 찾는다. 이러한 계기가 불확실할 경우, 응답자는 그 대신 최적화를 선택할 것이다.

Krosnick(1991)는 여러 계기 중 척도 중간점을 언급하였다. 구체적으로 그는 중간점이 현상 유지를 나타낼 때(예를 들면, "정부의 국방비 지출을 현 수준으로 유지"), 면접자가 선택 이유를 캐묻는 경우 변명하기 쉬울 것으로 생각해서 응답자가 중간점을 선택한다고 주장하였다. 따라서 중간점이 제시되면 응답자는 어느 한편도 선택하지 않고 만족화를 하게 되지만, 중간점이 제시되지 않은 경우에는 최적화할지 모른다. 따라서 중간점을 제시하면 유용한 자료를 수집할 수 없을지도 모른다.

이 분야에 관한 경험적 연구들에서는 척도 중간점들을 제시함으로써 나타나는 다양한 효과들을 입증하였다. 첫째로, 사람들은 중간점의 선택이 정당하다고 생각하지 않으면 자발적으로 선택하여 응답하는 일이 드물지만, 상당수의 사람들은 질문에 명시된 경우 중간점을 선택한다(Ayidiya & McClendon, 1990; Bishop, 1987; Kalton 등, 1980; Schuman & Presser, 1981). 중간점이 제시되었을 때 이에 끌리는 것은 인지 능력이 약하고(Narayan과 Krosnick, 1996), 관련된 주제가 응답자 개인에게 별로 중요하지 않은(Krosnick와Schuman, 1988; Schuman과 Presser, 1981, pp. 173-175; Stember와Hyman, 1949-1950도 참조할 것) 경우 가장 높게 나타난다. 중간선택문항이 제시되지 않으면 사람들은 마음속에 동전을 던져 나온 결과대로 응답을 할 것이라고 생각하기도 하지만, 이런 사람들에 의한 의견의 분포는 평평한 형태와는 유의적인 차이가 있다(Ayidiya와McClendon, 1990; Bishop, 1987; Kalton 등, 1980; Schuman과Presser, 1981). 따라서 그들이 동전 던지기 결과대로 응답하는 것 같지는 않다.

또 다른 연구들에서는 중간점의 존재가 신뢰도와 타당도에 미치는 효과를 검증하였지만 서로 상반되는 결과를 얻었다. 예를 들면, Alwin 및 Krosnick(1991)와 Andrews(1984)가 각기 행한 이차 분석에서는 중간점을 가진 척도가 중간점이 없

는 척도보다 신뢰도가 더 낮다는 결론을 얻었다. 하지만 대체로 실험 연구에서는 중간점의 유무에 따라 신뢰도의 일관성이 증가 또는 감소 형태를 보이는 것을 발견하지 못했다(Masters, 1974; Jacoby & Matell, 1971; Matell & Jacoby, 1971). Kalton 등(1980)은 중간점을 포함시키는 것이 상관타당도(태도와 인적사항(demographics) 간의 관련성으로 측정한)에 아무 영향을 미치지 않음을 발견하였지만, Schuman과 Presser(1981)는 중간선택문항이 포함된 경우 태도 간의 관련성이 높아졌음을 발견하였고, Stember & Hyman(1949-1950)은 중간점을 포함시킬 경우 면접자 편향효과가 감소하였음을 밝혀내었다. 따라서 중간점을 포함시키는 어떤 경우는 자료의 질을 향상시킬 수 있고 또 다른 경우에는 그렇지 않을 수 있다.

4.3.5 조사 실시의 용이성

길이가 각기 다른 척도들의 효과성을 평가하는 또 다른 기준은 응답자들에게 실시할 수 있는 용이성 및 그 성공도이다. Matell & Jacoby (1972)는 척도의 길이가 길수록 이것을 실시하는 데 많은 시간이 소요되며, 특히 척도의 길이가 13점 이상인 경우에 소요 시간이 많이 증가한다는 것을 발견하였다. 응답자가 질문지 한 세트의 모든 평정 척도를 제대로 작성하는 비율을 조사한 한 연구에서 2점 척도보다 4점 척도의 경우 제대로 작성한 비율이 더 높다는 사실을 보였지만(Ghiselli, 1939), 다른 연구에서는 3점과 7점 척도 간에 이러한 비율의 차가 없다는 것을 발견하였다(Smith & Peterson, 1985).

4.3.6 결론

이상의 연구들에서는 5-7점의 평정 척도가 이보다 더 짧거나 긴 척도보다 신뢰도와 타당도가 더 높은 것으로 보이기 때문에 이 길이가 최적이라고 제시하고 있다. 중간점의 경우, 연구자들이 중간점을 포함시킬 여부는 이보다 덜 명확하다. 타당도의 경우에는 상반된 증거가 나타났고, 만족화 이론에 따르면 중간점을 포함시키는 경우 측정의 질이 감소할 것이라고 한다. 따라서 우리는 이 문제를 명쾌하게 해결할 수 있는 추가 연구를 기대한다. 그 때까지는 중간점이 하나의 개념으로 필요한 경우(예를 들면, 국방비 지출의 증가, 감소 또는 현상 유지 여부에 관한 질문)이를 포함시키는 것이 현명하며, 질문지상의 다른 부분에서 응답자가 만족화 할 가능성을 최소화하기 위한 조치들을 취해야 한다(Krosnick(1991) 참조).

4.4 척도점 명칭

척도 구성 시 내려야 할 또 다른 중요한 결정은 모든 척도점의 명칭을 문자로 할 것인가 또는 일부 척도점의 명칭을 숫자로만 할 것인가에 관한 것이다. 평정척도가 의미를 가지려면 최소한 척도의 끝점에는 명칭을 붙일 필요가 있다. 그러나 많은 연구자들은 나머지 척도점에는 명칭을 붙이지 않고 있다.

끝점의 척도만 문자를 사용하는 것에는 두 가지 중요한 이점이 있다. 첫째, 언어 자체가 갖는 모호성 때문에 문자 명칭보다는 숫자가 더 정확할 수 있다. 둘째, 숫자로 된 척도 값은 복잡한 문자 명칭보다(전화 면접 중에) 기억하기 쉽다. 따라서 끝점의 척도의 명칭만 문자를 사용한 경우 모든 척도점의 명칭을 문자로 한 경우보다 응답의 인지적 부담이 적을 것이다.

하지만, 몇 가지 이유에서는 모든 척도점의 명칭을 문자로 하는 경우에 보다 향상된 자료의 질을 기대할 수 있다. 사람들이 일상적인 대화에서 숫자를 통해 복잡한 개념을 표현하는 일이 없기 때문에, 문자명칭은 응답자가 자신의 의견을 표현하는 자연스러운(따라서 더 쉬운) 방법이 될 수 있다. 게다가 숫자척도점은 문자 명칭에 의해 설정된 개념을 동일간격으로 구분하는 점이라는 것 외에는 고유한 의미가 없다. 따라서 모든 척도점의 명칭을 문자로 하는 것은 척도점의 의미를 명확하게 하여 응답의 용이성을 증가시키고 정확성을 제고하는 데 도움을 줄 것이다.

4.4.1 신뢰도

척도점 전체의 명칭을 문자로 한 척도와 일부의 명칭만 문자로 한 척도를 신뢰도 측면에서 비교한 연구들이 있다. 횡단 신뢰도에 관한 연구에서는 신뢰도가 문자명칭의 영향을 받지 않거나(Finn, 1972; Madden & Bourdin, 1964), 문자표시가 더 많은 척도일수록 신뢰도가 낮아진다(Andrews, 1984)는 것을 밝혀냈다. 하지만, 종단 신뢰도에 관한 모든 연구들에서는 문자명칭이 많을수록 신뢰도가 증가한다는 일관성 있는 결과를 보였다(Alwin & Krosnick, 1991; Krosnick & Berent, 1993; Zaller, 1988). 게다가 신뢰도 증가는 학력 수준이 낮거나 중간 정도인 응답자의 경우 두드러지게 나타났는데, 이들은 명료한 명칭으로 인해 많은 도움을 얻었을 것이라 추정된다(rosnick & Berent, 1993). 문자명칭을 사용하여 평정 척도의 끝점들의 사이가 벌어진 것처럼 해도 신뢰도가 더 높아진다(Bendig, 1955). 문자명칭은 이런 점에서 장점이 된다.

4.4.2 타당도

문자명칭에 관련된 연구들은 문자명칭이 얻어진 자료의 타당도를 증가시킨다고 제시하였다. Krosnick & Berent(1993), Dickinson & Zellinger(1980)는 더 많은 문자 명칭을 사용하여 태도를 측정한 경우, 특히 공식교육을 비교적 적게 받은 응답자의 태도와 다른 변수 간 상관계수가 높다는 것을 발견하였다. Andrews(1984), Alwin & Krosnick(1991)은 각자 행한 이차 분석에서 일부의 명칭만 문자를 사용한 문항보다 전체의 명칭에서 사용한 경우 더 큰 참점수분산(true score variance)을 발견하였는데, 이 결과 후자가 타당도가 더 높았다. 동일한 사물을 평정할 경우 사용되는 평정 척도가 더 많은 문자명칭을 갖고 있으면 평정 결과가 서로 일치하는 경향이 있다(Barrett 등, 1958; Bendig, 1953; Peters & McCormick, 1966). 게다가 더 많은 문자명칭을 사용하면 여러 평정 차원과 사물이 더 세세하게 구별된다(Barrett 등, 1958; Bendig & Hughes, 1953; Bernardin 등, 1976). 마지막으로, 더 많은 문자명칭을 사용하면 척도상의 등급들이 문맥 효과의 영향을 덜 받는다(Wedell 등, 1990). 따라서 이상의 많은 증거들을 고려해 볼 때, 전체의 명칭에 문자를 사용한 척도는 측정 타당도를 높인다고 할 수 있다.

4.4.3 응답자 만족도

응답자 만족도를 평가한 몇몇 연구들은 대부분의 사람들이 많은 문자명칭을 갖고 있는 평정 척도의 사용을 선호한다는 것을 밝혀내었고(Dickinson과 Zellinger, 1980; Wallsten 등, 1993; Zaller, 1988), 이러한 척도들이 타당도가 더 높은 측정 도구라는 것을 증명하였다(Dickinson과 Zellinger, 1980).

4.4.4 문자 명칭의 선택

전체의 명칭에 문자를 사용한 척도는 일부만 문자를 사용한 척도보다 신뢰도와 타당도가 더 높은 것으로 보이지만, 문자명칭의 이점은 적절하게 선택하기 나름이다. 문자명칭이 유용한 것이 되려면 응답자들에게 논리적으로 정확한 의미를 나타내야 한다. 특히 분석자가 잠재적 구성개념의 모든 변수를 통계 분석에서 구간수준의 변수로 처리할 계획이라면, 그가 선택하는 문자명칭들은 연속선상에서 비교적 동일한 구간을 나타내야 한다.

몇몇 연구에서는 조사에서 보통 사용하는 네 가지 차원의 문자명칭인 기호, 수량, 빈도 및 가능성에 대해 사람들이 부여하는 의미를 연구하였다. 이들은 응답자에게 여러 수식어에(우수한, 양호한, 보통) 수치를 부여할 것을 요청하였다(예

를 들면, 0에서 100까지의 척도나 크기측정척도를 사용하여). 이러한 문자명칭 중 다수는 여러 표본들에 걸쳐 매우 안정적인 의미를 가질 뿐만 아니라, 여러 해에 걸쳐 다양한 평정 방법을 사용한 경우에도 의미 있는 일관성을 갖는 것으로 보인다(몇 가지 결과들의 설명에 대해서는 표 4.1 참조). 이 척도 값의 종단 신뢰도는 매우 높은 것으로 보였고, 검사-재검사 상관계수도 .90을 넘기는 경우가 많았다. 따라서 문자명칭은 유용한 척도를 구성하기에 충분하고 정확한 의미를 갖는다.

몇몇 연구자들은 위의 문헌들을 검토하고 결과물로 제안된 문자명칭들이 부정확하다는 점을 강조하였다(Pepper, 1981). 어떤 용어의 의미는 다양한 요인에 따라 유의하게 변화된다. 예를 들어, 빈도수에 대한 절대적 의미는 이미 설명했던 사건의 유형에 따라 변할 수 있다(이를테면, 1년에 한 번 발생하는 지진은 "자주"라고 묘사되는 반면, 1년에 한 번 영화관에 가는 것은 "드물게"로 묘사될 수 있다). 게다가 사람들은 척도가 동일한 간격이라고 가정하기 때문에 한 명칭의 의미를 척도 내에서 제시하고 있는 다른 명칭의 문맥에서 해석한다. 서로 다른 사회 집단들은 때로 체계적으로 다른 방식으로 명칭을 해석한다(예를 들어, Schaeffer(1991) 참조). 하지만, 우리는 이 증거들을 다른 관찰자들처럼 우려할만한 것으로 보지 않는다. 응답자들이 이해하는 문자명칭의 의미는 무작위적인 차이나 체계적인 차이를 보이기도 하지만, 많은 명칭들은 태도 측정을 발전시키는데 매우 유용할만큼 충분하고 보편적인 의미를 갖고 있는 것 같다.

표 4.1 좋아함(기호)을 평가하는 문자 명칭의 척도값

언어 명칭	Myers & Warner (1968)	Wildt & Mazis (1978)	French-Lazovik & Gibson (1984)	Stone & Schkade (1991)
우수	93	91	91	99
매우 양호	78	82	80	
양호	67	70	58	73
보통	43	49		48
불량	21	17	16	23
매우 불량	12	5	11	1

4.4.5 숫자 명칭의 선택

척도 전체가 문자명칭이지만, 후에 자료입력을 쉽게 하기 위해 숫자도 포함시키고 싶을지 모른다. 자계식 질문지의 경우, 숫자들은 일부만 문자명칭을 사용한

척도에 영향을 줄 수 있다. 응답자들이 구체적인 수치들을 사용하여 척도점이나 문자명칭의 의미를 추론하기 때문이다. 예를 들면, 끝점의 명칭을 "전혀 성공하지 못함"과 "대단히 크게 성공함"이라고 정한 0에서 10까지의 11점 척도를 사용하여 응답자들에게 자신의 인생에서 성공한 정도에 대해 질문하였을 때, 약 34%의 응답자들은 0에서 5까지의 수치를 선택하였다(Schwarz와 Hippler(1991), Schwarz 등(1991) 참조). 대조적으로, 동일한 끝점의 명칭을 사용하고 수치의 범위가 -5에서 +5인 11점 척도의 경우에는 13%만이 -5에서 0까지를 선택하였는데 이 범위는 이전의 것과 논리적으로 동일한 것이다. "전혀 성공하지 못함"이라는 명칭과 숫자 "0"을 대응시켰을 때에는 성공한 적이 없다는 의미로 받아들이지만, 숫자 "-5"에 대응시켰을 때에는 실패가 있었다는 의미로 추론하였기 때문에 이와 같은 차이를 보이는 것 같다. 따라서 숫자들을 평정 척도에 사용하려면, 척도점의 의미를 의도한대로 강화시킬 수 있도록 숫자들을 세심하게 선택해야 한다.

4.4.6 결론

전체의 명칭에 문자를 사용한 척도는 일부만 문자를 사용한 척도보다 신뢰도와 타당도가 더 높은 것으로 보인다. 이를 근거로 연구자에게 그렇게 하는 것이 유용할 경우에만 모든 척도점에 문자명칭을 사용할 것이며 숫자만 사용하는 것은 삼가할 것을 권고한다. 연구자들은 응답자에게 비교적 정확한 의미를 전달하면서 관심의 연속선상에서 동일한 구간을 나타낼 수 있는 명칭을 선택해야 한다. 이를 위해서는 이미 측정된 바 있고 정신측정학적으로 우수한 특성을 갖는 것으로 알려진 명칭들을 사용해야 한다. 이 명칭들을 구할 수 없는 경우, 연구자들은 적절한 명칭을 찾기 위해 사전 조사를 실시해야 한다. 수치는 명칭의 의미를 전환시킬 수 있기 때문에, 연구자들은 함께 사용하는 것을 피하고 문자 선택문항을 간단하게 해야 한다.

4.5 의견없음 필터

조사에서 태도에 관한 질문을 할 때, 일반적으로 응답자의 응답은 그들이 전에 기억해 둔 의견이라고 가정한다. 조사가 관심을 두는 태도 대상에 대해서 기존의 의견을 정확하게 기억해두지 않은 경우, 질문 자체는 그가 기억하는 관련

정보들을 사용하여 태도 대상에 대해 새롭고 논리적인 평가를 하도록 자극한다. 따라서 기존의 태도이건 새로 형성된 태도이건 간에 이에 근거한 응답은 이 대상에 대해서 응답자가 우호적인지 비우호적인지를 나타낼 것이다.

4.5.1 무관심 가설(Non-Attitude Hypothesis)

Converse(1964)는 매우 다양한 가능성을 제시하였다. 응답자는 때때로 조사를 할 때 태도 질문에 아무렇게나 간단히 응답한다고 그는 주장했다. 그가 제시한 증거는 처음은 1956년, 다음은 1958년, 마지막으로 1960년에 응답자패널이 각각 응답한 동일태도 간 상관계수였다. 이 상관계수는 아주 낮았는데, 여덟 가지의 정책 문제들에 있어서 타우베타의 범위는 약 .28(연방 주택 건설 지원의 경우)에서 약 .47(흑백 분리 학교 폐지의 경우)까지였고 평균은 .37이었다. 측정의 시간 간격이 2년에서 4년으로 증가했어도 상관계수가 두드러지게 낮아진 것은 아니기 때문에, 그는 진정한 태도변화가 일어나지 않았다고 결론지었다. 게다가 응답자들이 표현한 의견의 분포가 4년간 꽤 일관적이었고, 이는 같은 결론이라는 걸 뒷받침해준다.

태도의 변화 때문이 아니라면 이와 같이 시간의 경과에 따른 태도보고서 간의 상관계수가 낮은 이유가 무엇 때문인가? Converse(1964)는 의미 있는 판단을 내리기 위한 관련 정보가 전혀 없을 지라도, 많은 미국인들이 기존의 의견이 없어도 태도에 대해 응답해야 하는 압박감을 가지고 있다고 주장했다. 면접자가 면전에 있다는 사실과 그가 집요하게 질문을 하는 것은 응답자가 어떤 의견을 가지고 있을 것이라 기대하고 있고 연구자는 그 의견이 무엇인지 알고 싶어 한다는 것을 암시한다. Converse는 전반적인 주제에 대해 무지하다는 것을 인정해서 바보로 보이기 싫기 때문에 응답자는 질문의 기본적인 것들에 대해서 알고 있는 바가 전혀 없어도 응답을 꾸며냈다고 주장하였다.

4.5.2 다른 보강 증거

사실, Converse(1964)의 연구결과가 발표되기 전후에 행해진 많은 다른 연구들에서도 그와 일치하는 결과들이 있다. 예를 들면, 일단의 연구들에서는 시간의 경과에 따른 태도보고서들의 일관성이 비교적 낮은 수준이라는 것을 밝혀내었다(예를 들면, Jennings와 Markus, 1984; Jennings와 Niemi, 1978). 시간의 경과에 따른 자료의 잠재계급분석(latent class analysis)에서는 평균 50%~ 70%의 응답자들이 공공정책문제에 관하여 아무런 태도를 갖고 있지 않는 것으로 나타났다(Brody, 1986; Taylor, 1983).

이러한 무관심(nonattitudes)에 관한 결론을 뒷받침하는 다른 연구에서는 태도 보고서들이 보통 다른 태도보고서들을 거의 예측하지 못하고(이를 검토하려면, Kinder & Sears(1985) 참조), 관련 행동을 잘 예측하지 못한다(Wicker, 1969)는 것을 밝혀냈다. 게다가 태도-행동 일관성이 높은 연구 결과들은 실제로 인위적인, 상관측정오차(correlated measurement error)일 수 있다(Budd, 1987; Budd & Spencer, 1986). Gill(1947), Hartley(1946, pp. 26-31), Ehrlich & Rinehart(1965), Schuman & Presser(1981), Bishop 등(1986)과 기타 학자들은 이상의 결론을 뒷받침하는 다른 발견점을 보고하였는데, 허구의 대상이나 너무 애매해서 사람들이 들어본 적도 없을 만큼 친숙하지 않은 대상에 대한 태도를 보고했음을 보여 준다.

Converse(1964)가 자신의 초기 증거를 해석한 결과는 다른 해석도 가능하다고 주장하는 평론가들로부터 상당한 비판을 받았다. 마찬가지로 우리가 앞서 무관심 가설로 기술한 증거들도 아주 다르게 해석할 수 있다(곧 출간될 Krosnick & Fabrigar 참조). 이 문헌들이 무관심 개념을 뒷받침하는 결정적인 증거 자료를 제공하는 건 아니다. 하지만, 이 가설에 대한 일치점은 연구자들이 무관심 보고서를 미리 방지할 수 있는 수단을 탐구하는 계기가 되었다.

4.5.3 의견없음 여과질문의 유형

사실, 유의도에 대해 경험적으로 고찰해온 유일한 방법은 의견없음을 여과하는 것이다. 이 개념은 조사 질문의 형식 때문에 응답자들이 부분적으로 무관심을 보고한다는 것이다. Schuman & Presser(1981, p. 299)가 지적하듯, 응답자는 대개 "규칙에 따라 경기를 하는"데, 이것은 응답자들이 제공되는 형식 외의 합리적인 응답을 제공하기 보다는 폐쇄형 질문에서 제시하는 응답선택문항에서 선택을 하고 있다는 의미이다. 어떤 질문에서 "모르겠음"(DK: Don't Know)이나 "의견없음"(NO: No Opinion)의 선택문항이 명시되지 않는 경우, 이것은 연구자들이 응답자에게 의견을 가지고 있을 것으로 기대한다는 의미로 해석되어 응답자의 무관심을 불러오게 된다. 그래서 응답선택문항이 명시적으로 인정되는 경우, 상당한 수의 응답자들이 DK를 시인할 것이다.

많은 연구에서 이와 일치하는 증거를 보고했다(Ehrlich, 1964; Bishop 등, 1980a, 1980b, 1982, 1983; 1986; Schuman & Presser, 1979, 1981; Schuman & Scott, 1989). 게다가 여과질문의 문안을 다양한 방법으로 구성하면, 각 방법에 따라 전혀 다른 결과들이 나올 수 있다. 준 여과질문(quasi-filter)은 응답자들에게 그들의 태도에 대한 질문을 주면서 질문의 마지막에 DK나 NO를 간단히 명시하는 것이다. 예를 들어, "낙태의 합법화를 찬성하십니까, 반대하십니까? 아니면 의견이 없습니까?"라는 질문을 제시할 수 있다. 반대로 완전 여과질문은 태도 측정

에 앞서 응답자가 이 문제에 대해 의견을 갖고 있는지 여부를 물어보고, 긍정적으로 대답하는 경우에만 태도 질문에 응답할 수 있다. 완전 여과질문은 DK응답을 더 많이 끌어낸다(Bishop 등, 1983; Schuman & Presser, 1981)

DK응답을 명시하거나 정당화하는 정도를 달리해서 완전 여과질문을 다양하게 나타낼 수 있는데, 이런 다양성도 중요하게 보인다(Bishop 등, 1983). 예를 들면, "이에 대해 의견을 갖고 있습니까, 아니면 의견이 없습니까?", "이 문제에 관하여 한쪽보다 다른 쪽을 더 선호할 정도로 충분한 관심이 있으십니까?", "이 문제에 대해 많이 생각해보셨습니까?" 혹은 "이에 관한 의견을 갖기 위해 충분히 듣거나 읽으셨습니까?"라고 질문을 할 수 있다. 관심사, 생각 또는 주제에 관한 정보를 접한 정도에 초점을 맞추는, 위의 질문 중 마지막 세 가지의 여과질문을 사용하면, 응답자들이 이전에 이 주제에 대해 고려해 본 일이 없기 때문에 더 쉽게 의견없음을 시인할 것이다. Bishop 등(1983)은 여과질문들을 강도(strength)로 순서를 정할 수 있다는 것을 밝혀내었다. 즉, 위의 여과질문 중 첫 번째 여과질문은 "관심"을 묻는 여과질문보다 더 적은 DK응답을 얻었고, "생각"과 "듣거나 읽은" 여과질문들보다도 더 적은 DK응답을 얻었다.

4.5.4 의견없음 응답의 의미

무관심 보고를 방지하는 의견없음 여과질문의 능력에 관한 증거는 고무적이다. 그러나 DK응답은 응답자들이 정말 의견이 없다는 것을 의미해서 억지로 의견제공을 요구받은 경우 순전히 아무렇게나 추측하거나 무의미한 응답을 한다고 가정하고 있다. 하지만 과연 그러한가?

사실, 사람들은 태도를 전혀 갖고 있지 않은 것보다 다른 많은 이유들로 인해 DK응답을 제공하는지도 모른다(예를 들면, Dunnette 등, 1956; Edwards & Ostrom, 1971; Klopfer & Madden, 1980; Krosnick, 1991; Smith, 1984). 첫째, DK응답은 척도 중간점이 제시되지 않으면 응답하기 상당히 어려운 양면적인 태도를 나타낼 수 있다. 둘째, DK응답은 어떤 사람이 대상에 대해 중립적 생각이나 감정을 가지고 있는데, 중립의 응답선택문항이 제시되지 않은 경우(예를 들면, 동의/동의 안 함 질문의 경우)일 수 있다. 셋째, DK선택문항은 응답자가 질문을 이해하지 못한 경우일 수도 있다. 넷째, 응답자들은 자신이 태도 척도 중 어디에 해당하는지 대략 알고 있지만(예를 들어, 1-7 척도에서 6이나 7 근처), 척도점의 의미가 애매모호해서 정확하게 어디를 선택할지 확신이 없을 수도 있다. 마지막으로, DK응답은 만족화를 나타낼 수 있다. 응답자는 이미 형성된 의견이나 즉석에서 적절한 의견을 만드는데 사용할 수 있는 정보를 갖고 있을지라도, 의견을 제

시할 동기나 능력의 부족으로 정신적 활동을 피하기 위해 DK선택을 골랐을지 모른다(Krosnick, 1991).

많은 연구들은 모르겠음이나 의견없음 응답이 다양한 의미를 가질 수 있다고 제안한다(Coombs & Coombs, 1976; Dunnette 등, 1956; Feick, 1989; Smith, 1984). 게다가 일부 연구들에서는 DK응답의 이유로 무관심보다 양면적 감정을 들고 있다(예를 들면, Ehrlich, 1964). 따라서 "의견없음"이라고 말하는 응답자들에게 억지로 실제적인 응답을 강요하면, 그 응답은 기본적으로 확률잡음(random noise)이 되지 는 못할 것이다.

4.5.5 의견없음 여과질문이 자료의 질을 향상시키는가?

이런 배경을 감안하여 볼 때, 흥미로운 점은 의견없음 선택문항을 제시하면 입수한 태도보고서에서 확률변이(random variation)가 감소되는지에 관한 것이다. 물론, 무관심보고서는 확률변이를 증가시킬 필요 없이 체계적 응답편향에 근거할 수 있다. 그러나 일부 무관심보고서는 순전히 임의라는 Converse(1964)의 주장이 옳다면, 임의성은 변수 관계를 약화시킬 것이다. 이 문제에 접근하는 다양한 방법들이 사용되었는데, 그런 활동은 모두는 상반된 증거나 임의적 무관심 관점의 명백하게 거부하는 것이었다.

예를 들면, Gilljam & Granberg(1993)는 조사 응답자들에게 원자력발전소 건설에 대한 태도를 알아보기 위해 세 가지 다른 질문들을 하였다. 이 질문들 중 처음 하나에는 의견없음 여과질문을 포함하였는데, 응답자의 15%가 이것을 선택하였다. 나중에 질문했던 나머지 두 개에는 의견없음 여과질문을 포함하지 않았는데, 각각 3%와 4%의 응답자만이 자발적으로 "모르겠음"응답을 선택했다. 따라서 처음에 "모르겠음"이라고 말한 응답자 대다수가 나중 두 개의 질문에서는 의견을 제시한 것이다. 그렇다면, 이제 문제는 나중에 제공한 응답이 진정한 의견이었는지 아니면 무관심이었는지의 여부이다.

이 문제에 접근하기 위해 Gilljam & Granberg(1993)는 두 개의 지표를 검증했다. 그것은 나중 두 개의 태도보고서 간의 상관관계 강도(strength)와 이것들이 다음에 실시할 실제 원자력발전소에 관한 국민투표에서 사람들의 투표를 예측할 수 있는 능력이다. 나중 두 개의 질문문항에 대한 응답 간 상관계수는 처음의 질문에서 "모르겠음"이라고 말한 사람들의 경우 .41(p<.001), 처음의 질문에서 실제 응답한 사람들의 경우에는 .82(p<.001)였다. 마찬가지로, 나중 두 개의 문항들에 대한 응답은 초기에 "모르겠음"이라고 말한 사람들의 결과에서 평균 76%를 정확하게 예측하였는데, 초기에 실제 응답을 한 사람들의 경우에는 94%였다.

따라서 무관심 가설을 뒷받침하는 증거는 여과질문이 타인의 의견보다는 평균적으로 질이 떨어지지만 순전한 임의가 아닌 의견을 가진 사람들을 분리해내는 결과를 보여 상반된 관점을 제공한다.

두 가지 비실험 연구에서도 상반된 증거를 얻었다. Andrews(1984), Alwin & Krosnick(1991)은 일부는 의견없음 여과질문을 포함하고 다른 일부는 포함하지 않는 여러 조사문항에 있는 임의측정오차의 양에 대한 상관을 메타분석하였다. Andrews(1984)는 의견없음 여과질문을 포함시켰을 때 확률오차의 크기가 유의하게 더 작은 결과를 얻었지만, Alwin & Krosnick(1991)은 정반대의 결과를 얻었다.

이러한 의견없음 여과질문에 관한 주장에 명백한 이의를 제기하는 연구에서는 대상 문항들과 기타 변수 간 관련성의 변화를 측정하였다. 의견없음 선택문항을 제공하고 일부 응답자들이 이것을 선택하게 되어 질이 나쁜 자료 제공을 방지하게 되면, 변수 간의 관련성은 더 높아질 것이다. Bishop 등(1979)은 여과질문이 있거나 없는 유사한 질문을 서로 다른 국가에서 추출한 표본에게 제시한 기존 조사들을 이차 분석하였다. 여과질문이 없는 질문보다는 여과질문이 있는 질문의 경우 관련성이 더 높은 경향을 발견하였지만, 그들의 비교에는 시간, 조사실시방식, 질문문맥 등의 혼란(confounding)이 포함되었다. Schuman & Presser(1981)가 실시한 의견없음 여과질문의 사용에 관한 19회의 실험에서는 대부분의 사례들에서 유의한 상관관계의 변화가 발생하지 않았다. 두 가지 사례에서는 유의한 관련이 있는 변화를 보였지만, 이 중 어느 것도 변화의 방향이라는 측면에서 볼 때 Converse의 예상과 일치하지 않았다. 더욱이 Schuman & Presser(1981)는 태도와 응답자들의 교육, 정치에 대한 관심, 나이나 성별 간의 관계가 여과질문 사용에 의해 변화한 사례들은 찾지 못했다. 따라서 이 증거는 의견없음 여과질문이 입수하는 자료의 질을 향상시킨다는 주장과 명백히 일치하지 않는 것이다.

마지막으로, 다음 두 연구에서는 NO여과질문의 유무라는 실험변수가 신뢰도에 미치는 효과를 탐구하였다. 여과질문이 무관심을 걸러낸다는 견해가 타당도가 있으려면 여과질문을 사용할 때 신뢰도가 증가해야 할 것이다. 하지만 McClendon & Alwin(1993)은 질문에 의견없음 여과질문을 포함시켰을 때에 신뢰도가 더 높아진 사례를 발견하지 못하였다. Krosnick & Berent(1990)는 의견없음 여과질문을 포함시켰는지 여부에 따라 시간의 경과에 따른 태도보고서들의 일관성이 유의하게 변화하는 사례를 발견하지 못했다.

4.5.6 선택문항 관점

이 모든 연구들은 의견없음 여과질문이 무관심을 제거한다는 주장에 이의를 제기하고 있다. Converse(1964)의 관점에서 볼 때 이 결과들은 놀라운 것이지만, 조사만족화의 개념에서는 꽤 이해할 수 있는 것이다(Krosnick, 1991). 의견없음 선택 문항을 제시하는 것은 입수하는 자료의 질을 향상시키지 못할 뿐만 아니라, 유용한 자료의 수집을 방해할 수도 있다. 구체적으로 설명하면, 이러한 응답 선택 문항은 그렇지 않아도 만족화 성향을 가진 응답자들이 "모르겠음"이라고 말하여 만족화하도록 조장하는 계기를 제공할 수 있다. 제시된 의견없음 선택 문항이 없고 다른 만족화 계기도 없는 경우, 응답자들은 만족화 대신에 최적화할 것이다. 따라서 이들로부터 유용한 자료를 수집할 수 있는데도 의견없음 선택 문항을 제시하면 자료를 얻을 수 없게 된다.

이 견해가 옳다면 NO 선택 문항에 대한 매력은 만족화의 조건에서 가장 클 것이다. 즉 그런 조건이란 최적화에 대한 응답자 능력과 동기부여가 낮고 응답 작업이 매우 어려운 경우이다. 이러한 추론과 일치하는 연구 결과로서, 가장 낮은 수준의 인지적 능력을 가진 응답자와(Bishop 등, 1980a; Schuman & Presser, 1981; Schuman & Scott, 1989) 질문지의 문제를 개인적으로 덜 중요하다고 생각하는 사람들이(Bishop 등, 1980a; Schuman & Presser, 1981, pp. 142-143) 의견없음 선택 문항에 가장 잘 이끌린다. 하지만, 이들은 주제에 대해 정보가 없고 무관심 보고의 가능성이 높기 때문에, 이런 경향들은 Converse(1964)의 무관심 주장과 일치하는 것으로 볼 수 있다. 따라서 관련성의 의미가 명확해지려면 증거가 더 필요하다.

만족화 개념을 검증하는 최상의 연구는 아니지만 몇몇 연구들은 하나의 질문에 대한 의견없음 응답들의 빈도와 응답 작업의 어려움의 여러 측면들의 상관관계를 파악함으로써 이를 뒷받침하는 증거를 제공하였다. Klare(1950) & Converse(1976)는 질문 문구가 복잡할수록 의견없음 응답률이 증가하였음을 입증하였다. 또한, Converse(1976)는 긴 설명 문구들을 포함하는 경우, 응답자들에게 과거나 현재의 일을 간단히 기술하는 것보다는 장래의 일을 예측하도록 요구하는 경우, 다항식 질문보다는 이분법적 질문의 경우(아마도 후자가 중립이나 질적 의견을 기술하는 측면에서 볼 때 더 어렵기 때문), 국내문제에 대한 질문 보다는 외국문제에 관한 질문의 경우(아마도 후자가 더 생소하고 주제 관련 응답자의 지식 수준이 낮기 때문)에 의견없음 응답률이 더 높다는 걸 발견했다. 게다가, Ferber(1966)는(질문 순서를 실험적으로 조작하지 않은 조사에서) 의견없음 응답률과 질문지상의 위치 간에 아무런 관계도 발견하지 못했지만, Culpepper 등

(1992)은 질문 순서를 실험적으로 조작한 질문지에서 높은 의견없음 응답률이 나타났음을 관찰하였다.

Hippler & Schwarz(1989)는 만족화 이론을 뒷받침하는 또 다른 경험적 연구들을 하였다. 그들은 강력한 어조로 구성된 의견없음 여과질문은 태도질문에 응답하려면 많은 지식이 필요하다는 것을 암시하여, 최적화를 원하는 응답자들이 실제 응답을 할 때 의욕을 저하시킨다고 주장하였다. Hippler & Schwarz(1989)는 응답자들이 실제로 의견없음 여과질문이 존재하고 그 어조가 강력하면 다음 질문이 더 심층적이고 더 많은 지식을 요하며 더 어려울 것으로 추론한다는 사실을 밝혔다. 또한 의견없음 문항이 제공될 때 이것을 선택하는 응답자는 이후에 의견없음 문항이 없는 조사 질문지에서도 같은 응답을 할 수 있다. 게다가 이런 나중의 반응은 마음속으로 동전을 던져 정한 것이라기보다는 체계적인 것 같다.

의견없음 응답의 타당도에 의문을 제기한 또 다른 연구들은 주제에 대해 의견이 없다는 사람들의 비율을 이 주제에 대한 국민 의지의 강도(strength)를 나타내는 지수로 사용하였다. 그들은 국민들이 더 강력한 의지를 표할수록 정부도 그 의지에 부응하는 조치를 취한다는 가정 하에, 하나의 정책 문제에 관한 국민의 선호와 정부의 조치 간의 상관성을 검증하였다. 국민들이 특정 주제에 대해 강하게 느끼는 정도와 와 이 문제가 그들에게 얼마나 중요한지를 묻는 직설적인 질문들을 사용하여 이 사실이 확인되었다(예를 들면, Conover 등, 1982; Schuman & Presser, 1981). 그러나 의견없음 응답률은 여론 및 정부 조치 간의 상관관계와는 관련이 없다(Brooks, 1990). Brooks도 의견없음 응답이 정말 의견이 없다는 것을 나타낸다는 주장에 의문을 제기하였다.

4.5.7 결론

종합해 볼 때 이 증거들은 무관심을 미리 방지하기 위해 의견없음 여과질문을 태도질문에 포함시킬 확실한 근거가 되지 못한다. 여과질문을 포함시키면 진정한 태도의 일부를 얻지 못할 수 있다. 따라서 이 증거들은 가능하면 여과질문을 생략해라는 권고라고 할 수 있다. 대다수의 응답자에게 전혀 정보가 없는 생소한 문제에 대한 질문을 할 경우는 여과질문을 제시하는 것이 적절하고 유익할 것이다. 그러나 응답자에게 생소한 문제에 대한 질문은 연구자들에게 유익하지 않을 것이다.

Converse(1964)의 무관심 가설에 대한 본질적인 개념은 정확하다. 즉 조사에서 태도를 보고하는 대다수의 사람들은 자신들의 사고와 행동을 형성하는 선호를

갖고 있지 않다. 그러나 우리는 의견없음 여과질문이 이들을 구분하는 최상의 방법이라고 생각하지 않는다. 다시 말하면, 아마도 이들은 자신의 태도의 강도를 직접 측정하는 방법을 잘 찾아낼 수 있을 것이다. 이런 방법은 방대하고 다양한데 (예를 들면, Krosnick & Abelson(1992), Petty & Krosnick(1995) 참조), 방법의 선택은 실제로나 이론상으로 중요할 수 있다. 예를 들면, 응답자들에게 자신의 의견을 얼마나 확신하고 있는지, 특정 주제가 그들에게 개인적으로 얼마나 중요한지, 또는 본인이 이 주제에 대해 얼마나 많이 알고 있는지 질문할 수 있다. 무관심 문제를 다루기 위해 태도 강도에 대한 문헌을 검토하고 이 책에서 개발한 기술을 사용하기 바란다.

4.6 맺음말

본 절에서는 우리가 고찰한 설계 결정에 관한 최종 결론을 확실하게 내리지 않았다. 이에 대한 연구는 최종 결론을 정당화시키기에 아직 충분하지 않고, 연구간에 일치하지 않는 경우가 있다는 사실은, 몇 가지 상호작용 변수들이 다른 접근방법들보다 더 나은 결과를 산출하는 조건들을 결정할 수 있다는 것을 암시하는 것이다. 따라서 다음의 연구에서 이 문제들을 탐구하기 위해 할 일이 많이 남아 있다. 질문지에 관한 연구를 하는 학자들이 방법론적 실험들을 통합하기 위해 필요한 모든 자료들을 수집하는 기회를 놓치지 않기를 바란다. 우리가 측정 접근방법들 간의 차이점에 대해 알면 알수록 미래의 측정도구를 더욱 효과적으로 설계할 수 있다.

제 2 장

자 료 수 집

제 2 장 자료 수집(Data Collection)

제 5 절 자료 수집방법과 조사품질: 개요

Edith de Leeuw
Vrije Universiteit, Amsterdam
Martin Collins
City University Business School, London

5.1 서문

행정적 인구조사의 기원은 로마시대로 거슬러 올라간다. 하지만 우리가 알고 있는 조사(survey)의 역사는 그 보다 훨씬 짧은 약 100년 정도이다. 1886년, Charles Booth는 『Labour and Life of the People of London』이라는 자신의 연구를 시작하였으나, 이는 School Attendance Officers를 대상으로 학생들과 이들 가족의 생활여건에 대한 면접조사에 불과했다. 그리고 1900년경 Rowntree는 『Poverty: A Study of Town Life』라는 제목으로 뉴욕시민에 대한 연구를 시작하였다. Rowntree는 뉴욕에 거주하는 가구들로부터 직접 면담자료를 수집함으로써 Booth 보다 진일보한 조사방법을 보여줬다. 이것은 최초의 개별방문 면접조사(face-to-face interview survey) 중 하나였다.(Moser and Kalton, 1979 참조) 그러나 최초의 우편조사는 이보다 앞선 1880년 Karl Marx가 프랑스 노동자에게 보낸 설문지로 기록되어 있다.

오늘날의 조사방법은 1930년대부터 시작되었으며 이 때 세 가지 중요한 조사과정이 새롭게 개발되었다. 여기에는 美인구조사국(U.S. Bureau of the Census)에서 시도·개발한 새로운 표본추출기법과 표본추출방법과 Gallup이나 Roper와 같은 투표 및 여론조사기관의 자료 수집기법의 개발, 그리고 Stouffer나 Lazersfeld와 같은 개척자들이 개발한 조사자료 분석방법이 있다.(Babbie, 1973 참조)

1960년대 초기에는 또 하나의 중요한 단계인 전화조사가 소개되었다. 그리고 이후 1970년경에는 컴퓨터를 이용한 전화조사 CATI(Computer Assisted Telephone Interviewing) 시스템이 미국에서 개발되었는데, 이는 여론조사기관에서 주로 이용되었다.(Nicholls and Groves, 1986 참조) 1987년에는 이러한 전화조사 방법론이 국제회의에 도입되었으며, 이후 1988년에는 이에 관한 연구논문이 발표되었다.(Groves et al., 1988) 이러한 과정에서 Dillman(1978)은 우편조사의 신뢰성(respectability)을 증대시키고, 조사자료의 품질을 개선하는데 크게 기여하였다.

그 뒤로, 최근에는 자료수집기법의 개발에서 측정 및 측정오차로 관심의 대상이 옮겨 왔다. Groves(1989)는 조사오차(survey errors)와 가능원인(possible causes)에 대한 포괄적인 검토를 위해 계량심리학과 통계학 및 사회과학을 종합화하는데 많은 공헌을 하고 있다. Groves는 오차추정 및 조사오차와 조사비용의 산입을 강조한다. 1990년의 한 국제회의는 조사의 측정오차에 대한 연구만을 목적으로 개최되었는데 이에 대한 결과논문이 다음 해에 발표되었다. Lyberg와 Kasprzyk(1991)는 이 논문에서 자료수집방법과 측정오차에 대한 개론을 피력하였다.

Caesar Augustus와 Karl Mark에서부터 Lars Lyberg와 Daniel Kasprzyk에 이르기까지 오랜 역사를 살펴보는 동안, 설문조사자들은 새로운 도구의 사용과 과거 도구의 개선에 대해 많은 것을 배울 수 있었다. 그러나한 조사자의 결과물은 단순히 하나의 결과가 아니며 기술에서 중요한 변화가 다시 한 번 올 것이고, 사회는 이를 받아들일 것을 우리에게 종용하고 있다. 새로운 컴퓨터 기술의 발전은 여러 가지 새로운 형식의 컴퓨터를 이용한 자료수집을 가능하게 해 왔다. 비단 CATI 뿐만 아니라, 컴퓨터를 이용한 면접조사 CAPI가 현재 대규모로 사용되고 있으며, 새로운 형식의 컴퓨터를 이용한 자계식 조사 CASI가 조사실무자들에 의해 수행되고 있다. 이러한 새로운 기술들은 면접원 오차의 조절과 면접과정의 모니터링에도 영향을 미친다. 동시에 특수한 모집단에 대한 조사와 특수 집단(예 : 노년집단, 아동집단, 인종집단)의 요구수용에 표본조사방법을 택해야 되는 당위성에는 중요한 사회변화가 반영된다.

조사유형과 자료수집방법에서의 이러한 변화와 더불어, 우리는 조사자들이 조사오차를 바라보는 방법에도 변화가 있음을 알 수 있었다. 다른 과학적 현상에 대한 연구에서와 마찬가지로 여기에도 뚜렷한 단계들이 있다. 첫번째는 Deming(1944)의 연구에 나타나 있는 기술(description)과 분류(classification)의 단계이고 두 번째는 원인조사와 이론형성의 단계이며 전술한 Groves(1989)의 저서가 좋은 예가 된다. _마지막은 _초고(coping), 조정(controlling), 개선(improving)의 단계이다. 현재는 이 세 번째 단계로 이동 중에 있으며, 조사오차를 처리하는 -조사품질과 품질관리의 연구 및 실천- 쪽으로 관심이 변하고 있다.

본 논문의 자료수집에 관한 절은 조사자료의 품질과(Nicholls et al.-9장), 조사연구에 음성인식론의 응용(Blyth-10장), CAPI 조사에서의 면접원 성과 평가방법(Couper et al.-11장), 타당성 연구에서 면접원과 응답자 행동의 상호작용 코드부여(Dykema et al.-12장), 면접방식과 민감한 질문(Jobe et al.-

13장) 및 아동 응답자(Scott-14장)에 새로운 자료수집기법이 미치는 영향에 관한 내용을 담고 있다.

본 서문에서는 사회조사와 자료수집방법에서의 품질에 대한 논제를 제시할 것이다. 본 논문은 Lyberg와 Kasprzyk(1991)의 연구를 바탕으로 하고 있지만 논문 제목의 변화 -1991년 측정오차에서 1996년 조사품질로- 에서 알 수 있듯이 강조하는 바에 뚜렷한 변화가 있다. 8.2절에서는 먼저 품질의 개념에 대한 논의를 사회조사에 한정시켰다. 그런 다음, 논의의 대상을 전술한 두 가지 중요한 변화의 효과 -조사자료수집에 대한 컴퓨터 기술의 효과- 로 이동하고(8.3장) 다음으로 사회가 변화하는 과정 속에서 특수한 모집단과 특별한 주제에 대한 조사의 필요성을 강조하였다.(8.4장) 그리고 수 년간 그 중요성이 지속되어 온 하나의 주제인 면접과정 특히, 면접원 평가와 교육 및 면접원과 응답자 간의 상호작용에 대한 주제를 다루었으며(8.5장) 앞으로의 연구과제(research agenda)에 대한 논의로 끝맺음을 하였다.(8.6장)

5.2 사회조사에서의 품질

양질의 조사를 목표로 해야 하는 것에는 누구도 이견이 없을 것이다. 그러나 양질(high quality)의 정의는 정확히 무엇일까? 여기에는 절대적인 기준이 있을 수 있을까? 그리고 자료수집의 품질은 전체 조사모델의 어느 곳에 적합한 것일까?

이러한 질문들은 새로운 것이 아니다. 일찍이 Deming(1944)은 오차에 대한 검토에서 절대적 정확도(absolute accuracy)는 상상 속의 개념이며, 가능 오차의 허용한도나 한계에 대해 말하는 것이 보다 실리적임을 강조하였다. 그는 또한, 가용한 자원과 함께 특정 자료의 사용목적에 필요한 정확성에 따라 허용한계가 달라져야 한다고 주장하였다. 여기서 우리는 조사설계자(survey designer)가 가용한 자원과 조사목적을 모두 고려해야 될 필요성에 대해서는 앞서 언급하였다.

Deming은 또한, 조사설계자 앞에 놓인 임무의 복합성에 대해서도 경고하였다. 비교적 잘 알려진 표본추출의 변이성(variability)과 함께 우리는 궁극적인 조사의 유용성에 영향을 미치는 몇 가지 다른 요인들도 함께 고려해야 한다. Deming은 이에 대해 13가지 요인을 제시하였는데 여기에는 표본오차 및 편향과 면접원 효과, 자료수집방법, 무응답, 설문지 결함, 처리오차 그리고, 해석상의 오차 등이 포함된다.

게다가 Hansen et al.(1953, 1961)의 조사오차 모델은 각 모집단 요소에 주어진 변수의 개별 참값(individual true value)이라는 개념에 중점을 두었다. 이 모델은 변수오차(variable error)와 계통적 편향(systematic bias)간에 차이를 두고 있으며 총오차(total error)의 개념을 제시하고 있다.

Kish(1965)는 총오차를 제곱근평균제곱오차(root mean square error)로 정의하고, 단순자(simpler)를 대신하는 것이지만 표준오차의 불완전한 개념으로 제안한다. Kish는 또한 비관측(예 : 무응답)과 관측(예 : 자료수집과 처리과정에서)의 오차를 구별하면서 조사자료에서 편향의 원천을 분류하였다.

조사오차에 대해서는 서로 다른 용어와 체계를 사용하는 다양한 분야 - 표본통계학, 심리학, 경제학 - 에서 논의되어 왔다. Groves(1989)는 이러한 다양한 시각들을 조정하고, Kish의 분류학에서 개발된 일반화된 분류방법을 통해 분야간 의사소통의 문제를 줄이려고 하고 있다. 중심에 서 있는 것은 평균제곱오차(mean squared error) - 오차분산(제곱가변오차)과 제곱편향(squared bias)의 합 - 의 개념이다. 가변(임의)오차(variable(random) error)는 동일한 표본의 반복시행에서 나타나는 서로 다른 값을 의미하는 반면, 편향(비임의 오차)(bias)은 그러한 반복 시 일정한 상수(constant)로 관측되는 오차 성분들이다. 두 가지 오차유형은 모두 비관측 오차(포함, 무응답, 표본추출)와 관측오차로 나눌 수 있다. 후자는 또한 응답오차 또는 측정오차로도 불리며, 응답자, 면접원, 도구 및 자료수집방식으로부터 발생될 수 있다. Groves의 체계에는 처리오차와 코딩오차는 포함되어 있지 않다. 하지만Groves는 오차 평가를 위한 설계를 검토하고, 가능 오차원인에 관한 여러 학계의 문헌을 요약하였다.

Groves에 있어 한 가지 중요한 측면은 조사품질의 논의에 비용을 산입시키려 한다는 사실이다. 이러한 비용산입을 모두가 인정하는 것은 아니기에 그는 다음과 같은 정당성을 제시하고 있다. "비용문제를 포함시키는 것은 조사자료의 품질에 대한 논의와 대조를 이루는 것으로 보일 수도 있다. 심지어 가장 나쁜 의미에서는 '안일하게 대처하는' 사고를 불러 일으킬 수도 있다. 이 책은 고비용과 저비용의 이러한 반작용의 혼돈에 대해서 논하고 있다. 따라서 여기서 취하고 있는 입장은(품질을 측정하는 것과 함께) 선택적 방법들의 비용을 정형적으로 산정하는 것만이 '최량의' 방법을 도출할 수 있다는 것이다."(Groves, 1989, p.6) 이러한 주장속에서 Groves는 비용이 조사설계자에게 하나의 제약조건이고 또한, 효용성은 어쩌면 설계에 있어서"품질"의 구성요소가 될 수도 있다.

비용은 조사설계에서 고려될 수 있는 수 많은 쟁점들 중 하나일 뿐이다. 다른 중요한 요인들은 어쩌면 정보의 무한성이나 적시성일 지도 모른다. 이를 위해서는 주어진 조사문항과 특정한 제약조건에서 최적의 자료수집방법을 위해 조사에서 고려해야 하는 다른 고려사항들의 범위를 지정할 수 있다. 이것은 종종 고객의 관점에 달려 있기도 하다.(Dippo, 20장 참조) 기본적인 조사질문은 대상 모집단과 질문의 유형을 정의한다. 조사의 윤리성과 사적인 규제에 대한 질문들은 설계를 제약할 수도 있다. 중요한 실제적 제약사항은 가용한 시간과 자금이며, 기타 자원들(기술력, 행정자원, 경험, 컴퓨터 하드웨어 및 소프트웨어)이다. 이러한 제약조건 속에서는 허용 가능한 무응답 수준이나 측정오차에 대한 결정들이 신중하게 내려져야 한다.

이와 같은 고려사항들이 조사설계를 반드시 제약한다는 사실은 쉽게 알 수 있다. 예를 들면 Alwin(1977, p.132)은 조사오차를 검토하는 과정에서 "조사자료의 최종 품질에 영향을 미치는 결정들은 대개 이론적 또는 방법론적인 것보다는 오히려 실제적인 고려사항의 영향을 받는다."는 사실을 인지하고 있다. 여기서 우리는 다음과 같은 하나의 또 다른 관점을 제기할 수 있다. 자원의 효율적인 활용은 조사 접근방식 그 자체의 "품질"로 간주되어야 한다. 조사과정을 가용 자원과 일치시키는 것은 설계자에 대한 하나의 제약으로 바라볼 것이 아니라, 설계에서 고려되어야 하는 필수적인 품질로 간주해야 한다는 오캄의 면도날 이론(Occam's razor)에 대한 설문조사자들의 해석을 믿는다.

자료수집 품질에 이러한 체계를 확립함에 있어, 우리는 자료의 품질은 의도된 자료의 용도에 따라 요구되는 품질을 반영해야 한다는 Deming의 지적을 상기해야 한다. 또한, 자료수집의 품질은 관심(attention)과 자원(resources)에 대해 나머지 모든 측면과 경합하는 전체 조사설계의 한 측면일 뿐이라는 사실도 인지해야 한다(본서 20장, O'Muircheartaigh, Introduction and Dippo 참조)

5.3 자료수집 수단과 전산화

5.3.1 기술적 변화: 자료수집에 미치는 영향

조사자료를 수집하는데 있어 컴퓨터를 이용한 자료수집 방법이 사용되어야 하는 지에 관한 사항은 더 이상의 논쟁의 대상이 아니다. 대부분의 전문조사기관들은 현재 최소 한 가지 이상의 컴퓨터를 이용한 자료수집 방식을 채택하고 있다. 이와 같은 새로운 기술의 활용은 완전히 새로운 약어들의 절정을 이끌었다. 다음에 나오는 분류법은 현재 주로 사용되는 약어와 용어에 대해 소개할 것이다. 아울러 자료 품질에 새로운 기술이 미치는 영향에 대해서는 Nicholls 등(9장)의 견해가 제시될 것이다.

자료수집의 방법이나 수단(예, 면접, 자계식 설문지)과 자료를 포획하는데 사용되는 기술은 서로 구별하는 것이 중요하다. 예를 들어 개별방문조사에서 면접원은 종이와 연필 또는 녹음기나 컴퓨터를 사용할 수 있다. Lyberg와 kasprzyk(1991)는 자료수집방법을 고찰하면서, 다음과 같은 조사에 6가지의 종이조사표식 자료수집 수단을 검토하였다. 개별방문면접조사, 전화조사, 자계식 설문지와 일기장식 조사, 관리용 기록(administrative records)의 이용과 직접관찰. 전산화된 등가방식(equivalents)은 모든 방법에 존재한다. 컴퓨터를 이용한 전화조사와 같은 방법들은 광범위한 곳에 비교적 오랜 기간 확산되어 있으며, 전자데이터교환(electronic data exchange)과 같은 다른 방식들은 여전히 새로운 것이다.

컴퓨터를 이용한 설문조사 방법들은 대개 CADAC(Computer Assisted Data Collection), CASIC(Computer Assisted Survey Information Collection), CAI(Computer Assisted Interviewing)와 같은 세계 공통의 용어들로 요약된다. 이러한 맥락에서 볼 때, 전통적인 종이조사표식 방법들은 PAPI(Paper And Pencil Interviewing)로 표현할 수 있다. 표 5.1은 여러 가지 컴퓨터를 이용한 면접기법에 대한 체계적인 개요를 보여준다.

표 5.1 컴퓨터를 이용한 면접방법의 분류

특정 방법	컴퓨터를 이용한 방식
개별방문면접	CAPI(Computer Assisted Personal Interviewing)
전화조사	CATI(Computer Assisted Telephone Interviewing)
서면 설문지 포함	CASI(Computer Assisted Self-administered Interviewing) 동의어 CSAQ(Computerized Self-Administered Questionnaire), PDE(Prepared Data Entry)
우편조사	DBM(Disk by Mail)과 EMS(Electronic Mail Survey)
패널조사 및 일기장식	CAPAR(Computer Assisted Panel Research), Teleinterview, Electronic diaries
자계식(면접원 없음)	TDE(Touchtone Data Entry), VR(Voice Recognition), ASR(Automatic Speech Recognition)
직접관찰	다양한 기법: people meters, TV appreciation panels, barcode scanning, computer assisted video coding
관리용 기록 (Administrative records)	Electronic data exchange, data mining

일반명: CADAC(Computer Assisted Data Collection), CASIC(Computer Assisted Survey Information Collection), CAI(Computer Assisted Interviewing)

컴퓨터를 이용한 전화조사(CATI)과 개별방문면접조사(CAPI) 방식은 이미 잘 알려져 있어 별도의 소개가 필요없는 방식들이다. 컴퓨터를 이용한 자계식 설문조사방식은 비교적 범용화되지 않은 방식이다. 여기에 대해서는 Nicholls 등(9장)이 간략히 소개하고 있으며, Blyth(10장)는 자동음성인식(Automatic Speech Recognition) 기술의 발전에 대해 전체적인 설명을 제공하고 있다.

관리용 기록이나 직접관찰을 이용한 자료수집에 관한 전자식 등가방식은 위 방법들과 마찬가지로 그다지 많이 알려져 있지 않다. 특히 직접관찰법에 대한 전자식 방법은 현재도 급속히 발전 중이다. 미디어 조사에서 "피플미터(people meters)"는 시청행위를 기록하는데 이용되고, 소비자 조사에서 자동화된 바코드 검색은 구매행위에 관한 데이터를 제공하며, 몇몇 통계기관은 휴대용 컴퓨터를 가진 숙련된 관찰자들을 이용해 월별 소비자 물가지수를 산출할 자료를 수집한다. 행동과학에서는 인간과 동물의 행태를 관측할 수 있는 정교한 도구가 개발되었다. 예를 들면, 현지에서 개개의 사건을 정확한 시간과 함께 기록할 수 있는 이벤트레코더(event recorder)가 있으며, 통합컴퓨터코딩(integrated computer assisted coding) 및 비디오시스템은 자세한 행동 코드부여와 이어지는 데이터 분석을 위해 개발되었다.(Noldus et al., 1989; Noldus, 1991 참조)

조만간 관리용 기록들도 전자식으로 바뀔 것이다. 통계기관에 의한 조사수립과정에서도 전자식 데이터교환은 관리용 데이터를 수집하면서 빠르고, 안정된 방법이 된다. 실제로 사회과학분야에서는 보관된 데이터가 전자식 수단을 통해 교환된 지 이미 오래다. 과거에는 데이터를 테이프로 운반하였다. 현재는 고퍼(Gopher)나 웹 브라우저 - 인터넷 사용자들이 전세계 정보를 검색할 수 있는 메뉴방식의 패키지 - 를 통해 라이브러리와 카탈로그를 검색할 수 있다. 뿐만 아니라 이러한 패키지들은 FTP(File Transfer Protocol)를 이용해 원격 컴퓨터에 액세스하여 데이터 파일을 검색할 수도 있다. (Crispen, 1995 참조) 예를 들면, 암스테르담에 있는 Steinmetz Archive는 인터넷을 통해 이러한 서비스를 제공한다. 그리고 노르웨이의 데이터 보관소인 NSD는 유럽의 모든 데이터 보관소의 중앙 액세스 또는 "탑 메뉴"의 역할을 한다. 실제로 메뉴와 문서들이 링크되어 있는 월드 와이드 웹 서버를 통해 누구나 유럽의 모든 데이터 보관소와 그 밖에 유럽 외 대부분의 데이터 보관소의 고퍼 및 웹 서버에 접속할 수 있다.

5.3.2 모드(Mode) 효과 비교

특정한 하나의 관리모드를 선택하는 것은 조사과정의 일부분이지만, 이것은 전체 조사설계에 중요한 영향을 미칠 수 있다. Schwarz(1장)는 설문지 작성시

특정한 선택사항의 결과에 대한 포괄적인 검토를 하였는데 특히, 자료를 제시할 때 의사전달 방법(청각 및 시각)과 시기상의 순서(순차적 접근과 임의적 접근)의 영향에 대해 각별한 관심을 가졌다.

Hox와 de Leeuw(1994)는 우편, 전화, 개별방문조사와 관련된 무응답을 연구하였다. 이들은 우편이나 전화, 개별방문을 이용하여 얻은 응답을 명확하게 비교하는 45가지 연구에 메타분석(meta-analysis)을 적용하였다. 완성도면에서 볼 때, 평균적으로 개별방문면접이 가장 높은 결과(70%)를 보였으며, 다음으로 전화조사(67%), 그리고 우편조사(61%)가 가장 낮았다. 이러한 차이는 Goyder(1987)의 관찰과도 일치한다.

자료품질에 대한 자료수집방식(모드)의 영향은 전통적인 지면을 이용한 개별방문면접과 전화조사 및 자계식(우편) 설문지를 통해 광범위하게 연구되어 왔다. De Leeuw(1992)는 이러한 자료수집방법의 실험적 비교에 대한 메타분석을 제시한다. De Leeuw는 우편조사를 전화조사 및 개별방문면접과 비교하면서, 우편조사의 경우, 질문에 응답하도록 하는 것이 다소 더 어렵다는 사실을 발견하였다. 즉, 전체 무응답과 항목별 무응답이 우편조사에서 더 높게 나타났다. 그러나 질문에 응답한 경우, 응답결과 자료의 품질은 더 나은 경향을 보이는 것으로 나타났다. 우편조사는 특히, 보다 민감한 질문에 대한 결과가 좋게 나타났다(예 : 음주행위에 대한 응답률이 높거나, 소득에 대한 무응답율이 낮게 나타남). 개별방문조사와 전화조사를 비교할 때는 미미한 효과만 발견되었다. 즉, 개별방문면접의 경우, 항목별 무응답이 약간 적은 것으로 나타날 뿐, 응답 타당성(기록체크)과 사회적으로 바람직한 응답에 대해서는 아무런 차이가 없었다. 일반적으로 이와 유사한 결론은 성공적으로 수행된 개별방문면접조사와 전화면접조사에서 도출될 수 있다(de Leeuw, 1992). 이러한 결론은 개별방문조사와 전화조사를 비교한 연구에서 가장 일정하게 나타나는 결과는 두 조사 간에 별다른 차이가 없다는 점이라는 Groves의 주장과 일치한다(Groves, 1989, p.551). De Leeuw의 모드효과에 대한 메타분석은 면접기술(예, 지면을 사용한 면접과 컴퓨터를 이용한 면접)의 잠재적 영향 조사에까지 확대되지는 않았다. Jobe 등(13장)은 면접방법과 기술이 민감한 질문의 응답에 미치는 영향에 대해 잘 정리된 연구논문을 발표하였다.

Schwarz 등(Schwarz et al., 1991)의 연구에 의한 영향으로 자료수집방식이 문맥(context)과 응답순서에 미치는 영향에 대한 관심이 이끌어졌다. 지금까지는 우편조사와 전화조사 사이의 차이점에만 초점을 맞추는 한정된 수의 비교분석만 행해져 왔으나, 이러한 연구들은 엄격한 순차적 질문제시로 인해 면접조사시

문맥효과(context effect)가 순서에 종속적임을 보여주었다. 그러나, 우편조사나 자계식 설문지에서는 문맥효과가 순서적이고 종속적이지 않다. 사실, 자계식 설문지에 대한 몇몇 연구에서는 후순위 질문이 선순위 질문에 대한 응답에 영향을 미쳤다(Schwarz, 1장 참조). 『청각만을 이용하여 의사를 전달한(예 : 전화조사, showcard를 이용하지 않은 면접조사) 몇 가지 면접조사에서는 최신효과(recency effect)가 발견된 반면(즉, 마지막 문항에 대한 응답율이 높음), 자극적인 시각적 설명을 이용한 방식에서는 초두효과(primacy effect)가 (즉, 최초 문항에 대한 응답율이 높음) 보다 일반적으로 나타났다.(Schwarz and Hippler, 1991 참조)』

몇 가지 컴퓨터를 이용한 자료수집방법(CADAC)의 직접적인 방식비교는 매우 드물다. 대부분의 문헌은 동일한 자료수집에 지면을 이용하는 방식과 컴퓨터를 이용하는 방식에 대한 비교만을 하고 있다. 이러한 사례에서 보면, 자료수집방법에 대한 서로 다른 기술의 영향들이 비교되어 있다(Nicholls et al., 9장 참조). 가까운 미래에는 새로운 컴퓨터 기술이 성행할 것이고, 이러한 컴퓨터를 이용한 방식들에 대한 비교가 가능할 것이다(즉, 모두 컴퓨터를 이용한 방식). 하지만, 그 때까지는 종이조사표 이용 조사와 컴퓨터를 이용한 조사간의 비교 결과만을 이용할 수 있을뿐이다.

전화조사 응답자의 경우에는 조사기관이 종이조사표 전화조사에서 CATI로 전환한다 해도 아무런 변화가 없다. 그러나, 면접원의 경우에는 관리업무가 컴퓨터에 의해 수행되기 때문에 업무의 복잡성이 훨씬 덜어진다. 그러므로, 차이가 있다면, 경로오차(routing error)가 적은 이유로 CATI에 약간의 장점이 더 있다고 할 수 있겠다. CAPI에서는 응답자 앞에 컴퓨터가 놓여 있어 응답자는 컴퓨터가 제시하는 것에 반응해야 한다. 그러나, 이 때에도 역반응(adverse reaction)은 거의 없었으며, 응답율의 감소도 나타나지 않았다. 따라서, 전화와 개별방문조사 간의 방식의 차이점에 관한 주된 결과는 이러한 조사기법을 컴퓨터를 이용한 경우에도 타당성이 있는 것으로 가정할 수 있겠다. 이것은 숙련된 면접원과 잘 작성된 동일한 설문지가 있다면, CAPI와 CATI는 모두 성공적으로 수행될 것이고, 자료품질의 차이 또한 미세할 것임을 의미한다. 물론, 종이조사표 개별방문면접이 동일한 방식의 전화조사에 비해 더 많은 가능성을 가지고 있었던 것처럼(예, 시각적인 자극과 카드 분류와 같은 특수한 기법의 활용) CAPI가 CATI보다 가능성이 더 많은 것임은 기억해야 한다.

컴퓨터를 이용한 자계식 설문지에는 몇 가지 형식이 있다(CASAQ, 8.3.1 참조). 이 형식들은 모두 화면을 통해 질문을 읽고, 컴퓨터에 응답을 입력한다는

공통점을 가지고 있다. 지면을 이용한 자계식 설문지와 마찬가지로, 응답자는 개인적인 상황에서 질문에 답하게 되고, 이것은 유리한 측면으로 자신을 내보이려는 의도를 감소시킨다. 민감한 질문에서 CASAQ가 CAPI보다 사회적으로 바람직한 응답을 잘 유도해 내지 못한다는 사실에는 몇 가지 증명된 바가 있다(Beckenbach, 1995). 뿐만 아니라, CASAQ에서는 면접원은 아니지만 응답자는 질문의 속도를 조절한다. 그러나, 응답자가 이를 조절하는 유일한 위치에 있는 것은 아니다. 즉, 질문의 순서와 주어진 시간에 보여지는 질문의 수는 컴퓨터 프로그램이 조절한다. 일반적으로 응답자는 종이조사표 설문에서와 같이 앞뒤의 질문을 마음대로 오갈 수 있는 상황이 아닌 것이다. 이러한 의미에서 CASAQ는 자계식 설문보다는 면접조사와 보다 유사하다 하겠다. 그러므로 질문의 순서에 따라 발생하는 문맥효과에서의 차이점은 더 이상 존재하지 않고, 질문을 제시하는 방법상의 차이(청각 대 시각)로 인한 효과는 지속될 것이다. 그리고, CASAQ는 이미 알려져 있는 문맥순차효과에 관해서는 CATI 및 CAPI와 유사할 것이다.

5.4 특수 모집단의 조사

5.4.1 사회적 변화와 전문화된 조사의 필요성

컴퓨터의 도입은 자료수집기술을 변화시켰다. 그리고, 현재는 인구통계와 사회의 주요한 변화로 인해 설문조사자들은 다른 측면으로 조사를 재설계해야 되는 상황에 놓여 있다. 전문화된 조사는 특별하거나 민감한 주제(예, AIDS, 성행위, 출산력)에 대한 조사든지 혹은, 특수한 모집단(예, 유년집단, 노년집단, 인종집단)에 대한 조사든지 조사산업에서 그 중요성을 점점 더 해가고 있다. 살펴보면, 전문화된 조사에 관한 일반적인 방법론적 문헌은 드물지만, 특정한 부모집단(subpopulation)에 대해서는 보다 많은 사실이 알려져 있다. 하나의 좋은 예가 노년층에 대한 조사다. 1960년대와 1970년대를 주도하였던 베이비 붐 세대는 이제 내일의 노년층이 되어 가고 있다.

선진국의 경우 65세 이상의 노년층은 1990년 12%에서 2025년에는 19%로 증가할 것으로 예상된다. 노년층은 비교적 동질적이고, 뚜렷이 증가하는 집단을 형성하므로, 조사자들은 조사방법의 개발에 많은 투자를 해 왔다. 노년층 조사방법에 대한 요약은 Herzog와 Rodgers(1992)를 참조한다. 이민집단 및 문화적, 인종적으로 다양한 집단을 바라보면, 서로 다른 양상이 나타난다. 각기 다른 나라에는 문화적으로 서로 다른 다양한 집단이 존재하고, 심지어는 한 나라에서도 각기 독특한 자신만의 문화를 가진 몇몇 구별되는 집단이 있다. 이러한 특수집단에 대한 조사방법은 알려져 있기는 하나, 산재되어 있고, 특정한 상황과 집단에 한정되어 있다. 따라서, 다음 절에서는 이러한 다양한 지식을 하나로 통합하고, 특수한 조사에 대한 일반적인 체계를 제시하고자 한다.

5.4.2 전문화된 조사의 설계

특수한 조사를 성공적으로 설계 및 실천하기 위해서는 조사목적을 분석하는 것이 무엇보다 중요하다. 이 특수한 조사로부터 원하는 것은 무엇인가? 기본적인 조사의 질문을 특별하게 만드는 것은 무엇인가? 지금 연구하는 집단은 왜 특별한가? 자동적으로 표준화된 개별방문면접으로 방향을 돌리고, 문제를 면접원에게 남기는 것만으로는 충분하지 않다. 특수한 모집단은 특별한 조사를 필요로 한다. 쟁점의 분석은 예를 들어, 설문지 작성이나 면접원의 선발 및 교육과 같은 조사의 일면에만 집중되는 것이 아니라, 전체 조사설계를 포괄하고 있어야 한다. 설계과정의 핵심 단계들은 다음과 같다. 특수한 모집단과의 친밀감을 형성한다. 양질의 조사를 위한 일반적인 지침을 도입한다. 특수한 모집단의 가능한 효과면에서 각 단계를 분석한다. 이에 따라 각 단계를 조정한다. 그리고, 식견이 높은 동료집단 및 중요 정보제공자들과 전체 조사설계의 모든 면에 대해 논의한다.

이 과정이 끝나고 나면, 조사설계의 전 단계들이 목적에 맞게 다듬어 질 것이다. 이것은 통상적인 표준조사에 비해 추가적으로 투입되는 작업이지만, 그다지 어렵지는 않다. 본 절의 다음 부분에서는 자료수집을 강조하면서 특수한 조사설계의 각 단계에 대해 논의하겠다. 여기서, 노년층 조사에 대한 예는 Herzog와 Rodgers(1992) 및 Broese van Groenou 등(1995)으로부터, 유년층에 대한 조사예는 Scott (14장)로부터, 그리고 소수인종(특히, 네덜란드에 있는 북아프리카 인종집단)에 대한 예는 Meloen과 Veenman(1990) 및 Schothorst(1993)로부터 인용하였다.

1단계: 대상 모집단을 정의하고, 이 집단을 특수하게 하는 것이 무엇인지 분석한다.

Scott(14장)는 유년층의 면접조사에 적절한 방법론은 반드시 이들의 연령을 고려하는 것임을 강조하였다. 서로 다른 연령집단은 인지 발달 뿐만 아니라, 감성 발달과 질의응답과정의 기초가 되는 인지력 및 의사전달 과정에 필요한 사회적 기술의 습득에 있어서도 서로 다른 양상을 보인다(Schwarz, 1장 참조). 연령층이 낮을수록 조사가 무엇인지 혹은 ***적절한*** 응답은 어떻게 나타내는지에 대해 알기가 힘들다. 취학전 아동에 대한 조사에서는 초등학생이나 사춘기 아동과는 완전히 다른 접근방식이 필요하다.

마찬가지로, 노년층에 대한 조사에서도 대상 모집단을 정의하는 것이 중요한다. 전기노년층(older adults)은 응답에 영향을 미치는 인지력이나 건강상의

문제가 있을 가능성이 많다. 이러한 문제들은 후기노년층(older old) (80세 이상)에서는 보다 보편적이다. 또, 인종집단을 조사할 때는 대상 모집단의 정의에 따라 다른 문제들이 발생한다. 예를 들면, 출생국이 기준변수일 때는 언어문제와 조사과정에 대한 무지의 문제에 종종 직면하게 된다. 그러나 부모의 출생국이 핵심 정의이고, 2세대 응답자들이 면접을 볼 때는 이러한 문제가 덜 나타난다.

2단계: 방식(Mode) 선택

대개는 개별방문면접이나 자계식 형식을 특별히 적응시킨 방식이 이용되며, 민감한 주제나 특수한 상황에서는 특히 더하다. 전화조사는 대상 표본을 찾기 위해 대규모 표본을 조사하는데 이용된다. 방식의 선택은 결코 간단한 문제가 아니며, 특수한 집단을 조사할 때는 반드시 표준방식이 적응되어야 한다. 이러한 적응은 성공적인 특수 조사에 핵심적인 열쇠가 된다. 예를 들어, 면접원들은 특수 집단의 관습에 특별히 숙련되어 있거나 예상되는 문제를 처리할 수 있어야 한다. 자계식 형식의 몇 가지 적응 사례로는 농아에 대한 조사에서 수화통역자와 함께 서면 질문지를 관리하는 것, 아동층에 대한 조사에서 "워크맨"과 함께 설문지를 제공해 질문 내용이 큰 소리로 읽혀지도록 하는 것, 그리고 성적 접촉에 대한 조사에서는 응답자에게 설문지를 소인이 찍힌 봉투와 함께 건네고, 가장 가까운 우편함까지 함께 걸어가는 것과 같은 복잡한 보안조치의 사용 등이 있다.

3단계: 표본추출

일반적인 조사에서는 가구(household) 표본이 가장 일반적으로 이용된다. 그러나, 특수한 모집단을 연구할 때는 잘 사용되지 않는다. 한 가지 예로 표본가구의 아동에 대한 Scott의 조사가 있다. 다른 경우로, 우리는 매우 다양하고, 대개는 창조적인 표본추출기법을 발견하는데, 이들 대부분은 비확률(nonprobability) 표본추출법이다. 이러한 예들은 기존의 목록표본체계를 이용하여 무작위표본선별(random sample screening) 단계에서부터 네트워크 표본(network sampling)과 포획-재포획 표본(capture-recapture sampling) 기법에 이르기까지 다양하게 걸쳐있다. 이 중 Sudman 등(1988)은 대체로 희귀 모집단의 표본추출에 대해 논하였다. Kalsbeek(1989)은 AIDS 연구를 위한 표본추출기법을 검토하였다. Czaja 등(1986)은 암환자 발견을 위한 네트워크 표본추출을 논하였다. Kalton과 Anderson(1989)은 노년층 연구에 도움이 되는 표본추출기법을 검토하였다.

4단계: 설문지

설문지는 반드시 기본적인 지침을 따라야 하는 것은 사실이지만, 예컨대 노년층이나, 혹은 언어장애를 겪을 수 있는 집단을 조사할 때 단순한 언어를

사용하고, 명확한 구조를 유지해야 되는 것과 같이 이 중 일부는 특히 중요한 사항일 수가 있다. 그러나, 관례적인 조사지침이 특정 집단에 적절하지 않을 때는 추가적인 적응과정이 필요할 수도 있다. 예를 들면, 다음과 같이 표준적인 평정기준이 적절하지 않을 수도 있다. 네덜란드에 거주하는 북아프리카 이주민에 대한 연구는 5점 척도가 yes/no로 응답되며, 따라서 2점 내지는 많아야 3점 척도의 사용이 바람직하다고 권고한다. 또한, 노년층 응답자의 경우에는 일련의 순서를 요구할 수 있어 7점 내지 9점 척도가 바람직할 수 있다. 또 다른 예를 들면, 표준적인 카드제시(showcard)는 면접원의 내용설명이 없으면 어쩌면 아무런 도움이 되지 못할 수도 있다는 것이다. 일례로, 북아프리카 이주민 집단의 1세대는 문맹이 많았으며, 따라서 노년층 응답자들은 글을 읽는데 어려움을 가질 수 있다.

질문에 사용된 용어와 질문순서에 대한 사전조사는 매우 중요하다. 이것은 검증된 표준조사 질문이라 하더라도 특정한 특수집단에는 적절하지 않을 수도 있기 때문이다. 예를 들어, 노년층은 연령이 낮은 집단에 비해 자신들을 다른 사람들과 비교하길 꺼린다("다른 사람과 비교할 때 당신의 나이는…") 또, 독실한 종교집단들은 대개 미래에 관한 질문에 답하려 하지 않는다 ("이것은 내가 답할 문제가 아니라, 신이 결정할 문제다").

설문지에서는 핵심적인 개념을 정의하고, 노년층이나 장애인을 조사할 때는 이 핵심적인 내용을 함축되게 표현한 설문지를 준비해 두는 것이 바람직하다. 왜냐하면, 건강상의 문제로 인해 긴 설문지에 답하기 어려운 경우에는 짧은 설문지를 응답자나 대리응답자에게 제시할 수 있기 때문이다. 비모국어 사용자를 대상으로 할 때는 면접원이나 통역자의 통역능력에 의존하기 보다는 번역된 설문지를 이용하는 것이 바람직하다. 컴퓨터 면접은 이러한 과정을 매우 간단하게 해 준다.(Bond, 1991 참조)

5단계: 면접원 교육 및 선발

소수인종을 연구할 때는 이중언어 뿐만 아니라, 이중문화를 가진 면접원을 채용하는 것이 바람직하다. 또한 면접원과 응답자의 성별이 동일하게 하는 것도 종종 필요하다. 만일 전문 면접원을 대규모로 채용하는 것이 불가능하다면, 특별한 경우에 투입될 수 있는 소수의 이중문화 및 이중언어 면접원을 최소한 교육이라도 시켜야 한다. 노년층 연구에 가장 효과적인 면접원 유형이나 연령은 없어 보인다. 하지만, 노년층은 피해자화(victimization)에 더 많은 관련이 있기 때문에 면접원은 매우 신뢰감을 주는 인상을 가져야 한다는 것이 결론이다.

모든 면접원은 기본적인 면접원 교육을 당연히 받아야 한다(예, Institute for Social Research, 1976, 1983; McCrossan, 1991). 그리고 추가적으로는 특수한 집단에 직면하는 문제들에 관한 교육이 바람직하다. 예를 들면, 노년층의 청각 또는 시각장애를 어떻게 극복해야 하는가, 노년층이 주제에서 벗어나는 것은 어떻게 방지하는가, 추가적인 면접원 지원이 필요할 때 응답자의 응답에 영향을 미치지 않으려면 어떻게 해야 하는가, 혹은 타인(남편, 자녀, 간호사)의 존재에 어떻게 대처하는가 등이다. 문화가 다른 집단을 조사할 때는 이중문화를 가지지 않은 면접원을 위해 문화적 차이에 대한 단기간의 교육과정도 필요하다. 관습적인 부분이 다를 수 있고, 소개를 보다 길게 해야 될 수도 있으며, 가장이 없는 아내와 자녀들은 면접하기 어렵거나 심지어 불가능할 수도 있고, 특정한 날이나 시간대에는 면접을 하지 못하는 경우도 있다. 결국, 일반적인 면접규칙 이면에 있는 조사방법과 이론에 관한 일부 배경지식은 비표준화된 상황에 도움이 되고, 기준을 벗어나야 되는 상황에서 최적화된 행동을 할 수 있게 해 준다.

특수한 모집단과의 면접에서 경험할 수 있는 어려움으로 인해, 면접원에게는 보다 강도 높은 감독이 필요하다. 이것은 보다 많은 접촉 뿐만 아니라, 조사감독원들의 보다 많은 지도가 필요함을 의미하는 것이다. 조사감독원들은 면접원들에게 좋은 지도자가 되어야 함은 물론, 힘들고 과중한 부담에 대처할 수 있는 가르침을 줘야 한다. 이것은 현지에서 겪게 되는 문제를 최소화하고, 면접원의 이직을 방지하며, 현지조사의 효율성을 증가시킬 것이다.

요컨대, 특수 모집단의 성공적인 조사에는 잘 다듬어진 양질의 조사가 요구된다. 적절한 사전조사와 더불어, 집단과 조사를 특별하게 만드는 것이 무엇인가에 대한 폭 넓은 논의와 분석은 특이한 어려움과 해결책을 발견하도록 해 준다. 그리고 무엇보다 가장 중요한 것은 성공적인 적응사례와 경험들의 공유가 필요하다는 것이다.

5.5 면접원 평가 및 교육

5.5.1 면접원 오차

기술과 사회의 변화는 자료수집의 변화를 가져왔지만, 한 가지 주제 즉, 면접원의 품질과 평가 및 교육에 대한 것은 여전히 높은 요구수준을 유지하고 있다. 면접조사에는 전화조사든 개별방문면접이든, 컴퓨터를 이용하든 조이조사표를 이용하든 두 종류의 중요한 행위자인 면접원과 응답자가 있다. 이 중 면접원은 설문조사자의 통제를 가장 직접적으로 받으며, 조사자는 이러한 면접원의 오차를 측정하고, 이를 최소화하려는 노력을 오랫 동안 해 왔다. 이 때

사용된 기본 전략들로는 면접원 선발과 교육, 수행능력감시 및 오차의 위험성을 감소시키는 확실한 설문지와 조사절차의 설계가 있다. 최고의 면접원이라 하더라도 질문이 이해하기 어렵고, 조사자의 기준절차 제시가 없을 경우엔 오차를 발생시키기 마련이다. 이러한 면접원 관련 오차를 감소시키는 방법에 대해서는 Fowler(1991)의 좋은 논평이 있다.

본 논문에서도 몇몇 저자들은 설문지 계획의 측면에서 면접원 오차를 감소시키기 위한 방안들에 대해 논의하고 있다. 예를 들면, Krosnick과 Fabrigar(6장)는 평정척도에 관한 문헌을 검토하였고, Jenkins와 Dillman(7장)은 설문지 계획에 대해 견해를 피력하였다. 또, Nicholls et al.(9장)은 컴퓨터를 이용한 자료수집기법이 면접원 오차를 방지하는데 어떤 도움을 줄 수 있는 지에 대해 논하였고, Couper et al.(11장)은 면접원의 수행능력을 감시하는데 어떤 기술들이 새롭고 유용한 도구를 제공하는 지를 보여주었다.

효율적인 오차감소전략을 위한 선행조건은 면접원 오차에 영향을 미치는 면접원의 개성과 수행능력에 대한 지각 및 탐지방법의 유용성이다. 따라서, 다음 절에서는 면접원 오차를 통계적으로 탐지하는 방법과 면접원 교육에 준한 감시방법 및 면접원 오차를 탐지 및 기술하는 하나의 도구로서 상호작용 코드부여(interaction coding)에 대한 논의를 집중적으로 하겠다.

5.5.2 면접원 오차의 통계적 탐지 및 모델링

면접원에 기인한 응답 변동과 응답자에 기인한 변동을 구별하는 것은 일반적으로 불가능하다. 따라서 면접원만의 독립된 영향력을 조사하기 위해서는 면접원의 임무에 깊이 관여하고, 이를 통제할 필요가 있다. 가장 간단하게는 면접원들에게 각 응답자를 무작위 할당할 수 있다. 대개는 소수의 면접원들에게 임의로 할당되는 하나 또는 그 이상의 일차표본단위(primary sampling unit)에서 응답자들에 따른 "지분"설계("nested" design)를 이용할 것이다 (사례는 Collins, 1990 참조). 그러나, 면접원 오차를 추정하는데 가장 일반적으로 사용되는 지수는 계급내상관(intraclass correlation)이다. 그리고, 근본적으로는 임의효과인 ANOVA 모델이 면접원 분산성분을 추정하는데 이용되며, 이에 따라 총 측정분산에 대한 면접원 분산의 비가 추정된다. 급내상관계수는 0에서 1까지의 값을 가진다. 여기서, 0은 특정한 질문에 대한 응답이 면접원과 무관한 것임을 의미하는 값이고, 1은 전적으로 면접원에 의해 결정되는 것임을 뜻한다. 보다 자세한 논의는 Groves(1989, 7장, 8장)를 참조한다.

적절히 잘 이루어진 개별방문면접에서는 일반적으로 급내상관계수가 평균 0.02 정도로 낮지만, 조사변량의 비율은 보다 높은 편이다(Collins, 1980 참조). 그리고, 중앙통제방식의 전화조사에서는 이 보다 낮은 평균 0.01 이하의 값을 보이는 것으로 나타나 있다(Hox 외, 1991 참조). 그러나, 조사통계의 총분산에 대한 면접원들의 전체효과는 계급내상관과 면접원의 업무부담 모두에 기인한다. 따라서, 면접원 효과가 적다 하더라도 평균적인 업무부담이 큰 경우에는 조사자료의 품질에 중대한 영향을 미친다.(Groves 외, 1988, p.364 참조)

면접원 분산에 관한 연구는 어느 정도의 분산이 면접원에 기인한 것인 지만 알려줄 뿐, 면접원의 어떠한 특성들이 이에 대한 역할을 하는 지에 관한 사항은 알 수 없다. 지금까지 이에 대한 통계적 접근방식은 두 가지가 있었다. 그 중 첫번째는 면접원 표본을 분할하여(예, 남성 대 여성, 경력자 대 초보자) 면접원 변수의 효과를 조사하는 것인데, 이것으로는 각 부표본에 대한 계급내상관관계를 추정할 수 있다. 그리고, 두번째는 면접원 변수를 (예, 성별, 나이, 경력) 응답자 레벨로 분배하여, 통계적 분석을 함으로써 면접원 변수의 효과를 추정하는 것인데, 이 때는 응답자들이 면접원에 할당되어(nested) 있다는 사실로 인해 발생되는 통계적 문제는 무시한다.

최근, 이러한 방법들은 계층적 선형 또는 다수준 모델로 대체되고 있는데, 이 모델들은 응답자들을 면접원에 할당시킨 자료의 계층적 구조를 고려하고, 면접원과 응답자 효과에 관해 통계적으로 정확한 추론을 가능하게 한다. 특히, 다수준 모델은 응답자 효과에 조건부적인 면접원 효과를 모델링함으로써 분석자들이 실험적인 방법 대신 통계적인 방법으로 관리할 수 있도록 해 주기 때문에, 실제 현지에서의 제약으로 응답자들을 면접원에게 무작위 할당하는 것이 불가능한 경우에 매력적이다. Hox(1994)는 면접원 효과를 분석할 때 다수준 모델을 사용하는 방법에 대해 명쾌한 지침을 제시하고 있다. 실제 적용에 관한 사항은 Hox et al.(1991)를 참조한다.

대부분의 조사에서 면접원의 인구통계학적 특성들은 자료의 품질과 무관한 것으로 나타나 있다. 그러나, 면접원의 인구통계학적 특성들은 조사의 주제나 질문과 관련해서는 영향을 미친다. 면접원 특성에 관해 유일하게 일관된 결과는 면접원의 인종(race)에 관한 것인데, 이것은 인종관련문제가 쟁점으로 되어 있는 경우에는 더욱 그렇다. 그리고, 성별에 관련된 주제를 다룰 때는 면접원의 성별 효과에 관한 보고서도 일부 있으며, 면접원의 나이와 경력에 대해서는 상호 엇갈린 결과들이 있다. 한편, 면접원에 대한 교육은 이에 대해 체계적인 조사는 거의 없지만, 면접원의 기교에 긍정적인 효과가 있을 것으로 보인다. 그리고,

이러한 교육 후에는 면접원들이 학습한 기교를 지속 사용할 수 있도록 하기 위한 업무 지향적인 감독이 필요하며, 중앙집중식의 직접감시나 녹음 테이프의 사용이 모두 이러한 목표를 달성하는데 도움이 될 것이다. Couper et al.(11장)는 어떤 종류의 컴퓨터화된 자료수집기법이 면접원 행동을 감시 및 평가하는데 새로운 방법을 제시하는 지에 대해 설명하고 있다. 면접원 관련 오차에 관한 일반적인 개요는 Groves(1989, 5장)와 Fowler(1991)를 참조한다.

5.5.3 면접과정과 상호작용 코드부여

면접원 오차를 통제 및 감소시키기 위해서는 면접원 오차를 통계적으로 탐지 및 모형화 할 수 있어야 할 뿐만 아니라, 이러한 오차가 발생하는 메커니즘에 관한 정보 또한 가지고 있어야 한다. 예를 들면, 상호작용 코드부여에 관한 연구는 면접원들의 준비가 미흡한 상황을 밝혀낼 수 있을 뿐만 아니라, 면접의 특성이 무엇인지, 응답자와 면접원 사이의 질문과 설문 내지는 상호작용이 무엇인지 그리고, 무엇이 면접의 표준화를 결정하는 지를 우리에게 제시해 줄 수 있다(Schaeffer, 1995). 또한, 면접에서 실제로 일어나는 사항에 대해 아는 것은 조사자들이 면접과정에 관한 이론을 정립하는데 도움이 되고, 면접원 교육과정을 개선하는데 지침이 된다.

면접과정 특히, 면접원의 행동효과에 대한 관심은 Hyman et al.(1954)의 연구로까지 거슬러 올라가며, 여전히 많은 관심을 요하고 있다. 이 분야에서는 면접원과 응답자의 언어행동을 분류하는 상호작용 코드부여가 일반적인 접근방식이다. 원래 사회심리학에서 개발된 이 접근방식은 Cannell 등(예, Cannell et al., 1968)에 의해 설문조사에 소개되었다. 그 후로는 면접원의 면접기여 및 교육훈련 참가 적합성 조사(예, Cannell et al., 1975)와 면접 스타일이 자료품질에 미치는 영향력 조사(예, Van der Zouwen et al., 1991) 그리고, 사전조사 면접에서 문제를 발생시키는 질문 파악(예, Morton-Williams and Sykes, 1984; Oksenberg et al., 1991) 등의 여러 가지 분야에 적용되었다. 이 접근방식에 관한 보다 자세한 논고 특히, 질의검정에 적용된 내용은 Fowler와 Cannell(1996)의 논문에서 확인할 수 있다.

초기에 이 기법을 사용할 때는 단편적인 방법으로 행동을 바라보려는 경향이 있었다. 그러나, 1990년 Tucson 회의 이후, Sykes와 Collins(1992)는 행동의 순차를 분석하고, 하나의 자극-응답 과정으로서 이것을 단순면접모형과 비교하며, 이러한 모형과는 달리 "단순명료하지 않은(non-straightforward)" 형식의 순차를 정의함으로써 그 분야를 확대시키고자 하였다. 당시에는 힘들고 귀찮았던 이러한

형식의 분석은 최근 Van der Zouwen과 Dijkstra(1995) 및 Dijkstra와 Draisma(1995)에 의해 발전되고 개선되었다.

이 연구의 가정은 면접과정의 단순모형과 주된 차이점이 면접원의 수행능력이나 문항설계 혹은 둘 다와 관련된 문제를 제기한다는 것으로 보인다. 하지만, 혹자는 이것이 반드시 사실은 아니며, 면접원과 응답자 간의 "교섭(negotiation)"이 어쩌면 자료품질의 개선에 더 기여할 지도 모른다고 주장하기도 한다. 드문 예로, Dykema et al.(12장)은 확인된 자료를 바탕으로 이 가설을 시험한 것이 있는데, 여기서 그는 - 경우에 따라서는 - 비표준적인 면접원 행동이 실제로는 긍정적인 기여를 할 수도 있음을 제안하였다. 여기에 대해서는 다른 측정분야로 확대하고, 순차분석의 개념을 통합하는 등 보다 많은 연구가 이루어져야 할 것이다.

5.6 연구 의제

본 절에서는 Lyberg와 Kasprzyk의 연구 및 1990년 Tucson 회의를 바탕으로 논의를 해왔다. 그러나, 1995년 Bristol에서 열린 조사측정 및 프로세스 품질(Survey Measurement and Process Quality)에 관한 회의는 1990년 Tucson 회의 이후 발생된 연구들에 대해 앞으로의 연구의제를 남기고 이를 재평가할 수 있는 계기를 마련해 주었다. 이것은 쉬운 일은 아니다 - 우리는 지난 5년간 발생한 변화들을 자신있게 예측하지 못했으며, 앞으로에 대해서도 더 많은 확신을 가질 수 없다. 단지 당장에 눈앞에 놓인 몇 가지 현안만을 지적할 수 있을 뿐이다.

CAI의 추가 개발

컴퓨터를 이용한 면접(CAI; Computer Assisted Interviewing)은 급속도로 번져왔다. 혹자는 지나치게 빠르다고도 한다. 이에 따라 CAI와 특히, CAPI에 대한 기술적 불평들이(예, 느린 응답시간, 긴 응답을 기입할 때의 문제, "휴대용" 컴퓨터의 무게) 하드웨어와 소프트웨어의 발전으로 개선되었다. 또한, 기자재의 성능을 극대화시키는 첨단 소프트웨어의 개발에도 많은 연구노력이 투입되었다. 하지만, 정작 인적 요인들은 경시된 것으로 보인다. 화면을 읽고, 응답을 입력하는 것은 지면 설문지를 통해 이루어지는 것과는 다른 지각과 운동기능을 필요로 한다. 모니터를 통해 질문을 읽는 것은 보다 어려운 일이며, 면접원들은 설문의 전체적인 구조를 파악하기도 더 어렵다고들 한다. 따라서, CAI 자료의 품질을 보다 개선시키기 위해서는 면접원의 시스템 지원을 극대화할 수 있는 경험적 연구와 소프트웨어 개발이라는 양쪽 측면에서 그러한 인적요인들에 대해

보다 많은 관심이 필요하다. 이에 대한 가능성은 Edwards et al.(1995)의 면접원 문제를 예측하는 CAPI 화면설계에서부터 열리기 시작하였다.

CAI 효과

위에서 Nicholls et al.(9장) 등이 말한 바와 같이, 주어진 자료수집모드에 컴퓨터를 이용할 경우 자료품질에 미치는 효과에 관한 직접적인 연구는 드물며, 더욱이 CAI가 자료에 미치는 효과에 대해서는 증명된 바가 거의 없다. 하지만, 변화를 겪고 있던 자료 사용자와의 논의는 시간 불연속성에 관한 불평 및 특정한 질문에 미치는 효과 등, 다소 다른 양상을 띠게 될 것으로 여겨진다. 이러한 모순은 대부분의 비교연구가 비교적 단순 명료한(straightforward) 자료수집과정을 고찰하였거나 혹, 어쩌면 너무도 다양하고 많은 질문 유형들을 살펴보았다는 사실을 반영하는 것인지도 모른다. 그렇다면, 고립된 문제(isolated problem)들은 간과되었을 것이다. 다른 한편으로 본다면, 고립된 문제들은 상당히 많을 수 있으며, 심지어 일부 사용자들에게는 극복할 수 없는 어려움들이 있을 수도 있다. 예를 들어, Sykes와 Collins(1988)는 일반적인 비교연구 분야에는 모드효과가 거의 없으며, 민감한 주제(음주) 및 특정한 질문유형(개방형 질문 및 평정척도)에 관한 특수한 조사에는 많은 효과가 있는 것으로 발표하였다. 우리의 연구의제의 관점에서 보면, 이것은 CAI 효과의 고립된 사례에 대해 더 많은 연구가 필요함을 암시하는 것이다. 이러한 사례를 조사하고, 발생배경을 파악하며, 이를 관리할 수 있는 방법 등에 관한 연구.

면접원 기교(skills)

면접에 새로운 기술이 활용될 때면 언제나, 면접상황에 미치는 영향과 면접원과 응답자 사이의 공감대가 유지되는 지의 여부에 많은 관심이 집중된다. 뿐만 아니라, 이러한 관심은 개방형 질문에 대한 응답을 처리하는데도 표명된다. 이와 같은 관심에 대한 답으로 경험 많고 노련한 면접원들은 그다지 어려움을 겪지 않는다는 주장이 있다. 이러한 맥락에서 Woijcik et al. (1992)는 CAPI에 관한 교육 프로그램을 설명하였는데, 컴퓨터를 이용한 면접의 특별한 수요와 관련하여 면접원들을 교육하는 방법은 더 많은 개발이 필요한 실정이다. 이에 따라, CAPI 면접원들을 위한 교육 프로그램의 개발 및 공유가 바로 본 의제의 요지라 하겠다. 새로운 기술 -CAI와 전화기술 모두- 의 활용에 필요한 면접원의 기교는 더 많은 관심을 가질 만한 사항이다.

면접원 행동

앞서 말한 바와 같이, 면접원과 응답자 간 상호작용에 관한 조사는 여전히 관심의 대상으로 남아 있다. Dykema et al.(12장)은 이 분야에서 몇 가지 새로운

질문은 던졌는데, 단순명료한 순차(straightforward sequence)를 가진 질의응답 행동과는 다소 다른 점들이 측정의 품질을 개선시키는데 공헌할 수 있다고 제안한다. 이들의 결과에서 드러난 한계는 행동 코드부여를 측정된 자료품질과 연계시키는 연구를 확대할 필요성을 극명하게 보여주고 있다.

전문화된 조사

전문화된 조사는 조사산업(survey industry)에서 그 중요성을 점점 더해가고 있으며, 현재도 계속해 이러한 특수한 조사를 실천하는 방법에 대한 지식들이 쏟아지고 있다. 그러나, 이 지식들은 간헐적이고, 산재되어 있다. 조사자들은 이러한 지식을 서로 공유해야 한다. 심지어는 직면하는 문제에 관한 단순한 문서에 불과한 것이라도 가치가 있다. 우리는 가까운 장래에 이러한 주제에 관한 이론적이고, 경험적인 연구가 많이 이루어지길 간절히 바란다.

민감한 질문

개인 프라이버시에 관한 대중의 관심과 민감한 주제에 관한 조사의 수요는 지속적으로 증가할 것으로 예상해야 한다. 또한, 많은 조사기관의 품질에 대한 이미지는 민감한 자료를 수집할 수 있는 각 기관의 능력에 달려 있으므로, 이 주제는 당연히 연구의제로 남아야 한다. 특히, 컴퓨터를 이용한 자계식 면접과 프라이버시를 보장할 수 있는 다른 방법들의 활용에 보다 많은 연구가 필요할 것이다.

산업구조에 대한 효과

컴퓨터를 이용한 자료수집과 여기에 필요한 대규모 투자에 대한 강조는 큰 규모의 조사대행기관(supplying agencies)들에게 이익이 될 것이고, 산업구조에도 중요한 효과가 있을 것이다. 그리고, 소규모 기관들은 대규모 기관들에 합병되거나 자료수집에 관한 하도급 계약을 맺어야 할 지도 모른다. 이러한 변화들은 자료품질에 반드시 영향을 미치는 것은 아니지만, 자료수집과 자료품질 분야에 대한 조사자들의 관심이 당연하다는 사실에 하나의 증거가 된다.

자료품질에 대한 조사기관(Agency)의 효과

정기적이거나 지속적인 조사에 점점 가중되는 경쟁의 압박은 조사기관에도 많은 변화를 불러올 수 있다. 예를 들면 현지조사 전략과 무응답자 응대전략, 면접원 교육 그리고 감독 내지는 편집 및 코드부여에 관한 측면에서 조사기관 사이의 편차는 자료품질에 영향을 미칠 수 있다. 따라서, 이러한 효과는 충분히 조사할 만한 가치가 있는 것이다. 하나의 단계로서의 조사기관 그리고, 조사기관에 속한 응답자들과의 계층적 또는 다각적 수준분석은 이러한 조사기관의 효과를 예측하고, 바로잡는데 유용한 도구가 될 것이다.

이 의제는 본 절의 주제인 자료수집에서 벗어나 도구설계와 자료처리과정으로 핵심이 전이되는 순간마다 등장할 것이다. 특히 자료수집에서의 많은 쟁점들이 조사과정의 다른 단계들과 상호작용할 때는 여지없이 나타날 수 밖에 없다. 또한, 이것은 자료수집의 품질은 전체 조사설계와 자원제약 및 계획된 활용이라는 맥락에서만 고려될 수 있다는 우리의 견해를 반영하고 있다.

감사의 글

본 소고에 관련 자료를 협조해 주신 Abby Israels, Liz McCrossan, Jeannette Schoorl, Yolanda Schothorst, 그리고 Ron Cjaza에게 감사드리며, 아울러 본 절의 초고에 식견높은 조언과 적극적인 제안을 마지않은 Joop Hox와 Norbert Schwarz에게도 감사드린다. 본 절은 이 많은 분들의 큰 도움을 받았지만, 최종 결과물과 특히 그것의 오류는 전적으로 저자의 책임이다.

제 6 절 조사자료 품질에 대한 새로운 자료수집기술의 효과

Willian L. Nicholls II
美 통계국
Martin Collins
시카고 대학, 여론조사센터
Jean Martin
영국 통계청

본 절은 조사자료품질에 새로운 자료수집기술이 미치는 효과를 조사한 것으로, 센서(census)와 서베이(survey)의 자료수집에 현재 활용되고 있는 가장 중요한 신기술에 대해 설명하였다. 조사범위, 무응답, 측정오차에 미치는 수집기술의 효과에 대한 연구 사례들을 요약하였으며, 기술 효과의 경향과 규모에 대한 일반적인 견해를 제시하고, 이러한 연구에 새로운 목적을 제안하는 것으로 끝을 맺었다.

6.1 새로운 자료수집기술의 개요

본 절의 핵심주제는 조사자료수집의 기술들이다. 이러한 기술들의 대부분은 자료수집(collection)과 포획(capture) 및 편집(editing)을 하나의 프로세스로 수행한다. 그리고, 종이조사표의 자료수집 및 편집절차 등, 종래의 개별방문면접과 전화 면접 및 자계식 설문에서 사용되던 종이조사표 작성을 대체하고 있다. 본 절에서는 광학문자인식(optical character recognition) 및 일괄편집(batch editing)과 같이 우편접수, 포획, 처리에만 사용되는 기술은 다루지 않았다.

가장 널리 알려져 있는 새로운 자료수집 기술군은 컴퓨터를 이용한 면접 혹은 CAI이며, 이것은 종이조사(종이조사표)를 소형 컴퓨터나 컴퓨터 단말기로 대체시켰다. CAI에서는 조사의 질문이 컴퓨터 화면에 나타나고, 면접원(또는 응답자)은 컴퓨터 키보드를 이용하여 답을 직접 입력한다. CAI 시스템들은 성능은 제각기 다르지만, 대부분 앞서 입력한 사항을 바탕으로 질문의 순서를 제어하며, 이미 입력된 정보를 반영하여 질의문항을 수정하고, 응답의 일관성과 완성도를 체크한다. 일반적으로 보면, 인구통계나 가구조사의 경우에는 기본 순서에 따라 한번에 하나씩 질문을 보여주고 답을 확인하며, 경제 또는 기업 조사의 경우에는 복수문항을 보여주고, 화면 하단에서 편집을 수행하며, 기업마다 각각의 기록을 유지하는 순서에 맞게 문항과 화면을 자유롭게 이동할 수 있게 한다.

CAI의 가장 보편적인 형식은 면접원이 컴퓨터를 조작하면서 화면을 보고 질문을 읽으면서 응답자가 말하는 내용을 기록하는 방식이다. 컴퓨터를 이용한 전화조사(CATI)는 1970년대 초 중앙전화설비에서 전화조사를 지원하기 위해 개발되었으며, 현재는 CAI 면접 뿐만 아니라, 전화일정 및 사례관리도 처리한다(Nicholls and Groves, 1986). 컴퓨터를 이용한 면접조사(CAPI)는 1980년대 후반 경량의 랩탑 마이크로컴퓨터의 등장으로 응답자의 가정에서 개별방문 CAI 면접이 가능해지면서 시작되었다(Baker, 1990). 또한, 면접원의 가정으로부터 분산화된 CATI에는 랩탑 마이크로컴퓨터를 갖춘 현지 면접원들도 채용되었다(Bergman et al., 1994).

1980년대에는 자계식 CAI도 등장하였다. 컴퓨터를 이용한 자계식 조사(CASI)에서는 응답자가 컴퓨터 화면에서 질문을 읽고, 자신의 답을 직접 기입한다. 이 때는 면접원이 응답자의 집으로 컴퓨터를 가져가기도 하고, 응답자가 컴퓨터가 있는 곳으로 나오기도 한다. 현지 작업자는 조사 시작을 돕고 응답자가 컴퓨터를 직접 다루는 것이다(Duffy and Waterton, 1984; O'Reilly et al., 1994; Mitchell, 1993). Scherpenzeel(1995)은 이러한 형식의 CAI에 대해 "면접원이 존재한다(interviewer present)"는 뜻으로 IP를 덧붙인 CASIIP를 제안하기도 했지만, 본 절에서는 간단히 CASI로 칭하였다.

자계식 CAI의 관련 형식으로, 면접원이 없는 경우에는 종종 컴퓨터를 이용한 자계식 설문(CSAQ; Computerized Self-Administered Questionnaire)이라 부른다. 여기서는 자료수집기관이 응답자에게 각자의 개인용 컴퓨터(PC)나 컴퓨터 계정에서 응답하도록 요청하며, 소프트웨어를 플로피디스크에 담아 우송할 때는 DBM(disk by mail) 방식이라고도 한다. 경우에 따라 CSAQ 소프트웨어는 로컬 네트워크나 모뎀, 전자우편을 통해 전송되거나 혹은, 전자 게시판이나 웹사이트에 업로드되어 원격지에서 접속할 수 있게 할 수도 있다 (Kiesler and Sproull, 1986; Downes-LeGuin and Hoo, 1994; Sedivi and Rowe, 1993). 자료수집기관은 또한 응답자에게 가정용 컴퓨터를 제공할 수도 있는데, 이 방식은 텔레인터뷰(TI; teleinterview)라 한다(Saris, 1990).

두 가지의 새로운 자료수집 기술은 면접원과 응답자 모두가 컴퓨터 화면과 키보드를 사용할 필요가 없다. 이것은 우선 하나의 자료수집 도구로서 일반적인 누름식 전화기를 이용하는 TDE(touchtone data entry) 방식이다. 이 때는 대부분의 경우, 응답자가 자료수집 기관의 무료전화서비스에 전화를 걸면 시스템이 음성을 디지털화하여 응답자에게 질문을 읽어주고, 응답자는 전화기 키패드를 이용하여 자신의 답을 입력한다(Clayton and Harrell, 1989). 디지털

음성은 응답자가 확인 또는 수정을 할 수 있도록 각각의 응답을 반복할 수 있다. 원칙적으로 TDE는 CAI와 유사한 방식의 복잡한 질문경로와 맞춤식 질의 및 편집 기능을 채택하고 있다. 그러나 실제로는 이러한 기능은 거의 사용하지 않고, 정부에서 수행하는 기업에 관한 간단한 패널조사에만 주로 이용되었으며, 일반적인 여론조사에도 시험되기도 하였다(Frankovic et al., 1994; McKay and Robinson, 1994).

음성인식(VR: Voice Recognition)은 TDE와 유사하지만, 키패드를 통한 응답보다는 구술응답 방식을 채택하고 있다. 즉, 컴퓨터가 디지털 음성으로 응답자에게 질문을 읽어주고, 말로 하는 응답을 분석하여 그 인식된 의미를 기록하는 것이다. 인식하지 못했을 때는 질문을 반복하거나 관련 프로브를 즉시 보여준다. 시스템의 정교함에 따라, 응답자는 간단히 예/아니오 내지는 한자리 숫자나 연속된 번호, 단어 구술 혹은, 명칭의 철자를 말할 수 있다. 소용량 어휘의 VR(예, 아니오 및 숫자)은 TDE를 대신하거나 이의 백업용으로 이용될 수 있고(Clayton and Winter, 1992), 대용량 어휘의 VR은 응답자가 조사기관으로 전화하는 반대 의미의 CATI 면접을 대신할 수 있다(Blyth and Piper, 1994). TDE와 마찬가지로, VR도 기업의 전화(call-in), 패널조사에 주로 이용된다. 뿐만 아니라, 소비자 패널의 전화(call-in) 면접(10장)에도 조사가 진행중이며, 2000년 미국 인구센서스에 이용되는 몇 가지 수단 중 하나로도 이용되고 있다(Appel and Cole, 1994).

CATI, CAPI, CASI, CSAQ, TDE 및 VR은 대부분 새로운 자료수집 기술을 주로 이용하므로, 이러한 방식의 자료품질 효과들이 가장 널리 연구되어 왔다. 반면, 전자자료교환(electronic data interchange) 혹은 EDI와 같은 기타 신흥 기술들의 자료품질 효과에 대해서는 거의 알려지지 않은 상태다 (Ambler and Mesenbourg, 1992). 다만, 추가적으로 새롭게 등장하거나 등장이 예상되는 기술들은 범용될 경우, 조사자료수집을 혁신적으로 바꿔 놓을 가능성을 가지고 있다. 여기에는 다기능 전화기(screen phone)와 비디오 폰, 쌍방향 TV 및 개인휴대용단말기(personal digital assistant)와 같은 새로운 개념의 전화기들이 포함된다(Baker, 1995; Ogden Government Services, 1993). 그러나, 이러한 기술들의 자료품질 효과에 대해서는 체계적인 증거가 전혀 없으므로, 여기서는 더 이상 고려하지 않았다.

6.2 조사자료품질과 기술

조사 관리자들은 여러 가지 이유에서 자료수집 기술을 선택할 수 있다. 예를 들면, 어떤 사람에게는 비용절감이 가장 중요할 수 있으며 또 어떤 사람에게는 적시의 평가물이 그리고 또 다른 사람에게는 조사자료의 품질이 중요한 사안일 수 있다. 이러한 조사의 비용과 적시성 및 자료품질에 새로운 자료수집 기술이 미치는 효과에 대해서는 앞서 Nicholls and Groves(1986)와 Groves and Nicholls(1986), Saris(1990), Weeks(1992), de Leeuw(1994), Baker et al.(1995), 그리고 Martin and Manners(1995)의 연구결과를 요약한 바 있다.

본 절은 조사오차의 부재와 같은 전통적인 의미에서 정의되는 조사자료품질에만 역점을 두고 있다. 비용효과와 적시의 결과 및 자료품질을 소비자 만족이라는 측면에서 정의할 때는 예산운영, 시간제한, 그리고 특정 조사의 우선순위에서 가장 잘 이해된다. 또한, 부분적인 이유이긴 하지만, 비용 및 적시성 보다는 조사오차에 새로운 기술이 미치는 영향에 대해 보다 일반화할 수 있는 경험적 증거가 존재한다. 심지어는 비용과 적시성이 수집기술을 선택하는데 주된 사안인 경우에도, 조사오차에 관한 기술의 효과를 파악하는 것이 여러 특성들의 오차를 방지하고, 그러한 오차에 가장 취약한 특성들을 줄여 나가면서 오차의 활용을 극대화시킬 수 있는 방안을 제시하기도 한다.

우리의 견지는 Bradburn and Sudman(1989)이 "조사의 과학적 근거"라고 부른 것에서 비롯한다. 이 모형을 따르는 조사는 신중하게 정의된 모집단으로부터 확률표본을 선별하고, 각각의 관련된 표본구성원에 동일한 정보가 수집되도록 표준화된 도구(또는 설문)를 이용하며, 선택된 표본의 대부분으로부터 정보를 수집하고, 표본자료에서 모집단에 관한 추론을 도출할 수 있는 표본이론을 채택한다. 이 모형은 자료수집을 면접을 통할 것인지 아니면, 응답자 자기기입방식을 통할 것인지를 결정하고, 또 수집방식(mode)이 지면인지, 마이크로컴퓨터인지, 혹은 전화기 버턴누름식인지를 결정하게 한다. 하지만, 혁신적인 수집기술의 활용은 적절한 조사절차를 대신할 수 있는 것은 아니다.

조사모형은 일반적으로 두 가지 부류의 조사오차 - 표본오차와 비표본오차- 를 인식한다. 표본오차는 표본 설계와 표본 크기에 주로 기인하고, 비표본오차는 보다 복잡한 양상을 가진다. Biemer(1988)와 Groves(1989)는 비표본오차를 포함오차(coverage error), 무응답 오차(nonresponse error), 측정오차(measurement error)의 세 가지 주요 범주로 나누었다. 이러한 오차들은 조사수집방법이나 조사도구, 면접원, 응답자 등, 다양한 원천에서 비롯된다.

조사자료는 이러한 비표본오차의 주된 형태들이 모두 제어 및 최소화되었을 때 높은 품질을 유지한다고 말할 수 있다.

조사자료품질에 미치는 조사수집방법의 효과에 관한 연구는 세 가지 선택적인 종이조사표 자료수집방식 - 개별방문면접, 전화 면접, 자계식(일반적으로 우편) 설문 - 에 주된 초점을 맞추고 있다(Sudman and Bradburn, 1974; Lyberg and Kasprzyk, 1991; de Leeuw, 1993). 그 예로, 모드효과에 관한 논문은 다음과 같은 결론을 보여주었다. (1) 전화가입이 보편적이지 않은 경우에는 전화조사가 개별방문면접조사에 비해 포함오차에 보다 취약하다. (2) 우편설문조사는 일반적으로 전화조사나 개별방문면접조사보다 단위 및 문항 응답률이 낮다. (3) 자계식 설문은 전화나 개별방문면접에 비해 민감한 행동에 보다 충실한 결과를 얻는다.

"모드효과(mode effect)"라는 표현은 컴퓨터를 이용한 수집방법과 종이조사표 수집방법 간에 적용범위, 무응답 또는 측정오차의 차이를 설명하는데 종종 이용된다. 물론 이것은 상호 혼재될 수도 있다. 하나의 자료수집방법은 수집방식(collection mode) (면접, 전화, 자기기입)과 수집기법(지면 혹은 컴퓨터 이용)을 모두 포함하고 있다. 따라서, 둘 중 한 가지가 바뀌면 자료품질에 영향을 미칠 수 있다. 예를 들어, 우편설문에서 CATI로의 변화는 방식(mode)과 기술(technology)을 모두 바꾸게 된다. 그러므로, 그 각각의 효과를 구별하기 위해서는 매우 조심스러운 조사설계가 요구된다.

수집기술의 조사자료품질에 관한 효과를 추정하기 위해서는 반드시 수집기술은 다르지만, 수집방식과 설문 및 기타 필수적인 조사조건은 모두 동일한 상태에서 조사처리를 비교해야 한다. 이러한 실험적 환경은 기술 효과를 평가하기 위한 대조연구에서만 근접될 수 있다. 보다 박약한 증거는 다음의 연구에서 확인할 수 있다. 통제된 환경에서 수집기술 및 설문과 같은 둘 이상의 설계요소들을 동시에 수정하는 연구, 조사기술이 변경되기 전후에 수행된 연구 그리고, 동일한 응답자들을 상대로 면접중에 수집방법을 바꿨을 때의 경험(예, CAPI에서 CASI로). 본 절에 기술된 정보는 이러한 증거 유형들을 기초로 하고 있다.

6.3 포함오차

포함오차는 연구대상 모집단의 구성원을 확인하는데 이용되는 목록이나 틀에서 일부 인원이 빠져 있는 경우에 발생한다(Groves, 1989). 전화나

개별방문면접조사의 경우에는 종이조사표 수집방법 보다는 CAI의 활용이 특별한 적용범위의 문제를 일으키지 않는다. 이 둘에는 정확히 동일한 표본틀과 현지표본방법이 이용될 수 있다. 이것은 또한 면접원이 소개하는 CASI의 경우에도 마찬가지다. 하지만, 이와는 대조적으로 CSAQ와 TDE 및 VR은 특수한 장비를 가진 응답자들만 참여시킬 경우 모집단의 적용범위를 제한하게 된다. 즉, 그러한 장비가 없는 응답자들은 조사에서 제외될 수도 있는 것이다.

TDE와 VR은 모두 전화를 필요로 하고, TDE의 경우에는 특히 누름식(touchtone) 전화기가 반드시 있어야 한다. 스웨덴과 같이, 전화기가 보편화되어 있고, 거의 대부분이 누름식인 나라에서는 이것은 문제가 되지 않는다. 그러나, 전화가입이나 누름식 통화가 보편적이지 않은 곳에서는(Trewin and Lee, 1988) TDE는 매우 선택적으로만 사용될 것이다. 예를 들어, 1993년도에는 미국 가구의 약 20%가 회전식 내지는 펄스 전화기를 가지고 있었고, 5%는 아예 전화기가 없었다. 회전식 전화기는 장년층 중에서도 특히 노년층이 거주하는 근린지구와 선택적인 지리적 위치에 집중되어 있었다 (Frankovic et al., 1994). 심지어는 미국 기업들의 약 10% 가량도 누름식 전화기가 없었다(Clayton and Werking, 1994). TDE를 소용량 어휘의 VR로 대체하면 누름식 전화기에 대한 필요성은 없어진다. 하지만, 여유있고, 효율적인 VR 소프트웨어가 모든 언어에서 활용될 수 있는 것은 아니므로, 이 또한 새로운 적용범위의 한계성에 직면할 수 있다.

CSAQ 포함오차는 잠재적으로 더욱 심각하다. PC 보유현황을 보면, 1994년 미국 가구의 26%(U.S. Bureau of the Census, 1995), 1993년 영국 가구의 24%가(Office of Population Censuses and Surveys, 1995) PC를 보유하고 있었지만, 그 외 다른 나라에서는 이보다 훨씬 낮은 보유율을 보였다. 또한, PC 보유자는 상위 소득계층인 도시 전문직에 집중되어 있었다. 이와 같이 중요한 적용범위의 편향에 대해 Sociometric Research Foundation (Saris and de Pijper, 1986)에서는 각 응답자의 가정에 소형 컴퓨터와 모뎀을 설치해 줌으로써 이를 직접 해결할 수 있는 방안을 마련하였다. 이러한 상황에서는 응답자의 TV가 컴퓨터 모니터의 역할을 하고, 조사 질문은 모뎀을 통해 전송되었으며, 응답자는 전화기를 통해 답을 얘기하였다. 물론, 적절한 TV와 전화기 그리고 응답자의 동기부여라는 요건이 적용범위를 제한할 수는 있지만, 아마도 종이조사표 방법으로 유사한 소비자 패널을 조사하는 것에 비할 바는 아닐 것이다.

CSAQ 기업 조사의 경우에는 아무래도 가구보다는 기업에 PC 보유 및 활용율이 더 높을 것이므로 그 효과가 더욱 크다. 개인용 컴퓨터는 이제 미국

대부분의 기업에 보편화되어 있다. 심지어는 100인 미만의 기업에도 2/3 가량이 PC를 보유하고 있다(Ogden Government Services, 1993). 이 정도의 PC 보급율은 다른 선진국에도 마찬가지일 것으로 생각된다. 그러나, CSAQ 조사에 대한 준비는 대규모 기업들 사이에서도 아직 미흡한 실정이다. PC는 용량이나 드라이브, 모뎀, 운영체제 등이 다르므로, CSAQ 응답자들은 일반적으로 CSAQ 소프트웨어를 구동 및 액세스할 수 있는 환경인지를 사전에 반드시 확인해야 한다. 그리고, CSAQ에 응답할 수 없는 환경은 반드시 다른 수집방법을 이용하든가 아니면 이들을 제외시킨 편향을 수용해야 한다.

6.4 무응답 오차

무응답 오차는 지정된 표본의 단위들로부터 정보를 수집하지 못한 결과에서 발생된다. 이 중 단위 무응답(unit nonresponse)은 지정된 표본단위로부터 거의 아무런 정보를 취득하지 못한 상태를 말하고, 항목 무응답(item nonresponse)은 참여 단위에 대한 특정 자료항목이 없는 상태를 말한다.

6.4.1 단위 무응답

단위 무응답률은 보통 비접촉율(noncontact rates)과 비협조(거부)율(noncooperation (refusal) rates)로 나뉜다. 과거의 일부 연구자들은 가능성이 있을 것으로 생각하긴 했지만, 새로운 수집기술이 비접촉율에 영향을 미친다는 증거는 거의 없다. 예를 들면, Groves and Nicholls(1986)와 Weeks(1988)는 CATI의 자동화된 통화일정이 종이조사표 전화조사에 비해 비접촉율을 감소시켜야 하지만, 이 가능성은 대조연구에서 전혀 확인되지 않았다고 주장하였다. 초기의 CATI 및 CAPI 시험에서는 또한 운영상의 문제로 면접원이 필요한 통화를 할 시간이 적어 현지업무가 지연되는 경우에는 비접촉율이 증가하는 것으로 나타났다. 반면, 운영상의 문제가 해결되었을 때는 그 효과가 사라졌다(Catlin and Ingram, 1988; Bernard, 1990; Kaushal and Laniel, 1993).

새로운 기술을 이용하여 조사에 참여하려는 응답자의 의지는 보다 중요한 사안이다. 초기에는 비록 일부 조사 관리자들이 CATI나 CAPI 응답자들이 컴퓨터에 답을 입력하려 하지 않을 것이라는 우려를 표명했지만, 연구결과 이것은 기우였던 것으로 드러났다. CATI 응답자들은 일반적으로 전화조사 중에 컴퓨터를 사용한다는 사실을 모르거나, 무관심한 것으로 나타난 반면, CAPI 응답자의 대부분은 종이조사표 방법보다는 CAPI를 선호하거나 아니면 사용된 기술에 아무런 관심이 없는 것으로 나타났다(Groves and Nicholls, 1986; Wojcik

and Baker, 1992; van Bastelaer et al., 1988). CATI와 CAPI의 거부율을 동일한 종이조사표 대조군과 비교했을 때는 이들 간에 유의한 차이가 발견되지 않았다(Catlin and Ingram, 1988; Baker et al., 1995; Bergman et al., 1994, de Leeuw, 1994; Manners, 1991).

면접원이 소개하는 CASI에 대한 응답자 수락율의 조사결과는 일관성이 없는 것으로 나타나 있는데, 이것은 아마도 연구마다 비교의 환경과 편향이 각기 다르기 때문인 것으로 사료된다. 초기 연구에서, Duffy and Waterton(19984)는 종이조사표 가구 면접에 비해 가구 CASI의 거부율이 높다고 보고하였다. 하지만, 실험실이나 교실 환경에서 최근 연구한 결과를 보면 응답자들이 종이조사표 자계식 보다는 CASI를 더 선호하는 것을 알 수 있다 (O'Relly et al., 1994; Mitchell, 1993). 한편, 현지 환경에서 CASI와 종이조사표 설문의 대조 응답률 비교는 많이 부족한 실정이다. Lessler and O'Relly(1994)의 연구보고서를 보면, CAPI 면접도중 CASI를 소개할 경우 임산부의 대부분은 별 문제없이 이를 받아들인다고 되어 있다. 그리고, Couper and Rowe(1995)는 노인 응답자들의 경우에는 많은 CASI 면접이 실제로는 면접원들에 의해(혹은 직접적인 도움을 받아) 완수된다는 사실을 발견하였다.

CSAQ는 PC(또는 다른 적절한 컴퓨터 접속)를 보유한 특수한 모집단에만 적용될 수 있기 때문에 CSAQ를 종이조사표 설문 응답률과 비교한 소수의 연구결과는 일반화하기가 어렵다. Kiesler and Sproull(1986)은 모두 컴퓨터 계정이 있는 대학생과 교직원들의 연구에서 CSAQ가 종이조사표 설문에 비해 응답률이 낮은 것으로 보고하고 있다. 또, 미국 정부기관은 기업(석유산업과 주정부기관) 응답자들을 대상으로 조사할 때 적절한 컴퓨터 자원을 가진 업체만을 사전 선별한 후에는 중대한 무응답 손실없이 종이조사표 방식에서 CSAQ 방식으로 성공적인 변화를 이끌어 냈다(Kindel, 1994; U.S. Energy Information Administration, 1989; Sedivi and Rowe, 1993). 그리고, Downes-LeGuin and Hoo(1994)는 disk by mail CSAQ 논문을 검토하면서 컴퓨터 지식이 있는 전문가들은 대개 우편 설문보다는 CSAQ 방식을 선호하는데 이것은 보다 흥미롭고, 쉽게 작성할 수 있으며, 시간이 그다지 소요되지 않기 때문이라고 주장하고 있다. Downes-LeGuin and Hoo는 DOS 기반의 컴퓨터에 액세스할 수 있는 선별된 전문가 모집단의 경우(예, LAN 관리자), 우편설문에 대한 응답률은 15-40%였던 반면, CSAQ 응답률은 40-60%였다고 보고했다. 하지만, 이들의 결론은 CSAQ의 이용보다는 연구 모집단의 특성과 설문의 길이 및 추적조사(follow-up) 절차가 응답률에 보다 중요한 변수라는 것이다.

미 노동통계청(BLS; Bureau of Labor Statistics)의 고용통계(CES; Current Employment Statistics) 조사에서 제기된 바에 따르면, TDE와 소용량 어휘 VR의 응답자 수락은 반드시 전체 조사설계의 측면에서 분석되어야 한다. 이 조사는 초기에는 매월 고용주 패널들을 대상으로 우편을 통해서만 실시되었으며, 이 때 응답률은 약 50% 정도를 유지하였다. 이후 응답자 표본은 응답률을 향상시키고, 적절한 결과산출을 위해 CATI 쪽으로 옮겨졌고, 6개월 후에는 CATI 표본의 응답자들은 필요할 경우 CATI 추적조사와 함께 TDE로 이동되었다. 누름식 전화기가 있는 거의 모든 CATI 응답자들은 TDE를 받아들였으며, 이러한 CATI의 응답률 또한 80% 이상을 유지하였고, 비용과 표본 이탈율은 오히려 감소하였다(Clayton and Harrell, 1989; Rosen et al., 1993). TDE는 그 자체만으로 응답률을 증가시킨 것이 아니었다. 높은 응답률은 초기에 CATI를 통해 오히려 효율적으로 유지된 것이다. 다음 번에 TDE 응답자들을 위한 선택사항으로 BLS가 소용량 어휘 VR을 제공했을 때는, 응답자들은 일반적으로 사용상의 편리성과 관리시간단축을 이유로 TDE보다 VR을 선호하였다(Clayton and Winter, 1992). BLS는 궁극적으로 모든 CES 우편 응답자들을 TDE와 VR로 옮겨놓기 위한 전략을 현재 개발하였지만, 각 전략에는 신중한 응답자 준비와 단계별 무응답 추적조사가 필요하다 (Clayton and Werking, 1994). 이와 유사한 TDE 절차들은 미 인구조사국(Census Bureau)에서도 개발 및 시험하였는데 비슷한 성공을 거두었다 (Bond et al., 1993).

미국에서는 현재 TDE와 VR이 지속적인 기업 패널(establishment panel)에 의한 소량의 수치자료 보고에 적합한 수집기술로 인식되어 있다. 비록 이와 동일한 기술이 가구 응답자들의 응답률에서는 문제가 보다 많은 것으로 보이기는 하나, 대안적인 수집방법과의 대조된 현지비교에서 시험되지는 않았다.

6.4.2 항목 무응답

항목 무응답은 다음의 경우에 발생한다. (1) 응답자가 조사항목에 실질적인 답을 하지 않거나 자계식 형식에서 항목을 빠뜨린 경우, 또는 (2) 면접원이 응답자에게 해당 질문을 하지 않거나 주어진 답변을 기록하지 않은 경우.

새로운 수집기술은 모두 컴퓨터를 이용하여 통제된 경로와 다음 문항을 보여주기 전에 각 항목에 입력된 사항의 확인을 통해 항목 무응답 오차를 최소화하기 위한 목적을 가지고 있다. 뿐만 아니라, 이전 문항으로 되돌아가기나 건너뛰고 나중에 돌아가기 등과 같이 이동형식을 선택할 수 있는 경우에도 본래의 논리적 결과의 유지가 가능하다. CASI, CSAQ, TDE 및 VR과 같은

자계식 기술들은 종이조사표 형식에서 누락된 자료를 빈 것으로 처리한 것과는 달리 "거부" 또는 "모름"이라는 분명한 의사표시를 할 수 있다는 장점을 가질 수 있다. 그러나, 이러한 특징들이 답하기 싫어하거나 답할 준비가 안 되어 있는 응답자들의 반응을 자극한다고는 반드시 볼 수 없다.

CAI 논문의 가장 일관된 결론 중 하나는 CAI가 궁극적으로는 응답자와 면접원 모두가 해당 문항을 빠뜨리지 않도록 할 수는 있지만, 응답거부나 모름, 의견없음 등의 의사표시율을 거의 또는 전혀 낮추지는 못한다는 것이다. 이러한 결과는 CATI, CAPI, CASI 및 CSAQ에도 적용되고 사실적 질문과 의사적 질문에도 모두 적용된다(Baker et al., 1995; Catlin and Ingram, 1988; de Leeuw, 1994; Groves and Nicholls, 1986; Manners et al., 1993; Martin et al., 1993; O'Reilly et al., 1994; Weeks, 1992). 항목 무응답률의 감소는 조사의 성격과 CAI의 효과적인 활용 및 조사도구의 복잡성에 따른다. 무응답은 종이조사표 항목 무응답에서 건너뛴 형태의 오차가 지배적인 경우에 가장 크게 감소된다(Baker et al., 1995; Catlin and Ingram, 1988; Manners et al., 1993). TDE와 VR의 경우에는 이와 유사한 항목 무응답 감소에 대해 제시된 바가 없는데, 이것은 아마도 비교적 간단한 설문과 자료수집 동안 편집확인(edit check)의 사용이 드물었기 때문일 것이다.

6.5 측정 오차

Groves(1989)는 측정 오차를 "조사도구에 기록된 응답의 부정확성"으로 정의하고 있다. 이러한 부정확성은 면접원이나 응답자 또는 설문에 기인할 수 있다. 수집기술은 이러한 종래 원천을 통한 조작이나 혹은 기술 자체의 취약함으로 인해 측정 오차에 영향을 미칠 수 있다.

6.5.1 측정 오차에 대한 일반적 접근방법

선택적 자료수집방법과 연관된 측정 오차의 수준을 평가하는 데는 전통적으로 세 가지 접근방법이 사용되었다(Biemer, 1988; Groves, 1989; Lyberg and Kasprzyk, 1991).

타당성(Validation)

첫째는 각 방법을 통해 수집된 자료를 다른 원천 - 예를 들면, 관리기록 - 으로부터 수집된 자료와 비교하는 것이다. 자료수집 기술을 비교할 때 기록확인을 이용하는 것은 드문 일이었으며, 특히 실제로도 유익하지 않다.

Mitchell(1993)은 대학생들을 연구하면서, 대학의 기록을 타당성 검증도구로 활용했을 때, 학점 평균과 학자적 성향에 관한 시험점수 및 F학점 과목수의 정확성 면에서 CASI와 종이조사표 설문간에 유의한 차이를 발견하지 못했다. 한편, Grondin and Michaud(1994)는 연속된 종이조사표와 CAI 전화조사를 통해 수집된 다양한 소득원에 관한 기록을 캐나다 국세청(Internal Revenue Service) 기록과 비교하였는데 CATI 결과가 더 일치하는 것을 발견하였다. 그러나, CATI와 동시에 이루어진 설문지 변경은 이 결과의 해석에 혼돈을 일으킨다.

분할 표본처리

타당성 기준이 적용되지 않을 때는 서로 다른 조사처리의 분포가 유의한 차이를 식별하는데 비교될 수 있다. 차이점의 성격은 조사방법이 편향과 오차의 영향을 최소한으로 받았음을 제시할 수도 있다. 대략 15가지의 연구가 이와 같이 다른 식으로 비교되는 상황에서 얻은 CAI와 종이조사표 조사분포를 조사하였다(Baker et al., 1995; Bergman et al., 1994; Bernard 1990; Catlin and Ingram, 1988; Groves and Mathiowetz, 1984; Groves and Nicholls, 1986; Heller, 1993; Kaushal and Laniel, 1993; Manners, 1990; Manners and King, 1995; Martin et al., 1993; Mitchell, 1993; Tortora, 1985; van Bastelaer et al., 1988). 이것을 보면, 단지 몇몇을 제외하고는 CAI와 종이조사표 분포간의 차이는 거의 없거나 작으며, 대체적으로 예상할 수 있는 평범한 것이었다. 이러한 결론은 CAPI, CATI, CASI, CSAQ와 각각에 상응하는 종이조사표 비교에도 적용된다. 뿐만 아니라, 인구통계 및 기업조사, 개인 및 가구의 배경 특징, 경제활동참가와 실업측정, 개인 및 가계소득 보고, 정치적 견해와 사회적 태도, 직업 만족도 측도, 심리학적 시험점수 등에도 똑같이 적용된다.

이 결론은 반드시 세 가지 면에서 평가되어야 한다. 첫째, 사회적으로 바람직한 응답에 관한 보고가 포함될 수 있는 민감한 문항이나 CAI로 변경하면서 설문도 변경되는 상황에는 반드시 적용되는 것은 아니다. 이러한 예외들은 다음 절에 고찰되어 있다. 둘째, 많은 CAI 대 종이조사표 비교는 불과 몇 백개의 사례표본을 기초로 했기 때문에, 이것으로는 각각의 분포간에 조그만 차이가 검출될 것으로 기대할 수 없다. 셋째, CAI와 종이조사표 자료의 비교는 일반적으로 중심성향에 대한 T-검정과 카이제곱 분석에 의해 측정된 광범위한 차이에 역점을 두었다. 하지만, CAI와 종이조사표 분포의 산포를 비교했거나 묵종(opinion acquiescence)과 극단(extremity)의 측도에 대해 조사한 연구는 거의 없다. 단 한가지 예외가 있다면, Martin et al.(1993)을 들 수 있는데, 그는 종이조사표 개별방문면접보다는 CAPI에서 태도척도에 극단값을 자주 사용한다는 사실을 발견하였다.

TDE와 우편설문 분포의 비교는 공통적이지는 않지만 유사한 결과를 도출한다. 미국 제조업체 선적량, 재고량, 주문량(M3) 조사의 응답자 표본은 3개월 동안 우편과 TDE 두 가지 방법으로 매월 미 인구조사국에 보고서를 제출하라는 요청을 받았다. Bond et al.(1993)은 이들의 분포에는 유의한 차이가 없다고 보고하였다. 또, Phipps and Tupek(1991)은 경시적 편집확인으로 TDE 응답에서 간혹 발생되는 오차를 모두 제거한 후에는 BLS Current Employment Survey에 응답한 TDE와 우편설문 간에 본질적으로 아무런 차이가 없음을 발견하였다.

응답 분산, 면접원 분산, 응답 안정성

측정 오차에 대한 두 가지 수집방법의 상대적 취약성을 평가하기 위한 세 번째 방법은 각각의 응답 분산이나 면접원 분산 또는 응답 안정성을 비교하는 것이다. 이 접근방식은 참값과 이에 대한 응답오차의 편차 및 편향으로 구성되는 조사 응답모형으로 시작된다. 응답오차는 최초 면접 후에 동일한 도구와 기준기간 및 조사조건을 이용하여 조사 응답자들을 간단하게 재면접함으로써 평가할 수 있다. 최초 면접과 재면접 값 사이의 차이는 응답오차에 대한 추정값으로 간주된다. 오차에 대한 면접원 기여도는 면접원에 응답자를 임의 할당한 조사에서 면접원간 분산에 대한 기여도를 조사함으로써 추정할 수 있다. 패널 응답 안정성은 패널조사에서 측정 오차의 관리를 의미하는 것으로, 여기서는 응답자들의 응답이 각 패널 파동(panel wave)에서의 응답오차와 면접원 오차 모두에 종속된다.

수집기술에 있어 응답 분산이나 면접원 분산 또는 패널 응답 안정성에 대한 비교 연구는 비교적 많지 않다. Bergman et al.(1994)은 스웨덴 경제활동인구조사(Swedish Labour Force Survey)에서 CAI와 종이조사표 방식의 분산화된 전화조사를 비교하였는데, 경제활동 항목에서 응답 분산간에는 아무런 차이가 없음을 발견하였다. 한편, 미국의 상시인구조사(U.S. Current Population Survey)에서는 경제활동 변수에 있어 CAI 응답 분산이 종이조사표 응답 분산과 같거나 혹은 그 보다 작은 것으로 나타났다(Rogers, 1991; Weller 1994). 그러나, CAI/종이조사표 방식과 두 번째 연구 전후의 설계간에 발생된 다른 차이점들은 CPS 결과에 대한 명확한 해석을 제한한다. Groves and Magilavy(1980)는 통계적으로 의미가 있는 수치는 아니었지만, 동일한 설문에 대해 종이조사표 전화조사보다는 CATI가 면접원 분산이 더 작게 나타나는 것으로 보고하였다. Martin et al.(1993)은 세 가지 방식을 연속하여(두 번 모두 종이조사표, 종이조사표 다음 CAPI, CAPI 다음 종이조사표) 순차적인 개별방문면접을 하면서 태도 항목에 대한 응답자 안정성을 비교한 결과, 응답 안정성은 CAPI를

이용했을 때 보다 커진다는 사실을 발견하였다. 이에 따라, Martin을 비롯한 연구진은 CAPI가 약간은 보다 안정적인 기술이라는 결론을 내렸다.

응답 분산, 면접원 분산, 그리고 응답 안정성에 관한 연구들이 비록 CAI가 자료의 품질을 향상시킴을 시사하는 증거 이상의 것을 제시하지는 못했지만, 이것으로도 CATI 및 CAPI 자료는 품질면에서 종이조사표 자료와 최소한 동일한 품질을 유지하고, 어쩌면 조금은 더 나은 것이라는 뜻은 전달된다. 이러한 새로운 기술의 사용이 늘어감에 따라, 특히 대규모 정부차원의 조사에서 이러한 분석들을 반복하고 확대하는 기회들을 가진다면 CAI가 측정오차의 감소에 대한 전반적인 영향력을 평가하는 최적의 수단임이 입증될 것이다.

6.5.2 면접원 오차

종이조사표로부터 CAI로의 변화는 면접원들의 활동에 가장 뚜렷한 효과가 있기 때문에 면접원들의 이 새로운 기술의 수용과 이에 대한 효과적인 활용 능력 및 종래의 현지 품질관리방법에 대한 이 기술의 영향은 공통적인 연구주제가 되어 왔다.

면접원 수용

1970년대 CATI가 처음 도입되었을 때, 조사자들은 전화 면접원들이 이것을 어려워하거나 자신들을 속박하거나 아니면, 자신들의 전문적 기술의 가치를 떨어뜨릴 수 있다는 생각을 할까 봐 걱정했다(Coulter, 1985; Groves and Mathiowets, 1984). 이에 따라 초기의 연구들은 면접원 수용에 대해 상반된 연구결과를 내 놓았다. 그러나, CATI 시스템이 검증력과 효율성을 증대시키고, 면접원들이 이 새로운 기술에 친숙해짐에 따라, 대다수의 전화 면접원들이 종이조사표 보다 CATI 방식을 더 선호하게 되었다. 많은 사람들은 CATI가 단순 사무직이라는 느낌을 줄이고, 자신들을 전문가적 입지를 향상시켰다고 믿었다(Nicholls and Groves, 1986; Weeks, 1992).

CAPI가 개별방문면접에 제안된 것은 1980년대였는데, 이 때 면접원 수용이라는 쟁점이 다시 부각되었다. 많은 국가에서, 현지 면접원들은 40대 이상의 중년 여성으로 비교적 컴퓨터에 대한 사전지식이 부족한 상태였다. 따라서, 초기의 휴대용 컴퓨터가 무겁고 가시도가 낮아 CAPI가 건강에 악영향을 미칠 수 있다는 걱정들이 제기되었고(Blom et al., 1989; Couper and Burt, 1993), 일부 조사관리자들은 CAPI로 인해 자신들의 경험많은 현지 요원들이 컴퓨터 면접에 익숙한 사람으로 대체되는 것은 아닐까 걱정했다. CAPI의 사전조사는

1980년대 중반에 이르러 시작되었는데 결과는 상반된 입장이 상존했다. 일각에서는 성공적인 운영이었으며, 면접원들도 의욕에 찼다는 결과를 내놓은 반면, 또 다른 쪽에서는 하드웨어와 소프트웨어 대한 면접원들의 불만이 많았다는 지적을 했다. 휴대용 마이크로컴퓨터의 초기모델은 무겁고 느렸으며, 일부 CAPI 소프트웨어는 원시적이고, 입력오류가 많았고, 디스크나 모뎀을 이용하여 전체 데이터를 본사로 전송하는 것 또한 그다지 안정적이지 못했다(Baker, 1990; Bernard, 1990; Brown and Fitti, 1988).

이후, CAPI에 대한 광범위한 면접원 수용은 가벼우면서도 강력한 하드웨어, 향상된 소프트웨어, 개선된 데이터 관리 및 통신, 그리고 신중하게 개발된 면접원 교육이 전체적으로 조화됨에 따라 이루어졌다. 최대한의 도움을 필요로 하는 피교육자들에 대한 철저한 준비와 각별한 관심으로, 조사기관들은 마침내 모든 숙련된 면접원들이 CAPI 방식으로 교육될 수 있음을 깨달았다. 일단 교육을 받고 나면, 대부분의 면접원들은 종이조사표 보다는 CAPI 면접을 더 선호하였다(Blackshaw et al., 1990; Couper and Burt, 1993; Edwards et al., 1993; Wojcik and Baker, 1992).

면접원 자료입력

컴퓨터를 이용한 면접의 주된 이점은 자료의 범위와 일치성 및 편집에 있어 수집되는 그대로의 완전함을 유지시켜 준다는 점이다. 따라서, 입력사항이 범위를 벗어나거나 이전 데이터와 일치하지 않는 경우, 면접원은 응답자와 함께 그러한 사항을 즉시 명료하게 밝힐 수 있다. 이와 대조적으로 종이조사표 면접의 경우 면접원은 응답자의 응답에서 명백한 오차를 찾고, 지면 양식에 정확히 기입되도록 하기 위해서는 자신의 주의와 기억력에 의존해야만 한다. 면접원이 주의를 했음에도 발생하는 오차들은 반드시 다음 번 면접항목이나 편집시에, 혹은 핵심 입력사항을 확인하거나 일괄적인 컴퓨터 편집시에 밝혀져야 한다. 그리고, 이러한 오차들이 검출된 다음에도, 이를 해결하기 위해서는 응답자에게 다시 연락하거나 사람이나 컴퓨터를 탓하거나 혹은 오류값들을 데이터 코드가 없는 것으로 변환하는 등 추가적인 조치가 필요하다. Manners(1990)가 관찰한 바에 따르면, CAI는 "성공할 가능성이 가장 큰 때 즉, 면접시에 이루어지는 자료 편집"으로 품질을 향상시킨다고 한다.

CAI 온라인 편집의 효과는 CAI와 종이조사표 자료를 동일한 컴퓨터 일괄편집을 통해 각각의 편집 오차율을 비교함으로써 평가되었다. 한 농업조사에서 이루어진 CATI와 종이조사표 전화조사의 초기 비교에서 Tortora(1985)는 일괄편집 결과, CATI 자료에서 보정이 필요한 문항은

종이조사표 자료에서 필요한 것보다 76%나 적은 것으로 나타났다고 보고했다. 또, Manners et al.(1993)는 복잡한 재정조사에서 CAPI를 종이조사표 개별방문면접과 대조비교한 결과, 보정이 필요한 문항이 75%정도 적은 것으로 보고했다. 이 밖에도 CAI 온라인 편집의 효과에 대해서는 Bernard(1990)의 경제활동인구 조사와 Sigman et al. (1993)의 기업 조사에도 언급되어 있다. 그러나, 대개 온라인 편집은 CAI의 선택적 특성이므로 이러한 장점은 온라인 편집이 면접 속에 구성되어 있는 경우에만 얻을 수 있다.

일반적으로 CAI는 문항 누락과 자료의 모순을 감소시키는 것으로 인식되어 있지만, 혹자는 이것이 면접원 기록 오차에는 유일하게 취약할 수 있다는 입장을 표명했다. Groves and Nicholls(1986)은 컴퓨터에 답을 입력하는 것은 지면 양식에서 박스에 체크 표시하는 것에 비해 오차가 발생할 가능성이 더 높을 수 있다는 주장을 제기했다. 또, Lyberg and Kasprzyk(1991)는 면접원들은 일반적으로 숙련된 타이피스트가 아니기 때문에 축어적인 응답을 생략하거나 혹은 긴 문장을 입력하는데 시간이 걸려 응답자들을 짜증나게 할 수 있음을 주지시켰다. 그러나, 이러한 입장들을 뒷받침해 줄 만한 증거는 없다. 지금의 면접원들은 이전에 생각했던 것에 비해 CAI에 대한 준비가 훨씬 잘 되어 있을 것이다. 실제로, 미인구조사국(U.S. Bureau of the Census) 종이조사표 면접원 표본을 조사한 결과, Couper and Burt(1993)는 약 절반 가량이 적절하거나 혹은 강도높은 컴퓨터 실습으로 CAPI 교육을 시작했으며, 이 중 거의 3/4이 스스로를 터치 타이피스트라고 자신하였고, 나머지는 일일이 보고 치는 타이피스트라고 응답하였다.

CAI 폐쇄형 문항의 면접원 자료입력에 관한 체계적인 연구를 보면 보다 고무적이다. 이중 모니터링 CATI 면접(Kennedy et al., 1990)과 테이프 기록 CAPI 면접(Dielman and Couper, 1995)을 기초로 한 연구에서는 입력 문항의 0.6%와 0.1% 만이 면접원 오차인 것으로 보고되었다. 또, 종이조사표 모의면접에 관한 초기 연구를 보면, 종이조사표 입력 오차가 CAI 입력 오차와 최소한 동일하며(Rustemeyer, 1977), 생략형태 오차를 포함하면 이를 훨씬 초과하는 것을 알 수 있다.

개방형 질문에 대한 CATI와 CAPI의 자료입력 품질에 대해서는 4개국에서 수행된 5가지의 연구에서 잘 조사되어 있다. 문장 길이, 다수응답 질문에서 각기 다른 응답횟수 및 코드 입력능력과 같은 기준을 바탕으로 할 때는 CAI와 종이조사표 간에 텍스트 응답입력에서 의미있거나 일관된 차이가 있는 것으로 보고된 것은 없다(Bergman et al., 1994; Bernard 1990; Catlin and Ingram, 1988;

Gebler, 1994; Kennedy et al., 1990). 이러한 연구에서는 직업과 업종에 관한 기술 및 광범위한 의견을 묻는 문항들이 모두 조사되었다.

조사감독원의 품질관리

종이조사표 면접원들의 현지 조사감독원은 면접원들의 업무 품질을 관리하는데 있어 전통적으로 다음과 같이 다양한 방법을 이용하였다. (a) 응답율과 면접원의 자료입력 오차율에 관한 기록보관 (b) 반환된 면접지의 완전성, 일치성, 허위작성에 관한 검토 기록 (c) 응답자 가정에서의 면접원 업무 관찰 (d) 전화시스템을 통한 중앙형 전화조사 감시.

CATI와 함께, 기록관리는 면접원 업무에 대한 일일 또는 실시간 업데이트가 가능하도록 자동화될 수 있다. 또한, 많은 CATI 시스템을 활용한 시청각 감시로 조사감독원은 전화조사의 대화를 들으면서 면접원의 CATI 화면을 눈으로 확인할 수 있다. 이것은 조사감독원이 면접과정의 문제를 파악하고, 정확한 조치를 취할 수 있게 도와준다(Couper et al., 1992). 뿐만 아니라, 체계적인 감시는 면접을 위조하거나 질문을 빠뜨리고 응답을 추측하는 것과 같은 심각한 면접원의 직무태만을 방지하는데 많은 기여를 한다. 그러나, 이러한 시청각 감시가 자료품질을 향상시킬 수 있는 CATI의 가장 강력한 도구 중 하나로 종종 인용되지만(Groves and Nicholls, 1986; Weeks, 1992; de Leeuw, 1994), 이것의 효과는 감시된 면접의 자료품질을 감시되지 않은 면접과 비교한 대조연구에서 증명된 바가 없다.

개별방문면접에 CAPI를 적용하게 되면 종래의 조사감독원의 관행을 상당히 깨뜨리게 된다. 지면 설문지의 기록검토는 가능하지도 않고, 필요하지도 않으며, 또 CAPI 면접원들에 대한 조사감독원의 현지관찰은 조사감독원들이 숙련된 CAPI 면접원들과 보조를 맞춰 이들의 면접과정을 동시에 기록할 수 없기 때문에 상당히 어렵다(Bernard, 1990; Blackshaw et al., 1990). 그 결과, 조사감독원들은 종이조사표보다 CAPI 면접원들의 관리가 더 힘들다고 느끼게 된다. 따라서, 제11장에 나와 있는 CAPI 면접의 자동 일시 확인과 완료된 면접의 본사 일일 전송 및 CAPI 컴퓨터 입력파일의 분석은 개별방문면접에 관한 새로운 품질관리방법을 제시하고 있다. 새로운 조사감독원 품질관리방법의 효과는 대조연구에서 입증되어 있지만, 이것은 다른 품질관리방법에도 마찬가지다. 조사감독원 품질관리의 가치가 종종 자명한 것으로 간주되거나 별도로 측정하기 매우 어려운 것으로 인식되어 있기 때문에, 자료품질에 관한 그 효과는 조사감독원의 관행이 변화되기 전후의 설계에서만 확인될 수 있을 지도 모른다.

6.5.3 응답자 오차

여기서는 새로운 기술이 자계식 자료입력과 응답자의 민감한 정보기입에 어떠한 영향을 미치는 지에 관해 고찰하였다.

응답자 자료입력

새로운 기술을 이용한 응답자의 입력오차와 종이조사표 양식을 이용한 입력오차를 비교한 연구는 그다지 많지 않다. O'Relly et al.(1994)는 종이조사표 자계식 보다는 CASI에서 응답자 오차가 훨씬 적은 것으로 보고하였다. 이와 유사한 결과는 Kiesler and Sproull(1986)과 Mitchell(1993)의 CASI 연구에도 잘 나타나 있다. 또한, 기업 패널조사에서 Bond et al.(1993)는 종이조사표 우편조사보다 TDE에서 응답자의 기록 오차가 훨씬 적은 것을 발견하였으며, Phipps and Tupek(1991)은 BLS CES에서 초기 TDE 입력오차율은 문항당 1.2~2.5% 정도라고 보고하였다. 이러한 오차율은 지나치게 자리수가 많은 숫자와 같이 쉽게 확인할 수 있는 TDE 입력실수에 관한 경시적 편집 후에는 영(0)으로 떨어졌다.

TDE 가구조사에서는 보다 많은 문제점들이 발견되었다. McKay and Robinson(1994)은 실험실 환경에서 처음 시도된 TDE 형식의 미국 상시인구조사(Current Population Survey)에서 새롭게 채택된 35가지 주제 중 5가지가 이 형식을 따를 수 없으며, 각자의 집에서 이루어진 TDE 보고서에는 23가지 중 2가지가 동일한 오차를 반복한 것을 발견하였다. 또, TDE로 TV 방송에 대한 대중의 반응을 조사하기 위해 모집된 표본에서 Frankovic et al. (1994)은 응답자의(특히 노년층 중의) 상당수가 현지 조작자에게 가야 되고, 이것은 이들이 TDE에 불만을 가지고 있거나 조작상에 어려움을 느끼기 때문이라는 뚜렷한 이유가 있다고 보고하였다.

음성인식 조사에서는 응답자의 실수는 시스템의 음성인식 오류보다 중요하지 않을 것이다. 여기에 관해서는 다음 절에 논의되어 있다.

사회적으로 민감한 행동에 관한 응답

성행위, 음주 및 약물복용, 불법행위 가담 및 소득과 같은 주제에 대해 응답자로부터 민감한 답을 이끌어내는 능력은 조사방법마다 다르다. 최근의 연구를 보면, 이러한 관점에서 개별방문면접과 전화조사 간에는 일관된 차이가 없는 반면, 자계식 방식은 면접원들이 주도하는 방식보다 민감한 답을 더 잘 이끌어낸다는 뚜렷한 증거가 제시되어 있다(제13장, de Leeuw, 1993). 본 절의 질문은 CAI 방법이 동일한 방식의 종이조사표 형식에 비해 민감한 행동에 관해 보다 많은 답을 유도해 내는 가이다.

전화조사에서는 종이조사표 형식보다는 CATI의 활용이 민감한 행동의 기록에 영향을 덜 미칠 것으로 보이는데, 이는 일반적으로 응답자들이 컴퓨터가 이용되고 있음을 인식하지 못하기 때문이다. 또, 개별방문조사에서는 CAPI 랩탑을 완전히 볼 수 있는데, 두 가지 연구에서 응답자들은 CAPI를 종이조사표 면접보다는 기밀한 사항을 제공하는 것으로 간주한다는 사실이 입증되었다 (Baker et al., 1995; van Bastelaer et al., 1988). 하지만, CAPI로 보다 정확한 기록이 전달되는 지 아닌 지는 확실하지 않다.

한 가지 연구 - Youth Cohort의 National Longitudinal Study(NLS/Y) - 에서 몇몇 민감한 주제들에 대한 보고는 CAPI가 종이조사표 개별방문면접에 비해 보다 성실히 제공된다는 사실을 발견하였다(Baker et al., 1995). NLS/Y의 11번째 조사에서는 종이조사표 면접에 비해 CAPI에서 상당히 많은 음주관련 문제가 보고되었다. 12번째 NLS/Y 조사에서는 이와 동일한 질문이 반복되지는 않았지만, 새롭게 추가된 문항들 중 피임유무에 관한 질문에서 종이조사표보다 CAPI에서 남성들의 응답률이 상당히 높은 것으로 기록되었다. 이와 동시에, 과거와 현재의 임신여부 및 임신 중 건강과 관련한 행위에 대한 여성들의 응답은 CAPI와 종이조사표 간에 뚜렷한 차이가 나타나지 않았다. 그리고, 소득과 유동자산에 대한 응답에 있어서는 통계적으로 유의한 차이가 발견되었다. 이러한 결과들은 아직까지 다른 연구에서는 반복되지 않았다. 따라서 만일, CAPI가 민감한 주제에 대한 응답을 유도한다면 그 효과는 작고, 일관되지 않을 것으로 보인다.

현재 많은 조사자들은 CAPI보다는 오히려 CASI가 민감한 주제의 응답을 유도할 가능성이 가장 높은 기술이라고 믿고 있다. 다른 자계식 수집방법과 마찬가지로, CASI도 프라이버시 관련사안을 줄이고, 사회적 요망성에 관한 질문을 최소화해야 한다. Kiesler and Sproull(1986)은 "줄어든 사회적 맥락의 질문은.. 전자식 조사에서는 조사환경을 특정개인과 관계없는 익명으로 만들 것이며.. 응답자들은 자기중심적이고, 사회적 규범과 다른 사람들에 주는 인상에 무관심하게 될 것이다"는 가설을 세웠다. 대학생과 교직원에 관한 대조연구에서 Kiesler and Sproull은 종이조사표 형식보다는 전자 설문형식에서 민감한 주제에 대해 상당히 높은 수준의 솔직한 응답을 발견하였다. Mitchell (1993)은 이후의 연구들을 요약하고, 대학생 표본을 대상으로 CASI와 종이조사표 자계식 설문에 대한 비교조사를 직접 수행한 결과, Kiesler and Sproull의 결과를 재현하는지는 못했다.

O'Relly et al.(1994)는 실험연구에서 민감한 주제에 대해 세 가지 기술의 응답을 비교하였다. 종이조사표 설문, 컴퓨터 화면에 질문이 나타나는 CASI, 헤드셋을 통해 디지털 음성으로 응답자에게 질문을 읽어주는 오디오-CASI 방식. 그레코-라틴 방격 실험설계는 세 가지 기술과 세 가지 민감한 주제 -성행위, 음주 및 약물복용, 개인소득- 로 각 응답자들을 로테이션시켰다. 여기서, 성행위와 약물복용 문항에서는 종이조사표와 두 가지 CASI 방식간에 유의한 차이가 (약 0.10 수준) 발견되었지만, 소득에 관한 질문에서는 차이가 없었다. 두 가지 방식의(표준 및 오디오) CASI는 민감한 행동에 관한 답을 얻는데 있어 거의 동일한 효과가 있었지만, 대부분의 응답자들은 답변방식으로 오디오-CASI를 선호하였다.

13장에는 Jobe 등이 유산과 성행위에 관한 여성들의 답변에 세 가지 조사 설계요소의 효과를 조사한 연구에 대한 설명이 있다. 여기서 세 가지 요소란 조사방식(자계식인지 면접인지)과 기술(CAI인지 종이조사표인지) 및 면접장소(응답자의 집인지 도서관이나 패스트푸드 레스토랑과 같은 제3의 장소인지)를 말한다. 여러 가지 민감한 주제의 분석을 바탕으로, 조사자들은 조사방식(면접보다는 자계식)이 민감한 행위에 대한 답을 유도하는데 가장 중요한 설계요소였다는 결론을 내리고 있다. 또한, 이들의 분석을 보면 기술(CAI 대 종이조사표)은 기껏해야 상황에 따라 작고, 미세한 효과만 있을 뿐이라고 주장하였다. 예를 들어, CAI는 각자의 집에서 여성들을 면접할 때는 성행위 대상자의 수가 상당히 높게 나타났지만 제3의 장소에서 면접한 경우에는 상당히 적은 수로 나타났다. 이는 아마 면접에 주위 사람들의 관심이 쏠리고, 이에 따라 개인의 프라이버시가 침해되기 때문인 것으로 사료된다. CAPI와 CASI 각각에 대한 분석을 뒷받침해 줄 표본이 충분하지 못한 관계로, 이 연구에서 CAI에 관한 결론은 이 둘이 혼재된 것으로 받아들이면 된다.

민감한 행위에 대한 답변을 유도할 수 있는 가장 효과적인 조사설계특징의 조합은 본 논문에서 해결되지 않았으며, 이에 대해서는 현재도 연구가 진행중이다. 하지만, 그 동안 미국의 몇몇 조사기관들은 민감한 주제에 대한 자료를 수집하는 방법으로 오디오-CASI에 집중해 왔는데, 이것은 읽는데 어려움이 있는 사람들에게 적절한 형식으로 자계식과 CAI가 조합된 형태이기 때문이다 (Lessler and O'Relly, 1995).

6.5.4 도구 효과

CAI 도구들은 면접원이나 응답자가 면접을 완료하는데 도움이 되는 여러 가지 표준 및 선택적 특성들을 제공한다. 이 중 자동 경로지정(automatic routing)은 다음에 질문할 문항을 찾기 위해 이전에 입력한 사항을 다시 살펴봐야 되는 종이조사표 형식의 복잡한 생략 지침을 생략하게 해 준다. CAI의 맞춤식 질문은 응답자에게 이전에 답변한 내용을 상기시키고, 응답자의 특정한 상황에 맞게 질문의 범위를 좁히고, 가족 구성원에 따라 질문을 개별화할 수 있다. 또, 이전 면접원들의 자료는 조사질문과 암시 및 편집확인에 포함되어 종속된 면접이 가능하게 해 준다. TDE와 VR에서는 이러한 특성을 원칙적으로는 사용할 수 있지만 빈번하게 이용되지는 않는다. 하지만, CAI에서는 자료 품질을 향상시키는데 이러한 설계특성들의 효과가 검증되지는 않았지만, 자주 사용되는 편이다.

최근의 몇몇 조사연구에서는 그러한 증거를 제공하고 있다. Baker et al.(1995)는 12회차 NLS/Y 조사에서 혼인여부와 학력, 군경력, 경제활동상황, 직업교육, 이민, 소득 등을 포함한 445개 문항에서 CAPI와 종이조사표의 분포상태를 비교하였다. 여기서 약 13.5%의 차이가 유의수준 0.05로 나타났다. 이 중 가장 빈번한 유의차는 종이조사표 면접원이 다음에 무엇을 어떻게 질문할지를 알기 위해 이전 응답을 체크할 필요가 있거나 각 가구의 기록을 참조할 필요가 있는 문항에서 발생하였다. 이러한 단계들은 시간 소모적이고 오차발생의 가능성이 높다. 따라서, CAI는 분명 이러한 단계들을 불필요하게 함으로써, 면접원들이 혼란을 일으키지 않도록 하는 본연의 임무를 수행하고 있다.

유사한 증거로는 스웨덴 경제활동인구 조사의 사전조사에서 Bergman et al.(1994)이 제시한 것이 있다. 이를 보면, 종이조사표 면접에서는 면접원들이 다음에 할 질문을 결정하는데 있어 기준이 되는 주의 근무시간과 일상적인 주의 근무시간 간의 차이를 계산해야 하지만, CATI에서는 그러한 계산과 경로지정이 자동적으로 이루어진다. Bergman과 그의 동료들이 종이조사표 면접에서 발견한 약 40시간 가량의 군집은 일부 종이조사표 면접원들이 어려운 수계산을 피한 값들을 선호했음을 제시하고 있다.

자동 경로지정과 맞춤식 질문 및 온라인 편집은 현재의 면접에서 이전의 입력사항을 기초로 할 뿐 아니라, 이전의 면접이나 관리기록 또는 기타 자료출처로부터 현재의 CAI 자료기록에 미리 설정되어 있는 정보를 기초로 할 수도 있다. 가구 조사에서는 이러한 과거자료의 활용을 일반적으로

"종속면접(dependent interviewing)"이라 부르고, 기업 조사에서는 "과거자료의 활용"이라 한다. 종속면접은 종이조사표 조사에서 사용되어 왔으며, Hill(1994)은 "패널조사에서 잡음이 많은 측도의 경험적 타당성을 향상시킬 것"으로 주목하였다. 그러나, CAI는 과거 자료의 보다 시기적절한 수집과 전송 및 현재 자료에서 제어된 공개를 용이하게 한다. Pafford and Coulter(1987)가 관찰한 바에 따르면, 종이조사표 면접에서는 면접원의 과거자료 활용이 조절될 수 없고, 응답편향과 면접원 변이성을 유발할 수 있다고 한다. CAI로 인해, 이전 자료는 현재값이 보고될 때까지 유보되어 있다가 이후 편집이나 암시, 주의 또는 의문 등 필요한 경우 공개될 수 있다.

Dippo et al.(1992)는 자료품질의 향상에 크게 기여한 미국 상시인구조사(CPS)의 재설계에서 CAI 종속면접의 활용을 기술하였다. 종래의 종이조사표 CPS 도구는 직업과 직종을 매월 독립적으로 물었다. 이 때, 동일한 직업과 업종에 대한 응답자의 설명에서 조그마한 변동은 직업 및 산업 분류 사이에 매월 큰 변화를 가져왔다. 하지만, 과거 자료의 조절된 공개를 이용한 CAI 사전조사는 세 자릿수 직업분류의 월간 변화를 39%에서 7%로 낮추었고, 이것은 실제 변화와 독립된 추정치에 더욱 가까웠다. 이러한 결과는 산업분류에서도 마찬가지였다. Grondin and Michaud(1994)는 캐나다 Survey of Labour and Income Dynamics(SLID)를 위한 CAI 방법론을 개발하면서 종속면접 활용의 효과에 대해 비슷한 주장을 했다.

종속면접은 CAI에서 유용할 수 있는 몇 가지 강력한 면접 선택사항중의 하나일 뿐이다. 시장조사에서는 여론 및 소비자 선호도 항목에서 이미 잘 알려진 순서 효과를 조절하는데 응답 범주체계나 질문순서의 무작위화가 이용된다. 이 때, Table look-up이나 pick-list는 온라인 데이터베이스에 접속해 소유하고 있는 자동차의 종류나 복용하는 약물의 특정한 이름내지는 지역의 패스트푸드 레스토랑에서 특정 음식에 지불한 금액에 관한 응답을 분류할 수 있다(Groves and Nicholls, 1986; Weeks, 1992). 이러한 특성들은 종래의 종이조사표 방법으로 시도할 수 없었던 조사를 가능하게 한다. 따라서, 조사자료의 품질에 대한 CAI의 가장 중요한 역할은 고품질로 수집할 수 있는 자료의 범위를 확대시킴으로써 발생할 수 있다.

조사설계자들은 면접을 신속히 진행하고, 자료품질을 향상시키기 위해 새로운 CAI 특성을 채택할 수 있지만, 반면 CAI에 종이조사표 설문을 도입할 때는 반드시 의도하지 않은 변화를 방지해야 한다. 설문 설계에 관한 방대한 문헌을 보면, 질의표현과 순서, 응답 범주의 위치 및 체계, 그리고 설문 형식지정에서

비교적 사소한 변화가 응답분포에는 실질적으로 상당한 효과를 미칠 수 있음을 반복적으로 지적하고 있다(Schuman and Presser, 1981; Schwarz and Hippler, 1991). CAI가 조사설계자들을 이러한 영향에서 벗어나게 해 줄 수는 없다. Bergman et al.(1994)은 분산화된 전화조사에서 일례로 종이조사표 응답옵션이 세 가지 행으로 그룹화되는 반면, CAI 응답옵션은 하나의 행에 배열되기 때문에 동일하게 표현된 문항은 CAI와 종이조사표 분포에서 유의적인 차이를 발생시킨다고 설명하고 있다. CATI 형식을 종이조사표 배치에 맞게 수정했을 때는 응답분포에서 차이가 없었다. Baker와 그의 동료들은 또한, NLS/Y에서 CAI와 종이조사표 분포간의 유의차를 조사하여 질의표현(wording)이나 형식(formatting)에서 비교적 미미한 차이가 있음을 발견하였다. 종이조사표 조사를 CAI로 전환하는데 있어 추정값의 일치성이 중요한 목표일 경우에는 가능한 한 CAI 설문을 종이조사표 설문과 나란히 만들어야 한다.

CAI와 종이조사표 응답분포 간의 차이는 정해진 선택항목 외의 응답을 기록할 때 "기타" 범주로 처리하는 등과 같이 현장 처리단계에서 뜻하지 않은 변화로부터 발생할 수도 있다. 종래의 종이조사표 방법에서는 일반적으로 면접원이나 사무원들이 무관한 응답은 삭제하거나 혹은 유의한 응답은 가능하다면 새롭거나 기존의 범주에 재분류하기 위해 "기타" 항목을 검토한다. 그러나, CATI와 CAPI 면접원 자료는 기록상의 편집없이 다음 처리단계로 직접 전송된다. 이처럼 "기타" 항목을 처리하는 방법상의 차이는 Cahoon and Russell(1991)과 Baker et al.(1995)가 보고한 경제활동인구조사에서 CAI와 종이조사표 간 분포상의 유의차를 발생시켰다. 이러한 예들은 종이조사표에서 CAI로의 불완전한 전이를 의미한다. CAI 시스템은 "기타" 응답의 편집을 포함하여 모든 종이조사표 조사단계를 재현하도록 설계될 수 있다. 혹은 CAI 제1단계에서 "기타" 응답은 정해진 선택항목의 두 번째 내지 세 번째 계층으로 이동하도록 경로를 지정할 수 있다.

6.5.5 기술적 오차

자료수집 기술은 그 자체만으로 측정 오차와 무응답의 원천일 수 있다. 몇 가지 기술들은 일부 형식의 자료를 수집할 수 없으며, 어떤 수집기술은 기술적인 문제로 오차를 발생시킬 수도 있다.

기술적 한계

CATI나 CAPI 면접은 일반적으로 종이조사표 면접에서 가능한 것이면 어떤 종류의 조사항목이든 모두 질문하고 기록할 수 있다. 여기에는 종이조사표로는 불가능한 확장된 pick-list 및 랜덤화된 문항을 비롯하여 답변확인

(check-answer), 복합 선택, 수치기입 및 개방형 문항이 포함된다. 그러나, 종이조사표 방법은 일반적으로 주관적 면접과 탐구적 연구에서 요구되는 장황하고 논술적인 답변에는 여전히 최상의 선택으로 남아 있다. 한편, CASI와 CSAQ에서는 자동으로 찾아내기, 입력, 확인 등이 복잡한 자계식 자료수집을 허용한다. 그러나 쉽게 수용되는 CASI 및 CSAQ 문항의 수와 종류는 응답자의 컴퓨터에 대한 숙련도에 따라 다를 수 있다.

TDE는 조사문항에 더 많은 제약을 둔다. 즉, 누름식 전화기를 사용함에 따라 숫자와 일부 특수기호에만 한정된 단일 키 입력만 가능하다. 이것은 이름이나 주소 직업에 관한 설명과 같은 알파벳 문장에는 사용이 극히 제한된다는 것을 의미한다. TDE의 활용은 또한 기업 조사에 참여한 응답자라 하더라도 5 내지 10분 정도로 제한된 시간만이 주어진다. 그리고, 아주 최근까지도 화자와 독립된 음성인식시스템은 숫자와 예, 아니오 정도만 받아들일 수 있었다. 하지만, 일부 대학연구소에서만 작동되는 보다 강력한 시스템은 현재 각 대기와 연속적인 숫자, 이름 철자 그리고, 질문을 반복하라는 명령에서 수백가지 가능한 응답을 인식할 수 있다. 응답자에게 가능한 VR 면접의 최대 길이는 아직까지 연구되지 않았다.

하드웨어 및 소프트웨어 오류

모든 자료수집 기술은 유의한 자료 손실을 유발할 수 있는 하드웨어 고장, 프로그램상의 실수 및 작동 오차에 취약하다. 따라서, 적절한 조치가 취해지지 않으면, 면접배정과 재조사 기록 및 완료된 면접원 자료는 하드웨어나 소프트웨어의 고장 내지는 중앙의 본사와 산재되어 있는 CAPI나 CATI 면접원들 사이의 송신오류로 인해 유실될 수 있다. 또, 작동 실수는 각기 다른 응답을 중복기입하거나 해당 응답자에게 결정적으로 중대한 질문을 빠뜨리는 결과를 초래할 수 있으며, 휴대용 컴퓨터에서 불충분한 컴퓨터 전력이나 예상하지 못한 정전 또는 충분하지 못한 배터리 수명은 면접 진행을 지체시키고, 응답률에 치명적인 영향을 미칠 수도 있다.

실제로 대규모 자료수집기관은 CATI와 CAPI의 가능성 및 운영 시험을 시작할 때 모두 한 두가지 이상의 이러한 문제들을 경험하였다(Bergman et al., 1994; Bernard, 1990; Brown and Fitti, 1988; Catlin and Ingram, 1988; Grondin and Michaud, 1994; Presser, 1983). 그러나, 각각의 인용된 예에서, 조사기관들은 기술적 문제를 예방하거나 그 효과를 최소화할 수 있는 방법을 발견하였다. 이후 몇 년을 거치면서, 컴퓨터 하드웨어와 범용적인 CAI 시스템 및 통신 프로토콜은 크게 개선되었다. 문제를 경험한 조사기관들은 중요한 곳에는 여분의 하드웨어를 설치하고, 정교한 백업 및 복구 경로를 이용하며, 개선된 진단 및 오류제거

도구로 CAI 시스템을 시험해야 한다는 사실을 알았다. 현재, 미 인구조사국(U.S. Bureau of the Census)에서는 기술적 오류로 인한 CAI의 조사건수 손실은 종이조사표 면접원이나 우편으로 인해 잘못 배치된 지면 형식의 손실보다 발생빈도가 훨씬 적은 것으로 평가되고 있다.

음성인식 오차율

음성인식은 분명 응답자가 말로 하는 답변을 인식하지 못할 때의 측정오차에 민감하다. Clayton and Winter(1992)는 미국의 소용량 어휘 VR 시스템에서(숫자, 예, 아니오) 나타난 0.6%의 오차율은 다른 수집 방법을 이용한 BLS의 결과와 비슷한 수준인 것으로 보고하였다. 또한, Appel and Cole(1994)은 자원 응답자를 대상으로 중용량 어휘 VR의 가능성 시험을 한 결과 인식 오차율이 약 1%였다고 보고하였으며, Blyth and Piper(1994)는 2~5%의 오차율을 제시하였다. VR 오차율의 이러한 차이는 그다지 놀라운 것이 아니다. 왜냐하면, 이러한 연구들이 각기 허용한 답변의 복잡성과 채택한 인식방법에서 크게 다르기 때문이다. Appel and Cole은 센서스 유형의 질문에서 예상되는 특정한 답변에 연구의 초점을 두었고, Blyth and Piper는 광범위한 조사에서 사용할 수 있는 음절을 바탕으로 일반화된 음성인식기를 만들고자 하고 있다. 이러한 최근의 추정치는 모두 가능성 검정시험에서 얻어진 것으로, 실제 조사에서는 어느 정도의 인식율이 될지는 예측하기 어렵다.

6.6 요약 및 결론

본 절에서는 새로운 자료수집 기술이 조사자료의 품질에 미치는 효과를 평가하기 위해 여러 연구문헌들을 조사하였다. 그 결과 CSAQ, TDE, VR 보다는 CATI, CAPI, CASI에서 다소 다른 결론들을 얻을 수 있었다.

6.6.1 CATI, CAPI, CASI

CATI, CAPI, CASI는 자료 품질을 향상시키기 위해 부분적으로 개발되었다. 심지어는 비용절감이나 시간단축 또는 처리과정의 편리성이 주목적인 때도 개선된 자료품질은 사실상 컴퓨터 제어 경로지정 및 온라인 편집으로 보장되는 하나의 기대된 부산물이었다. 이와 같이 기대된 자료품질의 이득은 현재 CAI의 성능이 효과적으로 사용되는 경우, 최소한 무응답 문항 및 면접 후 편집오류가 상당한 감소된다는 사실로 입증되고 있다. 마찬가지로 중요한 사실은 이러한 새로운 기술들이 어떠한 유의하거나 일관된 방법으로도 조사자료의 품질을 떨어뜨리지 않는다는 것이다. 또, 처음 소개되었을 때 CATI와 CAPI에 관한

극단적인 예측들은 가능성 검증단계에서 생산단계로 넘어간 후에는 사실로 입증된 것이 아무것도 없다. 공개된 증거들을 바탕으로 할 때, 이러한 기술의 사용을 반대한 응답자들은 실제로 아무도 없었으며, 자신들의 지위가 떨어질 것으로 생각하는 면접원들도 거의 없었고, 실제로 입력 오차도 증가하지 않고, 주관식 응답의 품질도 감소하지 않았으며 조사건수와 문항 자료의 손실도 종이조사표 방법보다 크지 않았다. 개개의 조사에서는 간혹 비정상적인 활용이나 적절한 예비조치가 취해지지 않은 경우 그러한 문제에 봉착할 수도 있으나, 지금까지 유용한 증거들은 이러한 문제들이 CAI 방법의 일반적인 문제는 아니라는 사실을 보여준다.

CATI와 CAPI의 일부 초기 주장자들은 컴퓨터 맞춤식 질문의 간편한 전달을 통해 면접원의 기록업무를 편리하게 하고, 면접과정에서 기타 예민한 변화를 통해 추가적인 자료품질의 향상을 기대하였다. 그러나, 이러한 예민한 효과는 아주 작았다. 또, CATI에서 예상되는 추가적인 자료품질의 이득은 자동화된 호출일정과 시청각 모니터링에서도 비롯되지만, 대조연구결과 이 또한 근거가 불충분했다. 따라서, 종이조사표에서 CAI로의 조사방법 변경이 자료품질을 향상시킨다 하더라도, 이러한 향상은 CAI 방법의 초기 제안자들이 예상했던 것만큼 크고, 광범위하지는 않았다. 초기 주창자들의 열의는 조사편향과 오차의 근원에 미칠 수 있는 CAI의 효과에 대해 비현실적인 기대를 촉발시켰을 수도 있다. 더욱이 종이조사표에서 CAI로의 변화과정에서 설문 설계나 현지 절차상의 의도하지 않은 변화는 조사 결과에서 작은 변화를 일으킬 수도 있다. 간단히 말해, 컴퓨터를 이용한 방법들은 조사의 비포함(noncoverage)과 무응답 및 측정 오차에 만병통치약(또는 일반적인 완화제 조차)이 되지 못한다. 뿐만 아니라, 조사관리자와 조사설계자들이 설문 설계와 현지 업무절차 및 공정 품질에 관한 광범위한 조사문헌들을 이해해야 될 필요성으로부터 해방시켜 주지도 못한다.

하지만 동시에, CAI 방법들은 조사 오차의 관리와 감소에 있어 미래의 중요한 이득을 산출할 수 있는 자료수집에의 새로운 접근방법에 대한 기반을 제공한다. 관리된 종속면접과 온라인 데이터베이스 및 오디오-CASI의 자료품질에 관한 영향은 충분히 평가된 것으로 남아 있다. 그리고 이것은 CAI가 방법론적 도구상자에 추가할 새로운 방법들 중 첫번째 파동에 불과하다. 목적 지향적 설문 설계와 인공지능의 응용, 그리고 멀티미디어 자료수집은 신중히 연구할 가치가 있다. 그 다음엔 지금으로서는 상상할 수 없는 컴퓨터를 이용한 자료수집 접근방법이 뒤따를 것이다.

이러한 전망이 옳다면, 특히 CAI 조사가 단순히 종이조사표 절차를 컴퓨터화한 수준에 머물 경우, CAI와 종이조사표 방법을 비교하는 연구는 이제 그만해야 할 때다. 그러한 연구들은 우리가 이미 알고 있는 것 이상을 제시해 주지 못한다. 또한, 종이조사표 조사가 점점 빠른 속도로 CAI로 바뀌어가고, CAI로만 가능한 절차들의 채택이 확대되어감에 따라 그 의미를 잃게 될 것이다. 새로운 연구는 최신의 CAI 도구의 효과를 평가하고, 선택적인 CAI 조사설계들을 서로 비교하는데 주력해야 할 것이다. CAI가 조사 자료수집의 새로운 표준이 되어감에 따라, 조사 방법론의 원칙과 지식베이스도 이 새로운 환경 속에서 수집방식과 설문설계 및 현지 업무절차를 선택할 수 있는 정보를 제공하는데 초점을 맞추어야 한다.

6.6.2 CSAQ, TDE, VR

CSAQ, TDE, VR의 보다 새로운 기술은 훨씬 복잡한 케이스를 제시한다. 조사기관이 전화조사에서와 같이 응답자 표본에게 CSAQ 소프트웨어와 하드웨어를 모두 제공할 경우, CSAQ의 장점은 아마 CAPI나 CASI와 비슷할 것이다. 그리고, 응답자들이 자신의 장비를 제공해야 하는 경우에는 조사설계자는 반드시 심각한 적용범위의 문제 가능성에 대처하고, 그러한 장비를 보유한 것으로 알려져 있는 특수한 연구 모집단으로 적용범위를 제한하거나 아니면, 하나의 대안으로 CATI나 종이조사표를 선택하여 복합적인 방법의 자료수집을 따르는 것에서부터 시작해야 한다.

CSAQ, TDE, 그리고 소용량 어휘 VR 자료수집의 성공적인 수행은 하나의 백업수단으로 우편설문과 CATI 또는 기타 다른 무응답 처리방법과 함께 정부기관에 약간의 수치자료를 보고하는 기업 패널의 월간 조사에서 발견할 수 있다. 이러한 응용을 볼 때 CSAQ, TDE, VR을 개발하는 주목적은 자료품질을 향상시키는데 있는 것이 아니라 자료품질의 저하없이 비용을 절감하고 수집 및 처리속도를 개선하는데 있다. 그리고, 이 기술의 개발자들 또한 높은 응답율을 유지하고, 패널의 이탈과 입력 오차의 문제를 최소화한다는 점에서 이것이 성공적이라고 보고한다.

컴퓨터를 보유한 것으로 알려져 있는 특수한 전문직 모집단의 조사에 disk by mail CSAQ를 이용하는 것은 보다 검증된 성공으로 보인다. 이것은 낮은 응답율의 종이조사표식 설문조사를 그다지 낮지 않은 응답율의 disk by mail CSAQ 조사로 대체하고 있으며, 이에 따라 응답자 자료입력 및 CAI의 컴퓨터 제어된 분기(branching)와 온라인 편집의 장점을 얻을 수 있다. CSAQ, TDE, VR을 일반적인 여론조사로 확대시키는 것은 당장에는 많은 문제가 있을 것이지만, 미래의 혁신적인 발전과 신중한 연구에는 분명 필요할 것으로 사료된다.

제 7 절 면접원의 CAPI 기술 활용평가

Mick P. Couper, Sue Ellen Hansen, Sally A. Sadosky
미시건대학 Survey Research Center

종이조사표 면접(PAPI)에서 컴퓨터를 이용한 면접조사(CAPI)로의 변화에서 보면, 면접원의 직무수행능력(Performance)을 평가하는 새로운 방법이 필요하다. 여기서 역점을 두는 것은 CAPI 도구가 면접원의 직무수행능력에 미칠 수 있는 효과이며, 협력과 효율성 및 면접기술에 관한 능력과 같은 면접원의 다른 중요한 업무적 특성들에 대해서는 다루지 않을 것이다. 사실, CAPI의 도입이 이러한 중대한 면접기술의 필요성을 감소시키지 않은 것만은 분명한 것으로 보인다(Couper and Burt, 1994). 그럼에도 불구하고, 여기서는 CAPI 시스템과 상호작용하는데 있어 면접원에게 요구될 수 있는 부가적인 기술에 초점을 맞추고 있다.

종이조사표 조사에서 면접원의 직무수행능력에 대한 종래의 척도는 CAPI 환경에서는 다소 불충분하다. 즉, 일부 도구들이 더 이상 무의미할 뿐 아니라, 업무의 성격도 많이 변해, 새로운 척도가 요구된다. 예를 들면, 많은 PAPI 조사에서 조사감독원들은 완성된 설문지의 가독성이나, 응답의 완전성, 범위내 답변, 그리고 정확한 생략 등을 검토한다. 이것은 자동화된 도구가 완전성을 보장하고, 정확한 건너뛰기 형태를 따르며, 범위와 기타 오차를 확인하기 때문에 CAPI를 이용할 때는 불필요한 사항이다(제9장 Nicholls, et al. 참조). 그렇다고, 이것이 이러한 기능들의 대부분이 자동화되어 있기 때문에 CAPI 도구를 사용하면 면접원들이 오차를 만들지 않는다는 것을 의미하는 것은 아니다. 이것이 의미할 수 있는 바는 남아있는 많은 오차들이 종래의 도구로는 탐지하기 더욱 어려워질 것이라는 점이다.

PAPI와 컴퓨터를 이용한 조사를 둘 다 수행하는 면접원의 직무수행능력을 평가하기 위한 일단의 종래의 도구로는 개별방문조사에서 테이프 녹음된 면접의 행동 코드부여(behavior coding) (Cannell et al., 1975)와 전화조사면접의 모니터링(Couper et al., 1992)이 있다. 이러한 방법들은 면접에 관여된 두 사람간의 상호작용에 대해서는 비교적 많은 정보를 도출하지만, 인간(주로 면접원)과 컴퓨터 사이의 상호작용에 대해서는 제시할 수 있는 바가 거의 없다. CAPI 시스템(하드웨어와 소프트웨어 둘 다)은 면접에서 드러나지 않는 존재로서, 면접의 원활하고 성공적인 완료를 용이하게 해 주는 하나의 도구로 있는 것이 가장 이상적이다. 컴퓨터와의 상호작용이 면접원에게 어려움으로 다가오는 한, 이것은 면접원과 응답자 사이의 상호작용에 방해가 될 것이다. 또한, 면접원과

응답자의 상호작용에서 어려움들은 획득되는 자료에 악영향을 미칠 수 있고(예, Suchman and Jordan, 1990 참조), CAPI 조사에서 면접원과 컴퓨터 사이의 상호작용에도 마찬가지일 것이다.

면접원과 컴퓨터 사이의 상호작용은 어떻게 하면 효과적으로 연구될 수 있을까? 사용자와 컴퓨터 사이의 상호작용을 관찰하기 위한 소프트웨어 평가에는 면접원-응답자 상호작용과 마찬가지로 실험실 절차가 주로 이용된다. 그러나, 이것은 사용자 또는 주체가 과장된 자각상태에 놓인 인공적인 환경으로, 결과를 현지설정에 일반화할 가능성이 우려되는 상황이다. 현지설정에서 자연스러운 상호작용을 관찰하는 것은 비용이 들 뿐 아니라, 제어되지 환경은 면접원간의 비교를 어렵게 만든다. 한편, 각본에 의한 모의면접은 비교적 자연스러운 설정과 모든 면접원에게 일관된 자극을 제공함으로써 이 두 가지 접근방식에 약간의 장점을 제공한다(Rustemeyer, 1977). 게다가, CAPI 활용의 특정한 면에 대한 면접원의 수행능력을 평가하는데 있어 도구에 내재된 시험을 허용하는 점에서 실험실 접근방식의 장점도 공유한다.

컴퓨터를 이용한 면접은 면접원이 누른 입력키의 기록을 모두 얻을 수 있다는 점에서 추가적인 자료의 출처를 제공한다. 비록 많은 CAPI 시스템에 의해 정례적으로 포획되었다 하더라도, 이러한 키입력 파일은(감사추적(audit trails) 또는 추적파일(trace files)이라고도 함) 면접원의 행동이나 직무수행능력을 평가하기보다는 시험 및 오류제거 목적으로 주로 이용된다(House and Nicholls, 1988 참조). 이러한 목적에서 키입력 파일을 이용할 때 한 가지 약점은 관리할 수 있고, 유의한 수준으로 줄이기 어려운 방대한 양의 자유형식 데이터가 생성되기 쉽다는 점이다. 그러므로, 면접원-컴퓨터의 상호작용과 발생하기 쉬운 오차 유형을 파악하는데 도움이 되는 이러한 데이터를 효율적으로 관리할 수 있는 방법의 개발이 중요하다 하겠다.

면접원의 직무수행능력을 평가하는데 있어 CAPI 조사가 PAPI 조사와는 다른 방법을 필요로 하기는 해도, 부분적으로는 그러한 평가를 용이하게 하는 도구를 제공하기도 한다. 그렇다면, CAPI 시스템과 면접원의 상호작용을 탐구하는데 있어 키입력 파일을 활용하는 것에 대해 논의해 보자. 우선은, 모의면접의 활용과 키입력 파일의 분석을 조합한다. 모의면접에서 일관된 질문과 답변의 사용은 키입력 파일의 상세한 분석을 용이하게 한다. 그리고, 두 번째로는 덜 제어된 현지설정에서 여러가지 CAPI 기능들을 활용하는데 있어 면접원간의 차이를 탐구하기 위해 현지면접으로부터의 키입력 파일을 조사한다.

컴퓨터화된 조사도구를 이용할 때 면접원들이 발생시키는 오차의 유형과 도수에 대해서는 거의 알려진 바가 없다. 이러한 오차의 도수와 유형 및 출처를 파악하는 것은 교육을 통해서든, 보다 효과적인 CAPI 도구의 설계를 통해서든, CAPI를 이용하면서 발생된 오차를 줄이기 위한 노력을 알리는데 이용될 수 있다(Couper, 1995 참조). 따라서, 핵심은 컴퓨터를 이용한 면접에서 면접원의 수행능력을 측정하고 평가하는데 이러한 도구의 유용성을 평가하는데 있다.

7.1 인간 - 컴퓨터 상호작용에서의 오차

CAPI 조사에서는 왜 면접원 오차에 관심을 가져야 할까? 컴퓨터를 이용한 면접은 지면을 이용한 자료수집에 비해 자료품질이 향상된다는 몇 가지 증거가 있으며(Nicholls et al., 제9장), 사실은 이것이 많은 이들이 새로운 기술을 채택하는 이유이다. CAPI 도구들은 자동 경로지정과 범위 및 일관성 확인, 그리고 적용되는 모든 문항의 완료를 보장하는 강력한 도구가 된다. 면접원들은 컴퓨터를 이용한 면접의 요건을 분명 따를 수는 있지만, 이러한 일을 효율적으로 진행하거나 혹은 조사설계자가 의도한 방법으로 하는 것에 관한 정보는 거의 없다.

CATI(Kennedy et al., 1990)와 CAPI(Dielman and Couper, 1995)에서 면접원의 입력 오차에 관한 연구들을 보면, 정해진 응답 내지는 폐쇄형 질문의 경우 면접과정에서 탐지되지 않은 입력 오차들은 거의 없는 것으로 드러나 있다. 그러나, 다른 설정에서(예, 스프레드쉬트 활용, 워드 프로세싱, 프로그래밍) 이루어진 컴퓨터 활용에 관한 연구들에는 예상했던 것보다 오차가 훨씬 자주 발생한다고 되어 있다(Shneiderman, 1992 참조). 이와 같이 뚜렷한 모순이 발생하는 한 가지 이유는 아마도 컴퓨터를 이용한 면접에서는 과정보다는 일의 결과만을(즉, 최종 데이터) 중시하는 협의의 오차 정의가 적용되기 때문일 것이다. 그리고, 또 다른 이유로는 컴퓨터의 다른 활용에 비해 컴퓨터를 이용한 면접업무의 복잡성이 상대적으로 낮은 것을 들 수 있을 것이다. 그러나, CAPI와 같이 면접원이 프로세스를 통제할 수 없고, 비교적 한정된 범위의 행위만 수행하는 폐쇄형 시스템이라 하더라도, 오차는 자료품질과 효율성 및 사용자 만족도에 영향을 미칠 수 있다. 특히, 면접원이 특정한 목표의 달성이 좌절되는 한(예, 이전 답변의 변경), 그 결과는 욕구불만의 증가에서 목표의 포기로까지(예, 답변을 변경하지 않고 그대로 방치) 나타날 수 있다.

이러한 상황에서는 면접원들이 가진 어려움이 무엇인지 또, 이들이 이것을 극복하려면 어떻게 해야 하는지를 탐구하기 위해 CAPI에서 면접원과 컴퓨터

간의 상호작용을 조사하는 것이 중요하다. 컴퓨터 면접에서 기록된 자료에만 초점을 맞추는 것은 도구 자체의 결점이나 면접원과 시스템간 인터페이스를 통해서건 아니면, 불충분한 교육이나 면접원의 역량부족을 통해서건, 면접원들이 당면하는 문제를 과소평가하는 것일 것이다.

따라서, 여기서는 광의의 견지에서 오차를 바라보고, 최종결과가 올바른 답변이든 아니든 상관없이, 조사면접 과정에서 컴퓨터와 상호작용함으로써 면접원들이 경험하게 되는 모든 종류의 어려움에 초점을 맞추고자 한다. 여기에는 목표달성을 위한 기타 차선책이나 면접원에게 유용한 CAPI 기능들의 비효율적 활용뿐 아니라, 탐지되어 수정되는 초기의 오차들도 모두 포함된다.

인간-컴퓨터 상호작용의 평가에는 몇 가지 오차 분류안이 제안되었다 (예, Reason, 1990; Rasmussen et al., 1987). 이 중 대부분은 Norman (1983)의 착오(mistake)와 실수(slip) 사이의 구별을 따른다. 실수는 올바른 의도가 잘못 실행되었을 때 발생하는 반면, 착오는 틀린 의도를 그대로 이행하는 경우를 말한다. 예를 들어, 어떤 사람이 키를 잘못 눌러 우발적으로 파일을 삭제했다면, 이것은 실수다. 그러나, 이 사람이 파일을 삭제할 의도를 가지고 있지만, 나중에 그것이 필요하다고 깨달은 경우에는 이미 착오가 발생한 것이다(Frese and Altmann, 1989 참조). 또, 이 안을 추가적으로 세분화한 안도 제안되었다. 예를 들어, Zapf et al.(1992)는 행동규정을 세 가지 수준으로 구분하고 있다. 지적 수준(사고 오차, 기억 오차, 판단 오차가 일어난 경우), 유연한 행동패턴의 수준(level of flexible action patterns, 습관이나 태만 또는 인식의 오차가 발생한 경우), 감각운동적 수준(sensorimotor level, 인쇄 오차 등이 포함). Zapf et al.(1992)는 비교적 복잡성이 낮은 시스템에서는(컴퓨터를 이용한 면접과 같은) 낮은 수준의 오차가(예, 감각운동적 오차) 발생할 가능성이 높다고 지적한다. 사실, 여러 등급의 오차들을 구분하는 것은 어려운 일이다. 왜냐하면, 그렇게 하려면 사용자의 의도를 알고 있어야 하기 때문이다. 이러한 의도는 행동데이터로부터 파악하기는 어렵다. 그렇지만, 여러 유형의 오차가 발생하는 수준을 이해함으로써, 우리는 적절한 오차감소 전략을 개발할 수 있을 것이다.

우리는 현재 어떤 종류의 오차가 발생하는지 혹은 면접원들이 CAPI 도구를 활용할 때 봉착하는 어려움이 무엇인지를 알기 위해 일반적인 Norman(1983)의 분류체계 속에서 CAPI 조사의 키입력 파일을 조사한다. 비록 지금은 면접원 오차 전반에 걸쳐 언급하고 있지만, 모든 오차가 반드시 면접원들에게서 기인한다는 것을 뜻하는 것은 아니다. 오차는 면접원에게 기인할 수도 있고(예, 교육부족이나 부주의 또는 동기부족), 또 컴퓨터나(예, 하드웨어나 소프트웨어의 고장 내지는 설계상의 결함) 인터페이스에(예, 면접원 요구와 컴퓨터 요건 사이의 불일치) 기인할 수도 있다.

7.2 설계 및 자료수집

본 연구는 미시건 대학 Survey Research Center(SRC)에서 70대 이상의 고령층을 대상으로 수행한 Study on Asset and Health Dynamics of the Oldest Old(AHEAD)의 일부분이다. 표본은 약 9,500가구와 12,000명의 개인으로 구성되어 있다. 현지 면접원은 전화조사와 개별방문면접 방식을 둘 다 이용하였고, 조사자료의 수집에는 랩탑 컴퓨터를 활용하였다. 우리는 자료수집의 분산된 성격을 강조하기 위해 CAPI로 두 가지 면접방식을 모두 다루었다. AHEAD 도구는 Microtab Systems of Australia에서 개발한 SurvayCraft를 이용하여 프로그램밍 되었다.

1993년, 총 137명의 면접원이 AHEAD 연구를 위해 교육을 받았고, 이러한 교육 후 실전면접이 실시되기 전에 각 면접원은 각자의 조사감독원들과 함께 동일한 상황에서 연출된 모의면접을 수행하였다. 면접원들에게는 이 면접을 실제 또는 "현장" 면접으로 생각하도록 지시하였고, 조사감독원들은 모든 면접원에게 대등한 자극이 주어지도록 응답자의 역할을 수행하는 교육을 받았다. 이것은 전화조사로 이루어졌으며, 면접원들은 랩탑 컴퓨터상에서 CAPI 도구를 이용하였다. 또한, 조사감독원들은 지시를 잘 따르는지 확인하기 위하여 각 모의면접을 테이프에 녹음하였다. 이와 같이 연출된 모의면접은 면접기술과(예, 탐색) CAPI 기능이라는(예, "모름" 또는 답변거부의 입력, 이전화면에서 답변 변경) 두 가지 면에서 특정한 몇 가지 시험을 위해 계획된 것이었다.

AHEAD 연구를 위해 교육받은 137명의 면접원은 모두 교육 후 모의면접을 완료하였다. 그러나, 전적으로 사용가능한 키입력 파일은 125명의 면접원으로부터만 수신되었다. 나머지 12개 중 6개의 키입력 파일은 제대로 전송되지 않았고, 6개는 수신되었지만 불완전한 상태였다.

유사한 케이스 손실은 실전 키입력 파일에서도 나타났다. 이러한 키입력 파일들은 각 면접자가 현지에서 정기적으로 전송하였다. 완료된 총 8,223건의 AHEAD 면접 중 입수된 키입력 파일은 7,222건(혹은 88%)에 불과하였다. 또, 이 중 754건은 면접시작과 함께 시작되지 않았다(즉, 보다 빨리 중단한 후 다시 시작하였다). 따라서, 이러한 건들은 모두 버리고, 나머지 6,468건의 키입력 파일만 분석에 들어갔다. 이것은 면접원 1인당 평균 47.6건에 해당한다(표준오차 = 2.03).

이것은 SurveyCraft 소프트웨어를 이용하여 실전조사(production survey)에서 키입력 파일을 저장 및 전송하는 것에 관한 첫번째 시험이었으며, SRC에서 개발한 CAPI 표본관리시스템(CSMS)의 최초의 실전적용(production implementation)이었다. 이후에는 SurveyCraft와 CSMS에 많은 개선이 이루어졌으며, 앞으로의 연구에서 키입력 파일의 수집, 검토 및 분석을 용이하게 할 것이다.

이러한 케이스 손실이 발생한 상황에서, 우리는 전체 AHEAD 면접에 대한 추론에 관해 신중을 기하였다. AHEAD 면접에서 상응하는 키입력 파일이 있는 것과 없는 것 사이의 차이를 탐구하기 위한 분석에서 면접원 특성에 따른 차이는 없는 것으로 드러났다. 즉, 케이스 손실은 면접원의 조사경험이나 컴퓨터 경험에 따라 발생되는 것으로 보이지 않았으며, 오히려 조사기간 후반부에 완료된 면접과 마지못해 조사에 응한 응답자로부터 입수된 키입력 파일이 손실되는 경향이 약간 있었다. 만일 그렇다면, 이것은 오차의 수준은 가용한 케이스로부터 약간 과소평가될 수는 있지만, 면접원 특성에 따른 키입력 행동의 비교는 그러한 손실에 아무런 영향을 받지 않는다는 사실을 제시하는 것이다.

7.3 분석

모의면접에 대한 키입력 파일은 두 가지 방법으로 조사되었다. 첫째는 도구 속에 몇 가지 시험이 내재된 상황에서 각각의 시험에서 면접원의 수행능력을 요약할 수 있는 코드부여 방안(coding scheme)이 개발되었다. 둘째는 면접원별로 여러 가지 기능키를 누른 횟수와 사용한 키조합에 관한 정보가 입수되었다. 이러한 횟수는 이후 가장 효율적인 경로를 나타내는 "템플릿" 또는 모델 키입력 파일과 비교되었다. 똑같이 수용가능한 접근방식을 선택할 수 있는 경우에는 면접원 교육에서 강조된 방법이 템플릿 파일을 생성하는데 채택되었다.

수신된 키입력 파일이 방대하고, 현지에서 면접원들이 경험한 상황이 다양한 경우, 이러한 케이스들의 분석은 여러가지 키입력 행동의 총횟수에서부터 시작되었다. 그 다음, 특이점(outlier)을 파악하여 보다 적은 키입력 파일의 상세한 조사에 들어 갔다.

키입력 파일에 대한 이 두 가지 접근방법은 상호보완적이다. 첫번째에는 행동 코드부여와 유사한 방식의 각 키입력 파일에 대한 상세한 검토와 코드부여가 포함되어 있다. 이것은 면접시 다양한 차원에서 어떠한 일이 발생하는 가에 관한 풍부한 자료를 도출한다. 우리는 이 방법의 사용을 모의면접에 확실하게 포함되어 있는 일련의 시험 및 특정한 문제를 내포한 것으로 확인되는 경우에만

한정한다. 두 번째 접근방법은 시간절약적이고 노동집약적이지만, 상세한 자료는 산출하지 못한다. 이 접근방법은 면접이나 면접원별로 지정된 키입력 행동의 총계를 집계한다. 집계된 키입력 횟수는 한 면접원이 특정한 키를 몇 회 눌렀는지를 나타내지만, 그러한 키의 사용이 그 시점에서 적절했는지 혹은 성공적이었는지에(즉, 원하는 목표를 달성하였는지) 관한 정보는 제공하지 못한다. 우리는 주변 키입력의 상세한 조사에 의해서만 면접원의 의도를 추측할 수 있다. 그러나, 총 횟수도 보다 상세한 절차를 이용하면 면밀한 주의를 필요로 할 수 있는 면접이나 면접원을 식별하는데 도움이 될 수 있다. 이와 같이, 각각의 분석방법은 서로 다른 유형의 키입력 행동에 초점을 맞추고 있다.

AHEAD 도구의 특수 기능으로 기능키와 다른 키조합에 관한 사항이 표 11.1에 나와 있다. 다른 기능들은 SurveyCraft 시스템과 같다(예, 되돌아가기(jump back) 옵션과 여러가지 편집기능). 그러나, AHEAD 연구의 면접원들은 여기에 대한 교육은 받지 않았다.

표 7.1 AHEAD 도구에 사용된 기능키

키	기능
[F1]	질문별 도움말 (QxQ). [Esc]을 눌러 빠져나옴
[F2]	설명/주석달기. [Enter][Enter]를 치면 질문으로 되돌아 옴
[F3]	가구명단(household roster) 화면에 있는 팝업 메뉴
[F4]	기능 없음
[F5]	자료를 저장하고, 면접을 계속함
[F6]	기능 없음
[F7]	보충질문(backup) [F9] 후 다음 질문으로 이동
[F8]	기능 없음
[F9]	보충질문
[F10]	이전에 입력한 응답 복구(보충질문시 사용).
[Alt D]	모름
[Alt R]	답변 거부
[Tab], [Shift Tab]	명단 화면에서 문항간 이동
[Up],[Down],[Left],[Right]	텍스트 입력, 주석달기 및 명단 화면에서 이동
[Backspace]	삭제
[Esc]	입력취소 또는 도움말 화면에서 빠져나오기

7.4 모의면접 키입력 파일의 평가

모의면접에 포함된 테스트들의 상세한 조사결과는 많은 곳에서 논의되어 있다(Couper et al., 1994). 그러므로, 여기서는 면접원의 수행능력평가에 있어 키입력 파일과 결합된 모의면접의 유용성에 대해 간략히 요약하도록 하겠다.

일반적으로, 모의면접 테스트의 키입력 파일은 면접원들이 교육받은 직후라 하더라도, 가장 공통적인 CAPI 기능을 사용하는데 있어서는 아무런 어려움을 겪지 않는다는 사실을 보여준다. 하지만, 모의면접에서 특정한 상황을 설정하여 지정된 CAPI 기능을 사용함에 있어서는 다소간의 차이가 나타난다. 예를 들어, 면접원들에게 [F1][Esc] 키조합을 이용하여 질문별로 지정된 도움말을 제공하도록 하는 세 가지의 테스트가 실시되었다. 이 세 가지 테스트 모두에서 도움말 화면을 이용한 면접자들은 거의 없었던 반면, 모의면접 내내 그러한 도움말을 한번도 사용하지 않은 면접원은 7%에 불과했다.

이와 유사하게 [F9] 키를 이용하여 보정(back up)하고, 이전 응답을 변경한 경우는 세 가지 테스트에 걸쳐 82~95%까지 나타났지만, 세 가지 테스트를 완료한 125명의 면접원 중 20명이 최소 한 번 이상 수정에 실패하였다.

모의면접의 키입력 파일은 각 면접원의 다양한 키입력 일치의 횟수를 집계하여 모의면접모델과 비교하는데도 이용된다. 이러한 비교는 특정한 기능들이(예, [F1], [F9], [F7]) 예상보다 자주 사용되지 않았음을 보여준다. 또 한편으로, 면접원의 51%가 모의면접에서 요구된 것 이상으로 [F2]를 자주 사용하여 코멘트를 입력한 반면, 43%는 그 이하로 [F2]를 사용하였다. 그 외에도 코드목록이 팝업되는 [F3]를 요구수준보다 자주 사용한 면접원은 60%였으며, "모른다"는 응답을 위한 [Alt D]는 74%의 면접원이 빈번하게 사용하였다.

모의면접 키입력 파일은 모델 키입력 파일과 비교하여 면접원들이 얼마나 "효율적으로" CAPI 기능들을 사용하고 있는 지를 파악하는데도 이용될 수 있다. 평균적으로, 면접원들은 모의면접을 완료하는데 1,563번의 키입력을 하였으며, 이것은 모델 키입력 파일(1,603)보다 약간 적은 수다. 그러나, 후자에는 두 가지 개방형 질문에(산업 및 직업) 대한 장문의 응답뿐 아니라, 상세한 설명이(질문반복, 탐색 등과 같은 면접원 행동에 관한 주의사항 포함) 포함되어 있음에 주의해야 한다. 사실, 이 파일에서 총 키입력의 1/3 (530/1,603)은 설명이나 개방형 질문에 들어 있는 텍스트 형식의 답변들이다. 비효율적인 키입력 사용의 또 다른 지표로는 면접원들이 커서 키를 평균 45.2회 사용한 반면,

템플릿 파일에서는 단 한 가지 경우에만 이 키의 사용이 요구되어 있다는 점이다. 따라서, 결론적으로 볼 때 면접원들은 교육을 받은 직후 중요한 CAPI 기능들을 사용할 수 있는 것으로는 나타났지만, 그 횟수와 효율성에 있어서는 다양한 차이를 보이고 있다. 게다가 텍스트 형식의 설명(주석)은 예상보다 훨씬 적게 입력했지만, 이것을 수정하는 데는 또 시간이 걸리는 것으로 나타났다.

7.5 실전 키입력 파일의 평가

면접원들이 제공된 여러 가지 CAPI 기능들을 어느 정도 사용하는 지에 관해 알려진 바는 거의 없다. Baker(1992)는 면접원 보고형식의 설문을 이용하여 면접들이 100회의 면접 당 1번 미만꼴로(0.6%) 도움말 화면을 이용하는 반면, 면접의 22.4%에서 답변을 수정하거나 변경한 것으로 보고한 사실을 발견하였다. 하지만, 이 외에는 여러 가지 CAPI 기능들이 어느 정도 빈번하게 사용되는 지에 관한 정보가 전혀 없다. 이러한 정보는 몇 가지 이유에서 매우 유익할 수 있다. 한 예로, 도움말 화면을 개발하고 면접원들에게 이의 사용법을 교육하는데 상당한 노력이 투입되는 것으로 알고 있는데, 만일 이러한 화면을 거의 활용하지 않는다면, 여기에 대한 투자를 다시 생각해 볼 필요가 있다. 또, 비슷하게는 주로 사용되는 면접원 활동들을 파악하여 교육의 초점을 적절히 맞출 수가 있다. 하지만, 불행하게도 이러한 척도는 다음과 같은 정보를 제공하지는 못한다. (a) 면접원이 기능을 제대로 사용하는지 아닌지, (b) 기능이 이들의 요구를 충족하는지 아닌지(예, 제공된 도움말이 유용한지), (c) 기능을 자주 사용하지 않는다는 것은 필요성을 못 느낀다는 것인지 아니면, 면접원이 그 기능을 사용하지 못한다는 것인지, (d) 면접원들이 좋아하거나 원하지만 제공되어 있지 않는 기능들은 무엇인지. 그럼에도 불구하고, CAPI 기능 활용도의 측도는 CAPI를 위한 최적의 도구설계에 관한 추가적인 연구개발의 지침이 되고, 면접원들의 효과적인 기능 사용을 평가하는데 있어 최소한의 초기정보를 제공해 준다.

표 7.2 실전면접에서 여러가지 CAPI 기능의 활용

키입력	한번 이상 사용한 면접 비율	한번 이상 사용한 면접원 비율	전체 면접에서의 평균	최초 5회 면접에서의 평균	마지막 5회 면접에서의 평균
[F1] 도움말	20.6	87.5	0.43	0.73	0.28
[F2] 설명	92.0	100.0	10.57	11.91	9.42
[F3] 팝업 메뉴	66.6	100.0	7.70	7.91	7.13
[F7] 보정 후 복귀	16.2	84.6	0.35	0.52	0.28
[F9] 보정	93.8	100.0	7.47	8.05	7.30
[F10] 현재 응답 복구	30.2	94.1	1.35	1.65	1.13
[Bs] Backspace	85.9	97.8	15.53	17.40	13.31
[Esc] 취소	84.7	97.8	5.98	6.15	6.08
잘못된 기능키	6.3	86.0	0.12	0.16	0.13
커서 키	-	-	47.67	55.96	41.51
모든 기능 키	-	-	28.40	31.47	26.00
텍스트 설명	-	-	258.00	284.70	232.20
전체 키입력	-	-	1322.35	1369.45	1245.29

표 7.2는 AHEAD 실전면접에서 여러 가지 키입력 행동의 측도를 요약한 것이다. 첫번째 행의 백분율은(전체 면접의 21%) 도움말 화면이 Baker(1992) 자료가 제시한 것보다 훨씬 자주 이용되고 있음을 의미한다. 이것은 보정기능(이전 응답 수정)에서도 마찬가지로 나타나 있다. 게다가, 이 두 가지 기능은 모의면접보다 실전면접에서 평균적으로 이용율이 낮다. 이것은 표 7.2의 마지막 두 행에 나타나 있는 전체적인 감소와 일치한다.

설명 또는 주석을 입력하는 [F2] 키의 사용은 전체 면접의 92%에서 사용되어 빈도가 높게 나타나 있다. 또, [F7] 키를 이용해 효율적으로 되돌아가는 기능은 그다지 자주 사용되지 않았지만, 대다수의 케이스에 응답을 변경하기 위한 보정기능이 포함되어 있다. 표 7.2의 두번째 행에서는 대부분의 면접원들이 최소한 번은 이러한 기능들을 사용할 수 있음을 확인할 수 있다. 예를 들어, 온라인 도움말을 사용하지 않은 면접원은 전체의 12.%에 불과하다.

표 7.2의 나머지 행들은 전체 실전면접에서 여러가지 키입력의 평균적인 사용을 나타내고, 특히 최초 5회와 마지막 5회 면접은 면접원 각자가 작성한 것이다. 일부 기능들은 면접원들이 랩탑 컴퓨터에 익숙해짐에 따라 시간이 지날수록 이용도가 증가할 것으로 예상된다. 반면, 다른 기능들과(온라인 도움말 등) 잘못되거나 비효율적인 사용은 연구가 진행되면서 점차 줄어들 수도 있다. 키입력 총 수는 최초 5회부터 마지막 5회까지 감소하고 있으며, 이것은 연구가 진행됨에 따라 면접 소프트웨어를 더 효율적으로 사용하고 있음을 의미하며, 특히, 커서 키와 백스페이스 키의 사용이(수정시에도 사용됨) 줄어드는 것은 이를 뒷받침한다. 반면, 조사기간 전반에 걸쳐 사용되어야 하는 기능들은([F2], [F9], [F3]와 같은) 시간이 지나도 줄어들지 않는데 이것은 고무적인 현상이다. 한편, 감지된 효과를 유발하는 원인에 대해서는 주의를 기울여야 한다. 예를 들어, 온라인 도움말의 사용이 감소하는 것은 (a)면접원들이 제공된 정보를 더 이상 이용할 필요가 없다고 느끼거나 (b)그러한 정보가 별 도움이 되지 않는다고 느끼거나 (c)시간이 갈수록 그러한 기능의 사용법을 잊어버렸기 때문일 지도 모른다.

잘못된 기능키의 도수가 낮은 것은 주목할 만하다. 이것은 [F5] 키를(면접 연장을 위해) 문맥에서 사용하지 않은 것을 비롯하여 AHEAD 도구에서 활성화되지 않는 기능키들을([F4], [F6], [F8]) 말한다. 이런 키들은 전체 면접의 6.3%에 불과하고, 평균적으로 보면 8회의 면접당 한 번꼴로(1/0.12) 나타났다. 그러나, 전체 면접원의 86%는 이 중 하나를 최소 한 번 이상 눌렀던 것으로 나타나 있다. 그러므로, 이와 같은 낮은 빈도에도 불구하고, 그러한 오류를 범하지 않은 면접원은 거의 없다는 결론을 내릴 수 있다. 이것은 앞으로의 조사에 귀중한 자료가 된다(7.7절 참조).

표 7.2로 다시 돌아가, 보충설명이나 주석에 텍스트를 입력한 평균량은(면접당 평균 문자수) 처음에서 마지막으로 면접이 진행되면서 점점 줄어들고 있다. 이것은 면접원들이 코멘트를 활용하는데 점점 효율적으로 되어 가고 있는지, 아니면 점점 손쉬운 방법을 택하고 있는지를 나타내는 지는 알 수 없다. 재미있는 것은 AHEAD 면접에서 전체 키입력의 3.6%가 커서 키이고, 2.1%는 기능키인 반면, 전체 키입력의 19.5%가 설명란의 텍스트 입력이라는 사실이다. 이것은 여러가지 활동이 일어나는 상대도수의 개념을 제시하는 것이다.

이와 같은 총 횟수는 바람직한 키의 사용은 증대시키고, 비효율적인 행동의 발생 및 잘못된 기능키의 사용은 감소시킬 수 있도록 설계와 교육자원을 할당하는 데는 유용한 반면, CAPI 도구와 상호작용하는 면접원의 수행능력 면에서의 차이는 발견할 가능성이 없다. 따라서, 이러한 쟁점은 다음 장에서 계속해 고찰하도록 하겠다.

7.6 면접원의 키입력 행동차

면접원의 수행능력을 평가하는 도구로 유용하기 위해서는 CAPI 수행능력과 관련하여 면접원 속성에 따라 체계적인 차이를 보이도록 키입력 파일을 예측해야 한다(Couper and Burt, 1994 참조). 이러한 차이는 면접 초기단계에서 가장 클 것으로 믿고 있으며, 따라서 우리의 분석은 각 면접원이 수행하는 최초 5회의 면접에 초점을 맞춰야 한다.

표 7.3 최초 5회 면접에서 조사경험별 평균 키입력

키입력	면접자 조사경험		
	1년 미만	1년~5년	5년 이상
[F1] 도움말	1.19[a]	0.80	0.48[b]
[F2] 코멘트	14.06	12.65	10.31
커서 키	61.39	57.16	52.58
전체 기능키	33.37	34.56[a]	27.66[b]
잘못된 기능키	0.11	0.24	0.11
코멘트 텍스트	351.9[a]	299.5	242.9[b]
전체 키입력	1428.2	1408.2	1307.7
(면접원 수)	(23)	(56)	(57)

주 : 윗 첨자가 다른 값들은 $p<.05$에서 유의한 차가 있음.

입력된 코멘트의 수는 조사경험에 비례할 것으로 예상할 수 있다. 초보 면접원들은 응답과정에만 집중하고, 추가적인 보충설명의 필요성은 인식하지 못할 수가 있다. 또 다른 한편으로, 경험이 많은 면접원들은 어떤 설명이 필요한지에 대해 더 나은 판단을 할 수 있고, 이에 따라 필요없는 설명은 삭제하기도 할 것이다. 표 7.3은 조사경험에 따라 세 부류로 나누어 지정된 키입력 행동을 비교한 것이다. 여기에 있는 데이터들은 온라인 도움말의 사용이 그랬던 것처럼, 조사경험에 따라 보충설명의 사용이 약간씩 감소하고 있음을 보여준다. 또, 경험이 많을수록 편집행위가 적게 나타난다(커서 키 사용의 감소가 증거). 비록 이러한 일반적인 경향이 세 가지 변수 모두에서 나타나기는 하지만, 상대적으로 적은 수의 면접원을 대상으로 할 때에는 통계적 유의도($p<.05$)를 항상 가지는 것은 아니다. 설명란에 입력된 텍스트의 양은 조사경험에 따라

현격히 줄어든다. 또, 조사경험이 가장 많은 면접원들은 경험이 없거나 중간정도인 면접원에 비해 기능키를 훨씬 덜 사용하는 것으로 나타나 있다(통계적 유의도는 후자의 경우에만 있지만). 이러한 경향들은 조사경험이 많을수록 키입력 총수가 감소하는 것을 설명하고, 또 이것은 경험이 많은 면접원들은 CAPI 도구를 보다 효율적인 이용하는 사용자일 수 있다는 사실을 제시해 준다.

컴퓨터 경험과 타이핑 능력에 관한 한은 둘 중 한 가지 방향의 가정을 세울 수 있다. 즉 한편으로는, 컴퓨터나 키보드에 익숙한 사용자들은 키입력 오류를 적게 범할 것으로 생각할 수 있고(즉, 잘못된 기능키는 잘 누르지 않고, 수정할 때 커서 키를 잘 사용하지 않을 것이라는), 다른 한편으로는 이 그룹은 컴퓨터에 거부감이 없으므로 수정하는데 여러가지 기능키와 커서 키를 오히려 더 많이 사용하려는 경향이 있을 것이다.

표 7.4 최초 5회 면접에서 컴퓨터 경험별 평균 키입력

키입력	면접자 컴퓨터 경험		
	없음	보통	많음
[F1] 도움말	0.36[a]	1.03[b]	0.76[b]
커서 키	42.39[a]	49.38	0.76[b]
백스페이스	13.47[a]	15.86	20.26[b]
전체 기능키	28.29	33.98	31.73
잘못된 기능키	0.094	0.18	0.19
전체 키입력	1308.2	1360.6	1405.8
(면접원 수)	(34)	(36)	(66)

주: 윗 첨자가 다른 값들은 $p<.05$에서 유의한 차가 있음.

표 7.4에 나타나 있는 컴퓨터 경험별 평균 키입력 횟수는 후자의 가정을 뒷받침한다. 컴퓨터 경험이 많은 면접원들은(즉, 정기적으로 컴퓨터를 사용하는 면접원들) 경험이 없는 면접원에 비해 수정 키를(커서 키와 백스페이스) 많이 사용한다. 이것은 컴퓨터 경험이 많을수록 여기저기 옮겨 다니면서 변경하는데 불편함을 느끼지 않는다는 사실을 제시한다. 전체 키입력 수도 마찬가지로 컴퓨터 경험이 많을수록 증가하고 있다.

스스로 평가한 타이핑 능력에 관한 경향도(표 없음) 컴퓨터 경험과 비슷한 방향성을 보인다. 커서와 백스페이스 키의 사용은 타이핑 능력에 따라 증가하는 현상을 보이며, 이것은 타이핑 능력이 뛰어날수록 편집행위를 많이 한다는 사실을 제시해 준다. 또한, 이들은 속도를 위해 초기의 정확성에는 그다지 신경쓰지 않는 등, 응답을 입력하는데 많은 주의를 기울이지 않는다는 것을 의미하기도 한다. 반면, 스스로 "독수리 타법"을 자처하는 면접원들은 나머지 두 그룹에 비해 기능키의 사용과 코멘트 입력, 코멘트란에 입력한 텍스트의 양, 그리고 전체 키입력 횟수에 있어 평균적으로 낮은 결과를 보이는 것으로 나타났다. 이것은 타이핑 속도에는 랩탑 키보드의 작은 크기가 제약이 될 수 있는 것으로 일부 설명될 수 있으나 키보드에 적당히 친숙해 지는 것이 최적일 것이라는 사실을 제시한다.

여러가지 면접원 속성과 잘못된 기능키의 사용에는 뚜렷한 성향이 없는 것으로 나타났다. 그 중 한가지 원인으로는 이러한 현상의 빈도가 상대적으로 매우 낮은 것을 들 수 있고, 또 다른 원인으로는 횟수를 집계할 때 잘못된 기능키 입력의 극히 일부만이 포착된다는 사실이다. 유효한 기능키가(예, [F1]) 잘못된 맥락에서(예, [F2]를 누르려 애씀) 눌러지는 경우는 허다하게 있을 수 있다. 한 가지 증거로는 [F1] 키의 총 도수와 [Esc] 키와 함께 사용된 도수 간의 차이를 들 수 있다. 표 7.2를 보면, 도움말을 활용하는데 [F1][Esc] 키는 최초 5회 면접에서 평균 0.73회 사용되었지만, [F1] 키만을 놓고 볼 때는 평균 1.02회 사용된 것을 알 수 있다. 이것은 [F1] 키 사용횟수의 약 1/4 가량이 도움말을 불러오는데 제대로 사용되지 않았다는 것을 의미한다. 비슷한 예로, [F2] 키의 사용도 약 1/12은 코멘트를 입력하는데 이용되지 않았다.

그렇지만, 이러한 비교들은 CAPI 도구를 사용하는데 있어 면접원 속성에 따른 차이들은 기껏해야 적당한 수준임을 밝혀준다. 게다가, 이러한 차이는 시간이 지날수록 점차 줄어든다. 실제로, 마지막 5회 면접이나 전체 면접에 대한 평균을 이러한 속성들과 비교해 보면, 표 7.3과 7.4에서 발견되는 유의차가 없어지는 경향이 나타난다. 그러므로, 초기에 논의된 오류와 비효율성은 특정한 특성들을 가진 면접원들 사이에서 나타나는 것이 아닌 것으로 결론 내릴 수 있다.

7.7 잘못된 기능키 사용에 관한 고찰

앞서 지적한 바와 같이, 실전면접 전반에서 키입력 행동의 총 횟수는 CAPI 시스템을 활용하는데 있어 광의의 성향만을 나타낼 뿐이다. 그러나, 이러한

횟수들은 개개의 면접을 상세히 검토하는데 도움이 될 수 있다. 따라서, 잘못된 기능키가 한번 이상 사용된 408건의 면접을 대상으로, 무효 키가 눌러진 배경을 고찰하고, 그 때 면접원이 무엇을 하려고 했는지를 파악하기 위해 각각의 파일을 하나하나 면밀히 검토하였다.

오류(error)의 종류는(실수든 착오든) 주변 키입력 행동으로부터 추측되고, 따라서 이것은 의도를 잘못 해석하는데 아주 민감하다는 사실을 주목해야 한다. 우리는 명백한 실수(slip)가 발생하지 않은 한, 오류(error)를 착오(mistake)로 분류하는 과오를 범한다. 다시 말해, 오류는 주변의 숫자나 기능키가 의도된 목표였던 경우에만 실수로 분류된다. 실수는 면접원이 올바른 의도를 가지고 있었지만 키 입력을 잘못한 경우를 말하는 반면, 착오는 올바른 기능키가 무엇인지를 제대로 알지 못하는 경우에 발생한다. 일부 오류들은 면접원이 의도된 행위를 포기함에 따라 의도한 목적이 무엇이었는지가 불분명한 것도 있다.

한 번 이상 잘못된 기능키가 입력된 408건의 면접에서는 잘못 사용된 횟수가 총 489회로 나타나 있다. 이것은 표 7.5에 요약되어 있다. 결과는 Zapf et al.(1992)과 일치하는 것으로 나타났는데, 여기서 우리는 낮은 단계의 기능일수록 더 많은 오류가 발생한다는 사실을 발견하였다. 다만, 4가지 기능키에서는 상당한 차이가 있었다. 예를 들면, [F8] 키를 이용한 오류 중 많은 것은 보정용으로 빈번하게 사용되는 [F9] 키 바로 옆에 있기 때문이다. 반면, [F5] 키에는 비슷한 오류가 거의 없는데 이것은 두 개의 무효 기능키([F4]와 [F6]) 사이에 끼어 있기 때문이다.

표 7.5 잘못된 기능키 활용의 분류

오류 유형	오류 발생율				
	[F4]	[F5]	[F6]	[F8]	전체
실수:					
옆 키 대신 기능키 잘못 누름	35.2	-	1.9	66.4	34.2
옆 숫자키 대신 기능키 누름	15.2	38.3	73.1	11.2	23.2
합계	50	38.3	75.0	77.6	57.5
착오:					
옆 키 아닌 기능키 대신 기능키 잘못 누름	5.6	3.3	-	-	3.5
ID 수정	1.1	31.7	7.7	0.9	5.5
의도 모름	43.0	26.7	17.3	21.5	33.5
합계	49.6	61.7	25.0	22.4	42.5
(오류 횟수)	(270)	(60)	(52)	(107)	(489)

AHEAD 면접에 사용된 랩탑 컴퓨터에는 기능키들이 수평상의 약간의 차이만 두고, 숫자키 바로 옆에 붙어 있다. 즉, [F6] 키는 [6] 키 우측상단에 있는 것이다. 이것은 기능키들이 인접한 숫자키 대신에 눌러지는 오류의 가능성이 크다는 사실을 설명한다. 랩탑 컴퓨터의 이러한 키의 근접성과 사이즈로 인해, 면접원들은 실제로는 동시에 두 개의 키를 누르는 경우가 많고, 다른 경우로는 숫자키 대신 기능키를 누르는 경우가 많이 생긴다. 잘못 눌러진 기능키가 아무런 기능이 없는 경우에는(예, [F6]) 그에 따른 영향도 전혀 없다. 하지만, 잘못 눌러진 키가 다른 결과를 초래하는 경우에는 그 영향이 보다 심각할 수 있다.

오류는 [F4] 키가 포함된 경우에 특히 분류하기가 어렵다. 의도를 알 수 없는 [F4] 키 사용의 대부분은 "예" 또는 "아니오" 응답에서는 다른 이상한 키입력 행동 없이 각각 [1]이나 [5]와 함께 혼합되어 나타나고, 이것은 근접성의 문제로 볼 수가 없다. 한 가지 설명할 수 있는 것은 질문란과 입력란 사이에 "[F4]-Search"라는 지시가 화면에 나타난다는 점이다. 비록 이것이 AHEAD 도구에서는 아무런 기능이 없다 하더라도, 거의 매 화면에서 나타나기 때문에 면접원들은 이를 사용하려고 시도할 수 있을 것이다. 이와 대조적으로 [F4] 키가 포함된 실수의 대부분은 명단(roster) 화면에서 팝업메뉴를 불러오기 위해 [F3] 대신 눌러진 경우에 발생한다.

면접 시작시에는 각각의 계정에 13자리 숫자의 표본 ID를 입력하게 되어 있는데 이 때에도 착오가 발생한다. 숫자를 입력할 때 이러한 오류는 발생하기 쉬운 반면, 면접원들은 ID 입력을 수정하는데 많은 어려움을 느낀다. 면접원들은 ID 입력을 취소하고 다시 시작할 때 주로 [F5] 키를 많이 누르는 것을 볼 수 있는데, 이것은 중지기능([F5])과 취소기능([Esc]) 사이에 혼돈을 일으키기 때문일 것이다.

이러한 분석은 유효한 기능키의 부적절한 사용에 초점을 맞추고 있는 것이 아님을 기억해야 한다. 다시 말해, 동일한 오류는 [F1] 키나 [F2] 키에서 발생할 수도 있고 [F4] 키나 [F8] 키에서도 발생할 수 있다. 여기에 관한 증거는 [F2] 키의 총 횟수와 실제로 작성된 코멘트 수 간의 차이에서 발견할 수 있다(7.6절 참조). 전체 면접에서 모든 키입력을 조사한 결과, 코멘트 없이 [F2] 키가 눌러진 경우가 1,523회인 것으로 나타났다. 이러한 오류들은 면접원들이 의도하지 않은 결과를 초래할 수 있다는 점에서 악영향이 클 수 있지만, 반면 여기서 논의된 4가지 기능키들은 대부분의 경우 단순히 CAPI 시스템의 영향을 받지 않는다.

7.8 논의 및 결론

본 절에서는 면접원의 CAPI 사용능력을 평가하기 위한 도구로서 키입력 파일에 대한 초기평가에 관해 다루었다. 이러한 관점에서 볼 때, 면접원들은 모의면접과 실전면접에서 모두 교육받은 CAPI 기능들을 사용할 능력은 분명 가지고 있는 것으로 보인다. 그러나, 두 가지 키입력 파일을 면밀히 살펴보면, CAPI 시스템을 사용함에 있어 거의 모든 면접원들에게서 다양한 오류와(실수와 착오 둘 다) 비효율적인 행동이 발견된다. 이러한 오류들은 자료품질에 대한 효과는 크지 않지만(대부분의 경우 면접원들은 오류를 복구하거나 어려움을 헤쳐나갈 수 있다) 효율성을 감소시키고, 욕구불만을 증대시키며, 면접원-응답자의 상호작용에 부정적인 영향을 미칠 수 있다. 이러한 어려움들의 대부분은 CAPI 조사에서 면접원 행동을 연구하기 위해 사용되는 현재의 도구로는 발견하기가 어렵다.

키입력 파일은 시스템을 사용하는 면접원의 능력차 만큼이나 도구 또는 시스템의 부적합성과 관련해서도 많은 것을 알려준다. 그러므로, 우리는 키입력 파일이 이러한 두 가지 목적에서 모두 유용한 도구라고 믿고 있다. 지금까지 논의한 바를 보면 먼저 키입력 파일 데이터의 분석으로 드러난 오류 감소의 쟁점을 다루었고, 다음으로 면접원의 평가에 대한 쟁점으로 돌아왔다.

키입력 파일에서 얻은 정보로 우리는 무엇을 할 수 있을까? 키입력 데이터에서 얻을 수 있는 정보로는 CAPI에서 오류를 감소시키는 일반적인 세 가지 전략 - 오류 방지, 오류 탐지, 오류 복구 - 이 있다. 이 중 일부는 실수에, 나머지는 착오의 감소에 목표를 둘 수 있다.

감지된 많은 오류의 발생을 방지하는 데는 여러 가지 조치를(소프트웨어와 하드웨어의 설계 및 교육 등) 취할 수 있다. 또, 분명한 것은 기능 키 사용에서 발생되는 많은 착오는 키 템플릿을 제공함으로써(화면상에나 랩탑 키보드에 탑재하여) 감소시킬 수 있다. 비록 대부분의 면접원들이 여러가지 기능들의 사용법을 알고 있다고는 하나, 특정한 CAPI 기능들에 대해서는 잊고 있거나 혼돈을 일으킬 때가 많으며, 이것은 드물거나 어려운 면접 상황에서 특히 더하다. 따라서, 기능을 상기하기 보다는 인식을 믿는 것이 CAPI 도구의 사용과 관련된 인지적 부담을 더는데 도움이 될 것이다(Couper, 1995 참조).

감지된 몇 가지 실수는 랩탑 컴퓨터 키보드의 설계에 기인한다. 이러한 결과들은 기능키와 숫자키의 근접성이 문제이며, 특히 숫자의 입력이 훨씬 많은

경우에 더 심각하다는 사실을 제시한다. 또한, 랩탑 컴퓨터 키보드에 두 줄로 되어 있는 키들을 분리하면 잘못된 기능키 사용이 훨씬 줄어들 것이다. 이러한 쟁점들에 대해서는 별로 다루어진 바가 없지만, 이것은 정상적 크기의 컴퓨터 키보드를 사용할 경우 그러한 실수가 많이 발생하지 않는다는 것을 의미한다.

소프트웨어의 설계도 특정한 오류를 방지할 수 있다. 그러므로, 키에 기능을 연결하는 것은 면접원들이 범하는 오류의 종류에 대한 신중한 고려와 함께 이루어져야 한다. 예를 들어, 키의 간격은(예 : "예", "아니오" 응답의 경우 각각 [1]과 [2]를 사용하기 보다는 [1]과 [5]를 사용하는 등) 조사응답을 입력할 때 발생하는 몇 가지 실수를 줄일 수 있다. 또, AHEAD에서는 면접을 일시중단할 때 사용하는 [F5] 키가 의도하지 않은 중단의 횟수를 감소시킬 수 있도록 정기적으로 사용되지 않는 기능키들 주위에 놓여 있다. 하지만, 그렇다 하더라도 유해하고, 어려운 것일수록 그것은 반드시 수행되어야 한다는 Norman(1993)의 주장에 따라, 일시중단 기능은 실수의 가능성을 줄이도록 [Shift-F5]로 바뀌었다.

키입력 파일의 분석은 면접자 교육에 관한 후행 연구에 정보를 제공함으로써 오류를 감소시키는데 도움이 된다. AHEAD 도구에서 면접원들이 다소 어렵게 느끼는 것에는 명단 화면에서의 이동과 편집, 응답 변경을 위한 전후 이동, 그리고, ID 입력 수정이 포함되어 있다. 면접원들은 다른 조사를 위한 교육 중에 이러한 키 조작에 대해 더 많은 실습을 할 수 있다.

오류 감소에 대한 두 번째 접근방법은 시기적절한 오류 감지에 있다. 이것은 범위 및 일치성 확인이나, 무효 입력에 대한 면접원 경고, 특정 기능키 사용불가 또는 기타 키 조합(예 : 컴퓨터 재부팅시의 [Ctrl-Alt-Delete]) 등, 이미 많은 CAPI 도구들의 일부가 되어 있다.

오류 복구는 최선의 노력에도 불구하고, 오류는 발생할 수 밖에 없을 것이라는 가정에서 비롯된다. Frese and Altmann(1989)은 오류관리에 역점을 둘 것을 주장하면서 피교육생에게는 단순히 오류를 피하는 방법이 아니라 그러한 오류를 다루는 방법을 가르쳐야 한다고 제안하였다. 오류는 발생된 종류뿐 아니라, 잠재적 영향에 따라서도 구분될 수 있다. 어떤 오류는(특히 비효율성) 자료품질에 미미한 영향을 미치지만, 다른 오류는(감각운동적 오류일지라도) 잠재적인 악영향을(예, 우발적인 파일 삭제) 미칠 수가 있다. 이것은 상황에 따라 각기 다른 오류복구 전략을 의미한다. 중대한 조치에 관한 확인 프롬프트의 적절한 활용과 의도하지 않은 조치에 대해 용이한 취소(undo)기능, 그리고 오류나 이전 입력사항을 수정하는데 대한 교육은 모두 CAPI 조사에서 오류관리를 용이하게 할 수 있는 접근방법들이다.

면접원 비효율성 -특히, 텍스트 입력의 편집- 에 대해 발견된 많은 사례들은 어떠한가? 이러한 편집은 대부분 시간 소모적일 뿐 아니라, 거의 불필요한 것으로 나타나 있다. 여기에 대한 한 가지 접근방법으로, 텍스트를 입력할 때는 인쇄상의 오류는 무시하고, 읽을 수 있고 코드 부여할 수 있는 데만 집중하라고 일러주고 싶다. 이것은 실제로 본 연구를 위한 교육에서 강조했지만, 면접원들은 여전히 텍스트 응답의 편집을 고집하는 것으로 나타났다. 이러한 현상은 텍스트 입력의 가독성과 완전성이 면접원 평가의 핵심요소로 작용하였던 지면 면접방식의 잔존 효과일 지 모른다. 또, 대부분의 기관들이 면접이 끝난 후 면접원들이 도구를 다시 들여다보는 것을 허용하지 않는 상황에서, 이들이 처음에 제대로 입력하려 하는 것은 어쩌면 당연한 일일 수도 있다. 한편, 또 하나의 접근방법은 그러한 편집행위를 인정하고, 면접원들에게 용이하게 편집할 수 있는 도구를 제공하는 것이다. 예를 들면, SurveyCraft는 다양한 편집기능이 있지만, 면접원들은 이것에 관한 정보를 얻지 못했거나 사용법에 관한 교육을 받지 못했다. 하지만, CAPI 사용에 친숙해짐에 따라, 그러한 도구의 활용은 점차 늘어갈 것이다.

마지막으로, 면접원의 수행능력을 평가하기 위한 도구로서의 키입력 파일이라는 쟁점으로 돌아가 보자. 몇몇 연구에서 CAPI의 성공적인 실천과 면접시 랩탑 컴퓨터 사용에 대한 면접원들의 긍정적인 반응에도 불구하고, 우리는 면접원들이 시스템과 도구를 어떻게 (잘) 사용하는지에 대해서는 여전히 모르는 바가 많다. 예를 들어, 우리는 면접원에게 온라인 도움말을 제공하기 위해 많은 자원과 노력을 투입하였지만, 여기의 데이터는 그러한 편리함이 그다지 자주 이용되지는 않는다는 사실을 보여주었다. 키입력 파일을 분석하면 면접원들이 여러가지 CAPI 기능들을 사용하는 범위와 성공적인 활용을 파악할 수 있다.

비록 한계는 있지만, 키입력 파일은 조사관리자들이 각자의 면접원들의 수행능력을 평가하는데 유용할 수 있는 많은 도구들 중 하나로 바라봐야 한다. 또, 이 파일들은 다른 척도들과 함께 CAPI 조사를 수행하는 면접원의 역량과 기술을 평가하는데 이용되어야 한다. 몇몇 평가방법들이 면접과정의 결과에만 집중하는 반면, 키입력 파일은(행동 코드부여와 같은 다른 방법들과 함께) 면접과정 그 자체에 중점을 둔다.

이와 같은 정보를 면접원 평가에 활용할 수 있는 방법에는 몇 가지가 있다. 첫째, 앞서 지적한 바와 같이, 키입력 파일은 연출된 모의면접과 함께 실전면접에 대한 면접원의 준비성을 판단하는 하나의 도구가 된다. 둘째, 조사감독원들은

각각의 면접과정을 검토하는데 감사추적(audit trails) 또는 키입력 파일의 "재생(playback)" 기능을 활용할 수 있다(또, 이미 그렇게 하고 있다). 셋째, 키입력 파일은 통계적 공정관리 원칙과 절차를 이용하여 면접원 행동을 수많은 면접의 경우와 비교할 수 있는 집계 데이터를 산출하는데 이용될 수 있다. 마지막으로, 이 과정은 추가적인 조사와 적절한 조정이 필요한 특이사례나 부적절한 키입력 일치를 감지하는데 이용될 수도 있다.

키입력 파일은 면접원의 수행능력을 평가하는 것 이상이라는 사실은 이미 입증하였다. 키입력 파일은 CATI와 CAPI 도구 설계시, 면접원의 작업을 용이하게 하고, 컴퓨터를 이용한 면접에서 인간-컴퓨터 상호작용의 과정을 이해할 수 있는 정보를 제공하는 자료를 제시해 준다. 이것은 여러가지 목적에서 유용할 뿐만 아니라, 이러한 자료는 적은 비용으로도(자료수집 비용에 견주어) 수집할 수 있다. 따라서, 이러한 자료에 대한 우리의 이해가 커지고, 키입력 파일을 요약 및 분석하는 도구가 개발됨에 따라, 이의 유용성은 훨씬 증대될 것이다.

조사 자료수집의 자동화는 면접원 수행능력의 평가라는 추가적인 과제를 만들었지만, 동시에 이것은 우리에게 그러한 평가를 용이하게 할 수 있는 도구도 함께 제공한다. 키입력 파일은 별도의 추가비용없이 자동화된 환경에서 기계적으로 수집될 수 있는 하나의 자료원이다. 또 다른 예에는 시간확인정보와 상세한 호출기록데이터 등이 있다. 이러한 것들은 모두 면접원들이 컴퓨터 도구와 어떻게 상호작용하는 지에 대한 우리의 이해를 도움으로써 오차를 줄이고, 컴퓨터를 이용한 면접의 효율성을 개선시킬 것이다.

감사의 글

본 절은 미시건 대학과의 협력약정으로 미 농무성 농업조사국에서 일부 지원하고, 조사방법론 프로그램에 대한 센터 내부의 지원을 받은 연구에 기초를 두고 있다. 키입력 자료수집과 모의면접 관리에 도움을 준 AHEAD 스탭과(Willard Rodgers, P.I.)과 현장 실무진, 키입력 자료의 준비와 분석에 도움을 준 Qian Yang; 그리고, 본 프로젝트에 전반적인 지원과 조언을 아끼지 않았던 Bill Connett에 감사의 말씀을 전한다. 아울러, 초안에 대해 식견높은 평을 해 주신 Edith de Leeuw에게도 감사드린다.

제 3 장

조 사 처 리 및 작 업

제 3 장 조사 처리 및 작업

제 8 절 조사후 처리의 몇 가지 측면

Lars Lybe
스웨덴 통계청
Daniel Kasprzyk
미국 국립교육통계연구소

8.1 서문

표본조사의 설계와 개발 및 이행은 수 많은 면면과 성분들로 이루어져 있다. 모든 구성성분들은 설문조사의 최종 결과물에 직접적으로 영향을 미치지만, 이 중 표본추출 및 설문지 설계와 같은 일부 성분들은 통계적이지도 않고 때로는 조사 지향적이지도 않은 양상을 띠면서 연구문헌에서 다른 무엇보다 많이 부각되어 왔다. 일단의 구성성분 즉, 작업(operation)과 프로세스(process)라는 집합은 조사, 표본, 설문이 설계되고, 자료가 수집된 후 이루어지는 과정이다. 이러한 작업들은 - 앞으로 조사후 자료처리작업이라 부르겠지만 - 수집된 자료를 처리하고 사용자가 활용 및 해석할 수 있도록 준비하는 몇 가지 프로세스들로 구성되어 있다. 처리작업들은 조사와 시스템에 따라 각기 다를 수 있지만, 대부분은 다음의 과정들을 포함하고 있다. 편집(editing), 코딩(coding), 데이터 입력(data entry), 대체(imputation), 가중치 부여(weighting), 추정(estimation), 분석(analysis). 이러한 작업들 중 일부는 전체 조사오차에 상당한 영향을 미칠 수 있다.

조사후 처리와 기타 조사 작업들 사이의 관계는 항상 명확한 것은 아니다. 예를 들어, 설문조사에서 컴퓨터 사용의 증가는 자료를 수집할 때 몇 가지 처리작업을 가능하게 한다. 이러한 이유로, 일부 통계전문가들은 조사처리를 분류할 때 자료처리과정에 자료수집이나 그 일부를 포함시키기도 한다. 본 논문에는 추정 및 분석뿐만 아니라, 자료수집과 관련한 품질사안에 대해서도 각 섹션에서 다루고 있다. 그 중, 본 절에서 다루고 있는 조사후 처리작업은 편집과 코딩 및 데이터 입력을 포함한다. 이에 관한 시각들은 지난 10년 동안 변화를 거듭해 왔다. 그래서 일찍이 매우 노동집약적이고, 오차 유발적이던 이러한 작업들은 점점 자동화되어 일단의 편집 및 코딩 규격이 일관되게 적용되는 상태에 이르고 있다. 앞으로 이러한 조사 후 자료처리활동의 전망과 실천은 다음 세 가지 발전방향으로 인해 점차 변화할 것이다.

첫째, 편집 및 코딩과 같은 작업의 자동화는 대규모 사회경제조사에서 일일이 수작업을 하던 많은 임시 노동력의 수요를 감소시켰다. 이와 같은 수작업들은 대개는 매뉴얼에 나와 있는 규격에 대한 약간의 교육과 학습에 따라 진행된다. 그러므로, 수작업 처리의 자동화는 사람보다는 컴퓨터가 보다 일관되고, 동일한 방법으로 편집 및 코딩 규칙을 이행함에 따라 자료의 품질을 향상시킬 가능성이 훨씬 크다 하겠다.

둘째, 컴퓨터를 이용한 자료수집은 몇몇 처리작업들을 통합하고, 이를 자료수집단계에서 수행할 수 있는 기회들을 창출함에 따라, 생산성과 자료품질을 개선하였다. 일반적으로, 컴퓨터를 이용한 자료수집(CASIC; Computer Assisted Survey Information Collection) 환경에서는 일부 편집 및 코딩 기능들이 자료수집 당시에 수행될 수 있으며, 따라서 응답자에게 직접 액세스함에 따라 시간과 비용이 절감되고, 품질이 향상된다. 자료수집에 대한 종래의 접근방식은 편집(editing)과 코딩(coding) 및 입력(keying)이 각각의 규격을 따르는 식으로 각기 다른 단계를 가지고 있다. 하지만, 오늘날의 많은 통계기관들은 이러한 작업들이 통합 및 수정된 다른 시스템들로 종래의 접근방식을 대체하였다.(Bethlehem, 제16장; Keller, 1995; Shanks, 1989 참조) 현재의 사고는 편집 및 코딩과 같은 개개의 성분보다는 하나의 처리시스템을 최적화할 필요성을 강조하고 있다. 이러한 사고는 자료수집 기관들이 자료의 수집, 관리 및 전달에 물리적, 정신적 통합을 요한다는 측면에서 상당한 시각의 변화를 의미한다. 따라서 이러한 맥락에서 여기서는 몇 가지 수집방식의 사용이 늘어가는 조사 후 처리성분들의 통합에 관한 논의를 하고자 한다.

셋째, 코딩 및 입력과 같은 작업에 대한 종래의 품질보증은 미리 지정된 평균발생품질(average outgoing quality)을 보증하는 합격판정 표본추출계획에 주로 의존해 왔다. Dodge and Romig(1994)가 최초 개발한 이 계획은 고용인 피드백을 강조하지 않는다. 따라서, 중요한 품질기능으로 이러한 방법을 사용하게 되면, 다른 견지 즉, 고비용, 비효율적일 수 있고, 고용인의 책임을 경감시킬 수 있다는 점에서 비난 받을 수 있다. 반면 Deming(1986)에서 비롯된 개념과 종합품질관리운동은 검사(inspecting) 보다는 초기단계에서 표준화된 프로세스를 이행하고, 고용인들을 점검하는 품질부서를 두기보다는 팀 내부에서 문제의 근본원인을 파악하려고 노력하며, 안정된 수준에서 품질을 유지하기보다는 지속적인 개선을 목적으로 하는 방법을 사용하는 것과 관련된 접근방식을 제시한다. 따라서, 이후 본 절에서는 이러한 발전방향에 대해 논의해 보도록 하겠다.

이에 따라, 본 절에서는 종래의 작업방법을 현재의 사고와 대조하면서 편집, 코딩, 데이터 입력이라는 조사후 처리활동에 대해 검토한다. 또한, 현재의 방법들과 특정 활동에 관련된 문제점을 설명하고, 마지막에는 이러한 활동들의 통합에 관해 다룰 것이다.

8.2 편집

편집이라는 것에 대해 지금까지 표준화된 정의는 없지만, 미연방 통계전문가 모임(Federal Committee on Statistical Methodology, 1990)에서는 다음과 같은 식의 정의를 내리고 있다.

자료대치 및 요약절차에 앞서 가능한 한 많은(반드시 이의가 있는 모든 자료는 아님) 에러성 데이터를(수작업 및 전자매체를 통해) 수정한다는 목표를 가지고 에러성(erroneous) 및/또는 의문성(questionable) 조사자료(조사응답자료 또는 식별유형자료)를 검출하는데 의도 및 사용되는 절차(들)

이 정의는 조사 수행과정에 들어 있는 몇 가지 단계를 포괄하며, 조사 주기동안 여러 번 나타날 수도 있다. 이것은 또한, (1) 자료품질에 관한 정보를 제공하고, (2) 앞으로의 조사과정 개선에 관한 정보를 제공하며, (3) 자료를 단순하게 "정제(clean up)"하기 위해 Granquist(1984)가 제시한 편집의 목적과 일맥상통한다. 다만, 위 정의는 세 번째 목적 즉, "정제"의 측면을 분명히 강조한다.

넓은 의미로 보자면, 편집은 잘못되었거나 의심이 가는 자료를 확인하기 위한 하나의 도구로 볼 수 있다. 여기에는 제각기 다른 많은 활동이 포함될 수 있다. Pierzchala(1990)가 지적한 바와 같이, 편집은 타당성을 검토하는 절차 내지는 통계적 절차가 될 수 있다. 이것은 기록단계에서나 혹은 어떤 집계단계에서 수행될 수 있으며, 자료를 수집하는 시기에서부터 자료를 요약하는 과정에 걸쳐, 서로 다른 조사단계에서 일어날 수도 있다. 결국, 이와 같이 편집은 수작업으로도 또, 컴퓨터 프로그램을 통해서도 이루어질 수 있는 것이다.

하나의 타당성 검토절차로 볼 때, 편집은 불일치와 의심스러운 값은 찾아내고, 필요할 경우 이를 수정까지 하는 일단의 방법들이다. 이러한 유형의 편집은 개개의 기록이나 개개의 설문단위에서 일어난다. 또 다른 한편으로, 하나의 통계적 절차로 볼 때, 편집은 응답자료의 통계적 분석에 기초를 둔다. 따라서, 이러한 유형의 편집은 일변량 또는 다변량 분포에서 특이점을 탐지할 목적으로 기록간 확인을 하거나, 과거자료와의 비교를 바탕으로 집계된 자료를 검토하거나,

혹은 과거자료를 바탕으로 각각의 표본단위에 맞는 편집범위를 지정하는 작업으로 특징지을 수 있다. 그러나, 어느 경우든 간에 에러성 데이터를 확인하기 위해 지정되는 편집규칙(edit rule)이나 점검(checks)은 자료와 개념 및 그러한 자료 내부의 관계에 정통한 전문가(subject matter expert)와 함께 개발되어야 한다.

타당성 검토절차는 일반적으로 다음 절차 중 하나 이상을 따랐던 종래의 미시적 편집(micro-editing) 접근방식을 말한다. (1) 사무직원은 누락이나 오류 및 잘못 이해한 지침에 피드백을 제공하려고 현지 면접원들이 수집한 자료의 품질을 점검하기 위해 설문지 표본을 검토한다. (2) 사무직원은 모든 설문지의 식별가능성과 일치성을 확인하기 위해 지정된 핵심 문항을 검토한다. (3) 컴퓨터 프로그램은 준비된 확인과정으로 의심스럽거나 모순된 데이터 및 무효 또는 누락된 입력사항을 일괄처리한다. 데이터의 수정여부를 결정하기 위해서는 대개 응답자와 연락하거나 기타 다른 정보를 수집하는 등의 조정절차가 필요하다. 반면, 거시편집(macro-editing) 점검은 데이터의 집계에 사용되며, 따라서 의심스러운 기록들의 하위집합을 확인할 수 있다. 이것의 목적은 개개의 기록에서 체계상의 불일치와 오류를 추적하는 것이다.

8.2.1 편집방법의 예

편집은 조사과정의 여러 단계에서 수동으로나 자동으로 수행될 수 있다. 일부 수동편집은 자료를 포획하기 전 자료준비의 일환으로 진행되는 것이 보통이다. 수동편집은 대개 면접원 설문지나 우편 설문지 표본에 대한 품질검토에서부터 이후의 식별가능성 점검과 핵심 문항에 대해 누락된 데이터 점검 및 소수점 이하 한 자리수 제한 등과 같은 특정한 응답기준에 관한 점검으로 이루어진다.

자동편집은 데이터가 입력된 후에 수행되거나 혹은 자료수집과 함께 이루어진다. 이러한 편집규격은 항상은 조사에 따라 구별되고, 완전성 확인과 변수범위 확인 및 생략 지침 확인에 관한 절차들을 포함한다. 일괄편집의 경우에는 모순되고 실용성이 없으며, 의심가는 경우에 관한 기록이 확인된다. 이 절차의 결과는 보통 잠재적 수정의 기초로 제공되는 에러 메시지 파일로 나타난다. 상호작용적 편집(interactive editing)의 경우에는 그 결과가 화면상에 보여지고, 편집자는 이러한 에러 메시지를 수정할 것인지 무시할 것인지를 결정하게 된다. 오늘날의 CASIC 기술은 조정이 필요할 때면 응답자를 호출하면서 면접 당시에 일부 편집이 진행되도록 해 준다. 이러한 미시편집 범위내에 속하는 그 밖의 일반적인 확인기능에는 "만일, 그렇다면"과 같이 미리

지정된 문구가 들어 있고, 각 문항의 합계가 보고된 총계와 같은 식의 변량의 허용범위가 포함된다.

다중기록(multirecord) 단계에서는 통계적 편집과 거시편집이 사용된다. 통계적 편집은 분포를 활용하여 정상치(지정된 범위내에 속하고, 의심스러울 정도로 일관된 데이터) 뿐만 아니라, 특이치를 파악한다. 거시편집은 기록집계 단계에서 수행된다. 만일, 일관되지 않고 의심스러운 사례가 발견되면, 이러한 사례는 집계에 포함된 개개의 기록단위로 추적이 이루어진다. 이 기법은 문항수가 적은 표본추출에 사용되는 기법들과 유사하다. Granquist(1994)와 Houston and Bruce(1993)는 거시적 다중편집(macro-multiediting)의 방식에 대해 다음과 같은 견해를 제시하고 있다.

1. Aggregate 방식 - 몇몇 핵심적인 집계가 검토 및 확인되고, 의심이 가는 집계에 신호를 보내, 집계된 개개의 기록이 검토되도록 하는 방식
2. Hidiroglou-Berthelot 방식 - 통계적 분석을 통해 합격기준을 도출하고, 보고된 데이터로부터 자동 계산되도록 하는 방식
3. Top-down 방식 - 입력값에 가중치가 부여된 편집값들이 목록상에 순서대로 분류되도록 하고, 추정값에 아무런 영향이 없을 때까지 이 목록의 위 또는 아래에서부터 수동으로 검토하는 과정을 계속하는 방식.

이러한 방식들을 수정한 것으로는 box-plot 방식과 gred(상호작용적인 그래픽 편집)가 있다. 후자는 특이점 검출과 같은 문제에 대해 탐색적 자료분석 개념과 그래픽을 이용한다.(예, Bienias et al., 1995 참조) 일반적인 개념은 최악의 데이터지점을 검토 및 편집하는 것으로, 편집으로 추정값에서 더 이상 의미있는 차이가 발견되지 않을 때 중단된다.

이와 같이, 편집의 첫번째 측면은 잠재적 오류(error)를 확인하는 것이다. 그리고, 두번째가 수정이나 조정단계다. 조정은 어떤 경우에는 응답자와 상의한 후 이루어질 수 있지만, 또 다른 경우에는 문항을 공란으로 처리하거나 대체 즉, 의심이 가거나 명백한 오류이거나 누락된 값들을 보다 합당한 값으로 대체할 수도 있다. 대체(imputation)에 관한 문헌은 아주 많다. 여기에 관한 사례는 Little and Rubin(1987), Kalton and Kasprzyk(1982, 1986), Madow et al.(1983),

Pierzchala(1990), Ferguson(1994) 등을 참조할 수 있다. 이러한 절차들은 목격횟수와 누락된 데이터의 양으로 인해 대개는 자동화되어 있으며, 연역(다른 기록값이나 과거 기록으로부터 값을 결정), 모형(함수값이 누락된 값을 결정), 기준(기준 값이 대체에 활용), 그리고 인간활동의 시뮬레이션(전문자료 편집시스템)에 의존한다.

이 중 마지막 유형의 절차에 대해서는 최근 비교적 새로운 패러다임이 시도되었다. 이 새로운 패러다임, 인공신경망(Artificial Neural Networks)은 인간의 두뇌를 시발점으로 이용한다. 이 접근방식에 따르면, 편집과 같은 작업은 네트워크 속에 신경세포(neuron)들을 연결함으로써 해결될 수 있다. 즉, 하나의 신경세포 모형이 주어지면, 서로 다른 신경세포망에 각기 다른 가중치를 부여함으로써 네트워크 지식을 표현하는 것이 가능하다. 실제로는, 이러한 편집모형은 전문편집 전후에 여러 쌍의 기록 예들을 관찰하는 식으로 학습하는 모형이 되어야 한다. 이러한 개념들은 Roddick(1993)와 Nordbotten(1995)이 논의하였다.

8.2.2 편집과 관련된 몇 가지 문제

편집은 비용이 많이 들고, 시간 소모적인 작업일 수 있다. 실제로 일부 조사를 보면, 편집이 전체 조사예산의 20~40%를 소비한다.(Federal Committee on Statistical Methodology, 1990; Granquist and Kovar(제18장)) 게다가, 광범위한 편집이 조사처리와 자료 출시를 지연시킨다면, 사용자 단체에게도 추가적인 비용이 발생하게 된다. 여기에 관해 잘 알려진 한 예로는 광범위한 편집으로 1년 가까이 결과물의 출간이 지연되었던 World Fertility Survey가 있다. 이처럼, 편집절차를 설계할 때는 전체 조사오차에 초점을 맞추는 것이 매우 중요하다. 만일 예산에서 편집이 차지하는 비중이 20~40%이고, 자료수집을 수행하는데도 고정비용이 발생된다면, 다른 오차원인을 처리하는데 드는 재정이 제한될 것이라는 것은 쉽게 알 수 있을 것이다.

이러한 사안은 많은 편집작업이 자료품질을 개선하거나 보다 나은 추정값에 기여하는 것은 아니라는 사실을 보여주는 연구결과에 비춰 볼 때는 훨씬 문제가 심각하다. 경제조사분야에서 주로 발견되는 이에 관한 예는 Bethlehem(1987), Boucher(1991), Linacre and Trewin(1989), Corby (1984) 등이 잘 설명하고 있다. 또 다른 평가 연구들은 파레토 원칙(Pareto principle)이 작용하고 있음을 보여준다. 즉, 대부분의 실제값이 편집의 결과로서 바뀌는데는 오류(error)의 책임은 없다.(예, Greenberg and Petkunas, 1986 참조) 이러한 결과들은

경제조사에서 편집기능은 보다 생산적이어야 함을 제시해 준다. 적중률(hit-rates) 즉, 값을 변화시키는 에러 메시지의 비율에 관한 데이터는 20~30% 정도로 낮은 적중률은 미시편집에서는 보기 드문 일이 아니라는 사실을 보여준다(Linacre and Trewin, 1989; Lindström, 1991). 학군, 학교, 교장 및 교사에 대한 최근의 조사를 보면, 예비파일과 사전편집파일(불일치, 무효입력, 중대한 누락데이터를 파악하기 위한 목적을 가진 파일) 사이에 한번 이상 응답이 변경된 비율은 공립학교 교장의 경우 1.6%에서부터 인디언 문제담당국(Bureau of Indian Affairs) 학교들의 경우 34%에 이르기까지 다양한 범위에 걸쳐있다.(Gruber et al, 1994)

실제로 많은 조사와 조사기관에서 상당한 과도편집(overediting)이 이루어지고 있음에 대한 걱정의 목소리가 높았다. 편집은 필수불가결한 것이긴 하지만, 보다 효율적으로 축소 시행되어야 한다. Granquist and Kovar(제18장)가 지적한 바와 같이, 편집의 남용은 조사품질을 조절하는 다른 노력들에 비해 비교적 간단하다는 사실에 일부 기인할 수 있다. 이러한 주장은 기업 조사과정에서 주로 제기되어 왔으며, 위에서 언급한 유형에 관한 연구들은 인구통계조사 문헌에서는 좀처럼 찾기 어려웠다. 때문에 인구통계조사 실시의 몇 가지 측면은 분명히 여러가지 편집유형에 초점을 맞추고 있음을 인식함에도 불구하고, 과도편집이 인구통계조사에서 의미있고, 비용이 드는 문제라는 사실을 확신을 가지고 말할 수는 없는 입장이다. 그러나 조사자들은 편집에서 다른 조사기능들로 자원을 이동시키면서 전체 조사오차에 보다 분명한 초점을 맞춰야 한다는 사실만은 틀림없음을 확실히 밝혀두고자 한다. 편집기능을 다른 조사작업에 통합시키는 것에 대해서는 앞으로 추가적인 조사가 진행될 것이다. 예를 들면, 컴퓨터를 이용한 자료수집에서는 면접, 코딩, 데이터 입력 및 편집이 모두 함께 이루어질 수 있다. 또한 거시편집 기법은 광범위한 예비편집을 줄이고, 많은 편집기능을 분석단계로 전이시킬 수 있다.

8.3 코딩

코딩은 대개 개방형 질문의 응답형식으로 된 원시데이터가 추정(estimation), 제표(tabulation), 분석(analysis)과 같은 조사의 최종단계에 사용될 수 있는 형식으로 분류 및 변환되는 과정을 말한다. 대부분의 조사에서 코딩 작업은 조사응답을 코딩하도록 교육받은 개개인의 기술과 지식 및 판단에 높이 의존한다. 코딩 작업은 광범위하고, 오차유발적이며, 비용이 많이 들고 시간 소모적이다. 코딩되는 변량의 예로는 직업, 산업, 교육, 사회경제적 상황 등이 있다.

일반적인 코딩 과정은 두 가지 측면을 가진다. 첫째, 조사원은 코딩에 사용되는 범주(category)를 반드시 정해야 한다. 둘째, 비구조적인 구술응답은 각각의 코드번호를 가진 미리 지정된 범주속에 코딩되어야 한다. 이러한 코드번호의 키(key)는 분류표(nomenclature) 또는 사전(dictionary)이라 부른다. 키는 보통 구술적 설명을 코드번호와 연관짓는 일단의 코딩 지침이 따른다. 따라서, 코딩이 사무적인 일인 한, 그 과정은 매우 복잡하고, 관리하기 힘들 수 있다. 오늘날 대부분의 조사기관들은 자동화된 코딩을 기반으로 조사를 수행하거나 혹은 몇 가지 자동화되었거나 컴퓨터를 이용하는 유형의 코딩을 사용한다.

8.3.1 분류표

분류표 또는 코드목록은 설문문항에 가능한 응답을 모두 나열할 목적을 가진 숫자화된 범주들의 목록이다. 분류표는 실제로 어떤 의미에서는 유사한 목적이나 응답을 그룹화할 목적의 분류시스템이다. 이러한 시스템들은 동물학과 천문학에서 종이나 별을 분류하기 위해 채택하는 등, 다른 학문분야에는 이미 존재하고 있는 것이다. 통계분야에서의 분류표는 통계자료의 수집, 제표, 표현을 개선하기 위해 개발되었다. 이러한 통계적 분류표는 통계자료를 이용한 사회현상의 분석 및 비교를 향상시키기도 한다. 대부분의 분류표는 계층적으로 구성되어 있으며, 여기서 코드목록은 세분화를 의미하는 몇 가지 단계로 이루어져 있다.

산업 및 직업과 같은 핵심적인 사회개념에 관한 분류표는 데이터 시리즈의 통일성(uniformity)과 비교가능성(comparability)을 촉진시킬 수 있어 강력한 정렬도구가 된다. 하지만 사회가 변화함에 따라 이러한 분류표도 반드시 개정되어야 한다. 예를 들면, 직업은 시작과 끝이 있고, 없어지는 업종이 있는가 하면, 새로운 형태의 산업이 생겨나고, 시간에 따라 생산공정과 개념이 변모하며, 통계시스템이 통합되는 등 분류표를 구성할 때는 반드시 조화(harmonization)와 연속성(continuity) 및 동질성(homogeneity)에 관한 몇 가지 기준을 고려해야 한다. 분류표의 구성과 수정은 지루하고, 힘든 작업이며, 최종 결과물은 서로 다른 이해집단 사이의 논의와 협상의 결과처럼, 항상 이론과 실제 사이에서 타협을 찾는다.

이 과정에 있어 심각한 문제는 분류표를 활용하는 통계전문가와 이 분야에 정통한 사람 및 코딩 전문가 사이에 커뮤니케이션이 없다는 점이다. 코딩에서의 일부 오류는 분류표의 이론적 측면이 실제와 일치하지 않는다는 데서 발생한다.

예를 들어, 직업에 관한 분류표는 전화회사에서 일하는 교환원과 다른 업종에서 일하는 교환원 사이를 구분할 수 있다. 그러나 응답자와 조사원들에게 이를 확실히 구분할 수 있는 조사질문을 요청하기란 매우 어려운 일이다. 오차 연구들을 보면, 코딩 오차들이 범주간에 균일하게 분포되지 않음을 알 수 있다. 여기서도, 파레토 원칙이 작용한다. 응답과 범주 사이의 몇 가지 연관성들은 다른 것보다 훨씬 분명해 범주당 오차 또는 범주당 취약성이 비대칭 분포로 나타난다. 이러한 연구에 대한 설명은 Lyberg(1981)와 Hansen et al.(1964)을 참조한다.

따라서 범주들은 범주에 대한 설명간에 구별이 어렵다는 사실에 기인한 오차를 피하기 위해 때때로 조합된다. 이론과 실제간에는 보다 많은 협조가 필요하다. 이러한 방향성에 대한 한 가지 노력은 분류과정에서 일관성을 향상시키는 도구로서 전문시스템에 대해 설명하고 있는 Conrad(제17장)에 의해 논의되어 있다.

8.3.2 코딩의 종류

코딩 시스템은 (1) 철저한 수작업 코딩에서부터 (2) 컴퓨터 코딩 즉, 컴퓨터를 이동한 코딩 시스템(CAC)과 혼합된 수작업 코딩 및 (3) 자동화 코딩에 이르기까지 세 종류가 있다. 수작업 코딩은 교육을 거친 코딩요원(coder)나 면접원 또는 응답자에 의해 진행된다. 이 때는 코딩되어야 하는 변량이 복잡할수록 특별한 교육을 받은 전문가의 필요성이 커진다. 수작업 코딩은 말 그대로, 종이와 연필 그리고, 매뉴얼을 가진 코딩요원에 철저히 의존해 프로세스를 완료한다. 반면, CAC에서는 코딩요원이 컴퓨터와 상호작용하면서 코드를 부여한다. 컴퓨터를 이용한 코딩은 도움말 화면이나, 의사결정표, 부가정보 등과 같이 컴퓨터가 제공해 주는 몇 가지 편리한 기능에 코딩요원이 액세스할 수 있다는 사실만 빼면 수작업 코딩과 비슷하다.

자동화 코딩에서는 컴퓨터 프로그램이 가능한 한도에서 코드를 부여하고, 나머지를 CAC나 수작업 코딩으로 핸들링한다. 지금의 많은 통계기관들은 다양한 차원에서 이 세 가지 코딩 방식을 모두 활용한다. 자동화 코딩을 위한 컴퓨터 프로그램은 조사응답의 비교적 많은 부분을 다룰 수 있다. 그 비율은 보통 70~80%까지 이른다(개괄적인 사항은 Lyberg and Dean(1992) 참조). 잘 설계된 프로그램은 저비용으로 낮은 오차율의 코드집합을 생성한다.

자동화 코딩시스템의 기본적인 특징은 다음과 같다. (1) 수작업 코딩에서 사용되는 분류표 대신 컴퓨터에 저장된 사전을 구축한다. (2) 다소 자유로운 형식의 입력방식으로 조사응답을 컴퓨터에 입력한다. (3) 응답을 사전과 매칭시켜 코드를 부여한다. (4) 코드가 지정되지 않은 사례는 수작업 코딩이나 CAC로 전달한다. (5) 코딩 품질을 평가하고, 그 결과 사전을 업데이트 하는 등의 작업을

진행한다. 자동화된 코딩시스템은 전세계 유수한 기관에 설치되어 현재 다양하게 응용되고 있다. 미인구조사국은 AIOCS(Automated Industry and Occupation Coding System)라는 전문가 시스템을 이용한다. 이 시스템은 정보력이 있는 용어와 없는 용어, 동의어, 오자 및 약어를 파악해 수작업 코딩을 시뮬레이션한다. 이 때 정확하게 매칭되면 코드지정이 진행되고, 매칭이 부정확할 때는 개연성있는 가중치를 사용하여 코드번호를 부여한다. 이 시스템에 대한 설명은 Chen et al.(1993)과 Gillman et al.(1993)에 잘 나타나 있다. 오차율에 관한 검토는 Lyberg and Dean(1992)을 참조한다.

스웨덴 통계청(Statistics Sweden)은 이것을 응용할 때 단순한 매칭 알고리즘을 사용한다. 사전 초기값들은 경험적 응답패턴과 주관적으로 선택된 설명의 혼합을 기반으로 한다. 주관적으로 선택된 설명은 분류표와 매뉴얼에 나와 있는 설명에서 직접 가져온 것이다. 수많은 연구들을 보면, 응답자들은 그러한 설명을 사용하려 하지 않고, 오히려 공식적인 목록에서 확인되는 것들을 변형한 다른 설명을 사용한다는 사실이 밝혀져 있다. 따라서, 컴퓨터에 저장되어 있는 사전은 응답자 자신들이 생성한 경험적 패턴을 포함하고 있다는 사실이 매우 중요하다 하겠다. 스웨덴 통계청은 응답자들의 속어 사용이나 일상적인 오자 등에 기인한 변형을 포함하여 신중하게 이를 처리하였다. 사전의 "경험적인" 부분은 공식적인 설명목록을 구성하는 부분보다 훨씬 효율적이다. 한편, 이 시스템은 CAC와 병용되기도 하는데, 그 결과는 매우 성공적이었다(Lyberg and Dean, 1992). 예를 들면, 1978년의 스웨덴 가계지출조사에서는 전체 코딩율이 65%였고 코딩 오차율은 1%였다. 그리고, 1985년 가계지출조사에서는 82%의 자동화 코딩율과 5%를 상회하는 비용절감을 이루었다. 이와 비슷하게, 1980년의 직업에 관한 코딩과 1985년의 스웨덴 총인구조사는 약 72%의 자동화 코딩율을 경험하였고, 수작업 코딩과 비교할 때 약 10%의 비용절감을 이루어 냈다.

캐나다 통계청(Statistics Canada)은 ACTR(Automated Coding by Text Recognition)이라는 시스템을 개발하였다. 이것은 영어와 불어의 2개 언어로 자동화된 코딩과 CAC를 모두 수행할 수 있는 일반화된 시스템이다. 이 시스템은 용어표준화 기법을 사용하여 입력된 텍스트 파일을 기준이 되는 구문파일과 매치시키고, 결과는 텍스트에 대한 코드로 출력된다. 사용자는 구문이나 텍스트의 기준파일과 관련된 코드를 제공해야 한다. 이 시스템에 관한 설명은 Wenzowski(1988)와 Ciok(1993)에 나와 있다. Miller(1994)는 ACTR의 개발 및 응용과 관한 간단한 역사에 대해 설명했는데, 메인프레임 버전인 Version 1은 1991년 캐나다 인구센서스와 1991년 캐나다 농업센서스, 교통조사 및 여행조사에 성공적으로 응용되었다. 이 모든 응용에서는 비용절감과 개선된 품질 및 코딩의 일관성이 발견되었다.(Miller, 1994)

호주 통계국(Australian Bureau of Statistics)은 ASCO(Australian Standard Classification of Occupations) CAC 시스템을 개발하였다. 이 시스템은 매우 효율적이며, 비교적 숙련되지 않은 코딩요원라 하더라도 수작업 코딩 전문가와 같은 수준으로 작업할 수 있는 많은 특성들이 있기 때문에, 완전 자동화 시스템보다 어쩌면 우수할 것이다. 이 시스템은 사용자 친화적인 인터페이스를 가지고 있고, 잠재 매칭율과 온라인 도움말 화면, 간략한 자료입력 설명, 그리고 빠른 검색 및 매칭절차를 제공한다. 1986년 호주 센서스에 이를 응용한 결과, 수작업 접근방식일 때 300만 호주달러였던 비용이 약 160만 호주달러로 절감되었으며, 전문가 수작업 코딩과의 일치성은 약 95%에 이르렀다(Lyberg and Dean, 1992). 시스템에 관한 자세한 설명은 Embury(1988 a, b)에 나와 있다.

프랑스 국립통계청 INSEE는 두 가지 시스템을 개발하였다. CAC 시스템인 COLIBRI와 완전 자동화시스템인 QUID(Lorigny, 1988). COLIBRI는 이름과 주소로 사업자 등기를 검색한다. 화면에는 후보명단이 순위별로 나타난다. 경제활동코드는 약 70%의 케이스에 지정되며, 경제활동의 전체 오차율은 대략 4%이다. QUID는 전문가에 의해 코딩된 경험적 응답에 대한 방대한 사전을 바탕으로 하는 CAC 시스템이다. 사회경제적 현황을 QUID를 응용해 코딩할 때는 기록의 약 90%가 5~10%의 오차율로 코딩되었다. CAC를 사용하는 다른 기관에는 뉴질랜드 통계청, 네덜란드 통계청, 영국 통계청 등이 있다.

8.3.3 코딩의 몇 가지 문제

자동화 코딩과 컴퓨터를 이용한 코딩(CAC)과 관련한 개발노력과 좋은 실무적 응용사례에도 불구하고, 여전히 문제점은 존재한다. 첫째, 수작업 코딩은 언제나 필요할 것이다. 자동화 코딩이 코딩할 수 없는 나머지 텍스트와 구문에 대한 코딩은 수작업이 필요할 것이다. 최고품질의 자료를 위해서는 수작업 코딩요원들은 응답을 정확하고, 일관되며, 통일성 있게 분류할 수 있는 코딩 전문가라야 한다. 그러나, 코딩이 점차 자동화됨에 따라, 경험많은 일단의 코딩요원들 - 일반적 응용에 충분한 규모의 - 을 한 기관에서 유지하기란 점점 어려워질 것이다. 둘째, 자동화 및 CAC 시스템은 지금도 개발 및 시험중이다. 통일되고 일관된 코딩과 잠재적 비용절감에 대한 큰 기대가 있는 반면, 기술은 획일적으로 채택되어 오지 않았다.

수작업 코딩에 있어서는 앞으로 더욱더 품질에 대한 사안이 반드시 강조되어야 한다. 종래의 코딩관리는 종속 또는 독립된 성격의 검증계획을 바탕으로 한다. 코딩 관리자가 지정된 코드를 검사하고, 수정여부를 결정하는

종속검증(dependent verification)이 매우 비효율적일 지라도, 많은 기관에서는 여전히 이를 활용하고 있는 실정이다. 이에 대한 대안으로, 둘 이상의 코딩요원이 독립적으로 코드를 지정하고, 다수의 코드를 채택하는 독립검증(independent verification)은 훨씬 효율적이기는 하나, 자주 사용되지 않는다. 편집에서와 마찬가지로, 이것은 독립검증과 비교해 볼 때, 종속검증을 관리하는 것이 비교적 간단하기 때문일 것이다.(Lyberg, 1981 참조)

종속이든 독립이든, 코딩 검증은 대개 전통적인 합격 표본추출을 통해 이루어지고, 여기서는 미리 지정된 평균발생품질을 보장하는 검사계획이 집행된다. 이러한 계획은 코딩요원에게 피드백을 강조하지 못하며 따라서, Deming(1986)을 비롯한 많은 사람들의 비판을 받았다. 이러한 비판은 수작업 코딩의 품질을 개선함과 동시에 비용을 절감하기 위한 미국 RTI(Research Triangle Institute)의 노력에 대한 Biemer and Caspar (1994)의 논의에서 찾아볼 수 있다.

비판은 다음을 기초로 한다.

1. 검사는 광범위하다. 종합품질관리원칙을 적용하고, 검사계획에 의존하기 보다는 최초에 정확한 코딩을 하는 것이 보다 바람직하다.
2. 프로세스에 품질을 적용시키려고 검사를 이용하는 것은 불가능하다. 품질은 반드시 시작단계에서부터 프로세스의 일부분이라야 한다. 표본추출 계획은 발생품질에 관하여 특정한 안정성만을 유지할 뿐이다. 이것은 품질을 개선시킬 수 없다.
3. 검사는 불완전하다. 첫번째와 두번째 오차는 표본추출 계획의 조정에 대한 필요성을 야기시킨다(Minton, 1972).
4. 검사집단에서 수행하는 검사는 코딩요원의 책임을 배제시킨다.
5. 경험적 연구들은 코딩요원이 코딩 문제의 비교적 작은 부분에 책임이 있음을 밝히고 있다. 시스템은 많은 책임이 있다. 그러므로, 코딩요원 개개인의 행동을 조정하려는 노력은 반드시 유용하거나 실용적인 전략이라고 볼 수 없다.
6. 코딩요원 피드백은 통계적 관리가 검사부서에서 이루어지는 환경에서는 도움이 되지 않는다. 피드백을 한다 해도, 이미 너무 늦다.

Biemer and Caspar(1994)는 코딩 성과를 보다 나은 성과와 비교하고, 일치되지 않는 결과를 유형별로 분류하여 파레토 분석을 적용하며, 코딩 팀과 함께 일치되지 않는 결과의 원인을 파악한 후, 마지막으로 이를 개선하기 위해 제시된 척도를 실행하는 순환적 프로세스에 바탕으로 둔, 코딩 개선의 새로운

패러다임을 제안한다. 1991년 RTI에서 조사한 산업 오차율과 직업 오차율은 각각 17%와 21%였다. 하지만, 이 새로운 프로세스를 도입한 후에는 각각 4%와 5%로 현격한 감소를 보였고, 동시에 비용은 절감되고 생산성은 증대되었다.

자동화 코딩에서의 품질 사안도 물론 보다 많은 강조가 필요하다. 자동화 코딩에 관한 대부분의 연구는 코딩율을 높이는데 집중해 왔다. 이러한 자원할당은 이해할만하다. 그러나 인정할 만큼의 높은 코딩율을 달성할 수 없다면, 품질에 대한 노력의 필요성도 없을 것이다. 물론 사전과 함께 정확한 매칭을 바탕으로 코드가 지정되는 경우만을 본다면, 품질은 그 정의에 따라 거의 완벽할 것이다. 문제는 부정확한 매칭에 바탕을 둔 코딩에서 발생한다는 것이다.

보다 면밀히 연구해야 될 필요성이 있는 한 가지 사안은 참값 코딩의 개념이다. 기본적인 가정은 모든 요소들이 오로지 단 하나의 범주에만 속한다는 것이다. 혹자는 하나의 요소가 해석에 따라서는 한 가지 이상의 범주에 할당될 수도 있다고 말할 것이다. 그러나 이것은 해당사항이 없다. 코딩은 하나의 요소에 한 가지 범주 외에는 허용하지 않는 일대일 맵핑이다. 다만, 실제로는 애매한 설명으로 이러한 맵핑을 수행하기가 매우 어려운 상황들도 있다. 그렇지만 정말 좋은 분류계획이라면, 애매한 경우에 적용할 수 있는 특별한 범주를 도입하여 이러한 문제를 해결할 수 있어야 한다. 요소의 코딩이 때로는 일대다 맵핑과정의 일부분으로 생각되어지는 이유는 분류표에 결함이 있는 즉, 분류표에 애매한 표현에 관한 범주가 들어있지 않다는 것이다. 실제로 참값 코딩은 예를 들면, 전문가들 사이의 다수 코드와 같이 종종 조작되어 정의되어야 한다.

보다 면밀한 관측이 필요한 다른 사안으로는 정확한 매칭, 부정확한 매칭, CAC 그리고, 순수 매뉴얼로 코딩된 사례들 간에 품질상의 차이, 자료분석에 미치는 코딩오차의 효과, 사전의 크기 및 내용과 관련된 품질 및 비용문제, "최선의" 전략을 결정하는 프로그램 영역이 최저의 비용으로 최고의 데이터를 산출하는 접근방식의 조합을 구현할 수 있도록 도와주는 보다 유기적인 의사결정 규칙.

8.4 자료포획

자료포획은 설문지에 기록된 정보가 컴퓨터에서 해석할 수 있는 형식으로 변환되는 조사단계를 말한다. 자료(데이터)는 지면 설문지에 있는 응답을 직접 입력하거나, 문자인식 및 문자판독기와 심지어는 touch-tone 및 음성인식을 통해 포획될 수 있다. 최근에는 새로운 형태의 자료포획기술이 개발되었는데, 여기에는 팩스전송, 전자문서교환(EDI), 인터넷을 통한 이메일 전송 등이 포함된다.

자료를 입력하는 것은 지루하고, 노동집약적인 작업이다. 많은 기관들은 컴퓨터를 이용한 전화면접(CATI)이나 컴퓨터를 이용한 대인면접(CAPI) 및 광학문자판독기(OCR)와 같이 보다 현대적인 기술에 의존하면서 이러한 작업을 피하려고 노력한다. 입력(keying)은 오차유발성이 매우 높은 작업은 아니다. 예를 들어, Fourth Followp Survey of the National Longitudinal Study에서 오차수준은 품질관리표본을 기초로 할 때 약 1.6%였고(Henderson and Allen, 1981), 1988년 미국의 Survey of Income and Program Participation에서의 오차수준은 이와 비슷한 0.1% 내외였다.(Jabine et al., 1990) 이러한 오차율은 각 변량에 들어 있는 오차와 관련되어 있다.

1990년 미국 센서스에서 긴 형식의 설문을 입력할 때의 품질을 결정하기 위한 한 연구에서, 입력자들은 입력 초기단계에 각 란의 0.62%에서 키입력 착오(또는 누락)를 범했다.(미인구조사국, 1993) 컴퓨터 기술은 컴퓨터를 이용한 대인면접을 수행하는 추측컨대, 자료입력에 능숙하지 못한 면접원들이 범하는 오차횟수에 관한 새로운 관심을 불러 일으켰다. 그러나 Dielman and Couper(1995)는 컴퓨터를 이용한 대인면접에서의 키입력 오차에 초점을 맞춘 연구에서 키입력 오차율은 0.095%에 불과하다고 제시하면서 키입력 오차는 의미를 둘만한 문제가 아니라는 증거를 보였다. 또, 스웨덴의 한 연구에서는 오차율을 기록에 대한 백분율로 표현하였는데, 여기서 기록당 평균오차율은 1.2%로 나타났다.(Lyberg et al., 1977 참조)

미인구조사국은 적절하게 설계된 설문지의 마이크로필름 사본으로부터 "정보"를 판독하고, 이 정보를 처리할 수 있도록 테이프에 담을 수 있는 장치를 이용하는 프로세스인 FOSDIC(Film Optical Sensing Device for Input to Computers)을 개발하였다. FOSDIC은 응답과 연관된 원이 표시되어 있는 지를 파악하기 위해 광선을 이용한다.(Brooks and Bailar, 1978) 이것을 이용한 결과, 평균오차율은 약 0.02%에 불과하였다.(Brooks and Bailar, 1978) 이와 비슷한 기술로, 미국의 국립교육통계센터(NCES; National Center for Education Statistics)에서는 High School and Beyond Study에서 설문지를 광학스캔하는 방법을 이용하였다.(Jones et al., 1986) 이러한 프로세스들은 다수선택이 가능한 질문에 매우 효과적이다. 그러나, 업무로드와 인건비가 증가함에 따라, 손으로 쓰여진 방대한 응답을 정확하고 비용 효과적으로 수집해야 될 필요성이 있다.

광학문자판독기(OCR; Optical Character Recognition)는 손으로 쓴 응답의 자료수집문제에 대한 하나의 해결책으로 제시되었다. 미 인구조사국은 광학문자판독산업의 현재 기술력을 제고하고, 새로운 연구를 자극하기 위한

목적을 가진 몇몇 회의를 후원하였다. 이 회의들의 일반적인 결론을 보면, OCR의 실현성은 어떻게 응용하느냐에 달려 있다는 것이다. 몇몇 응용요건은 다른 것에 비해 상당히 낮은 복잡성을 요구한다. 또, 한 가지 회의요약문에는 수기록을 위한 광학문자판독시스템의 정확성은 지난 몇 년간 엄청난 발전을 거듭했으며, "단어와 구문을 판독하는 장치의 성능은 이제 결과물의 정확성을 떨어뜨리지 않으면서, 센서스를 수행하는데 필요한 비용과 시간을 절감하기에 충분한 정도에 이른 것으로 보인다"고 되어 있다(Geist et al., 1994). 미인구조사국은 1995년 Census Test의 일환으로 OCR 기술의 비용, 정확성 및 속도에 관한 연구계획을 세워 놓고 있다.(U.S. Bureau of Census, 1994)

OCR 오차는 두 가지 유형이다. 치환(substitution)과 거부(rejection)이다. 거부는 반드시 수정되어야 하고, 따라서 다루기에 광범위하지만 또, 적절히 수정된 다음에는 오차가 없다. 치환은 물론 오차를 수반하지만, 대개는 경미한 오차다. 수기의 코드를 스캔한 1970 스웨덴 센서스와 관련한 한 연구에서는 어림잡아 0.14%의 숫자가 치환되었다. 이것은 언뜻 보기엔 아주 훌륭한 결과인 것 같지만, 앞서 말한 실현성에 있어서는 응용에 달려있다. 이 경우에는 숫자를 쓰는 과정이 고도로 표준화된 것이었다. 숫자는 OCR를 위한 숫자 쓰는 법에 대해 특별교육을 받은 코딩요원들이 작성하였다. 분명, 이것은 말하자면 일반적인 모집단이 OCR 처리를 위해 소득 등에 관한 데이터를 써 내놓도록 요청받는 것과는 거리가 먼 상황이다.(Lyberg et al., 1977 참조)

최근에는 스웨덴 통계청과 일부 타기관들이 스캐너에서 찍힌 전체 형식을 스캐닝하는데 대한 실험을 시작하였다. 이 정보는 컴퓨터를 통해 해석된다. 일반적인 의미에서의 데이터 입력은 더 이상 필요하지 않으며, 스캐닝 과정에서 일부 편집도 가능하다. 문서들은 편집 및 수정과정에서 저장되고 검색된다. Blom(1994)에 의하면, 이 연구는 여전히 초기단계에 있다. 자체 생성된 OCR 숫자에서부터 자유형식으로 응답자가 쓴 정보까지는 하나의 큰 단계이며, 스웨덴 통계청에서 수행된 초기연구에서 거부율은 7%~35%에 이른다.

종래의 자료수집방법과 관련된 오차율이 비록 크지는 않지만, 통계기관들은 이러한 프로세스에서 비용을 절감하고, 품질을 향상시킬 수 있는 방법을 계속해 찾고 있다. 응용될 수만 있다면, 새로운 기술들은 조사의 자료입력 측면의 비중을 축소하거나 아예 생략하려고 시도하고 있다. 예를 들어, touchtone 방식의 자료입력에서 응답자는 누름식(touchtone) 전화기를 이용해 자료를 수집하는 사람의 컴퓨터를 시스템을 호출하고, 전화기 버턴을 눌러 질문에 숫자로 응답한다. 마찬가지로, 음성인식 시스템은 화자독립적(speak-independent)이며, 응답자의 특정한 구술응답을 인식한다. 또, 인공지능 문자인식 시스템을 통해 팩스로 들어오는 응답을 해석하는 팩스전송도 자료입력과정을 피하고 있다.(Werking and Clayton, 1995) 결론적으로 말해, 미 국립교육통계센터(Kindel, 1994)와 미 인구조사국(Ambler et al., 1995)에서 데이터를 전자적으로 수집하는 것은 통계기관들이 자료수집의 비용절감에 대한 필요성을 어느 정도 느끼고 있는지를 시사하는 것이라 하겠다.

8.5 통합활동

Bethlehem(제16장), Shanks(1989), Keller(1995), Pierzchala(1990), Baker(1994), 그리고 Weeks(1992)가 지적한 바와 같이, 데이터 처리과정은 그 자체들 간에 그리고, 다른 조사작업들과 통합해야 될 필요성이 있다. 지금까지 데이터처리는 각각의 작업에 별도의 프로세스를 하나의 중앙화된 방식으로 수행되어 왔다. 이것은 각기 다른 참여그룹과 그룹간에 많은 데이터 전송 및 상당수의 규격들로 순차적으로 진행되었다.

새로운 기술의 출현은 분산된 방식으로도 데이터처리를 보다 쉽게 해주며, 앞서 논의한 바와 같이, 단조로운 수작업의 수요를 감소시킬 수 있다. 또 다른 한편으로, 조사과정에서 새로운 기술을 이용한다는 것은 조직의 업무방식을 변화시킬 것이다. 즉, 일괄처리과정이 상호작용적 처리과정으로 바뀔 것이다. 그리고, 처리주기 또한 규모면에서 크게 감소될 것이다. 각각의 처리단계에는 저마다의 소프트웨어가 있으며, 따라서 이러한 소프트웨어들 간의 관계가 거의 구분되지 않는다는 사실이 매우 중요하게 될 것이다. 이것은 여러가지 방법으로 이루어질 수 있다. 네덜란드 통계청에는 조사자료처리에 필요한 소프트웨어 주변에 사용자 친화적인 쉘(shell)을 구성하는 제어센터가 있다(제16장). 이 제어센터는 필요한 데이터 및 메타데이터를 어떤 형식으로도 입력할 수 있으며, 사용자들은 데이터형식과 이 형식들간의 변환을 지정할 필요가 없다. 이와 비슷한 것으로, 일반화된 조사처리 소프트웨어 개발에 관한 연구들이 현재 캐나다 통계청(Turner, 1994)과 스웨덴 통계청(Blom, 1994)을 비롯한 다른 많은 기관에서 수행중이다.

8.6 맺음말(後註)

편집, 코딩, 자료포획이라는 종래의 자료처리작업은 급속한 정보기술의 발전으로 많은 도움을 받게 될 것이다. 여러가지 형태의 컴퓨터를 이용한 조사정보수집은 제각기 다른 절차의 이행과 몇몇 작업의 통합을 가능하게 한다. 하지만, 여전히 기본적인 조사과정과 오차구조에 대해서는 이해가 부족하고, 특히 비효율적인 편집절차와 코딩오차에는 여전히 많은 문제점들이 존재한다. 따라서, 이러한 새로운 기술의 충분한 잠재력을 이용할 수 있기에 앞서, 기본적인 작업을 이해하고 이에 대한 실질적인 개선이 이루어져야 할 것이다.

제 9 절 조사처리를 위한 통합관리시스템

Jelke Bethlehem

네덜란드 통계청

9.1 서문

개인, 가계, 농가, 기업 등으로부터 자료를 수집하고, 이러한 자료를 유용한 통계정보로 바꾸는 것은 국가 통계청의 임무다. 점점 늘고 있는 통계수요는 자료의 수집, 분석, 발간의 전체 프로세스에 대해 조직화되고 효율적인 접근방식을 필요로 함과 동시에, 정확성과 적시성 및 관련성 측면에서 고품질의 기준이 반드시 유지되어야 한다.

종래의 조사과정은 통계의 품질에 영향을 미치는 몇 가지 결점을 가지고 있다. 따라서, 본 절에서는 이러한 통계정보의 품질을 개선하는데 있어 정보기술의 역할에 대해 논의하고자 한다. 오늘날에는 조사과정의 거의 전단계에 충분한 소프트웨어가 있다. 그러나, 이러한 소프트웨어 패키지들은 저마다의 데이터 형식과 제어언어를 가지고 있어 효율적인 프로세스 진행에 걸림돌이 되고 있다. 이를 해결할 수 있는 한 가지 방법은 모든 것을 처리할 수 있는 이상적인 패키지를 설계하는 것이다. 하지만, 이러한 패키지는 존재하지 않을 뿐 아니라, 앞으로도 불가능할 것이다. 이에 따라, 본 절에서는 색다른 접근방식 즉, 통합관리시스템(Integrated Control System)을 제안하고자 한다.

통합관리시스템은 조사자료를 처리하는데 필요한 모든 소프트웨어를 포괄하는 지능적이고, 사용자 친화적인 쉘(shell)이다. 이 관리시스템은 데이터와 메타데이터(데이터에 대한 설명)를 표준화된 형식에서 처리한다. 또한, 이러한 데이터와 메타데이터를 시스템에 들어 있는 개개의 프로그램에서 요구하는 어떤 형식으로도 변환할 수 있다. 따라서, 사용자들은 몇 가지 데이터형식을 지정해야 되는 부담에서 벗어날 수 있고, 데이터형식 간의 변환에 대해 걱정할 필요도 없다. 실제로는 데이터가 여러 방식으로 포맷됨에도 불구하고, 사용자측에서 보면 단 한 가지의 방식으로 데이터를 지정하는 것처럼 나타난다. 나머지는 모두 관리시스템이 알아서 한다.

통합관리시스템은 여러가지 데이터 처리작업을 위한 소프트웨어 모듈들을 가지고 있다. 이 중에서도 없어서는 안 될 모듈은 컴퓨터를 이용한 자료수집, 자료입력, 자료편집을 위한 프로그램들이다. 이러한 모듈은 하나의 데이터집합을 생성하도록 요구된다. 또 그밖에, 제표, 대체, 가중치 부여 및 분석용 프로

그램들과 같은 모듈도 포함될 수 있다. 관리시스템은 사용자들이 소프트웨어 모듈을 덧붙이거나 배제할 수 있도록 유연하고 개방적이다. 이런 식으로 관리시스템은 특정한 환경에 맞게 사용할 수 있도록 수정할 수 있다. 다시 말해, 관리시스템에 적절한 모듈을 둠으로써, 조사를 처리하는 통계청 전부서에서 이를 사용할 수 있고, 이것으로 표준화와 통합이 촉진되는 것이다.

다양한 조사를 처리할 수 있는 하나의 시스템에는 반드시 여러가지 조사를 상세히 설명할 수 있는 하나의 편의장치가 반드시 있어야 한다. 여기서, 메타데이터의 설명은 결정적인 역할을 한다. 메타데이터는 수집, 처리, 출판될 데이터를 설명한다. 일단, 한 조사에 대해 적절한 메타데이터 설명이 컴퓨터가 읽을 수 있는 형식으로 되어 있다면, 관리시스템은 데이터와 메타데이터를 여러 소프트웨어 모듈에서 요구되는 형식으로 변형하는데 필요한 지식을 가지게 된다.

현재까지, 완전한 통합관리시스템을 구현한 통계청은 없다. 그러나, 몇몇 통계청들은 그러한 방향을 지향하는 쪽으로 발전을 거듭해 왔다. 이와 같은 측면에서 본 절은 통합관리시스템의 개념과 이것이 조사결과의 품질에 미치는 영향에 대해 논의하고 있다. 그 중 9.2절은 조사자료품질과 그것의 많은 차원들에 초점을 맞추었다. 종래의 조사과정의 총괄적인 설명을 이용하면, 정보기술의 새로운 발전이 그러한 프로세스를 향상시키는데 어떤 도움을 어떻게 주는 지를 알 수 있다. 그리고, 9.3절에서는 업무과정에서 품질을 향상시키기 위한 몇 가지 접근방식에 대해 설명하였다. 업무과정과 조사과정 간에는 유사점이 많기 때문에, 이러한 접근방식은 조사의 품질을 향상시키는데 적용될 수가 있다. 9.4절은 집중과 표준화 및 통합의 개념에 대해 소개하고 있으며, 9.5절은 이러한 개념들이 세계의 몇몇 메이저 통계청에 의해 어떻게 구현되고 있는 지를 설명하고 있다. 또한, 9.6절은 새로운 접근방식을 구현하고 있는 조직에 가능한 영향에 대해 논의하고, 9.7절은 통계자료의 품질에 가능한 영향을 확립하고자 하였으며, 끝으로 9.8절은 향후에 가능한 몇 가지 도전으로 결론을 맺었다.

9.2 조사에서의 품질

조사는 특정한 모집단에 관한 정보를 수집하기 위해 수행된다. 그리고, 그 결과는 통계의 형식 즉, 지금까지 알려지지 않은 모집단의 특성에 대한 추정치로 나타난다. 그러나, 통계는 모집단 특성의 미지의 값에 정확하게 부합되지는 않을 것이다. 모든 조사업무는 오차의 영향을 받고, 이러한 오차의 크기는 결과의 정확성에 영향을 미친다. 이것이 바로 현재 논의하고 있는 품질의 첫번째 측면이다.

일반적으로, 오차는 두 가지 광의의 범주인 표본오차(sampling error)와 비표본오차(nonsampling error)로 나눌 수 있다. 표본오차는 전체 모집단이 아닌 표본만이 관측된다는 사실에 기인한다. 모든 조사실험의 반복은 각기 다른 표본을 사용하는 결과를 초래하고, 따라서 제각기 다른 추정치를 낳게 된다. 조사자들은 적절한 표본설계와 충분한 양의 표본을 통해 표본오차를 관리하고, 이것은 미미한 표준오차를 가진 편향되지 않은 추정값을 산출한다. 물론, 예산의 따라서는 표본의 규모가 제한될 수도 있다.

비표본오차는 표본 메커니즘의 사용과는 다른 차원의 근원에서 비롯되는 모든 조사오차를 포함한다. 비표본오차는 질문에 대해 잘못된 답변을 얻거나(관측오차) 아니면, 답변을 얻지 못하는(비관측오차) 경우에 기인한다. 관측오차(observation error)의 예로는 측정오차(measurement error), 처리오차(processing error), 과포함오차(overcoverage error)가 있고, 비관측오차(nonobservation error)의 예로는 무응답오차(nonresponse error)와 미포함오차(undercoverage error)가 있다. 그리고, 비표본오차는 관리가 어렵다. 이러한, 비표본오차를 처리하는 가장 좋은 방법은 설계단계에서의 예방책이다. 사실 표본체계의 신중한 선별과, 설문지 설계의 세심한 배려, 면접원에 대한 적절한 교육 및 철저한 사전 조사에 대해서는 누구든 생각할 수 있을 것이다. 그럼에도 불구하고, 비표본오차가 발생하는 많은 원인은 조사과정에 내재되어 있으며, 이런 것은 실제로 피하는 것이 불가능하다.

조사연구에서 품질의 또 다른 측면으로는 적시성(timeliness)을 들 수 있다. 정확한 정보에 의한 의사결정을 위해서는 정책결정자들에게 최신의 통계정보가 필요하다. 이것은 조사설계와 결과물 출판 사이의 시간간격이 가능한 한 짧아야 함을 의미한다. 지금까지 보면, 통계적 정보를 생산한다는 것은 복잡하고, 시간소모적인 과정이다. 이 과정에는 각기 다른 부서와 다른 컴퓨터 시스템들이 참여할 수 있다. 또한, 이 과정은 순차적이다. 이것은 이전 단계가 완료되기 전에는 다음 단계가 진행될 수 없음을 의미한다. 특히, 대규모 조사일 경우에는 결과물이 출판될 준비를 갖추는데 몇 년은 아닐지라도, 최소한 몇 개월은 걸릴 수 있다. 이러한 결과물의 출판이 늦어지면 늦어질수록, 그것의 유용성은 점점 감소하게 된다.

여기서 논의되고 있는 품질의 마지막 측면은 조사의 비용이다. 조사 설계자들은 항상 최소의 비용으로 정확하고 시기적절한 통계를 얻고자 노력해야 한다. 돈은 한번 쓰여지면 끝이기 때문에 그것이 쓰여질 때는 반드시 품질을 향상시키는데 도움이 되어야 한다.

여러가지 조사처리활동이 결과물의 품질에 미치는 영향을 보다 면밀히 살펴보기 위해, 네덜란드 통계청에서는 1984년에 Data Editing Research Project를 추진하였다(Bethlehem, 1987 참조). 여기서는 크고 작은 조사와 사회적, 경제적 조사와 같은 몇몇 다른 조사에서 수행된 자료편집활동에 대한 신중한 평가가 내려졌다. 비록, 이러한 조사들간에 상당한 차이점이 발견되긴 했지만, 일부 일반적인 특징도 확인할 수 있었다. 종래의 자료편집과정은 그림 9.1에 요약되어 있다. 네덜란드 통계청에서는 현재, 이 과정이 현대화되어 있지만, 몇몇 다른 나라의 통계청에서는 여전히 종래와 비슷한 과정이 진행되고 있다.

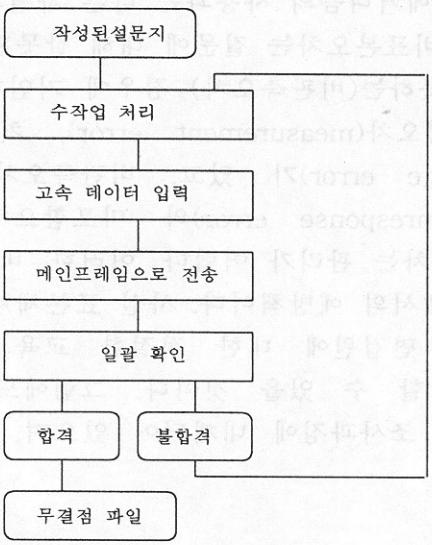


그림 9.1 종래의 자료편집 접근방식

그 과정을 살펴보면 먼저 설문지가 수집되면 분야별 전문가들이 작성된 설문지를 수작업으로 확인한다. 필요한 경우에는 건너뛴 질문에 답을 얻고, 명백한 오차를 설문지에 수정한다. 때로는 다음 단계의 빠른 자료입력을 준비하기 위해 설문지를 새로운 형식으로 복제한다. 그런 다음, 설문지를 자료입력부서로 전달한다. 자료입력자는 오차의 확인없이 빠른 속도로 데이터를 컴퓨터에 입력한다. 컴퓨터는 자료입력에 전용되는 시스템이다. 데이터가 입력되고 나면, 파일은 메인프레임 컴퓨터 시스템으로 전송된다. 메인프레임에서는 맞춤식 오차탐지프로그램이 가동되고, 탐지된 오차목록이 출력된다. 이 오차목록은 주관부서로 송달된다. 전문가들은 에러메시지를 조사하고, 해당 설문지를 참조하여, 명백한 오답은 보다 더 적절한 답으로 대체한다. 수정이 가해진 목록이 다시 자료입력부서로 송달되고, 자료입력자는 컴퓨터에 수정된 내용을 입력한다. 이렇게 해서 수정된 파일은 메인프레임 컴퓨터에 전송되고, 수정된 기록과 기존의 올바른 기록이 합쳐진다. 이와 같은 일괄처리방식의

오차탐지 및 수작업 수정의 순환과정은 불합격 설문지의 수가 충분히 적다고 판단될 때까지 반복된다. 이 모든 과정이 끝나면, "무결점" 데이터집합을 제표 및 추가분석할 준비가 갖춰진다. 지정된 조사에서 이 과정을 면밀히 조사해 보면, 몇 가지 결론을 도출할 수 있으며, 아래와 같이 요약된다.

- 각기 다른 부서에서 많은 사람들이 참여하였다. 이 많은 사람들이 정보를 처리하였다. 즉 응답자는 설문지를 작성하고, 분야별 전문가들은 설문지를 확인해 오차를 수정하며, 자료 입력자는 컴퓨터에 데이터를 입력하고, 전산담당부서의 프로그래머들은 편집프로그램을 짰다. 한 개인/부서에서 다른 개인/부서로의 자료 전송은 오차와 오해, 시간지연의 원인이 될 수 있었다.
- 각기 다른 컴퓨터 시스템들이 가동되었다. 대부분의 자료입력은 전용의 미니컴퓨터 시스템에서 수행되고, 자료편집 프로그램은 메인프레임 컴퓨터에서 가동되었다. 뿐만 아니라, 여러 대의 데스크탑 PC(MS-DOS 환경에서 가동)와 그 외 다른 시스템들도 있었다. 시스템 간의 파일전송은 시간지연의 원인이었고, 부정확하거나 불완전한 규격 및 문서는 오차의 원인이 될 수 있었다.
- 모든 활동이 품질개선을 목표로 하지는 않았다. 많은 시간이 오차의 수정이 아닌 자료입력을 준비하는데 허비되었다. 분야별 전문가들은 자료입력 중에 문제가 발생하지 않도록 설문지의 결점을 없애야 했다. 가장 충격적인 예가 무응답 질문에 "모른다"는 코드를 부여한 것이었다.
- 프로세스가 순환적이었다. (한 부서에서 다른 부서로 또, 한 컴퓨터 시스템에서 다른 시스템으로) 자료입력, 자동확인, 수작업 수정의 순환 과정이 대개 세 번 이상 반복되었다. 이러한 순환주기 때문에, 자료처리는 매우 시간 소모적이고, 비용이 많이 발생하였다.
- 거의 전 단계에서 저마다의 데이터 규격을 필요로 하였다. 비록 기본적으로 같은 것일지라도, 그러한 규격의 "언어"는 참여한 전부서 또는 컴퓨터 시스템마다 완전히 다를 수 있었다. 설문지 자체는 그냥 첫번째 규격일 뿐이었고, 다음 규격은 자료입력 프로그램에 관한 것이었다. 그리고, 자동확인 프로그램은 또 다른 자료규격을 필요로 했으며, 통계패키지 SPSS를 활용하는 것과 같은 제표 및 분석에도 다시 다른 규격이 필요했다. 이 모든 규격에는 변량과 유효응답, 경로 및 유효관계에 대한 설명이 필요했다.

이러한 결론들을 품질의 관점에서 풀이해 보자면, 자료편집단계의 문제점들은 품질에 실질적인 영향을 미친다는 사실이 분명해질 것이다. 즉, 프로세스는 시간 소모적이고, 비용 발생적이며, 오차수정은 너무 늦은 단계에서 시도됨에 따라 효과적이지 못하다.

9.3 조사과정의 개선

통계적 생산과정과 업무과정 사이에는 일반적으로 유사한 점이 많다. 그러므로, 업무과정의 개선에 관한 접근방식은 조사과정에도 똑같이 적용될 수 있다. 여기서는 이 두 가지 접근방식에 대해 논하도록 하겠다. 그 중 하나는 산업생산의 품질개선에 관한 Deming의 이론이고, 또 하나는 보다 최근에 나온 업무과정 재설계(BPR; Business Process Redesign)이다.

품질관리는 산업생산에서 언제나 중요한 쟁점이었다. 이에 관해서는 산업의 품질 및 생산성 개선에 관한 Deming(1986)의 이론이 잘 알려져 있다. 그의 유명한 14가지 경영포인트는 많은 부분이 통계정보의 생산에도 적용된다. 이 중 한 가지는 대량검사에 대한 종속을 중단해야 한다고 지적하고 있다. 최종산물의 검사는 품질개선을 위해서는 시기적으로 너무 늦고, 비효과적이며, 비용발생적이다. 품질은 반드시 설계단계에서 확립되어야 한다. 이러한 지적은 특히 자료수집과 자료편집에 적용된다. 오차가 발생한 다음에 설문지에서 오차를 검출하고, 이를 수정하려고 노력함에 따라, 오히려 오차의 원인을 찾는데 실패하게 되고, 결과적으로 이러한 오차를 제거할 수 없게 되는 것이다.

Deming이 말하는 몇 가지 문제점들은 컴퓨터를 이용한 면접기법을 활용하면 해결할 수가 있다. 지난 10년간 마이크로컴퓨터의 급속한 발전으로 이제는 컴퓨터를 이용한 면접에 마이크로컴퓨터를 활용할 수 있게 되었다. 따라서, 지면 설문지는 질문이 담긴 컴퓨터 프로그램으로 대체된다. 컴퓨터는 면접과정을 관리한다. 그 두 가지 중요한 활동은 다음과 같다.

1. 경로관리 : 컴퓨터 프로그램은 다음 질문을 결정하고, 이를 화면상에 보여준다. 이러한 결정은 앞에서 질문한 것에 대한 응답에 따라 달라질 수 있다. 이것은 설문지에서 올바른 순서를 따라야 하는 면접원의 부담을 덜어준다. 결과적으로, 이제 더 이상 경로오차(routing error) 즉, 무관한 질문을 하거나 관련된 질문을 하지 못하는 경우는 발생하지 않을 것이다.
2. 오차확인 : 컴퓨터 프로그램은 질문에 대한 응답을 확인한다. 각각의 응답이 입력된 후에는 즉시 범위확인이 이루어지고, 관련된 모든 응답이 입력된 후에는 일관성 확인이 진행된다. 이 때, 오차가 탐지되면, 컴퓨터는 경고신호를 보내고, 이에 따라 입력된 응답을 수정할 수가 있다. 프로그램은 탐지된 오차가 모두 수정될 때까지 다음 질문을 진행하지 않는다.

컴퓨터를 이용한 면접은 1970년대에 CATI(Computer Assisted Telephone Interviewing)에서부터 비롯되었다(Nicholls and Groves, 1986; Groves and Nicholls, 1986 참조). 그러던 것이 1980년대 들어 소형의 휴대용 컴퓨터가 등장하면서 CAPI(Computer Assisted Personal Interviewing)가 가능하게 되었다. 초기의 실험은 Bemelmans-Spork and Sikkel(1985)와 Van Bastelaer et al.(1987)가 수행하였다. 그리고, 보다 최근으로 와서는 CASI(Computer Assisted Self-Interviewing)가 활용되고 있다. 이 면접방식은 기업경기실사에 특히 중요하다. CASI 응용에 대해서는 Wieringa and Fokké(1992)와 Stol(1993)이 잘 설명하였다.

컴퓨터를 이용한 면접은 사회 및 인구통계 조사에서 매우 성공적으로 검증되었다. 여기에는 종래의 면접과 비교해 다음과 같은 세 가지 중요한 장점들이 있다.

- 이러한 면접방식을 적용하면, 면접을 진행하는 중에 자료입력 오차와 비일관성에 대한 탐지가 가능하다. 이와 같이 탐지된 문제점들은 면접원과 응답자가 모두 확인할 수 있으며, 이 때가 이것을 수정할 수 있는 최적기가 되는 것이다. 이런 식으로, 컴퓨터를 이용한 면접은 보다 정확한 자료를 산출한다.
- 컴퓨터를 이용한 면접은 조사과정에서 자료의 수집, 입력, 편집 등 세 가지 단계를 통합한다. 면접원들은 질문에 대한 응답을 기록하는데 컴퓨터를 이용할 수 있으므로, 면접하는 도중에 자료입력을 처리할 수 있다. 또한, 면접이 진행되는 동안 필요한 모든 확인이 이루어진다면, 많은 오차들이 탐지 및 수정될 수 있고, 따라서 별도의 자료편집단계가 불필요하게 될 것이다. 또한, 무엇보다 면접원들의 파일은 하나의 데이터 파일로 통합되고, 그러한 정보는 "무결점" 정보로서 다음 처리를 준비하게 된다. 이런 식으로, 컴퓨터를 이용한 면접은 조사과정의 시간과 비용을 절감시킨다.
- 컴퓨터를 이용한 면접은 면접원들의 작업을 보다 쉽게 해 준다. 컴퓨터는 설문 전체의 적절한 경로를 결정하기 때문에 면접원들은 각각의 질문에 보다 집중할 수가 있다.

컴퓨터를 이용한 면접은 조사과정의 효율성과 결과물의 품질을 향상시켰다. 그 예는 Tortora(1985), Groves and Nicholls(1986), Catlin et al.(1988)이 CATI에 대해, Van Bastelaer et al.(1987), Sebestick et al.(1988)이 CAPI에 대해 잘 설명하였다. 그리고, Week(1992)는 CAPI와 CATI 기법에 관한 연구개요를 제시하였다.

경제조사에서 컴퓨터를 이용한 면접의 응용은 사회조사 만큼 많이 파급되어 있지는 않다. 기업경기실사에서 응답자들은 대개 즉시 활용할 수 있는 정보를 가지고 있지 않다. 이들은 관리기록, 컴퓨터 파일을 참조하고, 다른 사람이나 부서와 협의할 시간을 필요로 한다. 사실, 이런 경우에는 지면 설문지가 보다 적절할 수도 있다.

하지만, 컴퓨터를 이용한 면접이 기업경기실사에는 불가능하다 하더라도, 자료입력과 자료편집의 통합으로, 효율적인 면에서는 소득이 있을 수 있다. 이 기법은 CADI(Computer Assisted Data Input) 또는 CADE(Computer Assisted Data Editing)라 한다. CADI는 분야별 전문가들이 지면 설문지의 자료입력과 자료편집을 통합할 수 있는 상호작용적이고, 지능적인 환경을 제공한다. 자료를 처리하는 데는 두 가지 전략이 적용될 수 있다. 그 중 첫번째 전략은 분야별 고용인들이 마이크로컴퓨터를 가지고 한 묶음의 설문을 건별로 처리하면서 작업을 수행하는 것이다. 이들은 자료를 입력하고, 설문이 완료된 후에는 모든 종류의 오차를 테스트하도록 체크 옵션을 실행시킨다. 그러면, 검출된 오차가 화면상에 나타나고, 이들은 다시 설문을 참조하거나, 응답자와 연락해 오차를 수정한다. 모든 오차를 제거하고 나면, 무결점 기록이 파일로 작성된다. 만일, 무결점 기록을 생성하지 못할 경우에는 문제가 있는 기록만 별도의 파일로 작성할 수가 있다. 이 문제가 있는 건은 나중에 전문가들이 따로 처리할 수 있으며, 이 때도 물론 CADI 시스템을 이용하게 된다. 이 접근방식은 비교적 소량의 복잡한 설문에 효율적이다.

두 번째 전략으로는 자료입력자들이 CADI 시스템을 이용하여 오차확인없이 자료를 미리 입력하는 것이다. 입력이 완료되면, CADI 시스템은 모든 기록을 일괄적으로 확인하고, 틀린부분을 표시해 준다. 그 다음은 분야별 전문가가 맡는다. 이들은 잘못된 기록을 건별로 처리하고, 검출된 오차를 수정하려 한다. 이 접근방식은 소량의 간단한 설문에 적합하다.

업무과정에서의 품질개선에 대한 또 다른 접근방식으로는 업무과정 재설계(BPR; Business Process Redesign)가 있다. 업무과정재설계의 개념은 Hammer and Champy(1994)의 저서에도 나타나 있다. 이들은 오늘날의 많은 기업들이 Adam Smith의 The Wealth of Nations에 의해 2세기 전보다 정형화된 원칙에서 운영되고 있다고 역설한다. 이러한 원칙들은 대량의 재화와 용역을 위해 팽창하는 시장속에서 작용되고 있었지만, 오늘날의 급변하는 시장과 경쟁하는 풍토에서 살아남는 기법이 되지는 못했다. 종래의 접근방식은 업무과정을 단순한 작업으로 세분화하는데 초점을 맞추고 있다. 이러한

작업지향적 접근방식의 장점은 비교적 단순한 작업은 선진기술이나 지식을 그다지 요하지 않는다는 것이다. 하지만, 단순작업은 하나로 묶는데 복잡한 과정을 필요로 한다. 200년 동안, 기업들은 단순화된 작업의 장점을 얻기 위해 이와 같이 복잡한 과정으로 인한 불편함과 비효율성 및 고비용을 감수해 왔다.

일반적으로 작업지향적 프로세스는 서로 다른 부서에 있는 제각기 다른 사람들을 참여시킨다. 이 프로세스는 어떤 한 부서에서 작업을 시작하기 위해서는 다른 부서가 작업을 완료할 때까지 기다려야 한다는 의미에서 순차적이라 할 수 있다. 더욱이, 몇몇 업무과정은 반복적인 성분을 가지고 있다. 따라서, 하나의 프로세스가 완료되려면 몇 번의 싸이클을 겪게 되는 것이다. 이러한 순차적이고 반복적인 프로세스들은 시간소모적이고, 오차발생적일 수 밖에 없다.

작업지향적 조직에 있는 사람들은 프로세스 지향적이지가 않다. 이들은, 작업과 일, 구조에만 중점을 둘 뿐 프로세스는 염두에 두지 않는다. 이러한 프로세스의 분열로 인해 누구도 프로세스를 전체적인 하나로 관리하지 않고, 또 그에 대해 책임을 지는 사람도 아무도 없다. 그 뿐만 아니라, 업무과정에 참여하는 사람들은 내향적(inward-oriented)이다. 부서는 그들만의 세계이고, 만족을 위해서는 부서의 장이 가장 중요한 사람이다. 이들은 고객의 존재를 망각하고 있는 것처럼 보인다.

품질에 대한 강조는 업무과정 개선만으로는 부족하다. 별개의 각 작업이 효율적으로 수행되고 있다 하더라도, 전체적인 프로세스는 그렇지 못하다. 따라서, 첨단정보기술의 응용은 현재의 작업을 자동화하는데만 사용된다면 아무런 도움이 되질 못할 것이다. 근본적인 개선을 이루려면 다른 무언가가 필요하다. 업무과정재설계는 비용이나 품질, 서비스, 속도와 같은 중요한 동시대의 성과척도에서 혁신적인 개선을 이룩해낼 수 있는 근원적인 재사고와 업무과정의 과감한 재설계를 의미한다.

BPR은 작업지향적 접근방식을 배제하고, 프로세스를 둘러싼 작업구성을 의미한다. 또, 이것은 과거의 인습과 전통적인 규칙을 깨는 것을 의미한다. BPR은 일맥상통하는 작업을 하나로 모음으로써 프로세스의 단절을 방지한다. 따라서, 종래의 부서간 경계를 넘나드는 것은 더 이상 프로세스를 지연시키지 않을 것이고, 종래의 순차적이고 반복적인 업무과정의 성격 또한 현격히 줄어들 것이다.

BPR은 여러가지 활동들이 specialist가 아닌 generalist에 의해 수행되게 함으로써 전문화의 포기를 증진시킨다. 몇 가지 다른 일들은 하나로 통합되고, 프로세스 내 여러 단계가 자연스러운 순서에 따라 진행될 수 있다. 이것은 결과물의 가치에 부가될 수 없는 광범위한 확인 및 관리단계들을 생략하거나 줄일 수 있게 해 준다. 또, 순차적인 성격을 없앰으로써 수평적 방식으로 업무를 수행하고, 그 결과 프로세스가 가속화되게 해 줄 것이다.

BPR은 정보기술의 창의적 활용의 중요성을 강조한다. 그러나, 정보기술의 새로운 발전은 종래의 프로세스에 들어 있는 작업을 단순화 하는데만 이용돼서는 안된다. 이것은 최소한의 성과개선만 이끌게 될 뿐이다 보다 큰 실질적인 개선은 컴퓨터 기술을 근본적으로 새로운 방식으로 활용함으로써 이룩될 수 있다.

종래의 조사과정에 대한 설명과 종래의 업무과정에 대한 설명간에는 상당한 유사점이 있다. 또, 두 과정 모두 동일한 문제점을 안고 있다. 그것은 프로세스 단절, 프로세스가 아닌 작업에 대한 중점, 여러 부서의 참여, 순차적이고 순환적인 성격에 관한 것이다. 네덜란드 통계청은 통계적 생산과정을 재설계함에 있어 BPR과 같은 접근방식을 따랐다. 그 결과, 새로운 방식의 조사처리가 생겨났고, 실제로 이것은 BPR "avant la lettre"의 한 사례였다.

9.4 새로운 조사과정에 대하여

Data Editing Research Project와 Business Process Redesign 원칙의 결론은 모두 작업지향적인 조사과정을 버리고, 대신 프로세스를 전체적인 하나의 것으로 바라보는 근본적으로 다른 접근방식으로 대체할 것을 강조하고 있다. 이러한 변화는 그에 따른 적절한 조직구조와 업무환경 및 도구를 필요로 한다. 이런 맥락에서 본 절에서는 네덜란드 통계청을 비롯한 여러 국가 통계청에서 발생된 변화에 대해 설명하고자 한다. 이 기관들은 모두 다음 조건을 만족하는데 목표를 두고 있다.

1. 집중 : 특정한 하나의 조사와 관련된 모든 자료처리활동은 가능한 한 하나의 부서에 집중되어야 한다. 특히, 주관부서에 집중하는 것이 보다 바람직하다.
2. 하드웨어 표준화 : 모든 자료처리활동은 가능한 한 동일한 유형의 컴퓨터 기반에서 수행되어야 한다.
3. 소프웨어 표준화 : 모든 자료처리활동은 반드시 맞춤식 소프트웨어 대신 표준 소프트웨어로 수행되어야 한다.

4. 통합: 자료처리에 필요한 모든 소프트웨어는 반드시 통합시스템의 일부분이어야 하고, 조사필요한 모든 정보가 담겨 있는 메타데이터 정의를 사용해야 한다. 이 메타데이터 정의는 조사에 관한 정보의 주된 원천으로 반드시 모든 시스템과 부서에서 사용되어야 한다.

이상의 조건들을 만족하게 되면, 종래 접근방식의 일부 단점이 사라진다. 즉, 모든 작업은 한 부서의 한 조사팀이 표준 소프트웨어와 하드웨어를 이용하여 수행하기 때문에 부서 및 컴퓨터 플랫폼간의 시간소모적인 자료전송이 없어진다. 조사팀은 여러 활동을 통합하고, 이에 따라 프로세스의 순환이 사라진다.

관리시스템은 조사과정에 필요한 여러 소프트웨어 패키지들을 통합한다. 오늘날, 컴퓨터는 조사과정의 거의 모든 단계에서 이용된다. 그리고, 여기에는 각 활동 유형에 따른 방대한 랩탑 컴퓨터 면접용 자료수집 프로그램, 데이터 저장 및 관리용 데이터베이스 프로그램, 특수 제표 프로그램, 복잡한 분석을 위한 강력한 패키지, 고품질 그래픽을 생성하는 프리젠테이션 프로그램 등 소프트웨어가 있다. 대부분의 프로그램들은 하나의 특정한 작업에 초점을 맞추고 있다. 어떤 패키지는 자료입력을 할 수 있는가 하면, 또 다른 것은 통계분석에 필요할 수가 있다. 또, 전부는 아니지만, 프로세스 내 몇몇 단계를 커버하려고 노력하는 패키지들도 있다. 그러나, 실제로 이것은 제대로 사용되지 않는다. 하나의 패키지는 하나의 작업에는 매우 유용할 수 있지만, 다른 작업은 또 다른 패키지를 필요로 할 수 있다.

하나의 조사에서 몇 가지 다른 컴퓨터 패키지를 가지고 일하는 것은 효율적이지 못하다. 각 패키지는 자체의 데이터 형식과 데이터 정의를 필요로 한다. 이것은 데이터 파일 및 셋업 파일의 반복된 변환을 의미하는 것일 수 있다. 또, 이것은 시간과 컴퓨터 자원을 낭비하고, 오차의 잠재적 원인이 된다. 따라서, 전체 조사과정을 관리할 수 있는 하나의 완벽한 패키지가 있으면 가장 이상적일 수 있으나, 이것은 환상에 불과하다. 통계자료의 처리는 하나의 패키지에서 모든 것을 해결할 수 없는 다양한 측면을 가지고 있다. 게다가, 정보기술 뿐만 아니라, 통계분야에서도 새로운 개념들이 급속히 생겨나고 있다. 그러므로, 이러한 개념들은 절대로 하나의 시스템에서 한 번에 실현될 수가 없다. 아마도, 궁극적인 조사처리 패키지는 결코 현실화될 수 없을 것이다.

통합관리시스템은 색다른 문제해결 접근방식을 제공한다. 그 개념은 조사처리가 기본적으로 데이터와 메타데이터의 조작을 위한 일단의 활동이라는 생각에 근거한다. 데이터는 설문형식 위에 작성된 질문에 대한 응답이고,

처리과정에서 유용한 통계정보로 변형된다. 관리시스템은 모든 조사 소프트웨어를 하나의 사용자 친화적인 셸(shell)로 통합한다. 조사원들은 이러한 셸(shell)을 이용하여 조사를 정의하고, 자료를 수집하며, 데이터를 처리한다. 종래의 자료수집을 보면, 시스템은 반드시 지면 설문지를 생성할 수 있어야 하고, 그에 상응하는 자료입력 및 자료편집 프로그램이 있어야 한다. 또한, 컴퓨터를 이용한 면접 프로그램들일 경우에는 시스템은 반드시 CAPI나 CATI 또는 CASI를 위한 프로그램을 생성할 수 있어야 한다.

통합관리시스템은 고정된 실체가 아니라 유동적인 시스템이어야 한다. 그리고, 사용자들은 각자의 성분을 포함시킬 수 있어야 한다. 따라서, 자체적인 소프트웨어 개발은 반드시 지양되어야 하는 고비용의 시간소모적인 활동이다. 즉, 유용한 표준 소프트웨어를 시장에서 구입할 수 있다면, 그것을 써야 한다. 그리고, 관리시스템은 반드시 그러한 소프트웨어를 셸로 통합하는 기능을 제공해야 한다.

통합관리시스템은 여러 구성성분들 간의 커뮤니케이션을 담당해야 한다. 따라서, 반드시 처리되어야 하는 조사자료에 관한 정보를 각 구성성분에 제공할 수 있어야 한다. 여기서, 메타데이터의 절대적인 역할이 중요한 것이다.

문헌들을 보면, 메타데이터의 개념에 대해서는 정의가 잘 되어 있지 않다. Berg et al.(1992)에 의하면, 일반적이지만 오히려 모호한 정의가 "데이터에 관한 데이터"라는 것이다. 통합관리시스템에서 사용하는 경우, 메타데이터는 조사자료의 수집, 처리, 출판에 필요한 정보에 관한 정형적인 설명과 같이 보다 실용적인 방법으로 정의된다. 이러한 메타데이터의 정의에는 아래 사항이 포함되어야 한다.

- ■ 조사변량의 정의 : 각 변량은 반드시 식별할 수 있는 이름과 유효값의 범위, 코드목록(정성변수일 경우), 필요한 질문 텍스트(가능한 여러 언어의) 및 변량을 기록하기 위한 기타 텍스트를 가지고 있어야 한다.
- ■ 자료모형 : 자료모형은 변량들간에 그룹이나 계층, 반복과 같은 측면의 관계를 설명한다. 예를 들어, 그룹은 다른 그룹속에 포함될 수도 있고, 여러 번 반복될 수도 있다.
- ■ 경로지침 : 경로지침은 질문(변량)을 묻는 순서와 그 상황을 설명한다. 이러한 지침은 관련된 질문만 하고, 관련없는 질문은 건너뛰게 한다.

■ 관계 : 어떤 관계들이 변량의 값을 구속할 경우에는 수집된 자료에 대한 일관성을 확인할 수 있도록 그 때마다, 이러한 제약요소에 대한 설명이 있어야 한다. 비 일관성은 대개 오차의 원인이 된다. 이러한 오차의 탐지와 수정은 자료품질을 향상시킬 수 있다.

■ 계산 : 모든 조사변량들이 응답의 직접적인 결과는 아니다. 때로는 산술적 표현을 통해 새로운 변량이 다른 변량에서 파생되기도 한다. 게다가, 이것은 누락되었거나 잘못된 답변을 "합성된" 답변으로 반드시 대체할 수 있어야 한다. 이러한 답변은 계산의 결과일 수도 있다.

■ 다른 파일과의 연계 : 어떤 조사에서는 데이터 파일을 다른 조사나 이전 조사와 연계하는 것이 매우 중요하다. 이것은 다른 조사에서 정보를 가져올 수도 있게 해 준다. 또한, 패널조사의 데이터를 비교하는데도 이용할 수 있다.

이러한 메타데이터의 정의는 표본설계의 정의 그리고, 조사변량과 다른 조사변량간의 관계와 같이 조사에 관한 보다 많은 정보를 포함하는 것으로 확대시킬 수 있다.

통합관리시스템으로 조사를 처리하는 시발점은 자료모형의 정의에 있다. 자료모형은 메타데이터 정의 뼈대가 된다. 그렇다면, 자료모형은 무엇과 닮았을까? 경험적으로 볼 때, 많은 조사자들은 세계를 여러 가구가 있으며, 각 가구에는 구성원들이 있고, 또 이 구성원들은 건강문제를 가지고 있다는 등과 같이 계층적으로 바라본다. 만일, 이것이 사람들이 조사에 대해 생각하는 자연스러운 방법이라면, 조사설계도구는 반드시 계층적 자료모형을 직선적이고 쉬운 방법으로 처리할 수 있어야 한다. 또한, 이것은 단순한 데이터구조를 통해 계층을 구현하는 식으로 자료모형설계에 모듈식 접근방식을 추구해야 한다.

대량의 복잡한 데이터구조가 가지고 있는 주된 문제는 전체적인 구조트랙을 유지하는 것이다. 수천가지 변량들을 동시에 고려하는 것은 불가능하다. 필요한 것은 이것을 하나의 변량이나 혹은 일단의 제한된 변량그룹으로 한번에 집중할 수 있고, 이들의 관련된 구조적 관계를 한눈에 파악할 수 있도록 그룹핑하는 방법이다. 바로 이런 문제를 해결하기 위해 도입된 것이 하위모형(submodel)의 개념이다. 하위모형은 하나의 모듈화된 방식으로 메타데이터를 그룹핑하기 위한 구성이다. 그 중, 내부로부터의 하위모형은 다른 변량들과 독립적으로 서로 밀접하게 관련된 변량그룹으로 확인될 수 있고, 외부로부터의 하위모형은 다른

하위모형속에 자리잡고 있거나 여러 번 반복되는 하나의 정보단위로 확인될 수 있다.

자료모형이 계층적이기 때문에, 그 구조는 나무(tree)와 같이 표현될 수 있다. 간단한 예를 들어 보자. MS-DOS(Dosshell) 또는 Windows(File manager) 환경에서 구동되는 파일관리자에 친숙한 사람들은 자료모형이 시각화되는 방법과 파일관리자가 디렉터리 구조를 시각화하는 방법 간에 유사한 점을 발견할 수 있을 것이다.

자료모형의 예로 가구조사를 설명해 보자. 이것은 두 가지 단계를 가진 계층모델이다. 최상위 단계는 세 가지 변량만 있는 가구를 의미하는데, 주소(street와 town)와 가구의 크기(size of household)가 있다. 가구의 구성원은 몇 가지의 특성을 가지고 있다. 이러한 특성들은(이름, 성별, 혼인여부, 나이, 자녀수(여성의 경우), 직업(유/무), 그리고 직업이 있는 경우 그에 대한 설명) 변량집합에 기록된다. 이 조사에서 가능한 자료모형은 그림 9.2와 같다.

그림 9.2의 좌측에 있는 데이터 트리는 최상위 단계의 자료모형을 나타낸 것이다. 이것은 일반적인 구조에 대한 적절한 시각을 제시하고 있다. 이를 보면, 세 개의 필드(Street, Town, HHSize)와 하나의 하위모형(PERSON)이 있다. 기본필드를 하위모형과 구별하기 위해 하위모형의 명칭은 대문자로 쓰여져 있다. 여기서, 플러스(+) 부호는 기본필드가 도시되지 않은 하위모형을 의미한다. 하위모형이 펼쳐지면 즉, 하위모형에 있는 모든 필드가 가시적으로 나타나면, 세부항목이 모두 드러나는 오른쪽에 있는 데이터트리로 표현된다. 자료모형에는 하부구조가 최대 10번까지 반복될 수 있음을 의미하는 LIST [1..10] OF 라는 구문이 있다.

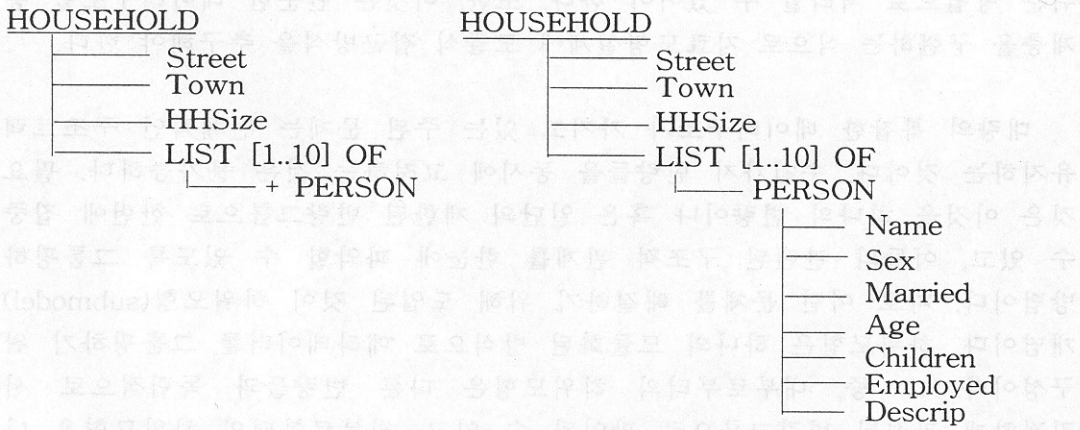


그림 9.2 계층적 자료모형

하위모형은 복잡한 자료모형을 하나의 모듈화된 방식으로 설계하는 데 있어 강력한 도구가 된다. 하위모형의 사용에는 몇 가지 장점이 있다. 첫째, 모형의 구조가 보다 분명해진다. 둘째, 개발과 유지관리가 간편하다. 셋째, 재사용이 가능하다. 제대로 설계되고 시험된 하위모형은 다른 조사의 자료모형에도 쉽게 적용될 수 있다. 이것은 개발에 드는 시간을 절감해 주고, 조사간 상호조정을 증대시킨다.

그림 9.2와 같은 자료모형에는 조사처리에 필요한 모든 정보가 들어있지는 않다. 메타데이터 규격은 자료모형의 구조를 반영해야 할 뿐만 아니라, 변량 자체에 관한 상세 정보와 변량을 처리하는데 적용되는 규칙 및 조사과정 전체에 필요한 기타 모든 정보도 제공해야 한다. 통합관리시스템의 메타데이터 규격 언어는 반드시 이 모든 것을 고려해야 한다.

9.5 몇몇 국가 통계청에서의 개발

대부분의 국가 통계청들은 조사과정의 개선을 목표로 한다. 그런 의미에서, 본 절에서는 몇몇 주요한 통계청에서 이루어진 개발에 대해 설명하고자 한다. 일반적인 경향은 가능하다면, 표준 소프트웨어를 사용하고, 소프트웨어 간 데이터 전송과 데이터 설명이 간단하다는 것이다. 혹자는 여러 단계를 통합함으로써 어떻게 통계적 과정의 효율성이 증대되고, 비용이 감소되는 지를 설명하기도 한다. 예를 들면, Baker(1987), Brakenhoff et al.(1987), Outrata and Chinnappa(1989), Bethlehem(1992)이 대표적이다.

1990년, 미인구조사국은 조사과정에 새로운 컴퓨터 기술을 적용하는 것이 뒤처져 있다는 결론을 내렸다. 여기에는 일반적인 자동화정책도 없었다. 고도로 분산화된 소프트웨어 개발은 각 조사에 특화된 시스템의 결과를 가져왔으며, 이것은 자원낭비의 원인이 되었다. 인구조사국의 답은 CASIC 개념이었다. CASIC란 컴퓨터를 이용한 조사정보수집(Computer Assisted Survey Information Collection)을 말한다. CASIC는 자동화 및 정보통신 활용의 극대화를 통한 조사과정 여러 단계의 통합을 목적으로 한다. 이로 인해, 인구조사국의 현재 활동은 자료수집 단계에 집중되어 있다. 컴퓨터를 이용한 면접기법(CAPI, CATI, CASI) 뿐만 아니라, TDE(Touch-tone Data Entry), VRE(Voice Recognition Entry)와 같은 유망한 기술들을 모두 고려하고 있다. 다만, 조사과정의 후행단계(코딩, 대체, 가중치 부여, 제표)를 통합하는 것은 아직 채택되지 않았다. 이것은 장기적 목표로 고려되고 있다. 통합과 표준화는 비용효과성을 개선하고, 자료품질을 향상시키며, 자료처리시간을 절감할 것으로 기대된다. 인구조사국의 CASIC 활동에 대해서는 Mattchett et al. (1994)이 설명하였다.

캐나다 통계청은 20년 넘게 표준적인 조사 소프트웨어를 개발해 왔다. 주목적은 바로 비용이다. 소프트웨어의 개발은 비용이 많이 드는 것으로 고려된다. 이것은 표준 소프트웨어에도 마찬가지로 적용된다. 그러나, 표준 소프트웨어 개발이라는 접근방식은 반복되는 관습적인 개발에 비하면 비용이 훨씬 적게 드는 것이다. 비용절감은 소프트웨어 개발에만 해당되지 않는다. 이것은 조사시스템 설계와 교육에도 영향을 미친다. 캐나다 통계청의 접근방식은 언제나 조사과정에서 지정된 단계를 자동화하는 도구를 제공하는 것이었다. 예로는, GSAM(표본추출용), DC2(자료포획용), GEIS(편집 및 대체용), ACTR(자동코딩용), GES(추정용), Oracle(관계형 데이터베이스 시스템), SAS(통계적 분석용) 등이 있다. 표준 패키지들은 함께 작동할 수 있지만, 통합된 시스템을 개발하려는 시도는 지금껏 없었다. 각 패키지는 저마다의 언어를 가지고 있다. 캐나다 통계청은 하드웨어 표준화에 대한 정책은 구현하지 않았다. 다만 플랫폼의 독립성을 제공하기 위해 복합 플랫폼에 대한 이식성(portability)을 강조하고 있다. 미래의 구조는 상업용 패키지와 자체 개발된 패키지가 하나로 통합된 것이 될 것이며, 여기서의 통합은 표준 그래픽 사용자 인터페이스와 표준 프로토콜 및 공통의 스크립팅 언어가 고려된 것을 말한다. 보다 자세한 내용은 Outrata and Chinnappa (1989)와 Turner(1994)를 참조한다.

영국 통계청(ONS)의 사례에 대해서는 Manners and Diamond(1994)와 Martin(1993)이 잘 설명하고 있다. ONS는 지면 설문지를 전자적인 형식으로 옮기는 것에 대한 중요성을 일찍이 인식하였다. 이에 따라, CAPI에 대한 개발연구도 1987년에 시작되었다. 비록, 이러한 변화의 초기비용이 상당하긴 했지만(예를 들면, 랩탑 컴퓨터상의 도구개발과 교육 등으로), 비교적 짧은 기간에 많은 절감을 이루어냈다. ONS는 자료수집단계에서의 여러 활동을 통합하는데 집중하고 있다. 그래서, 코딩을 면접과정에 통합하는데 표준 소프트웨어(Blaise)를 사용한다. 이에 따라 자료입력과 자료편집, 코딩이 더 이상 필요없어 많은 비용이 절감되고 있다. 자료수집과정의 표준화 및 통합은 ONS의 경쟁력을 높여준다. 현재, ONS의 경쟁력으로는 보다 낮은 개발비용과 보다 짧은 개발기간 그리고, 각 조사마다 새로운 도구를 재창조할 필요가 없다는 점, 그러면서도 자료의 품질은 보다 높다는 것이다. 다만, 자료수집활동을 조사과정의 다른 부분과 통합하려는 시도는 아직까지 없다.

스웨덴의 개발에 대해서는 Blom(1994)이 다루었다. 스웨덴 통계청은 통계시스템의 분산에 대한 정부의 결정이 필요하였다. 현재는 몇몇 다른 기관에서 자금을 조성하고, 대행기관에 조사를 위탁할 수 있다. 스웨덴 통계청의 경우에는 이것이 통계시장에서의 경쟁력을 높여 주는 것이다. 또한, 예산상의

제약도 보다 비용효과적인 자료수집 및 자료처리방법에 대한 연구를 수행하는 원동력이 되었다. 그리고, 컴퓨터를 이용한 면접기법(CAPI와 CATI)이 특히 강조되었으며, 다른 자료수집기법에 대한 연구로는 CASI, TDE, 스캐닝, 팩스 출력 등이 있다. 이러한 기법들은 아직까지 통합되지는 않고 있다.

호주 통계국의 CASIC 기술도입은 집약적인 키보드 작업에 투입될 수 있는 시간상의 제약으로 인해 "heads down" 자료입력활동을 없앰으로써 다소 가속화되었다. CADI 기법을 이용하면, 자료입력은 다기능 환경속에서의 하나의 작업이 된다. 표준환경은 CADI 작업을 위해 개발되었다. IPS 시스템은 Windows 기반의 프론트엔드와 Oracle 데이터베이스 기반의 백엔드를 가진 클라이언트/서버 아키텍처를 가지고 있다. 또, 최근에는 Blaise가 CAPI와 CATI용 표준 패키지로 선정되었다. 일부 조사에서는 OCR과 OMR 기법도 사용되었다. 다만, TDE에 대한 시도는 성공적이지 못했다. 그리고, 여러 자료수집기법들과 다른 단계에 사용되는 소프트웨어를 조사과정에 통합시킨 것은 없다. 이에 대한 개괄적인 내용은 Tozer and Jaensch(1994)가 잘 설명하고 있다.

네덜란드 통계청의 정책은 언제나 조사과정 속 전단계의 표준화 및 통합이었다. 각 활동에는 표준 소프트웨어가 있고, 통합은 Blaise Control Centre에 의해 현실화된다. 이것이 바로 통합관리시스템 개념의 구현인 것이다. 이것은 컴퓨터를 이용한 조사처리를 위한 Blaise Ⅲ 시스템에 기반하고 있다. 이 시스템은 필요로 하는 소프트웨어에 따라 각기 다른 많은 형식으로 자동변환 될 수 있는 장치 가독성 메타데이터를 사용한다. Blaise Ⅲ는 컴퓨터를 이용한 면접과 종래의 지면 면접방식을 모두 지원한다. 뿐만 아니라, 멀티모드 조사도 가능하다. Blaise Control Centre에는 데이터 조작용의 Manipula, 가중치 조정용의 Bascula, 그리고 제표용의 Abacus와 같은 몇 가지 추가모듈이 탑재되어 있다. 자료분석에는 상업용 패키지 SPSS와 Stata가 사용된다. Control Centre는 데이터와 메타데이터의 포맷을 신속히 처리한다. 이와 같은 네덜란드 통계청에 관한 상세정보는 Bethlehem and Keller (1992), Bethlehem (1992), Schuerhoff (1993)의 논문에서 확인할 수 있다.

9.6 조사기관에 대한 효과

조사과정의 과감한 재설계는 조사기관과 그 곳에서 일하는 고용인들의 구조에도 영향을 미친다. 이런 맥락에서, 본 절은 집중(concentration), 표준화(standardization) 및 통합(integration)의 정책을 구현할 때 가능한 효과에 대해 논하도록 하겠다.

모든 조사활동의 한 부서로의 집중은 (가능한 한) 새로운 노동 분화를 의미한다. 컴퓨터 부서에 의해 이전에 수행되던 조사처리활동은 주관부서로 이관될 것이다. 그리고, 그 결과 컴퓨터 부서는 EDP(electronic data processing) 부서에서 IT(information technology) 부서로 바뀌게 되고, 조직내에서 근무하는 사람들을 위해 적절한 인프라(하드웨어 및 소프트웨어)를 제공하고 유지하는 책임을 맡게 될 것이다.

집중은 비교적 컴퓨터에 경험이 많지 않은 통계전문가들과 분야별 전문가들에게 조사처리시스템을 관리하는 책임이 주어짐을 의미한다. 이들은 반드시 이러한 작업을 성공적으로 수행할 수 있는 적절하고 사용자 친화적인 환경을 유지해야 한다. 이런 면에서 통합관리시스템은 결정적인 역할을 한다.

집중은 컴퓨터 전문가들이 조사팀에 포함되어야 함을 의미하는 것은 아니다. 이것은 통계전문가들이 specialist로서의 역할을 그만두어야 됨을 의미한다. 이들은 컴퓨터를 포함하여, 조사과정의 전반적인 측면에 대해 폭넓은 지식과 경험을 가진 generalist가 되어야 한다. 이러한 접근방식의 결과는 많은 노력이 교육활동에 투자되어야 한다는 것이다. 이에 따라, 주관부서의 근무자들은 자신의 업무에서 보다 다양한 경험을 하게 된다. 이들은 더 이상 작업지향적이지 않고, 보다 프로세스 지향적으로 될 것이다. 이들은 각자의 업무에 있어 조직 전반을 관리함에 따라 보다 고품질을 결과를 산출할 수 있을 것이다.

물론, 새로운 접근방식의 성패는 궁극적으로 주관부서 근무자들의 수용력에 달려있다는 사실은 깨달아야 한다. 그리고, 이러한 변화를 뒷받침하고, 근무자들이 각자의 업무에 있어서 뿐만 아니라, 조직 전체에 있어서도 이의 장점을 확신할 수 있도록 하는 통계기관의 총괄적인 관리가 매우 중요한 몫을 차지한다.

표준화는 하드웨어와 소프트웨어 모두의 표준화를 의미한다. 그 중, 하드웨어 플랫폼 측면에서 보면, 분명한 선택은 마이크로컴퓨터이다. 조사처리는 비교적 컴퓨터에 미숙한 사용자들에 의해 분산화된 방법으로 수행되고, 이것은 비교적 저렴한 가격의 사용자 친화적 환경을 필요로 한다. 뿐만 아니라, 가용할 수 있는 유익한 소프트웨어들도 충분히 있어야 한다.

마이크로컴퓨터의 도입은 신중한 사고없이 이루어져서는 안된다. 마이크로컴퓨터의 활용이 비록 효율적인 정보처리에 새로운 방식을 열게 될 것으로 보이긴 하지만, 새로운 문제의 근원이 될 수도 있다. 만일, 모든

조직단위가 각자의 컴퓨터 유형을 자유롭게 선택하고 구매할 수 있다면, 자동화의 인프라는 쉽게 제어불능 상태가 되어 혼란에 빠질 수 있다. 따라서, 여기에는 표준화에 관한 강력한 정책이 요구된다.

또한, 조직 내에서 마이크로컴퓨터를 사용하는 방법에 대해서는 많은 주의를 기울여야 한다. 언뜻 보면, 많은 독립형 마이크로컴퓨터의 분산배치가 간단한 해답일 것 같지만, 실제로 이것은 적절한 방법이 아니다. 여기에는 다음 문제발생 가능성이 있다.

- 로컬 하드디스크에서 플로피 디스크로의 (기밀) 데이터 복제가 너무 쉽다. 이것은 데이터 보안문제를 야기시킨다.
- 백업 또는 아키빙과 같은 작업은 주관부서 사용자들이 종종 게을리 한다. 이것은 귀중한 자료의 손실을 초래할 수 있다.
- 부서간 커뮤니케이션이(예, 데이터파일 공유) 플로피 디스크 교환으로만 가능하다. 이것은 비효율적이다.
- 고립된 많은 마이크로 컴퓨터를 가진 대규모 조직에서는 새로운 버전의 소프트웨어 패키지를 배포하기가 종종 어려운 일이다.

전부는 아니지만, 이러한 문제들의 많은 것은 마이크로컴퓨터를 근거리 통신망(LAN; local area network)에 연결하는 것으로 해결될 수 있다. 이것은 IT 부서가 중앙화된 방법으로 데이터를 백업 및 아키빙하는 것을 가능하게 한다. 뿐만 아니라, 버전관리와 소프트웨어 업데이트도 훨씬 쉽다. LAN상에서 새로운 소프트웨어를 배포 및 설치하는 것은 매우 빠르게 진행되고, 한 번의 명령으로 모든 파일서버에 새로운 버전을 업로드시킬 수도 있다.

새로운 자동화 인프라의 효율적인 사용은 주관부서의 소프트웨어 사용에 관한 규칙을 필요로 한다. 만일, 이러한 부서들에게 통합관리시스템에 들어가는 소프트웨어를 자유롭게 선택할 수 있게 한다면, 작업불능의 상황이 연출될 수도 있다. 이 시스템이 비록 몇 가지 문제는 해결할 수 있다고 하나, 많은 비효율적이고, 비효과적인 상황은 그대로 남을 것이다. 즉, 서로 다른 소프트웨어가 동일한 작업에 사용되거나, 동일한 소프트웨어의 서로 다른 버전이 동시에 사용되어 결국 데이터 파일의 호환이 불가능한 상황에 이를 수 있는 것이다. 또한, 각기 다른 부서의 사용자들은 더 이상 동일한 "언어"로 의사소통할 수 없고, 이에 따라 교육 및 지원에 대한 필요성이 엄청나게 증대할 수 있다. 그리고, 여러 패키지를 구매하는 것은 한 패키지의 사이트 라이센스에 비해 훨씬 많은 비용이 드는 것은 당연하다.

통합관리시스템으로 고품질의 통계를 효율적으로 생산하는 데는 자동화 도구에 관한 표준화 정책이 요구되는 것은 분명하다. 표준화의 주요 장점은 다음과 같이 요약할 수 있다.

- 교육에 대한 투자에서 보다 많은 환원이익
- 보다 폭넓은 지식과 경험의 교류
- 결과적으로, 지원 및 유지관리에 보다 적은 자원 소요
- 하드웨어, 소프트웨어, 데이터의 보다 나은 상호교환성
- 조사시스템에 관한 보다 적은 개발비용
- 새로운 하드웨어 / 소프트웨어 선정시 보다 적은 문제발생
- 하드웨어 / 소프트웨어의 대량 구매에 따른 할인

표준화의 올바른 선택은 매우 중요하고 그 영향도 크다. 선택은 앞으로 응용소프트웨어 도구가 대규모의 컨버젼 없이 다른 인프라에 이식될 수 있는 가능성을 고려해야 한다. 표준화 정책의 한 가지 결과는 유동성 감소일 수 있다. 모든 변화는 많은 양의 잉여업무를 발생시킬 수 있고, 따라서 잘못된 선택은 심각한 결과를 초래할 수도 있다. 그러므로, 소프트웨어 표준을 너무 자주 바꾸는 것은 현명한 일이 아니며, 기존의 표준을 새로운 버전으로 바꾸는 것 또한 자주해서는 안된다.

통합의 개념은 일반적으로 조사과정의 효율성과 품질을 개선하기 위한 수단으로 받아들여 진다. 통합관리시스템은 궁극적으로 이러한 개념을 구현하는 것이다. 사용자 측면에서의 장점은 조사를 정의하고, 처리하는데 한 가지 "언어"를 사용하는 것이고, 그것은 메타데이터 정의언어라 한다. 메타데이터 정의는 조사설계단계에서 생성되며, 장치가 읽을 수 있는 형식으로 저장된다. 관리시스템의 모든 구성성분은 이 메타데이터 정의로부터 조사에 관한 정보를 검색한다. 필요할 경우, 관리시스템은 메타데이터를 조사과정에 필요한 형식으로 변환할 수도 있다. 관리시스템은 또한, 데이터 저장형식에 관한 모든 기술적인 측면을 사용자에게 숨긴다. 따라서, 사용자들은 데이터를 어떻게 저장할 것인지만(예, 관계형 데이터베이스 시스템) 지정하면 되고, 기술적인 상세는 모두 관리시스템에서 알아서 한다.

집중, 표준화, 통합과 같은 개념의 도입은 업무과정의 재설계를 의미한다. 이것은 조직 자체와 그 속에서 근무하는 사람들 모두에 영향을 미친다. 그러므로, 통계기관의 총괄적인 관리가 이러한 변화를 뒷받침하는 것이 중요하다 하겠다. 경영진은 조사처리의 새로운 접근방식을 허가하고, 유도해야 할 뿐 아니라, 직원들이 재설계가 가져다 줄 급격한 변화를 수용하도록 설득해야 한다.

집중, 표준화, 통합의 개념은 매우 매력적이다. 하지만, 조직내에서 이러한 개념의 구현을 방해할 수 있는 실무적인 문제들은 인식해야 한다. 그 중 한 가지 문제는 일반적으로 사람들은 자신의 업무방식을 바꾸길 꺼려한다는 것이다. 이들은 현재의 시스템과 절차를 이용하고 있으며, 이를 바꿀 이유가 없다고 생각한다. 그리고, 또 다른 잠재적인 문제는 맞춤식 시스템을 표준 패키지로 바꿔야 한다는 것이다. 맞춤식 시스템은 즉각적인 응용에는 언제나 조정이 쉽다. 그러므로, 표준 패키지로의 변경은 시대에 역행하는 것으로 보이게 된다. 물론, 이러한 관점은 맞춤식 시스템의 개발 및 유지보수 비용을 간과하고 있을 뿐만 아니라, 표준 시스템 사용의 범조직적 장점의 가치 또한 깨닫지 못하는 것이다.

9.7 품질에 대한 효과

일부 통계기관들이 다양한 각도에서 최소한 개념의 일부를 깨닫고 있긴 하지만, 아직까지 완벽하게 구현된 통합관리시스템의 사례는 없다. 그러므로, 통합관리시스템이 조사의 품질에 미치는 효과에 대해 확고한 그림을 제시하기란 어려운 일이다.

조사비용에 관한 정보는 통합관리시스템이 전체 조사비용을 절감할 수 있음을 시사하는 것으로 보인다. The Subcommittee on Data Editing in Federal Statistical Agencies(1990)의 보고서를 보면, 종래의 조사과정에서 편집작업이 상당한 비용을 소비하고 있다고 되어 있다. 많은 조사에서 편집은 전체 비용의 최소 20%를 발생시키며 특히, 경제조사에서 있어서는 40%를 상회하는 것으로 나타났다. Blom(1994)은 스웨덴의 경우에서 비슷한 결과를 제시하였는데, 이를 보면, 조사비용의 40%가 데이터 인식에 할애되고, 20%가 편집과 코딩에 할애되는 것으로 되어 있다.

비용절감으로의 중요한 변화는 컴퓨터를 이용한 면접기법의 도입일 수 있다. Clayton and Werking(1994)은 이 변화를 노동집약적 활동에서 자본집약적 기술로의 전이로 특징짓는다. 비노동비용은 감소하는 반면, 노동비용은 계속해 증가하기 때문에 이러한 변화는 조사비용의 감소에 긍정적인 효과가 있다. Weeks(1992)와 Nicholls and Groves(1986)와 같은 일부 학자들은 CAPI와 CATI 조사의 초기개발비용은 종래 조사의 초기개발비용보다 결코 적지 않다고 지적한다. 이것은 주로 하드웨어와 소프트웨어 및 교육에 대한 높은 투자가 원인이다. 반면, Martin(1993)은 초기투자의 회수기간은 ONS의 경우 2년에 불과했음을 강조한다. 많은 비용절감은 자료입력과 코딩 및 일괄편집을 없앰으로써 성취된 것이었다. Nicholls and Groves(1986)는 컴퓨터를 이용한

면접은 지속적이고 확장적인 조사에 특히 비용 효과적일 수 있다고 주장한다. Weeks(1992)는 조사도구의 재사용과 대규모의 복잡한 조사에서의 비용절감을 강조한다. 이 모두는 조사도구의 용이한 개발, 유지관리, 재사용이 허용되는 환경은 전체 조사비용을 실질적으로 절감시킨다는 사실을 의미한다. 통합관리시스템은 이러한 환경을 제공하고자 하는 시도하는 것이다.

네덜란드 통계청은 이미 통합관리시스템을 상당히 구현해 놓고 있다. 이는 지난 10년간 집중, 표준화, 통합에 대한 개념을 인식하고 있었다는 것이다. 메인프레임에서 마이크로컴퓨터로의 변화와 하드웨어, 소프트웨어의 표준화 정책은 조사비용의 실질적인 절감을 가능하게 했다. 예를 들어, 자동화 인프라의 비용은 1987년부터 1990년까지 약 20%가 절감되었다. 현재도 그 비용은 계속해 줄어들고 있다. 새로운 환경은 조직이 상당한 예산 및 직원감축에 직면해 있을 당시, 동일한 생산레벨을 유지할 수 있게 해 주었다. 상세한 사항은 Bethlehem(1992)을 참조한다.

자료의 정확성에 있어서는 통합관리시스템의 장점에 관한 확실한 증거를 얻기가 더 어렵다. 컴퓨터를 이용한 면접기법의 장점은 지나 10년을 거치면서 꽤 분명해 지고 있다. Weeks(1992)가 여러 문헌들을 고찰한 결과를 보면, 임계오차와 경로오차, "모른다"는 응답과 무응답이 상당히 감소되었음을 알 수 있다. 종래의 조사에 관한 몇몇 연구들은 오차의 90%가 응답기록을 잘못한 것에 기인하고, 그러한 오차는 컴퓨터를 이용한 면접에서는 불가능하다고 강조한다.

통합관리시스템의 사용으로 자료의 정확성이 증대되는 데는 다음 두 가지의 이유가 있다.

- 조사과정의 한 단계에서 다음 단계로의 데이터 및 메타데이터의 전환은 독특한 메타데이터 정의를 사용하여 자동적으로 구현되기 때문에, 각 단계간 일관성이 보장된다. 또한, 여러 소프트웨어 인자들의 환경설정 및 관리파일의 수작업으로 준비함으로써 발생가능성이 있는 오차를 피할 수 있다.
- 통합관리시스템은 조사도구의 재사용을 촉진한다. 특히, 조사정의언어의 모듈화는 여러 조사에서 사용될 수 있는 각각의 조사정의를 개발, 시험 및 유지 관리할 수 있게 해 준다.

자료품질의 마지막 측면은 조사결과의 적시성이다. 대부분의 저자들은 집중, 표준화, 통합이 조사결과의 적시성을 향상시킨다는 데 동의한다. 어떤 경우는

전체 처리시간의 50%가 줄어드는 것으로 확인되었다. Bethlehem (1992), Blom(1994), Clayton and Working(1994), Tozer and Jaensch (1994) 참조 바란다.

통합관리시스템의 활용은 적시성에서도 상당한 개선을 이끈다. 새로운 조사도구의 개발은 표준화된 조사정의 모듈이 사용될 때 더욱 빨라진다. 아울러, 표준화된 환경에서 작업할 때는 면접원을 교육시키는데도 시간단축이 요구된다. 이것은 조사과정에 참여한 기타 모든 사람에게도 마찬가지로 똑같이 적용되는 사실이다. 결국, 하나의 처리단계에서 다른 단계로의 원만한 전이는 전체처리시간을 단축시킨다.

Bethlehem(1992)은 주관부서에 근무하는 사람들에게 뚜렷한 장점이 있다고 보고했다. 이런 변화는 각자의 업무 범위와 다양성을 증대시켰다. 이들은 자신의 업무를 구성하는데 자율성과 책임이 높아진 것을 매우 기쁘게 여겼다. 이것은 전문화의 감소, 보다 광범위한 기술의 습득 그리고, 목적에 따라 시스템을 채택하고, 개조할 수 있는 가능성 등을 이끌었다. 이러한 장점들은 보다 긴 시간의 교육에 대한 필요성을 점점 부각시켰고, 심지어는 컴퓨터로 작업하는 데 대한 그 동안의 오랜 거부감을 극복시켰다.

9.8 미래

통합관리시스템은 통계기관들이 저마다의 생산공정의 능률성을 향상시킬 수 있다. 공통의 데이터 언어와 한 프로그램에서 다른 프로그램으로 데이터 및 메타데이터 전이의 가능성을 통해 조사과정의 효율성이 증대된다. 이것은 적시성을 향상시키고, 생산된 통계정보의 품질을 개선할 것이다.

통합관리시스템은 종래의 조사를 보다 나은 방법으로 처리할 수 있을 뿐만 아니라, 복잡한 조사를 보다 쉽게 수행할 수도 있다. 그러므로, 이러한 시스템들은 앞으로 보다 많고, 보다 나은 통계정보를 약속한다.

통합관리시스템은 통계기관의 업무를 개선할 것이지만, 그렇다고 만병통치약을 말하는 것은 아니다. 관리시스템들은 귀중한 해결방안을 제시해 줄 수 있다. 그렇다면, 두 가지의 다소 연관된 문제점을 간략히 언급하고, 이러한 문제를 처리하는데 있어 관리시스템이 어떤 역할을 하는 지를 제시해 보도록 하겠다.

첫번째 문제는 통계정보의 배포다. 최근 발전한 데이터통신은 전자형식의 통계정보에 대한 수요를 증대시켰다. 전자고속도로(인터넷)와 같은 매체는 이런 방식으로 정보에 접근할 수 있게 해 주는데 매우 적절하다. 이 새로운 전자배포식 접근방식은 접근가능한 모든 데이터와 메타데이터를 포함하고 있는 통합된 출력 데이터베이스의 구현을 필요로 한다. 통합관리시스템의 메타데이터 성분은 시스템의 출력적인 측면에서 필요한 특성을 포함하도록 확장될 수 있다. 고려할 필요가 있는 측면으로는 예를 들면, 집계작업과 정보유출방지 등이 있다.

두번째 문제는 통계의 통합이다. 대부분의 통계기관들은 여러 조사에서 얻은 통계정보를 출판한다. 이 때 조사정의에서의 비일관성은 각기 다른 조사의 결과들을 통합하기 매우 어렵게 만든다. 특히, 용이한 전자적 통계정보 접근은 이러한 문제를 더욱 부각시킨다. 이런 경우, 통합관리시스템은 조사정의에 표준화된 환경을 제공하는데 있어 중요한 역할을 할 수 있다. 특히, 모듈화된 조사정의 언어와 같은 특성들이 이러한 측면을 도울 수 있다. 모듈은 표준적인 질문집합을 개발할 수 있게 해주고, 이러한 질문집합들은 신중한 검증을 거친 후 승인된 설문모듈 라이브러리에 포함될 수 있다. 또한, 모든 설문설계자들에게 이러한 표준집합을 활용할 수 있게 해 줌으로써, 기존에 완전히 새로운 질문집합을 개발하던 것을 대신할 수 있어 이들에게는 매우 매력적으로 다가설 것이다.

제 4 장

품 질 평 가 및 관 리

제 4 장 품질 평가 및 관리

제 10 절 조사측정 및 프로세스개선: 개념과 통합

Cathryn S. Dippo
美노동 통계청

10.1 서문

조사측정(survey measurement)과 프로세스개선(process improvement)은 공통점이 매우 많다. 이론적으로는 둘 다 Fisher의 철학을 능가하는 것으로, 정밀도의 척도인 무작위표본과 일관된 추정량 및 신뢰구간을 하나로 묶은 Neyman의 임의화(randomization) 철학을 따르고, 실제로는 통계관리의 생산공정에 종속되어 있다. 그러나, 공통의 근원과 가정에도 불구하고, 조사측정과 프로세스개선의 개념은 엄격한 의미에서는 통합되어 있지 않다. 즉, 통계측정의 과학적 원리가 조사프로세스의 자연발생적 부분이 아닌 것이다.

최근 몇 년간, 조사측정작업을 맡은 많은 기관들은 자신들의 작업에 품질관리의 형식을 도입하였다. 제 4 절에 있는 여러 절들에는 이와 같이 품질개선의 관행을 조사작업(survey operation)에 통합시키고 있는 좋은 선례들이 소개되어 있다. 이것은 프로세스개선을 실천 또는 운용적 차원에서 조사측정에 통합하는 방향에 있어서는 중요한 단계이긴 하지만, 이론적 차원에서 개념이나 철학을 통합하는 것은 아니다. 따라서, 이와 같은 통합의 관점에서 품질측정과 메타데이터 생성은 초기의 개념논의에서부터 프로세스 상품의 배포에 이르기까지 조사측정의 전과정에 걸쳐 조사측도 또는 모집단 추정값의 산출만큼이나 많은 관심을 받곤 했다.

본 절의 목표는 조사측정과 프로세스개선의 개념적 통합에 대한 제고를 공유하는 것이다. 동일한 철학적 근원을 가진 두 개념이 지금껏 통합되지 않았던 이유를 이해하는 것은 어쩌면 거꾸로 바람직한 통합을 촉진시키고, 궁극적으로는 새로운 이론적 배경이 실제 조사작업에 표현되도록 도와 줄 통찰력을 불러일으킬 것이다. 이에 따라, 본 절에는 "과거가 무엇인지가 서막이다…,"고 믿는 많은 역사학자들의 깨우침에 따라, 과거에 대한 고찰이 들어 있다. 나는 또한, 단어와 용어의 의미가 이해의 바탕을 제공하고, 이해없이는 개념을 파악하거나 정의할 수 없다는 의미론(semantics)에 많은 비중을 둔다. 나는 우리 앞에 놓인 수수께끼의 답을 알지 못한다. 나의 목적은 오로지 조사측정 "속에" 프로세스개선을 "통합" 또는 "구축"해 나가야 될 필요성에 대한 인식을 증대시키는 것뿐이다.

10.2절에서 이 두 가지 견지로부터 조사측정과 프로세스개선 및 품질측정이라는 용어에 대해 간단히 살펴본 후, 10.3절과 10.4절에서는 또 다시 이 두 가지 견지로부터 오차라는 용어의 의미를 짚어볼 것이다. 오차라는 단어의 용도와 품질과의 관계를 이해하는 것은 조사측정과 프로세스개선의 이론적 구조를 파악하는데 결정적인 역할을 한다. 그 다음엔 이 두 가지 견지로부터 조사측정의 실제 프로세스에서 측정오차라는 것에 대해 고찰한 후, 마지막 절에서는 다시 통합의 개념으로 돌아올 것이다. 많은 쟁점들은 조사측도 산출을 위한 품질의 측도와 측정프로세스를 위한 품질의 측도 사이의 구별에서 비롯된다. 이러한 구분을 이해하는 것이 프로세스품질의 유용한 측도 즉, 프로세스를 개선하는데 이용될 수 있는 측도를 정의하는 첫번째 단계다. 이러한 측도들은 고객이 통계상품의 사용적합성을 결정하는 데는 도움이 되지만, 반면 여러 가지 방면에서 그 속에 들어 있는 많은 정보를 빠뜨리게 된다. 이 측도들이 전달하지 못하는 것은 통계적 의미를 제공하는 정보 즉, 프로세스와 통계상품 양측면에 관한 데이터와 메타데이터다.

10.2 몇 가지 용어들의 의미와 용도

조사측정(Survey Measurement)은 자주 사용되는 용어로, 그 의미와 해석은 사용자와 상황에 따라 매우 다양하다. 예를 들어 어떤 경우에는 모집단에 속해 있는 몇몇 단위로부터 관측된 자료를 바탕으로 한 인구량 척도에서처럼 그 용도에서 비롯되는 표본의 의미를 함축하기도 한다. 이런 식으로 하면, 상시인구조사(Current Population Survey)에서 수집된 자료로부터 추정된 미국 중심도시에 거주하는 편부모가정의 비율이 조사척도(survey measure)가 된다. 또, 혹자는 이러한 인구량에 대해 10년 단위 센서스를 바탕으로 한 추정값도 조사 척도가 된다고 말한다. 이 집단의 경우에는 직접관측 형태보다는 자계식 설문의 관측방법이 은연중에 추정 값을 조사 척도로 정의 내린다. 반면, 토지측량인(land surveyor)은 질문보다는 기계장비를 이용한 자신의 관측결과를 조사척도로 정의하기 쉽다.

이 모든 의미론적 해석의 공통된 맥락은 측정하는 행위이다. Random House College Dictionary (1980, p.829)를 보면, 척도(measure)의 첫번째 정의가 "특히, 표준과의 비교를 통해 어떤 것의 범위와 치수, 수량 등을 확인하는 행동 또는 프로세스"로 되어 있다. 이것은 척도를 "어떤 목적을 향한 일련의 체계적인 행동"(Random House, 1980, p.1055)으로 정의되는 하나의 프로세스(process)로 정의하고 있으며, 이 프로세스를 통해 우리는 조사측정-프로세스개선의 통합된 개념의 핵심에 접근할 수 있다. 다시 말해, 조사기반의 통계를 산출하는 것과

관련된 모든 작업 및 업무들은 반드시 통계적 측정의 원칙이 적용되어야 하는 하나의 프로세스로 바라봐야 한다.

통계의 산출을 측정프로세스로 고려하는 것은 이제 더 이상 새로운 관점이 아니다. 하지만, 이러한 관점에 대해 언급되어 있는 통계문헌은 그다지 많지 않다. Eisenhart(1952, 1963)는 기계장치의 보정시스템과 같은 의미에서 측정에 대해 논하였다. Bailar(1983)는 자신의 통계자료의 광범위한 논의에 표본조사와 센서스를 이용하는 인구특성의 측정을 포함시켰다. Bailar는 품질을 정의하고, 측정프로세스의 통계적 관리의 필요성을 촉구하는데 역점을 두었다.

그러나, 조사측정을 일련의 행위로만 정의하는 것은 뭔가 충분하지 않다. 측도를 유용하게 하려면, 반드시 단순한 수치이상의 것이어야 한다. 그리고, 반드시 전후관계(context)가 있어야 한다. 어떤 개체의 속성을 수치로 변환하는데 사용되는 맵핑에 대해 약간이 지식이 없다면, 수치는 아무런 의미가 없다. 예를 들어, John은 4, Mary는 6이라는 정보는 John은 생후 4개월, Mary는 생후 6일일 수 있는 것처럼 누가 더 나이가 많으냐는 질문에 대한 답으로 적절치 않다. 비슷한 조사측정으로, 종업원 수의 증가가 어떤 한 조사에서는 25만 명으로 측정되었고, 다른 조사에서는 15만 명으로 측정되었다면, 이것도 불충분한 정보다. 이러한 두 가지 통계치를 적절히 해석하려면, 사용자에게 메타데이터나 데이터에 관한 데이터가 필요하다(데이터, 메타데이터, 정보 사이의 관계에 대한 논의는 Hand, 1993 참조). 이와 같이, 조사측정이라는 용어에 의미를 부여하기 위해서는 반드시 프로세스와 그 프로세스에 관한 메타데이터(input, output 포함) 두 가지 모두에 관심이 주어져야 한다.

반면, 프로세스개선(process improvement)은 품질철학과 관련하여 사용되는 많은 용어들 중 하나일 뿐이다. 종합품질관리, 지속적인 품질 또는 프로세스개선 등, Deming의 14가지 포인트는 모두 고객에 대한 포커스를 통해 재화와 용역의 품질을 개선하는데 목적을 두고 있다. 따라서, 어떤 철학을 탐구하던 간에 성공의 열쇠는 고객의 역할, 프로세스 그리고, 통계적 품질측도로 정의된다.

조사측정의 맥락에서 품질(quality)이란 최근 몇 년간 상당히 많은 관심을 받아온 용어이다. 혹자는 품질의 정의에 대해 다루었고(예, Norwood, 1987), 또한, 혹자들은 조사기관의 품질관리를 고찰하였다(예, Fesco, 1989; Colledge and March, 1993). 그리고, 지금도 많은 사람들이 조사품질의 지표에 대해 논하고 있다(예, Groves and Tortora, 1991). 품질을 추구함에 있어서는 표준을 바탕으로 하거나(Freedman, 1990), 지속적인 품질개선을 바탕으로 하는(Linacre, 1991) 등,

다양한 접근방식이 시도되고, 논의되어 왔다. 그 중, Colledge and March(제22장)는 몇몇 국가통계기관들을 대상으로 품질정책과 표준, 지침 및 권고관행을 검토하였고, Morganstein and Marker (제21장)는 지속적인 품질개선과 품질을 개선함에 있어 현재 최상의 방법들의 역할에 중점에 두었다. 그리고, 이후 이어지는 절에는 여러 가지 조사측정 상황에서 품질개선이 어떻게 성취되었는가에 관한 몇 가지 좋은 예들이 나와 있다.

조사척도에서의 품질의 개념은 하나의 관념(notion)이다. 이것은 모호하고, 불명확하다. 품질의 척도들은 조사측정과 프로세스개선의 두 개념 모두에 내재되어 있다. 이 두 개념에는 또한, 통계의 품질은 반드시 사용자와 상황에 따라 다른 효용성으로 정의되어야 한다는 생각도 포함되어 있다. 두 가지 모두 오차측정에 초점을 맞춘다. 하지만, 이와 동시에 오차의 의미와 오차의 측도 및 오차척도의 활용은 서로 극명하게 다르다.

조사측정 문헌들을 보면, 조사품질의 척도로 평균제곱오차(mean squared error)가 가장 많이 사용되었다. 여기서의 오차란 수학적 기대치라는 의미로 정의된다. 오차라는 용어는 프로세스에서의 착오(mistake)를 의미하는데도 사용된다. 조사연구자들은 응답자 협조, 질문, 자료입력, 문항 코딩과 같은 조사측정의 특정한 하위프로세스에서 오차의 가능성을 오랫동안 인식해 왔으며, 정의된 절차의 이행을 측정 및 감시하기 위해 품질관리절차를 개정하였다. 이에 따라, 면접원들이 질문을 읽을 때 실질적인 의미를 바꾸는 비율, 응답률, 입력오차율 그리고, 문항 코딩 오차와 같은 프로세스오차 척도들이 조사가 진행중일 때 정기적으로 산출된다.

조사측정의 맥락에서 오차라는 용어의 사용과 관련한 의미론적 문제를 해결하기 위해, 대개는 총계가 아닌 평균제곱오차의 성분 중 하나가 측정오차로 명명된다. Groves(1991)는 총계오차(total error)는 관측오차와 비관측오차 모두의 작용임을 지적하면서, 통계학과 심리학 분야에 걸친 관측오차(observation error)로서의 측정오차의 의미에 대해 논하였다. 그 결과, 측정오차는 조사척도의 전반적인 품질과 관련이 없거나 혹은, 조사측정에서의 오차는 측정오차와 동일한 것이 아니라는 결론을 내렸다.

보다 중요한 사실은 대부분의 프로세스오차 척도들이 평균제곱오차와 직접적인 작용관계가 없다는 것이다. 예를 들어, 응답률의 증가는 반드시 평균제곱오차의 감소로 해석할 수는 없다. 마찬가지, 평균제곱오차의 감소도 반드시 특정한 프로세스오차 측도의 개선을 의미하는 것은 아니다.

10.3 조사품질의 척도로서 평균제곱오차

조사표본의 기본적인 법칙은 적은 것으로 보다 많은 일을 하는 것이다. "표본조사는 단기간에 적은 비용으로 똑 같은 관측절차나 측정절차를 이용하여 전체 적용범위로부터 얻을 수 있는 추정값을 제공한다. 잘 설계되고 관리된 표본조사는 전체인구조사에서 달성될 수 있는 것보다 높은 품질을 성취하게 해 준다."(Hansen and Madow, 1976, p.76). 하지만, 1930년대 중반, 미통계청에서 유한모집단에 무작위표본의 적용가능성에 대해 조사한 결과, 이러한 진술을 뒷받침할 만한 이론적 근거나 실증적 증거가 거의 없었다. 특히, 어떤 표본이 보다 고품질을 제공한 것인지 파악할 만한 표준적인 품질척도는 전혀 없었다.

이 초창기 조사에서는 확률표본에서 제외된 것으로 추정되는 숫자와 자발적으로 우편 접수한 숫자를 단순히 비교하는 것으로 충분하였다. 그러나, 조사표본이 개념적으로 보다 복잡한 사회경제적 특성들을 측정하는데 사용되기 시작함에 따라, 보다 정교한 조사품질측도와 조사비용이 점점 늘어나게 되었다.

표본 변이성에 의해 정의되는 조사품질은 Neyman(1934)에 의해 조사비용에 처음으로 연계되었다. 제약조건에 따라(즉, 고정된 전체표본크기) 분산 추정량을 최소화하기 위한 Neyman의 최적할당의 개념은 오늘날에도 지속 개발중인 보다 효율적인 표본설계와 저분산 추정량 연구에 초석이 되었다. 그러나, 현상황은 Hansen and Hurwitz(1943, p.335)가 "..실무에 적용가능한 이론으로 "최량"의 표본시스템을 보여줄 만한 이론은 아직까지 개발되지 않았다.."고 말한 1943년과 똑 같은 실정이다. 실제로, Godambe(1955)는 전체 모집단을 추정하기 위한 최량의 선형 비편향 추정량은 없음을 증명하였다. 즉, 누구나 반드시 가정을 정하고, 실현 가능한 집합에 대한 선택항목의 수를 줄이며, 평가기준을 정해야 한다.

일찍이, Hansen and Hurwitz(1943, p.337)는 "비선형적이고 일관되지만, 평균제곱오차가 보다 적은 추정치"를 도출하기 위해서는 양질의 추정치에 기준이 오로지 "최량선형비편향"만을 포함하는 것으로부터 바뀌어야 한다고 주장하였다. 비록, 이들이 제곱평균오차와 분산을 상호 엇갈리게 사용하긴 했지만, 이들을 포함하여 그 누구도 표본오차가 평균제곱오차와 동일하다거나 혹은, 비표본오차가 존재하지 않는다고 생각한 사람은 없었다. 1940년도 초, Stephan과 Deming 그리고, Hansen(Stephan et al., 1940)은 1940년 인구센서스에서 가구 및 가구에 속한 구성원들에 대한 목록화 절차와 관련된 편향을 없애거나 줄이기 위한 논의를 하였다. 그리고, Hansen, Hurwitz, Pritzker(1967)의 비표본오차의 관리 및 측정에 대한 관심은 결국 평균제곱오차의 개념을 이끌어 내게 되었고, 앞으로 이것을 HHP 모델이라 부르겠다.

HHP 모델에는 조사추정값(x)의 품질을 평가하는 세 가지 요소인 요구사항(requirements), 규격(specifications), 작업(operations)이 있다. 요구사항은 조사측도의 이상적인 목표(Z)에 반영된다. 종종 일반적인 사회 또는 경제적 개념에 반영되는 이러한 목표들은 반드시 상세한 규격이나 근로정의(working definition)를 통하여 정의되는 목표집합(Y)으로 정제되거나 전이되어야 한다. 그런 다음, 규격은 반복 적용되어 x의 기대값(X)을 반영하는 조사작업으로 반드시 전이되어야 한다.

예를 들어, 경제활동인구의 측정을 가정해 보자. 개념적으로, 개인의 경제활동분류(Z)는(즉, 경제활동에 종사하는지, 안 하는지) 매월 12일이 속한 한 주 동안의 노동시장활동을 기초로 한다. 취업의 근로정의(working definition)(또는 규격) (Y)는 급여를 받고 최소 1시간 이상을 일했거나, 자신의 사업이나 전문직 또는 농장을 최소 1시간 이상 운영했거나, 혹은 가족이 운영하는 농장이나 사업체에서 무급여로 최소 15시간을 일했거나, 일하지는 않았지만 특정한 사유로 임시 결근한 직업이나 사업체를 가진 사람들이다. 반면, 취업의 작업정의(operating definition)(X)는 "지난주에 급여를 받는 어떤 일을 했습니까?"라는 질문에 "예"라고 응답하거나, 혹은 이 질문에 "아니오"라고 응답한 사람 중에 "지난주에 임시 결근했던 것을 포함하여 풀타임 또는 파트타임의 어떤 직업이 있었습니까?"라는 질문에 "예"라고 응답하는 등의 사람을 말한다. 더욱이, 작업정의는 일단의 면접원 교육자료와 면접원 매뉴얼, 이러한 지침서를 보충하고 세분화하는 여러 가지 문서 그리고, 기록된 자료의 모순을 해결하기 위한 컴퓨터 편집과 대치 등에 의해 유지된다.

HHP 모델의 이 세 가지 요소들은 다음과 같이 평균제곱오차를 정의하는데 사용된다.

$$MSE(x) = E(x - Z)^2$$
$$= E[(x - X) + (X - Y) + (Y - Z)]^2$$
$$= E[(x - X)^2 + (X - Y)^2 + (Y - Z)^2 + 2(X - Y)(Y - Z)$$

여기서, $E(x - Z)^2$는 총분산, 표본, 비표본이고, $(X - Y)$는 규격과 비교했을 때의 실제 조사작업의 편향이며, $(Y - Z)$는 요구사항과 비교했을 때의 조사규격의 편향(적절성)이고, $2(X - Y)(Y - Z)$는 두 편향 간의 상호작용을 말한다.

HHP는 정확도(accuracy)를 정의된 목표와 실제 성과의 작용으로 정의하다. 다시 말해, 정확도는 작업(operations)이 반복되는 조사에서 규격과 얼마나 일치하고, 총분산 및 제1 편향과 얼마나 똑 같은가라는 의미로 정의된다. 이것은 조사측정에서 용어들이 일반적인 이해와 전적으로 동일하지 않은 의미를 전달하는 또 다른 경우이다. 즉, Random House College Dictionary (1980, p.10)에는 정확도가 "참되거나, 올바르거나 혹은 정확한 상태 내지는 품질, precision, exactness, 또는 correctness."로 정의되어 있다.

적절성(Relevance)은 현재의 논의에서 특히 중요하다. 고객과 조사척도의 활용은 품질 척도에 포함된다는 것도 바로 HHP 모델에서의 적절성이라는 용어를 통해서다. Morgenstern(1963)은 정확한 통계가 주어지지만, 직접적인 목적과 관련이 없는 경우 그가 경제 모델링에서 자주 직면하였던 "허위의 정확도(specious accuracy)"가 발생한다고 지적하고 있다. 또, Dalenius (1985)는 특히, 공적인 통계와 그 목적성의 맥락에서 연관성 있는(relevant) 통계의 관념에 대해 논하고 있다. 적절성을 평가하거나 특정하는 방법들은 조사측정 문헌에서 거의 찾아보기 힘들다. Moore and Shiskin(1967)은 요구사항에 가중치를 부여하고, 규격에 평점을 매기는 것을 기초로 하는 방법론을 이용하여 경제적 지표로서 조사 추정값의 적절성을 평가하려고 노력하였다. 그리고, Dalenius(1977)는 HHP 모델의 맥락에서 D_{ij} 행렬의 근사치를 구하는 논의를 했으며, 여기서 각각의 D_{ij}는 잠재적인 사용자 집단에 대해 i번째 정의된 목표 Y_i와 j번째의 이상적 목표 Z_j 간의 차를 반영한다. 그러나, 실제로는 이것이 보편적이지 않으며, 대부분의 저자들은 평균제곱오차를 단순히 분산에 편향제곱(bias squared)을 더한 것으로 정의하고, 요컨대 고객의 조사척도 활용을 무시하면서 논의되고 있는 적절성에 대해서는 고려하지 않는다.

우리는 전체 조사품질의 척도로서 이와 같이 축소된 평균제곱오차의 정의를 주로 이용해 왔는데, 이는 HHP 모델에서 참값(truth) 혹은 Z가 극히 드문 상황에서만 제대로 정의될 수 있으며, 또 실제로 이것으로는 총평균제곱오차의 HHP 정의를 측정하는 것이 일반적으로 불가능하기 때문이다. 그러나, 전반적인 조사품질의 측도에 고객의 요구를 반영하려 노력하고, Z가 고객에 의해 정의된 참값(truth)이라면, 추정값의 적절성은 결코 무시될 수 없는 사안이다. 그리고, (Y - Z)가 측정될 수 없는 경우라면, 평균제곱오차의 축소된 정의의 추정값으로는 불충분하고, 통계의 적절성에 관한 메타데이터를 제공하는 다른 방편이 반드시 제공되어야 한다.

10.4 조사측정 프로세스

이와 같이 명백한 조사품질의 정의로는 1980년대에 왜 그토록 조사품질을 정의하거나 설명하고자 하는 많은 활동들이 있었으며, 또 적은 것으로 보다 잘 할 수 있는 방법이 왜 명백하게 밝혀지지 않고 있나 하는 의구심을 품을 수 있다. 여기에 대한 답은 조사상품 또는 측도와 조사측정 프로세스 사이의 구분으로 되돌아 간다. 평균제곱오차나 어떤 분산 또는 편향성분을 추정하는 것과 연관된 모든 수학은 현재 관리하에 정의되어지고 재현가능한 프로세스를 가정한다.

하나의 프로세스로서 조사측정을 논의할 때면, 우리의 관심은 오차 즉, 보다 보편적으로는 착오(mistake)로 돌아간다. 물론, 두 가지 종류의 오차 HHP 용어에서는 규격(specification)에서의 착오와 작업(operation)에서의 착오를 서로 구분하는 것 또한 중요하다. HHP(1967, p.52)는 "유익한 조사설계는 알맞은 조사절차 규격과 적당한 작업관리를 통해 조사결과의 정확도에 대한 효과적인 관리를 필요로 한다"고 지적하였다. 또, Juran et al. (1979)은 설계의 품질과 부합(conformance)의 품질 간에 유사한 구분을 하고 있다. 오차가 착오를 지칭하는데 사용될 때면, 우리는 더 이상 반복되는 사건에서의 임의화 이론과 기대값을 사용하지 않는다. 우리는 지금 특정한 실현(realization)과 다른 것보다 하나의 표본을 선택하기 위한 일단의 규격 내지는 다른 사람의 응답기록보다는 면접원의 응답기록을 다루고 있다.

총 평균 제곱오차를 분배하는 방법에는 여러 가지가 있듯이, 조사측정에 관한 프로세스모델도 여러 유형이 있다. 하나의 프로세스로서 조사측정은 Hansen and Steinberg (1956)의 관점과 같이 다각적인 관점으로 바라볼 수 있다. 이들은 조사오차의 잠재적 근원으로 정의(개념, 문제, 연구대상 모집단)와 규격(단위의 적용범위, 조사문항연구, 코딩 절차) 및 실행에 오차를 포함시켰다. 이러한 관점은 물론, 조사척도와 HHP 모델의 광범위한 해석과 일맥 상통한다. 경우에 따라, 조사측정 프로세스는 설계 또는 방법에 주어진 관측프로세스로 바라볼 수도 있다. 그리고, 이 관점은 적절성을 무시하는 축소된 평균제곱오차 모델과 일치한다.

그림 10.1과 10.2에는 폭넓은 기반의 두 가지 조사프로세스모델이 나와 있다. 이 중 그림 10.1은 캐나다 통계청 모델이고,(Statistics Canada, 1987) 그림 10.2는 캐나다 통계청 모델을 기점으로 한 미국 노동통계청(BLS) 모델이다(BLS, 1994). 따라서, BLS 모델이 보다 최신 모델인 바, 여기서는 이 모델의 몇 가지 특징에 대해 논하도록 하겠다.

모델의 중심에는 고객이 있다. Deming(1982, p.225)이 지적했듯이, "고객은 생산라인의 가장 중요한 부분이다." 비록, 모든 조사나 센서스가 조사수행기관 외부에 있는 누군가의 요구 표출로부터 시작되긴 하지만, 그 기관의 각 구성원과 응답자들도 때로는 조사프로세스의 직접적인 상품에 대한 고객이 될 수 있다. 기관 외부인이든, 내부인이든, 이 모든 고객의 요구는 생산과정의 일부분으로 처리되어야 한다. 고객들은 종종 질적인 용어인 적절성, 적시성, 정확성, 포괄성 등의 품질기준을 표현한다.

BLS 모델에 개괄적으로 나와 있듯이, 조사측정의 실제 프로세스는 일단의 6가지 작업 개념화(conceptualization), 계획수립(planning), 설계(Design), 개발(development), 이행(implementation) 및 타당성검증(validation)으로 구성된다. 개념화는 측정의 목적(이유)과 요구사항(무엇을)을 정의하는데 역점을 둔다. 그리고, 측정프로세스에 필요한 자원(누구)과 시기(언제)의 정의는 계획수립의 초점이다. 설계는 개념화에서 정의된 목적과 요구사항이 계획수립 중에 설정된 시기와 자원의 범위 내에서 어떻게 충족될 것인지에 대한 규격을 정하는 작업이며, 개발은 설계를 이행하는데 필요한 도구들을 준비한다(예, 소프트웨어, 하드웨어, 지침서, 수집도구, 교육자료). 끝으로, 타당성검증 중에는 실제의 성과수준과 계획된 성과수준이 서로 비교된다. HHP 모델에서 이상적 목표는 개념화 과정에서 내 놓게 되고, 정의된 목표는 계획수립과 설계 및 개발 과정에서 상세히 정해진다. 그리고, 조사추정값의 기대값은 이행작업을 반복함으로써 이론적으로 정의되며, 타당성검증에는 조사추정값을 이상적 목표 및 정의된 목표와 비교하는 과정이 포함된다.

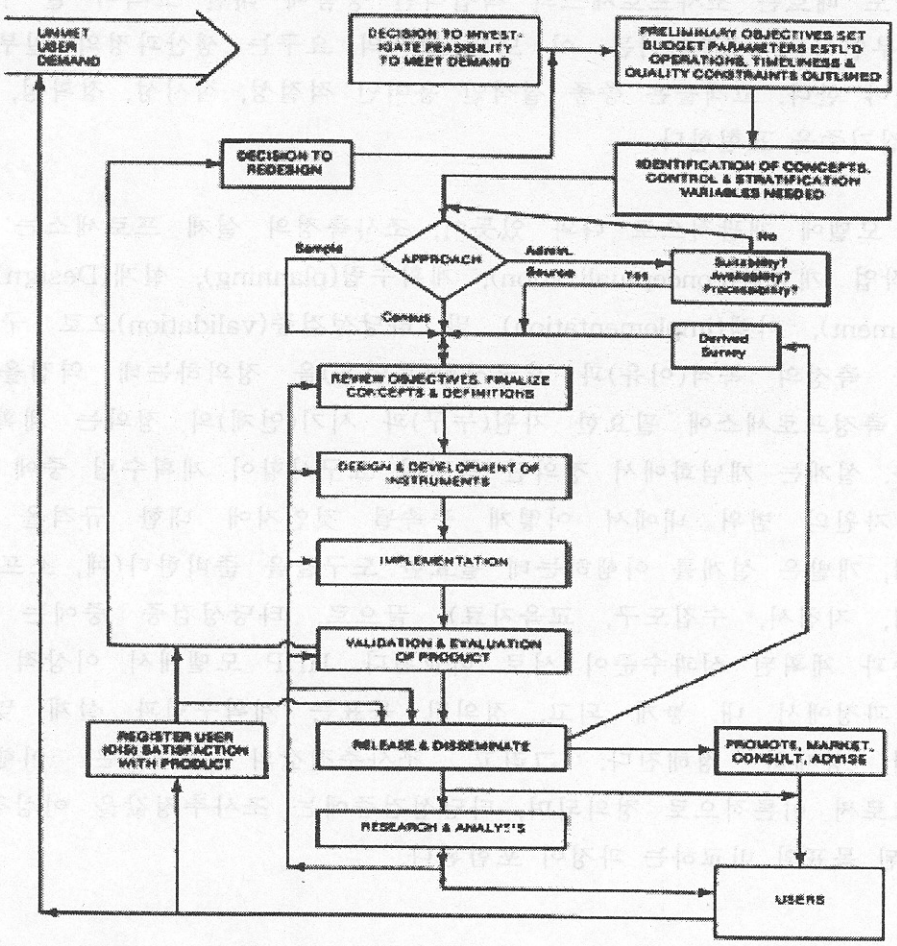


그림 10.1 캐나다 통계청의 일반적인 통계조사프로세스모델.
Statistics Canada, Quality Guidelines, April 1987, Page 3.

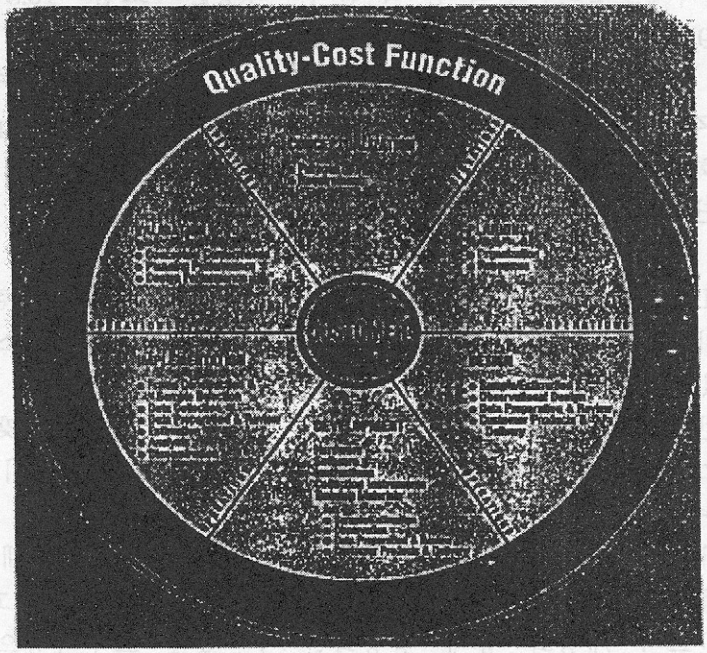


그림 10.2 미노동통계청 품질측정모델

BLS 조사측정프로세스모델에서는 작업들 간의 관계가 순차적이지 않다. 즉, 개념화를 끝내고, 계획수립으로 넘어가는 것이 아니며, 아직까지 개발에 어떤 기술을 활용할 수 없음을 발견한 후에야 설계를 준비해서는 안된다. 이 모델에는 작업들간에 계속 되풀이되는 흐름이 있다. 각각은 나머지 작업들의 지원을 받고, 고객의 요구를 만족시킬 수 있도록 수행된다.

프로세스모델(process model)은 또한 순환적이다. 작업(operation)들은 차례로 6가지의 작업을 가진 것처럼 보일 수 있는 여러 활동(activity)들로 구성된다. 예를 들어, 소비지출을 수집하기 위해 설문지를 계획하는 것에는 선택적인 설문형식을(예, 일지와 대본에 따른 면접) 개념화하고, 설계, 개발 등을 누가 할 것인지에 관한 계획을 수립하고, 선택하고, 설계하고(예, 하루하루 작성할 페이지 분량을 정하고, 구매유형별로 면접을 구분하는 등), 도구를 개발하며(예, 전산화된 자료수집도구를 위한 언어저작), 설계이행(BLAISE 나 CASES를 사용해서 자동화된 자기관리 질문지와 컴퓨터가 조력하는 개인용 인터뷰 도구를 만들어 냄). 두 가지 방식의 통계상품이 고객이 필요로 하는 정보를 산출하는지에 대한 타당성 검증의 과정들이 포함될 수 있다. 또한, 각 활동은 예컨대, 선택적 도구의 타당성을 검증하는데 타당성 기준의 개념화를 필요로 하는 등, 이와 유사하는 것으로 볼 수도 있다.

BLS 품질측정모델의 반복적, 순환적 특성은 그림 10.3에 나와 있는 Shewhart(1939)의 대량생산공정모델과 일치한다. 그는 대량생산공정의 3단계를(규격, 생산, 검사) 과학적 방법에서의 단계들로(가설수립, 실험수행, 가설검증) 생각할 것을 제안한다. 그가 지적한 바에 따르면(p.1), "경제적인 관리와 극대화된 품질평가를 달성하기 위해서는 통계적 이론과 기법들이 품질관리의 3단계에 모두 반드시 들어가야 한다."

고객이 BLS 모델의 조사측정프로세스의 초점이라면, 품질-비용 함수는 구속력이 있다. 그것은 조각들을 하나의 전체로 모으는 원소들의 집합으로 표현할 수 있다. 품질-비용 함수는 Rosander(1985)가 열거한 품질의 다양한 개념과 차원인 규격 부합성, 사용 적합성, 구매자 만족, 알맞은 가격에서의 수용가능성, 그리고 수명주기비용의 가치가 있는 수명 주기이익을 반영한다. 품질-비용 함수는 또한 자원이 제한적이라는 사실도 반영한다. 고객의 요구는 정해진 시간과 예산 등의 범위 내에서 처리되어야 하고, 통계상품은 적절성, 적시성, 정확성 등과 관련된 품질기준을 만족해야 한다. 그리고, 고객의 비정량적 품질기준은 반드시 개념화 중에 정량화할 수 있는 요구사항으로 변환되어야 한다. 프로세스상의 변화가 이와 관련된 비용만큼의 가치가 있는 지를 결정하는 것 또한 품질-비용 함수를 통해서 이루어진다.

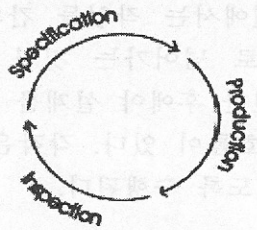


그림 10.3 Shewart의 프로세스모델

그러나, 품질-비용 함수는 거의 정량화할 수 없는 이상을 반영한다는 점에서 총 평균 제곱오차와 비슷하다. 이것은 HHP 모델에서 이상적 목표와 거의 동일한 방법으로 정의되는 선택적 품질기준의 정의인 근사치(approximation)와 대리값(proxy)의 사용을 필요로 한다. 게다가, 평균제곱오차의 경우에서처럼, practitioner들은 종종 계산하는 구성성분이나 부분적인 척도로 되돌아간다. 정리하자면, 한편으로는 규격이나 방법에 초점이 맞춰지고, 첫번째 의미 즉, 수학적 기대값의 측면에서 정의되는 오차 측도인 응답분산, 응답편향, 무응답 편향 등이 있고, 또 다른 한편으로는 관리작업이나 방법이행에 초점이 맞춰지고, 두번째 의미의 오차 측도 즉, 착오가 되는 무응답률, 편집실패율, 과소포함율 등이 있다. 첫번째 측도집합의 추정이 특수한 상황에서만 가능한 것인 반면, 두번째 집합은 대부분 쉽게 추정할 수 있다.

10.5 조사측정에서의 오차측정

일반적으로, 조사측정에서의 오차를 측정 및 관리하는데 사용되는 접근방식은 물리학에서 사용되는 접근방식과 비슷하다. Mahalanobis(1944)는 조사오차를 관리 및 연구하는데 있어 상호관입표본(interpenetrating samples)의 사용을 주장하면서 통계공학(statistical engineering)이라는 표현을 사용하였다. 그리고, 런던대학의 Jerzy Neyman과 함께 연구하면서 나중에 미국 국립표준청(National Bureau of Standards) 통계공학연구소 1대 소장을 역임한 Eisenhart는 측정의 개념을 하나의 생산공정이라고 믿었다(Olkin, 1992). Eisenhart(1952, 1963)는 측정방법과 측정프로세스를 서로 구분하였는데, 여기서의 프로세스는 방법의 실현으로 정의된다.

상호관입표본에 기초한 것을 포함하여 많은 실험적 연구들이 수년간 광범위하게 선택적 방법의 효과를 조사하거나 측정하는데 이용되었다. 미국에서는 이러한 것이 1950년의 10년 단위 센서스와 관련하여 처음으로 이루어졌다(Hanson and Marks, 1958). 이러한 유형의 연구들에서는 Groves (1989, p.307)가 지적한 바와 같이, "[측정에서] 차이점의 설명이나 원인이 추구되고 있다." 하지만, 이러한 연구들의 결과를 해석하는데 있어서는 많은 주의가 필요하고, 여기에 관해서는 나중에 다시 논하도록 하겠다.

작업품질 관리를 위한 표본추출은 일찍이 1940년 미국 센서스에서 Deming에 의해 제기되었다(Olkin, 1987). 반면, 1940년경에 시작된 상시인구조사(Current Population Survey)를 위한 프로세스관리절차는 일차표본단위의 수가 68개에서 230개로 변화하여 실업인구 추정치에 상당한 차이가 나타난 1950년 중반까지는 아주 극미한 상태였다(Hansen et al., 1955). 이에 대해 연구원들은 "68개 영역의 원 표본에서는 면접원의 감독과 교육이 적었던 반면, 훨씬 많은 수의 새로운 일차표본단위에서는 새로운 면접원들의 교육에 많은 관심이 집중되었기 때문에 이러한 차이가 발생했다."고 결론내렸다(Hansen, 1987, p.185). 그리고, 그 결과 프로세스관리절차가 확대되었다(상세한 내용은 Hansen and Steinberg (1956) 참조).

그러나, 품질관리절차를 조사측정에 적용함에 있어 그 뒤에 숨은 가정은 무엇인가? Shewart(1939)는 무한한 작업의 반복횟수에 상응하는 무한수열을 산출하기 위해 한 작업에서의 재현성(reproducibility)에 대한 필요성을 언급하였다. 조사측정의 맥락에서 재현성에 대해 다루었던 HHP는 재현가능한 방법을 다른 연구원들이 따를 수 있고, 최소한 표본과 응답분산의 범위내에서는 동일한 가정에서 동일한 결과를 얻을 수 있는 방법이라고 정의하였다. 즉, 규격은

변이성(variability)의 원인이 될 수 없고, 해석상의 차이는 있을 수 없다는 것이다. 사실, 이것은 달성할 수 없는 목표이다. 하나의 프로세스가 통계적으로 관리되는 시기를 결정할 때는 Shewhart의 권고를 따를 수 있을 것이다. 그리고, 이러한 프로세스관리챠트 방식은 어떤 프로세스의 단계와 변이성을 예측할 수 있을 때 그 프로세스가 통계적으로 관리된다는 개념을 바탕으로 한다. 다만, 자료입력이나 코딩과 같은 조사작업을 평가할 때는 이 개념을 어떻게 사용할 것인지에 대해서는 쉽게 결정할 수 있겠지만, 조사질문을 개발하거나 계층간의 표본할당을 결정하는 프로세스들이 통계적으로 관리되는 시기를 결정할 수 있는 방법은 아직까지 개발되지 않았다.

품질관리를 서비스에 적용하는 것은 제조분야에 적용하는 것보다 훨씬 폭넓은 접근방식을 필요로 한다. 품질관리는 반드시 물질적인 통계상품보다 많은 곳 개인의 성과, 의사결정, 처리해야 될 데이터와 같은 곳에 많이 적용되어야 한다. 그러므로, Shewhart 챠트에다 부가적으로 "발견, 추정, 비교, 시험효과, 업무 및 다른 작업분석을 위한 무작위표본, input-output 분석, 학습곡선분석, 절차서와 규격, 대기 또는 지연시간 분석, 그리고 현장 시험 및 실험설계 등을 위한 표본추출"을 비롯한 많은 기법들이 필요하다 (Rosander, 1985, p.6).

또 한가지, 선택적 방법이나 규격을 평가하는데 필요한 가정들은 무엇인가? 선택적 방법에 관한 예측이나 추론을 하는 것에 비해 그 결과를 이해하는 데는 지금이나 앞으로도 선택적 규격의 해석에 아무런 변동이 없고, 각각의 작업세트들이 언제나 통계적으로 관리된다는 가정을 필요로 한다. Deming(1987, p.361)이 주장한 바대로, "분석적 문제들에 대한 통계이론은 아직까지 개발되지 않았다. 유일한 예외가 있다면 그것은 통계적 관리에서 증명될 수 있는 안정된 시스템으로부터 나오는 성과나 상품에 대한 것이다." 방법이나 절차에 관한 추론을 할 때는 계수적 목적보다는 분석적 목적을 위한 결과를 이용하고 있다.

조사의 계수적(enumerative) 목적은 사실적 설명 즉, 실업인구가 몇 명이냐, 미국에서 현재 보유중인 석유량은 얼마냐, 혹은 뉴욕시에 거주하는 세대의 월평균 가계지출은 얼마냐 등과 같은 질문에 답하는 것이다. 반면, 분석적(analytic) 목적은 원인이나 효과 즉, 실업자들은 왜 직업이 없느냐, 석유보유량이 왜 작년 보다 적거나 많으냐, 혹은 뉴욕시에서는 왜 가계비용이 많으냐 등과 같은 질문에 이유를 답하는 것이다. Deming(1953, p.245)은 "개념은 낡은 것이다"라고 했지만, 그도 조사측정문헌에 이러한 개념들이 존재하는데 대해 책임이 있는 사람 중의 하나다. Hahn and Meeker(1993)는 계수적 연구와 분석적 연구 양면으로부터 추론을 하는데 필요한 가정들을 검토하였다.

관리챠트의 사용은 기본적으로 분석적이다(Deming, 1953). 그리고, 선택적 방법론의 시험도 분석적이다. 시험이 수행될 때 지배적인 조건들은 앞으로는 두 번 다시 존재하지 않을 것이다. 또한, 시험은 판단표본(judgment sample)으로 간주해야 한다(Deming, 1987). 앞으로의 측정에 대한 예측은 초 모집단(superpopulation)이나 확률프로세스(stochastic process)의 범주 추정이라는 면에서만 표현될 수 있고, 이것은 모델 종속적(model-dependent) 접근방식을 필요로 한다(Hansen et al., 1983). 반드시 가정해야 하는 모델은 기본적으로 프로세스가, 우리의 경우에는 전체 조사측정프로세스가, 관리되고 있다는 것이다.

대부분 반복적인 조사가 수행되고 있는 상황에서는 유한모집단 표본이론에서 가정된 바와 같이 조사측정의 프로세스가 관리되고 있는 지가 확실하지 않다. 조사측정은 그 속에서 기계가 아닌 많은 사람들이 광범위한 작업을 수행하는 하나의 프로세스이다. 시간이 갈수록 응답자와 면접원, 자료검토요원, 코딩요원, 조사감독원 등은 변하기 마련이다. 그리고, 비록 작업을 수행하는 사람들은 그대로라 하더라도, 이들의 행동은 변한다. 왜냐하면, 사람은 경험과 사건을 바탕으로 자신의 행동을 수정하려는 것이 자연스러운 현상이기 때문이다. 따라서, 평균제곱오차(또는 어떤 하위성분)나 프로세스척도의 추정은 특정한 시간과 상황에 대한 추정일 뿐이다. 이러한 척도들은 여러가지 프로세스가 관리범위 내에 남아있는 지를 파악할 수 있도록 반드시 반복적으로 생산 및 감시되어야 한다. 이와 같이 반복된 척도들은 사용자들이 통계의 사용 적합성을 평가하는데 필요한 메타데이터의 없어서는 안 될 중요한 부분이다.

10.6 프로세스개선과 조사측정의 통합

Deming의 14가지 포인트(1982, p.343) 중 첫번째는 "상품(product)과 서비스(service)의 개선을 위한 쪽으로 목적의 항구성(constancy)을 창조하라"는 것이다. 통계상품의 개선은 프로세스의 개선으로부터 비롯되기 때문에 많은 품질개선 주창자들은 지속적인 프로세스 개선이라는 표현을 사용한다. Deming의 세번째와 네번째 포인트에는 "품질이 쌓이고 있는 통계적 증거가 필요하다"와 "가격과 함께 의미있는 품질척도에 달려 있다"는 구문이 있다. 이와 같이, 프로세스 개선은 생산시스템에 짜 맞춰진 좋은 품질척도가 있느냐에 달려 있다. 생산의 일면을 바꾸고자 할 때의 결정은 비용과 품질에 관한 사안들을 기초로 해야 하고, 이것은 다시 말해 변화는 입증할 수 있는 개선을 이끌어야 한다.

Deming의 철학에서 품질관리는 견고한 통계적, 과학적 원칙의 기반에서 확립된다. 상품의 경우, 생산공정과 품질척도를 정의할 수 있다면, 철학을 실천하고 상품을 개선할 수 있다. 조사척도는 쉽게 식별할 수 있는 상품이고, 조사측정에는 몇 가지 프로세스 모델이 존재하며(10.4절에 기술한 바와 같이), 여기에는 대체로 수많은 품질 척도들 즉, 조사척도 및 프로세스들이 있다. 따라서, 조사측정과 프로세스개선 사이에는 자연스러운 연결고리가 있는 것으로 보인다.

현실적으로 볼 때, 프로세스개선은 사업을 하는 방법과 같이 조사측정에서 왜 자동적으로 이루어지지 않을까? 즉, 일반적으로 말하자면 조사측정작업이 있는 많은 기관들은 왜 각자의 품질관리 프로그램을 개발하고, 실행해야 될 필요성이 있는 것일까? 혹은 보다 현실적으로 볼 때, 예를 들면 무응답을 감시하기 위한 척도와 절차들은 왜 대부분의 조사에서 적절하게 들어맞지 않는 것일까? 만약, 조사측정과 프로세스개선이 통합되어 있다면, 프로세스개선척도의 산출은 조사척도의 산출에 병합될 수 있고, 각각의 품질관리 인력들이 존재하지 않을 수 있을 것이다.

문제는 품질척도를 정의하는데 있다고 말하는 사람들이 있다. 평균제곱오차를 정량화하는 것은 불가능하고, 응답편향을 정량화하는 데는 너무 많은 비용이 든다. 반면, 저렴한 비용으로 쉽게 계산할 수 있는 무응답률과 같은 간접적인 성과척도의 개선은 반드시 평균제곱오차를 감소시키는 결과를 가져다 주지는 못한다. 또, 문제는 비용에 있다고 말하는 사람들도 있다. 조사측정프로세스에 관한 자료를 정의하고, 수집하고 평가하는 데는 비용이 든다. 프로그램 관리자들은 불가능하다고 인식하는 상황에 놓여 있다. 이들의 예산은 축소되고, 생산비용은 증가하고 있다. 그리고, 여전히 문제는 일차적인 측정목표 조사척도 그 자체에서 비롯된다고 말하는 사람들도 있다. 만약 상품이 시계열 통계라면, 일부 고객들은 측정프로세스 보다는 오히려 현상을 유지하는데 더 많은 관심을 가질 수도 있을 것이다. 다시 말해, 통계적으로 관리되고 있지만 부정확한 프로세스가 보다 정확한 프로세스보다 나은 것이 되고, 이에 따라 서로 다른 기대값을 가지게 된다.

Deming(1982)은 이러한 각각의 정서들을 다루었다. 자료 사용자, 응답자, 조사방법론자 등을 포함하여, 조사측정과 관련한 모든 고객의 요구를 충족하기 위해서는 전체 조사측정 프로세스인 개념화에서부터 타당성 검증에 이르기까지의 관리에 초점을 맞추는 것이 필요하다. 이행활동(예, 체계생성, 표본선별, 면접행위, 코딩 등)에 관한 프로세스 관리에 대한 사고만으로는 부족하다. 평균제곱오차와

같은 통계상품 품질의 단일척도는 조사측정 프로세스를 개선하는데 있어 충분한 정보를 제공할 수 없다. 따라서, 앞으로의 계획을 세우고, 연구, 교육에 자원을 투자하는 것이 하나의 팀으로서 움직이는 관리직의 역할이다(고객 포함). 그리고, 프로세스 품질의 척도와 평균제곱오차(또는 그 하위성분) 간의 직접적인 연결관계를 지금 설명할 수 없다고 하여 잠재적인 관계 속에서 척도를 산출하고, 조사하는 것을 배제해서는 안된다 (Groves, 1990). 비록 단순한 모델이라 하더라도 의사결정에 중요한 정보를 제공할 수 있는 것이다 (Linacre and Trewin, 1989).

품질을 측정하는 일차적인 도구는 1940년에 사용되었던 표본조사와 같다. 잘 설계된 표본조사 측정프로세스는 센서스보다 나은 추정치를 제공할 수 있으며, 이것은 순환적인 조사측정 프로세스모델의 각 단계에서는 틀림없는 사실이다. 게다가, 선택적 방법론과 자료품질을 평가하기 위한 Mahalanobis의 상호관입표본은 여전히 비용효과적이다. 몇몇 종류의 반복된 부표집을 이용한 분산 추정은 복잡한 표본설계로 진행되는 조사에는 일반화되어 있다. 상호관입표본은 면접원 및 응답 분산을 측정하는데 이용된다. 그리고, 임베디드(embedded) 실험은 선택적 설문지나 절차 등을 평가하는데 유용한 정보를 제공할 수 있으며(Fienberg and Tanur, 1989), 부표집(subsampling)은 자료입력 및 코딩을 감시하기 위한 통계 프로세스관리에 사용된다. 통계적으로 관리되는 프로세스를 가정할 수 있게 해 주는 적절하게 설계된 시험이나 실험에 자금을 투자하는 것은 종국에는 부적당하거나 무용한 자료를 취득하는 위험을 감행하는 것에 비해 훨씬 비용효과적인 것이 될 것이다.

이론적으로 보자면, 조사측정과 프로세스개선의 통합은 단순히 프로세스에 역점을 두고, 프로세스 개선을 측정할 수 있는 도구를 가지는 것 이상이다. 조사측정 프로세스와 관련된 모든 사람들은 반드시 통계를 산출하는 궁극적인 목적이 기관의 사내직원을 비롯한 모든 고객에게 의미있는 자료를 제시하는 것이며, 이 자료가 정보를 제공하기 위해서는 반드시 메타데이터가 수반되어야 한다는 사실을 인식해야 한다. 예컨대 개념화 중에 정의된 요구사항에는 모든 고객의 요구를 충족할 데이터와 메타데이터가 모두 들어있어야 한다. 따라서, 조사측정 프로세스의 상품을 일단의 통계치나 모집단 추정치로 정의하는 것으로는 불충분하다. 배포되는 통계상품에 추정된 표준오차와 출처 및 신뢰도 문구를 삽입하는 계획도 마찬가지로 불충분하다. 물론, 무응답률이나 품질 개요를 추가하는 것이 도움은 되겠지만, 이 또한 여전히 불충분하다. 내부고객들은 어떤가? 핫덱 대체(hot-deck imputation)에 대한 특정한 알고리즘이나 특정한 질의어법을 사용하는 이유에 대해 알고 싶어 하는 고객들은 또 어떠한가?

조사측정과 프로세스개선의 진정한 통합을 위해서는 데이터와 메타데이터 모두에 통계측정의 과학적 원칙들이 반드시 적용되어야 한다. 조사측정의 관점에서 비롯된 것이든 프로세스개선의 관점에 비롯된 것이든 오차의 척도는 그 자체만으로는 부족하다. 즉, 메타데이터를 산출하고 분석하는 프로세스가 개념화에서부터 타당성 검증에 이르기까지 조사측정의 전 프로세스에 반드시 절대적으로 필요한 요소가 되어야 한다.

감사의 글

조사측정과 프로세스개선이라는 개념들을 통합하는데 있어 어려움을 함께 한 BLS 품질측정팀 동료 팀원들에게 감사한다. 조사연구에 사용된 영어의 여러 가지 표현에 대한 어려움에 Wesley Schaible의 높은 식견은 본 절을 준비하는데 헤아릴 수 없는 귀중한 도움이 되었다. 그리고, 수없이 많은 참고문헌들을 힘든 내색하지 않고 조사해 준 Daphne Van Buren과 기술적 도움을 준 Dan Rope에게도 감사드린다. 끝으로, 본 절에 기술된 견해들은 저자 혼자만의 견해이며, BLS 전체의 견해를 반영하는 것은 아님을 밝혀두고자 한다.

제 11 절 통계기관들의 지속적인 품질개선

David Morganstein and David A. Marker
Westat, Inc., Rockville Maryland

11.1 서문

지난 15년 동안, 산업현장에서의 품질개선에 관한 많은 글들이 쓰여졌다. 그리고, 지난 5년 간에는 품질개선을 정부부처에 적용하는 것에 대해 논하는 논문과 저서들이 많이 간행되었다(Osborne and Gaebler, 1993; Cohen and Brand, 1993; Gore, 1993). 그 중 일부 논문에는 통계기관의 품질개선에 관한 일반적인 방안들인 검토되었다(Colledge and March, 1993; Linacre, 1991). 하지만, 이 논문들은 대개 체계적인 응용방법에 대한 사안을 반영하지 않았다. 뿐만 아니라, 성공적인 품질개선 프로그램은 기초적인 통계방법들이 통합되어야 하고, 이것은 특히 통계기관에게서 기대할 수 있는 사안이다(Simmons, 1972; Morganstein et al., 1993; Morganstein and Hansen, 1990; Hapuarachchi et al., 제26장).

분명, 통계기관들은 품질에 항상 관심을 가져왔다. 문헌들을 보면, 정확도를 향상시키기 위해 재설계된 자료수집방법과 기록된 자료의 정확도(그리고 오차)를 전달하기 위한 보다 나은 방법을 보여주는 많은 예가 나와 있다 (Hansen et al., 1955; Groves, 1989). 그럼에도 불구하고, 전체 기관 차원에서 품질개선을 위한 체계적인 접근방식은 간과되어 왔다. 본 절은 프로세스의 중요성을 강조한 Dippo(제20장)의 견해를 바탕으로 조사기관 자체를 최종 상품이나 조사, 자료등록 또는 기타 상품을 생산하는 하나의 프로세스 시스템으로 바라보는 체계적인 접근방식에 대해 논하고 있다(이 새로운 접근방식은 보다 나은 프로세스 관리가 결과를 개선할 것이라는 사실을 인식하지 못한 채 거의 모든 초점을 조사 상품의 품질측정에만 맞추고 있는 종래의 품질관리방안과 대조적이다). 따라서, 본 절에서는 통계기관들이 통계 상품을 산출하는 프로세스들을 어떻게 시스템적으로 구분하고, 보다 잘 관리할 수 있는지에 대해 설명하고자 한다.

먼저, 11.2절은 지속적인 품질개선을 위한 확고한 계획을 설명하고, 이것은 몇몇 세부단계로 구성되어 있다. 대부분의 단계에는 통계기관들의 사례가 제시되어 있다. 그리고, 11.3절은 당시최량의 방법(CBMs; current best method)에 대해 논하고 있는데 이 방법들 중 하나는 여러 표준들 속에 포함되어 있다. 16개 기관들의 Colledge and March(제22장) 조사는 목적과 정의를 주로 다루는

표준들을 확립하였다. 이 조사에 따르면, 조사 프로세스들의 일상적 작업에 대한 최량의 방법들이 광범위하게 사용된 사례는 거의 없는 것으로 드러났다. 11.3절에는 CBMs들의 활용을 위한 정의와 근거, 그리고 이를 개발하는 방법과 피해야 하는 함정에 대한 내용도 들어 있다.

11.2 지속적인 품질개선

본 절에는 조사기관에서 지속적인 개선방안을 이행하는데 필요한 특정 단계들이 논의되어 있다. 우리는 이 방법을 우리 조직뿐 아니라, 여러 서비스 및 제조기관들에게 사용해 보았으며, 서비스 기관의 일부는 정부기관들이었다. 그림 11.1, 지속적인 품질개선 프로세스에 관한 흐름도에는 조사의 고객(외부 또는 내부)에서부터 비롯되는 계획의 몇 가지 중요한 측면과 이 측면들의 상호관계가 나타나 있다. 이것은 조사고객의 수요로 다시 되돌아오는 끊임없는 하나의 루프이며, 통계상품(조사결과)에 대한 초점을 조사 프로세스의 이해와 이를 이용한 작업으로 빠르게 옮기고 있다. 바로 이러한 것들이 종합품질관리의 핵심요소들임을 반드시 기억해야 한다(Cohen and Brand, 1993).

이 계획을 상세히 살펴보기 전에 다음 용어에 대한 정의를 먼저 확인해보기로 하자.

외부고객(External customer) : 외부의 자료 요청자 또는 사용자 (의회, 주정부 내지는 다른 기관이나 주 또는 정부기관, 연구단체 등).

내부고객(Internal customer) : 타부서나 개인의 통계상품 또는 서비스를 수령하는 동 기관내 한 부서나 개인.

측정능력(Measurement capability) : 측정프로세스의 평균제곱오차는 전반적인 오차 요구사항에 비해 작다.

프로세스 안정성(Process stability) : 프로세스 변동이 전적으로 무작위 성분으로 구성되는 상태.

프로세스 능력(Process capability) : 안정적인 프로세스의 무작위변동은 고객의 요구사항이나 그 변동의 한계보다 작다. 안정적인 프로세스는 고객의 변동 요구사항을 충족하지 못한다면 불가능하다.

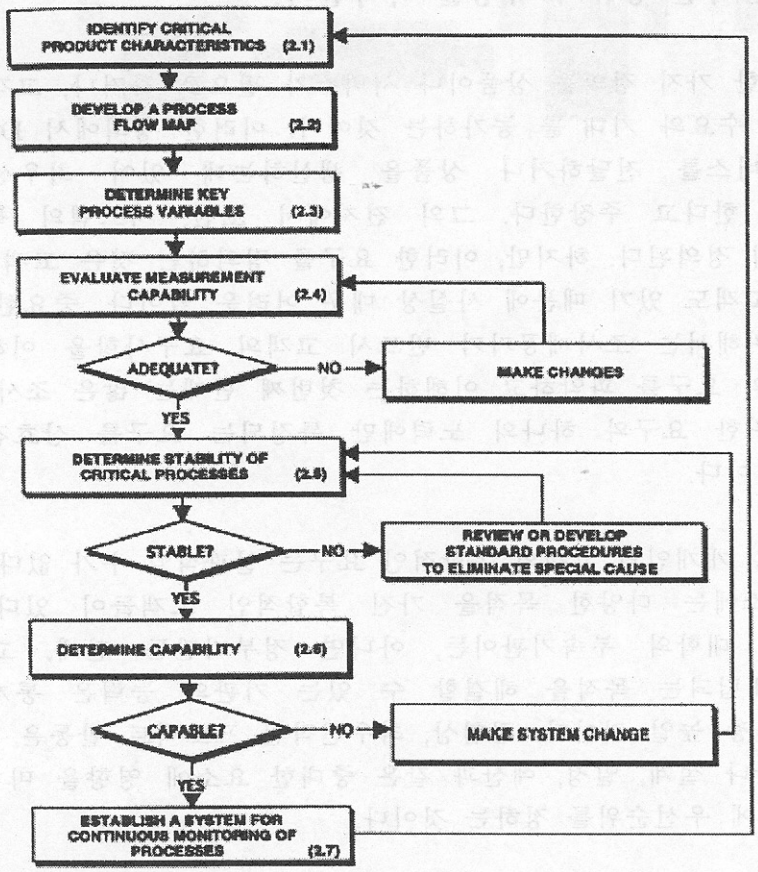


그림 11.1 지속적인 품질개선을 위한 계획

계획의 제1단계는 조사고객의 요구와 고객에게 가장 중요한 조사특성 확인 및 중요하면서도 관리 가능한 조사 프로세스의 측면들에 초점을 맞추고 있다. 다음으로 부원들은 이러한 프로세스를 성공적으로 측정할 수 있는 능력을 평가한다. 이와 같이, 적당한 측정능력이 보여질 때, 조사 프로세스의 안정성을 결정하기 위한 신뢰도 있는 데이터가 수집 가능하다. 만일, 프로세스가 안정적이지 않다면, 수정된 운영절차나 재교육이 안정화에 도움이 될 수 있다. 조사 프로세스가 안정적으로 되면, 그 결과를 고객의 요구와 비교할 수 있다. 이 때, 프로세스가 고객의 요구사항을 지속적으로 충족하지 못하다면, 이것은 안정적이라 하더라도 무용한 것으로 간주된다. 또, 프로세스가 무용하다면(예, 소요 응답률이나 정확도를 지속적으로 얻을 수 없거나, 특정 기한이 있는 데이터가 정기적으로 기록되는 경우), 시스템의 변경이 요구된다. 이러한 평가 및 수정/재교육의 순환과정은 단순한 조사수행이 아니라, 통계기관의 모든 다른 활동에도 적용될 수 있다. 이 계획의 각 단계는 아래에 보다 상세히 설명되어 있다.

11.2.1 핵심적인 상품의 특성을 파악한다.

품질의 한 가지 정의는 상품이나 서비스가 필요로 하거나, 고객의 "경쟁적인 가격에서의 수요와 기대"를 능가하는 것이다. 이러한 정의에서 Deming(1982)은 고객은 서비스를 전달하거나 상품을 생산하는데 있어 최우선적인 초점의 대상이어야 한다고 주장한다. 그의 견지에서 보면, 시스템의 목적은 고객의 요구에 의해 정의된다. 하지만, 이러한 요구를 정의하는 것은 고객이 내부고객도 있고, 외부고객도 있기 때문에 사실상 매우 어려운 일이다. 중요한 모든 목적을 충족하기 위해서는 조사제공자가 반드시 고객의 요구사항을 이해하고 있어야 한다. 고객의 요구를 파악하고 이해하는 첫번째 단계는 많은 조사에 확대할 수 있는 광범위한 요구와 하나의 노력에만 특정되는 요구를 상호간에 주의깊게 구분하는 것이다.

불행히도, 개개인의 고객의 복합적인 요구는 일관적일 수가 없다. 많은 조사와 데이터베이스에는 다양한 목적을 가진 복합적인 고객들이 있다. 통계기관이든, 민영회사든, 대학의 부속기관이든, 아니면 정부기관든 간에, 고객의 요구를 이해하고 대립되는 목적을 해결할 수 있는 기관의 능력은 통계상품의 성공 가능성을 한층 높일 것이다. 경험상, 최우선적인 프로젝트 활동은 (하나 이상의) 고객들과 만나 설계, 일정, 예산과 같은 중대한 요소에 영향을 미치는 상충되는 조사목표 간에 우선순위를 정하는 것이다.

특정한 조사의 목표를 명확히 하기 위해 외부고객과 만날 때는 조사고객과 설계자들이 사용하는 언어에 많은 관심을 가져야 한다(Groves, 1989). 예를 들면, 하나의 정부기관에서 정확도(accuracy)란 단어는 여러 부원간에 분산과 평균제곱오차를 모두 의미하는 것으로 사용되었다. Marker and Ryaboy(1994)는 정확도라는 표현은 환경통계학자들에 의해 세 가지 대립하는 개념인 편향, 총오차(편향 + 분산), 성과평가감사 수행 중에 확인된 오차(vs 품질관리표본에 의해 확인된 오차) 의미하는 것으로 사용되어 왔다고 지적한다. 용어의 정의 없이는 고객의 목표를 충족하는 것과는 상관없이 "정확한" 조사를 설계하게 된다.

고객만족도조사(CSS; customer satisfaction survey)는 고객의 광범위한 요구에 대한 개요와 지난 실적에 대한 고찰을 제공할 수 있다. CSS는 고객의 품질정의(즉, 정말 중요한 특성들이 무엇인가)와 특정한 상품 및 서비스에 관한 고객의 인식을 결정하는데 사용된다. 또 다른 것으로는 이용가능한 상품 및 서비스(예, 데이터베이스, 컨설팅 등)에 대한 고객의 친밀도를 평가하고, 고객에게

보여줄 수 있는 다른 상품 및 서비스를 확인하고자 하는 목적도 있을 수 있다. 놀라운 사실은 고객들은 그들이 매우 귀중하다고 여기는 상품이나 서비스의 특성에 대해 알지 못할 수도 있다는 것이다(예, 15년 전의 자동차 고객들은 에어백의 제공 가능성에 대해 알지 못했다).

11.2.2 프로세스 흐름도를 개발한다

통계상품을 개선하는 다음 활동 즉, 프로세스 흐름도를 개발하는 것은 관련 프로세스에 대한 이해를 확대시킨다. 이 단계는 활동내역에 대한 충분한 인식 이상의 결과를 가져다 줘야 한다. 대개의 경우, 이 단계는 프로세스가 이미 잘 이해되어 있다는 가정을 우선 거부한다. 프로세스에 참여한 사람들은 프로세스의 내역을 상세히 잘 알고 있음은 의심의 여지가 거의 없다. 하지만, 많은 기관들에 대한 우리의 경험에 비춰보면, 이러한 생각들을 비교했을 때는 반드시 상당한 차이가 나타난다(Gross and Linacre, 제23장). 조사프로세스 참여자들은 모두들 각자의 관점에서 프로세스를 바라본다. 따라서, 상호간에 대화가 없을 경우에는 이러한 차이들이 지속적으로 존재할 것이며, 결국에는 프로세스를 제약하는 차선의 프로세스들로 될 것이다.

프로세스 흐름도에는 세 가지 구성성분을 포함시킬 것을 제안한다. 첫째, 프로세스의 시이퀀스에는 의사결정 시점, 프로세스의 흐름 그리고, 각 단계별 고객(들)에 관한 설명을 덧붙인다. 둘째, 프로세스 제공자와 고객들의 의사소통을 개선하면서 각 프로세스의 주체들을 확인한다. 셋째, 중요한 프로세스 변수나(즉, 프로세스를 반복할 때마다 다를 수 있고, 핵심적인 상품특성에 영향을 미칠 수 있는 요인들) 의사결정 또는 프로세스 참여자들이 취할 수 있는 행동들의 목록을 작성한다. 프로세스 변수들의 예로는 설문지 어법, 인원구성, 표본설계, 면접원 교육프로그램 내용 등이 있다. 각 프로세스를 소유한 사람은 누구이며, 중요한 프로세스 변수는 무엇인지를 결정하기 위해서는 프로세스에 참여한 인원들이 프로세스 흐름도를 개발하는데도 참가해야 한다. 흐름도가 도출되는 프로세스는 최종 상품만큼이나 중요하다.

도시된 바와 같이, 한 정부통계기관은 2년에서부터 적게는 6개월 미만의 기간 동안 국가의 연간 대기가스 배출량을 추정하는데 필요한 시간을 단축하기 위해 품질개선팀(QIT)을 발족시켰다. 팀의 구성원에는 추정치를 생성하는 부원과 이러한 추정치의 기초자료를 제공하는 부원들도 포함되어 있었다. 팀은 먼저 프로세스 흐름도를 작성하였다. 프로세스에는 데이터 테이프에 적힌 ID 번호와 발전소명을 비교확인하는 작업도 포함되어 있었다. 팀은 자료제공자인 팀원이

다른 데이터 소스에 동일한 명칭과 ID가 있는 것을 확인했을 때는 이 모든 단계를 생략할 수 있었다. 프로세스 흐름도 생성과정에서 팀은 목적을 달성한 기존 프로세스들에서 간단한 개선이 있었음을 확인하였다.

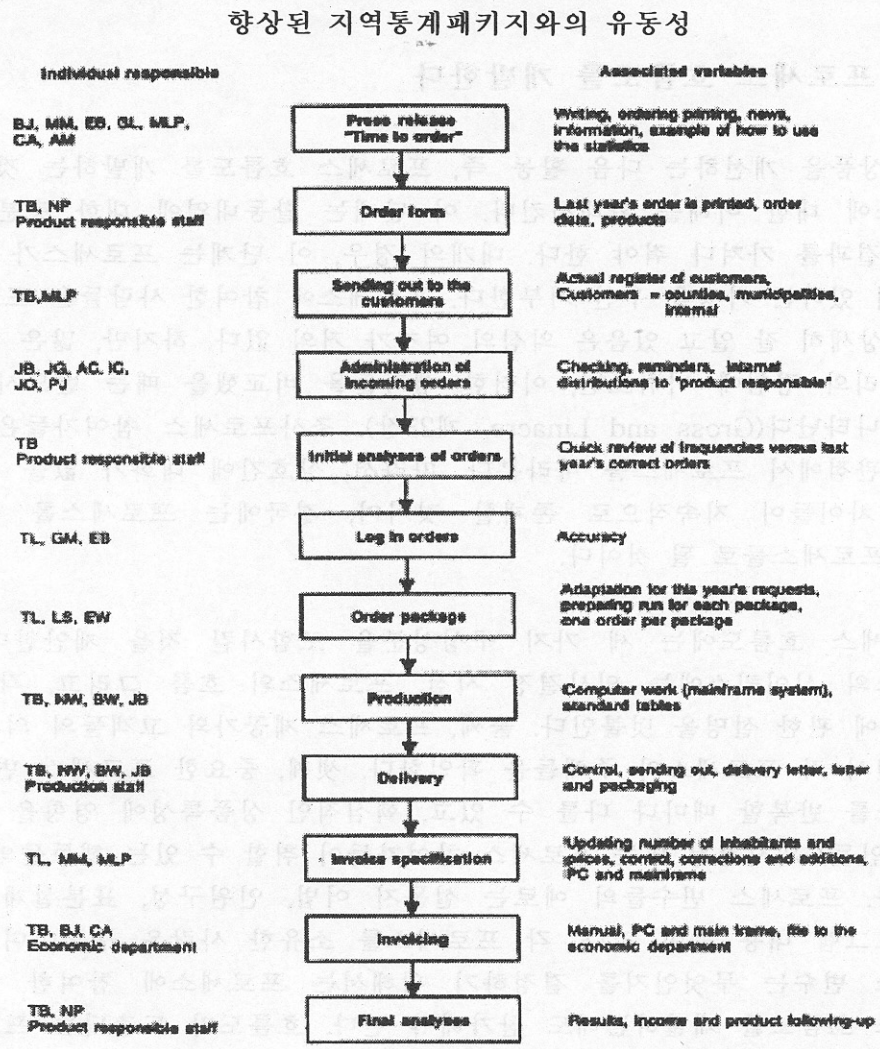


그림 11.2 흐름도의 세 가지 구성성분

그림 11.2는 흐름도의 일례를 보여준다. 흐름도를 개발한 팀의 목적은 고객의 요청에 대응하는 기관의 유동성을 증가하는 것이었다. 중앙 아래는 지역의 통계패키지 생산에 관계된 각각의 프로세스들이다. 그리고, 그림의 왼쪽 아래부분은 각 프로세스 책임자의 명단이며, 오른쪽에는 프로세스 참여자들이 이해할 수 있는 용어로 표현된 각 프로세스와 관련된 변수들이 나와 있다.

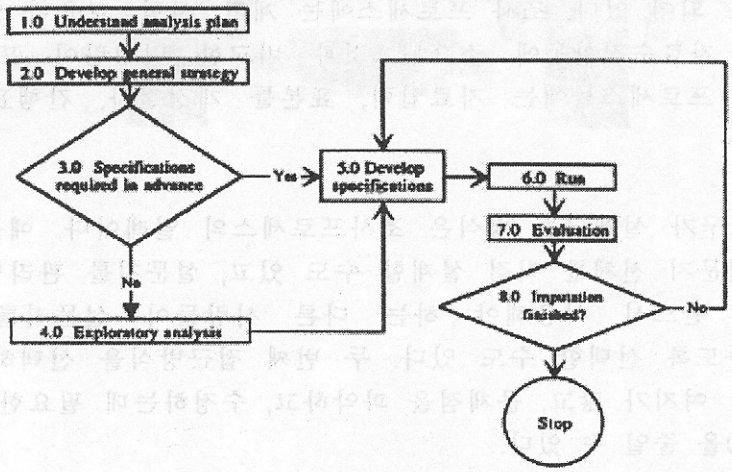


그림 11.3 매크로 수준의 대치(imputation) 프로세스 흐름도

어떤 흐름도에는 상세내역 간에 적당한 균형이 필요하다. Cohen and Brand(1993)는 관리자들 앞에 97단계의 긴 흐름도가 제시된 상황에 대해 설명한다. 관리자들은 회의적인 입장이었고, 작업자체가 시간낭비라는 반응들을 보였다. 이 예에서는 두 가지 교훈을 얻을 수 있다. 흐름도의 개발에는 관리자들을 참여시켜라. 그리고, 매크로와 마이크로 수준의 챠트를 이용하여 흐름도를 단순화시켜라. 무응답에 대한 대치 프로세스를 보여주는 매크로 수준의 챠트(그림 11.3)는 새로운 부원이 전체 프로세스를 이용하는데 유용하고, 마이크로 수준의 챠트는 개개의 프로세스들이 어떻게 상호작용하는 지를 이용하는데 필요하다.

프로세스의 한 부분이 나머지 다른 부분과 어떻게 관련되어 있는지와 프로세스 자체가 다른 활동들과 어떻게 연관되어 있는 지를 파악하는 것은 중요하다. 각 구성성분 책임자들 뿐만 아니라, 하급부원들이 프로세스의 시작과 끝을 이해할 때 장벽은 없어진다. QIT는 시스템에서의 변동의 원인을 파악한 후, 그러한 원인을 반영하는 변수들이 현재 측정되고 있는 지의 여부를 결정하는데 프로세스 흐름도를 이용할 수 있다.

11.2.3 중요한 프로세스 변수를 결정한다

우리는 이 단계가 가장 어렵다는 사실을 알았다. 서비스 활동의 프로세스들은 생산환경보다 유형적인 형태가 아니며 따라서, 확인하고 측정하는데 보다 많은 노력이 든다. 기초가 되는 프로세스들은 현실적이만, 그럼에도 결과 데이터의 품질을 통제한다. 대개 프로세스들은 의사결정과 이러한 의사결정이 이루어지는

경로의 형태로 되어 있다. 조사 프로세스에는 계획, 시험, 교육에 할당된 시간과 자원을 실제 자료수집활동에 소요된 것과 비교한 비교량이 포함되어 있다. 그리고, 다른 프로세스들에는 자료입력, 표본틀 개정절차, 간행표 확인절차가 포함되어 있다.

자료수집도구가 설계되는 방식은 조사프로세스의 일례이다. 예를 들면, 도구 설계자들은 설문지 전체를 직접 설계할 수도 있고, 설문지를 관리할 면접원들과 같이 그것을 반드시 사용해야 하는 다른 사람들이 설문지를 검토하게끔 여유시간을 두도록 선택할 수도 있다. 두 번째 접근방식을 선택하면, 설문지의 품질이 개선될 여지가 많고, 문제점을 파악하고, 수정하는데 필요한 재작업량(예, 편집, 대치 등)을 줄일 수 있다.

이 단계의 목적은 프로세스 결과와 상품 또는 서비스 품질 및 이들의 일관성과 비용에 가장 큰 영향을 미치는 중대한 요소들을 확인하는 것이다. 그리고, 바로 여기서, 팀 구성원들은 질문을 묻고, 각 프로세스 요소와 변수에 관한 응답을 기록하기 시작한다. 이들은 각각의 축적된 지식과 경험을 바탕으로 품질에 영향을 미칠 가능성이 클 것 같은 프로세스들을 선별한다. 이러한 중대한 측면들 중 상당수가 일상적으로 측정되지 않거나, 혹은 심지어는 전혀 측정되지 않는다는 사실은 실로 놀라운 일이다.

중요한 프로세스 변수들을 확인하는 가장 간단하면서도 효과적인 방법 중 하나는 파레토 그림이다(Gitlow, et al., 1989). 이 도구는 Vilfredo Pareto(Juran et al., 1979)가 "중대한 소수(vital few) 대 사소한 다수(trivial many)"을 분류하는 것으로 표현한 원칙 즉, 모든 오차의 대다수에 책임이 있는 비교적 소수의 오차(또는 결함) 유형을 찾으려 노력하는데 기초를 두고 있다. 일례로, 몇몇 정부 통계기관들이 하나의 조사품질의 척도로서 추적하는 편집오차율을 한번 생각해 보자. 그림 11.4의 파레토도를 보면, 오차원인이 빈도수별로 내림차순으로 정렬되어 있어, 소수의 중대한 사안을 쉽게 확인할 수 있다. 나머지 다수의 사소한 편집은 기타항목에 모두 함께 들어 있다. 이 특정한 경우를 보면, 97회의 편집오차 중 과반수 이상이 B3 문항에 집중되어 있고, 86%가 B3, B2, B5 문항에 해당될 뿐, 나머지 문항의 오차는 총 오차의 14%에 불과하다. 파레토 챠트는 자원을 어떻게 할당할 것인가를 결정하는데 보다 효과적이다. 이 경우에는 편집 프로그램이 진정한 특이점(outlier)만을 발견할 수 있도록 B3, B2, B5 문항에 할당할 수 있는 범위가 재평가되었다. 품질개선을 위해서는 편집오차의 원인을 파악하고, 개선된 질의어법이나 허용범위의 수정 또는 기타 다른 방법을 통해 앞으로의 발생빈도를 줄이기 위한 노력이 요구된다.

파레토 챠트는 빈도 외 다른 데이터를 분석하는데도 이용될 수 있다. 챠트의 행은 대개 각 자원에 소요된 비용이나 일수를 나타내게 되는데, 이렇게 표현한다면 그러한 결과의 척도들이 단순히 점으로 나타낸 빈도보다는 오히려 낫다. 그러나, 불행히도 결과들을 하나의 자원에 직접 연관짓기란 상당히 어려운 일이다.

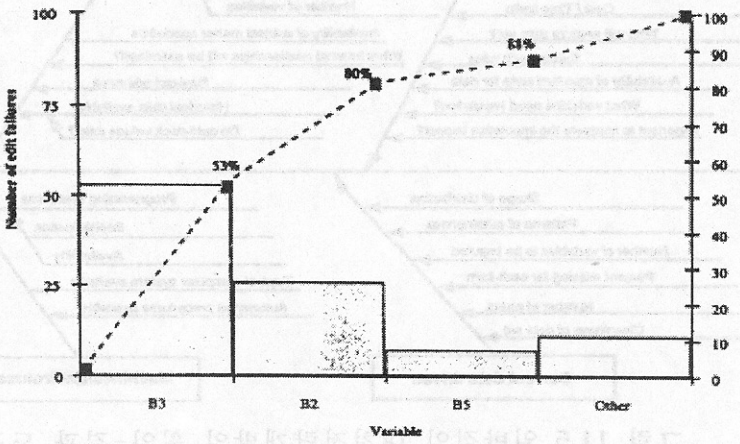


그림 11.4 범위오차 편집의 파레토 챠트

빈도 또는 다른 유형의 데이터를 활용할 수 없을 때에도 물론 중요한 프로세스 변수들은 확인할 수 있다. 원인-결과 도표(fishbone 또는 Ishikawa 도표라고도 함)는 이러한 목적을 위한 것이다(Ishikawa, 1982). 그림 11.5는 QIT가 대치 프로세스에서 개발한 것으로, 일반적인 대치전략개발에 영향을 미치는 요인들을 묘사해 놓은 fishbone 다이어그램이다(그림 11.3도 QIT의 상품이었다. "일반적인 전략"이란 그림에서 두 번째 상자를 말한다). Fishbone상에 있는 모든 요인들로부터 팀은 가장 중요하다고 생각하는 5-6가지를 선택한다. 이것은 측정대상 요인들이자, 변동성을 줄여야 하는 요인들이다. 복합적인 요인들을 선택하는 이유는 전문지식을 가진 팀은 일반적으로 몇 가지의 중요한 요인들을 확인할 수 있을 것이기 때문이다. 그러나, 이들이 최선의 판단에서 단 하나만을 고를 경우에는, 단 하나만을 선택하는데 대한 객관적인 기준이 거의 없을 것이기 때문에 이것은 비교적 경미한 요인이 될 수도 있다. 몇 가지 요인들을 선택함으로써 이들은 이러한 불행한 상황을 미연에 방지한다. 일단, 몇 가지 중요한 요인이 선택되고 나면, 다음 단계는 시간을 거치면서 이것을 측정하는 것이다. 이 내용은 다음 두 절의 주제로 넘긴다.

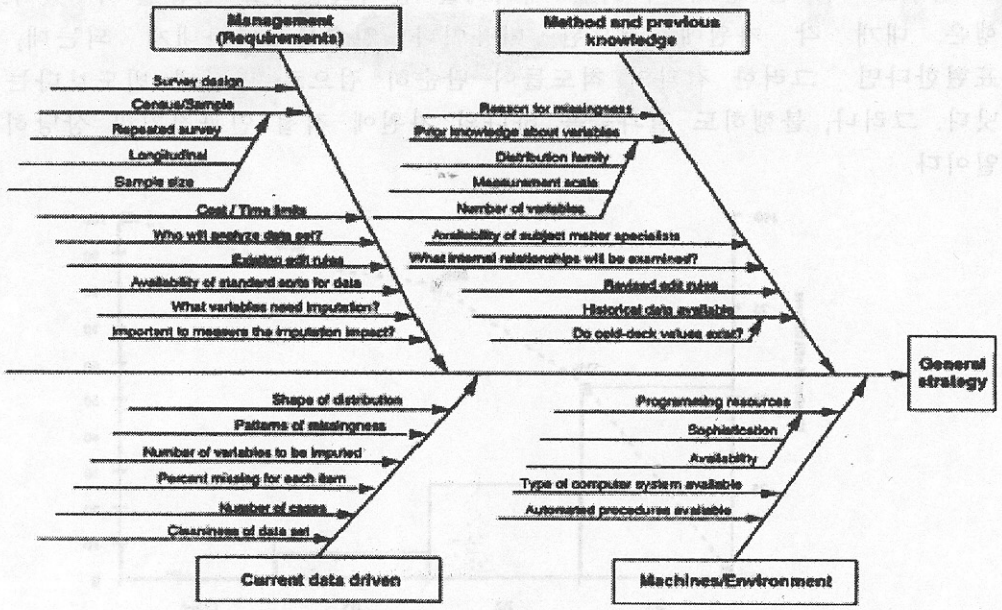


그림 11.5 일반적인 대치전략개발의 원인-결과 도표

11.2.4 측정능력을 평가한다

통상적인 착오는 측정오차에 대한 지식없이, 자료를 수집하고, 프로세스 안정성에 관한 결론을 내린다는 것이다. 선택적으로, 연구자들은 종종 중요하지만, 측정이 어려운 프로세스보다는 오히려 쉽게 측정할 수 있는 프로세스를 선택한다. 경험에 비추어 볼 때, 측정오차는 대개의 경우 품질개선에서 가장 낮은 비중을 차지하는 측면 중 하나였다.

고객만족도조사는 통상적으로 제한된(3점 내지 5점) 척도를 사용하고, 그 결과는 거의 모든 응답자들이 동일한 값을 선택하는 것으로(예, 매우 만족) 나타난다. 이와 같이 빈약한 측정시스템은 만족도의 상관관계에 거의 아무런 견해를 제시하지 못한다. 또한, 많은 조사에서 극단값만 있는 척도를 사용하고, 그 결과 중간값 응답의 의미에 관한 정보를 얻지 못한다. 이러한 경우들에 있어서는 측정오차가 지속적인 품질개선의 목적을 위한 조사의 활용을 제약한다.

미국 National Health Survey 재설계의 일각에서는 대인면접비용을 절감하는 방법을 파악하는데 도움이 되고, 제안된 재설계 방법론의 비용에 관한 효과를 측정하기 위해 대인면접과 관련된 활동들의 비용을 검토하였다. 비용데이터는 몇 년 동안 면접원들이 기록한 것으로 전산화된 데이터베이스에 입력되었다. 그러나,

재설계자들이 이러한 조사를 수행하면서 기록된 데이터는 바람직한 분석을 뒷받침할 수 없다는 사실을 알았다. 즉, 필요한 데이터를 수집하는 데는 기존의 면접원 절차를 수정하는 것이 필요하였다. 이에 따라, 비용데이터는 재설계자의 요구사항을 충족할 수 없었고, 조사 재설계방법에 관한 결정은 대인 자료수집과 관련된 비용에 미치는 효과에 관한 정확한 정보없이 이루어져야 했다.

프로세스의 품질을 개선하는 것은 가능하나, 개선을 정량화하는 것은 불가능하다. 위 건강조사의 예에서는 재설계된 조사가 수집된 건강데이터의 정확도 및 비용 양면에서의 품질을 개선할 것이라고 주장할 수 있을 것이다. 그러나, 적절한 비용데이터가 없으며, 변화의 비용효과를 정확하게 측정할 수 없고, 여러가지 선택적 설계를 평가하는데 비용을 쓸 수 없게 된다. 게다가, 개선의 정도를 측정할 수 없으면, 상부관리자로부터 품질개선업무에 대한 허가를 얻기가 더 어려워 진다.

11.2.5 핵심적인 프로세스들의 안정성을 결정한다

중요한 프로세스 변수들이 일단 확인되고 나면, 이러한 변수들에 안정성 시험이 가해진다. 조사 프로세스들이 합당하게 예측가능할 때(또는 안정적), 변수들은 개선의 노력이 시작된 후의 비교 근거를 제공할 수 있다. 중요한 프로세스 변수의 안정성에는 자료를 취득 및 분석하는 체계적인 방법이 필요하다. 이것은 주로 관리챠트와 파레토 분석 및 기타 여러 가지 통계적 도구와 방법론들을 이용하여 성취된다. 이 단계에서는 프로세스 및 상품 변동의 근원과 프로세스 변화의 변동에 대한 효과가 확인된다.

프로세스 변이성은 주로 두 가지 유형 즉, 특수원인(special cause)과 일반원인(common cause)으로 분류된다. Deming은 Shewhart의 이상원인(assignable cause)을 특수원인으로 개명하였다(Deming, 1982; Shewhart, 1931). 어떤 경우에는 특정하고 식별가능한 문제가 상품이나 서비스 또는 프로세스 상태의 불합격의 원인이 될 수 있다. 예를 들면, 한 묶음의 설문지가 한 사람에 의해 부적합하게 코딩되었거나, 디스켓의 결함이 하나의 표집틀로 사용된 파일의 기록상의 오차의 원인이 될 수 있다. 또, 원재료들이 (예, 설문지나 표지글을 작성하는데 사용된 종이나 잉크의 품질) 규격에 안 맞을 수도 있다. 특수원인들로 알려진 이 각각의 변동원인에 대해서는 특정한 일시간대의 지역적 행동만으로도 충분히 상황을 개선할 수 있다. 변동의 특수원인을 제거하게 되면, 프로세스 변이성이 감소하고, 조사 신뢰도가 증대된다.

이와 대조적으로, 일반원인 변동은 모든 프로세스에 침투한다. Deming (1993)에 따르면, "변동의 일반원인들은 장기간에 걸쳐 관리범위 내로 스며들어 관리챠트상에 여러 점들을 생성한다. 변동의 일반원인들은 매일 그 자리에 그대로 있다." 이에 따라, 잘못된 인원채용과 교육이 체계화되고, 모든 조사에서 모든 인원들에게 적용된다. 시대에 뒤떨어진 절차와 낮은 수준의 측정능력이 모든 연구에 영향을 미친다. 그리고, 지역적인 문제해결 노력으로는 이와 같은 일반원인 변동을 처리할 수 없다. Deming(1982)은 일반원인 변동을 줄이기 위한 "조직적인 행동"은 관리의 책임이라고 설명하였다.

많은 특수원인의 예들은 일반원인의 근원으로 바뀔 수도 있다. 표본틀이 잘못된 데이터는 한 가지 경우를 통해 들어올 수 있지만(특수원인), 이것은 표본틀을 조사하는 규정이 잘못되어 있는 프로세스에서는 전형적인 것이 될 수 있는 것이다(일반원인). 스킵 패턴을 따르지 않은 면접원은 개인적인 문제로 어려움을 겪겠지만(한번의 부주의로 특수원인을 초래하는), 만일 그러한 slip-up들이 보편화되어 있다면, 이것은 보다 나은 교육에 대한 필요성이나 혹은 새로운 설문에서 스킵 패턴을 확인해야 할 필요성을 의미할 수 있다(특수원인).

Juran et al.(1979) 그리고 다른 사람들(Gitlow and Gitlow, 1987)은 프로세스 변동의 약 15%만이 특수원인에 기인하고, 나머지 85%(대다수 변동)는 일반원인에 의한 것이라고 추정하고 있다. 그러면서, Juran은 대부분의 기관들이 최소 50% 이상의 노력을 특수원인에 대해 고민하는데(예, 하급자의 메모 검토) 소진하고 있다고 주장하였다.

일반원인과 특수원인을 정정하는 책임은 서로 다른 부원에게 있다. 그러므로, 두 가지 변동원인을 서로 구분하면 바람직하지 않은 결과를 방지할 수 있다. 정해진 시간에 무엇이 잘못되고 있는 지를 확인하는 것은 단지 첫번째 단계에 불과하다. 종종 간과되는 두번째 단계는 일반원인으로부터 특수원인을 구분해 내는 것이고, 그런 다음 어떤 부원이 어떤 조치를 취해야 하는 지를 결정하는 것이다. 이와 같은 구분은 관리챠트를 이용하는 주된 이유다(Rosander, 1985). Simmons(1972)는 작업자의 지엽적인 조치가 필요한 지 혹은 시스템을 변화시키는 관리자의 조치가 올바른 답인지를 정하기 위해 조사의 각 싸이클에 대한 오차들을 비교할 때 관리챠트를 사용하는 것에 대해 논하고 있다.

그림 11.6은 환경데이터를 수집하는 무작위전화선별조사의 관리챠트이다. 이 챠트에는 중요한 프로세스 변수인 카운티별로 완료된 전화통화 비율이 나와 있다. 횟수가 지나치게 낮은 스크리너(screener)에 대해서는 해당 카운티의 전체 조사를 지연시키면서 알맞은 가구를 확인하기 위한 추가적인 전화통화를 요구하였다. 낮은 횟수는 추가적으로 선별된 전화번호의 통화시간이 초기와 같은

시간동안 완료되도록 할당되지 않을 경우, 편향의 잠재적인 원인을 발생시킬 수도 있다.

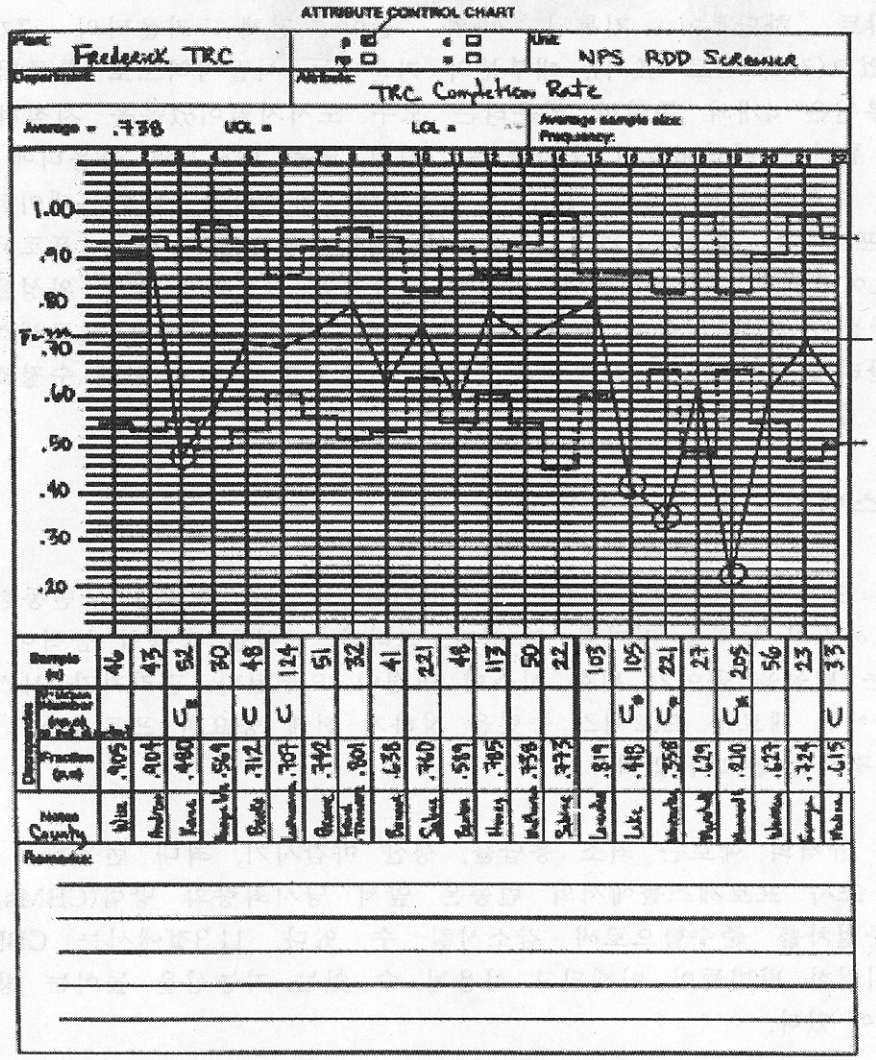


그림 11.6 완료된 telephone screener에 대한 관리챠트

챠트에서 검은색 실선은 관측된 완성율을 나타내고, 두 개의 점선은 모두 동일한 비율을 가진 카운티들 간의 무작위변동을 가정한 상태에서 관측된 완성율의 범위를 나타내는 것이다. 이것은 관리한계들이다(표본크기가 다르므로 카운티별로 다르게 적용됨). (표준관리챠트 방법론을 이용하면, 이러한 상하위의 관리한계는 전체 평균 표준오차의 ±3 범위내에서 설정된다. 따라서, 안정적인 프로세스에서는 400개중 1개 정도만 이 한계를 초과할 것으로 예상된다) 완료된 카운티가 있으며, 프로젝트 부원은 곧바로 관측된 선별율을 챠트에 표시한다.

그리고, 전부는 아니지만 많은 수의 카운티가 완료되고 나면, 관리한계가 결정된다. 팀은 4개 카운티(챠트상의)의 예상밖의 낮은 완성율에 대한 원인을 찾는다. 챠트 하단부에 기록된 바와 같이 전체 카운티의 7개만이 도시지역이었고(각각 U로 표시), 대부분의 카운티는 시골지역으로 분류되었다. 프로젝트 부원은 4개의 특이한 카운티는 모두 도시지역이었음을 지적하고는, 스크리너에 통과된 전화번호의 숫자를 증가시켜 모든 도시지역 카운티에 대한 프로세스를 수정하였다. 그림 11.7은 시골지역 카운티에 대한 데이터만을 보여준다. 프로세스 평균과 관리한계는 시골지역 카운티에 대한 프로세스만 구분하여 보여주도록 재산출되었다. 이러한 시골지역 카운티에 대한 완성율에는 예상밖의 변동이 전혀 없음에 주의해야 한다. 즉, 이것은 안정된 프로세스에서 비롯된다. 관리챠트의 활용은 잠재적인 문제를 예측하고, 고객과 함께 수정조치를 미리 계획할 수 있게 해 준다.

11.2.6 시스템성능을 결정한다

안정적인 프로세스에서는 규격을 충족하거나 예상된 프로세스 변동한계를 정하는 프로세스의 능력을 평가할 수 있다. 그러나, 규격을 충족할 능력이 없는 프로세스에는 변동을 줄이기 위한 시스템 변경이 요구된다. 관리자가 시스템을 변경한 다음에는 새로운 프로세스 능력을 정하기 전에 중요한 프로세스 변수의 안정성을 재평가할 필요가 있다.

프로세스 규격의 예로는 최소 응답률, 생산 마감시기, 최대 변동계수 등이 있다. 주요 조사 프로세스들에서의 변동은 앞서 당시최량의 방법(CBMs)이라 말했던 표준절차를 준수함으로써 감소시킬 수 있다. 11.3절에서는 CBMs를 정의하고, 이러한 방법들이 이해되고 사용될 수 있는 가능성을 높이는 방법에 대해 논의되어 있다.

11.2.7 프로세스를 지속적으로 감시할 수 있는 시스템을 확립한다

안정적이고 유능한 프로세스를 달성하는 것은 프로세스 개선의 시작에 불과하다. 과거에 능력이 있는 것으로 확인된 프로세스는 그대로 방치해 둔다면, 여전히 유능한 것으로 기대해서는 안 된다. 기술은 점점 진부해 지고, 공급자들은 새로운 방법을 실천하고, 조사감독원과 고용인들은 무의식적인 오차와 부주의한 오차를 만들며, 고객의 요구는 계속해 변화한다. 그러는 동안, 지속적이고 끊임없는 개선의 목표는 여전히 그대로이다. 그러므로, 부원들이 정보를 유지하고, 피드백을 제공하기 위해서는 그리고, 프로세스 변동을 관리하고,

개선된 방법을 통해 지속적인 프로세스 변동을 감소시키고자 하며, 프로세스 변경을 평가하는데 있어 지원을 하기 위해서는 감시시스템이 필요하다. 다음 절에서 논의될 당시최량의 방법(CBMs)은 변동을 최소화할 수 있게 도와주고, 이에 따라 품질을 개선하고, 최량의 관행을 따를 가능성을 증대시키며, 감시 프로세스를 보다 쉽게 해 준다.

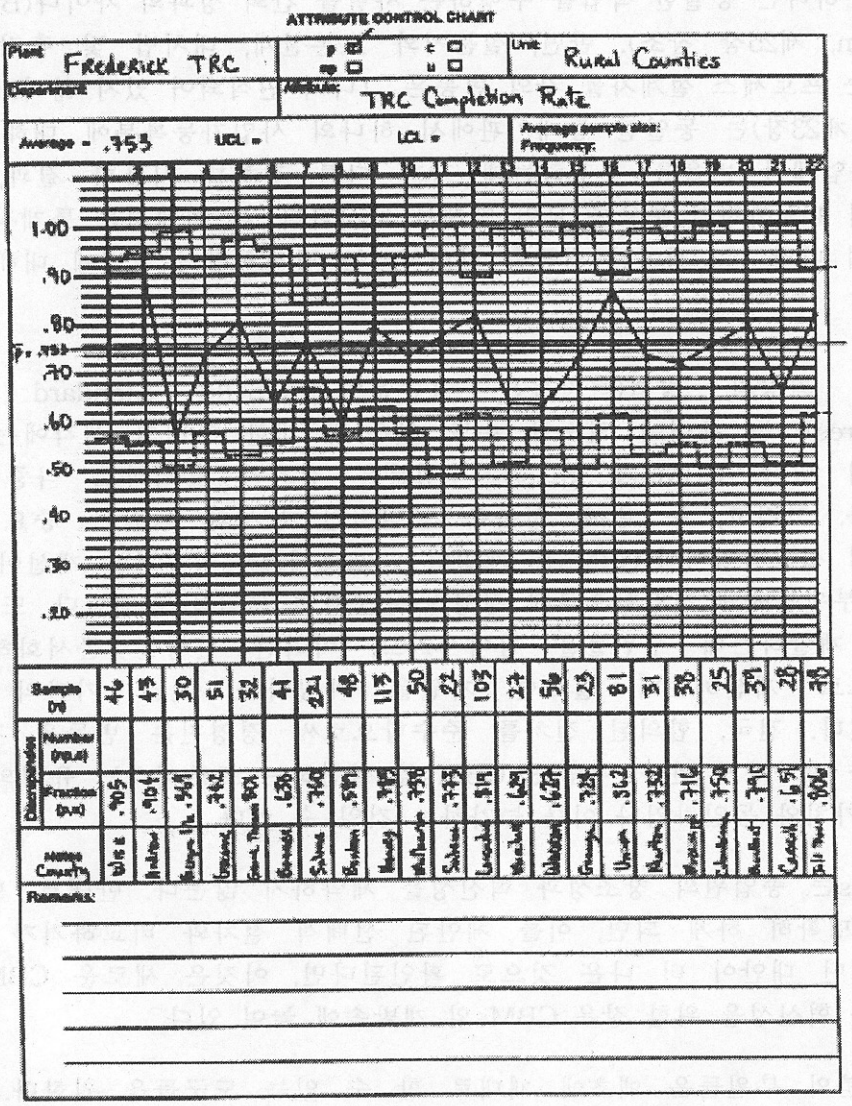


그림 11.7 시골 지역 카운티의 관리챠트

11.3 품질개선에서의 CBMs의 역할

"최량의 관행이 표준화되어 있지 않고, 관련된 모든 사람에게 전달된다면, 지속적인 개선을 얻기란 불가능할 것이라는 것은 말할 것도 없다" (Kristensen et al., 1995). 변동의 축소는 품질개선의 중심이다. 가장 빈번하게 확인된 변동의 원인 중 하나는 동일한 작업을 수행하는 사람들 간의 성과의 차이다(Batcher and Scheuren, 제25장 참조). 반면, 설문지와 표본설계, 대치법 및 추정절차 등을 개발하는 프로세스 설계자들 간의 변동은 그다지 인식되어 있지 않다. Gross and Linacre(제23장)는 동일한 통계기관에서 하나의 사업자등록부에 대해, 종업원이 없는 사업체의 비율을 추정하는데 각기 다른 절차를 사용한 결과, 추정치가 31%에서 44%까지의 분포를 보인 일례를 제시한다. 본 절에서는 통계작업의 많은 측면에서 나타나는 개인간 변동을 감소시키는 당시최량의 방법에 대해 논하도록 하겠다.

당시 최량의 방법(독자들은 표준운영절차(SOP; Standard Operating Procedures)라는 용어에 익숙할 수도 있음)에 대한 우리의 시각에는 몇 가지 성분들이 있다. CBMs의 한 가지 성분은 절차의 문서화로 나중에 자세히 논하겠다. 그리고, 두 번째 성분은 물질적인 문서와 똑같이 중요한 것으로 경영진의 접근방식이다(CBMs를 개발하는 프로세스). 지속적인 개선이 가치있는 일임을 부원들에게 전달함으로써 경영진은 중요한 목적을 확립한다. 또한, 최량의 관행을 개발하는데 부원들을 참여시키고, 이러한 관행을 문서화하며, 이를 지속적으로 개선하는데 필요한 자원을 제공하는 것은 기관의 경쟁력을 강화시킨다. 결국, 합의된 절차를 준수함으로써 경영진은 변동을 감소시키는 행동들을 강화하게 된다. 이렇듯, CBMs를 실천하는 프로세스는 유형의 상품 즉, 문서의 가치와 동일하거나 이를 능가하는 것일 수 있다.

CBMs는 종업원의 창조성과 혁신성을 제약하지 않는다. 반대로, 당시최량의 방법을 명확히 하게 되면, 이를 제안된 선택적 절차와 비교하기가 쉬워진다. 만일, 어떤 대안이 더 나은 것으로 확인된다면, 이것은 새로운 CBM이 된다. 창조성과 혁신성을 위한 장은 CBMs의 개발속에 놓여 있다.

대부분의 부원들은 애초에 제대로 할 수 있는 도구들을 원한다. 재작업은 혼돈스럽고 무의미한 활동이다. CBMs은 이러한 재작업의 양을 감소시키는 방법을 제공한다. Batcher and Scheuren(제25장)이 기술한 바에 따르면, 미국세청(Internal Revenue Service)의 부원과 경영진은 모두 면접원들을 위한 표준조사응답지침(standard probe and response guide)의 개발과 이의 준수의 중요한 가치를 발견하였다.

11.3.1 CBM의 선택

당시 최량의 방법이라는 용어는 표준운영절차라는 구문과는 다른 메시지를 전달한다. 전자는 방법들이 지속적으로 개선되고 있다는 개념을 전달한다. 현재 문서화되어 있는 것들은 지금까지 알려져 있는 최량의 것들이다. 그러나, 향후에는 보다 나은 것들을 기대할 수 있다. CBM이라는 용어는 개개인이 현재의 운영방법을 결정하지 못하게 하면, 방법에 대한 지속적인 개선과 CBM을 이용한 운영에의 꾸준한 노력의 의미가 담겨 있다. 반면, SOP는 정적인 의미를 전달한다. 기록된 바에 의하면, SOP는 개성 상실에 대한 걱정을 앞세우고, 지속적인 개선을 위한 팀 접근방식의 철학을 뒷받침하지 못한다.

11.3.2 CBMs의 두 종류: 반복적 프로세스와 창조적 프로세스

CBMs은 요약하는 작업의 종류를 바탕으로, 반복적이고, 예측가능한 업무나 혹은, 독창적이고 창조적인 업무의 두 가지 광의의 범주로 분류될 수 있다. 전자의 작업에 대한 CBMs은 각 작업을 실행할 때 반드시 정확한 순서에 따라야 하는 체크리스트(CBM의 한 부분)로 쉽게 요약할 수 있다. 예를 들면, 비행기 조종사는 이러한 종류의 선형 CBM을 따라야 한다. 체크리스트 각 항목은 비행기가 이착륙할 때마다 매번 동일한 순서로 검토된다. 조사작업의 경우에는, 절차가 다소 복잡할 수는 있지만 가구 기록부를 작성하고, 수정하는 것과 같이 하나의 표본틀에서 표본을 선별하는(설계가 아님) 절차가 이 부류에 속한다. 또한, 영수증 관리작업도 아마 이 종류에 속할 것이다.

CBMs의 두 번째 부류는 의사나 과학자들이 수행하는 것과 같이 예측하기 어려운 프로세스들에 적용된다. 통계기관에서 고도로 숙련된 부원들은 체크리스트를 활용하면 도움을 얻을 수 없을 정도로 지나치게 창조적이라고 처음엔 스스로들 생각할 수 있다. 이와 같이 많은 전문가들이 자신의 작업은 너무 독창적이라 CBM은 아무런 관련이 없다고 생각할 수 있으나, 진실로부터 더 이상 벗어날 수는 없다. 이러한 작업들은 몇 가지 면에서 CBMs을 보다 필요로 한다. 왜냐하면, 이들의 프로세스는 너무 독창적이고, 많은 의사결정 가지들이 있는 비선형적이어서 중요한 요소를 잊어버리기가 매우 쉽기 때문이다. 예를 들어, 조사를 설계하거나 재설계하는데는 수 많은 중요한 단계들이 필요하다. 이 때, CBM은 반드시 처리해야 될 요소들을 상기시켜주는 역할을 한다. 정부조사를 수행하는 것과 같이 짧은 시간-프레임의 작업에 종사하는 전문가들은 틀림없이 복잡한 프로세스에서 중대한 단계를 간과하지 않도록 도와 줄 수 있는 모든 도구를 필요로 한다.

두 종류의 프로세스들은 모두 11.3.3절에 기술되어 있는 비슷한 CBMs 개발절차를 필요로 한다. 그리고, 두 종류의 CBMs은 보다 나은 것으로 알려진 방법으로 정기적으로 갱신되어야 한다. 반복적 프로세스를 위한 CBMs은 따라야 하는 단계들의 순서를 명확히 정의한다. 반면, 창조적 프로세스를 위한 CBMs은 처리방법이 아니라, 처리해야 되는 요소들을 정의할 가능성이 보다 크다.

11.3.3 CBMs 준비

11.3.3.1 지도팀의 역할

CBMs에 관한 작업을 시작하기 전에, 기관에서 CBMs의 개발 및 활용에 책임이 있는 상위 관리지도팀을 먼저 파악하는 것이 바람직하다. CBM 지도팀은 전사적 품질관리의 책임을 지고 있는 그룹이 될 수도 있고, 혹은 전체 품질팀에 보고하는 개별 팀이 될 수도 있다. 규모가 큰 통계기관의 경우에는 전체 기관과 각 부서에 지도팀을 두는 것이 바람직하다. 아래 기술된 바와 같이 이들의 역할은 서로 비슷하지만, 전사적 그룹은 부서 그룹의 업무를 개발 및 조정하는 책임을 지고 있다.

전체적인 책임

상위 관리자와 부원으로 구성된 지도팀은 CBM을 보다 쉽게 개발할 수 있다. 왜냐하면, 변화에 대한 많은 장애들은 기관의 사명을 정할 수 있는 권한을 가진 이 팀에 의해서만 극복될 수 있기 때문이다. 대부분의 통계기관들은 각자의 일상적인 상품을 생산할 능력이 있지만, CBMs을 개발하고, 이를 기관의 일상적 활동에서 활용하는 것은 매우 어려운 프로세스다. 이러한 활동들은 동시에 발생하지는 않는다. 여기에는 기관의 경영진이 지시하는 시간과 최선의 노력이 투자되어야 한다.

지도팀은 CBM의 영향에 속한 기관내 모든 사람들에게 적절한 역할을 확인한다. 이들은 의견을 유도하는 의사소통이 쟁점을 처리한다. 이들은 또한 수용을 이끌어내고, 친근감을 증대시키며, 행동을 변화시키기 위한 프로세스를 검토한다. 지도팀은 이러한 프로세스를 고려하고 계획한다.

참가자 선별

지도팀은 때로는 시간과 자원의 충돌을 해결하면서 각 CBM 팀을 선별하고, 지원한다. 지도팀은 진행상황을 검토하고, 업무량이나 기한을 변경한다. 팀원들은 CBM 개발과 부원교육 및 감시에 필요한 인적자원에 대해서는 의사결정의 권한을 가지지 않는다. 일반적으로 이것은 지도팀의 대표만이 할 수 있다.

변화에 위한 계획수립

특정한 행동들은 성과의 변화를 지원 및 촉진한다. 그 중 첫 번째 단계는 CBMs의 개발을 위한 주제선정이다. 초안의 전사적 검토는 새로운 절차에 대한 모든 부원의 수용을 가속화하는데 도움이 된다. CBM을 활용할 수 있을 때에는 몇 가지 형식의 프로세스 참여자 교육이 필요하다. 여기서, 지도팀은 다시 이러한 변화를 위한 계획이 확실히 수립되도록 한다.

준수 촉진

CBMs의 활용을 지원함으로써, 경영진은 새로운 방법을 준수할 것을 촉진시킨다. 왜냐하면, 보다 나은 방법의 활용성만으로 그것을 받아들일 충분한 사유가 되지 않기 때문이다. 부원들은 일반적으로 변화를 북돋우지 않는 한, 시도되고 실존하는 절차를 이용하려 한다. 지도팀은 변화하는 사람들을 위한 보상전략을 개발한다. 어떤 사람들은 보다 빠르게 방법을 변화시킬 것이다. 또 어떤 사람들은 개선된 방법을 채택하는데 경영진의 격려와 교육 및 리더쉽을 필요로 할 것이다.

11.3.3.2 CBM 개발프로세스

그림 11.8의 단계들은 신중한 CBM 개발프로세스를 제시하고 있다. 그림 왼쪽에는 CBMs 개발 각 단계의 책임자들이 명시되어 있고, 오른쪽에는 각 단계에서 고려해야 하는 주요 프로세스 특성들이 나열되어 있다. 이 플로우챠트는 한 기관의 품질평가에 사용되는 CBMs을 체계적으로 개발하기 위한 단계들을 보여준다. (1)전사적 CBM 활용으로 이익을 얻게 될 그룹의 사명에 절대적으로 필요한 프로세스들을 확인한다. (2)CBM을 작성할 팀을 선정한다. (3)CBM 문서를 개발한다. (4)CBM의 세부내역을 부원들에게 전달한다. (5)절차를 지키는데 도움이 되는 계획을 수립한다. 이러한 단계들을 자세히 살펴보면 다음과 같다.

중요한 프로세스 확인

11.2절에서 논의한 바와 같이, 통계기관의 각 부서에는 수없이 많은 프로세스들이 있긴 하지만, CBMs은 중요한 소수(vital few)에 대해서만 개발되어야 한다. 지도팀은 부서의 사명을 완수하는데 절대적으로 필요한 프로세스를 선별한다. 예를 들어, 표본틀(frame)이나 기록부(register)를 유지해야 될 책임을 지고 있는 부서에서는, 결정적인 프로세스에 입력자료의 갱신을 위한 절차와 중복자료의 삭제를 위한 절차들이 포함된다. 또, 표본을 추출하는 부서의 경우, 중대한 프로세스들로는 표본선별과 조사결과의 가중치 부여 및 대치가

있다. 자료입력 부서에서는 자료검색을 위한 규칙이나 코드 충돌을 해결하기 위한 절차 등이 포함될 수 있다. 이 때, 너무 많은 프로세스들이 선정되면, CBMs을 적용하는데 많은 노력이 소요되어 부원들을 지치게 만들 수 있다.

지도팀은 부원들의 참여가 물론 제안을 하는 수준에서 제한적이긴 하지만, 프로세스를 선별하는 과정에 전체 부서를 참여시킬 필요가 있다. 새로운 방법의 전폭적인 수용과 이행은 거의 전부원의 지지를 바탕으로 한다. 따라서, 전부서는 의견을 모으고 있는 CBMs의 개념을 알아야 하고, 각 팀이 모여서 선정된 프로세스들을 문서화할 것이라는 사실을 알아야 한다. 또한, 작성된 초안에 대한 각자의 생각과 의견이 중요하다는 사실도 인식해야 한다. 기관이 지속적인 품질개선을 기본적인 운영방침에 통합함에 따라, CBM 프로세스 선정의 역할은 자연적으로 프로세스에 참여한 부원들쪽으로 옮겨간다. 그리고, 초기에는 경영진이 프로세스를 선정하는데 지도적인 역할을 담당하게 되지만, 품질개선의 분위기가 무르익어 감에 따라 그 역할은 점점 적절한 자원을 확보하는 쪽으로 많은 비중을 두게 된다.

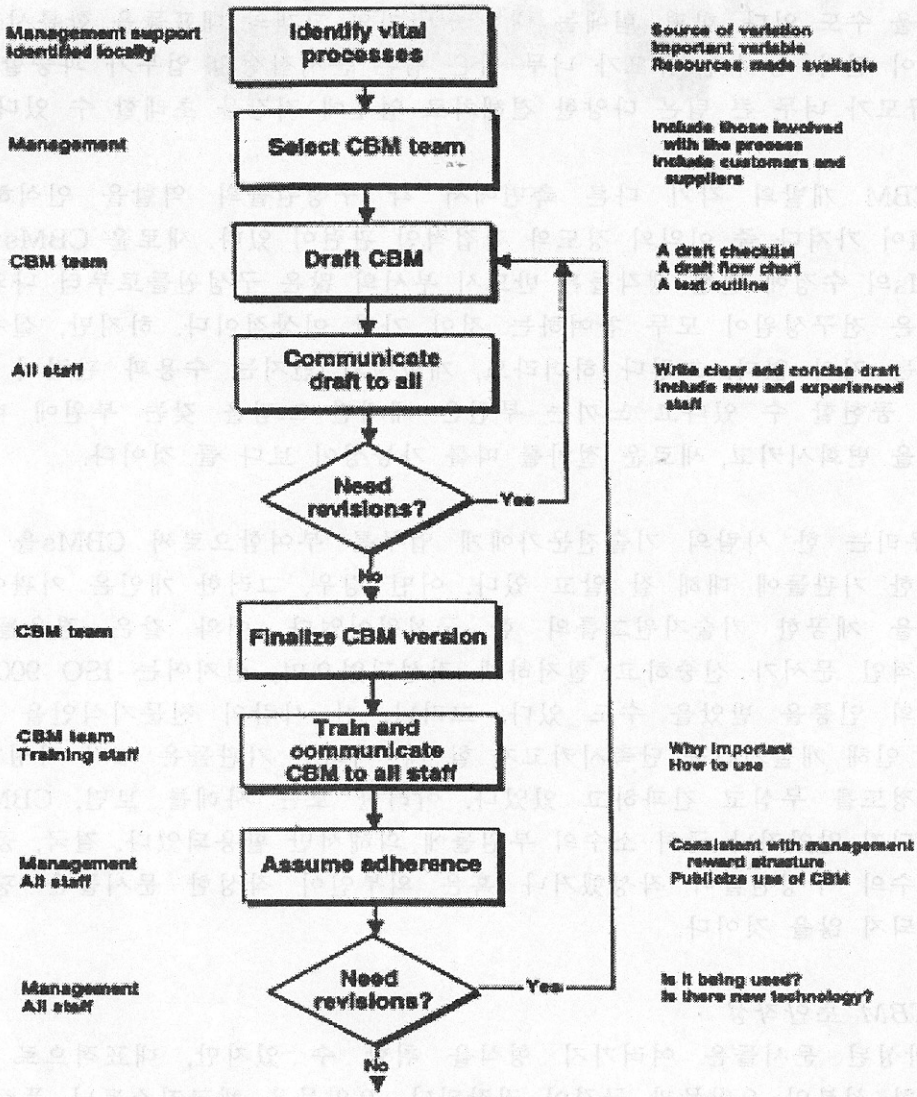


그림 11.8 CBMs 개발을 위한 플로우챠트

CBM 팀 선정

CBM 팀은 대개, 변동의 원인을 파악하고, 이를 제거하며, 그럼으로써 최량의 관행이 증명되어야 하는 프로세스를 결정하는 업무들을 해 온 기존의 품질개선팀이 맡는다. 경력자와 초보자로 고루 구성된 소규모 팀은 개발프로세스에 중요한 균형감을 가져다 준다. 부서의 문제를 맡아 처리해 왔던 이 팀의 부원들은 도전하는 어려움을 인식하고, 또 이에 대처하는 몇 가지 방법을 알게 될 것이다. 반면, 신입부원들은 신선한 시각을 제공한다. 고도의

지식을 동원하여 작성된 문서들은 이와 같이 경험이 부족한 부원들은 읽기 어려울 수도 있다. 한편, 팀에는 내부 공급자와 고객의 대표들을 합류시키는 것도 도움이 된다. 동시에, 규모가 너무 작은 팀은 문서작성의 업무가 과중할 수 있고, 또 규모가 너무 큰 팀은 다양한 견해차로 업무에 지장을 초래할 수 있다.

CBM 개발의 각기 다른 측면에서 각 구성원들의 역할을 인식하는 것은 CBM이 가져다 줄 이익의 정도와 직접적인 관련이 있다. 새로운 CBMs이나 기존 CBMs의 수정에 관한 생각들은 반드시 부서의 많은 구성원들로부터 나와야 하고, 사실은 전구성원이 모두 참여하는 것이 가장 이상적이다. 하지만, 실제로 이런 경우는 거의 없다. 그렇다 하더라도, 개방성의 인지는 수용과 관련이 있다. 즉, 뭔가 공헌할 수 있다고 느끼는 부원은 배제된 느낌을 갖는 부원에 비해 과거 관행을 변화시키고, 새로운 절차를 따를 가능성이 보다 클 것이다.

우리는 한 사람의 기술전문가에게 업무를 부여함으로써 CBMs을 신속하게 개발한 기관들에 대해 잘 알고 있다. 어떤 경우, 그러한 개인은 기관에 기술적 지원을 제공한 기술지원그룹의 한 구성원이었다. 이와 같은 경우들은 모두, 최종적인 문서가 신중하고 철저하게 작성되었으며, 심지어는 ISO 9000과 같은 표준의 인증을 받았을 수도 있다. 그러나, 한 사람의 전문지식인을 선별하고, 이로 인해 개발기간을 단축시키고자 할 때, 이러한 기관들은 그가 작성한 문서의 수용정도를 무심코 간과하고 있었다. 이러한 모든 사례를 보면, CBM이 거의 활용되지 않았거나 극히 소수의 부원들에 의해서만 활용되었다. 결국, 공감대없이 극소수의 구성원들이 작성했거나 혹은 외부인이 작성한 문서들은 광범위하게 활용되지 않을 것이다.

CBM 초안작성

작성된 문서들은 여러가지 형식을 취할 수 있지만, 대표적으로 두 가지 중요한 성분인 요약문과 규격이 권장된다. 요약문은 체크리스트나 플로우챠트의 형식일 수 있고, 프로세스의 중요단계에 대한 개요를 제공하며, 연대순으로 표현된다. 그림 11.9는 1단식 목록 표본추출 체크리스트로, 여기에는 관련 절의 세부내용을 확인할 수 있도록 항목번호가 참조되어 있다. 규격 항목에는 각 단계들의 배경과 타당성이 제시되어 있으며, 여러 개중 선택해야 될 사항에 대해서는 그 선택항목과 지침이 나와 있다.

프로젝트 번호_____ 착수일_____ 프로젝트명_____

설계자(팀원이 있는 경우 모두 기재)_____

참조 항	조치 또는 결정	확인	비고
	표본설계		
1.1	표본틀(들) 확인		
1.1.1	범위 적정성 결정		
1.1.2	중복계획 개발		
1.2	층화 여부(및 방법) 결정		
1.2.1	층화요원 선별		
1.2.2	계층 수 선정		
1.2.3	층화표본 할당		
1.2.4	추정치에 대한 층화효과 결정		
1.3	소트여부 결정		
1.3.1	소트변수 선택		
1.3.2	등급간격 결정 (필요시)		
1.3.3	소트명령설정		
1.4	계수와 다른 MOS가 있는지 결정		
1.4.1	MOS 확인		
1.4.2	최소 MOS 선택		
1.5	확실성 확인		
2.1	프로그래머 초기규격		
2.2	프로그래머 지침서		
2.2.1	적부 기록 정의		
2.2.2	누락 분야 조정방법 선택		
2.2.3	WESSAMP 규격 작성		
2.2.4	무작위 착수 지정		
2.2.5	사용될 MOS 선택지정		
2.2.6	목표표본의 크기 또는 비율 지정		
2.2.7	선택 재생허용에 관한 설명		
2.2.8	치환 선정		
2.3	QC 검증계획 개발		
2.3.1	최소 MOS 미만의 표본단위 없음		
2.3.2	층별, 하위영역별 도수 카운트		
2.3.3	층별 및 전체 MOS 합계		
2.3.4	지시된 표본이 있는 표본틀 비율 목록		
2.3.5	가중치가 부여된 수 조사		
3.	문서작성		
3.1	설계수정사항		
3.2	표본선택		
3.3	메모		
3.4	워크시트		
3.5	아키브 작성파일		

그림 11.9 1단식 목록 표본추출 체크리스트

요약문은 경력자든, 초보자든 할 것 없이 전부원들에게 유용한 것이며, 반면 규격은 새로운 구성원이나 경험이 부족한 부원들에게 가장 도움이 된다. 작성된 문서에는 활동에 익숙하지 않은 사람도 필요한 작업을 훌륭히 수행할 수 있도록 충분한 상세내용이 제공되어야 한다.

11.2절에서는 중대한 프로세스의 플로우챠트가 가지는 가치에 대해 논의하였다. 플로우챠트나 체크리스트에는 의사결정 참여자가 명확하게 명시되어야 한다(예, 그림 11.8의 왼쪽 행). 많은 경우, 이러한 의사결정 참여자는 1인 이상일 것이다. 그러나, 효율성의 개선은 명확한 의사결정체계가 합의되고, 문서화되어 있을 때 가능할 것이다. 플로우챠트는 여러가지 하위 프로세스들의 주체들도 확인한다. 전형적인 예를 들자면, 정기보고서 작성자가 있을 것이다. 따라서, CBM 팀의 업무는 종종 주체없는 보고서와(또는 고객이 더 이상 없는 것으로 확인된 보고서) 부서에 더 이상 아무런 가치가 없는 진부한 기타 행동들을 밝혀낸다.

팀에서 일단 CBM의 초안을 작성하고 나면, 주어진 프로세스에 참여한 나머지 전부원에게 이를(체크리스트, 플로우챠트, 상세 배경) 회람하는 것이 중요하다. CBM 초안에 대한 이들의 의견은 두 가지 이유에서 매우 중요한 의미를 가진다. 프로세스에 관한 지식을 기반으로, 보다 나은 대안적 절차를 발견할 수 있고, CBM의 효과적인 면에서 볼 때는 CBM이 진정으로 프로세스를 운영하는 최량의 방법을 제시하는 것이라는 사실에 전부원이 반드시 동의해야 한다. 전부원이 부서에 참여하고 있을 때, 이들은 CBM의 주체라는 느낌을 가지게 되고, 이것은 그 유용성을 확실히 하는데 도움이 된다. CBM 팀은 다른 부원들의 권고사항에 대해 논의하고, 영향을 받는 전부원의 합의가 이끌어질 때까지 필요하다면 계속해 반복적으로 문서를 수정해 나가야 한다.

정보공유 및 교육부원

중대한 프로세스에 관한 문서작성이 비록 해 볼만한 일이고, 상당한 노력을 필요로 하긴 하지만, 이 작업만으로는 부원들의 업무수행방법을 변화시킬 수 없다. 따라서, CBMs을 개발하고 그 중요성을 고취시킬 목적을 가진 정보공유 및 교육부원도 똑같이 중요한 역할을 차지한다. 문서작성 후에 이루어지는 부원들의 교육과정은 질의에 응답하고, 세부사항을 논하는데 도움이 될 수 있다.

이러한 정보공유활동의 결과는 CBM 개발에 직접적으로 참여하는 것 이상의 효과를 얻을 수 있다. 이러한 활동들은 장벽을 허물고 전사적 팀정신을 북돋우는데 도움이 되며, CBM 팀에 프로세스의 고객과 공급자가 모두 포함되어

있을 때는 특히 더하다. 부원들은 다른 프로그램(예, 표본틀이나 기록부를 수정하는 프로그램)을 통해 CBMs에 대해 학습하면서, 몇 가지 자신이 도전하는 일을 해결하는데 도움을 받을 수 있다. 전 기관의 품질은 이와 같은 식으로 개선된다.

준수 확인

부원들의 CBM 문서활용에 관한 피드백은 CBM의 적절성(relevant)과 유동성(current)을 보장한다. 하지만, 이것은 유용성의 측정없이는 달성될 수 없다. 이를 위해서는 법적/계약적 준수요구가 부여되지 않은 한, 비공식적인 감사(auditing) 프로세스가 권장된다. 이 때는 부원들에게 감시의 주목적이 절차를 평가하고, 절차의 가치를 증대시키기 위한 것임을 확인시켜야 한다. 감시의 목적은 징계가 아니다.

이러한 비공식적 감사방법의 몇 가지 예는 다음과 같다. 부원들에게 CBMs이 유용한 것이라고 생각하는지를 정기적으로 묻거나, 부원들로 하여금 CBMs의 사용과 제한에 관한 내부 프리젠테이션을 하도록 하거나, CBMs의 가치를 인정하지 않는 부원에게 그 이유를 설명하도록 한다(예, 새로운 기술로 CBM이 시대에 맞지 않는다는 등). 이러한 방식을 통하면, CBMs의 활용이 장려되고, CBMs에 필요한 수정사항들도 명확하게 드러날 수 있다.

CBM 팀원들은 준수확인업무를 수행하기에 적합하다. 팀에게는 종종 수정이 필요한 시기를 결정해야 하는 경우가 생기는데, 이 때는 활용도의 감시가 많은 도움이 된다. 관리자들은 CBM의 활용이 부서에 중요하다는 메시지를 전달하는 것 외에 이러한 감시에도 참여해야 한다. 뿐만 아니라, 관리자들은 부원들의 CBMs 활용을 북돋울 수 있는 보상체계의 검토책임도 맡고 있다(예, 다른 작업에 긍정적이거나 방해가 되는 것으로 간주되는 CBM 팀에 적용하는 등).

11.3.4 CBMs: 관리도구

CBMs는 지속적인 개선문화를 고취시키는데 있어 강력한 도구가 될 수 있으며, 관리자들은 CBMs의 개발과 활용을 가속화하기 위한 많은 조치를 취할 수 있다. 그 중 한 가지가 보상체계다. 경영진은 CBMs의 활용을 증진시키기 위해 부원들을 위한 여러 가지 보상제도를 마련할 수 있다(즉, 공지, 승진, 보상금 등). 만약, 경영진이 현재의 환경을 부원들이 지속적인 절차전개에 약간의 시간을 할애하고, 프로세스 중에 새로운 질의를 하면서 프로세스에 능동적으로 참여하도록 북돋우는 환경으로 변화시킨다면 부원들은 CBMs의 활용에 지원과 양해가 있음을 알게 될 것이다.

앞서 말한 바와 같이, 문서화된 CBMs의 존재만으로는 품질을 개선시키기에 충분하지 않다. 이것과 똑같이 중요한 것은 바로 개발프로세스다. 정상적인 사람들은 CBMs의 작성과 수정에 참여하길 원한다. 그러므로, 프로세스의 주체가 CBMs의 작성과정에서 배제되어 있다면(예, 외부 "전문가"를 고용해 CBMs를 작성하게 하는 등), CBMs의 활용가능성은 크게 줄어든다. 그 밖에, CBMs 준수에 대한 내부감사시스템은 CBM 문서가 부원들의 실무에 적합하도록 하는데 도움을 준다.

CBMs의 실효성의 한 가지 척도는 새로운 통계부원의 피드백 형식에서 비롯된다. CBMs을 제공하지 않은 다른 기관에서 전임온 신입부원들은 단시간에 별 어려움없이 팀에 참여하고, 절차를 학습하며, 설명서를 작성할 수 있었다고들 말한다. 이것은 CBMs가 이러한 프로세스에서 중요한 요소였음을 말해주는 것이다. 또 다른 사례들을 보면, CBMs은 유실된 자료의 대치절차에 대해 부원들의 동의를 쉽게 이끌어 냈다. CBMs의 개발은 표준화된 소프트웨어에 개선이 필요함을 확인시켰다. CBMs 체크리스트의 활용은 조사자료의 표집 및 가중치 부여와 같은 프로세스에서 중요한 단계들이 무심코 간과될 가능성을 감소시켰다. 그리고, 왜 특정한 절차를 따라야 하는가에 대한 CBM의 상세한 설명은 신입부원이 권고된 프로세스의 필요성을 이해하고, 이를 인정하는데 도움을 주었다. 뿐만 아니라, 외부 고객들은 이러한 자세한 설명서를 통해 이행되는 프로세스를 보다 잘 이해할 수 있었다.

11.4 결론

Colledge and March(제22장)가 수행한 조사에서 지적된 바와 같이, 통계기관들은 자신들의 사명에 있어 품질을 중요한 한 가지 요소로 만들었다. 기술혁신과 개선된 기법들은 계속해 등장한다. 그러나, 조직화 및 관리에 대한 오늘날의 접근방식은 품질개선노력에 새로운 원동력을 부가할 수 있다. 많은 기관들은 본 절에서 제시한 제안에 대해 두 가지 중요한 방식으로 벗어나 있다. 이들은 CBMs 문서의 개발과 같은 새로운 도구를 가진 체계화된 계획을 활용하지 않고, 전 부원들의 참여와 함께 최고경영진이 이끄는 지속적인 품질개선에 대한 체계적인(systematic) 접근방식을 가지고 있지 않다. 어떤 한 부원이 현재 사용하고 있는 것에 비해 보다 새롭고 나은 절차를 통합하는 것은 비록 시간소모적인 과정이라 하더라도 전체 기관을 위해서는 개선된 효율을 창출할 수 있는 한 가지 단계다.

기관의 구조를 지속적인 품질개선을 위해 노력하는 것으로 변화시킨다는 것은 쉬운 일이 아니다. 어쩌면, 경영행동을 수정하는데도 1, 2년이 걸릴 수 있다. 그러나, 그렇게 되어야만 전 부원들이 체계적인 방식으로 품질개선에 접근할 것으로 기대할 수 있다. 그러므로, 이러한 변화에는 다년간에 걸친 부원들의 시간, 비용의 투자와 특히, 무엇보다 개개인의 경영참여가 들어 있다. 이러한 구성요소들은 기관이 지속적인 품질개선의 목표를 기본적인 모든 작업에 통합하기 전에 요구되는 것이다.

제 12 절 국가 통계기관들의 품질정책, 표준, 지침 및 권고관행

Michael Colledge
호주 통계청
Mary March
캐나다 통계청

12.1 서문

본 절에서는 통계정보의 수집, 편찬, 분석, 간행을 담당하는 정부 통계기관들의 품질에 관한 정책(Policy)과 표준(Standard), 지침(Guideline) 및 권고관행(Recommended Practice) 간단히 말해 "품질관행들(Quality Practices)"에 대해 다루었다. 여기에는 세계 16개국의 국가 통계기관들의 조사를 바탕으로(별첨 참조) 기존의 품질관행들을 요약하고, 이러한 관행들의 기원과 실효성에 대한 논의가 들어 있다. 이러한 맥락에서 본 절에서의 "품질"은 고객의 관점에서는 통계상품과 서비스의 적절성, 접근성, 정확성, 적시성 및 비용성을 말하는 것이고, 응답자의 시각에서는 자료수집의 용이성, 비용성 및 비밀성을 의미하는 것이다. 다시 말해, 품질은 자료의 정확성만을 의미하는 것이 아니라, 이보다 훨씬 폭넓은 의미로 해석된다. 한편, 본 절은 가장 중요한 통계적 성격과 어떤 의미에서는 통계기관들에 독특한 관행들에도 역점을 두고 있다. 따라서 본 절에서는 품질관리와 관련된 일련의 모든 활동을 다루지는 않았으며, 특히 종합품질관리에 관한 문헌들에는 주로 많이 등장하는 것이지만, 통계기관과 그 밖의 기관에 광범위하게 적용될 수 있는 관행과 관련된 계획이나 조직 및 인적구성에 대한 사항들은 모두 배제하였다.

품질관행들은 "표준"과 "지침"을 통해 강력하게 시행될 수 있는 "정책"에서부터 선택적인 "권고관행"에 이르는 하나의 연속선상에 놓여져 있는 것으로 보여진다. 보다 자세히 말하자면, 각각의 용어들은 다음과 같은 의미로 사용된다.

정책 : 기관 경영진의 지시사항으로, 법적으로 요구되거나 기관 자체에서 정할 수 있으며, 예외 없이 모두 따라야 한다.

표준 : 예외 없이 거의 모두가 따라야 하는 관행이다. 이를 어기는 것은 바람직하지 않으며, 이 때는 상급관리의 승인이 필요하다. 표준이 충족되지 않았을 때는 시정조치를 취해야 한다.

지침 : 특별한 사유서가 없을 경우 따라야 하는 관행

권고관행 : 목적에 따라 기관이 장려하는 관행으로 강제적인 것은 아니며, 문서작성이나 변명 없이 어겨도 상관없다.

통계기관들 간에는 서로의 품질관행에 관심이 상당히 많다. 이러한 품질관행들은 모든 측면, 특히 그 중에서도 조정과 조화가 필요한 측면에서 통계 프로세스와 상품의 품질을 개선하고, 규모의 경제를 뒷받침하는데 중요한 하나의 메커니즘으로 보여진다. 품질관행들은 예를 들면, 상품이 고객의 요구에 부합되도록 하고, 기관이 보유한 자료를 고객이 원하는 어떤 형식으로든 확실하게 볼 수 있고 쉽게 접근할 수 있도록 하며, 각각의 조사작업에서 얻은 통계치들이 상호 일관성이 있도록 하고, 자원과 자료에 대한 내부의 효율적인 공유를 통해 비용을 최소화할 수 있게 하는데 도움을 줄 수 있다. 그 밖에도 품질평가를 제공함으로써 기관의 프로세스와 상품에 대한 고객의 인식을 증대시키고, 이로 인해 기관의 명성을 높이는데도 기여한다. 미국 농업통계청(NASS; National Agricultural Statistical Service) (1983)의 장기계획수립에 관한 보고서를 인용하면, "앞으로 있을 통계보고서비스의 모든 활동들에 최우선적인 요소는 반드시 통계표준이라야 한다."

그러나, 품질관행들이 표준화의 의미를 내포하고 있는 한, 여기에는 잠재적인 침체(downside) 가능성이 있다. 즉, 개인고객의 입장에서 볼 때는 "비표준적인" 지리적 영역에 관한 통계와 같이, 품질표준이 고객의 특정한 요구사항과 일치하지 않을 수도 있다. 따라서 만일, 품질이 고객의 편에서 정의된다면, 오늘날의 품질관리이론에 따라, 품질관행들의 도입과 시행은 반드시 개인고객들의 요구와 균형을 이루어야 한다. 그리고, 통계기관 내부의 시각에서 보자면, 시행되는 품질관행들은 국지적인 최적화(local optima)보다는 광역적인 최적화(global optima)를 지원하고, 개인의 의사결정을 축소시키기 때문에, 부원들은 불필요한 관료주의와 분명한 소득이 없는 과외의 업무를 포함하고 있는 것으로 인식할 수 있다. 앞서 말한 보고서에서 NASS(1983)는 "한 사람의 표준은 다른 사람에게는 장해가 된다"고 지적하고 있으며, 업무표준 제거에 관한 Deming의 유명한 14가지 포인트 중 하나는 비록 종업원의 업무 할당량에 관한 것이긴 하지만, 이 또한 개인의 창의력을 허용하면서, 품질관행들 간의 균형을 깨뜨릴 필요성을 지적하고 있다.

실행상의 문제에 대비하여 잠재적 이익에 가중치를 부여하는 것은 통계기관들이 품질관행들에 대한 각자의 접근방식에 유동적인 입장을 취할 수 있게 해 준다. 일례로, 캐나다 통계청은 자체의 품질지침(Statistics Canada, 1987)에 대해 다른 기관들의 수많은 요구들을 받아들이고, 1993년 조사에서 증명된 바와 같이 이것을 내부의 중요한 문서로 관리하고 있다. 이 조사에서는 품질지침에 대해 단순히 상투적으로 말한 사람은 거의 없는 것으로 나타났다. 이 예는 품질관행을 바라보는 태도의 전형이다. 이 품질관행들은 많은 수요가 있지만, 대개는 활용되지 않고 있다.

12.2 품질관행 조사

통계기관들의 비확률표본을 선택하였다. 21개 설문지를 우편발송하였으며, 15개의 응답을 받았다. 그리고, 무응답에 대한 추적조사는 없었다. 이후, 통계규정을 제정하는 두 기관이 접촉하였으며, 이 중 한 곳에서 추가적인 정보를 제공하였다. 각각의 기존의 문서화된 품질관행에 대해 응답자들에게 다음과 같은 내용의 질의를 하였다.

- 관행의 명칭, 주목적의 요약, 간략한 설명
- 적용분야(들) (목록에서 선택)
- 유형(법률, 정책, 표준, 지침, 권고관행에서 요구되는 정책)
- 도입시기, 원천과 개발에 대한 설명
- 준수도, 적용 및 시행에 대한 설명
- 해당 관행이 기관에서 가장 의미있고, 효과적인 관행들 중 하나인지 여부

응답자들에게는 효과적이지만, 아직 문서화되지 않은 관행들과 현재 개발중인 관행들에 대해 목록을 작성하고 간략히 설명하도록 하였으며, 기타 다른 설명도 덧붙이도록 요청하였다.

처음부터, 품질관행들에 관한 질문이 순탄하거나 쉽게 답할 수 있는 것이 아니며, "품질관행"을 구성하고 있는 것이 무엇이며, "정책", "표준", "지침", 그리고 "권고관행" 사이의 차이점을 무엇인지에 관해 상당한 변동이 있을 거라는 사실은 예상하고 있었다. 따라서, 이러한 문제에 대처하기 위해 호주 통계청과 캐나다 통계청에 실제로 있는 관행들을 기초로 설문지에 답변예시를 기재하였다.

12.3 결과의 개괄적 요약

16개 기관들이 답한 관행은 총 179가지였으며, 그 중 41가지를 "자신들의 5가지 가장 의미있고 효과적인 관행"에 포함되는 것이라고 지적하였다 (아래서부터는 간단히 "가장 효과적인 관행"이라고 함). 설문에서는 전체 적용범위에 대해 요청하지 않았기 때문에 이 기관들이 평균 11가지의 품질관행을 가지고 있다고 해석할 수는 없다.

품질관행의 유형 및 적용분야별 분포는 표 12.1에 나와 있다. 이 표를 보면, 가장 일반적으로 등장하는 유형은 "표준"이고, 가장 적게 등장하는 유형은 "권고관행"임을 알 수 있다. 일반적으로(단, 예외는 있음), 정책들은 아마 시행 및

감사에 대한 부수적인 요건을 반영하기 때문에 보다 세분화된 초점이 맞춰진 경향이 있으며, 반면 지침들은 대개 보다 일반적이어서 광범위한 적용분야를 포함하고 있다.

관행 유형 및 적용분야별 준수도 분포는 각각 표 12.2와 12.3에 나와 있다. 예상한 바와 같이, 준수도는 관행 유형에 따라 달랐으며, 법률에서 요구하는 관행의 준수도가 가장 높고, 권고관행의 수준이 가장 낮았다. 또, 비밀유지 및 자료배포 분야에서도 상당히 높게 나타났는데, 이 두 가지는 일반인과의 접촉과 관련되어 있다. 반면, 분류와 텍스트 내용 및 표현방식 분야에서는 낮게 나타났는데, 이 모두는 오차의 개념적 통합과 기록에 목적을 둔 관행과 관련이 있다.

도입시기별 관행분포는 그림 12.1에 나와 있다. 이를 보면, 중간시기는 1990년이고, 평균시기는 1980년임을 알 수 있다. 평균시기는 일부 기관들이 장기수립된 관행들을 포함시켜 중간시기보다 상당히 빠른 것으로 나타나 있다. 가장 오래된 관행은 미국 농업통계청의 비밀유지 절차에 관한 관행으로 이것은 1909년으로 거슬러 올라간다.

표 12.1 주요적용분야 및 유형별 관행

주요 적용분야	유형						
	정책(법)	정책	지침	표준	권고관행	지정없음	합계
고객 요구사항 수립, 시장조사	0	2	1	4	1	1	9
규격/사용 모집단, 자료항목	0	1	0	9	0	0	10
규격/사용 분류	2	3	1	11	0	1	18
수집도구 선정, 시험	0	0	0	0	0	0	0
설문설계	0	4	0	1	0	1	6
전자형식설계 (CATI, CAPI)	0	0	0	0	0	0	0
표본틀, 사업자등록부	3	0	1	5	0	0	9
표본설계, 선정	0	0	0	2	1	0	3
조사시스템 및 절차 시험	0	0	0	0	0	0	0
응답자 증첩 최소화, 측정	1	7	2	0	0	1	11
응답오차 최소화, 측정	0	0	0	0	0	0	0
비밀유지	6	1	0	2	0	0	9
자료수집, 포획, 추적	0	6	0	4	1	0	11
편집, 대치	0	0	0	1	0	0	1
가중치, 추정, 특이점	0	0	1	0	3	0	4
절기별 조정, 추세 산출	0	1	2	0	1	0	4
자료분석	0	0	1	0	0	0	1
택스트 내용, 표현	0	8	0	4	0	1	13
도표 표현	0	1	0	2	0	0	3
그래픽 표현	0	2	1	1	0	0	4
공간적 표현	0	0	0	0	0	0	0
자료 배포	3	10	6	3	0	0	22
데이터 및 메타데이터 관리	2	2	2	1	2	2	11
조사 및 프로그램 평가	0	1	0	3	0	0	4
조사설계 (일반)	1	0	5	2	1	0	9
시스템 설계	0	1	0	1	1	1	4
조직 계획	0	1	0	2	3	0	6
인원, 교육	0	2	0	0	1	0	3
지정 프로그램	2	0	1	0	1	0	4
전분야	20	53	24	58	16	8	179

12.3.1 통계법

각 기관은 기본적인 통계법의 체계 내에서 운영되며, 이 법은 예컨대 인구센서스의 시기 및 내용을 정해놓은 것과 같이 특별한 통계프로그램에 관한 하나 이상의 부가적인 법령들과 함께 운용된다. 통계법들은 모두 통계기관과 일반인 특히, 일반인과 응답자들 간의 상호작용에 관한 일부 품질관행을 구현하기 위한 것으로 생각할 수 있다. 또, 전부는 아니지만 대부분의 법령들은 자료배포에 관한 기관의 의무사항, 개인 응답의 비밀유지, 자료수집의 목적 명시에 관한 요구, 그리고 답변 및 정확한 정보제공에 관한 응답자들의 권리와 의무에 관한 사항들을 명시하고 있다.

표 12.2 적용분야 및 준수도별 관행

주요 적용분야	준수도				
	높음	보통	낮음	지정없음	합계
고객 요구사항 수립, 시장조사	3	1	0	5	9
규격/사용 모집단, 자료항목	2	0	0	8	10
규격/사용 분류	3	4	0	11	18
수집도구 선정, 시험	0	0	0	0	0
설문설계	4	1	0	1	6
전자형식설계(CATI, CAPI)	0	0	0	0	0
표본틀, 사업자등록부	5	2	1	1	9
표본설계, 선정	1	1	0	1	3
조사시스템 및 절차 시험	0	0	0	0	0
응답자 중첩 최소화, 측정	5	3	0	3	11
응답오차 최소화, 측정	0	0	0	0	0
비밀유지	7	0	0	2	9
자료수집, 포획, 추적	5	3	1	2	11
편집, 대치	1	0	0	0	1
가중치, 추정, 특이점	1	2	1	0	4
절기별 조정, 추세 산출	1	2	0	1	4
자료분석	1	0	0	0	1
택스트 내용, 표현	2	6	1	4	13
도표 표현	1	0	0	2	3
그래픽 표현	0	1	1	2	4
공간적 표현	0	0	0	0	0
자료 배포	10	1	1	10	22
데이터 및 메타데이터 관리	5	3	0	3	11
조사 및 프로그램 평가	1	0	0	3	4
조사설계 (일반)	1	2	1	5	9
시스템 설계	2	0	0	2	4
조직 계획	3	2	0	1	6
인원, 교육	1	1	0	1	3
지정 프로그램	1	2	0	1	4
전분야	66	37	7	69	179

몇몇 국가의 통계법에는 품질관행에 대해 추가적으로 직접적인 언급이 있다. 이러한 면에서 보면, 최근에 개정된 Statistics Act (Statistics Finland, 1994)가 아마 가장 포괄적인 법령 중 하나일 것이다. 여기에는 다음에 관한 내용이 명시되어 있다.

표 12.3 준수도 및 유형별 관행

준수도	유형						
	정책(법)	정책	지침	표준	권고관행	지정없음	합계
높음	14	23	7	18	3	1	66
보통	4	7	9	6	10	1	7
낮은	0	2	2	2	1	0	7
지정없음	2	21	6	32	2	6	69
전체	20	53	24	58	16	8	179

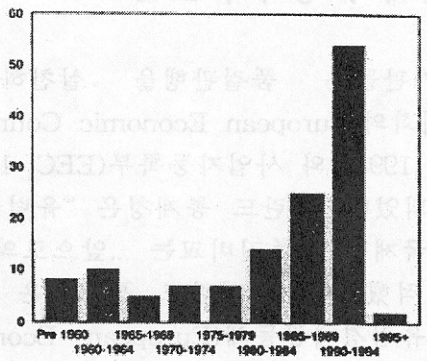


그림 12.1 도입시기별 관행분포

- *자료수집도구:* "통계를 위한 자료수집은 일차적으로 다른 관계에서 수집된 자료를 이용한다".
- *정의 및 분류:* "자료의 호환성을 위해서는 통계의 편찬(compilation)에 가능한 한 통일된 개념과 정의 및 분류체계를 사용한다."
- *응답자 부담 가중 최소화:* "통계기관은 수집할 자료를 정하기 전에 ..응답자와 먼저 교섭한다.. 교섭은 응답자의 관점이 고려될 수 있도록 가급적 빨리 한다.. 통계를 편찬하는 기관은 응답자들에게 통계편찬에 필요한 자료만 제출하라는 요청이 이루어지도록 한다. 자료는 경제적이고, 가능한 한 응답자에게 불편을 끼치지 않는 방법으로 수집한다."

■ *자료내용 및 배포:* "통계는 가능한 한 신뢰가 있어야 하고, 사회적 상황과 발전에 관한 올바른 설명이 제시되어야 한다. 통계는 완료 후 가급적 빨리 간행한다."

■ *시스템 설계 및 자료관리:* "자료는 좋은 통계적 관행에 따라 처리한다. 통계를 편찬하는 기관은 자료의 도난뿐 아니라, 자료가 무단으로 처리, 사용, 파괴 및 변경되지 않도록 적절한 보호조치를 취한다."

다음과 같은 두 가지 공통적인 품질관행은 대개 법에 의해 지탱되지만 그렇다고, 반드시 통계법에 그와 같은 사항이 들어 있는 것은 아니다. (1) 지정된 상위 공공단체에 새로운 조사나 진행중인 조사의 중대한 변경에 관한 제안서 제출요건, (2) 통계프로그램 특히, 산출되는 통계의 적절성에 관한 것을 검토하고, 이에 따라 통계기관을 자문하기 위한 일정한 형식의 국가 "감시"체계 수립이 그것이다. 각각의 특정한 예는 12.4절에 논의되어 있다.

12.3.2 기관내부 및 국제적 정책과 표준

유럽의 몇몇 응답기관들은 품질관행을 실천하는 유럽통계 사무국(EUROSTAT)의 유럽경제회의(European Economic Council) 규정을 참조하였다. 그 중, 표준통계단위(EEC, 1993a)와 사업자등록부(EEC, 1993b)의 사용에 관한 두 가지 규정은 그대로 인용되었다. 핀란드 통계청은 "유럽공동체는 품질에 중요한 효과를 미친다," 그리고 "국제적인 품질비교는 ...앞으로의 통계에 매우 중요하게 될 것이다."는 사실을 지적했으며, 네덜란드 통계청은 표준산업분류의 활용에 관한 자신들의 정책은 유럽경제공동체(European Economic Community)에서 비롯된 것으로, 장기적인 국가적 관행이 되어 있다고 하였다. 또한, 비슷한 방식으로 미국의 통계기관들은 행정관리예산국(Office of Management and Budget)이 품질관행으로 정한 표준들을 참고하였다.

품질관리 및 품질평가에 관한 표준으로 국제표준화기구(International Organization of Standardization) (1994) 9000 시리즈를 참고한 예는 없었다. 이러한 표준들이 특별히 사회봉사단체(service organization)에 관한 지침을 다루고 있다는 사실에 비하면 이것은 실로 놀라운 사실이다.

12.3.3 개발, 이행, 효과 및 준수

많은 기관들에서 품질관행을 맡고 있는 독자적인 그룹이 없는 것으로 나타났다. 또한, 법에서 요구하거나 혹은 분류와 관련된 정책을 제외하면, 품질관행의 개발과 도입은 대개 체계적인 프로세스의 결과이기보다는 오히려 우연한 현상 기관 내부에서 개인과 그룹 중 주도권을 쥐는 기회의 문제처럼 보인다. 예를 들면, 호주 통계청은 한 부원이 지난 10년 동안의 회의기록에서 고위관리직의 의사결정을 발췌하여 2년 넘게 이를(몇몇 품질정책을 포함하여) 공식적인 정책으로 전환하였으며, 이는 일괄 수용되었다고 지적하였다. 이것은 어쩌면 하나의 상급단체에서 개개의 건으로 논의했던 정책보다 훨씬 많은 수일 수도 있다.

응답자들이 가장 효과적이라고 답했던 품질관행들은 명백한 비밀유지에 관한 진술과 함께 여러 적용분야에 광범위하게 걸쳐 있다. 품질관행의 효과는 관행 고유의 유용성과 준수도라는 두 가지 상관요소로 바라볼 수 있다. 놀랍게도, 응답자들은 높은 준수도를 가장 효과적인 관행이라고 지목할 수 있는 요건으로 간주하지 않았으며, 이는 응답자들이 판단의 기준으로 고유의 유용성에 보다 많은 역점을 두고 있음을 제시해 준다. 이러한 결론은 "품질관행은 그다지 잘 이용되지 않고, 개념개발에서 실행으로의 변화를 만들지도 않는다"는 말과 일치한다. 뿐만 아니라, 품질관행은 초기의 열정이 사라진 후에는 종종 낡은 것으로 전락한다는 지적도 있었다. 이를 방지하기 위해, 미국 농업통계청에는 정책과 표준의 문서화 및 관리에 관한 정책이 있다. 그리고, 누락된 입력자료의 수로 판단하건대, 많은 응답자들은 준수도를 평가하는데 어려움을 느꼈으며, 이는 정책에 대한 평가도 마찬가지였다. 이것은 품질관행의 감시와 시행은 대개 제한적이거나 실재하지 않음을 제시해 준다.

12.3.4 품질관행과 윤리

몇몇 품질관행들은 조사전문가들에게 요구되는 윤리적 행동양식에 대해 언급하고 있는데, 예를 들면 표본단위는 자료수집비용과 상관없이 표집층 내에서 무작위로 선별해야 하고, 면접원은 회수하지 못한 응답을 만들어내거나 응답에 영향을 미치려 해서는 안되며, 각각의 응답은 비밀을 유지하고, 통계는 간행되기 전에 미리 유출되거나 개인의 이익을 위해 사용되어서는 안된다 등의 내용이 있다. 일례로, 미 노동통계청에는 "고용인들은 자료원이나, 자료수집방법, 응답자로부터 받은 자료 또는 행정보고서 기재사항 등에 대해 고의로 허위진술해서는 안된다"고 규정해 놓은 자료수집 보전정책이 있다. 또, 미

인구조사국에는 동일한 주제로 하나의 목적이 면접 위조를 검출하려는 것이라는 재 면접 정책이 있다.

이와 같이, 윤리적 요건과 품질관행의 일치가 부분적으로 발생하는 이유는 자료수집 및 처리에서의 높은 윤리적 기준이 편향되지 않은 즉, 정확한 통계를 산출하기 때문이다. 그리고 또 한 가지 이유는 "품질"은 매우 광범위하게 정의되어 있으며, 따라서 비밀유지 위반이나 사전유표와 같은 비윤리적 행위로 인해 응답자와 사용자에게 발생하는 비용까지도 고려하고 있다는 것이다. 실제로, 핀란드 통계청에서는 직업윤리 정책을 하나의 품질관행으로 기재하였다.

12.4 적용분야별 품질관행

다음 절에는 고객 요구사항에 관한 것을 시작으로 선택된 적용분야에서의 몇몇 관행에 대해 간략이 설명되어 있다. 여기서는 응답기관들이 "가장 의미있고 효과적인" 관행으로 지적한 것은 간단히 "가장 효과적인" 관행이라 하였다.

12.4.1 고객 요구사항 정립

사용자/자문그룹

프랑스 국립통계조사연구소(INSEE; Institut National de la Statistique et des Etudes Economiques)에는 산업영역별 고객그룹 도입에 관한 권고관행이 있다. 그 목적은 기관의 요구가 사업체에서 유용한 자료와 일치하고, 기관의 상품이 사업체의 요구에 적합하도록 함으로써 사업체와 통계기관 사이의 관계를 개선하기 위한 것이다. 비록, 이러한 점에서 공식적으로 문서화된 관행은 없지만, 캐나다 통계청에도 광범위한 "사용자 자문위원회"들이 있으며, 미국 인구조사국도 특정한 기법을 활용하거나 특별한 관심을 표명하는데 있어 외부의 전문가 자문그룹을 정기적으로 초빙하고 있다.

고객만족도조사 수행

몇몇 기관들은 고객만족도를 측정하기 위한 관행들을 기재하였다. 예를 들어, 1994년 미국 노동통계청은 각 연방기관들의 고객서비스표준 제정을 요구하는 행정명령(Executive Order) 12862(The White House, 1993)의 일환으로 지속적인 거래기반(transaction-based)의 고객서비스조사를 처음 실시하였다. 그리고, 2년마다 노동통계청 사무국에서는 각각의 고개표본에 서비스제공 설문지를 보내고 있다(전화와 우편연락).

핵심고객 파악 및 연락

매월 핵심고객을 파악하여 방문하는 원칙으로 시행되는 뉴질랜드 통계청의 핵심고객 연락에 관한 표준은 가장 효과적인 관행으로 손꼽혔다. 이와 유사한 관행은 호주 통계청에도 있다.

안내데스크 질의문 작성 및 분석

뉴질랜드 통계청에는 고객 서비스부원이 특별히 제작된 중앙데이터베이스 시스템을 이용하여 2년 단위로 고객의 요구사항을 기록하고, 코딩하는 관행이 있다. 여기서, 알려진 것은 활동이 최고조에 달하는 시간대가 보다 한산한 시간대보다 코딩의 신뢰도가 떨어진다는 사실이다(이것은 품질관리에 있어 일반적인 문제에 관한 하나의 특정한 예다).

12.4.2 통계단위와 자료항목 및 분류

보고된 통계단위와 자료항목에 관한 대다수의 관행들은 기업경기 통계와 관련이 있다. 이러한 관행들은 효과적으로 이행되기 어렵다고 알려져 있다. 예를 들어, 네덜란드 통계청의 응답을 보면, 기업경기 통계단위에 대한 내부의 합의를 도출하는데는 약 5년의 논의가 필요하며(국가 고객 및 기타 사용자들의 요구를 충족하는) 조정된 개념과 정의에 관한 표준도 역시 협의하는데 상당한 시간이 소요된다고 되어 있다. INSEE(프랑스)는 국가회계회의(National Accounting Council)에서 발표한 가장 효과적인 관행으로 간주되는 사전행동적(proactive) 정책을 가지고 있다. 이 정책의 목적은 통계적 요구사항이 국가회계계획 즉, 일반적인 회계원칙을 수정하는데 고려되고, 사업체와 기관의 통계전문가들이 동일한 용어를 사용하도록 하기 위한 것이다. 영국 국립통계청을 포함한 몇몇 기관들은 기업경기 통계에 관한 표준분류의 사용은 중앙사업자등록부(central business register)의 이용을 통해 강화될 수 있다고 지적하였다.

12.4.3 설문설계

설문설계에 관한 대다수의 관행들은 기업경기 통계의 수집과 관련이 있다. 여기에는 준수도가 높고, "형식은 쉽게 작성할 수 있도록 하고, 응답검증을 위한 추적조사의 필요성이 줄어들도록 사용자와 함께 협의하여 설계하자"는 취지의 호주 통계청의 관행이 있다. 이 관행에는 또한, 형식(즉, 설문지)은 "응답자들에게 전문적이고 효율적인 기관의 이미지를 제시하고, 다른 정부부처와 차별화될 수 있는 통합된 기관의 접근방식을 촉진해야 한다"고 명시되어 있다. 한편, 체코 중앙통계청에도 많은 기업들이 전체 요구사항이 담겨있는 하나의 (연간) 설문지를 받아보도록 하기 위한 기업경기조사요건의 조정에 관한 절차가 있다.

전자형식 설계에 관한 응답이 없는 것은 전자식 수집방법 및 기술이 품질관행이 개발되기에는 너무 실험적이고 새로운 것임을 말해 준다.

12.4.4 기록부(Registers)와 표본틀(Frames)

핀란드 통계청에는 개인의 모든 기록부에 대해 보관처, 내용, 용도, 삭제 및 보관절차가 문서화되고, 이를 공공에 알리도록 법적으로 요구되는 정책이 있다. 그 외, 이러한 일반분야에서 보고된 기타 관행들은 사업자등록부의 내용과 용도 및 유지관리에 대해 명백히 다루고 있다. 예를 들어, INSEE(프랑스)에는 사업체를 다루는 모든 정부기관들의 공통적인 참고단위로 기업과 지방단위에 하나의 상수인 등록번호를 부여하도록 법에서 강제하는 매우 효과적인 정책이 있으며, 이로 인해 등록정보의 저장공간이 축소되고, 행정자료의 교환과 사용이 용이하게 되어 있다. 또한, 네덜란드 통계청은 "동일 분야의 통계는 동일한 모집단을 기초로 하고, 다른 분야의 통계는 모집단을 중복하거나 빠뜨리지 않으며, 통계단위는 통일하여 정의되고 분류되도록" 하는 목적의 사업자등록부 표준을 정해 놓고 있으며, 영국 국립통계청(ONS)에는 경제공동체 규정에 따라 기록부 구조와 최소한도의 내용을 규정하는 정책이 있다(EEC, 1993a, 1993b).

12.4.5 응답자 부담 가중

조사요구의 검토

일부 기관들은 상부기관의 검토와 승인이 필요한 신규조사나 근본적으로 수정되는 조사에 대해 제안서를 요구하는 정책들을 기재하였다. 예를 들면, 프랑스에서는 INSEE가 조사를 반드시 National Council for Statistical Information에 등록하도록 되어 있다. 이 정책은 최근, 지정된 적절성과 통계품질규범에 대한 일치성을 만족하는 조사에 "일반적인 관심도 표시"라고 일컫는 품질지표를 부여하도록 하는 내용을 포함한 다른 정책으로 보완되었다. 한편, 호주의 경우에는 호주통계청(ABS)이 법적으로 반드시 새로운 조사들을 의회에 상정하도록 되어 있다. ABS 정책은 "이유를 불문하고, 의회에 상정되는 통계수집은 반드시 명백히 자발적인 원칙에서 수행되어야 한다"고 명시하고 있다. 또한, 영국 국립통계청은 사업체나 지방기관의 모든 신규조사나 혹은 기존조사의 상당한 변경에 Survey Control Unit의 허가와 장관의 승인을 받도록 하는 정책이 있다. 이 때, 제출되는 정보에는 조사에 응함으로써 응답자에게 발생하는 비용에 관한 평가가 포함되어 있다. 반면, 뉴질랜드에서는 통계에 관한 책임을 맡고 있는 장관이 새로운 통계조사를 승인하는 요건이 1993년에 폐지되었으며, 다만 "독립된" 검토는 여전히 남아 있다.

응답자 부담 가중 균일화

5개 기관은 응답자 부담 가중을 표집 모집단 전체에 균일하게 확산시키는 관행들을 기재하였고, 6번째 기관(핀란드 통계청)은 이러한 관행을 개발중이라고 답했다. 이 중, 4개 기관은 기업경기조사와 관련이 있고, 나머지 한 기관은 기업경기조사와 가계조사 양면에 모두 관련이 있다. 하나씩 살펴보면, 호주 통계청에서 가장 효과적으로 간주하는 정책에는 "동기화된(synchronized)" 표본추출에 관한 내용이 있다. 즉, 사업자등록부에 있는 각 통계단위에는 연속적인 주기를 갖는 조사 및 여러 조사들 간의 표본 중복을 관리하기 위해 무작위수가 부여된다. 그리고, 이와 유사한 정책은 스웨덴 통계청에도 있다. 네덜란드 통계청의 경우에도 가장 효과적이라고 답한 정책에 아주 다른 기술적 접근방식이 포함되어 있다. 즉, 신규조사의 표본추출은 진행중인 조사들 간에 응답자 부담을 최소화하여 사업체 단위를 선별할 수 있도록 편향없이 조정된다. 한편, INSEE(프랑스)는 연간 기업체조사를 총괄하여 표본 선별시의 협의조정에 관한 절차가 있으며, 뉴질랜드 통계청 지침은 소규모 사업체 단위가 주어진 시점에서 의무적으로 보고해야 하는 조사횟수를 2회로 제한하고, 하나의 조사에 주어진 가구가 연속으로 참여하는 시간간격을 2년 이상으로 규정하고 있다.

12.4.6 비밀유지

개개의 응답의 비밀을 유지하기 위한 정책에는 6개 기관이 답했다. 이 중 4개 기관의 정책은 가장 효과적인 것으로 평가되었고, 모두 준수도가 높은 것으로 나타났다. 전형적인 예는 미국 노동통계청의 정책인데, 여기에는 "비밀유지 헌장에 따라 BLS가 직접 또는 BLS의 후원아래 수집 또는 유지되는 자료는 통계적 목적으로만 사용되고, 허가된 사람만 접근할 수 있게 하는 식으로 관리한다. 공공을 위해 작성된 것으로 사전배포된 경제계열 자료는 공식으로 공표되기 전에 누설하거나 허가받지 않은 방식으로 사용되지 않을 것이며, 허가받은 사람만 접근할 수 있을 것이다"고 규정되어 있다. 비밀유지가 통계자료를 수집하는데 있어 중심적인 특성이 되어 감에 따라, 모든 기관들이 이러한 정책을 채택할 가능성은 매우 높다 하겠다. 이 정책이 보편적으로 기재되지 않았다는 사실은 무엇이 "품질관행"을 구성하는 가에 대한 해석상의 차이를 보여주는 것 같다. 예를 들면, 스웨덴 통계청은 "우리는 품질적 차원에서 특별히 비밀유지를 고려하지는 않는다. 우리가 알기로는 비밀유지에 관한 엄격한(외부의) 법령이 있다는 것이다. 그러므로, 우리는 아무런 선택권이 없는 것이 아니라, '게임의 법칙'을 고수할 뿐이다."라고 하였다.

12.4.7 수집, 추적, 포획 및 코딩

과세서류의 표본추출 및 자료포획으로 자체 자료를 수집하는 미국세청의 소득통계과(Statistics of Income Devision)는 문서관리에 관한 정책과, 모든 자료가 두 번 포획되고, 교정확인이 이루어지도록 요구하는 표준, 그리고 운영지침 뿐 아니라, 품질관리와 품질감사 체크리스트에 관한 접근방식의 개요도 함께 제공하는 전반적인 처리관리표준이 있다. 이 모든 관행은 가장 효과적인 것으로 고려되고, 준수도도 높다. 그리고, INSEE(프랑스)와 미국 인구조사국은 자료포획 및 코딩 품질관리정책이 있다. 또한, 이와 같은 효과를 가진 공식적인 정책문은 없더라도 품질관리는 캐나다 통계청에서도 널리 사용되고 있다. 캐나다 통계청에는 기업경기조사와 사회조사에 관한 무응답 측정의 기록 및 활용에 관한 정책도 있다. 여기에는 16가지의 무응답율이 정의되고 있는데, 그 중 13가지는 자료수집에 활용되고, 나머지 3가지는 추정에 활용된다.

12.4.8 계절별 조정 및 추세 산출

일부 기관들은 계절별 조정 및 성향 산출분야에 지침이나 권고관행이 있거나 혹은 현재 개발중이라고 답했다. 이들 대부분은 X11-ARIMA라고 불리는 하나의 특정한 방법에 대해 언급하였다. X11에 대해서만 언급한 호주 통계청의 응답을 보면, 이모든 관행들의 목적을 "다음을 통해 매체가 주요 경제지표들의 의미있는 특징들을 정확하게 보고하도록 지원하기 위한 것"으로 나타내고 있다. 계절별로 조정된 추정치보다는 추세의 움직임을 강조한다. 추세 추정치를 역사적 관점에 둔다. 충분한 증거가 쌓이기 전까지 반환점(turning point)에 관한 주장을 피한다. 민감도를 분석한다. 그리고, 추세가 깨지는 것을 확인한다.

캐나다 통계청은 "앞으로의 기준 기간에 대한 추정치는 다음을 전제로 수용할 수 있는 기관의 상품이다. 추정치는 구체적인 내용이 대중의 철저한 검증을 받을 수 있도록 확연히 드러나 있는 수학적 모델을 이용하여 산출된다. 이 모델에서 사용된 가정은 추정치와 함께 명확하게 정의되고 표현된다. 모델의 가정과 입력데이터는 누가 이용하더라도 동일한 결과를 얻을 수 있다."고 진술하는 예측에 관한 정책이 있다.

12.4.9 결과 자료품질

몇몇 기관들은 자료의 품질에 관한 정보가 고객에게 전달될 것을 요구하는 정책이 있다고 했다. 예를 들면, 모든 결과물에 방법론과 자료품질에 관한

설명서가 수반될 것을 요구하는 장기 수립된 캐나다 통계청 지침이 있다. 현재는 이것을 표준으로 상향조정하고, 이의 준수에 더 많은 관심을 기울도록 하는 검토가 진행중이다. 이와 마찬가지로, 미국 인구조사국의 정책은 표본조사에서 표준오차가 추정치와 함께 제시되도록 요구하고 있다. 그리고, 포괄적인 호주 통계청의 정책은 주로 비신뢰도를 근거로 자료의 금지에 관한 기준을 다루고 있는데, 이 또한 "모든 간행물에는 자료품질과 방법론에 관한 설명서가 포함되고, 표집오차와 비표집오차의 크기가 제시되어야 한다"고 규정하고 있다.

12.4.10 문자, 표, 그림 및 공간적 표현

대부분의 기관들은 표현방식(presentation)에 관한 정책들이 있었다. 예를 들면, 미국세청의 소득통계과(Statistics of Income Devision)는 문서작성지침(written guideline)과 스타일지침(style guideline)이 있고(Statistics of Income Division, 1991, 1994), 미국 농업통계청과 노동통계청에도 편집, 도표 및 챠트작성 지침이 있다. 이러한 관행들의 예에는 고객과 매체가 유의한 특징들을 정확하게 해석하는데 도움을 주는 핵심포인트 처리 등, 모든 간행물에 컴퓨터 시스템으로 지원되는 스타일 지침의 활용과 주요 경제지표의 특정한 형식에 관한 내용이 포함되어 있다. 하지만, 놀라운 사실은 공간적 표현에 관해 특별히 다루고 있는 관행은 하나도 없었다.

12.4.11 자료배포

자료배포 분야는 품질관행에서 많이 고려되고 있었는데, 이는 공적인 사항을 다룰 때 개선된 일관성에 대한 필요성을 반영하는 것으로 보인다.

정확성

몇몇 기관들은 통계결과를 발표하기 전에 상급기관의 검토와 승인을 요구함으로써 한결같이 높은 품질을 유지하도록 하는데 목적을 둔 관행들이 있었다. 예로, 미국 노동통계청은 출판을 책임지고 있는 Associate Commissioner의 사전승인을 통해, 자체의 편집, 도표, 챠트작성 지침 및 연방 출판물표준(publication standard)에 부합되도록 하는 출판정책이 있다. 반면, 네덜란드 통계청은 그 책임을 출판물 발행인에게 돌리는 체제를 유지하고 있다.

호주 통계청과 캐나다 통계청은 오차가 있는 자료의 발간 또는 금지에 관해 유사한 정책들이 있다. 전자의 목적은 "배포된 모든 자료에 품질표시가 있어 사용자들의 이해를 돕고", "표집오차 때문에 개개의 항목이 하나씩 금지되는 것은 아니고, 편향이 심한 자료는 발간되지 않게" 하도록 하기 위한 것이다.

배포일자

기관들이 배포일자를 사전에 공표하고, 지정된 일시 전에는 어떤 통계도 배포될 수 없다는 금지정책(embargo policy)을 따른 것은 일반적인 관행이다. 금지정책에 입각한 일자의 사전공시에 관한 조건 및 일반적인 매체 처리와 관련된 관행에 대해서는 7개 기관이 응답하였다. 그 중 전형적인 예는 호주 통계청의 일반적인 금지정책에 두 가지 예외를 규정하고, 다음에 관한 조건을 명시해 놓은 사전배포 및 "lock ups"에 관한 정책이 있다. (1)정부부처에 지정된 출판물의 사전배포, (2)"lock-up" 상황에서 정부관리의 지정된 출판물 확인, 이 때 참석자는 금지가 해제되기 전까지는 방을 떠나거나 외부와 연락할 수 없다. 미국 노동통계청은 "민감한 배포의 비밀을 유지하는 것과 그것을 널리 알리는 것 사이의 긴장상태를 객관적으로 해결하기 위한" 유사한 정책이 있다. 여기서는 공식적인 발표 전에 lock up 상황에서 매체가 제어하는 제한된 자료접근이 허용된다. 발표의 동시성에 관한 일반적인 규칙의 하나의 예외는 미국세청의 소득통계과(Statistics of Income Devision)에서 조세기록을 다른 사용자들보다 60일 먼저 주요 고객(Office of Tax Analysis, Congressional Joint Committee on Taxation)에게 공개하는 것이 있다.

접근

미국 노동통계청(1994a)은 새롭게 제정된 고객서비스지침에 대해 답했는데, 여기에는 배포일정, 정보 가용성, 상품의 사용 및 제한, 담당자 연락처에 관한 사항이 명시되어 있다. 이러한 지침들은 각 연방기관들의 고객서비스표준 제정 및 이행을 요구하는 행정명령(Executive Order) 12862(The White House, 1993)에 따라 개발되었는데, 국제기구에 통계정보를 제공하는 것은 행정관리예산국(Office of Management and Budget) 표준을 따른다. 한편, 스웨덴 통계청은 고객들이 필요한 자료를 빠르고, 효율적으로 찾도록 해 주는 통합출판시스템정책을 자신들의 가장 효과적인 관행이라고 하였고, 호주 통계청은 데이터 가시성 및 접근성을 크게 개선하는데 목적을 둔 지속적인 데이터관리에 대해 설명하였다(12.5절에 상세 설명). 호주 통계청은 또한, 마이크로데이터 배포에 관한 조건을 다루고 있는 정책이 있다고도 하였다. 이 정책에는 "식별이 불가능한 단위 기록들은... 비밀유지의 제약으로 유용한 파일을 생성할 수 없거나 이에 대한 요구가 없다면 대부분의 사회조사에서... 그리고, 고객의 수요와 의도된 자료의 용도에 대한 엄격한 평가를 전제로 하는 경우 다른 수집자료들로부터... 배포될 것이다"고 명시되어 있다. 이 정책은 또한, "의도된 용도에 다른 자료원이 보유하고 있는 자료와 일치하는 자료가 있는 경우에는 이의 배포가... 허가되지 않을 것이다"는 설명을 포함하여, 계속해 배포에 관한 조건을 규정하고 있다.

12.4.12 데이터 및 메타데이터 관리

연속성

조사절차를 변경하는 동안 중요한 시간대의 연속성이 유지되도록 하는 것을 기본적인 목적으로 가진 관행은 두 가지가 기재되었다. 호주 통계청 정책은 조사의 "병렬 시행(parallel run)" 내지는 변경에 기인한 통계적 불연속성의 규모 추정을 요구한다(변경의 일련성과 규모에 따라).

문서화

메타데이터 작성과 관련한 관행들 중에는 공유 데이터베이스에 조사문서의 저장 에 관한 INSEE(프랑스)의 권고관행이 있는데 이것은 개념과 데이터항목 정의 및 처리방법의 합리화와 표준화를 목적으로 하고 있다. 이 시스템은 현재 가구조사에만 제한되어 있다. 캐나다 통계청은 이와 유사한 목적을 가진 "통계자료문서화시스템"의 사용에 관한 정책이 있다고 하였다. 그리고, 미국세청 소득통계과(Statistics of Income Devision)는 행정처리기록의 상세 및 처리와 포괄적인 통계자문활동 기록의 제공에 관한 정책이 있다.

중앙공유 데이터베이스

영국 국립통계청은 기관에서 발행한 모든 데이터를 일부 메타데이터와 함께 보관하기 위한 중앙공유 output 데이터베이스의 활용에 관한 지침이 있다. 이 지침의 목적은 다른 형식과 다른 매체로 생성된 output이 일관된 정보를 가지도록 하고, 표준형식테이블 생성시설을 제공하며, 공공의(기밀하지 않은) 자료접근을 허용하는 것이다. 현재 메타데이터 관리시스템을 개발중이라고 답한 호주 통계청과 뉴질랜드 통계청을 비롯한 다른 기관들은 output 데이터베이스의 활용에 점점 많은 역점을 두고 있다고 하였다.

12.4.13 조사 및 프로그램 교육

스웨덴 통계청은 기관에서 생산하는 각 상품의 전반적인 품질의 변화에 대해 매년 자기평가를 요구하는 정책이 있으며(Eklöf and Lindström, 1995), 이를 가장 효과적인 관행이라 하였다. 이 정책에 따르면, 결과물은 중앙단에서 총괄적으로 검토 및 추적되고, 품질개선 우선순위를 정하는 기본원칙을 제공한다. 캐나다 통계청은 프로그램 목적과 이 프로그램이 고객의 요구를 충족하는 정도를 검토하고, 프로그램 설계와 그 손익을 잠재적 대안과 비교평가함으로써 상품의 적절성과 효과가 지속될 수 있도록 하기 위해 현재 시행중인 프로그램 교육과정이 있다. 영국 국립통계청의 조사관리정책은 "정기적으로 수행되는

사업체 또는 지방기관의 모든 조사가 지속되기 위해서는 기관장(minister)들의 정기적인(조사횟수에 따라 최소 3년 내지 5년 단위로) 철저한 검토와 승인이 있어야 한다... 조사의 필요성은 매년 검토되어야 한다"고 규정하고 있다.

12.4.14 일반적인 지침

많은 기관들 특히 그 중에서도 예를 들어, 미국 국립교육통계센터(National Center for Education Statistics) (1991, 1992)와 같은 도급자를 이용하는 기관들은 광범위한 조사 프로세스에 적용되는 일반적인 조사표준/지침이 있다고 하였다.

12.5 일반적인 관측 및 주제

품질관행의 조사는 몇 가지 한계가 있다. 첫째, 이것은 확률표본을 기초로 하지 않았으며, 모든 통계기관들을 대표하는 것으로 간주할 수도 없다. 다만, 통계자료수집과 처리 및 배포에 있어 선도적인 위치로 인정되는 기관들은 몇몇 있었다. 둘째, "품질관행"을 구성하는 것이 무엇인가에 대해 기관들 간에 해석상의 차이가 있었다. 품질이 갖춰짐에 따라, 오늘의 품질관행은 내일의 진부한 활동이 되어간다. 예를 들어, 실제로 세계의 모든 기관은 확률표본을 사용하고, 응답자들의 데이터를 편집하며, 무응답에 대해 조정한다. 이러한 것들은 본질적으로 품질관행으로 고려되지 않는 일상적인 절차의 공통된 일부분이다. 그러나, 프로그램 개발단계와 비교해 다른 단계에 있는 기관들은 품질관행이 무엇이며, 진부한 활동의 일부는 무엇인가에 대해 다른 인식들을 하고 있다. 셋째, 기관내부에서 조차도 품질 설문에 관한 응답들이 부원들의 특정한 관심사에 따라 다소 다르다는 것을 추정할 수 있다. 이러한 제한 속에서 조사결과는 다음과 같은 몇 가지 일반적인 관측을 뒷받침한다.

정보공유에 대한 응답과 의지를 기초로 판단하건대, 통계기관들은 통계적 통합과 조정에 관한 하나의 원칙으로 그리고, 규모의 경제와 품질개선을 위해, 품질관행을 수립 및 활용하는데 많은 관심이 있다. 기관에서 정한 관행이 없는 경우, 부원들은 대개 "지난 번에 했던 것"을 바탕으로 맹목적으로 지역적인 관행을 답습한다.

통계기관들의 "품질"을 구성하는 것에 관한 시각은 주로 정확성에 초점을 두는 것에서부터 적절성, 적시성, 가용성, 비용요소(응답자에 대한 비용 포함) 등을 포함하는 보다 폭넓은 견해에 이르기까지 최근 몇 년간 상당히

다각화되었다. 또 이것은 보다 고객서비스 지향적으로 되었다. 품질은 "용도에 영향을 미치고 사용자들이 중요하다고 생각하는 통계서비스의 모든 측면이다"(Eklöf and Lindström, 1995). 하지만, 기관들은 여전히 결과데이터가 고객이 대개 스스로 할 입장에 있지 않은 용도에 적합한 정확성을 가져야 한다는 각별한 책임감을 여전히 인식하고 있다.

기관 내부의 품질관행의 수와 성격은 다음 예에 나타난 바와 같이 조직의 구조와 관련이 있다.

- INSEE(프랑스)와 같은 일부 기관들은 모든 부원들이 통과해야 하는 강도높은 교육프로그램이 있으며, 그 결과 기술적 표준과 지침에 대한 요구가 줄어든다.
- 미국과 같이 분산화된 통계시스템을 가진 국가에서는 각각의 통계기관들의 활동을 조정할 목적을 가진 몇몇 관행들이 있을 가능성이 많다. 이러한 관행들은 미국 행정관리예산국(Office of Management and Budget)과 같이 자체적으로 조사를 수행할 수도 있고, 그렇지 않을 수도 있는 중앙기구에 의해 개발 및 공포된다. 하지만, 캐나다, 네덜란드, 뉴질랜드와 같이 중앙집중식 통계시스템을 가진 나라들은 각 기관들 간에 품질관행을 정의해야 될 필요성이 전혀 없다.
- 한 기관의 여러 사무국들의 지리적 산포도도 영향이 있다. 사무국이 넓게 분포된 기관은 조정(coordination)의 목적을 가진 품질관행을 도입해야 하는 동기가 더 크다.
- 입력데이터에 대해 행정기관과 같은 다른 조직에 크게 의존하거나 (예 : 미국세청 소득통계과(Statistics of Income Devision) 계약을 체결하여 조사활동을 수행하는(예, 미국 국립교육통계센터(National Center for Education Statistics)) 기관들은 계약서를 유지하기 위해 자료수집 품질관행이 있을 가능성이 많다.
- 미국 에너지국(Energy Information Administration)과 같이 비교적 새로 구성된 기관들은 Freedman (1990)이 개괄적으로 설명한 일반적인 원칙을 수립하기 위한 품질관행을 정의할 가능성이 크다.

기관들은 사회조사통계보다는 기업경기조사 통계에 관한 품질관행을 보다 많이 기재하였다. 이것은 사회조사가 보다 다양하고, 표준화의 구속을 덜 받는 경향이 있기 때문일 것이다. 아니면, 단순히 설문지를 작성한 사람들의 특정 관심사를 반영하는 것일 수도 있다. 앞서 지적한 바와 같이, 설문지는 소표본에게 발송되었고, 작성자에 관한 뚜렷한 지시사항은 없었다.

품질관행의 개발 및 확대라는 측면에서는 두 가지의 주요 주제가 있는 것 같다. (1) 응답자와 고객도 마찬가지, 외부 세계와 잘 정의된 인터페이스, (2) 개념적 차원과 실질적 차원 양면에서 자료입력, 처리, 출력에 관한 통합. 이 두 가지 주제는 메타데이터 및 새로운 컴퓨터, 통신기술의 출현과 응용에 관한 증가된 관심으로 뒷받침된다. 예를 들면, 스웨덴 통계청은 "우리는 메타데이터 시스템 개발을 너무도 강조하고 있다. 새로운 통계법령에(April 1994) 따라 스웨덴 통계청은 모든 공식적인 통계에 포괄적인 메타데이터 시스템을 준비해야 한다."고 응답했다. 또한, 호주 통계청은 현재 진행중인 중요한 데이터관리 프로젝트에 대해 말했는데, 여기에는 보다 나은 분류체계와 보다 큰 가시성 및 보다 많은 접근성을 가진 output 데이터를 통해 고객서비스를 개선하고, 개념과 절차를 통합하여 output 데이터의 정보내용와 상호 통일성을 향상시키고, 시스템 유지관리비용을 절감한다는 두 가지 목표가 있다고 하였다. "이러한 목표들은 전부는 아닐지라도 대부분의 기관들의 데이터 상품이 궁극적으로 생성되게 될 기업정보저장소의 개발, 정착 및 활용을 통해 이루어진다. 이 저장소는 기관에서 생산하는 모든 데이터 및 이에 상응하는 메타데이터를 저장하고, 분류하고 접근할 수 있는 시설을 가질 것이다. 저장소는 합리화와 통합을 증진시킬 수 있는 데이터관리정책 및 관행의 도입과 연계하여 구현된다."(Colledge and Richter, 1994).

일부 기관들은 전체의 틀 속에 품질관행이 자리잡는 품질관리 그 자체의 통합된 접근방식에 관한 표현도 하였다. 스웨덴 통계청은 "우리는 품질업무를 단순히 몇 개의(많을 수도 있고, 적을 수도 있는) 독립된 관행이 아닌 하나의 통합된 방식으로 바라본다."고 강조하였다. 이 기관은 최근, 기관에서 일어나는 모든 품질활동에 관한 통합된 체계를 제공하는 광범위한 품질정책을 도입하였다(Statistics Sweden, 1993). 이 정책의 목적은 "사용자들이 요청하는 객관적인 통계정보를... 전문 통계기관으로서의 역할을 수반하는 정확성에 대한 요구를 충실히 이행하면서... 사용자가 지정하는 형식과 지정하는 시간에.. 생산 및 제공하는 것"이다. 이와 비슷한 것으로는 현재 개발중인 "품질측정모델"(U.S. Bureau of labor Statistics, 1994b)이 있는데 그 목적은 세 가지 개념을 하나의 모델에 혼합하여 프로그램의 개선을 돕는 것이다. 고객중심의 종합품질관리의 원칙, BLS "업무프로세스"의 일반적인 시각은 즉, 통계청 프로그램 전반에 이용된 핵심활동 그리고, 제한된 자원 속에서 고객만족과 통계품질을 최적화하는데 바탕을 둔 의사결정. 이 모델은 품질지침 및 측정에 관한 체계를 제공한다. 요약하자면, 비록 실행은 어렵지만 개념적 통합에 대해서는 강렬한 욕구가 있는 것 같다.

품질관행에 관한 욕구는 단순히 기관 내의 조직단위에 한정되는 것이 아니라, 통계기관들과 다른 조직에 걸쳐 두루 적용된다. 예를 들면, INSEE(프랑스)는 국가회계기관이 수행하는 회계원칙을 개정하는데 처음부터 끝까지 참여하고, 이를 통해 사업체 자료수집에 대한 자신들의 요구사항이 충분히 고려되도록 하는 정책을 소개하였다. 통계기관들 간에 국제적인 데이터 공유는 점점 늘어나는 추세인 것 같다(Ryten, 1995). 이것은 앞으로 공통적인 데이터 모형에 관한 필요성으로 이어질 것이고, 따라서 국제적 품질정책이 점차 중요하게 될 것이다. 예를 들면, 유럽경제공동체 규정은 현재 통계단위와 사업자등록부에 역점을 두는 것에서 한 단계 넘어 점차 확대될 것이다.

감사의 글

여기의 저자들은 많은 응답자들에게 감사를 드리고, 이 조사가 상호이익을 위해 사용되었으면 하는 바람을 가지고 있다. 아울러, 편집에 많은 도움을 Cathy Dippo와 표, 그림을 작성하고 원고 교정을 도와준 호주 통계청의 Alistair Hamilton에게 감사드린다.

제 13 절 기업경기조사 추정치의 비교 가능성 개선

Bill Gross and Susan Linacre
호주 통계청

기업경기조사(business survey)에서 품질관리 및 오차조절에 관한 전략은 대개 개개의 조사차원에서 결정된다. 그러므로, 표본틀이나 선별단위, 기록단위, 무응답 처리 및 추정과 같은 사항에 대한 방법론적 선택들도 일반적으로 조사별로 이루어진다. 이상적으로는 이와 같은 선택들은 비용과 데이터 가용성의 제약 속에서 해당 조사의 사용자 요구사항에 대해 달성할 수 있는 최상의 품질을 제공한다. 그러나, 보다 폭넓은 상황을 구성하기 위해, 여러 조사에서 수집된 자료를 통합해야 될 필요성이 있는 사용자들에게는 그러한 결과가 항상 기대되는 것이 아니며, 어쩔 수 없이 각각의 조사에 서로 다른 전략이 선택되고, 이것은 통합된 자료의 사용자에게는 자료품질에 매우 중대한 영향을 미친다.

13.1 서문

어떤 조사에서든 품질관리에 대한 근본적인 요건은 사용자 요구사항의 철저한 이해와 상품 및 서비스의 품질특성과 이것이 사용자들에게 가지는 의미 파악이다. 그리고, 두 번째 요건은 상품과 서비스를 생산하는데 실제로 이용되는 프로세스에 관한 충분한 지식이다. 이러한 지식은 품질의 측정과 개선을 가능하게 한다. 품질관리의 세 번째 측면은 팀웍과 공동의 목표 그리고, 사용자 요구를 만족할 수 있는 상품 및 서비스의 지속적인 개선을 확고히 하는 것이다. Colledge and March(1993)는 이상의 내용을 품질관리의 근본적인 측면으로 분류하고 있다.

기업경기조사 설계 및 이행의 측면에서는 사용자 요구를 이해하는 것이 항상 간단한 일은 아니다. 왜냐하면, 대개는 다양한 사용자들이 조사의 서로 다른 측면에 관심을 가질 것이기 때문이다. 여기서, 어떤 사용자는 예컨대 한 산업의 포괄적인 척도나 혹은 전체 경제에서의 임금과 급여에 관한 정보 등, 단일 조사의 주제에만 뚜렷한 관심을 보일 수 있고, 반면 다른 사용자는 보다 큰 맥락의 일부분으로서만 그러한 정보에 관심을 가질 수도 있다. 일반적으로, 첫번째 사용자그룹은 직접적으로 그리고, 간접적이라 하더라도 매우 가시적이거나 두드러진 방식으로 자료를 사용하여(예, 매체배포) 자신들의 요구사항을 명시하면서 조사와 밀접한 관계를 유지할 것이다. 그러나, 몇몇 조사의 정보를 통합하는데 관심이 있는 사용자들은 이와 대조적으로 보다 어려운 작업을 해야 한다. 즉, 이들은 반드시 몇 개의 조사그룹들과 관계를 구성해야

한다. 때로는 몇 개 기관이 될 수도 있다. 교차조사(cross-survey) 사용자들에 대한 어려움은 이들의 output에 사용된 데이터가 종종 -원자료 생산자가 때때로 인식하지 못할 정도로- 추가적으로 처리되거나 조작됨에 따라 더욱 악화될 수도 있다. 자료 생산자들은 이 사용자들이 조사의 자료상품에 의존하는 범위를 파악하지 못할 것이다.

통계기관들은 몇 년 동안 통합된 경제통계개념의 개발을 이끌면서 교차조사 사용자들을 위해 상당한 노력을 해 왔다. 이러한 통합은 표준화된 분류/정의와 같은 인프라와 하나의 국가 사업체단위 등록증에 의해 뒷받침되고, 이 구조는 또한 산업분류 및 산업단위 등에 관한 국제표준에 의해 인도된다 (U.N. Statistical Office, 1990). 그러나, 통합의 개념을 구현하는 것은 단지 일부만의 성공을 거둘 뿐이다. 개개 조사의 자료 사용자들에게 해법을 제시하는 것과 몇몇 조사의 자료를 필요로 하는 사용자들에게 해법을 제시하는 것 사이에는 서로 간에 여전히 밀고 당기는 긴장감이 존재한다. 게다가, 통합정보시스템의 생산을 뒷받침하는 방법론적 관행에 대한 요구는 충분히 인식되어 있지 않다. 이러한 상충된 관심을 관리하기 위해서는 반드시 여러 사용자들의 요구가 균형을 이루어야 한다.

본 절에서는 호주 통계청(ABS)의 조사통합프로젝트(Survey Integration Project)에 대해 기술하였다. 이 프로젝트는 초기단계에서는 비교적 간단한 것처럼 보였다. 교차조사 사용자들의 조사결과 품질개선에 대한 욕구에서 비롯된 이 프로젝트의 목표는 ABS 조사에서 사용되는 각기 다른 방법론들을 검토하고, 그 차이점을 조사하며, 이러한 차이점들의 효과를 평가하는 것이었다. 통계청은 다양한 방법들을 사용하고 있고, 이 중에는 오랫동안 사용되어 온 것도 있다. 또한, 검증과 평가 및 문서화가 잘 되어 있는 방법들이 있는 가 하면, 충분히 검증되었거나 평가와 문서화가 이루어지지 않은 방법들도 있다. 따라서, 필요한 방법론적 변화를 관리하는 것은 기술적 사안만큼이나 중요한 쟁점임이 점점 확실히 드러났다.

일반적으로, 방법론을 검토하는 과정은 몇 가지 이익을 제공하는 것으로 알려져 있다. 이 중에는 "최상의 관행" 해법의 확인과 정립, 현재 관행들의 보다 나은 문서화, 그리고 인프라(시스템 구성 성분 뿐만 아니라, 문서작성과 교육도 포함) 공유에 관한 기회들도 들어 있다.("최상의 관행"의 개발은 제21장 Morganstein and Marker 편에 논의되어 있다)

우리는 현재 일관된 방법론들을 통한 품질개선에 초점을 두고 있지만, 어떤 상황에서는 통일된 절차들이 개개의 조사에서 비용을 발생시키거나 품질을 떨어뜨릴 수 있음을 인식하고 있는 것도 중요하다. 예를 들어, 무응답자를 추적하기 위해 동일한 절차들을 활용하는 것은 비현실적이다. 왜냐하면, 조사들은 대치법의 정확도가 각기 다르고, 그 결과 무응답 효과의 정도도 차이가 있기 때문이다. 조사통합은 맹목적으로 적용되어서는 안되고, 오히려 적정하지 못한 방법론상의 차이를 없애는 데 목적을 두어야 한다.

조사통합프로젝트(Survey Integration Project)는 캐나다 통계청에서 1985년부터 1991년까지 수행한 기업경기조사 재설계프로젝트(Business Survey Redesign Project)와 유사한 점이 많다(Colledge, 1987; Dodds and Colledge, 1990; Worton and Plat다, 1995 참조). 재설계 프로젝트는 "자료원과 경과시간 사이에서 정렬 및 비교될 수 있는 경제자료"를 생산하는 목적을 가지고 있었다(Dodds and Colledge, 1990, p.56). 이것이 주는 이익 중 하나는 "개념, 정의, 분류체계, 그리고 조사절차의 표준화"였다(Colledge, 1987, p.550). ABS 조사통합프로젝트는 이러한 목표들 중 많은 것을 공유한다. 그러나, 몇 가지 다른점도 지적될 수 있다. 조사통합프로젝트는 비교적 소규모였고, 한번에 몇 가지 측면의 방법론에 집중하였으며, 광범위한 조사에서 그러한 방법론들의 일관성을 개선하고자 했던 반면, 기업경기조사 재설계 프로젝트는 대규모로 사업자 등록증 및 행정자료 접근 등을 포함하는 새로운 경기조사 시스템 인프라의 도입과 그러한 인프라를 이용한 몇몇 기업경기조사의 동시 재개발을 포함하고 있었다.

아래에서는 지금 방법론에서 가능한 몇 가지 차이점과 이러한 차이가 기업경기조사 추정치의 비일관성을 어떻게 이끌 수 있는 가에 대해 논하고(13.2절), 일관된 방법론들을 통해 품질개선에 관한 프로세스의 개요를 설명하며(13.4절), ABS의 사례연구를 프로세스의 실례로 들어보겠다. 다른 통계기관들은 13.3절에서 고찰하겠다.

13.2 방법론상의 차이

서로 다른 조사의 추정치들 간에 비교가능성을 개선하는 첫번째 단계는 비일관성이 발생하는 방법론의 측면들을 이해하는 것이다. 방법론들은 자료항목과 분류, 단위, 표본틀, 표본 및 표본틀 유지절차, 그리고 편집, 대치, 추정의 범주들로 그룹화 된다.

13.2.1 자료항목 및 분류

조사는 대개 표준화된 정의와 분류의 체계 속에서 진행되지만, 이러한 체계의 사용은 몇 가지 방식에서 다를 수 있다. 먼저, 자료항목의 내역이 서로 다를 수 있다. 예를 들면, 하나의 자료항목에 관한 기준기간이 다를 수 있다. 종업원의 수를 특정일을 기준으로 기입하거나 혹은 특정한 지불기간을 기준으로 기입할 수도 있다. 후자가 약간 높은 수치를 산출한다. 기재되는 항목간에도 차이가 있을 수 있다. 예를 들어, 퇴직금은 임금과 봉급에 포함될 수도 있고, 안될 수도 있다. 또한, 동일한 항목을 측정하기 위한 조사문항이 다른 경우, 그 응답도 다른 결과를 낳고, 한 항목의 다른 세부내용에 관한 질문은 더 큰 차이를 보일 수도 있다. 예를 들어, 종업원 수에 관한 정보를 수집하면서 풀타임인지 파트타임인지 각각 구체적으로 기재하도록 요청하게 되면 모든 종업원을 세는데 있어 파트타임 종업원들이 누락될 가능성이 줄어들고, 전체 종업원 수를 증가시킨다. 자료수집시기도 기록되는 값에 영향을 미친다. 통계기관과 같은 대규모 사업체들은 초기버전의 데이터와 최종버전의 데이터를 가지고 있을 수 있다.

분류코드는 서로 다른 수단으로 부여될 수 있다. 예를 들면, 하나의 단위는 두 가지 다른 조사에서 두 가지 다른 산업으로 코드가 부여될 수 있다. 즉, 한 조사는 사업자등록부에 있는 코드를 사용하는 반면, 나머지 다른 조사는 조사 중에 수집된 정보로부터 코드를 도출한다. 코딩과정이 단순히 사무적인지 아니면 컴퓨터를 이용한 것인지에 따라 부여되는 코드번호에 영향이 미친다.

13.2.2 단위

표본틀(frame) 즉, 일반적으로 사업자등록부(business register)는 위치, 사업체, 기업 등과 같은 여러 단위들의 계층을 포함하고 있다. 그리고, 표집단위는 조사마다 서로 다를 수 있고, 기록단위(reporting unit)도 (즉, 자료가 수집되는 단계) 이와 마찬가지다(Nijhowne (1995)은 이러한 유형의 단위들을 정의했다). 기록단위가 다르면 적용범위도 달라진다. 예를 들어, 어떤 조사에서는 슈퍼마켓 체인에 속해 있는 상점들이 기록단위이지만, 그 본사는 다른 조사에 기록되어 있을 경우, 그러한 상점들의 목록이 예전 것이라면 동일한 질문에 상당히 다른 응답이 나타날 수 있다.

하지만, 동일한 표집단위 및 기록단위가 사용된 경우에도 하나의 대규모 조직에서 접촉점이 서로 다르다면 그 조직의 활동에 대해 각기 다른 시각을 가질 수 있고, 이에 따라 서로 다른 정보가 제공될 수 있다. 예를 들어, 급여와 직원이라는 분야는 재무회계분야 보다 훨씬 제한적인 시각을 가질 수 있다.

13.2.3 표본틀

각기 다른 원천의 표본틀이 사용되면, 분명 적용범위도 다를 것임을 예상할 수 있지만, 같은 원천, 전형적으로는 중앙사업자등록부라 하더라도 표본틀은 서로 다를 수 있다. 또한, 같은 기준기간 내에서도 표본틀은 서로 다른 시기에 추출될 수 있으며, 사업자등록부에서 추출된 표본틀들은 조사마다 다양한 방식으로 나타낼 수 있다.

13.2.4 표본 및 표본틀 유지관리절차

표본 및 표본틀 유지관리절차는 관측된 구조 또는 현황과 표본틀에 기록되어 있는 것이 서로 다른 단위들을 파악하고, 이를 처리하기 위한 규칙들이다. 이러한 규칙들의 차이는 ABS에서 이루어진 조사들간 추정치의 비교가능성에 상당한 영향을 미치는 것으로 확인되었으며, 이에 관해서는 13.5절에서 보다 상세히 논하도록 하겠다. 일단, 변화를 감지하는데 있어서의 변동성 및 변화에 대한 각기 다른 처리방법에서 비롯되는 차이는 검출되었다. Struijs and Willeboordse(1995)는 사업자등록부에서의 그러한 변화를 처리하는데 대해 논하였다.

하나의 단위가 기록되어 있는 현황을 바꾸는 가장 극단적인 경우는 그 단위의 활동이 중단되는 때 - 사망 - 이다. 이러한 경우를 확인하고, 처리하는 것은 때로는 간단한 일이지만, 대개는 추적조사에도 불구하고 일부 단위의 사망을 확인할 수 없는 경우가 많다. 그러므로, 추정치에서 이러한 단위들을 어떻게 처리해야 할 것인가에 관한 문제는 그 현황에 관한 최상의 추측을 필요로 한다. 사망을 확인하기 위해 추적조사에 투자하는 노력 그리고, 여전히 생존하는 것으로 되어 있는 미확인된 사망자의 비율에 관한 가정들은 조사마다 각기 다르다. 하지만, 이 중 어떤 경우라도 조사 추정치에는 서로 다른 결과를 초래한다.

하나의 단위가 산업을 변화시키는 것도 흔히 있는 일이다. 비록 여러 조사들이 그러한 단위를 처리하는데 동일한 방법을 채택할 수도 있지만, 일부는 그와 같은 변화를 겪을 가능성이 많다. 예를 들어, 상품에 관한 자료를 수집하는 제조업체들의 한 조사는 모든 단위에 산업코드를 부여하고, 이것으로 산업의 어떠한 변화를 검출할 수 있지만, 분기별로 행해지는 고용율 조사에서는 산업의 변화를 나타낼 수 있는 정보를 수집하지 못할 수도 있다

13.2.5 편집, 대치, 추정

표준화된 분류 및 자료항목 정의가 점점 보편화되어 가고는 있지만, 편집과 대치 및 추정방법에 대한 근본적인 차이는 여전히 존재한다. 편집의 경우를 보면, 까다로운 편집은 변경사항이 더 많고, 이로 인해 총계가 달라지고 때로는 단위값들의 분포도 달라지게 한다. 그리고, 추정치에 상당한 영향을 미칠 수 있는 단위에 관한 전략은 자주는 아니지만, 규모로는 상당히 큰 변화를 초래한다. 마찬가지로, 대치법이 서로 다르면, 단위값들이 달라지고 이로 인해 총계도 달라진다. Kovar and Whitridge(1995)는 이와 같이 서로 다른 대치법에 대해 논하였다.

추정치는 기준의 원천이나 유통과 같이 특정한 가중치 부여방식을 이행하는 방법에 따라 영향을 받을 수 있다. 그리고, 조사간에 상당한 차이를 초래할 수 있는 추정치의 또 다른 측면으로는 특이점의 탐지와 처리방법이 있다. 특이점에 대한 의사결정은 주관적인 요소이며, 동일한 일반원칙을 따른다 하더라도 조사분야마다 제각기 다른 의사결정을 할 수 있다.

13.2.6 비표집오차와의 관계

추정치 비교가능성의 개념은 비표집오차의 개념과 관련이 있다. 왜냐하면, 두 개념 모두 표본에 대한 추정량의 기대값이라는 측면에서 정의되기 때문이다. 따라서, 두 가지 조사의 추정치를 비교할 수 없다고 말할 때는 그 기대값들이 같지 않다는 것을 의미하는 것이다. 또한, 방법론이 달라도 동일한 비표집오차가 발생할 수 있으며, 이에 따라 추정치 비교도 가능할 것이라는 것은 논리적으로는 가능하긴 하지만, 실제로는 비표집오차가 다른 경우가 더 많다. 이런 식으로 비교불가능성의 원천은 비표집오차의 원천과 관련이 있다.

Biemer and Fecso(1995)는 비표집오차의 원천을 규격, 표본틀, 무응답, 처리 및 측정오차로 분류하고 있으며, 우리의 범주도 이러한 분류와 일치한다. 자료항목 및 분류의 다른 선택에서 비롯되는 모순은 Biemer and Fecso의 규격 또는 측정오차에 해당되고, 단위의 선택에서 비롯되는 모순은 무응답오차나 측정오차에 해당된다. 또한, 두 가지 목록에서의 표본틀 범주는 Biemer and Fecso의 처리오차에 해당되고, 우리의 표본 및 표본틀 유지관리 절차와 편집, 대치, 추정이라는 범주도 이와 관련이 있다. 한편, 우리의 범주는 Griffiths and Linacre(1995)의 기업경기조사에 포함된 프로세스 목록과도 유사하다.

13.3 통계기관들의 방법론적 차이

13.2절 주어진 방법론상의 차이에 관한 사례는 ABS의 경험에서 비롯된 것이었다. 이에 관한 조사는 다른 통계기관들의 상황을 평가하기 위해 수행되었으며, 15개국의 16개 기관이 이에 응답했다. 목적은 기관의 절차들이 여러 조사에서 어느 정도의 일관성을 유지하고 있는 지를 평가하는 것이었다. 일관성은 기관의 구조 및 추정치 비교가능성을 위해 사용하는 전략들과 관련이 있었다. 이 조사는 정부 통계기관들의 확률표본에 기초한 것이 아니었으며, 따라서 다른 기관에 대한 추론에는 주의가 요구된다.

13.3.1 방법론상의 모순

조사대상 기관들에게 주요 기업경기조사에 관한 정보를 요약하고, 특정한 방법론적 측면들이 이러한 조사에 동일하게 적용되었는지 그리고, 차이가 있었다면 몇 가지 예를 들도록 하였다. 표 13.1은 방법론에 관한 응답들을 요약한 것이다.

표 13.1 방법론에 관한 응답의 요약

질문	응답횟수 "예"	응답횟수 "아니오"
1.하나 이상의 주요 조사에서 얻은 중요한 자료항목이 있습니까?	16	0
2.조사에서 동일하게 수집된 공통 항목이 있습니까?	4	12
3.자료는 항상 표집단위에서 직접 수집됩니까?	3	13
4.사업자등록부를 이용하는 주요조사의 경우, 기준기간이 같은 조사들은 정확히 같은 시기에 표본틀을 정합니까?	6	10
5.사업자등록부는 몇 가지 주요조사를 위한 다른 원천들로 보완됩니까?	11	5
6.주요 조사들은 표본틀에 기록된 것과 다른 구조나 상태를 가진 단위들을 동일한 방식으로 정의하고, 확인 및 처리합니까?	11	5
7.하나 이상의 조사에서 공통적인 주요 자료항목들에 대해서는 같은 항목에 동일한 편집을 적용합니까?	3	13
8.하나 이상의 조사에서 공통적인 주요 자료항목들에 대해서는 무응답의 대치나 조정에 같은 방법이 사용됩니까?	3	13

모든 기관들은 하나 이상의 조사에서 자료항목을 수집하였다. 설문지에는 직업, 임금 및 급여, 자본지출, 재고 등의 항목이 있었는데, 거의 모든 기관들이 이러한 항목들을 한 번 이상 수집한 것으로 나타났으며, 예외는 주로 분산화된 통계구조의 일부에 소속된 기관들에게서 나타났다.

대부분의 기관들은 일부 항목을 수집하는 방법에서 몇 가지 모순이 있었다. 그리고, 그러한 예의 대부분은 항목의 내역(예, 각기 다른 유형의 고용인들이 기재한 직업) 및 기준기간(예, 분기 또는 연단위)과 관련이 있었다. 또, 일부 관련 활동들은 항목에 포함되어 있거나 빠져 있기도 했다(예, 재판매를 위한 상품재고가 생산재고에 포함된 것인지 아닌지의 여부).

대부분의 기관들은 통계단위들이 자료를 직접 보고하지 않는 경우가 일부 있었다. 그 예는 소규모 단위들의 보고에서 인용할 수 있는데, 하지만 이런 경우는 통상적으로 본사에서 각 지사에 관한 보고를 하는 경우가 많았으며, 여기에는 대규모 단위들로 소수 포함되어 있었다. 또, 다른 예들은 협회에서 제공하거나 행정기록에서 확인된 많은 단위들의 자료와 관련되어 있었다.

절반이 안 되는 기관들이 동일한 기준기간 동안 같은 시기에 표본틀을 구성하였으며, 나머지는 각기 다른 조사의 요구사항을 충족하기 위해 다른 시기에 표본틀이 갖추어졌다(예, 표본을 준비하는데 필요한 시간이 각기 다름). 다른 시기에 구성된 표본틀 간에 차이점의 정도에 대해서는 아무런 정보가 없었다. 그리고, 대다수의 기관들이 일부 조사에 자체의 사업자등록부를 보완했지만, 대개는 한 두 가지의 주요 조사에 한해서만 이루어졌다.

대부분의 기관들은 주요 조사에 동일한 표본 및 표본틀 유지관리 절차를 이용하였고, 그렇지 않은 경우에는 모집단에서 출생을 인정하는 전략이 달라 일부 차이점이 발생되었다.

대부분의 기관들은 조사마다 각기 다른 편집방법을 적용하였으며, 이것은 일반적으로 조사간 비교에는 항목의 조합을 달리 하는 것이 유용하기 때문이었다. 또, 대부분의 기관들이 조사마다 무응답 대치에 약간의 차이가 있었다. 여기에는 셀평균(cell mean)을 대치하고, 이전값을 조정하며, 다른 항목 내지는 다른 조사로부터 값을 계산하는 방식들이 포함된다. 편집과 대치의 경우는, 몇몇 기관에서 조사마다 다른 방식으로 사용하는 표준화된 도구가 있었다.

요약하자면, 대부분의 기관들이 표본 및 표본틀 유지관리 절차에서는 일관된 접근방식이 있는 것으로 보고했지만, 그 외 다른 방법론적 측면에서는 일관성을 찾아보기 드물었다. 이 많은 모순들은 특정한 조사에 특정한 사용자 요구사항과 특징들로 설명된다. 그렇지만, 일관성을 증대시킬 가능성이 얼마나 큰 가를 평가하는 것은 본 절의 범위를 넘는 사안이다.

13.3.2 구조

기관들에게 자신들의 구조에 관해 질문하였다: 기능적으로 전문화되어 있는지 그리고, 기업경기조사 방법론 전문가들의 중심그룹 내지는 중앙의 기업경기조사 분류그룹이 있는지.(모든 기업경기조사가 기관의 한 부서에서 처리되고, 다른 부서들은 이러한 조사에서 얻은 자료들을 분석 및 발행만 하는 경우, 그기 관은 기업경기조사에 기능적으로 전문화되어 있다고 할 수 있다) 질문의 요지는 기관의 일관성 정도가 이러한 측면의 구조와 연관이 있느냐는 것이다. 우리는 일관성 정도를 표 13.1의 2~4번 문항과 6~8번 문항에서 "예"라는 응답수를, 5번 문항에서 "아니오"라는 응답수를 세어 측정했다. 또, 이를 기초로 기관들을 낮음(0~2회), 보통(3~4회), 높음(5~7)으로 분류했다. 이것은 비록 미숙한 척도이긴 하지만, 관계에 대한 몇 가지 분석을 가능하게 한다. 표 13.2에서 표13.4까지는 일관성 정도와 구조적 측면 (1) 기능적 전문화, (2) 방법론에 관한 중심그룹, (3) 분류에 관한 중심그룹에 따라 기관들을 분류한 것이다.

기능적으로 전문화된 기관의 수가 적어 어떤 결론을 내리기에는 미약하지만, 이를 통해 기능적으로 전문화된 기관들은 다른 기관들만큼의 큰 차이를 보이지 않는 일관도 분포를 가진다.

표 13.2 방법론적 일관성 정도와 기능적 전문화 여부에 따라 분류된 기관의 수

기능적 전문화	일관성 정도			
	낮음	보통	높음	합계
예	2	1	0	3
아니오	8	1	4	13

표 13.3 방법론적 일관성 정도와 기업경기조사 방법론 전문가 중심그룹의 존재여부에 따라 분류된 기관의 수

방법론 전문가 중심그룹	일관성 정도			
	낮음	보통	높음	합계
예	3	1	2	6
아니오	7	1	2	10

방법론 전문가들의 중심그룹의 존재와 일관성 정도 사이의 관계에는 몇 가지 증거가 있다. 방법론 전문가들의 중심그룹이 있는 기관들은 그렇지 않은 기관들에 비해 일관성 정도가 높을 가능성이 큰 것으로 보인다.(표12.3)

중앙분류그룹의 존재와 일관성 정도 사이의 관계에는 아무런 근거가 보이지 않는다. 기능적 전문화와 마찬가지로 중앙분류그룹이 없는 기관의 수가 적어 어떤 결론을 내리기에는 미약하다.(표 12.4)

13.3.3 전략

추정치 비교가능성을 증대시킨 전략에 관한 내용을 물었다. 응답들은 "우리는 정의에 대해 설명하도록 하여 조사와 항목을 정확하게 정의하는 법적체계가 있음"이라는 고도로 중앙집중화된 접근방식에서부터 "각 조사에 대한 전략은 가능한 한 완전하고, 정확한 추정치를 산출하는 것이고, 각 조사의 목표는 조사에서 개발 및 설계된 질문에 가능한 최적의 응답을 제공하는 것임."이라는 고도로 분산화된 접근방식에 이르기까지 다양하게 나타났다.

표 13.4 방법론적 일관성 정도와 중앙분류그룹의 존재여부에 따라 분류된 기관의 수

중앙분류그룹	일관성 정도			
	낮음	보통	높음	합계
예	8	2	4	14
아니오	2	0	0	2

각 기관들은 조직구조와 표준 및 인프라와 관련된 전략들을 나열하였다. 조직의 전략에는 다음을 구성하는 것이 포함되어 있다. 모든 응답자의 연락을 맡은 한 그룹과 자료의 처리를 맡은 또 한 그룹, 분류 및 표준에 관한 단일그룹, 방법론에 관한 단일그룹, 대규모 사업체들로부터의 조사자료를 조정하고, 그들의

구조에 관한 정보를 유지하는 한 단위, 중앙의 설문설계그룹. 표준들은 분류, 자료항목, 질문, 코딩, 표본 및 표본틀 유지관리 절차와 관련이 있었다.(이러한 표준의 활용에 관한 정책들에 대해서는 Colledge and March(제22장) 편에서 논의되었다) 인프라의 유지에는 중앙의 사업자등록부와 일반화된 처리시스템 및 표준의 참고번호, 문서화, 교육과정이 포함되어 있다. 여러 조사들의 자료를 단위별, 총계별로 비교하는 것과 같은 통계프로세스들도 사용되었다.

표준의 준수와 보편적인 인프라의 사용을 확고히 하는 데는 서로 다른 접근방식이 취해질 수 있다. 제시된 응답들을 보면, 표준들과 표준 인프라의 사용을 어떻게 이뤄나갈 것인지를 설명하는데 "선전한다(promulgate)"에서 "촉진한다(promote)"를 거쳐 "적극 권장한다(strongly encourage)"에 이르는 용어들을 사용함으로써 이러한 사실을 반영하고 있었다.

13.4 일관성을 증대시키기 위한 프로세스

방법론적 일관성을 증대시키는 프로세스는 기술적인 프로세스인 만큼 관리적인 프로세스이기도 하다. 관리프로세스는 한 조사에 특정하거나 여러 조사에 걸쳐있는 사용자들의 요구에 관한 이해를 다룬다. 또한, 관리자들은 우선순위에 따라 이러한 요구를 충족해야 하는 통상적인 의무를 반드시 부여해야 한다. 만일, 일관성 향상이 최우선적인 요구라면, 예전에 지역차원에서 이루어졌던 의사결정은 중앙그룹 내지는 전체그룹 차원으로 변화될 필요가 있는 것이다. 개개의 조사들은 프로세스에 대한 제어력이 없고, 조사에 특정된 사용자 요구사항의 변경에 응대할 능력이 떨어진다. 이러한 요구사항 변화에는 표본틀 생성과 같이 지역의 기능들을 중앙으로 대체하고, 관리자들에게는 조사에 대한 제어력을 줄이고, 교차조사팀의 다른 팀원들에 의존도를 높이는 사항 등이 포함될 수 있다. 기술적인 해결책은 반드시 여러 조사에 관련될 수 있는 추정치의 품질목표와 대비해 각 조사의 품질의 균형을 이루어야 한다.

아래의 개괄적인 프로세스는 여러 조사에 걸친 결과의 통합을 개선하는데 있어 기술적, 관리적 측면을 모두 고려하고 있다. 이것은 ABS 조사통합프로젝트(Survey Integration Project)의 성패와 예상되는 향후의 요구에 관한 분석에서 비롯된 것이다. 여기서는 5가지 단계가 논의되었으며, 이 단계들은 조사들 간의 일관성을 증대시키는데 있어 절대적으로 중요한 것으로 간주된 것들이다. 5가지 단계는 다음과 같다.

1. 현재의 방법론들을 보다 잘 이해하기 위한 조사와 분석, 교차조사 및 조사특정한 사용자 요구를 충족하기 위한 최상의 방법론적 접근방식이라는 점에서 이와 필적하는 요구사항, 이러한 집합적 요구 각각에 대해 선택적 방법론적 접근방식의 영향.
2. 사용자 요구라는 측면에서의 우선순위에 관한 합의, 이러한 요구에 관한 조사 및 분석결과의 의미, 여러 가지 요구를 최대한 충족하는데 적절한 방법론.
3. 합의된 방법론 수정의 이행
4. 조사결과의 개선된 비교가능성의 이득을 유지하고, 조사특정한 사용자를 비롯해 교차분석 사용자들의 의식을 유지하기 위한 감시 등의 메커니즘.
5. 자료간 비교가능성을 유지하는 체계내에서 방법론들의 개선을 이룰 수 있는 실험 및 평가.

이상의 단계들과 제12장에서 Morganstein and Marker가 기술한 당시최상의 방법(Current Best Methods)의 개발은 서로 함께 한다. 이 둘은 모두 일관된 절차를 개발하기 위한 모든 단계에서의 부원 참여와 상급관리의 참여 및 절차 감시와 검토를 필요로 한다. 다만, ABS에서의 조사통합은 다양한 조사에서 일관된 방법론들을 성취하는데 목적이 있는 반면, Morganstein and Marker는 조사간에 일관적으로 적용될 수 있는 프로세스를 정의하는데 주된 역점을 두고 있다.

13.4.1 연구조사 및 분석

몇몇 조사에 걸친 연구조사는 문서와 자료의 분석을 필요로 할 것이다. 여러 조사에 사용된 절차들은 제대로 문서화되어 있지 않거나, 오래된 것일 수 있으며, 문서화되어 있더라도 그것을 얼마나 정확하게 따르는지를 확인해 볼 필요가 있을 것이다. 문서들은 일관된 용어가 사용되지 않았을 수도 있고 또, 중앙 보관소에 보관되지 않았을 수도 있다. 연구조사단계에서는 기존 관행들에 관한 종합적인 기록이 필요하지 않으며, 이는 변경될 가능성이 확실시 되는 경우에는 특히 더하다. 경험에 비추어 보면, 연구조사는 중요한 차이점들을 측정하고, 기존 관행의 존재이유와 특정한 사용자 요구를 충족하는 점에서의 중요성을 밝혀내며, 현실적으로 가능한 대안적 절차에 관한 의견들을 수렴하는데 보다 적합할 수 있다.

분석의 목적은 다양성이 조사간의 추정치 비교가능성에 어느 정도의 영향을 미치는 지를 파악하고, 일관된 절차를 통해 이러한 비교가능성을 개선하기 위한

체계를 개발하는데 있다. 분석은 문제의 그룹들을 확인하고, 이러한 그룹들에 적용가능한 해법을 제시할 수 있다. 또한, 예정된 절차상의 변화가 일련의 자료와 특정 조사의 사용자에게 미치는 영향에 관한 시각도 제공한다.

13.4.2 합의

이행단계 전에는 우선 교차조사 사용자와 조사특정 사용자들에 대한 우선순위에 관한 협의가 필요하다. 방법들의 적절한 균형이 이루어지고, 나아가 이러한 균형이 유지되고 있다면, 보편적인 이해는 매우 중요한 사안이다.

사용자와 함께 이루어지는 협의과정에 덧붙여, 합의에 이르는 과정에는 요구를 이해하고, 제안된 절차가 수용 및 진행가능 하도록 하며, 어떠한 다양성이라도 정당화되고 수용되기 위한 방법론과 조사들 사이의 광범위한 협의가 포함된다. 상급관리자의 승인은 일관성이 회사의 목표라는 사실을 명확히 하고, 가능하다면 필요한 자원을 제공하고, 인프라를 지원하면서 합의에 이르는 과정을 더욱 쉽게 해야 한다. 그러나, 상급관리자의 승인이 필요한 만큼, 하나의 팀으로 작업할 부원들이 모든 단계에 참여하는 것 또한 절대적으로 필요하다.

13.4.3 이행

새로운 절차를 이행하는데 걸리는 시간은 각기 다를 수 있다. 사무적인 절차의 변경은 쉽게 이행할 수 있지만, 대개는 몇 가지 변경이 이루어지기 전까지 복잡한 시스템의 사소한 변경은 뒤로 미루는 것이 나을 때가 많다. 또, 사무시설이나 소프트웨어 설비가 중앙에서 제공될 때는 개개 조사의 변경은 순서를 기다려야 한다. 어떤 변경이 일시적으로 중단된 시간들을 필요로 할 때는 사용자 요구사항이 이러한 시기를 결정하는데 중요한 역할을 한다. 변경과정은 조정과 관리가 필요하며, 몇 가지 측면 및 몇 가지 조사에서 장기간에 걸쳐 이러한 변경사항이 발생할 경우에는 특히 더하다.

몇 가지 중앙설비는 변경을 도울 수 있다. 이것은 지역서비스를 중앙서비스로 대체함으로써 일관성이 증대되는 경우에는 확연히 드러난다. 절차가 통일되어 있지만, 지역적으로 이행되는 경우에는 문서와 교육 및 일반적인 조언과 지원을 통해 변질되는 문제를 중앙에서 뒷받침할 수 있다. 또한, 새로운 절차를 위해서는 일반화된 컴퓨터 시스템들도 개발 또는 수정될 수 있다.

13.4.4 이득을 유지하기 위한 메커니즘

교차조사 사용자 요구를 이해하고, 이를 실천하기 위해서는 지속적으로 뒷받침되는 메커니즘이 필요하다. 이러한 메커니즘들은 교차하는 사안들이 이해되고, 처리되도록 하는 책임을 맡은 교차조사 프로젝트팀이나 위원회의 형식을 취할 수 있다. 또한, 개개의 조사와 중요한 교차조사 사용자들 사이의 관계를 개선하고 특히, 조사관리자들이 교차조사 사용자를 핵심 사용자로서 구분하도록 하기 위한 프로세스들도 있을 수 있다. 조사 및 교차조사 분석가 사이의 관계는 교차조사의 관점에서 볼 때 나타나는 조사결과의 몇 가지 모순을 해결하기 위한 협업에 따라 구성될 수 있다.

이행단계에서는 새로운 절차를 감시하는 것 외에 개선된 일관성도 반드시 유지되어야 한다. 만일, 사무절차에 변경사항이 발생한다면, 직접적인 외부감시가 어려워지고, 이것은 표준절차를 이행해야 하는 조사에까지 영향을 미치게 된다. 또, 중앙서비스가 제공되는 경우에는 각 조사의 관리자들은 사용자에 대한 적절성 측면에서의 품질에 대해 감시 및 보고할 중앙그룹을 필요로 할 것이다. 예를 들어, 일반적인 분기별 표본틀이 사용되는 경우에는 이전 분기에서 시작된 변경사항들의 조정이 이루어질 수 있다.

13.4.5 실험 및 평가

방법론의 통일성은 변화된 환경에 대한 개혁과 적응을 가로 막을 수 있다. 물론, 중앙의 연구그룹에서 비롯되는 개혁도 있을 수 있지만, 이 역시 개개의 조사그룹에 의한 실험에서 시작될 수도 있다. 일관된 품질을 유지하고자 하는 욕구로 개개의 조사에는 실험이 있을 가능성이 적다. 이러한 위험을 최소화하기 위해서는 생산환경과 거리가 먼 실험을 장려하고, 몇몇 조사의 견지에서 나온 문제점들에 대한 검토를 권장하는 메커니즘이 있어야 한다.

새로운 접근방식을 이행하는데 소요되는 시간은 통일성을 수반할 수 있는 보다 느린 의사결정과정을 감안해야 한다. 그러나, 일단 새로운 접근방식이 성공적으로 개발되고 나면, 이후에는 광범위한 조사에서 최상의 관행으로 빠르게 수용되고, 이러한 조사에서 동시에 이행될 수 있다. 결국, 점진적인 변경이 없도록 하고, 몇몇 작은 의사결정을 통해 조사간 방법론의 일관성을 낮추기 위해서는 정기적으로 절차를 검토하는 것이 필요할 것이다.

13.5 사례연구 - 표본 및 표본틀 유지관리 절차

본 절에서는 ABS 조사통합프로젝트에 속한 하위프로젝트에 대해 기술하였다: 표준화된 표본 및 표본틀 유지관리 절차의 개발과 이행. 13.2절에서 언급한 바와 같이, 표본 및 표본틀 유지관리 절차는 관측된 구조 또는 현황과 표본틀에 기록되어 있는 것이 서로 다른 단위들을 파악하고, 이를 처리하기 위한 규칙들이다. 본 사례연구의 주요 측면들은 다음과 같다. (1) 이것은 조사추정치의 실질적인 문제에서 비롯되었다. (2) 이것은 통합된 경제통계를 뒷받침하기 위한 장기적인 개념체계와 인프라 및 그룹들이 있는 통계기관에서 발생되었다. (3) 다각적인 차원에서 관리자의 참여는 일관성을 성취하기 위한 전략에 핵심적인 부분이었다.

이 연구는 사업자등록부 변경이 국가회계에 중요한 요소인 주요경제지표들의 추정치에 상당한 변화를 불러 일으킨 상황을 개선하고자 하는 시도에서 출발하였다. 연구조사자들은 중대한 요인이 표본 및 표본틀 유지관리에 사용된 절차가 서로 다르기 때문이라는 사실을 밝혀냈다. 비록, 변화의 분석은 보상적 조정을 가능하게 했지만, 절차의 다양성을 감소시키는데 중요한 것으로 고려되었다.

1969년 이후부터 ABS는 표준단위정의 및 표준분류의 개념적 체계와 통합된 경제통계를 뒷받침하기 위한 사업자등록부의 인프라 하에서 운영되었다. 또한, 조직으로는 경제통계에 관한 표준을 개발 및 감시하기 위한 중앙그룹과 방법론적 조언을 제공하기 위한 중앙그룹이 있었다. 그럼에도 불구하고, 이러한 개념적, 조직적 인프라가 문제가 있는 추정치의 원인이 되는 관행의 다양성을 막지는 못했다. 관행의 다양성에 대한 일례로 ABS는 기능적으로 전문화되어 있지 않다. ABS에서는 조사관리자가 조사처리를 직접 책임지고 있으며, 조사개발과 분석 및 결과물의 간행도 함께 맡고 있었다. 방법론 전문가들이 표본 및 표본틀 유지관리 절차를 개발하는데 일부 참여하고, 공통의 이론적 체계를 가지고 있긴 했지만, 이들은 일관되거나 통일된 접근방식을 유지하는 것보다는 특정한 조사목표를 달성하는데 더 많은 역점을 두었다.

관리자의 참여는 전략의 핵심적인 부분이다. 조사관리자들과 그 부원들의 참여는 진행가능한 일단의 절차를 개발하는데 필요하였고, 상급관리자의 지속적인 지원은 일관된 절차에 관한 초점을 유지하는데 도움이 되었다.

13.5.1 연구조사 및 분석

ABS의 통일된 표본 및 표본틀 유지관리 절차는 ABS 사업자등록부에 있는 단위들에 대한 대규모 조사에서 개발되었다. 이 조사는 최근의 조사에 참여하지 않은 단순한 구조를 가진 단위들을 검토하였다. 주 목적은 사망한 단위들을 삭제하는 것이었다. 표집오차의 효과를 제외하고는 사업자등록부에 기초한 추정치에는 아무런 변화가 없을 것으로 예상했으며, 이는 삭제된 단위들이 조사추정치에 아무런 기여를 하지 않았기 때문이었다. 그러나, 실제로는 주요조사의 추정치에 상당한 영향이 있었던 것으로 확인되었다. 이러한 차이는 조사와 사업자등록부가 단위의 사망여부를 확인하는데 서로 다른 절차를 사용했고, 또 일단 사망한 것으로 확인된 단위들을 처리하는 방법도 달랐기 때문에 발생한 것이었다. 앞으로 등록부를 유지관리하는 것에서 비롯되는 조사추정치의 영향을 방지하지 위해서는 사업자등록부와 모든 조사가 동일한 방법으로 사망을 확인해야 한다. 뿐만 아니라, 경험 자체가 비록 사망처리에만 한정되어 있긴 했지만, 이와 같은 사안은 대부분의 표본 및 표본틀 유지관리 절차에 똑같이 적용될 수 있는 것으로 인정되었다.

연구조사의 첫번째 부분은 위에서 기술한 문제점을 확인하고 그에 관한 전반적인 해결책을 찾는 것이었다. 그리고, 두번째는 광범위한 조사에 걸쳐 모든 표본 및 표본틀 유지관리 절차를 상세히 연구조사하는 것이었다. 또, 각기 다른 조사의 부원들과 함께 절차의 전범위에서 실제 관행을 문서화하고, 표준규칙의 가능성을 가늠하는 견지에 대해 논의하였다. 기존 절차의 문서를 수집하긴 했지만, 세부내용과 유통상에 상당한 차이들이 있는 것으로 나타났다. 비록 다른 조사들도 이러한 절차의 해당되는 일부를 사용하도록 권장되긴 했지만, 분석은 사업자등록부를 사용하는 조사에 집중되었다.

일관성을 권장하는 방법론 전문가들의 중심그룹이 있는데도 이와 같은 다양성이 어떻게 발생했는가 하는 것은 한번쯤 검토해 볼 만한 일이다. 이에 대한 설명은 조사관리자들의 우선순위와 방법론 전문가들의 고객과의 관계에서 가능해 진다. 조사관리자들은 대개 자신들의 조사의 특정한 요구사항과 문제점들을 다른 조사와의 일관성보다 훨씬 중요한 것으로 생각하였다. 또, 일부는 조사특정한 사용자들의 요구에 응하고 있었지만, 중요한 교차분석 사용자들이 있는 경우에도 특정조사에 적합한 관행을 그대로 적용하면 되는 것으로 생각했다. 고객서비스 관점에서 볼 때, 방법론 전문가들은 조사관리자들의 우선순위를 따랐다. 뿐만 아니라, 더 재미있는 것은 이러한 방법론 전문가들이 표준방법을 적용하기 보다는 오히려 특정한 문제에 대한 해법을 개발하고 있었다는 사실이다.

13.5.2 합의

절차를 조사하고, 개발하면서 완벽하게 통일된 일단의 절차들은 실현 불가능하다는 사실이 점점 분명해졌다. 일부 조사들은 표준절차의 변형을 필요로 하곤 했다. 또, 이런 경우에는 다양성의 정도를 제한하기 위해 이러한 변형들을 표준절차의 구조 속에서 유지하였다. 이러한 변형들은 승인과정을 거쳤으며, 아래 기술된 바와 같이 매뉴얼로 문서화되었다.

절차서 초안은 모든 조사에 회람되었다. 편집그룹은 초안과 최종본에 대해 언급을 하였다. 이 접근방식은 4~6명의 한 팀이 이러한 문서를 개발하는 Morganstein and Marker(제21장)의 제안과 일치한다. 절차서가 거의 완성되어 갈 즈음에는 각각의 조사에서 이행여부를 점검하고, 표준절차서의 변형에 대한 필요성을 논하기 위해 이를 검토하였다. 조사그룹과 지원그룹의 관리자들을 포함하는 집행 위원회에서는 절차서와 그 변형을 승인하였다. 위원회는 상급관리자에게 진행사항을 보고하고, 원칙에 입각한 지원을 유지할 수 있도록 하였다.

이러한 프로세스는 전폭적인 지지를 얻었다. 사업자등록부에서 사망자의 삭제는 개별 조사 및 교차조사 사용자 모두에게 두드러진 문제를 일으켰다. 이에 따라, 일관성은 하나의 단순한 추상이 아니라, 개별조사의 시계열에 대한 동요를 방지할 수 있는 것으로 인식되었다.

13.5.3 이행

새로운 절차의 이행에는 두 가지 측면 -사무적인 프로세스의 변경과 컴퓨터 시스템의 변경- 이 있었다. 각각에 대한 일정은 개별조사와 협의하여 정했다. 방법론 그룹의 일반적인 지원은 하나의 단위에 절차가 사용되었음을 기록하기 위한 표준표시(standard flag)에 관한 매뉴얼과 교육패키지 및 규격의 형태로 제공되었다. 이러한 표준표시들은 일반화된 입력처리시스템에서 이행되었다. 한편, 사무적인 변경은 이를 지원하는 일반시스템이 구성되거나 혹은 개별시스템이 충분히 수정되기 전에 발생할 수 있으므로, 과도적인 조정이 이루어졌다. 또, 변경의 효과가 측정되기도 했지만, 그 방법은 기존 절차들이 다른 만큼 조사간에 서로 달랐다.

시스템에서의 이행은 문제가 있었다. 처리시스템은 연령과 유동성이 다른 광범위한 시스템들이 있다. 보다 오래된 시스템에서의 부분적인 이행은 비용이

많이 들고, 위험하기도 하지만 어떤 경우에는 시스템이 재개발되기 전까지 기다려야 하는 경우도 있다. 이러한 문제점들은 일관된 방법론들을 성취 및 유지하는 것과 일반화된 컴퓨터 시스템을 개발 및 활용하는 것 사이의 관계를 부각시킨다. 일반적인 시스템으로는 몇몇 조사의 일관성을 향상시키거나 절차를 개선하기 위한 변경을 쉽게 할 수 있다. 또, 한편으로 일관된 방법론들은 일반화된 시스템을 보다 간단하고 보다 저렴하게 할 수 있다.

13.5.4 유지관리

앞서 언급한 집행 위원회는 개별조사 분야들간의 방법론적 사안을 처리하기 위해 정기적인 모임을 지속할 것이다. 위원회는 절차의 지속적인 적용가능성을 보장하고, 절차를 변경 또는 변형하는데 대한 승인을 맡는다. 그리고, 매뉴얼은 전자식 기업문서로 활용할 수 있고, 방법론 그룹은 이것을 유지하면서 지속적인 교육을 제공한다.

절차의 사용을 감시하기 위해, 관리자들의 일상적인 사무작업 감시 외에 추가로 두 그룹의 측정이 제안되었다. 첫째, 각 절차에서 확인된 단위들의 표본수는 갑작스런 변화를 감지할 수 있도록 조사관리자들이 감시한다. 둘째, 각 절차에서 확인된 단위들의 모수 추정치는(예, 사망 확인) 개별조사에서 도출하고, 이를 조사간에 비교한다.

13.5.5 실험 및 평가

실험 및 평가단계는 아직 이르지 못했지만, 전조가 된 활동은 있었다. 그 중 한 가지는 사업자등록부 관행이 변경된 경우의 전반적인 검토일 것이다. 또, 한 단위의 사망 또는 생존여부에 관한 최상의 추측법칙이 개발 및 시험되었으나 이 또한 정기적으로 재검정되어야 한다.

13.6 결론

지금까지 여러 조사에 자료를 관련시키는 사용자들에 대한 조사결과의 품질 관리에 대해 살펴보았다. 그 결과, 중요한 측면들은 교차조사 분석가들의 요구를 포함하여 하나의 조사에 대한 전체 사용자 요구를 이해하는 것, 현재의 절차와 방법론 및 이것이 사용자 요구를 충족하는데 있어 가지는 의미를 이해하는 것, 그리고 이러한 요구를 충족할 수 있는 보다 나은 방법론들을 지속적으로 찾는 것이었다. 여기서, 우리는 교차조사 분석가들을 위한 보다 나은 해법의 완성에는

적절한 기술적 해법을 개발하는 것뿐만 아니라, 이들의 요구를 보다 잘 충족할 수 있도록 조직의 문화를 정립해 나가는데 필요한 전략을 찾는 것도 포함될 것임을 지적하는 바이다. 변화에는 실제, 노력과 시간단절이라는 면에서 비용이 발생하며, 조사특정한 사용자들의 요구와 중요 프로세스에 대한 합당한 제어력을 유지하려는 조사의 욕구라는 면에서의 장벽들도 있다. 이러한 비용을 감당하고, 장벽을 극복하기 위해서는 기업과 조사관리자들의 상당한 노력이 필요하다. 또한, 절차를 개혁하고 개선하고자 하는 의지도 반드시 있어야 하지만, 이는 여러 조사분야들간에 팀 접근방식의 맥락에서 유지되어야 한다.

감사의 글

본 절의 견해는 저자의 소견일 뿐이며, 반드시 호주 통계청의 견해를 반영하는 것은 아니다.

제 14 절 조사자료 평가: 사전조사에서 품질평가까지의 변천 과정

James L. Esposito
미국 노동통계
Jennifer M. Rothegb
미국 인구조사국

14. 1 서문

일반적으로, 이 절의 목적은 측정오차를 포착하고 최소화하려고 노력하는 조사연구원의 문제-해결 행태를 다루는 것이다(Biemer et al., 1991; Grover, 1987, 1989; Grover et al., 1988; Turner and Martin, 1984). 특히, 우리는 질의 평가와 설계의 개선을 통해 조사연구원들이 양질의 조사 자료를 획득하고자 하는 노력에 초점을 맞추고 있다. Dippo(제20장)에 따르면, 지속적 처리과정 개선을 위한 첫 번째 단계는 적절한 척도를 결정하고, 조사 측정 시스템 내에 이들을 구축하는 것이다. 조사연구원이 조사 측정의 인지적 측면(CASM; Cognitive Aspects of Survey Measurement) 활동을 채택할 경우 (Jabine et al., 1984; Jobe and Mingay, 1991), 그들은 인지적 심리(Tourangeau, 1984)와 사회적 심리(Cannell et al., 1981; Esposito and Jobe, 1991) 개념을 사용하는 일련의 질의 응답 과정에 관한 모형을 개발하고, 측정오차를 평가할 수 있는 기술적 방법을 개발해야 한다. 이러한 기술이 조사 설문지를 평가하는데 어떻게 사용되는 가에 대한 실례와 예시를 제공함으로써, 우리는 사용자들이 자료의 질적 측면을 향상시키고, 사전적 예비조사를 넘어 사후 실행의 품질평가 역시 개선할 수 있을 것으로 확신한다. 우리가 강조하는 것은 평가의 과정이고, 특별한 기법의 강점이나 약점이 아니다(이러한 기법의 강점이나 약점에 대해서는 Presser and Blair, 1995; Esposito et al., 1992; Cannell et al., 1989; Fowler and Roman, 1992를 참조).

본 절은 아래와 같이 구성되어 있다. 14.1절에서는 조사의 질적 문제에 대해 간단히 언급한 후, 조사 설문지의 효과를 평가하기 위한 몇 가지 기법에 대해 확인하고 평가하며, 설문을 관리한 면접원들로부터 획득한 자료의 질적 측면을 평가할 수 있는 또 다른 기법들에 대해 검토한다. 14.2절에서는 이러한 몇 가지 기법들이 미국 상시인구조사(U.S. Current Population Survey; CPS) 자료의 질적 측면에 대한 문제를 확인하고 개선 사항을 평가하는 데 어떻게 사용되는지에 관해 보고한다. 그리고, 14.3절에서는 주요한 사회적·경제적 조사에 있어 이상적인 품질평가 프로그램을 제안하고, 이러한 프로그램과 관련한 실질적인 문제를 제기하며, 품질평가조사와 관련한 비용과 편익에 대해 토의한다.

14. 1. 1 조사의 품질

다양한 형태의 조사 관련 분야에 종사하던 연구원들이 받아들일 수 있는 조사품질에 대한 정의를 도출하는 것은 상당히 불가능한 과업으로 보일지 모른다. 그러나, 광범위한 부분을 포함할 수 있는 정의가 결여되어 있다는 것은 개념이 불완전하게 이해된다는 것을 의미하는 것은 아니다. Bailar(1984)는 조사품질을 다중적 의미로 해석하고 있다. 그 하나는 주어진 자료 세트를 산출하는 데 포함된 연결 수단이나 단계의 관점에서 해석된다. 그녀는 많은 특성들이 양질의 조사 자료의 수집을 가능하게 한다고 설명한다.(예를 들어, 확률표본추출, 개념적 명확성, 개념에 적합한 조작적 정의, 가장 식견 있는 응답자에 의한 보고, 정확하게 코드 및 가중치가 부여된 자료, 소표본 분산, 모순이 거의 없는 검증절차(Bailar, 1984, p. 43)). 다른 연구원들은 (Anderson et al., 1979; Groves, 1989) 조사 품질을 자료 정확도를 떨어뜨리는 다양한 형태의 오차의 관점에서 설명하고 있다. 예를 들어, Groves (1989)는 비관측 오차(예를 들어, 포함오차, 무응답오차, 표본오차)와 관측오차 (예를 들어, 면접원, 응답자, 설문, 자료수집방식)를 구분하고 있다.

접근방법은 다르더라도, Bailar와 Groves의 조사품질에 대한 개념화는 개념을 정의하고 이 절의 한계를 설정하는 데 매우 유용하다. 후자와 관련하여, 우리는 연구원들이 측정오차의 원천을 포착하고 잠재적으로 줄이는 것을 가능하게 하는 방법과 기법에 관해 집중적으로 설명한다.(우리는 조사품질에 있어 비관측오차의 효과에 대해서는 언급하지 않는다.) 예를 들어, 인지적 면접과 응답자 디브리핑 기법은 조사 개념이 잘못 전달된 때를 결정하는데 사용된다. 또한, 면접원들이 설문을 읽기 어렵거나 응답자들이 적절한 응답을 하기가 어려운 경우에는 대화식 코딩방식으로 면접성과를 감독하고 방식효과(mode effect)를 평가할 수 있다. 그러한 기법에 의해 포착될 때, 문제점은 자료의 품질에 부정적인 효과을 가져다 줄 것으로 추측할 수 있다. 그러나, 몇몇 경우에, 효과의 크기를 정량적으로 평가하기는 어렵다. 그럼에도 불구하고, 일단 정의가 되고 나면, 설문지 설계자(즉, 주요사건 전문가, 행동에 관한 과학자, 조사방법론자들)들은 설문 조정(예를 들어, 표현이나 질문 구조의 변경)이나 여타 전략적 방법(예를 들어, 면접원 훈련, 형식 변경)을 통해 문제를 감소시킬 수 있다. 가정이 사실이라면, 결과적으로 측정오차를 줄임으로써 자료의 질적 측면을 향상시킬 수 있을 것이다.(조사 문항으로 문제점을 확인하고 교정하는 주제는 14.3.1절 참조).

14. 1. 2 조사 설문지 평가, 트레이드 수단

조사문항의 사전검정에 사용되는 방법과 기법에 대해서는 방대한 문헌이 존재한다.(예를 들어, DeMaio, et al., 1993; Forsyth and Lessler, 1991; Willis, 1994). 이러한 기법은 평가 자료가 수집되는 장소(사무실, 실험실, 혹은 현장), 평가의 목적(사전검사 (PT), 혹은 품질평가(QA)), 분석자료의 원천(S)과 목표(T)(응답자, 면접원, 기타)로 구분된다. *목적*의 중요성은 문헌상으로 잘 다루어지지 않는다. 다수의 동일한 기법이 개발되었고, 현지 시험조사 설문지는 잘 작성된 설문 항목이 설문지가 완성되고 생산에 들어간 후 어떻게 작용하는지를 평가하기 위한 목적으로 사용될 수 있으며 또한 사용되어 왔다. 표 14. 1에 포함된 정보는 설문지 평가의 계획 단계에 사용될 수 있다. 다중기법 품질평가 프로그램을 개발하고자 하는 연구원들은 다양한 원천(즉, 면접원, 응답자, 전문가)으로부터 도출된 정보를 확인하기 위해 *원천과 목표*에 있는 정보를 사용할 수 있으며(예를 들어, 디브리핑, 행동코딩, 인지적 형태 평가), 그렇게 함으로써, 잠재적으로 문제가 될 수 있는 조사 문항을 확인할 수 있는 다중적 시각을 가질 수 있다.

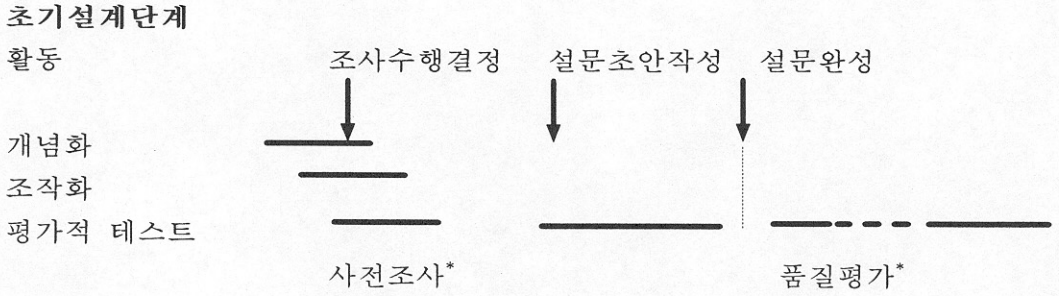
우리는 사전조사와 품질평가조사 간의 개념적 구분은 *설문지의 위상*(status of the questionnaire)에 근거한다고 믿고 있다. 평가조사 시점에, 설문지가 초기 개발단계에 있거나 현지시험단계에 있다면, 그리고 설문지를 변경하는 것이 아직 가능한 상태라면, 우리는 평가 작업을 *사전조사*로 본다(그림 14.1). 예를 들어, 만약 설문지가 현재 사용되고 있고 평가조사의 목적이 결정된 상태라면, 우리는 평가 작업을 *품질평가조사*로 분류한다.(기타의 사용은 14.3절에 있다). 이렇게 두 가지로 규정하는 것이 철저하게 빠짐없이 규정한 것은 아니다. 예를 들어, 위에서 고려되지 않은 것은 특별한 목적을 위한 특정 기법(예를 들어, 행동코딩)의 효용을 증명하기 위해 수행된 방법적 조사이다.(예를 들어, 면접원 성과 평가; Mathiowetz and Cannell, 1980을 보라).

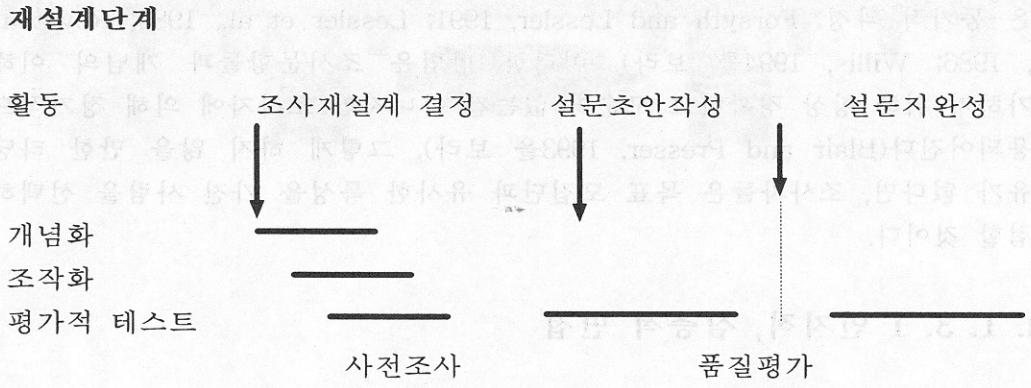
표 14. 1 조사 문항 평가에 사용되는 몇 가지 방법과 기법

방법 / 기법	자료 수집 장소	목적	분석 자료의 원천(S)과 목적(T)			
			면접원	응답자	기타S/T	조사Qs
인지적 / 심층적 면접						
(1) 생각 말하기 기법 사용	실험실	PT		S		T
(2) 탐문기법 사용	실험실	PT		S	T (개념)	T
(3) 기타 기법 사용	실험실	PT		S		T
디브리핑						
(1) 조사 후 추적 조사	현장	PT/QA		S	T (개념)	T
(2) 디브리핑 설문	현장	PT/QA	S	S		T
(3) 표적 집단	현장/실험실	PT/QA	S	S		T
실험 (예를 들어, 분할 투표 테스트)	현장/실험실	PT/QA	S	S	S (상황)	T
전문가 검토						
(1) 전문가 패널	사무실	PT			S(전문가)	T
(2) 인지적 형태 평가	사무실	PT/QA			S(전문가)	T
대화식 코딩						
(1) 행동코딩	현장	PT/QA	S	S	S(상황)	T
(2) 대화 분석 코딩	현장/실험실	PT/QA	S	S	S(상황)	T
(3) 프로토콜 코딩	실험실	PT/QA	S	S		T
항목 무응답 분석	현장	PT		S		T
재면접	현장	QA	S	S	S(상황)	T
응답 분포 분석	현장	PT		S	S(설계팀)	T
vignette	실험실/현장	PT	S	S	T(개념)	T

주: PT는 사전조사를 의미하고, QA는 품질평가를 의미함.

원천; DeMaio et al., 1993; Forsyth and Lessler, 1991; Willis, 1994.





* 사전조사는 특정 설문지를 완성하기 전에 수행되는 평가적 테스트를 의미하며 개발적 테스트와 현장 테스트를 포함한다. **품질평가**(예를 들어, 재면접)는 특정 설문지를 완성하고 난 후의 평가적 테스트를 의미한다. 품질평가는 지속적 과정일 수도 있고 아닐 수도 있다.

그림 14. 1 조사 설문지 작성과 재설계 과정 요약

더욱이, 이러한 두 가지 규정이 항상 구분하기 쉬운 것은 아니다. 예를 들어, 지속적인 질적 감시 프로그램은 일상적 사용(예를 들어, 목표가 있는 면접원 훈련)보다 더 많은 개입을 필요로 하는(예를 들어, 질문표현의 변화) 특정 조사문항과 관련한 심각한 문제점을 밝힐 수 있다. 마지막으로, 그림 14.1에서와 같이 품질평가조사와 사전조사는 동일한 조사에서 동시에 발생할 수 있다(예를 들어, 설문지 재설계 노력을 하는 동안). (이 그림은 14.3.1절에서 다시 언급한다.)

표 14.1에 언급한 대로 조사평가자료는 다양한 장소에서 수집될 수 있다. 실험실, 사무실, 현장 등 다음절에서 우리는 첫 번째 실험실 및 사무실에 근거한 방법과 조사를 고려한다. 그리고 다음으로 현장조사방법을 고려한다.

14. 1. 3 실험실 및 사무실에 근거한 방법과 조사

실험실에서는 조사자들이 설문지 관련 평가 조사를 실시함에 있어 통제 가능한 환경을 제공 받을 수 있다.(미국 정부조사실험실에서 수행된 조사의 검토와 사례는 Demaio and Rothgeb 1996; DeMaio et al., 1991; Dippo et al., 1993; Dippo and Norwood, 1992; Sirken 1991; Willis et al., 1991, 참조). 실험실 근거방법(예를 들어, 인지적 면접)은 일반적으로 질의-응답 과정의 심리적 측면에 초점을 맞춘다.(예를 들어, 이해와 같은 인지적 과정, 만족 또는 선택적 보고와

같은 동기적 과정; Forsyth and Lessler, 1991; Lessler et al., 1989; Royston et al., 1986; Willis, 1994를 보라). 이러한 방법은 조사문항들과 개념의 이해를 평가하기 위해 항상 정확하고 모순이 없는건 아니지만 조사자에 의해 정기적으로 사용되어진다(Blair and Presser, 1993을 보라). 그렇게 하지 않을 만한 타당한 이유가 없다면, 조사자들은 목표 모집단과 유사한 특성을 가진 사람을 선택하여 면접할 것이다.

14. 1. 3. 1 인지적, 심층적 면접

조사 응답과정에 적용될 때 인지적, 심층적 면접 방법은 인간의 정보처리 과정의 다단식 모형에 근거한다.(Willis, 1994; Willis et al., 1991을 보라). 이 모형은 응답자들이 질문을 이해하려고 시도하고, 기억 속에서 관련 정보를 회상하고, 노력 및 자기표현 수준과 관련한 의사결정을 수행하고, 궁극적으로 응답을 제공한다고 설명한다. 선형적 과정을 가정하지는 않는다. 몇 가지 질문에 대해서는 응답자들은 단계를 반복한다. *인지적면접*에서 조사자들은 응답자들에게 답변을 하면서(동시에) 또는 답변을 하고 잠시 뒤에(회고) "크게 소리내어 생각하게" 하여 조사문항으로 문제점들을 파악하려 한다. 종종, 조사자들은 응답자들에게 한정된 수의 일반적이거나 항목 특정적인 질문에 답하도록 함으로써 탐문기법을 면접에 통합한다. *심층적면접*은 인지적 면접의 형태를 취하지만, 다양한 기타의 기법 -부연하기, 신뢰 수준, 잠재 응답 측정과 같은(Royston, 1989; Ryoston et al., 1986; DeMaio and Rothgeb, 1996)- 을 사용한다는 점에서는 인지적 면접과 차이가 난다. 이러한 과정을 통해 조사자들은 응답자들이 어디에서 인지적 어려움을 경험하는지, 그리고 어디에서 질문의 의도(조사 스폰서, 문항 설계자)와 질문의 이해(응답자)간에 모순이 존재하는지를 판단할 수 있다.

14. 1. 3. 2 기타 실험실 및 사무실에 근거한 방법

인지적, 심층적 면접 외에도 조사 실험실에 적합한 다양한 여러 방법들이 있다(예를 들어, 전문가 평가, 순위 결정). 여기에 관심이 있는 독자들은 이러한 방법을 논하기 위해 Forsyth and Lessler(1991)을 참고해야 한다. 사무실 근거 조사방법은 특정한 분석적 기법(예를 들어, 언어 분석, 컴퓨터 코딩)의 개발을 반영한다고 나타난다(예를 들어, 대화 분석적 코딩, 자동 코딩). 그리고 다른 방법은 설문지 설계에 전문적 의견을 제공하는 전문가의 도움을 받는다(예를 들어, 인지적 형태 평가, 전문가 패널).

CASM 운동의 상대적 서광과 미국 연방통계시스템(국가건강통계센터, 노동통계청, 인구조사국) 내의 세 가지 인지적 실험의 수립으로 연방 조사에 있어 인지적 조사는 급격히 증대해 왔다. 이러한 방법들이 사용되는 방법을 독자들에게 보여주기 위해 우리는 문헌에서 찾은 혁신적인 학설을 간략히 서술한다.

14. 1. 3. 3 사전조사 사례

Blixt and Dykema(1993)는 소위 *체계적 심층면접*이라 불리는 혁신적 사전조사방법을 개발했는데, 이는 문제성 조사문항을 확인하는 방안으로써 행동코딩과 여러 심층 인지적 면접 -생각 말하기, 부연하기, 기억 탐구, 기타 탐구 기법- 을 통합하는 기법이다. 본 연구에서는 인지적 면접 동안 여러 종류의 탐구가 요구된다. 정의 탐구("내가 '영양'이라는 단어를 사용하면, 당신에게 어떤 의미로 들리는가?"), 참조체계 탐구("내가 '다른 사람들'이라고 하면, 당신에게는 어떤 의미인가?"), 그리고 동기관련 탐구("얼마나 자주 술을 마시는가에 대해 말하면 당신은 당황스러운가?"). 숙련된 조사자들은 심층면접을 수행하고, 음성녹음한 후 다음을 사용하여 코딩하였다. (1) 표준 응답자 행동 코드(예를 들어, 양호한 응답, 설명 요청) (2) 위에서 언급된 탐구에 대한 답변을 반영하는 전문적 코드(예를 들어, 조사 문항에 대한 비판과 응답 옵션). Blixt and Dykema는 두 가지 문제성 문항에 대한 데이터를 탐구한다. 이들은 분석을 통해 체계적 심층면접은 조사와 관련한 인지적 동기적 문제 모두를 확인하는 효과적인 방법이라고 주장한다.

Bolton(1993)은 인지적 면접과 전산화된 내용 분석을 통합하는 *자동 코딩*(automated coding)이라 불리는 조사문항 분석방법을 개발하였다. 이 방법은 인지론에 잘 부합하며, 인지적 면접 사본의 내용을 포함한 다섯 가지 언어적 카테고리와 매치된다(예를 들어, 반복, 망각, 신뢰, "말할 수 없음", "알 수 없음"). 그리고 인지적 면접 내용들을 구성하고 있는 4개의 비언어적 카테고리와 매치한다. 그녀는 목표 문제와 관련한 특정한 인지적 문제를 확인하기 위한 수단으로 요인 분석법(다시 말하면, 이해를 위한 요인 원천, 검색, 판단, 응답)을 사용한다. 본 연구에서 Bolton은 대체적 고객 만족 조사를 평가하고 관측적 감시를 통한 자동 코딩과 대비한다.(예를 들어, 행동 코딩과 유사한 기법). 그녀는 자동 코딩은 이해, 검색, 판단 문제를 더 잘 포착할 수 있지만, 관측적 감시는 응답의 어려움을 더 잘 포착할 수 있다는 것을 발견했다.

14. 1. 3. 4 품질평가 조사 사례

Forsyth 등(1992)과 Forsyth and Hubbard(1992)는 소위 *인지적 형태* 평가라 불리는 기법을 개발했는데, 우리는 좋은 설문지 평가 도구 -품질평가와 사전조사 양쪽 모두- 라 믿고 있다. 설문지 항목은 코딩도구를 가진 전문가에 의해 코딩되는데, 이는 조사 응답 과정의 인지적 모형에 부합한다. 자세한 코딩 카테고리는 개인의 이해, 해석, 기억, 판단, 응답 생성 과정에 관한 요구를 평가하기 위해 개발되었다. Forsyth 등은 이 코딩 기법을 약물 남용에 대한 전국 가구 조사(National Household Survey on Drug Abuse)의 문항을 평가하는 데 사용하였으며, 몇 가지 유형의 문제를 발견했다(예를 들어, 모호한 용어, 감추어진 정의가 있는 응답 카테고리) 그 뒤, 조사자들은 소수의 소리내어 생각하기(think-aloud) 면접을 통해 자신들의 결과를 증명할 수 있었다.(품질평가 조사의 다른 예에 대해서는 제12장 Dykema et al., Chapter 참조).

14. 1. 4 현장 근거 방법과 조사

현장 근거 방법은 면접원과 응답자(DeMaio et al., 1993) 두 가지 주요 원천으로부터 정보를 도출하는데, 이러한 기법은 조사가 실제로 이루어지는 장소나 이와 아주 근접한 현장에서 수행된다. 우리는 방법을 면접원 디브리핑, 응답자 디브리핑, 상호작용 분석 및 재면접 프로그램 등 4개로 나눌 것이다.

14. 1. 4. 1 면접원 디브리핑

숙련된 현장 면접원은 어떻게 조사 문항을 응답자에게 잘 이해시키는 가를 아는 사람 중의 하나이다(Converse and Schuman, 1974; DeMaio, 1983; cf. Bischoping, 1989). 이러한 전문적 지식은 다양한 기법에 도입될 수 있다. 그 중 보다 일반적인 세 가지 기법은 평정형식, 표적 집단, 면접원 디브리핑 설문을 사용하는 것이다.(Fowler, 1989; Fowler and Roman, 1992; Esposito and Hess, 1992) 디브리핑 설문은 대규모 면접원의 많은 대표들을 두는 이점을 가진다. 표적 집단은 응답자와 면접원에게 문제를 야기시킬 수 있는 문항을 심층적으로 관찰 할 수 있는 이점이 있다. 유용하기는 하지만, 면접원 보고자료의 속성을 잘 알아야만 하고, 문제성 조사 문항을 확인하고 수정하고자 할 경우 다양한 원천(예를 들어, 면접원, 응답자, 전문가, 실제 면접을 통해 코딩된 상호작용 자료)으로부터 획득하는 과정을 거쳐야 한다.

14. 1. 4. 2 응답자 디브리핑

아마도 조사 문항이 얼마나 잘 이해되었는가에 관한 자료 수집 방법은 사람 -응답자- 에게 질문을 해 보는 것이다. Belson(1981)은 개념과 용어를 조사 스폰서나 설문 설계자가 이해할 수 있을 때를 결정하기 위해 추적 탐구를 사용하여 몇 가지 흥미로운 조사를 연구하였다. 다른 조사자들은 응답자 디브리핑 방법으로 vignette를 사용하였다.(Martin, 1986; Polivka and Martin, 1992) 동일한 목적을 위해 또 다른 접근이 표적 집단에 이루어 졌다. (Palmisano, 1989) 연구의 공통점은 응답자들이 조사 문항에 대한 자신의 이해(혹은 오해)를 평가하기 위해 사용할 수 있는 정보를 제공하고 있다는 것이다.

14. 1. 4. 3 상호작용 분석

상호작용 분석은 조사행동 감시(자연적 조사 상황에서 자주 있는)와 면접원과 응답자 간의 행동교류 코딩, 그리고 특정 조사문항에 대한 사전설정 행동코드의 빈도를 표로 만드는 것을 포함한다. 빈도가 높은 문항은 문제가 있는 것으로 표시된다.(예, Dykema et al., Chapter 12; 리뷰, Fowler and Cannell, 1996; Exposito et al., 1994 참조) Cannell과 그 동료들이 개발한 것으로 상호작용 분석에 가장 흔히 적용되는 행동코딩(behavior coding) 역시 이 문헌에 큰 공헌을 했다.(예, Cannell et al., 1968; Marquis, 1969; Cannel et al., 1975). Morton-Williams and Sykes 또한, Morton-Willams, 1979; Morton-Williams and Sykes, 1984; Sykes and Morton-Williams , 1987)

14. 1. 4. 4 재면접 프로그램

재면접은 주어진 그룹에서 두 번째 면접을 수행하는 것을 포함한다. (Biemer and Forsman, 1992; Cantwell et al., 1992; Forsman and Schreiner, 1991) 목적(예, 현장작업평가 응답분산 및 응답편향 측정)과 두 번째 면접을 수행할 방법(예, 동일 응답자 대 가장 식견있는 응답자, 동일 면접원 대 다른 면접원, 동일 질문 대 개념적으로 유사한 질문)이 다양하게 존재하는 한, 면접 프로그램은 밀접한 관련이 있는 일단의 기법을 포함한다고 생각할 수 있다. 목적과 범위 면에서 재면접 프로그램은 상대적으로 다른 평가 방법보다 더 많은 자원의 투자가 필요하고, 몇몇 응답자를 제외할 수도 있다.(Blair and Sudman, 1993)

14. 1. 4. 5 사전조사 사례

Fowler(1992)와 그의 동료들(Cannell et al., 1989)의 작업은 명백하지 않은 조사 문항을 평가하기 위한 행동 코딩(그리고 응답 분포 분석)의 사용에 관해 특별한 교훈적 사례를 제공한다. 이들 조사자들은 여러 건강조사들의 질문들로 구성된 60개 항목의 설문을 개발했다. Fowler의 연구는 표준 행동코딩 절차를 통해 확인된 일곱 가지 이상의 문제성 문항을 표적으로 하고 있다.(예를 들어, "당신은 정기적으로 운동이나 스포츠를 즐깁니까?") 이러한 질문은 용어나 개념(예, 운동)이 제대로 정의되어 있지 않다. 그 문항들은 결과적으로 재관리되어진 새로운 질문지에 의해 수정되어 지고 구체화 되어진다.길고 어려운 질문은 몇 가지 짧은 질문으로 대체되어야 한다. 행동 코딩과 응답 분포 분석은 많은 문항을 유효하게 개선시킬 수 있는 것으로 드러났다. 그러나, 수정 후에도 몇가지 문항은 상태적으로 부적절한 응답을 가져올 확률이 높아, 명확한 수정이 요구되었다. 이는 조사 문항을 "결정"하는데 있어 두 가지 중요한 점을 지적한다. (1) 때로는 하나의 문제를 단정짓는 것이 다른 문제를 야기시킬 수 있다. 그리고, (2) 몇몇 질문은 정정하기가 매우 어렵고(예, 복잡한 질문), 그렇게 하는 것은 조사 문항 수를 지나치게 증가시켜 응답자의 부담을 가중시킬 수 있다.

행동 코딩(면접이 진행되는 동안 수행된)과 면접원 디브리핑 및 관측자 디브리핑에 포함된 건강 설문지의 초고에 관한 또 다른 재미있는 암시를 주는 연구는 Willis(1991)에 의해 수행되었다. 49개의 현장 면접 후, 관측자와 면접원은 특정 조사 문항의 문제점에 관한 자신의 의견을 보고했다. Willis는 디브리핑(debriefing)이 행동 코딩(62)보다 더 많은 문제성 문항(87)을 제공한다는 것을 발견했다. 또한, 문제가 있는 것으로 확인된 94개의 질문 중 59%는 두 가지 방법이 일치하였다. 두 방법 간 상관관계가 적은 점에 지나치게 집중하기 보다는 Willis는 매우 자극적인 발언을 했다. "행동 코딩이 디브리핑 보다 보존적인 방법이라는 상황에서는 두 가지 방법 모두에서 문제가 있는 것으로 드러난 질문들은 명백한 수정대상으로 보고, 디브리핑에 의해서만 확인된 문제성 질문들은 추가적인 가능성을 고려하는 것이 최선이다"(p.12, 이탤릭체 표시). 이에 대해서는 본 절의 후반부에서 다시 자세히 논하도록 하겠다.(14.3.1절 참조)

14.1.5.6 품질평가조사 사례

이들의 조사가 상이한 목적을 가지고 수행되었다고 하더라도(예를 들어, 두 가지 사전 조사 방법의 효율을 증명하기 위해), 최근 Oksenberg, Cannell, 그리고 그의 동료들(Oksenberg et al., 1991; Cannell et al., 1989)의 연구에서는 어떻게 훌륭한 품질평가 조사를 수행할 것인가에 관한 유용한 사례를 제시하고 있다.

Oksenberg 등이 보고한 연구에서는 조사자는 광범위한 의학적 사안과 질문 유형으로 된 60개 항목의 설문지를 만드는 것에서부터 시작했다. 모든 질문은 당시의 주요 건강조사에 사용되고 있었고, 아마도 몇 가지 방법으로 사전조사 되었을 것이다. 조사팀은 164명의 응답자에게 전화로 질문했다. 모든 면접에서 면접원과 응답자의 상호작용은 표준절차로 사용되는 행동코딩으로 되어 있었다. 결과는 놀랍다. 사전조사된 문항의 60%가 15% 이상의 부적절한 응답 코드를 가지고 있었다. 이 코드는 응답자의 답변이 질문 목적을 만족하기 못할 때 부여된다. 또한, 질문의 50%가 10% 이상의 설명을 요구받았다. 104회의 면접에서 목표 질문을 한 후에 4가지 유형의 특별한 추적 탐구 질문이 이루어졌다. 상이한 유형의 탐구가 이루어졌는데, 이해 탐구가 가장 유용한 것으로 나타났다. 이러한 탐구는 몇몇 응답자들이 최초에 별 어려움 없이 응답한 질문에 대한 문제를 드러낸다. 그러한 문제는 행동 코딩으로는 포착할 수 없다. 이러한 조사자들은 문항평가방법의 상호보완적인 측면을 인식하고, 제시하면서 문헌에 큰 공헌을 한다. 그리고, 이와 똑같이 중요한 사실은, 위의 사례 및 그들의 보고서에서 제시된 자료에 의해 직접적으로 묘사되는 것으로 전통적으로 제시되는 기법이 항상 조사 문항과 관련한 심각한 문제를 포착하지는 못한다는 것이다.

14.2 사전조사부터 품질평가까지의 과정: 사례연구로서 상시인구조사(CPS)의 재설계

1980년대 중반, 컴퓨터를 이용한 면접의 조사측정 및 진행에 인지심리학의 적용이 늘어나는 것에 고무된, 미국 노동통계청(BLS)과 미국 인구조사국은 사용자 편의를 위해 컴퓨터를 이용한 면접환경으로 상시인구조사를 재설계하기로 결정했다. 재설계에는 사전조사와 품질평가 방법론들이 모두 사용되었다. (Campanelli et al., 1991; Esposito et al., 1991, 1992; Rothgeb et al., 1991) 1986-89년 동안 먼저, 재설계 과정에 적용될 필요가 있는 CPS의 핵심적 문제를 확인하기 위한 예비조사가 수행되었다.(Copeland and Rothgeb, 1990 참조, 리뷰로는 Martin, 1987 참조). CPS 설문지의 문제를 확인하기 위한 연구에서는 BLS와 인구조사국 조사자들에 의해 다양한 기법들이 사용되었다.

- CPS 면접원들의 표적집단을 이용한 면접원 디브리핑(미국 노동통계청, 1988)
- 표적집단, 추적탐구문항 및 vignette를 이용한 응답자 디브리핑 (Palmisano, 1989; Campanelli et al., 1989; Campanelli et al., 1991)
- 카테고리 분류작업(Fracasso, 1989)
- 현장실험(Westat / AIR, 1989a, 1989b)

이러한 연구들의 결과와 다른 자료원들은(예, National Commission on

Employment and Unemployment Statistics, 1979; 미국 노동통계청, 1986, 1987, 1988) CPS 재설계 단계를 설정하는데 도움이 되었다.

14. 2. 1 CPS 재설계: 사전조사(1단계와 2단계)

CPS 재설계는 세 가지로 단계로 되어있다. 처음 두 단계는 CPS 설문의 대안적 항목을 평가하기 위해 재설계되었다.

14.2.1.1 제1단계

1단계 현장조사에서는 두 개의 대안적 CPS 설문지가(B, C형) 기존 CPS(A형)와 함께 현장검증 되었다. 그리고, 이러한 초기 사전조사 후에는 대안적 설문지들의 최상의 문항들이 하나의 대안 설문지로 통합되었는데(D형), 이는 제2단계에서 검증되었다. 이 초기단계에서는 (a) 면접원 디브리핑, (b) 현장근거 응답자 디브리핑, (c) 행동코딩, 그리고 (d) 항목근거 응답분석과 같은 사전조사 방법론이 사용되었다.

면접원 보고서(디브리핑)

1단계 조사에서는 면접원 디브리핑의 두 가지 측면을 이용하였다. (1) 자계식 설문지 작성, 그리고 (2) 다른 면접원과 함께 표적집단 토의에 참여가 그것이다. 이 두 가지 면접원 디브리핑 기법은 상이한 체제를 이용하다 하더라도, 그들은 동일한 정보를 수집하고자 하고, 결과적으로 동일한 하부 구조를 공유한다. 설문지와 운영자의 표적집단지침은 둘 다 특정 설문지(A, B, 혹은 C)에 대한 면접원의 일반적인 선호도에서 특정 질문이나 연속되는 질문의 특정한 평가쪽으로 진행되도록 구성된다. 면접원 디브리핑을 통해 조사자들은 특정 설문항목의 문제점에 대한 면접원들의 견해를 얻어낼 수 있게 된다.

응답자 디브리핑

응답자의 CPS 문항에 대한 이해 정도를 알아보기 위해, 네 번째와 마지막 월별 면접 후 가구별 응답자를 대상으로 현장근거 응답보고가 수행되었다. 이러한 사후 면접은 vignette 또는 일련의 적절한 추적탐구질문 중 하나로 간주된다. 예를 들어, 직업을 가진 사람이나 실직자의 vignette를 평가하기 위해서 주고(예, "Bill(jan)은 지난 주 정규수업에 출석하면서 남학생(여학생) 파티에 바를 봐주고 돈을 벌었다"), 응답자들에게 대상이(Bill 혹은 Jan) 직업인인지 혹은 실직자인지 구분하도록 요청하였다. 또한, 가구에서 보고한 사업이 BLS 사업의 기준을 만족하는지를 평가하기 위해 응답자들에게 몇 가지

추적탐구질문이 주어졌다.(예, "당신은 간판을 걸거나, 전화번호부에 등재하거나 혹은 신문을 통해 회사의 상품이나 서비스를 광고하는가?") 추적 질문을 수행하는데 대한 응답자의 적격성은 주요 면접 수행 도중 그들의 응답에 의해 결정된다. 응답자 디브리핑을 통해 조사자들은 개념과 질문에 대한 응답자의 이해도의 정량적 자료를 얻고, 이에 대해 적절한 수정을 할 수 있었다.

행동코딩

BLS와 인구조사국 조사자들은 특별히 개발된 형식과 코딩 절차를 이용해 (Cannell et al., 1975, 1989; Shepard and Vincent, 1991의 연구에 기초), 면접원과 응답자의 상호작용을 감시 및 코딩하였다. 이 절차들은 실제 면접 중에 면접원과 응답자의 상호작용을 코딩할 수 있도록 설계되었다. 모니터 요원은 질의표현(wording)의 사소한 변경 또는 중대한 변경으로 면접원이 질문을 있는 그대로 정확하게 읽었는지를 기록하였다. 질의표현은 의미나 의도가 바뀐 경우에 중대한 변경으로 코딩되었다. 또한 응답자의 행동들은 적절한 답, 양호한 답, 부적절한 답, 설명이 필요함, 간섭, 모름, 응답거절로 구분하였다. 행동코딩은 면접원(예, 정확한 질의표현과의 편차로 증명되는 것)이나 응답자(예, 부적절한 답변이나 설명요청, 간섭 등으로 판단되는 것) 에게 문제가 될 수 있는 항목을 확인하는 데 도움을 준다.

항목근거 응답분석

응답과 무응답을 구분하는 목적은 상이한 유형의 응답자에게 주어지는 질의표현이나 질의순서 상의 어려움을 알아보기 위한 것이다. *항목 무응답률*은 실질적인 답변을 제공하지 않은 사람들의 비율로 정의되었다. 여기에는 응답을 거부하거나 모른다고 답한 사람들도 여기에 포함된다. 또한, 응답거부율은 특정 질문에 대한 민감도 자료를 제공하고, "모름"이라는 응답률은 응답자의 난이도를 나타내며, *응답 분포*는 유사한 질문에 대한 응답분포에 주목하면서, 모든 조사문항에 대해 표로 작성하였다.

표 14.2는 2단계에서 테스트되는 D형 설문지의 CPS "근로(work)" 문항을 선택하기 위해 위에서 기술한 방법들이 1단계에서 어떻게 사용되었는지를 요약한 것이다. 나머지 CRP 질문들도 유사한 방식으로 평가되었다. 표를 보면서(뒤의 표 14.3) 독자들은 두 가지 점에 주목해야 한다. 첫째, 응답자 디브리핑 자료를(즉, 추적탐구질문) 제외하면, 평가적 방법은 조사자료품질과 관련한 간접적인 정보만을 제공한다. 근로에 관한 선택적 질문들에 대한 평가적 자료를(예, 행동코딩자료) 비교함으로써, 분석가는 어떤 질의표현이 양질의 자료를(예, 정확한 질문낭독율과 적절한 응답율이 높은 질문) 제공할 수 있는가를 추론하게

된다. 둘째, 분석적 방법은 항상 명쾌한 결론을 가져오지는 못한다. 개념이 복합적인 면을 측정할 경우에는 특히 그러하다.(예, 직업 개념) 우리는 현재의 사례가 조사연구원들이 일상적으로 직면하고 해결하는 보다 대표적인 유형의 문제들이라고 믿는다.

표 14.2 D형 설문지를 위한 근로(Work) 문항선택

선택 문항

A형: 집에서 한 일을 빼고, 당신은 지난 주 내내 어떤 일이라도 했습니까?

B형: 지난 주, 당신은 임금을 받거나 수익이 발생하는 일을 했습니까?

C형: 지난 주, 당신은 어떤 일이라도 한 것이 있습니까? 여기에는 임금을 받거나 아니면 기타 다른 종류의 보수를 받는 일이 포함되어 있습니까?

목적: D형 설문에서 근로의 개념을 가장 잘 조작할 수 있게 하고, 응답자와 면접원에 대한 문제들을 최소화해 주는 근로 문항을 선택하기 위함. (근로의 개념에 관한 기준에는 다음도 포함됨 기준 주간에 임금 내지는 임금과 비슷한 종류의 보수를 받거나 수익이 발생하는 일을 1시간 이상 한 경우 혹은, 가족기업이나 농장에서 무보수로 15시간 이상 한 일)

측정 사안: 응답자들의 "근로"에 대한 개념 해석과 근로활동의 보고어 질의표현이 미치는 영향을 결정한다.

평가적 방법론 및 기법의 결과

A. 행동코딩: 분석자료는 B형 질문을 선택하는 기초를 제공한다.

- 면접원이 질문을 정확하게 읽은 시간비율에 관한 선택적 근로문항 간의 유의차(A = 94.3%, B = 98.8%, C = 93.9%)
- 질문에 적절하게 답한 응답자의 비율에 관한 선택적 근로문항 간의 무의차(A = 90.9%, B = 95.6%, C = 91.9%)

A. 면접원 디브리핑: 보고사항을 보면, 면접원(및 응답자)들이 세 가지의 근로문항 모두에서 몇 가지 어려움을 겪었음을 알 수 있다.

(1) 표적집단. 일부 면접원들은 A형 질문이 주부들의 일을 무의미하게 만드는 것으로 들리고, 몇몇 응답자들을(예, 자원봉사자) 혼란스럽게 하기 때문에 선호하지 않는다고 보고하였다. B형 질문에서 "수익"이라는 단어의 사용은 일부 응답자들을 혼돈시켰다 - 특히 사업을 하지 않는 사람들. C형 질문에서 "기타 다른 종류의 보수"라는 구문은 일부 응답자 뿐만 아니라, 일부 면접원들에게도 혼돈을 가져왔다.

(2) 디브리핑 설문(면접원수 68명). 어떤 문항이 질문하기에 가장 어려웠냐는 질문에 2명의 면접원이 A형이라고 답했고(말이 장황하거나 세련되지 않다), 3명이 B형(헷갈리고, 애매하고, 이해하기 어렵다), 그리고 3명은 C형을(B형과 같은 이유) 택했다. 이번에는 어떤 질문이 응답자들이 답하기에 가장 어려운 것으로 보였냐는 질문에는 5명의 면접원들이 B형을(헷갈리고, 애매하고, 이해하기 어렵다) 택했고, 3명이 C형이라고(B형과 같은 이유) 하였다. 그리고, 응답자들이 주로 어떤 표현이나 개념을 잘 이해하지 못했느냐는 질문에는 6명이 "임금을 받거나 아니면 기타 다른 종류의 보수를 받는 일"이라고 답했고, 4명은 "임금을 받거나 수익이 발생하는 일" 또는 그냥 "수익"이라고 했으며, "일" 또는 "일과 직업의 차이"라고 한 면접원도 4명 있었다.

B. 응답자 디브리핑: 분석자료는 세 가지 형식의 질문 모두에 약간의 기초를 제공한다.

(1) 추적탐구문항. 세 가지의 근로 문항들은 모두 고용인들을 파악하는데 효과적이었다. 가능한 모든 질문 페어에서 누락된 고용인들의 비율간에는 유의한 차이가 없었다(A = 2.0%, B = 1.7%, C = 1.1%).

(2) Vignette. 세가지의 근로 문항 중 CPS 정의와 일치하는 응답을 잘 이끌어내는 것은 없었다(즉, 나머지 두 질문보다 뚜렷하게 나은 것으로 보이는 질문은 없었다). B형 질의표현이 모든 vignette 시나리오에 "아니오" 라고 답하는 응답자의 비율이 높다는 점은 다른 선택 문항들에 비해 총괄적이지 않을 수도 있음을 제시하는 증거가 될 수도 있다. B형 질문은 주변 근로활동들을(예, 가정에서의 일) 정확하게 분류하지 못한다는 점에서는 다른 질문들에 비해 성공적이지 못하나, 비근로 활동은(예, 자원봉사) 다른 질문에 비해 정확하게 분류하고 있다.

C. 항목근거 응답분석: 분석자료들은 서로 간에 더 나은 질문이 없음을 제시해 준다.

(1) 응답-분산 분석. 세 가지 근로 질문은 모두 일한 것으로 보고한 개인의 비율이 거의 동일하게 나타났다(A = 59.16%, B = 57.95%, C = 58.71%, 가능한 모든 질문 쌍에서 전술한 비율의 차이는 무의미했다).

(2) 무응답 분석. 세 가지 질문간에 항목 무응답은 거의 없었다(A = 0.18%, B = 0.18%, C = 0.22%)

권고 (R) 및 근거 (J)

R: D형 설문에는 B형 질문을 약간 수정한 형식을 택한다. "지난 주, 당신은 임금을 받는(혹은 수익이 발생하는) (것 중 하나의) 어떤 일이라도 했습니까?" - 가족기업이나 농장에 관한 앞의 질문에(즉, "가족 중에 사업을 하거나 농장을 운영하는 사람이 있나요?) "예"라고 답한 응답자들에게 읽어주는 삽입구. 면접원들에게는 "지난 주"라고는 기준 기간과 "어떤"이라는 단어를 강조하도록 한다.

J: 응답분석과 응답자 디브리핑은 결론에 이르지 못했다. 다시 말해, 어떤 질문이 다른 질문들보다 더 낫다거나 더 나쁘다는 것을 제시해 줄 증거나 거의 없었다. 그리고, 행동코딩 분석은 B형 질문을 선택하는데 기초가 되었다. 한편, 면접원 디브리핑은 세 가지 질문 모두에 문제가 있음을 보여주었고, B형 질문에서 "수익"이라는 단어에 대한 혼돈은 가족 중에 누군가가 사업을 하거나 농장을 운영하는 경우에만 적용되게 하면 쉽게 해결된다.

14.2.1.2 제2단계

2단계 현장시험에서는 새롭게 합성된 설문지(D형)가 기존 CPS와 대조 검정되었다. 2단계의 주목적은 수정된 CPS 설문지의 개발을 완료하기 위해 질의표현에서 문제가 되는 부분을 파악하는 것이었다. 자료의 평가는 D형 설문지에 "치명적인 결함"이 있는지를 확인하고(예, 질의나 개념이 이해되는 과정에서의 총오차) 이를 수정하는데 사용되었다. 그 결과, 몇 가지 방법론들이 다르게 사용되었다. 예를 들면, 2단계에서는 표적집단에 대해서만 면접원 디브리핑이 이루어졌고, 분석범위도 보다 협소했다. 한편, 가장 큰 관심은 경제활동인구(labor force) 분류에(예, 근로 및 실직과 관계있는 항목) 직접적인 영향을 미치는 조사문항이었으며, 이러한 관심은 조사지침(프로토콜)에 반영되었다. D형 설문에 문제성 항목이 있는지를 파악하는데는 행동코딩이 수행되었다. 대조 설문에(A형) 있는 항목들은 평가되지 않았다. 또한, 응답자 디브리핑과(추적탐구 및 vignettes) 응답분포분석은 1단계와 비슷했다. 그리고, 이러한 자료들을 활용하면서 A형과 D형간의 차이를 검토하였다. 상기된 여러 가지 사전조사 기법들에 의해 생성된 자료를 바탕으로 조사자들은 D형 설문에서 심각한 결함의 증거를 찾았지만, 아무 것도 발견하지 못했다(표 14.3 참조). 경미한 질의표현 변경만 권고되었다.

14.2.2 CPS 재설계: 품질평가조사 (제3단계)

내용(content) 수정 및 새로운 자료수집기술이 경제활동인구 추정치에 미치는 영향을 파악하기 위해 수정된 CPS 설문이 1992년 7월부터 기존 CPS(월마다 60,000가구)와 병행하여 개별조사(월마다 12,000가구)에서 시험되었다. 이 병행조사는(제3단계) 이전 단계들과 몇 가지 면에서 달랐다. 예를 들어, 3단계 표본은 주소록에 있는 가구들로 구성되었으며, 1월과 5월에는 개인면접을 그리고, 이후에는 전화조사를 실시하였다. 또한, 새로운 설문과 새로운 자료수집기술이 경제활동인구 추정치에 미치는 효과를 평가하는 것 외에 3단계에서는 수정된 설문에서 생성된 자료의 품질을 평가하기 위해 면접원 디브리핑, 행동코딩, 응답자 디브리핑 및 항목근거 응답분석을 다시 수행하였다. 사실, 지금이라도 14.1.2에 제시된 구분이 있었다면, 조사자들은 사전조사없이 수정된 CPS 설문에 가해진 변경의 품질평가만 수행했을 것이다.

14.2.2.1 3단계 방법 및 결과

이 단계에서 면접원의 보고를 듣는 데는 표적집단과 표준화된 면접원 디브리핑 설문이 모두 이용되었다. 면접원 디브리핑의 주목적은 수정된 설문의 문제점과 추가적인 면접원 교육이나 매뉴얼 개선이 필요한 절차에 관한 문제점들을 확인하는 것이었다. 표적집단에게는 면접원들이 수정된 설문과 새로운 자료수집기술을(예, 랩탑 컴퓨터) 최소한 3-4개월 정도 사용한 후에 수행되었다. 면접원들에게는 응답자로부터 적절한 답을 얻기가 가장 힘들었던 질문을 파악하도록 요청하였다. 또한, 문제점의 정도에 관한 척도를 얻기 위해 파악된 문제성 항목 각각에서 어려움이 있었던 시간비율을 평정형식을 이용하여 추정하도록 요청하였다. 그런 다음, 표적집단에 참여한 면접원들로부터 확인된 가장 문제성 있는 질문들을 표준화된 디브리핑 설문에 포함시켜 CPS 전 면접원들에게 보냈다. 그리고, 이 자계식 디브리핑 설문에서는 면접원들에게 특정한 문항을 질문했을 때 적절한 답을 얻는데 어려움이 있었던 시간의 비율을 추정하도록 요구하였다.

3단계에서는 행동코딩을 이용하여 여러 가지 조사문항에서 면접원-응답자 상호작용을 조사함으로써 수정된 설문의 품질을 평가하였다. 여기서 가정한 것은, 면접원이 질문을 있는 그대로(혹은 조그만 변경하여) 낭독하는 비율이 높고, 응답자가 적절한(혹은 양질의) 답을 제공하는 것이 자료품질에 긍정적인 공헌을 할 수 있다는 것이었다. 지정된 일단의 면접원들은 응답자의 동의를 얻어 364개의 병행조사 면접을 테이프에 녹음하였다. 이와 같이 녹음된 면접의 수가

너무도 방대한 까닭에, CPS에 경험이 있는 4명의 감독요원이 행동코딩을 수행하는 교육을 받았다. 코딩요원은 사무실 기반의 코딩에 적용되는 표준화된 코딩형식을 사용하였다. 내용면에서 이 형식은 이전 시험단계에서 사용된 것과 아주 흡사하다.

3단계에서 응답자 디브리핑은 문항 이해도와 응답 정형성을 평가하기 위해 수행되었다. 또한, 다른 단계에서와 마찬가지로 추적탐구문항은 가구 구성원 중 해당되는 사람에게만 질문하였다. 그러나, vignette은 사용하지 않았다. 이러한 디브리핑 질문으로부터는 응답자들이 "근로" 문항을 어떻게 해석하는지에 대한 제시 외에 몇 가지 응답 정확성의 척도를(예, 실업자료) 얻을 수 있었다(Rothgeb, 1994).

응답분포분석은 수정된 설문과 새로운 자료수집기술이 2단계와 동일한 응답패턴을 산출했는지를 파악하는데 부분적으로 수행되었다(이 자료들은 기존 설문(A형)과 수정된 CPS 설문에서 산출된 경제활동인구 분류간의 차이를 비교하는데도 중요하다). 응답거부 또는 "모름"이라는 응답이 보통 보다 높게 나타날 것으로 의심이 가는 항목에 대해서는 항목 무응답 자료도 산출하였다. 그리고, 질의경로가 긴 자주 묻지 않는 항목에 관한 분석에는 충분한 사례가 쌓일 수 있도록 12개월의 기간에 걸쳐 자료를 누적시켰다. 그 외, 다른 방법을 통해(예, 응답자 디브리핑) 수집된 자료는 응답패턴간의 차이가 발견됐을 때, 이러한 차이점을 설명하는데 유용하였다(Rothgeb, 1994).

사전조사와 함께 품질평가의 척도는 정성적 자료와 정량적 자료를 모두 제공한다는 사실을 깨닫는 것이 중요하다. 자료품질 합격선은 평가되는 문항간에 서로 다를 수 있다. 즉, 개념구성에 절대적으로 중요한 항목들은 허용가능한 자료품질에 보다 엄격한 기준이 필요할 것이다. 아울러 여기서는, 위에서 논의된 품질평가방법들은 대개는 상호보완적인 자료를 제공하지만, 때로는 맞지않는 자료를 산출할 때도 있다는 사실을 기억해야 한다. 그러나, 다양한 자료원들의(즉, 코딩요원, 면접원, 응답자 그리고, 설문설계자) 독자적인 시각들이 있다는 것을 가정하면, 일부 부적절한 사항들은 그 방법들이 품질평가목적에 부적합하다는 것을 의미하는 것으로 간주해서는 안 된다.

사전조사에서 품질평가까지의 과정을 보여주기 위해 우리는 계속해 앞서 기술한 방법들이 CPS "근로" 문항의(수정된 문항: "지난 주, 당신은 임금을 받는(혹은 수익이 발생하는) 어떤 하나의 일이라도 했습니까?") 평가에 어떻게 사용되었는지에 관한 예를 들 것이다. 우리가 "근로" 문항을 예로 들어 설명하는 이유는 분석자료가 앞에서 말한 모든 평가기법에 유용한 몇 안되는 CPS 문항 중

하나가 이것이기 때문이다. 또한, 이것은 가장 다양한 사람들에게 질문된 CPS 문항이기도 하다.

표 14.3에는 세 가지 다른 시험단계 각각에서 얻은 CPS "근로" 문항에 관한 평가자료가 나와 있다. 그 중 3단계의 결과는 이 문항에서 얻은 조사자료가 꽤 높은 품질의 것이었음을 제시한다. 즉, 행동코딩자료는 면접원들이 거의 모든 시간에 수정된 "근로" 문항을 정확하게 또는 약간 변경하여 읽었음을 보여준다. CATI를 통해 조사한 응답자들은 98%의 시간 동안 적절한(혹은 양질의) 응답을 제공한 반면, CAPI를 통한 개별면접에서는 93%의 시간에 적절한(혹은 양질의) 응답을 제공하였다.

표 14.3 CPS 재설계의 사전조사(PT) 및 품질평가(QA) 단계들로부터 얻은 "근로" 문항에 대해 선별된 결과

					응답자 디브리핑	행동코딩		응답분포
시험/설계	날짜	표본크기	설문 버전	면접원 디브리핑	누락 취업율	INT 코드 (%E + mC)	RSP 코드 (%AA+qA)	"예" 응답 % (% NR)
제1단계 CATI/RDD (PT)	1990. 7 - 1991. 1	70,000hhlds 누적 모든 버전	A	표 14.2	2.0%	94% CATI	91% CATI	59.16% (0.18%)
			B	표 14.2	1.8%	99% CATI	96% CATI	57.95% (0.18%)
			C	표 14.2	1.1%	94% CATI	92% CATI	58.71% (0.22%)
제2단계 CATI/RDD (PT)	1991. 7 - 1991. 10	32,000hhlds 누적 두 버전	A	-	3.8% (2.2% 유급)	-	-	57.74% (0.15%)
			D	FG: "단지 내 직업"	2.6% (2.0% 유급)	100% CATI	95% CATI	57.01% (0.08%)
제3단계 CATI/CAPI 주소목록 표본 (QA)	1992. 7 - 1993. 12	144,000 hhlds 누적 (1993년만)		FG: 35.5% IDQ: 18.2%	2.9% (1.6% 유급)	100% CATI 99% CATI	98% CATI 93% CATI	58.58% (0.16%)

약어: FG는 표적집단 자료 ; IDQ는 면접원 디브리핑 설문자료 ; INT는 면접원 ; RSP는 응답자 ; (% E + mC)는 정확하거나 조금 변경한 질문 낭독율 ; (%AA + qA)는 적절한 양질의 응답율 ; (%NR)은 무응답율(즉, 응답거부와 "모름") ; hhlds는 가구

면접원 디브리핑에서 얻은 자료는 혼재되어 있다. 표적집단 참여자들은 응답자들이 수정된 "근로" 문항에 답할 때 어려움을 느끼는 것으로 보인다고 보고하였다. 실제로, 평가된 모든 문항 중에 이 문항이 적절한 응답을 얻는데 가장 어려운 문항이라고 평정되었다. 표적집단의 과정에서는 면접원들에게 다음과 같이 요청하였다. "이 문항을 질문할 때 응답자들로부터 적절한 답을 얻는데 어려움을 겪은 시간은 몇 퍼센트였습니까?" 그 결과, 평균평정은 35.5%, 중간평정은 34%의 값을 얻었다. 이와 똑 같은 질문은 총 400명 중 345명의(86%) 현지 면접원들이 평정한 자계식 면접원 디브리핑 설문에도 주어졌는데, 여기서는 18.2%의 평균평정과 10%의 중간평정이 산출되었다. 여기서 흥미로운 사실은 표적집단 참여자들이 평균적으로 거의 36%의 시간 동안 근로 문항에 적절한 답을 얻는데 어려움을 겪은 것으로 보고한 반면, 행동코딩자료에는 응답자들이 "적절한 답"을 제공한 시간의 비율이 매우 높게 나타났다는 사실이다(CATI의 경우, 98%, CAPI의 경우 93%). 이러한 모순은 실제 CPS 면접을 들어보면 해결할 수 있다. 일부 응답자들은 관련정보를 제공하거나("글쎄요, 뭐 저의 일정한 직업이겠죠.") 질문을 반문함으로써("저의 일정한 직업 말입니까?") "근로" 문항에 대한 답변을 한다.(이것은 "보고하는(reporting)" 행동 즉, 직접적인 답변보다는 관련성 있는 정보를 제공함으로써 질문에 답하는 하나의 예로 볼 수 있다) 이러한 응답을 듣게 되면, 대부분의 면접원들은 탐구(probe) 하거나 "예"라는 항목에 체크한다. 우리의 훈련된 행동코드 부여자들도 이러한 응답들을 적절한 응답으로 코딩했다. 왜냐하면, 둘 중 어떤 답이든 응답자가 하나의 직업을 가지고 있고, 아마 대부분이 지나 주에 그 직장에서 일했을 것이라는 사실을 암시하기 때문이다. 그러나, 되돌아보면(면접원들이 기록한) 그러한 응답들은 어쩌면 단지 최소한도에서 적절한 답으로만 간주해야 한다. 그렇다면, 지금의 이러한 맥락에서는 질문 자체가 일부 응답자들을 혼돈시킨 것으로 보인다. 그러나, 다른 자료는 우리가 이 질문으로부터 정확한 정보를 얻고 있음을 제시해 준다.

위에서 논의된 결과들은 한 가지 이상의 평가방법을 활용하는 것이 얼마나 중요한 것인가에 관한 좋은 예를 제시한다. 면접원들은 자신들이나 응답자들이 조사문항에서 느끼는 문제점들에 관한 귀중한 정보원이기도 하지만, 때로는 이들이 제공하는 정보는 가장 특징적인 면접상황에 의해 윤색되기도 한다. 만약, "근로" 문항에 관한 자료품질의 평가가 면접원 표적집단의 결과에만 의존해 왔다면, 후원자(BLS)는 그 결과에 놀랐을 수도 있다. 하지만, 다행히도 행동코딩을 비롯하여 응답자 디브리핑 및 응답분포와 같은 기타 평가방법들로부터 보다 객관적인 품질평가자료를 얻을 수 있었다. 면접원 표적집단으로부터 얻은 평균 및 중앙값과 표준 자계식 설문에서 도출된 평균 및 중앙값 간의 차이는 집단적 행동(collective behavior)이 문제점 보고에 영향을

미칠 수도 있음을 시사한다. 즉, 표적집단 참여자 중 일부는 처음부터 "근로" 문항이 문제가 있다고 생각하지 않았을 수도 있다. 하지만, 다른 사람들이 문제가 있는 것으로 보고하는 것을 듣고는 그것을 각자의 평정에 반영했을 수도 있다. 또, 일부 면접원들은 이 문항에 어려움이 있었다는 사실을 잊고 있다가 그룹토의에서 그 문항이 문제가 있는 것으로 파악한 다른 사람들의 얘기를 듣고 그러한 어려움이 다시 생각났을 가능성도 있다. 이와 반대로, 디브리핑 설문은 각자 독립적으로 작성하였다. 면접원들은 다른 사람들의 의견을 듣지 않고, 각자가 경험한 문제의 빈도를 과소평가했을 수도 있다. 따라서, 우리는 "근로" 문항의 실제 난이도(어려움의 정도)는 디브리핑 설문과(10%) 표적집단에서(34%) 얻은 중앙값들의 사이에 있지 않을까 하고 추측하는 바이다.

응답자 디브리핑을 분석한 결과는 수정된 "근로" 문항이 고품질의 자료를 생산하고 있음을 제시해 준다. 주변근로활동과 가족기업 또는 농장에서의 무보수 근로 및 지하경제에서의 근로는 지난 몇 년 동안 CPS에 지나치게 적게 보고됐을 수 있다는 우려의 목소리들이 있었다(Martin and Polivka, 1992). 수정된 설문은 이러한 근로 유형의 보고를 개선시키기 위해 설계되었다. 누락된 취업인구 추정치를 얻기 위해 본조사에서 취업된 것으로 보고되지 않은 사람들에게는 응답자 디브피링에서 기준 주간에 단 몇 시간이라도 한 일이 무엇이라도 있는지에 대해 물었다. 이것은 본조사에서 일하는 것으로 보고되지 않았지만, 실제로는 기준 주간에 어떤 유형의 근로활동을 했던 사람들의 추정치를 얻기 위한 목적이었다. 그 결과, 응답자 디브리핑에서는 본조사에서 취업된 것으로 보고되지 않은 사람들 중 단 1.6%만이 기준 주간에 어떤 보수를 받는 일을 했던 것으로 보고되었다.(무보수 근로를 포함하면 2.9%). 이러한 자료는 누락된 취업율이 1.6%와 2.9% 사이의 값임을 보여 준다. 여기서, 우리는 누락된 취업율의 보고는 사전조사 기간보다 3단계 중에 약간 높았다는 사실에 주목해야 한다. 그러나, 1,2단계에서의 표본은 3단계의 표본에서와는 다르게 전화없는 가구는 포함하지 않는다. 일상적이거나 비공식적이거나 혹은 주변적인 근로활동에 종사하는 사람들의 비율은 전화가 있는 가구들보다는 전화가 없는 가구들이 높을 가능성이 많다. 이러한 활동들은 본조사에서 근로로 보고되지 않을 수 있고, 그 결과 누락된 취업율의 보고건이 보다 많아지는 결과를 초래한다.

응답분포자료를 보면 초기의 근로문항에 답한 사람들 중 58.6%가 지난 주에 일을 한 것으로 보고되어 있다. 이러한 자료는 이전 단계에서 얻은 추정치와 일치했다. 이 문항에서 항목 무응답은 없었다.

정리하자면, 3단계에서 수집된 분석자료는 다음을 제시해 준다 : (a)면접원들은 수정된 설문에 있는 항목들 대부분을 제대로 읽었다. (b)응답자들은 대부분의 항목들을 의도대로 이해하고 있었다. 그리고 (c)경제활동(labor force) 오분류는(예, 모비율에 대한 누락된 취업율과 취업율의 비로 측정되는 것) 수정된 설문에서 비교적 낮았다(Rothgeb, 1994; Polivka, 1994; Dippo et al., 1995). 우리는 여기서 더 많은 다른 표준척도들도 자료수집 프로세스를 감시하는데 이용되었음을 기억해야 한다(예, 재면접의 총차율과 순차율 ; 설문, 항목 및 면접원별 무면접율).

14.3 맺음말

앞선 두 절에서 우리는 조사설문을 평가하는데 이용되는 여러 가지 방법들에 관한 배경자료를 제시하였다(즉, 사전조사와 품질평가조사를 통해). 아울러, 여러 방법론들이 재설계된 CPS 설문을 평가하는데 실제로 어떻게 활용되는지에 관한 사례를 제공함으로써 그러한 자료를 보다 구체화하려고 노력하였다. 본 마지막 절에서는 주요 사회경제조사에 이상적인 품질평가 프로그램의 폭넓은 개요를 기술하고, 그러한 프로그램과 연관된 몇 가지 실제의 문제를 처리하며, 품질평가조사와 관련된 비용편익을 고려하고자 한다.

그러나, 본론으로 들어가기에 앞서 품질평가 프로그램의 몇 가지 용도에 대해 미리 언급하도록 하겠다. 그 첫번째는 면접원 교육개선에 관한 방안을 제시하는 용도다. 두 번째는 전체 생산표본을 이용하여 사전조사 결과의 신뢰도를 점검하는 용도다(예, CPS 재설계 3단계). 그리고 세 번째는 기존 설문이 제공하는 자료품질의 정도를 평가하는 용도다. 심각한 자료품질 문제들이 해결되지 않는다면, 조사 후원자들은 (1) 재설계된 설문을 사전조사하기 위한 공식적인 계획을 세우거나, 아니면 (2) 품질평가자료만을 기초로 수정을 할 것이다.

14.3.1 이상적인 품질평가프로그램

설문자료의 타당성과 신뢰도에 기여하는 한 가지 수단으로 우리는 중요한 사회경제지표를 제공하는 포괄적이고, 지속적이며, 다방법, 다기법의 조사품질평가 프로그램을 주장하는 바이다. 여기서, 먼저 포괄적(comprehensive)이라는 것은 측정오차가 다양하고, 상관적인 원천들을(즉, 면접원, 응답자, 자료수집 문맥; Esposito and Jobe, 1991 참조) 조사하는 조사프로그램과 실험실 조사, 사무실 조사, 현지 조사를 통합하는 조사프로그램을 의미한다. 그리고, 지속적(on

going)이라는 것은 조사도구의 정기적이거나 주기적인 평가를 포함하는 프로그램을 의미한다. 이것은 Dippo(제20장)가 기술한 지속적인 측정의 개념과 일치한다. 또한, 다방법론(multimethodology)이라는 것은 다각적인 방법과 원천 - (1)면접원, (2)응답자, (3)조사 후원자 및 설계자, (4)면접원과 응답자 간의 자연스러운 문맥 상호작용 - 으로부터 분석자료를 활용하는 프로그램을 말한다. 끝으로, 다기법(multitechnique)이라는 것은 앞서 말한 원천들로부터 정보를 수집하는데 자원을 활용할 수 있는 다각적인 기법들을(예, 표적집단, 디브리핑 설문) 포함하는 프로그램을 의미한다. 우리의 관점에서는 다방법의 기준이 매우 중요하게 보인다. 왜냐하면, 이것은 조사자들에게 네 가지 핵심적인 시각에서 평가자료를 제공하기 때문이다. 조사문항에 관한 잘못된 의사결정의 기회는 단지, 한 두 가지가 아니라 모든 시각이 고려될 때 최소화된다. 아울러, 다기법 기준도 조사자들로 하여금 각자의 기법들에 약점을 보완하면서 동시에 실력을 쌓아나갈 수 있게 해 주기 때문에 이 역시 중요하다.

앞서 말한 지속적인 조사활동 프로그램은 고가일 수도 있음은 인정된다. 그러나, 품질자료를 생산하는 비용을 고려한다면, 품질자료를 생산하지 못하는 비용도(예, 사용자에 대한 비용) 반드시 고려해야 한다. 비록 그렇다 하더라도, 우리는 주어진 현재의 재정적 상황에서 품질평가프로그램은 반드시 합당하게 경제적으로 구현되어야 한다고 생각한다. 따라서 이러한 제약조건을 생각하면서 우리는 실험실과 행동코딩을 공생하는 방식으로 통합하는 유연성 있는 프로그램을 제안하는 바다. 이 프로그램의 핵심은 세 가지의 현장방법론과(면접원, 디브리핑, 응답자 디브리핑, 행동코딩) 다양한 실험실 및 사무실 기반의 기법들로 구성되어 있다는 것이다. 이러한 프로그램은 기존 품질평가척도를(예, 재면접 프로그램) 대체하기보다는 보완하는 쪽으로 설계된다. 품질평가프로그램에서 사용되는 현장방법론은 (a)조사자들에게 설문과 관계된 문제점이 어디에 존재하는지를 알려주고, (b)그러한 문제점들의 설명을 제공하며, (c)조절된 실험실 환경에서 검정될 수 있는 문제의 원인에 관한 가설을 제시해야 한다. 또한, 원인가설을 검정하는 것 외에 실험실 및 사무실 기반의 조사는(예, 인지적 면접, 인지적 형태의 평가) 자원이 허용하는 경우, 후행 현지조사에서 처리될 수 있는 가설집합을 도출할 수 있다.

이러한 권고와 함께 우리는 여기서 앞서 말한 품질평가프로그램에 관한 세 가지의 실제사안을 처리하고자 한다.

1. *특정한 평가방법이 주어졌을 때, 조사자는 주어진 설문항목에 "문제"가 있는 지를 어떻게 결정하는가?* 실제로, 특정한 설문항목에 무엇이

"문제"인가에 관한 결정은 조사개념 및 그러한 개념들이 주어진 문항에서 조작화된 방법에서부터 출발한다(그림 14.1 참조). 만일, 기본적인 개념이 잘못 정의되어 있거나, 혹은 정의되어 있지 않다면, 조사자는 특정한 개념들이 주어진 문항에서 어떻게 조작화되었는지를 판단한 방법이 없고, 응답자들이 질문을 제대로 이해하는 지를 정확하게 파악할 수단이 없다. 또 만일, 구체적인 질문목적이 없다면, 조사자들은 응답자의 답변을 평가할 근거를 가지지 못한다(이 사안에 관한 확대논의는 Schwarz(제1장)와 Hox(제2장) 참조). 그러나, 명확한 질문목적이 있다 하더라도, 주어진 문항에서 문제가 되는 것의 기준을 정하는 것은 주관적인 과정이 되기 쉽다. 게다가, 이러한 기준은 종종 조사자들마다 그리고, 방법들마다 각기 다른 경우가 많다. 예를 들어, 행동코딩의 경우, 문제성 문항을 파악하는 기준은 해당 기법에 실질적인 경험이 있는 조사자들에 의해 제시된다. Cannell(1989) 등은 특정 행동코드가(예, 설명요청) 질문시간의 15% 이상에서 나타나는 경우를 문제성 문항이라고 하였다. 많은 실무자들은 이러한 경험규칙을 따른다. 반면, 다른 조사자들은(예, Esposito 등, 1991) 보다 엄격한 기준을 사용했다(예, 응답자로부터 질문시간의 최소 90% 이상에서 적절한 응답을 얻지 못한 경우를 문제성 문항이라고 했다). 추적탐구문항으로부터의 자료 분석은 일반적으로 탐구문항이 현재 고려중인 개념의 선택적 척도로 설계되고, 탐구문항과 목표문항 사이의 실질적인 응답 분포차가 문제의 신호로 받아들여진다는 점에서 볼 때 보다 객관적이라고 할 수 있다. 다른 기법들을 보면 면접원이나 응답자가 주어진 설문항목에서 경험한 어려움을 상기하고, 이후 조사자들은 그러한 판단에 어느 정도의 가중치를 부여할 것인지를 반드시 결정해야 한다. 요약하자면, 경험과(Esposito 등, 1991, 1992) 문헌고찰을(예, Cannell 등, 1989; Presser and Blair, 1994; Willis, 1991) 바탕으로 우리는 모든 평가방법들이 다들 약점을 가지고 있다고 믿는다. 더욱이, 문제성 문항인가 아닌가에 대한 결정은 주관적인 과정이기 쉽다. 그리고, 이것이 바로 우리가 하나의 평가방법계획은 매우 위험스럽다고 생각하는 이유다.

2. *다각적인 평가방법들을 활용할 때(예, 행동코딩, 추적탐구문항, 면접원 디브리핑 설문), 조사자들은 "문제성" 설문항목을 어떻게 확인하는가?*
최소한 다른 전략들이 개발되기 전까지는 유용한 한 가지 전략은 우리가 상대신뢰모델(relative confidence model)이라고 하는 것에 의존하는 것이다. 이 모델은 Willis(1991)가 제안한 개념을 확대시킨 것으로, 다음과 같이 해석한다. 각기 다른 평가방법들은 서로를 상호 보완하는 것으로 바라볼 수 있다. 그리고, 함께 사용되는 경우 복수의 방법은 단일 방법에

비해 문제성 문항을 파악하는 보다 정확한 수단을 제공한다. 이 모델은 벤다이어그램으로 설명된다(그림 14.2 참조). 각각의 원은 *다른 원천*에서(예, 면접원, 응답자, 전문가) 정보를 도출하는 서로 다른 설문 평가방법을 나타낸다. 먼저, A 모델을 보면 하나의 평가방법이 특정 *그룹의 문항*들을 문제가 있다고 파악하였다(영역 1). 이러한 한 가지 방법만이 주어졌을 때는 원 밖에 있는 문항들을 문제가 있는 것으로 간주할 아무런 근거가 없다(영역 0). 다음으로 B모델을 보면, 두 가지 평가방법들이 사용되었고, 세 영역이 원으로 구획되어 있다. 그 중 영역 2에 있는 문항들은 두 가지 방법 모두에서 문제가 있는 것으로 확인된 것들이다. 영역 1에 있는 문항들은 한 가지 방법에 의해서만 문제가 있는 것으로 확인된 것이다. 그리고, 이 영역 밖에 있는 것은 어느 방법으로도 문제가 있는 것으로 확인되지 않은 문항들이다(영역 0). 검토 및 수정가능한 문항을 선택할 때, 우리는 영역 2에 있는 문항들을 문제성으로 선택할 때 가장 확신에 차 있었으며, 영역 1에 있는 문항들은 그 확신이 다소 떨어진다는 것을 느꼈다. 끝으로, C모델은(세 가지 평가방법) 동일한 논리체계를 따른다. 검토 및 수정을 위한 문제성 문항을 정확하게 선정할 때의 우리의 확신은 영역 3에서 가장 높았고, 영역 2와 영역 1로 갈수록 점차 낮아졌다(상대신뢰모델을 뒷받침하는 논리체계는 하나의 원천으로부터 평가정보를 도출하는 다기법 비교에도(예, 디브리핑 설문과 표적집단을 이용하여 면접원에게서만 정보를 수집하는 것) 일반화될 수 있는 것은 아님을 주의해야 한다. 우리는 원 사이에 중복이 더 많을 것으로 기대하지만, 분석은 단일 원천의(즉, 면접원) 시각과 사용된 기법들의 특정한 약점에만 한정되었다.)

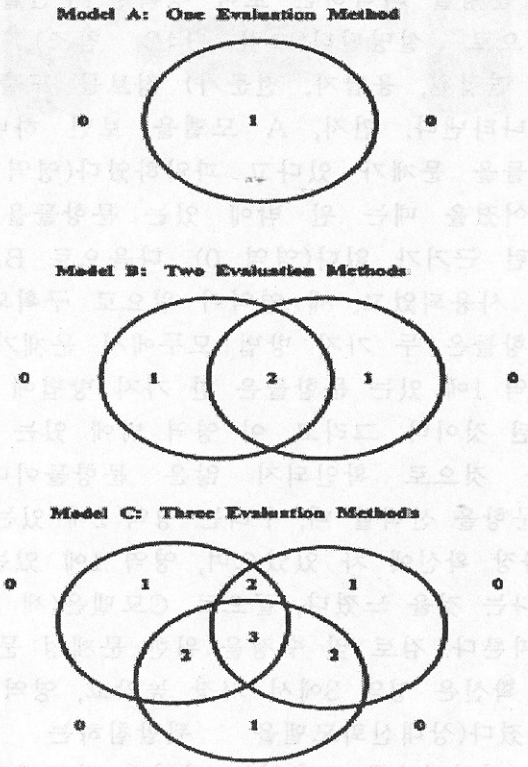


그림 14.2 문제성 조사 문항을 확인하기 위한 상대신뢰모델

3. *문제성 문항이 정확하게 파악되었을 때(설문에 있어서 수정을 원할때), 조사자는 그러한 문항의 수정을 어떻게 진행하는가?* 우리의 전략은 앞서 말한 CPS 재설계 1단계과정과 비슷한 시험-확정-재시험 모델(test-fix-retest model)을 이용하는 것이다. 이 모델에서는 초기 사전조사 단계에서 확인된 문제들을 바탕으로 문항을 수정한 후 두 번째 평가단계에서 그 문항을 재시험한다. 그러나, 대개 문항설계자는 하나의 문제를 해결하기 위해 하나의 문항을 수정하면서 다른 것들을 생성시킨다. 우리에게는 실증적 자료를 바탕으로 양질의 조사문항을 설계할 수 있는 일단의 지침들이 절실하게 필요하다. 이러한 점에서 Belson (1981, p.389)은 몇 가지 유용한 안을 제시한다. 예를 들면, 그는 우리가 피해야 하는 것을 제시한다. "용어 정의가 아주 다르거나 많은 문항을 올려놓는 것, 응답자의 일상적인 작업도구가 아닌 용어들의 사용, 응답자에게 하기 어려운 임무를 주는 것, 응답자에게 많은 기억이 필요한 일을 주는 것" 또 다른 가능성은 본 설문에 덧붙이기 전에 인지적 형태의 평가를 이용하여 수정된 문항을 평가하는 것이다(Forsyth 등, 1992; Forsyth and Hubbard, 1992).

14.3.2 품질평가조사의 비용과 편익

조사품질은 비용이 발생하지 않고는 달성될 수 없다(Groves, 1989). 앞서 말한 종류의 주요 사회경제조사를 위한 품질평가프로그램은 지속적인 인적, 재정적 자원의 투입을 포함하고 있다. 더욱이, 조사품질의 개선에는 일반적으로 자료수집 프로세스의 몇 가지 변화가 들어 있다(예, 문항 또는 절차상의 수정). 프로그램 관리자들은 비록 품질개선을 위한 척도는 환영하지만, 또한 반드시 자료 사용자들을 위한 상품의 유용성도 함께 고려해야 한다(예, 시계열 보전). 이와 같이 경합하는 수요들은(품질수준을 유지하는 것과 시계열 통계의 비교가능성을 보장하는 것) 프로그램 관리자들을 곤란하게 만든다. 바꿀 것이냐 아니냐는 이들의 딜레마이며, 조사자들이 가볍게 받아들일 사항이 아니다. 프로그램 관리자들은 품질 관련한 개선이 권고될 때마다, 조사자들이 그러한 변화의 예상된 결과에서 자신들의 통계상품과(예, 실업률, 소비자물가지수, 빈곤지수) 조사의 조작적 측면에(예, 자료처리편집) 자료를 제공할 것을 기대할 모든 권리가 있다.

물론, 품질평가조사에 자원을 투자할 때는 실질적인 편익들도 있다. 그 중 가장 분명한 사용자 편익은 보다 타당성 있고, 신뢰도 높은 통계와(즉, 보다 정확한 자료) 그러한 통계에 기초하여 이루어지는 의사결정의 보다 큰 확신이다. 또한, 이러한 사회경제자료를 생산하는 조직의 입장에서 분명한 편익은 향상된 공신력일 것이고, 조사단체의 편익도 마찬가지일 것이다. 이와 비슷한 것으로는 임상심리학에서의 한 가지 원칙이 있다. 정상적인 인간의 행동을 명확하게 이해하길 원한다면, 비적응 행동패턴을 연구하고, 그 기원을 파악하려고 노력해라. 이와 비슷한 원칙은 조사응답 모델과 설문설계에도 적용될 수 있다. 응답자가 조사문항에 어떻게 응답하는가와 양질의 자료를 생산하는 설문을 어떻게 설계하는가를 이해하길 원한다면, 문제성 문항을 확인하고, 그 문제가 왜 존재하는 지를 파악하려고 노력해라. 우리는 품질평가조사의 주된 편익 중 하나가 조사자들이 조사문항에 응답하는데 포함된 프로세스들을(사회적, 심리적) 이해하는 것을 돕고, 또 그렇게 함으로써 보다 나은 설문을 설계하는데 있어 우리 스스로를 돕는 것이라고 믿는다.

감사의 글

본 절의 일부는 미 인구조사국과 노동통계청에서 수행한 방법 및 조사에 관한 보고서다. 우리는 재설계를 성공적으로 이끈 두 기관의 모든 개인의 노력을 치하하고자 하며, 특히 여기에 인용된 연구를 수행한 사람들에게 깊이

감사드린다. 여기서 표현된 관점은 저자 개인의 것으로 인구조사국이나 노동통계청의 관점을 반영하는 것은 아님을 밝혀둔다. 아울러, 본 절의 초고에 유익한 도움의 말을 해 준 Cathryn Dippo와 Elizabeth Martin에게 감사의 말을 전하는 싶다.

제 5 장

추정, 분석 및 해석에 대한 오차의 영향

제 5 장 추정, 분석 및 해석에 대한 오차의 영향

제 15 절 측정오차의 조사자료분석에 대한 영향 검토

Paul P. Biemer

Research Triangle Institute, Research Triangle Park

Dennis Trewin

호주통계청(Australian Bureau of Statistics)

15.1 서문

이 절에서는 여러 가지 표준 일변량 및 다변량 분석 기법에 대한 측정 오차의 영향을 검토하기로 한다. 조사 데이터의 측정 오차, 보다 양호한 조사 설계를 통해 측정 오차를 줄이기 위한 방법, 분석 단계에서 오차 영향을 보상하는 방법 및 최종 결과를 해석할 때 측정 오차를 기각하는데서 발생할 수 있는 위험성을 사용자에 알리기 위한 표준은 여러 문헌에서 언급되었다. 이 절에서는 측정 오차에 대해 이러한 모든 항목에 주의를 기울어야 할 필요성에 대해 다룬다.

측정 오차에 관심을 가지는 이유는 무엇인가? 예를 들어 관측치에 독립적인 무작위 오차를 추가할 때 변량은 증가하지만 평균이나 합계의 기대값은 무작위 오차가 0의 기대값을 가지는 한 어느 한쪽으로 치우치지(편향되지) 않는다는 사실은 잘 알려져 있다. 세 가지 유형의 오차가 모두 표준 오차를 증가시킬 수 있지만 표준 오차는 표본 크기를 증가시키면 감소한다. 비용 증대의 가능성 이외에 이들 오차에 관심을 가지는 이유는 무엇인가? 이 절에서는 위와 같은 질문에 대한 답을 제공하며 0의 평균값이 있는 독립적 오차라도 분석자가 사용할 수 있는 결과를 만들어낸다는 사실이 설명된다. 사실 모집단 평균이나 합계 및 그 표준 오차는 이들 오차에 의해 편향되지 않고, 보편적으로 사용되는 유일한 두 가지 통계이다. 또한, 이 절에서는 면접원 또는 기타 조작원의 역할에 의해 발생된 오차가 자료 분석 방법에 미치는 영향에 대해 자세히 설명한다.

15.2절에는 범주형 자료 분석에 보다 적절한 모형은 물론 연속적인 조사 자료의 분석을 위한 모형도 설명된다. 또한 오차가 표본 단위 사이에 연관성을 가지지 않거나 단위 사이의 연관성이 면접원, 코딩요원, 감독원 또는 기타 조사 조작원에 의해 발생할 경우 이들 각각의 모형 형식에 대해 측정 오차의 영향을 고려하기로 한다. 이 절이 이 절에서는 앞으로 설명할 오차 영향의 검토를 위한 단계가 구성된다.

15.3절에는 일변량에 대한 연관성 및 비연관성 오차의 영향이 설명된다. 특히 평균, 비율, 합계 및 모집단 수량과 같은 기술 통계가 설명된다. 15.4절에는 다변량 분석에 대한 다음과 같은 오차 영향이 설명된다. 회귀 분석, 연관성 분석, 분산 분석, 공분산 분석 및 범주형 자료 분석(특히 두 변수 사이의 적합성 검정 및 독립 검정). 이 설명의 목적은 오차 영향의 성격을 파악하고 자료 분석 오차 기각과 관련된 잠재적 위험을 이해하도록 하는 것이다. 여기에서는 자료 분석의 측정 오차 영향 연구를 위해 제안된 모형의 완벽한 검토는 제공하지 않는다. 이들 영향력 보충을 위한 모든 방법이 설명되는 것도 아니다. 이들 방법을 모두 설명하기에는 이 절의 길이가 너무 짧기 때문이다. 그러나 15.5절에는 가장 보편적으로 사용되는 오차 보충 방법과 이 문헌과 이들 문제를 취급하는 다른 문헌에 대한 기사에 대한 참조의 요약이 제공된다.

15.2 측정 오차 영향 연구를 위한 모형

조사 오차 모형은 통계에 대한 측정 오차의 영향과 통계적 추론의 이해에 필수적이다. 오차 모형을 사용하면 고려되는 오차의 성격, 오차를 유발하는 조사 조건, 분산 및 분산 추정량 및 이들 영향의 규모를 간결하고 정밀하게 설명할 수 있다. 설명을 쉽게 할 수 있도록 측정 오차에 두 가지 일반 모형이 사용된다. 하나는 연속 자료이고 다른 하나는 이진 자료이다. 또한 주요 결과의 설명을 쉽게 할 수 있도록 표본 분수 n/N이 무시되는 대체 없이 단순 임의 표본(SRS)을 사용한다.

15.2.1 연속 자료의 오차 모형

이 절에서는 고려되는 연속 오차 모형 중 가장 일반적인 형식에 따라 임의로 선택한 단위 j의 관측치 y_j가 다음과 같은 두 가지 구성요소의 합계라고 가정한다. 즉, 단위에 대한 참값 μ_j 및 면접원, 설문지, 응답자, 면접방식, 면접 설정 등을 포함하는 측정 프로세스에 영향을 주는 오차 dj가 그것이다.

$U = \{1, 2, \cdots, N\}$은 N 단위를 구성하는 대상 모집단의 레이블 종류를 나타낸다. j번째 단위에 관련된 참값 μ_j의 기대값과 분산은 각각 (15.1)과 (15.2)에 설명된다.

$$\overline{M} = \frac{\sum_{j=1}^{N} \mu_j}{N} \qquad (15.1)$$

$$\sigma_{\mu}^{2} = \frac{\sum_{j=1}^{N}(\mu_j - \overline{M})^2}{N} \qquad (15.2)$$

가장 일반적인 형식의 오차 구조 d_j는 면접원, 감독원, 코딩요원 또는 기타 조사 조작원에 의해 발생된 측정 오차 사이에 잠재적 연관성을 제공한다. S = {1, 2, …., n}은 $n = Im$ 단위의 표본을 나타내며 $m = n/I$ 단위의 I할당으로 분할되는 것으로 가정한다.(동일한 할당 크기는 오차 영향을 쉽게 설명하는 목적으로만 사용되며 결과의 일반화를 제한하지 않는다.) S_i는 i번째 조작원에게 할당된 단위의 집합를 표시한다. 이제 dj를 dij로 다시 표시한다. 여기에서 아래 첨자 ij는 j = 1,…, m, I = 1,…, i에 대해 Si의 j번째 단위를 나타낸다. 단위 $j \in S_i$인 경우, d_{ij}는 두 가지 오차 b_i와 ε_j의 합계이다. 여기에서 b_i는 i번째 조작원에게 할당된 모든 단위에 동일한 것으로 간주되는 조작원 오차이고 ε_j는 응답자 및 조작원과 같은 오차 발생원에 의해 발생하는 단위특정 또는 기본 오차(elementary error)이다. 따라서 i_j번째 관측치에 대한 모형은 (15.3)이다.

$$\begin{aligned} y_{ij} &= \mu_{ij} + d_{ij} \\ &= \mu_{ij} + b_i + \varepsilon_{ij} \end{aligned} \qquad (15.3)$$

조작원 오차를 추정할때는, 만일 우리의 주요한 관심사가 현 조사에 고용된 I명의 조작원에 있다면, 오차 b_i는 고정된 상수이거나 조작원 편향이라고 가정할 수 있다. 그러나, 종종 조사를 계획할때는 우리는 현재 조작원 집단을 선택한 잠재적인 조작원들의 모집단에 더 관심을 가지게 된다. 그 이유는 만일 이 조사가 반복될 경우에 향후 조사를 위한 조작원도 이 모집단에서 모집하게 될것이기 때문이다. 이 경우에 무작위 조작원 효과의 가정은 더 적절해 진다. 무작위 조작원 효과이기 때문에 우리는 본질적으로 그 조작원 오차를 b_i라 가정한다. 그리고, 각각 평균 B_b와 분산 σ_b^2인 잠재 조작원 오차를 갖는 잠재 모집단으로 부터 무작위 표본을 구성한다. 즉, 평균 B_ε, 분산 σ_b^2, Cov(μ_{ij}, d_{ij}) = 0 을 가지게 된다. 이때 d_{ij}의 공분산 구조는 (15.4)과 같다.

$$\begin{aligned} Cov(d_{ij}, d_{i'j'}) &= \sigma_b^2 + \sigma_\varepsilon^2 \quad for\ i = i';\ j = j' \\ &= \sigma_b^2 \quad for\ i = i';\ j \neq j' \\ &= 0 \quad for\ i \neq i' \end{aligned} \qquad (15.4)$$

고정 조작원 효과의 가정하에서 공분산 구조는 σ_b^2=0로 수정된다. 그러므로 고정 조작원 효과는 일반적인 무작위 조작원 효과 모형의 특수한 경우이다. 우리가

광범위하게 고려할 일반적인 모형의 또 다른 특수한 경우로 조작원이 없는 경우의 효과가 있다. 즉, 모든 i 에 대해서 $b_i \equiv 0$. 이 모형은 종종 무상관 오차 모형(uncorrelated error model)으로 일컬어지며, $j \neq j'$일때 $\mathrm{Cov}(d_{ij}, d_{ij'})=0$이다. 이 가정하에서 (15.4)의 공분산 구조는 $\sigma_b^2=0$으로 수정된다.

15.2.2 이진수 자료의 오차 모형

이진수 자료의 응답의 경우에(예를 들어, "예/아니오" 즉 0 또는 1로 코딩하는 응답의 경우), 연속형 자료 모형의 가정은 적절하지 않다. 예를 들어, $\mathrm{Cov}(\mu_i, d_{ij}) = 0$ 가정은 일반적으로 적용되지 않는다. 그러므로, 이러한 종류의 자료에 대해서는 대체 모형이 개발되었다. 연속형 자료 모형에 관해서 우리는 모집단이 I개의 조작원 집단으로($S_1, \ldots, S_I$) 분할되었다고 가정한다. y_{ij} 가 $Si(i = 1,\cdots, I)$의 j 번째에 대한 관측치라고 하고, 이때 '예'인 경우의 y_{ij} 값은 1이며 '아니오'의 경우에는 0이다. μ_{ij}는 응답이 예일 때 단위 $(i,j) = 1$, 아니오 일때 0으로 대응하는 참값이라고 하자. 이진수 자료의 경우에, (15.1) 과 (15.2)는 $\pi = \mathrm{P}(\mu_{ij} =1)$, $\pi\ (1 - \pi)$로 각각 주어지며, 이때 π 는 무한 모집단 비율이다. 이때, $j \in S_i$에 대해, 우리는 다음의 오분류 확률을 정의한다(15.5).

$$\begin{aligned} \theta^i &= P(y_{ij} = 0 \mid \mu^{ij} = 1) \\ \phi^i &= P(y_{ij} = 1 \mid \mu^{ij} = 1) \end{aligned} \qquad (15.5)$$

여기서 θ_i와 Φ_i는 거짓 0값의 확률과 거짓 1값의 확률로 각각 간주된다. 표15.1은 방정식 (15.3)에 대한 오차 분산을(d_{ij}) 나타낸 것이다. (15.6)도 가정한다.

$$\begin{bmatrix} \theta^i \\ \phi^i \end{bmatrix} \sim \left(\begin{bmatrix} \theta \\ \phi \end{bmatrix}, \begin{bmatrix} \sigma^{\theta\theta} & \sigma^{\theta\phi} \\ \sigma^{\phi\theta} & \sigma^{\phi\phi} \end{bmatrix} \right) \qquad (15.6)$$

이진 자료 모형은 코크란(Cochran's (1968))의 모형을 일반화한것이다(표 15.1). 이진 자료 오차 모형에 대한 오차 분산은 $\theta_i = \theta_i' = \theta$ 이며 $\Phi_i = \Phi_i' = \Phi$와 같다.(종종 $1 - \theta_i$와 $1 - \Phi_i$는 각각 측정치에 대한 민감도(sensitivity) 및 특이도(specificity)로 불려진다는 점에 주의한다.) 여기서 우리는 조작원에 따른 모형에 오분류 확률(misclassification probabilities)을 허용하고, 따라서 모형에 상관 오차가 포함된다.

표 15.1 이진수 자료 오차모형의 오차분포

y_{ij}	μ_{ij}	d_{ij}	$P(y_{ij} \mid \mu_{ij})$
1	1	0	$1 - \theta_i$
0	1	-1	θ_i
1	0	1	ϕ_i
0	0	0	$1 - \phi_i$

연속형 자료의 경우에, 만일 조작원 효과가 고정된다고 가정하면, 비록 오분류 확률이 조작원 전반에 걸친 다른 확률을 설명하기 위해서 i로 표기었다 하더라도 (15.6)에서 $\theta_i = \theta_{i'} = \theta = 0$ 이다. 조작원 효과가 없다고 가정한다면, 코크란의 모형에 의해 우리는 $\theta^i = \theta^{i'} = \theta$ 이고 $\phi^i = \phi^{i'} = \phi$로 정의할 수 있으며, 이는 $\sigma_{\theta\theta} = \sigma_{\phi\phi'} = \theta_{\theta\phi} = 0$라는 것을 의미한다. 이 모형은 $(i,j) \neq (i'\ j')$에 대해 $\mathrm{Cov}(d_{ij}, d_{i'j'}) = 0$ 이므로 비상관 이진 자료 모형이라고도 한다. 그러나, 표15.1에서 알 수 있듯이 오차는 참값과 음의 상관관계를 가진다는 점에 주의한다.

15.3 일변량 분석에 대한 측정 오차(measurement errors)의 효과

관심 모집단을 설명하는데 도움이 되는 평균, 합계, 비율, 중심값, 또는 기타 다른 통계량을 추정하는 것은 조사 자료 분석의 일반적인 요소이다. 따라서 이러한 기술 통계량에 대한 측정 오차 효과에 대한 연구는 모든 조사 작업에 관련된다. Hansen 등(1951), Sukhatme and Seth(1952), Kish and Lansing(1954)의 논문에 모집단 평균에대한 추정량에 대한 상관오차와 무상관 오차의 효과가 고찰되어 있다. 그후 수많은 저자가 개념을 다듬고 추정량과 분산에 대한 비표본 오차 효과를 실험적으로 조사했다. Hansen 등(1961, 1964)은 이진수 자료에 보다 적절한 모형을 사용하여 표본 비율에 대한 측정오차의 효과를 조사했다. 코크란(Cochran)은 이진수 자료에 대한 측정 오차 모형을 재정의했으며 평균, 비율 및 수많은 표준 통계 기술의 효과에 대해 고찰했다. 이 절에서 우리는 15.2절의 모형과 관련하여 이러한 초기 논문의 결과를 소개한다.

15.3.1 연속형 자료에 대한 평균과 합의 추정량 및 표준 오차

15.3.1.1 무상관 오차 효과(Uncorrelated Error Effects)

$\sigma_b^2 = 0$ 일때 (15.4)절의 분산 공분산 구조를 가지는 모형(15.3), 즉 무상관 모형을 가정한다. SRS의 경우, M에 대한 일반적인 추정량은 (15.7)이다.

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{m} y_{ij} \qquad (15.7)$$

이제, 합계 M은 $\overline{NM}$ 이므로, $\overline{M}$에 대해서 측정 오차효과를 고려하는데 충분하다. 이는 (15.3)로부터 B_d가 표본 평균의 편향(bias)인 (15.8)임을 알수 있다.

$$E(\bar{y}) = \overline{M} + B_d \qquad (15.8)$$

$B_d = B_d + B_\varepsilon$는 조작원들의 편향와 기타 다른 자원들에 의한 편향들의 합계인점에 주의한다. $B_d = 0$인 경우 표본 평균은 모집단 평균의 불편 추정량이다. 이 조건은 조작원 오차와 기본 오차 분포가 모두 0이거나, 조작원과 기본 오차 편향이 서로 상쇄될수 있다는 것을 내포하고 있다. 그러므로, $B_d = 0$일때, 측정 오차가 있더라도 모집단 평균에 대한 표본 평균은 비편향(unbiased)이다.

y 분산은 (15.9)과 (15.10)에서 알수 있다.

$$Var(\bar{y}) = \frac{\sigma_\mu^2 + \sigma_\varepsilon^2}{n} = \frac{1}{R}\frac{\sigma_\mu^2}{n} \qquad (15.9)$$

여기서

$$R = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2} \qquad (15.10)$$

우리는 R을 신뢰도비(reliability ratio)로 간주한다(Sukhatme ans Seth, 1952 참조). 그러므로, 우리는 (15.9)로부터 무상관 오차의 결과로서 y의 분산은 증가하고 유의적인 표본 크기는 $n' = Rn$이라는 것을 알 수 있다. 만일 c가 한단위를 관찰하는데 드는 비용이라고 한다면, $c(n/R - n)/cn = (1 - R)/R$은 추정량 y의 정확도의 고정 수준에 대한 측정 오차로 인해 조사 비용에서 비례하여 증가한다. 효과를 설명하기 위해서,몇 설문 문항에 대해 $R = 0.7$ 이라고 가정해 보자. 그렇다면 $(1 - R)/R = 0.43$이며, 또한, 총 조사 비용은 무상관 측정 오차로 인해서 43%가 증가하게 된다.

우리가 알아본바와 같이 신뢰도비는 측정 오차 연구에서 흔히 볼수 있는 모수이다. 이는 한 관측치의 총 분산에 대한 참값의 분산의 비율이며, 이는 무상관 모델에서 참값의 분산과 오차 분산의 합계이다. R에 대한 보다 일반적인 식은 (15.11)에 나타나있다.

$$R = \frac{VE(y_{ij}|ij)}{V(y_{ij})} \qquad (15.11)$$

이 식의 분자는 측정 오차 분포에 대한 조건부 기대값(i,j)의 무한 모집단 분산이다. 종종 기대값 $E(y_{ij}|ij)$는 단위 (i,j)의 참 점수(true score)로 불린다(Lord and Novick, 1968). 이는 참값 μ_j와 측정 편향 B_d의 합과 동일하다. 그러므로 우리가 가정한 모형하에서 참 점수의 분산은 참값의 분산과 동일 한 σ_{μ}^2이다.

이제 무상관 오차 모형하에서 $\bar{V}(y)$의 추정량을 살펴보자. 보통 추정량은 (15.12)와 같다.

$$v(\bar{y}) = \frac{s^2}{n} = \frac{\sum_{i,j}(y_{ij} - \bar{y})^2}{n(n-1)} \qquad (15.12)$$

이것은 (15.13)과 같이 나타낼 수 있으며(예로, Cochran, 1968를 참고한다),

$$E[v(\bar{y})] = \frac{\sigma_{\mu}^2 + \sigma_{\varepsilon}^2}{n} = Var(\bar{y}) \qquad (15.13)$$

그러므로 보통 추정량은 무상관 측정 오차하에서는 비편향이다. 그러므로, 비록 표본 평균의 분산이 증가하더라도 분산의 보통 추정량도 동일하게 증가하게 된다. 우리가 앞으로 볼수 있듯이, 추정량, $\bar{V}(y)$는 $B_d \neq 0$일때 조차도 무상관 오차가 있고 비편향인, 우리가 연구하는 유일한 추정량이다.

15.3.1.2 상관오차 효과

이제 $\sigma_b^2>0$일때 (15.4)에 의해서 주어진 분산-공분산 구조와 함께 모형(15.3)을 살펴보자. 이 모형하에서는 우리가 조작원 오차 분포의 평균이 B_b라고 가정하고 있으므로 y의 기대값은 여전히 (15.8)로 주어진다. 그러나, 추정량의 분산은 변한다. Sukhatme and Seth (1952)는 처음으로 다음을 제시하였다.

$$Var(\bar{y}) = \frac{\sigma_\mu^2 + \sigma_\varepsilon^2}{n} + \frac{\sigma_b^2}{I} \qquad (15.14)$$

여기서, 가변 조작원 오차로 인해 오른쪽의 마지막 항이 증가값이 되는 것을 보여주었다. 그러므로 분산에 대한 조작원의 기여도는 조작원 수 I가 증가할수록 감소하며, 반면에 기본 오차 기여도는 표본의 수가 증가할 수록 감소한다. 대안으로 (15.14)는 다음과 같이 쓸 수 있다.

$$\begin{aligned} Var(\bar{y}) &= \frac{\sigma_\mu^2 + \sigma_b^2 + \sigma_\varepsilon^2}{n}\left[1 + (m-1)p_y\right] \\ &= \frac{\sigma_\mu^2}{n}\frac{1}{R}\left[1 + (m-1)p_y\right] \end{aligned} \qquad (15.15)$$

여기서

$$p_y = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_b^2 + \sigma_\varepsilon^2} \qquad (15.16)$$

그리고, (15.11)로 부터

$$R = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_b^2 + \sigma_\varepsilon^2} \qquad (15.17)$$

조작원이 면접원일때 항 ρ_y는 동일한 면접원의 할당량의 두 관측치의 오차간의 상관이므로 면접원간의 상관계수(intrainterviewer correlation coefficient)라고 한다. $[1+(m-1)\rho_y]$항은 층화 표집에서 표집 설계 효과와 대응되는 항과 유사한 면접원의 설계 효과(design effect) 내지는 *deff*로 간주된다. 조작원 *deff*에 대한 면접원은 조작원 분산의 결과로서 분산에 대한 증가량을 반영한다. 예를 들어, 만일 평균 면접원의 할당량이 50단위, ρ_y=0.01로 구성되었다면 이때, $[1+(m-1)\rho_y]$ = 1+49(0.01)]=1.49이며 표본 평균 분산은 상관 면접원 오차로 인해 49%가 증가한

다. 게다가, R = 0.7 인 신뢰도와 합쳐져서 측정 오차로 인해 분산의 증가량은 1.49/0.7 = 2.13 이다. 그러므로 이 예제에서 측정 오차 분산의 효과는 분산의 두 배 이상이된다. 이러한 경우는 조사연구에 있어서 특이하지 않다.(Groves, 1989, pp.365-368 참조). 그림 15.1에서 우리는 $0.0 \leq \rho_y \leq 0.1$에 대한 상관 면접원 오차와 m의 대체값으로 인한 분산의 증가량($(m-1)\rho_y$)을 점도표로 나타내었다. 면접원의 할당량이 증가할 수록 면접원 분산의 효과 또한 상당히 증가한다는 점에 주의한다. 이 그림은 작은 면접원 효과와 적절한 할당량에도 면접원 분산은 표본 분산의 몇배가 커질 수 있다는 것을 설명하고 있다.

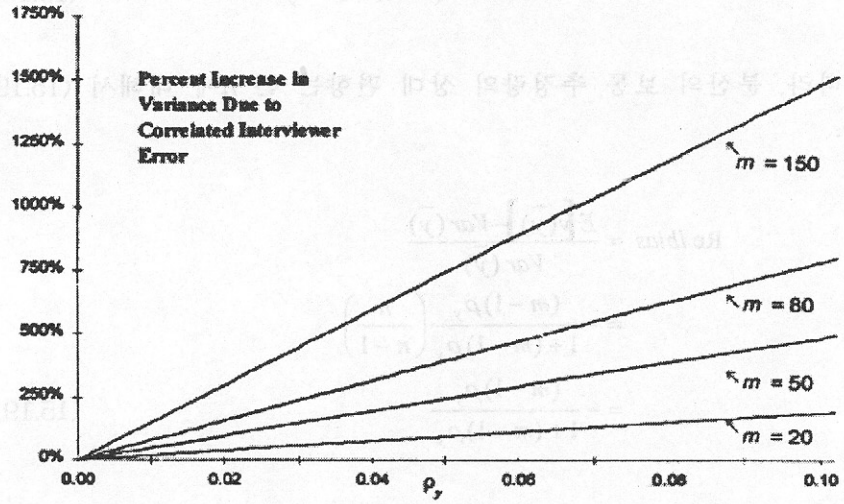


그림 15.1 면접원간 상관 ρ_y와 할당량 크기 m의 함수로 상관 면접원 오차로 인한 분산의 증가량

면접원과 다른 조작원 효과에 대한 연구는 조사 방법을 개발하는데 매우 중요한 역할을 해왔다. 예를 들어 미 인구조사국에서 실시한 1950년 인구조사에서는 면접원이 조사에 끼치는 효과에 대한 당국의 우려가 일련의 인구조사에서 (Hanson and Marks, 1958) 자기 조사 인구조사 설계를 채택하도록 하는데 주요 원인이 되었다. 그 이유는 자기 조사(self-enumeration)에서는 측정 오차가 가구간에 무상관(uncorrelated)이기 때문이었다. ρ_y를 감소시키는 것은 또한 표준화된 면접법(Fowler and Mangione, 1990)과 코딩 방법을 위한 주요한 동기부여(Biemer and Caspar, 1994)이다.

Groves (1989, p. 365)는 10회의 개별면접조사에서 추정된 면접원에 대한 ρ_y 값을 보고하고, 여기서 측정된 수많은 조사 특성치에 대한 평균 ρ_y값이 0.005에서 0.102로 나타났음을 보여주었다. 문헌에서 반견된 대부분의 면접원 분산에 대한 연구에 의하면 ρ_y은 이 범위내에 포함되어 있다. 예를 들어, 집중화된 전화 면접

의 경우 ρ_y의 대표적인 값은 0.01이다. 개별방문면접 조사의 경우에는 보통 조금 큰 ρ_y값이 생성된다.

이제 (15.12)에서 $V(\bar{y})$의 기대값에 대해 고찰해 보자. Hansen 등 (1961)은 다음을 증명하였다.

$$E\left[v(\bar{y})\right] = \frac{\sigma_\mu^2 + \sigma_b^2 + \sigma_\varepsilon^2}{n}\left(1 - \frac{m-1}{n-1}\rho_y\right) \tag{15.18}$$

이에 따라, 분산의 보통 추정량의 상대 편향는 큰 n에 대해서 (15.19)와 같이 나타난다.

$$\begin{aligned} \text{Re}\,lbias &= \frac{E\left[v(\bar{y})\right] - Var(\bar{y})}{Var(\bar{y})} \\ &= -\frac{(m-1)\rho_y}{1+(m-1)\rho_y}\left(\frac{n}{n-1}\right) \\ &= -\frac{(m-1)\rho_y}{1+(m-1)\rho_y} \end{aligned} \tag{15.19}$$

이 식으로부터 우리는 편향는 ρ_y 단조 감소 함수임을 알 수 있으며, 뿐만 아니라, 과소 추정량의 크기는 대략 상관 면접원 오차로 인한 분사의 증가량과 동일하다는 것을 알 수 있다. 그러므로, 보통 분산 추정량은 상관 오차로 인한 분산의 증가량을 반영하지 않는다. 대표적인 ρ_y값이 0.01이라고 하면, 상대편향은 m=20인 경우 -16%에서 m=100인 경우에 -50%에 이른다. (15.19)로부터 편향의 최대 절대값은 약$(m-1)/m \times 100\%$이라는 점에 주의한다.

이 결과는 면접원 분산이 모집단 평균에 대한 추정량의 분산과 편향를 상당히 증가시킬 수 있다는 것을 시사한다. 사실 면접원 분산의 크기가 표본 분산의 크기를 초과하는 것은 그리 특별한 것은 아니다. 게다가 단순 임의 표본 추출과 면접원 오차가 존재하는한 평균과 총합의 표본 분산을 추정하기 위한 일반적인 서적에 기술된 공식은 음으로 편향되었으며 그 편향는 분산의 증가량 즉. $(m-1)\rho_y$과 거의 동일하다. 월터(Wolter(1985, pp.381-392))는 복합 표본 추출법(complex sampling)을 사용해서 유사한 결과를 이끌어냈으며 임의 집단과 다른 표준 오차를 추정하기 위한 유사 반복 방법은 $Var(\bar{y})$ 추정량의 편향를 상당 부분 제거할 수 있다는 것을 증명하였다.

15.3.2 사분위수 추정량

사분위수는 사회 과학과 사회 정책의 개발시 종종 사용된다. 예를 들어, 인구 소득 분포의 사분위수는 종종 정부의 부양 정책에서 빈곤을 나타내는 경계로 사용된다. 게다가, 표본 분포의 90%, 95%, 99% 사분위수는 흔히 가설 검정을 위한 임계값와 신뢰 구간의 오차 한계를 결정하기 위해서 사용된다. 사분위수 추정에 대한 무상관 측정 오차의 효과는 모집단이 정규 분포를 따른 다는 가정하에 풀러 (Fuller (1995))에 의해서 고찰되었다. 풀러의 표기법을 사용하면 확률 변수 X에 대한 사분위 함수는 (15.20) 및 (15.21)과 같다.

$$Qx(a) = F_x^{-1}(a) \qquad (15.20)$$

여기서,

$$F_x(x) = P(X \le x) = a. \qquad (15.21)$$

그러므로 $Q_x(a)$는 변수에 대한 누적 분포 함수의 역함수이다. 사분위수에 대한 몇가지 예는 다음과 같다. 중앙값$Q_x(0.5)$, 제1사분위수 $Q_x(0.25)$, 사분위 범위 $Q_x(0.75)-Q_x(0.25)$.

누적 분포 함수에 대한 측정 오차의 효과를 설명하기 위해서 참값이 $N(0,1)$인 확률변수로서 모집단에 분포되어있으며 관측치 y_{ij}는 $N(0,1)$을 따르는 d_{ij}를 가지는 무상관 오차 모형을 따른다고 가정하자. 즉, $B_d = 0$ 이며 $\sigma_D^2 = 1$ 이라 가정하자. 그러므로, y_{ij}는 $N(0,2)$을 따르는 확률 변수이다. 그림 15.2에서 이 부분이 설명되어 있으며 이는 $N(0, 2)$ 분포(y_{ij}의 분포) 곡선상에 $N(0,1)$ 분포곡선이 중첩되어 있다는 것을 보여준다.(μ_{ij}의 분포) 또한 각 분포의 제 95백분위수도 나타난다 y-분포의 제 95백분위수는 μ분포의 백분위수보다 상당히 크다.

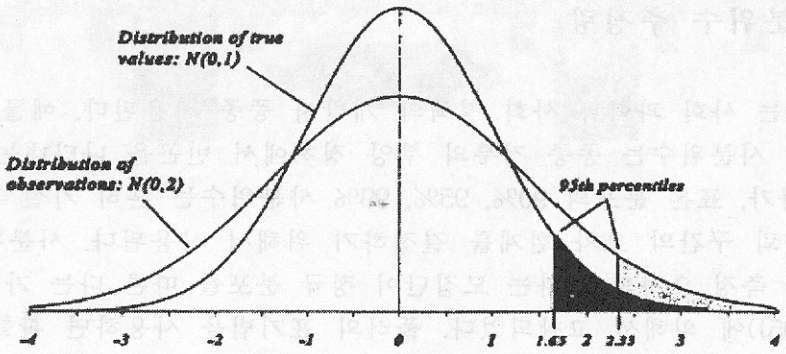


그림 15.2 정규 분포의 제 95백분위수에 대한 측정 오차 효과

이제 $Y=\mu+d$ 라고 가정한다. Y와 μ이 정규분포를 따른다면 $P(Y\leq y_0)=a=P(\mu\leq\mu_0)$는 다음을 의미한다.

$$\frac{y_0-\bar{Y}}{\sigma^y}=\frac{\mu^0-\bar{M}}{\sigma_\mu} \tag{15.22}$$

여기서, y_0와 μ_0은 y와 μ의 분포에서 각각 제 a 사분위수이다. 그러므로, 참값의 중심 분포 $\mu-\bar{M}$에서 보면, 관측치의 중심 분포 사분위수 $y-\bar{M}$은 다음과 같이 주어진다.

$$Q_{y-\bar{Y}}(a)=R^{-1/2}Q_{\mu-\bar{M}}(a) \tag{15.23}$$

이때 $Q_{y-\bar{M}}(a)$와 $Q_{\mu-\bar{M}}(a)$는 각각 $y-\bar{M}$과 $\mu-\bar{M}$에 대한 사분위 함수이며 R은 15.2절에서 정의된 신뢰도비이다. 그림 15.2로 설명하면 $N(0, 1)$ 분포에 대해서 $Q_{\mu-\cup}(0.95)=1.65$이다. $\sigma_d^2=1$, R=0.5이므로 그림에서 보이는 바와 같이 관측치 y의 분포에 대한 $Q_{\mu-\cup}(0.95)=\sqrt{0.5)^{-1}}\times 1.65=2.33$이다. 뿐만 아니라,

$$\begin{aligned} Q_y(a)&=Q_{y-\bar{Y}}(a)+\bar{Y}\\ &=R^{-1/2}Q_{\mu-\bar{M}}(a)+\bar{T}\\ &=R^{-1/2}\left[Q_\mu(a)-\bar{M}\right]+\bar{Y} \end{aligned} \tag{15.24}$$

여기서 B_d는 15.2절에서 정의된 측정 편향이다. 그러므로, 만일 $\hat{Q}_y(a)$가 측정 오차가 없을 때 $\hat{Q}_y(a)$의 비편향 추정치라면, 측정 오차 모형에서 다음과 같다.

$$E[Q_y(a)] - Q_\mu(a) = (R^{-1/2} - 1)[Q_\mu(a) - \overline{M}] + B_d \qquad (15.25)$$

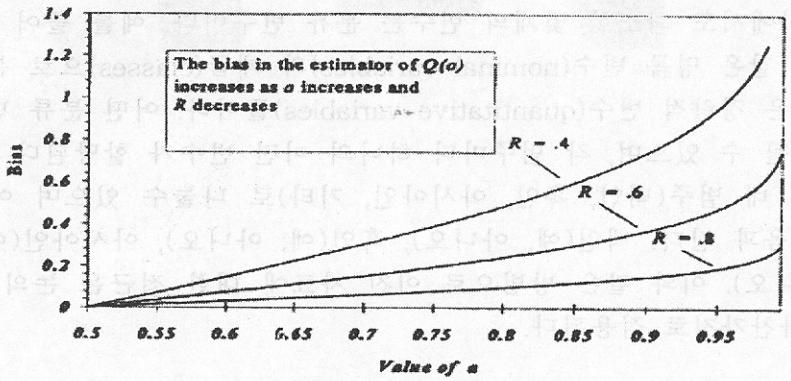


그림 15.3 y가 신뢰도 R로 측정되었을때 $N(0,1)$ 분포의 평균을 추정하기 위한 $\hat{Q}_y(a)$에서의 편향

그림 15.3은 μ_i가 $N(0,1)$ 분포를 따르고 y_i가 신뢰도 R=0.4, 0.6, 0.8일때 관측된 함수를 나타낸다. $a \geq 0.5$때, 편향는 a가 증가하고 R이 감소할때 증가한다한다는 점에 주의한다. 무상관 오차 모형하에서 정규 분포의 사분위수에 대한 결과는 (15.17)에 주어진 신뢰도 정의와 함께 상관 오차 모형에 대해 여전히 적용된다는 것은 쉽게 증명된다.

표본 중앙값(a=0.5)의 편향와 표본 평균의 편향을 비교해 보면 (15.25)로 부터 모형 (바꿔말하면, B_d)하에서 정규 분포의 평균과 중앙값이 동일하기 때문에 서로 동일하다는 점에 주의한다. 우리는 이제 정규 분포의 평균을 추정하기 위해서 표본 평균에 관련하여 표본 중앙값의 효율성에 대해서 고찰해본다. 수리 통계학에 잘 알려진 결과로부터(Wilks, 1962, pp. 364 참고), $\hat{Q}_y(05)$의 점근적 분포 분산은 큰 n에 대해서 $N(\overline{Y}, \pi\sigma_y^2/2n)$이다. 어떤 n에 대해서 y의 분포는 $N(\overline{Y}, \pi\sigma_y^2/n)$이다. 그러므로 $\hat{Q}_y(05)$의 y에 대한 점근적 상대 효율은 (15.26)과 같다.

$$\frac{\sigma_y^2}{(\pi/2)\sigma_y^2} = 0.637. \qquad (15.26)$$

예를 들면, 크기 1,000인 표본에 대해서 표본 평균은 표본 크기가 1,000/0.637 또는 1,570인 표본 중앙값 만큼 효율적이다. 이는 편향의 측정 오차 분산과 편향의 크기와 무관하다.

15.3.3 이진수 자료의 비율과 합계의 추정량 및 표준오차

어떤 조사에서도 최소한 몇개의 변수는 분류 변수이다. 예를 들어 인종, 성별, 결혼 여부와 같은 명목 변수(nominal variables)와 계급(classes)으로 분류되는 수입과 나이같은 정량적 변수(quantitative variables)들이다. 어떤 분류 변수도 이진 변수로 변환될 수 있으며, 각 범주마다 하나의 이진 변수가 할당된다. 예를 들어, 인종 변수는 네 범주(백인, 흑인, 아시아인, 기타)로 나눌수 있으며 이진 변수로 나타내면 다음과 같다. 백인(예, 아니오), 흑인(예, 아니오), 아시아인(예, 아니오), 기타(예, 아니오). 이와 같은 방법으로 이진 자료에 대한 접근은 논의된 다항 변수에 대해 마찬가지로 적용된다.

15.3.3.1 무상관 오차 효과

15.2.2절에 주어진 이진 자료에 대한 무상관 오차 효과를 가정한다. SRS에 대해서 $\overline{M} = \pi$에 대한 보통 추정량은 이진수 자료에 대해서 디

$$p = \frac{\sum_{i=1}^{I}\sum_{j=1}^{m} y_{ij}}{n}. \qquad (15.27)$$

코크란(Cochran (1968))은 다음을 증명하였으며

$$\begin{aligned} E(p) &= \pi(1-\theta) + (1-\pi)\phi \\ &= \pi^{*} \end{aligned} \qquad (15.28)$$

표본 비율에서 편향은 다음과 같다.

$$\pi^{*} - \pi = -\pi\theta + (1-\pi)\phi. \qquad (15.29)$$

그러므로, Φ>0인 θ에 대해, $\pi\theta=(1-\pi)\ \Phi$이면 π의 추정량에 대한 편향은 0이다. $\pi\theta=(1-\pi)\ \Phi$가 되려면 π의 크기에 상관없이 거짓인 양의 값의 개수가 거짓인 음의 값의 개수와 정확하게 일치해야 한다. π가 작으면, π의 편향에 대해서 상대적으로 작은 값인 Φ의 영향력도 매우 심각할 수 있다는 점에 주의한다. 예를 들어, π=0.01, θ=0.05, Φ=0.05인 경우에 상대편향는 (0.059 - 0.01)/(0.01)=5.9이며, 즉 590%가 된다. 이때 상대편향에 대한 $E(p)$는 (0.01)(0.95)+(0.99)(0.05)=0.059이다.

코크란이 관찰한 바와 같이 만일 μ_{ij}의 분포가 모수 n과 π를 가지는 이항분포를 따른다면 y_{ij}의 분포는 모수 n과 π^*를 가지는 이항 분포를 따르게 된다. 이항 확률변수 분산에 대한 일반적인 공식을 사용하여 우리는 y_{ij}의 분산이 (15.30)이라는 것을 알 수 있다. 그러므로 y_{ij}에대한 분산은 π가 0.5에 가까울때 보다 π^*가 0.5에 가까울수록 μ_{ij}에 대한 분산보다 크다.

$$Var(y_{ij}) = \pi^*(1-\pi^*) \tag{15.30}$$

$p = \bar{y}$ 인Var(p)의 보통 추정량은 다음과 같으며

$$v(p) = \frac{pq}{n-1} \tag{15.31}$$

이는 다음의 기대값을 가진다.

$$E[Var(p)] = \frac{\pi^*(1-\pi^*)}{n}. \tag{15.32}$$

그러므로 연속형 변수의 경우에 있어서는 n에 대해서 Var(p)는 (15.30)와 같기 때문에 분산의 보통 추정량은 무상관 오차 모델 하에서 비편향이다.

15.3.3.2 상관오차 효과

이제 15.2.2절에서 설명한 이진수 자료에 대한 상관 오차 모형을 가정한다. 미국 통계청(1985)에서는 이 모형하에서 p의 기대값은 (15.78)의해서 주어진다는 것을 증명하였다. 계속해서 조작원 할당량은 상호관입(interpenetrated)되었다고 가정해 보자. 즉, 단위들은 조작원에게 무작위 할당되므로 $E(\sum_j \mu_{ij}/m|i)=\pi$ 이다. 여기서 $E(\cdot|i)$는 i번째 조작원 할당량이 주어졌을때의 조건부 기대값이다. Biemer and Stokes(1991)는 연속형 자료에 대해서 형태상으로 (15.15)와 유사한 이러한 조건하에서 (15.33)을 증명하였다.

$$Var(p) + \frac{\pi^*(1-\pi^*)}{n}[1+(m-1)\rho_y \tag{15.33}$$

단, 지금은 다음과 같다.

$$p_y = \frac{\pi^2\sigma_{\theta\theta} + (1-\pi)^2\sigma_{\phi\phi}^2 - 2\pi(1-\pi)\sigma_{\theta\phi}}{\pi^*(1-\pi^*)}. \qquad (15.34)$$

그러므로 $Var(p)$은 상관 조작원 오차의 결과로 요인 $[1+(m-1)\rho_y]$에 의해서 증가된다. 분산의 증가량은 $(m-1)\rho_y$이며 이는 그림 15.1에 있는 함수와 동일한 것이다. 그러므로 연속형 자료에서 평균의 분산에 대한 면접원 오차효과에 대한 논의는 이항 반응 변수에도 적용된다. 만일 오분류 분산 성분이 모두 0이라면 $\rho_y=0$이며 면접원 *deff*값은 1인점에 주의한다. 이는 무상관 조작원 오차에 대한 조건인 모든 i에 대해서 $\theta_i=\theta$와 $\Phi_i=\Phi$인 것과 동일하다. 모든 i에 $(1-\pi)\Phi_i=\pi\theta_i$라면, 즉, 만일 모든 조작원에 대해서 거짓인 양의값의 개수가 거짓인 음의 값의 개수와 동일하다면, 면접원 *deff*가 1인것도 증명될 수 있을 것이다. 이 조건이 유지된다면 각 조작원의 할당량에 대한 양의 값에 대해서 기대되는 수는 π이며 그러므로 각 조작원의 할당량의 평균은 π에 대해서 비편향이라는 점에 주목한다. 이제 상관 오차 모형하에서 $Var(p)$의 추정량에 대해서 고찰해 보자.

$Var(p)$의 보통 추정량은 다음과 같다.

$$v(p) = \frac{pq}{n-1}. \qquad (15.35)$$

Biemer and Stokes(1991)는 다음을 증명하였다.

$$E[Var(p)] = \frac{\pi^*(1-\pi^*)}{n}\left[1 - \frac{m-1}{n-1}p_y\right]. \qquad (15.36)$$

그러므로, 분산의 보통 추정량에 있어서 상대 편향는 다음과 같으며,

$$\text{Re}\,lbias = \frac{(m-1)p_y}{1+(m-1)p_y}\left(\frac{n}{n-1}\right)$$
$$\approx -\frac{(m-1)p_y}{1+(m-1)}p_y \qquad (15.37)$$

이는 이 ρ_y가 (15.34)에 의해 주어지는 것을 제외하고는 (15.19)와 동일하다. 이와같이 relbias 즉 ρ_y값과 m값간의 관계는 $\bar{y}$에 대해서 논의되 바와 같다.

15.4 다변량 분석시의 측정 오차의 영향

이절에서 우리는 회귀 계수와 상관 계수 및 분산과 공분산 분석에 대한 측정 오차 효과를 고찰해 본다. 여기서 우리의 목표는 자료 분석에대한 측정 오차 효과의 성질을 파악하는 것이다. 그러므로 우리는 강한 가정 보다는 실제적 중요성을 가지는 단순 모형에 집중한다. 우리는 오차 효과가 보다 현실적이고 복잡한 모형의 가정에서는 어떻게 변경될지에 대해 알아본다.

15.4.1 회귀분석과 상관분석

Madansky(1959), Cochran(1968)과 최근에 들어서는 fuller(1991)를 포함하여 수많은 저자들이 회귀 계수의 추정에 대한 측정 오차의 효과에 대서 고찰하여왔다. 고려되는 가장 단순한 경우는 설명 변수 μ에 대한 종속변수 z의 회귀이다. 여기서 z는 오차 없이 측정되었으며 y는 연속형 자료 무상관 측정 오차 모델을 따르는 μ의 측정치라고 가정한다. z가 또한 오차로 측정되며 측정 오차가 상관 오차인 경우를 이어서 고려해보자.

우리는 모형 (15.38)에서 β_0와 β를 추정하기를 원한다고 생각해보자.

$$Z_{ij} = \beta^0 + \beta\mu^{ij} + \zeta^{ij} \tag{15.38}$$

여기서 β_0와 β는 각각 절편(intercept)와 기울기 계수이다. 그리고 ξ^{ij}는 독립적이며 평균은 0 분산은 σ_ξ^2인 정규 분포 모형를따른다. μ_{ij}, 즉 y_{ij}에 대한 관측치는 15.2.1의 무상관 오차 모형을 따르고 측정 오차 d_{ij}는 정규 분포를 따르며 ξ^{ij}에 독립적이라 가정하자. 보통 추정량 β는 다음과 같다.

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{15.39}$$

$\hat{\beta}$의 기대치가 다음과 같다는 것은 잘 알려져 있으

$$E(\hat{\beta}) = R\beta \tag{15.40}$$

여기서, R은 (15.10)에서 주어진 신뢰도비이다. 그러므로 무상관 측정 오차 모형에서 β의 추정량은 0쪽으로 기울어져 있거나 치우쳐져 있다. 뿐만 아니라, 보통 통계 검정(H_0: β=0)과 관련된 검정력(power)은 이 편향의 결과로서 감소된다.

이제 β_0에 대한 보통 추정량 즉, 다음을 고려해 보자.

$$\hat{\beta}_0 = \bar{z} - \hat{\beta}\bar{y} \qquad (15.41)$$

(15.38)과 (15.40)으로 부터 다음을 증명할 수 있다.

$$E(\hat{\beta}_0) = \beta_0 + \beta\overline{M}\left[-R - B_d / \overline{M}\right]. \qquad (15.42)$$

(15.42)의 오른쪽에 있는 두번째항의 부호가 음또는 양일수 있기 때문에 절편의 편향도 음또는 양일수 있다. 이들의 효과를 설명하기 위해서는 그림15.4의 회귀를 고려한다. 진한 타원은 (μ_{ij}, z_{ij}) 점들의 모임을 나타내고 이 점을 통과한 선은 μ_{ij}에 대한 z_{ij}의 참 회귀임을 나타낸다. 밝은 타원은 (μ_{ij}, z_{ij}) 점들의 모임을 나타낸다. 설명 변수가 오차로 측정될때 참 점에 대한 효과는 가로축을 따라 증가된 변동분에 대응하여 오른쪽과 왼쪽으로 퍼지게 된다는 점에 주의한다. 밝은 타원의 중심을 통과하는 추정된 회귀선은 어두운 타원의 중심을 통과하여 그려지게 되는 참 회귀선보다 작은 기울기를 갖는다. (15.40)에서 수학적인 결과를 그래픽으로 확인할 수 있다. (15.40)에서의 결과는 이 그림에서 볼수 있다. 15.4의 그림에서 β와 $\overline{M}$ 모두 양이며 $B_d = \bar{y} - \overline{M} = 0$ 이다. 그러므로 (15.42)에서 괄호가 쳐진 항은 양이며 β_0의 추정량에 양의 편의와 관찰된 편차에 증가를 초래한다. 마지막으로 그림의 B_d를 주목한다.

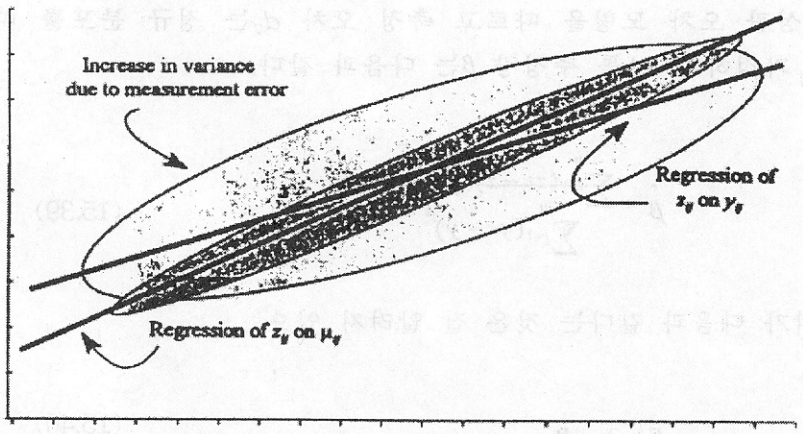


그림 15.4 μ_{ij}에 대한 z_{ij}의 회귀선에 대한 무상관 오차의 효과

설명 변수가 가로축을따라 양으로 이동할때 밝은 타원이 오른쪽으로 이동할 것이다. 관찰 회귀선이 오른쪽으로 이동하게 되면 절편은 점점 작아진다. 마찬가지로 Bd가 점진적으로 음이 될때 그림에서 회귀는 추정 절편의 기대값을 증가시키면서 왼쪽으로 움직일 것이다. 이것은 (15.42)에서도 볼 수 있다. 큰 음의 상대 편향, $B_d / \overline{M}$ 은 양의 괄호 항과 편향을 초래하는 반면에 $B_d / \overline{M}$ 은 (15.42)에서 음의 괄호 항을 초래하고 음의 편의값을 초래한다.

종속변수 z_{ij}가 오차로 측정될 수 있을때 $z_{ij} = \eta^{ij} + \delta^{ij}$ 이며 여기서 η^{ij} 는 참값 성분이고 δ^{ij} 는 측정 오차이다. 이는 (15.3)과 유사하다. 이때 $\xi^{ij} = \xi^{ij} - \delta^{ij}$ 라고 하면 우리는 $\eta_{ij} = \beta_0 + \beta\mu_{ij} + \xi_{ij}$ 로 (15.38)을 다시 쓸 수 있다. 이제, η_{ij}가 z_{ij}를 대체할 수 있고 ξ^{ij} 가 ξ^{ij} 를 대체할 수 있을때 앞선 결과는 동일하게 적용된다. 모형 분산은 $\mathrm{Cov}_{(\xi^{ij}, \delta^{ij})} \geq 0$ 일때 증가하며 다른 경우에는 감소한다는 점에 주의한다.

Chai(1971)는 측정 오차가 변량내에서 및 변량간 상관관계에 있는 경우에 대해서 고찰했다. Chai(1971)는 상관 오차가 있을때 기울기의 계수는 오차간의 공분산의 크기와 부호에 따라서 상당히 과소 추정되거나 과대 추정될 수 있다는 사실을 증명했다. 그러므로 회귀분석에서 기울기 계수는 항상 약하다는 것을 나타내는 무상관 오차 모형의 결과는 비와 비제로 모두에 대해서 측정 오차의 실제 효과를 과대 단순화 한것이다.

z가 오차 없이 관측되었을때 z와 y간의 상관관계 추정에 대해 무상관 측정 오차의 효과를 고찰한다. (15.43)과 (15.44)에 주의한다.

$$\begin{aligned} [Corr(z,y)]^2 &= \frac{[Cov(z,y)]^2}{Var(z)Var(y)} \\ &= \frac{[Cov(z,\mu)]^2}{Var(z)Var(\mu)} \frac{Var(\mu)}{Var(y)} \\ &= [Corr(z,y)]^2 R \end{aligned} \qquad (15.43)$$

그리고, 이에 따라

$$Corr(z,y) = \sqrt{R}\, Corr(z,\mu). \qquad (15.44)$$

그러므로, 무상관 오차 모형하에 상관계수는 신뢰도비의 제곱에 의해 약해진다(attenuated). z와 y 모두가 오차에 의존할때, 우리는 다음을 알게된다.

$$Corr(z,y) = \sqrt{R_z R_y}\, Corr(\eta,\mu) \qquad (15.45)$$

여기서, R_y는 y에 대한 신뢰도 비이며 R_z는 z에 대한 신뢰도 비이며 η와 μ은 각각 z와 y와 연관된 참값이다.

15.4.2 분산 분석과 공분산 분석

코크란, Cochran(1968)은 측정 오차의 결과를 분산분석(analysis of variance (ANOVA))에서 고찰하였다. 이 절에서 우리는 두 독립 표본의 비교를 고찰한다. 그러나 효과는 일반 ANOVA의 경우에도 동일하게 적용된다. 집단A와 집단B가 있을때 이 두 집단의 평균을 비교하기 위해서 15.3.1절에 있는 결과를 평균과 표준 오차를 추정하기 위해서 적용할 수 있다. $\overline{y_A}$와 $\overline{y_B}$를 각각 집단A와 집단B로의 단순 무작위 표본의 평균이라고 하자. 15.3.1로부터 우리는 귀무가설 $H_0 : M^A = M^B$ 하에서 $\bar{y}_A - \bar{y}_B$의 차이의 기대값이 0이 아니고 $B_{dA} - B_{dB}$와 동일하다는 것을 알 수 있다. 여기서는 두 집단과 연관된 측정 편향이다. 그러므로 귀무가설 H_0은 명목 유의 수준 보다 참일때 기각된다. 게다가 우리는(15.15)로 부터 평균의 표준 오차가 무상관 및 상관 분산 성분에의해 팽창된다는 것을 알 수 있다. 그러므로 만일 측정 오차가 집단간 독립적이라면 차이의 분산은 분산의 합이며 값또한 커진다. 이 효과는 귀무가설 H_0 검정력을 감소시키기는 경향이 있다.

두 집단에 대한 분산과 측정 오차 편향에 대한 지식 없이는 ANOVA에 대한 오차는 대개 예측하기 어렵다. 공분산 분석(Analysis of covariance (ANOCOV)) 은 종종 ANOVA에서 오차 분산을 감소시키기 위한 조사 자료의 분석에 사용되거나 집단간 통제되지 않는 변수로 인해서 발생하는 차이를 제거하여 몇 분석 변수와 비교될 수 있도록 하기 위한 조사 자료의 분석에 사용된다. ANOCOV의 패턴 사용은 독립변수가 실험적으로 통제될 수 없으며 조사 응답자들에게 무작위하게 할당되지도 않는 관찰 연구에서 특히 중요하다. 여기서 우리는 통제변수에서 측정 오차의 중요성에 대해서 고찰해본다.

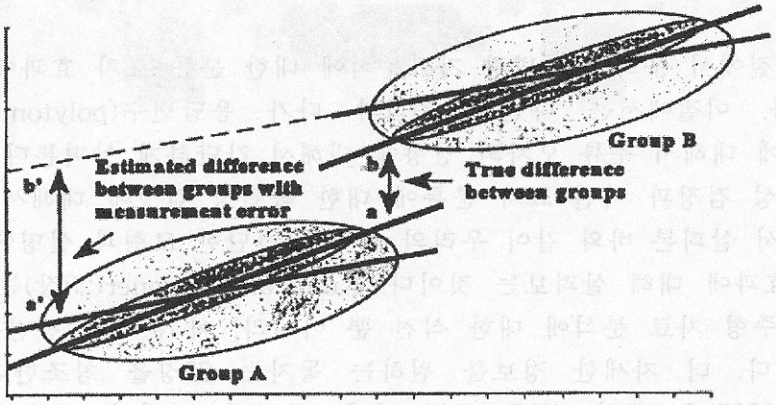


그림 15.5 두 집단의 공분산 분석에 대한 측정오차의 효과

오차 없이 측정된 종속변수 z_{ij}에 대한 ANOCOV에서의 측정 오차 효과는 그림 15.5에서 설명될 수 있다. 두 쌍의 타원은 집단 A와 B 점들의 모임을 나타낸다. 각 집단에 대하여 어둡고 밝은 타원을 해석하는 것은 그림 15.4에 설명된것과 동일하다. 즉 통제 변수(가로축에 위에 있는)의 오차는 일련의 실제의 점들에 대해서 관찰된 점을 좌우로 퍼지게하여 회귀선의 경사를 약하게 한다. 각 집단이 다른 측정 오차 분포를 따른다 하더라도 회귀선의 경사는 각 집단에 대해 동일하다고 가정된다.

ANOCOV에서 회귀선 간의 거리는 통제 변수를 설명한 후의 z_{ij}에 대한 집단 평균간의 차이를 의미한다. 측정 오차가 없을때, 어두운 타원을 지나가는 선은 각 그룹의 종속변수와 통제변수간의 관계를 정의한다. 그러므로 두 집단간의 실제 차이는 거리(a-b)이다. 측정 오차가 통제변수에 추가되었을때 통제변수에 대해 추정된 기울기의 계수는 두 밝은 타원을 통과하는 회귀선에의해 설명된것 처럼 약해진다. 게다가 선 a'-b'에 의해 그림 15.5에서 설명되는 두 집단간의 추정차는 편향되어 있다. 이 경우 통제변수를 설명한 후에 집단간 관찰 차이가 실제 차이보다 크다. 그러나, 일반적으로 비교시의 편향은 측정 오차 분산의 크기에 따라서 양이거나 음이 될 수도 있으며 두 집단 모두의 측정 편향의 크기와 방향에 따라서 양이거나 음이 될 수도 있다. ANOCOV에 대한 측정 효과가 Lord(1960)에 의해 도식화되어 설명되었다. 이 주제에 대해서 보다 깊은 고찰을 하고자 하는 독자는 그 논문을 참고로한다.

15.4.3 범주형 자료 분석

15.3.3.1절에서 우리는 일변량 자료분석에 대한 분류 오차 효과에 대해서 고찰해 보았다. 이절에서는 다변량 분석과 다가 응답변수(polytomous response variables)에 대해서 분류 오차의 영향에 대해서 간단하게 살펴본다. 특히, 우리는 표준 적합성 검정과 이원 교차 분류에 대한 독립성 검정에 대해서 고찰해 본다. 이미 앞에서 살펴본 바와 같이 우리의 목적은 간단한 모형과 설명에 주로 의존하는 오차 효과에 대해 살펴보는 것이다. Kuha and Skinner(28장)는 분류 오차가 있을때 범주형 자료 분석에 대한 식견 뿐 아니라, 이 분야의 광범위한 참고자료를 제공했다. 더 자세한 정보를 원하는 독자는 28장을 참조한다. Mote and Anderson(1965)은 또한 SRS 표본 추출 방법에 대해서 고찰하고 Rao and Thomas(1991)는 그들의 많은 결과물들을 복합 표본 추출 설계로 확장시킨다. 이 절은 이러한 작업을 총체적으로 다룬다

15.4.3.1 적합성 검정

π_i는 K 범주 분류 변수 μ의 i번째 범주에 속하는 모집단의 비율을 나타낸다. $\pi^{0i}, i = 1,...., k,$ 는 $\sum_i^K \pi^{0i} = 1.$ 인 것과 같이 [0, 1]사이의 상수를 나타낸다. 적 설은 다음과 같다.

$$H_0 : \pi^i = \pi^{0i}, \quad i = 1,...., K \tag{15.46}$$

이전과 같이 우리는 크기 n의 SRS이 모집단에서 선택되었다고 가정하고 i번째 계급에서 표본 비율을 p_i로 표기하며 표본 추출 및 측정 오차 분포 모두에 대해서 p_i의 기대값을 π^{i*}로 표기하기로 가정한다. 예제에서 μ은 K=4 범주를 가지는 연령 변수이다 즉, 21세이하, 21세-35세, 36세-65세, 65세 이상의 4개의 범주로 분류되는 변수를 말한다. 뿐만 아니라, 이 범주에서 이전의 인구 조사 비율은 $\pi^{i*}, i = 1,2,3,4$ 라고 가정하며 우리는 현재의 타켓 모집단 비율 π_i가 인구조사 비율과 동일한지 여부를 검증하기를 원한다.

보통 적합성 검정은 (15.47)에서 주어지는 피어슨 검정 통계량을 사용한다.

$$X^2 = n\sum^K \frac{(p_i - \pi^{0i})^2}{\pi^{0i}} \tag{15.47}$$

피어슨 검정 통계량은 분류 오차가 없을때 귀무가설 H_0하에서 x^2_{K-1} 변수로 점근적으로 분포한다. 분류 오차가 있을때 Mote and Anderson(1965)은 적합성 검증이 검증에 대한 명목 수준보다 더큰 제I종오차의 확률을 이끈다는 것을 증명하였다. 뿐만 아니라, Mote와 Anderson은 오분류를 보완하기 위해서 X^2를 수정하기 위한 여러 프로시져를 제안하였다. 이들 절차들은 검정 크기를 수정하면서도 제한적인 검정력을 감소시키는 것들이었다.

X^2에 대해서 분류 오차의 효과를 설명하는 것은 최초에 Mote and Anderson이 제안한 오분류에 대한 단순 모형을 고려한다. 순서형 범주에 대해서, 인접한 범주에 대해서만 오분류가 발생한다고 가정하는 것이 타당할지도 모른다. 위의 연령 예제를 고려해보면 이것은 응답자들이 그들의 옳은 연령 범주의 좌측 또는 우측으로 한 범주씩 오분류 되었을지도 모른다는 것을 의미한다. 그러나 그 이상으로 오분류되었을 가능성은 거의 희박하다. r을 구간 [0, 1]사이의 상수라 하고 한 범주에서 인접한 범주로 오분류된 미지의 비율로 정의하자. 뿐만 아니라, r은 모든 범주에서 동일하다고 가정하자. 그러므로, 다음과 같다.

$$\begin{aligned}
\pi_1^* &= \pi_1(1-r) + \pi_2 r \\
\pi_1^* &= \pi_{i-1} r + \pi_i(1-2r) + \pi_{i+1} r, \quad i = 2,\ldots, K-1 \\
\pi_1^* &= \pi_{K-1} r + \pi_K(1-r).
\end{aligned} \tag{15.48}$$

이 모형하에서 분류 오차로 제1종 오차의 증가된 확률의 효과를 설명하기 위해서, 우리는 K=4일때 $0 \le r \le 0.25$에 대한 H_0하에서 π_{01}=0.4, π_{02}=0.3, π_{03} =0.2, π_{04}=0.1으로 임의로 설정한 후에 (15.47)에 X^2을 계산했다. 그림 15.6은 n: 500, 2,500, 5,000에 대한 세값에 대한 결과를 보여준다. 그림에서 수평선은 =Pr(Type I error)=0.05, 즉, c23, 0.95를 가지는 검정의 임계값이다. r = 0.36, n = 500(그래프에서는 보이지 않음), r = 0.16, n = 2,500, r = 0.11, n = 5,000일때, 귀무가설 H_0는 틀리게 기각될지도 모른다(수평선 위쪽에 X^2값이 표시되어)는 점에 주의한다.

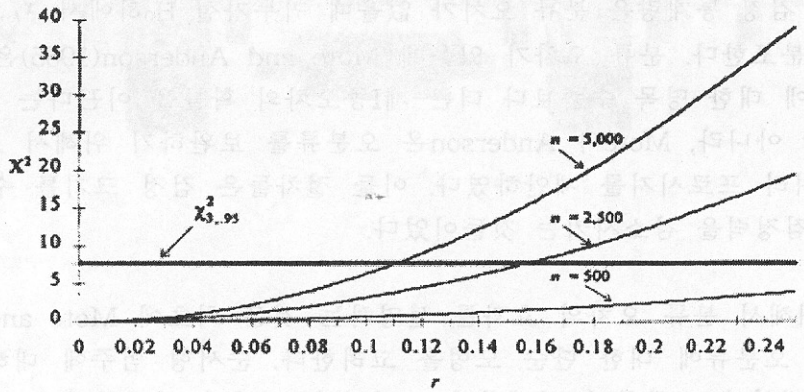


그림 15.6 표본 크기(n)와 인접 범주로의 오분류 비율의 함수로서 X^2에 대한 분류 오차의 효과

검정의 α 수준은 현실화된 α 수준 보다 상당히 작다. 게다가 이 효과는 기타 요인이 모두 같을때 n과 함께 증가한다. 분류 오차에 대해서 X^2을 수정하기 위해서 오분류 기제를 나타내는 모형을 적합할 필요가 있다. 모형을 적합시키기 위해서는 외부의 소스에서 추가 정보가 필요하다. Kuha and Skinner(28장)는 관심 모집단의 적어도 한개의 표본에서 알려진 두 가지의 가능한 외부 정보를 확인했다.

a. 골드 스탠다드 측정(기준 측정)과
b. 오분류 변수에 대한 반복 측정

골드 스탠다드는 검증용으로 사용할 수 있는 외부의 자원을 사용해서 오차 없는 측정을 하는 것을 말한다. 대안적으로 최초에 조사에 참여했던 응답자들에게 보다 정확하고 개선된 조사 과정을 사용한 재면접을 통하여 얻어질 수 있다. 예를 들어, 최초 조사 자료는 한 수준 낮은 면접원을 사용하여 중앙 방식의 전화조사를 통해 수집하는 한편 최고의 면접원들을 사용해서 일대일 재면접을 실시하는 방법이다. 이 분석의 목적에 대해서 골드 스탠다드가 옳다고 여겨진다.

반복측정은 재면접을 통해서 얻어질 수 있으며, 재면접시 개선된 방법을 사용해야 할 필요는 없다. Kuga and Skinner는 오분류 모수에 대한 정보는 통계적 모형과 특성의 실제 값간의 관계와 특성의 반복측정을 통해서 얻어질 수 있다고 지적했다. 반복 측정치들 간에 서로 독립적이라는 것이 가장 보편적인 가정이다. Mote and Anderson(1965)은 (15.48)의 가정이 유효하다는 가정하에 조사 자료로부터 직접 r을 추정하는 과정을 설명하였다. Mote and Anderson(1965)은 결과로서 생기는 검정의 점근적 검정력이 분류 오차가 없는 검정의 검정력 보다 약하다는 것을 증명했다.

15.4.3.2 연관성 검정

I와 J의 범주를 가지는 변수 μ^A와 μ^B의 이원 교차 분류에 대해서 고려해 보자. p_{ij}는 ij번째 모집단의 비율을 나타낸다고 하자. $\pi_i = \sum_j \pi_{ij}$ 및 $\pi_j = \sum_i \pi_{ij}$ 라고 하고 이들은 각각 행과 열의 주변 확률을 각각 나타낸다고 하자. μ_A와 μ_B은 독립적이라는 가설은 다음과 같다.

$$H_0 : \pi_{ij} = \pi_i . \pi_j, \quad i = 1,..., I; j = 1,..., J. \tag{15.49}$$

보통 x^2 검정 통계량은 다음과 같다.

$$X^2 = n \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(p_{ij} - p_i \cdot p_{\cdot j})^2}{p_i \cdot p_{\cdot j}} \tag{15.50}$$

여기서 p_{ij}는 ij번째 셀의 표본의 비율이며 $p_{i\cdot}$와 $p_{j\cdot}$이 각 행과 열의 주변확률이다. μ_A 에 대한 분류 오차가 μ_B에 대한 분류오차에 대해서 독립적이면, Assakul and Proctor(1967)에 따르면, 귀무가설 H_0는 검정 $H_0^* : \pi^*_{ij} = \pi^*_{i} \cdot \pi^*_{j}$와 동일하다. 여기서 $x^2_{(I-1)(J-1)}$에 대해서 검정에 대한 임계값을 사용하고 $\pi^*_{ij} = E(p_{ij})$이다. 그러나 이 검정의 점근적 검정력은 만일 측정오차가 없다면 작아질 것이다.

효과를 설명하기 위해서, 우리는 그림 15.6의 적합성 검정 설명에 사용된 분류오차에 대한 단순 모형의 확장된 형태를 고려한다. 그러므로, 우리는 오분류는 μ_A에 대한 비율 r_A 과 μ_B에 대한 비율 r_B로 인접한 범주에만 일어날 수 있다고 가정하자. 논의를 보다 단순화 하기 위해서 우리는 단지 하나의 정사각형 표만을 고려한다 즉, ij,k는 동일하다. 그리고 우리의 목적은 오분류에서 오는 검정력을 작게하는 것을 설명하는 것이기 때문에 우리는 μ_A와 μ_B 간의 완전한 상관관계를 가정한다. 그러므로 우리는 $i \neq j$에 대해서 $\pi^{ii} > 0$ 이고 $\pi^{ij} = 0$ 이며 다음의 기대 셀 확률을 정의한다.

$$\pi^*_{12} = \pi_{11} r_A + 0.5 \pi_{22} r_B \qquad \pi^*_{21} = 05 \pi_{22} r_A + \pi_{11} r_B$$

$$\pi^{K,K-1} = 1_{\pi^{K,K} r_A} + 0.5 \pi^{K-1,K-1} r_A + \pi^{K,K} r_B$$

$$\pi^{*ii} = \pi^{ii}(1 - r_A - r_B)$$

$$\pi^{*i+1,i} = 0.5(\pi^{i+1,i+1} r_A + \pi^{i-1,i-1} r_B)$$

$$\pi^{*}_{i,i+1} = 0.5(\pi_{i,i} r_A + \pi_{i+1,i+1} r_B), \quad i = 2,..., K-1. \qquad (15.51)$$

이 가정에 대한 추론은 (15.48)에 대한 추론 '셀의 오분류는 인접한 셀로만 일어난다.'와 유사하다. 이원표에 대해서, 이는 사실 $2 \le i \le K-1$에 대해서 셀(i,i)에 속해있는 개인이 셀$(i-1, i)$, $(i+1, i)$, $(i, i-1)$, $(i, i+1)$로 분류될 수 있는 양의 확률을 가진다는 것을 의미한다. 예로써, $i = 1, \ldots, 4$, $r_A = r_B = 0.10$에 대해(즉, 대각셀들의 20%가 인접한 행과 열로 오분류되었을때), $K = 4$, $\pi_{ij} = 0.25$라고 가정해 보자. 그리하여, 실제의 관찰된 표는 표 15.2에서 볼 수 있다.

(15.51)의 모델에 대해서 $X^2(r_A, r_B)$는 오분류 모수 r_A, r_B로 y_A, y_B에 대해서 계산된 X^2값을 나타낸다고 하자.

그러므로, $X^2(0,0)$은 오분류가 없을때 X^2 값(즉, $r_A = r_B = 0$)이다.

표 15.2 실제 비율이 π_{ii}=0.25, π_{ij}=0일때 분류 오차를 가지는 두변수에 대한 관찰된 셀 비율

실제 연관 μ^B					관찰 연관 y_B				
μ^A	1	2	3	4	y_A	1	2	3	4
1	0.25	0	0	0	1	0.20	0.0375	0	0
2	0	0.25	0	0	2	0.0375	0.20	0.025	0
3	0	0	0.25	0	3	0	0.025	0.20	0.0375
4	0	0	0	0.25	4	0	0	0.0375	0.20

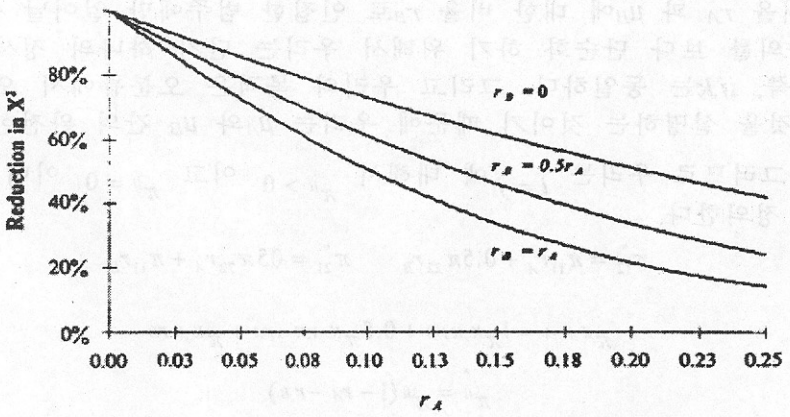


그림 15.7 오분류가 없을 때 X^2값에 대한 오분류가 있을때 X^2값의 비율.

표 15.2의 예제에서, $X^2(0, 0)$에 대한 $X^2(0.10, 0.10)$은 0.94인 비율에 대해, n = 1,000, $X^2(0, 0)$은 2,625이며(왼쪽 표 계산) $X^2(0.10, 0.10)$은 2,469 이다(오른쪽 표 계산). 그러므로, 관찰된 X^2는 오분류가 없는 실제 X^2의 94%이며 이는 오분류의 결과로 X^2이 작아진다는 것을 설명한다. 그림 15.7에서 비율 $R_X^2 = X^2(r_A, r_B)/X^2(0,0)$가 $0 \le r_A \le 0.25$와 세가지 조합(r_A와 $r_B : r_B = 0, r_B = 05r_A$와 $r_B = r_A$)에 대해서 그래프로 나타나 있다. 모든 경우에, r이 증가함에 따라 R_X^2는 감소하며 이리하여 X^2에 대해 희박해 지고 연관성 검정에서 검정력의 손실을 나타낸다.

Mote and Anderson(1965), Rao and Thomas(1991), Kuha and Skinner (28장), Van de Pol and Langeheine(29장) 모두 두 변수간에 상관관계를 가지는 측정 오차를 포함한 보다 복잡한 분류 오차의 경우를 고찰했다. 이 경우 타당한 검정을 하기위해서 모형 기반의 방법을 사용할 필요가 있다. 두 단계의 표본 추출 방법을 통해 얻어진 검증 자료를 사용하여 X^2를 수정하는 것이 일반적으로 X^2가 제한적인 분포를 따르지 않는다는 것이 증명되어 왔다. 그러나, 만일 모형 기반의 방법이 실제값의 일관적인 추정량을 제공한다면, 점근적으로 옳은 검정이 개발될 수 있다.

15.5 측정오차 효과 보완방법

이미 이전의 논의로 명확히 증명된 바와같이 측정 오차 분석의 효과는 매우 중요하다. 일변량 분석에대해서 상관 오차는 무상관 오차보다 더 심각하다. 그러나 사분위 추정과 다변량 분석에 대해서 무상관 오차라도 결과치에 상당한 편향을 초래할 수 있다. 측정 오차가 있을때 추정량은 편향될것이며 분산은 커질것이며 가설 검정은 타당하지 않거나 설득력이 없게 될 것다. 뿐만아니라 다른 형식의 분석으로 부터의 결론도 오도될 수도 있다. 실제로 조사 방법론을 사용하는 현장에서는 좋은 조사 설계 방법을 통한 조사 자료에 대한 측정 오차를 줄이는데 주로 집중한다. 이것은 설문지 설계라는 지적 방법을 포함한다. 예를 들어 개선된 채용방법, 훈련, 관리, 감독을 통한 면접원과 다른 조작원의 오차를 줄이는 방법, 응답자에게 친근한 면접 방법사용 등등(Groves, 1989 참고). 오차를 줄이고 통제하려는 이러한 노력에도 불구하고 측정 오차는 자료 수집에 있어서 문제가 되고 있으며 오차 효과를 보완하기 위한 방법이 최종적으로 요구되어지고 있다.

조사문헌에는 조사 자료분석에서 측정 오차를 조정하려는 방법을 제안하는 수많은 논문들이 들어 있다. 이 절에서는 이러한 방법들에 대해서 살펴본다. 게다

가, Biemer 등(1991, Section E)은 우리가 여기서 논의한 종류의 분석에 대한 것 뿐만 아니라, 구조 방정식 모델링, 패스 분석 및 이벤트 히스토리 분석에 대한 오차의 보완방법에 대해서 아홉장에 걸쳐서 설명한다. 오차 보완방법은 두개의 필수 요건에 따라서 특징지워진다. (a)오차 구조의 모형과 (b)모형에 의한 오차 모수를 추정하기 위한 외부 자료 및 보조 자료 모형의 측정 오차에 대해서 15.2절에 명기된 모형들이 보통 대부분의 측정 프로세스를 설명하기 위해서 수정될 수 있다. 그러나 몇 경우에는 이러한 일반적인 모형의 계통적 서술이 불충분하여 대안이될 모형이 반드시 명기 되어야 하는 경우도 있다. Pullum and Stokes(31장)는 출생률 추정의 오차를 모델링하여 발표했으며 van de Pol and Langeheine(29장)는 장기적인 범주형 자료를 분석하기 위한 잠재 계급 모형을 사용하는 적용법을 제공했다.

외부 자료의 필요 조건은 보통 기록 확인, 보다 정확한 측정자료를 얻기위한 두 단계 표본 추출 방법, 응답자 부표본 재면접을 통해 얻어진 반복측정 자료, 그리고 이전인구 조사 자료로 만족된다. Kuha and Skinner(28장)는 외부정보(즉, 관리 기록 자료, 정확도가 높은 재측정 자료, 분석의 목적에 입각하여 오차가 없다고 판단되는 기타 자료)와 응답자들과의 재면접을 통해서 얻어진 반복측정 자료를 확인자료로서 분류했다. Rao and Sitter(33장)에 의해서 제안된 방법은 검정자료가 측정 편향에 대한 평균과 합계의 추정량을 수정하는데 사용할 수 있다고 가정한다. 이 절에서는 수정된 추정량의 표준 오차를 추정하기 위해서 재표본 추출 방법에 초점을 맞춘다. van de Pol and Langeheine에 의한 29장은 반복 측정의 적용과 장기 자료 분석에 대한 잠재 계급 모형의 사용방법을 제공한다. 노동시장의 범주에서 실제의 움직임으로부터 측정 오차를 분리함으로써 그들은 비용변화 뿐 아니라, 총액 흐름도 측정하려 했다. Pullum and Stokes(31장) 또한 장기자료를 사용할때 측정 오차 분석 적용방법을 제공했다.

표 15.3 추정량과 자료 분석에 대한 오차 효과 요약

통계량/모수	오차 구조	오차 효과
$\bar{y}$ p	상관 또는 무상관	Bias = B_d Bias = $-\pi\theta + (1-\pi)\phi$
Var($\bar{y}$) Var(p)	무상관	$Var(\bar{y}) = \sigma_\mu^2 / (R_n)$ $Var(p) = \pi^*(1-\pi^*)/n$
v($\bar{y}$) v(p)	무상관	비편향 (fpc 없는 경우) 비편향 (fpc 없는 경우)
Var($\bar{y}$) Var(p)	상관	$Var(\bar{y}) = \sigma_\mu^2 / R_n(1+(m-1)p_y)$
v($\bar{y}$) v(p)	상관	Relbias = $\dfrac{(m-1)p_y}{1+(m-1)p_y}$
$\hat{Q}_y(a)$	상관 또는 무상관	bias = $=(r^{-1/2}-1)\left[Q_\mu(a)-\overline{M}\right]+B_d$
β β^0	무상관	$E(\hat{\beta}) = R\beta$ $E(\beta_0) = \beta_0 + \beta\overline{M}\left[1-R-B_d/\overline{M}\right.$
β, β^0	상관	한쪽방향에서 편향될수 있음
Corr(z,y), y with error Corr(z,y), z and y with error	무상관	Corr(z,y) = $\sqrt{R}$ Corr(z,μ) Corr(z,y) = $\sqrt{R_x R_y}$ Corr(η,μ)
Corr(z,y), y with error Corr(z,y), z and y with error	무상관	한쪽방향에서 편향될수 있음
ANOVA ANOCOV	상관 또는 무상관	요인 차이가 한쪽방향에서 편향될수 있음
적합성에 대한 X^2	무상관	X^2 편향됨. 검정크기가 명목 수준보다 큼
연관성에 대한 X^2	무상관	X^2 편향됨. 검정력 손실

인구학적 분석 모델링 기법과 일반적인 추정 방정식(GEE)를 사용할때 그들은 파키스탄에서 출생 신고가 누락되어 있거나 바뀌어져 있는 것에 대한 유용한 분석을 제공했다. Van de Pol and Langeheine(29장)와 Nusser and Fuller(30장)는 모두 반복 특정을 그들의 측정 오차 분석에 병합시켰다.

Bassi and Fabbris(32장)는 외부 자료가 필요하지 않은 ANOVA 기법을 사용하여 측정 오차 요인 추정의 예를 제공했다. 오히려, 면접원과 다른 조작원의 할당량을 실험 계획법을 사용해서 상호 관입(무작위하게 분포)시킴으로써, 조작원 분산과 면접원간 상관계수(ρ_y)가 가능하다. 이 절은 이러한 추정량의 정확성 뿐만 아니라, 면접원 분산을 추정하기 위한 상호 관입 표본과 반복측정을 합하여 발생하는 추가적인 이점도 고찰한다

측정 오차를 보완하기 위한 몇 가지 방법이 있다. 범주형 자료에 대해서는 Kuha and Skinner(28장)가 확인 자료와 반복 측정 자료의 유효성과 함께 수정 방법을 요약한다. 각각의 방법은 사용의 용이성, 효율성, 가설 검정의 타당성의 면에서 각각 장단점이 있다. 반복측정에 대한 중요한 새로운 개발은 실제 값에 대한 정보를 추출하기 위해서 잠재 계급 모형을 점점 많이 사용하고 있다는 것이다. 비록 모형의 잠재력이 상당기간 인정은 받아왔더라도 잠재 계급 모형이 실제로 사용된것은 최근에 일이다.

회귀와 상관분석은 측정 오차로 인한 편향에 대해서 수정하기 위한 몇 가지 잘 알려진 방법이 있는 다른 분야이다. 도구 변수를 사용하는 것은 흔한 방법중 하나이며 이 변수들은 대부분의 계량 경제학에 요약되어있다. (Fuller, 1987 참조). 최대 우도 추정법(Maximum likelihood estimation methods)은 측정 과정에 대해서 강한 가정을 필요로 하고 이러한 가정들이 직접 실행가능하지 않으면 방법이 제한적이 되지만, 이미 제안되었다. Fuller(1991)는 보다 유용한 방법을 개발해 반복 측정이 가능하게 했다.

이 절에서 논의된 오차 효과의 요약이 표 15.3에 제공된다. 이 표에서는 측정 오차가 어떤 경우에는 단지 미미한 효과로 끝나는 반면, 가끔 자료 분석에 대한 표준 방법을 무효로 만들 수 있다는 것을 명확히 보여준다. 측정 오차가 문제인지 결정하는 것은 분석이 잘 수행될 것인지에 대한 고려가 필요하며, 해당 분석에 대해 오차를 야기하는 편향의 크기에 대한 지식이 필요하다. 의심이 가는 부분에 대해서 자료를 조사하는 것은 조사 자료 처리단계에서 필요한 단계일지라도, 조사만으로는 측정 오차의 편향 효과에 대한 정보를 얻기가 힘들다. 오히려, 오차 성분을 추정하고 자료 분석 동안 오차 효과를 수정하기 위해서 필요한 것은 잘 명기된 모형과 충분한 외부 자료이다.

Fuller(1991)는 설문 예산에서 회귀 계수 추정량을 수정하기 위해서 응답자들을 재면접하기 위해 사용되는 비용을 조금씩이라도 할당 받아야 된다고 주장했으며, 이 정책은 정책이 없을 때 보다 추정량에 작은 평균제곱에러를 제공한다는 것을 증명했다. 그러나 이러한 정책이 조사 스폰서들에게 많은 지원을 받았다는 증거는 거의 없다. 아마도 조사 분서의 결과를 수정하기위한 반복 측정이나 골스스탠다드 측정법을 사용하는데 지원이 부족했던 이유는 수정 방법의 가정을 만족시키는 재측정 자료를 얻기가 힘들어서 일 것이다. 우리의 희망은 이 절이 모든 종류의 자료 분석에 대해서 측정 오차의 편향 효과에 대한 우려와 의식을 증대시킴으로써 이 분야에서 필요한 사람들이 더 큰 주목을 받을 수 있도록 일조하는 것이다.

감사의 글

이 절의 초고에 유용한 조언을 주신 James Chromy와 Siu-Ming Tam에게 감사의 말씀을 전합니다.

제 16 절 상호 관입 할당량으로 감독되는 면접 - 재면접 조사에서 비표집 오차의 추정량

Francesca Bassi and Luigi Fabbris
University of Padua

16.1 서문

면접원에 의해서 자료가 수집되는 조사에서는 비표집오차의 주요 원인은 응답자, 면접원, 감독원과 다른 현장 훈련자, 코딩요원, 입력자들이다. 응답자들은 설문지 미회수, 사회적인 조건, 혼동 등으로 인해서 부정확한 응답을 줄 수 있다. 만일, 응답이 완전히 무작위하게 실제 값에서 벗어난다면 소위 무상관 응답 오차가 발생한다. 그러나 오차가 계통적이라면 편향이 발생한다. 면접원 오차는 면접동안 면접원의 성격, 태도, 행동에 관련될 수 있다.

오차가 한명의 면접원에게서 발생한다면, 이 오차는 편향이다. 이에 반해서 오차가 표본 모집단과 관련이 있다면 상관 무작위 오차이다. 코딩요원과 자료에 영향을 주는 다른 조작원의 오차는 통계적으로 면접원의 오차와 유사하다.

감독원 오차는 편향의 다른 잠재적인 원인이다. 훈련과 커뮤니케이션을 통해서 감독원은 직접 면접원에게 영향을 주고 그러므로, 결과적으로 면접원이 수집하는 응답에 간접적인 영향을 끼친다. 특정 감독원 집단에 대한 계통적 오차는 고정 편향으로 모형화될 수 있을것이다. 반면에 큰 규모의 감독원 모집단의 오차는 상관 무작위 오차로 적절히 모형화 된다. 앞의 경우에서 조사에 유효한 특정 감독원 집단은 큰 규모의 잠재 감독원 모집단으로부터 추출된 무작위 표본으로 고려된다.

훈련은 설문조사 연구원이 지시한대로 계속된다. 우리는 이런 종류의 오차를 지시 오차(instruction error)라고 한다. 조사 연구원의 부주의로 정보나 태도 마음가짐이 그냥 통과될 수도 있다. 이런 오차는 후에 설문 조사 과정에서 나타나게 된다.

이 절에서 우리는 면접원 투입되는 조사에서 비표본 오차에 대한 연구에 모형을 제공한다. 우리의 모형은 면접원과 감독원 효과 모두를 병합한 것이다. 표본 평균과 표준 오차들의 분산의 다양한 성분의 추정량이 제안되고 이들 성분의 효율성이 비교된다. 뿐만 아니라, 한 면접을 토대로 상관 응답 분산의

추정량인 평균 제곱 오차(mean squared error(MSE))가 면접 자료와 재면접조사로부터 얻어진 자료를 합한 추정량과 비교된다.(가끔 이중 표본 추정량이라고도 한다.) 비록 면접- 재면접 자료가 응답자에 대한 더 만은 정보를 포함하고 있다 하더라도, 재면접 자료가 모조사(parent survey)로부터의 자료와 상관관계에 있을지도 모르기 때문에 이 자료가 상관 응답 분산의 추정량의 효율성을 개선시키지 못할 수도 있다.

이 절에서 우리는 이중 표본 추정량이 단일 표본 추정량 보다 개선된 추정량이 될 수 있는 충분한 조건을 제공한다.

16.2 모형

면접원이 투입되는 면접법을 사용하여 모집단에 대해 표본 조사가 실행되는 경우를 고려한다. 후에 재면접 조사가 평가를 목적으로 조사 응답자의 부표본에 대해 실행된다. 모집단은 크기가 N^*이며 H층으로 나뉜다. 각 층에서 반복(replacement)없이 표본을 추출한다. 우리는 각 층이 동일하게 N개 이며 각 층의 표본의 크기 $n=km$이다. 여기서 k는 평가에서 면접원의 수를 나타내며, m은 각 면접원의 할당량의 크기를 나타낸다.

게다가 층내의 면접원에 대한 할당량은 무작위하게 주어진다. 면접원에대한 무작위 할당량은 상관관입(Mahalanobis, 1946)으로 알려져 있으며 하나의 표본 조사 설계에서 상관 오차 분산의 추정이 요구된다. 마지막으로, 우리는 각 H층이 한 감독원에게 할당된다고 가정한다. 그리고 동일한 감독원과 면접원 할당량의 단순 가정이 우리가 결론을 보다 복잡한 조사의 경우로 일반화하는 것을 제한할 수는 없다.

우리는 모조사에서 사용되었던 동일한 상관관입 할당량 설계가 재면접조사에도 또한 사용된다고 가정한다. 게다가 보편성이 없지 않다면 모조사의 모든 단위가 재면접 조사에도 있다고 가정한다.

표 16.1 모조사(parent survey)와 재조사에서의 면접원과 감독원의 할당량

	모조사				재면접			
감독원	1		2		1		2	
면접원	1	2	3	4	1	2	3	4
부표본	A	B	C	D	B	A	D	C

그러나, 재면접 조사에서는 면접원 할당량이 동일한 층(stratum) 내의 다른 면접원과 바뀌어서 각 재면접원이 본인의 할당량이 아닌 것을 하게 된다. 면접원은 모조사와 재조사에서 동일한 감독관을 가지게 된다는 점에 주의한다.

설명하기 위해서 두명의 감독원(H = 2)과 4명의 면접원(한 층에 k = 2)이 있는 경우를 생각해 보자. 표 16.1에서 문자 A, B, C, D는 모조사와 재면접 조사를 위한 크기 m인 면접원의 할당량을 나타낸다.

우리의 가정은 추가적인 계층 오차의 효과, 즉 감독관의 효과를 포함하므로써 Fellegi(1964) 가정을 확장한다. 이중 상호 관입인 면접- 재면접 설계는 H=1일때 Fellegi(1964)와 동일하다는 것은 쉽게 증명될 수 있다.

우리의 설계하에서는 층 h의 한 단위에 대한 응답 모델은 다음과 같다.

$$y_{hij} = \mu_{hij} + b_h + a_{hj} + r_{hij} \qquad (16.1)$$

여기서 y_{hij} (h=1, . . . , H; i=1, . . . , k, j=1, . . ., m)는 정규 분포 확률 변수 Y를 나타내며, μ_{hij}는 층 h내의 면접원 i의 할당량 내에서 대응하는 참값이거나 단위 j에 대한 오차 없는 값이다.

a_{hi}는 hi번째 면접원으로 인한 오차이며, b_h는 h번째 감독원의 오차이다. 그리고, r_{hij}는 μ_{hij} + a_{hi} + b_h부터 y_{hij}의 편차이다. 이 값은 응답자 오차라고 불려질 것이다. 그러므로, 우리는 응답은 응답자, 면접원, 감독원의 오차에 의해서 외곡된 각 개인의 참값 μ_{hij}의 결과라고 가정한다.

성분 추정을 위해서 다음과 같이 가정한다.

1. N 표본 크기가 n = km이며 크기 N인 각 층에서 반복없이 무작위하게 추출된다. 잠재 면접원이 포함된 대규모의 모집단에서 무작위하게 선택된 H_k 면접원들은 하나의 표본을 구성한다. 면접원들 각각은 m개로 구성된 하나의 표본을 할당 받는다. k명의 면접원이 한 그룹이며 한 층 내에서 단독으로 운영된다.
2. 모형 (16.1)의 오차 성분 분포는 평균이 0이고 분산이 알려진 정규분포를 따른다. 응답자 오차 r_{hij}는 정규분포 $N(0, \sigma_a^2)$부터의 i.i.d. 값이며, 면접원 오차, a_{hi}는 $N(0, \sigma_b^2)$로 부터 i.i.d이다. 감독원 오차, b_h,는 정규분포 $N(0, \sigma_r^2)$를 따르며 i.d. 이다.(즉, 동일한 분포를 따르나 독립은 아니다.) 이는 관측치에 대해서 결합 분포는 다변량 정규분포라는 것을 의미한다.

3. 면접원과 감독원의 오차는 서로 독립이며 응답자의 오차도 서로 독립이다. 표본 편차 및 응답 오차에도 서로 독립이라고 가정한다.

층 h의 i번째 할당량의 j번째 단위의 관측치는 y_{hij}로 다음과 동일할 것이다.

$$y_{hij} = \mu_h + d_{hij} + c_{hij} \qquad (16.2)$$

(Hansen 등, 1961 참조) 여기서,

$\mu_h = \sum_{i=1}^{k}\sum_{j=1}^{m}\mu_{hij}/n$ 은 변수 Y에 대한 층의 평균이며

$d_{hij} = r_{hij} + b_h + a_{hi}$ 는 단위 j의 응답 편차로 면접원 i에 의해서 열거되고 h에 의해서 감독된다. $C_{hij} = \mu^{hij} - \mu^h$ 는 동일 단위의 표본 편차이다.

E( )와 V( )는 각각 비표본수집 오차 분산과 표본 설계에 관련해서 기대값과 분산을 나타낸다. 가정 (1)-(3) 하에서 층 h에 대한 관찰 평균의 분산은 (16.3)에 주어지며, (16.3)은 Fellegi(1964) 방식의 비표본 오차 구성을 정의한다.

$$\bar{y}_h = \frac{1}{n}\sum_{i}^{k}\sum_{j}^{m} y_{hij} = \frac{1}{k}\sum_{i}^{k}\bar{y}_{hi}$$

$$V(\bar{y}_h) = \frac{N-n}{N-1}\frac{\sigma_{sh}^2}{n} + \frac{\sigma_{rh}^2}{n}\left[1 + (m-1)\delta_{zh} + m(k-1)\delta_{3h}\right. \qquad (16.3)$$

여기서 σ^2_{rh} 다음과 같이 정의되는 무상관 응답 분산이다.

$$\sigma^2_{rh} = E\left\{\frac{1}{km}\sum^{k}\sum^{m} d^2_{hij}\right\}$$

여기서 σ^2_{Sh}는 층 h의 변수 Y의 표본 분산(즉, 층 내의 실제값의 분산)이며 다음과 같이 정의된다.

$$\sigma^2_{sh} = E\left\{\frac{1}{km}\sum^{k}\sum^{m} c^2_{hij}\right\}$$

우리는 δ_{2h}가 면접원의 할당량내의 응답자 편차간의 상관계수이고 다음으로 정의되며,

$$\delta_{2h} = \frac{1}{\sigma_{rh}^2} E\left\{\frac{1}{n(m-1)}\sum_{i}^{k}\sum_{j\neq j'}^{m} d_{hij}d_{hij'}\right\}$$

$\delta^{2h}\delta^{2}_{rh}$는(상관 관계에 있는)면접원 분산이라는 것을 알 수 있다. 마지막으로, 우리는 감독원의 할당량 내에서 응답자 편차간의 상관 계수, δ^{3h}를 정의한다.

$$\delta_{3h} = \frac{1}{\sigma_{rh}^2}\left\{\frac{1}{nm(k-1)}\sum_{i\neq i'}^{k}\sum_{j,j'}^{m} d_{hij}d_{hi'j'}\right\}$$

여기서, $\delta^{3h}\delta^{2}_{rh}$은 상관 관계에 있는 감독원 분산이다.

$$\bar{y} = \frac{1}{H}\sum^{H} \bar{y}_h$$

전체 모집단 평균 추정량은 분산 (16.4)를 가지는 다음과 같다.

$$V(\bar{y}) = \frac{N-n}{N-1}\frac{1}{H^2 n}\sum_{h=1}^{H}\sigma_{sh}^2 + \frac{1}{H^2 n}\sum_{h=1}^{H}\sigma_{rh}^2\left[1+(m-1)\delta_{2h}+m(k-1)\delta_{3h}\right]+\frac{1}{H^2}\sum_{h\neq h'}^{H}\sigma_{4hh'}\sigma_{rh}^2\sigma_{rh'}^2$$

(16.4)

여기서, 항 $\sigma^{4hh'}$는 교수 효과, 즉 두 다른 층에서 얻어진 응답 ㅍ 상관 계수이다.

$$\delta_{4hh'} = \frac{1}{\sigma_{rh}\sigma_{rh'}} E\left\{\frac{1}{k^2k^2}\sum_{i,i'}^{k}\sum_{j,j'}^{m} d_{hij}d_{h'i'j'}\right\}$$

(16.3)과 (16.4)의 오른쪽의 첫번째 항은 표본 추출론 인한 분산을 나타내는 반면 나머지 항들은 응답자 오차에 기인한다는 점에 주목한다.

목표 모집단 Y에 대해서 평균과 분산을 각각 (와 (2 로 각각 표기하자. 그러면

$$\sigma^2 = E\left\{\frac{1}{Hn}\sum_{h=1}^{H}\sum_{i=1}^{k}\sum_{j=1}^{m}(y_{hij}-\mu)^2\right\} = \sigma_s^2 + \sigma_r^2$$

여기서 $\sigma_s^2 = 1/H\sum_{h=1}^{H}\sigma_{sh}^2 + 1H\sum_{h=1}^{H}(\mu_h-\mu)^2$ 그리고, $\sigma_r^2 = 1/H\sum_{h=1}^{H}\sigma_{rh}^2$ 이다. 게다가, 모든 층에 대한 응답자 분산은 층에 걸친 산술 평균이고, $\delta^4\sigma^{r^2}$는 다음과 같다.

$$\delta_4\sigma_r^2 = \frac{1}{H(H-1)}\sum_{h\neq h'}^{H}\delta_{4hh'}\sigma_{rh}^2\sigma_{rh'}^2$$

그러므로 (16.4)는 다음과 같이 쓸 수 있으며

$$V(\bar{y}) = \frac{N-n}{N-1}\frac{1}{Hn}\left[\sigma_s^2 - \frac{1}{H}\sum_{h=1}^{H}(\mu_h-\mu)^2\right] + \frac{\sigma_r^2}{Hn}\left[1+(m-1)\delta_2 + m(k-1)\delta_3 + n(H-1)\delta_4\right] \qquad (16.5)$$

여기서 첨자 h가 없는 항은 모집단 모수를 가르킨다. 이제 재면접 조사와 관련해 응답 오차를 고찰해 보자. 우리는 모형 (16.1)과 (16.2)이 모두 재면접 자료에 적용된다고 가정한다. 분산 성분이 만일 모조사와 관련이 있다면 1로, 재면접 조사와 관련이 있다면 2로 표기한다.

재면접과 관련있는 분산 성분 뿐만 아니라, 우리는 또한 조사간의 분산 모형 오차의 상관관계를 정의할 필요가 있다. 그러므로 우리는 다음과 같이 모수를 정의한다. $(_{1h}$는 다른 조사에서 동일한 단위에 대한 응답자 오차간의 상관관계를 나타내며, 만일 응답자가 첫번째 조사에서 주어진 부정확한 답을 기억해 내서 재면접 조사에서 반복한다면 상관관계는 커지기 때문에 리콜 효과라고도 부른다. 그러므로, $(_{1h}$는 다음과 같다.

$$\beta^{1h} = \frac{1}{\sigma^{rh1}\sigma^{rh2}}E\left\{\frac{+}{n}\sum^{k}\sum^{m} d_{hij1}d_{hij2}\right\}$$

$(_{2h}$는 동일한 한 면접원이 두개의 다른 조사에서 질문한 두개의 다른 단위에 대한 응답자 오차간의 상관계수이다. 이는 다음과 같이 주어진다.

$$\beta_{2h} = \frac{1}{\sigma_{rh1}\sigma_{rh2}} E\left\{\frac{1}{nm}\sum_{i}^{k}\sum_{j\neq j}^{m} d_{hij1}d_{hij'2}\right\}$$

($_{3h}$는 두가지 조사에서 동일한 부표본에 속한 단위에 대한 응답자 오차간의 상관 계수이다. 이는 다음과 같이 주어진다.

$$\beta_{3h} = \frac{1}{\sigma_{rh1}\sigma_{rh2}} E\left\{\frac{1}{n(m-1)}\sum_{i}^{k}\sum_{j\neq j}^{m} d_{hij1}d_{hioj'2}\right\}$$

여기서, i(는 재면접 조사에서 면접원을 나타내며 i는 모조사에서 면접원을 나타낸다. ($_{4h}$는 다른 조사에서 다른 면접원에게 할당된 다른 단위의 응답 오차간의 상관관계이고, 다음과 같이 주어진다.

$$\beta_{4h} = \frac{1}{\sigma_{rh1}\sigma_{rh2}} E\left\{\frac{1}{k(k-2)m^2}\sum_{i'\neq i,io}^{k}\sum_{j\neq j'}^{m} d_{hij1}d_{hi'j'2}\right\}$$

면접원들이 두번째 면접에서 할당량을 서로 바꾸게되므로 k=2라면 ($_{4h}$는 의미가 없다는 점에 주의한다. 마지막으로 ($_{5hh'}$는 다른 조사에서 다른 감독원에게 할당된 단위의 응답 편차 사이에 상관계수이며 이는 다음과 같이 주어진다.

$$\beta_{5hh'} = \frac{1}{\sigma_{rh1}\sigma_{rh'2}} E\left\{\frac{1}{nkm}\sum_{i,i'}^{k}\sum_{j,j'}^{m} d_{hij1}d_{h'i'j'2}\right\}$$

($_{4h}$와 ($_{5hh'}$는 대략 ($_{3h}$와 ($_{4hh'}$와 거의 같은 크기라는 점에 주의한다. 우리는 ($_{4h}$은 동일한 층에 속한 단위들간 상관계수이며 ($_{5hh}$는 동일한 지시를 받는 현장 인력에 의해서 면접을 받는 단위간 상관계수라는 것을 알 수 있다. 식 (16.5)에서와 같이, 이러한 상관계수는 층에 걸쳐서 상관계수를 평균화함으로써 모집단 전체에 대해서 정의할 수 있다. 다음에서는 이러한 모집단 수준에서 상관계수는 첨자 h 없이 동일한 의미로 사용될 것이다.

16.3 오차 성분의 추정량

이 절에서는 무상관 분산의 추정량(16.3.1절), 면접원 분산의 추정량(16.3.2절), 감독원 분산의 추정량(16.3.3절)을 분산분석(analysis of variance (ANOVA))을 사용해서 구한다. 재 면접 조사가 있을 때와 없을 때 자료 수집설계에 적합한 추정량을 얻을 수 있다.

16.3.1 무상관 분산

Fellegi(1964)에 의해서 제안된 층 h, σ^{2}_{rh}에 대한 무상관 응답 분산의 추정량은 다음과 같다.

$$s^{2}_{rh} = \frac{1}{2k(m-1)} \sum_{i}^{k} \sum_{j}^{m} (y_{hij1} - y_{hi^{\circ}j2} - \bar{y}_{hi1} + \bar{y}_{hi^{\circ}2})^{2} \qquad (16.6)$$

Fellegi는 모조사 응답자의 부표본에 대해서 두번째 측정 없이 무상관 응답 분산은 표본 분산과 혼동되어 분리하여 추정할 수 없다는 것을 증명했다. Kish(1962)가 제시한 표본 추출 및 무상관 응답 분산 즉, σ^{2}_{s}과 σ^{2}_{r}의 합에 대한 추정치는 다음과 같다.

$$s^{2}_{sh} + s^{2}_{rh} = \frac{1}{k(m-1)} \sum_{i}^{k} \sum_{j}^{m} (y_{hij} - \bar{y}_{hi})^{2} \qquad (16.7)$$

16.3.2 면접원 분산

재면접조사 자료가 이용가능할때, 층h의 면접원 분산의 추정량, $h, \delta^{2h}\sigma^{2}_{rh}$은 Fellegi(1964)에 의해 다음과 같이 주어지며,

$$\begin{aligned} {}_{1}\hat{\delta}_{2h} s^{2}_{rh} &= \frac{1}{2(k-1)} \sum_{i}^{k} (\bar{y}_{hi1} - \bar{y}_{hi^{\circ}2} - \bar{y}_{h1} + \bar{y}_{h2})^{2} \\ &\quad - \frac{1}{2km(m-1)} \sum_{i}^{k} \sum_{j}^{m} (\bar{y}_{hi1} - \bar{y}_{hi^{\circ}2} - \bar{y}_{hi1} + \bar{y}_{hi^{\circ}2})^{2} \end{aligned} \qquad (16.8)$$

재면접조사가 아닌 경우의 추정량은 Kish(1962)에 의해 다음과 같이 주어진다.

$${}_{11}\hat{\delta}_{2h} s^{2}_{rh} = \frac{1}{k-1} \sum_{i}^{k} (\bar{y}_{hi} - \bar{y}_{h})^{2} - \frac{1}{km(m-1)} \sum_{i}^{k} \sum_{j}^{m} (y_{hij} - \bar{y}_{hi})^{2}. \qquad (16.9)$$

16.3.3 감독원 분산

마지막으로, 두 조사의 모든 층으로 부터 정보를 합친 감독원 분산의 추정량은 Bassi (1991)에 의해서 다음과 같이 주어진다.

$$
{}_1\hat{\delta}_3 s_r^2 = \frac{1}{2(H-1)}\sum_h^H (\bar{y}_{h1} - \bar{y}_{h2} - \bar{y}_1 + \bar{y}_2)^2
$$

$$
- \frac{1}{2Hk(k-1)}\sum_h^H \sum_i^k (\bar{y}_{hi1} - \bar{y}_{hi\circ 2} - \bar{y}_{hi1} + \bar{y}_{h2})^2 . \quad (16.10)
$$

재면접 조사 자료가 사용되지 않는 경우의 추정량은 Fabbris (1991) (16.11)에 의해서 다음과 같이 제공된다.

$$
{}_{11}\hat{\delta}_3 s_r^2 = \frac{1}{H-1}\sum_h^H (\bar{y}_h - \bar{y})^2 - \frac{1}{Hk(k-1)}\sum_h^H \sum_i^k (y_{hi} - \bar{y}_h)^2 . \quad (16.11)
$$

16.4 추정량의 효율성

추정량 (16.6)-(16.11)의 통계 분포를 계산하기 위해서는 다음의 결과가 필요하다. 첫번째 결과는 Bhat(1962)에 의한 것으로 정규 분포 오차의 가정하에서 표본 제곱합과 표본 분산은 카이제곱(chi-squared) 분포를 따른다. 게다가 상수 제곱은 별도로 보더라도 이것은 관측치들이 공통 분산과 공분산을 가지는 한 관측치가 상관관계를 가질때도 유효하다. 두번째 결과는 Matern (1949)에 의한 것이다. 정규분포를 따르며 상관관계를 가지는 변수들로 구성된 두 음이 아닌 무상관 정방형은 독립분포를 따른다. 이 두가지 결과는 면접원과 감독원 표본 평균이 동일한 분산과 공분산을 갖고 정규분포를 따른다고 가정하기 때문에 (16.1)의 모형에 적용할 수 있다.

16.2절에서 제안한 추정량으로 구성된 성분들은 서로 독립이며 상수는 별개의 카이제곱 확률 변수로 분포하기 때문에 추정량 (16.6)부터 (16.11)는 또한 카이 제곱 분포를 따른다. 카이제곱 분포의 평균과 분산에 대한 단순한 관계는 우리가 분산을 계산할 수 있게 해주고 그러므로 각 추정량의 평균제곱오차도 계산할 수 있다.

정방형에서 평균과 분산사이의 관계는 관측치가 분포의 네번째 적률(moment)이 두번째 적률의 제곱의 3배라는 조건하에서 비정규, 연속분포에서 추출되었을때 조차도 유효하다. 그러므로 정규 분포의 가정에서 약간 벗어나는 것도 가능하다. 추정량 (16.6)-(16.11)의 기대값과 분산은 부록에 제공되며 상관 응답 분산의 대체 추정량의 상대적인 효율성은 16.5절에서 논의된다.

16.5 추정량에 대한 비교

이 절에서 우리의 목표는 i = 1, 2, 3, 4, 5에 대해서 상관계수 ρ_i의 가정하에 재면접조사 자료를 토대로한 추정량과 모조사의 자료와 추정량의 효율성을 비교하는 것이다. 우리는 이러한 상관관계 사이에 계통적인 관계를 가정한다. 즉 i = 2, 3, 4에 대해서 ρ_i (ρ_{i+1}로 가정한다. 이렇게 가정함으로서 우리는 논의를 단순화 시킬 수 있으며 직관적으로 효율성의 비교를 이해할 수 있다. 이 가정은 감독원들은 평균 면접원들 보다 많은 지식을 가지고 있으며 연구원들은 조사에 대해 보다 많은 전문적 기술을 가지고 있다는 가설과 같은 맥락으로 볼 수 있다. 전문적 기술(지식)에 대한 계층은 재면접 조사에도 다시 반영된다.

비교에 들어가기 앞서 다음과 같은 추가적인 가정이 필요하다.

1. 재면접 조사는 모조사와 동일한 필수 조건하에서 수행된다. 즉 두 측정치 모두에 대해서 신뢰도는 같다. 이때, 우리는 $\sigma_{r1}^2=\sigma_{r2}^2=\sigma_r^2;\sigma_{s1}^2=\sigma_{s2}^2=\sigma_s^2;\delta_{21}=\delta_{22}=\delta_2;\delta_{31}=\delta_{32}=\delta_3;\delta_{41}=\delta_{42}=\delta_4;$ 라고 가정한다. 이 단순화는 만일 면접원, 감독원 교육 효과가 서로 같을때 유효하다.
2. 면접원 오차 상관관계는 사건내에서 서로 같다($\delta^2 = \beta^2$). 감독원 오차($\delta^3 = \beta^4$) 와 학교 효과($\delta^4 = \beta^5$)도 동일하게 유효하다. (Fellegi, 1964 참조).
3. $\beta^2 = \beta^3$ 이다. 즉, 두 조사 모두에서 동일한 면접원의 할당량내의 응답자 편차 간에 상관관계는 각 조사의 동일한 부 표본에 속한 단위간 상관관계와 같다는 것을 의미한다. 이 가정은 우리의 평가에서 모수적인 공간을 축소시킨다.
4. 면접원, 감독원, 교수법에 대해 할당량이 중첩구조를 가지므로, 우리는 $\delta^2 \geq \delta^3 \geq \delta^4$ 와 $\beta^2 = \beta^3 \geq \beta^4 \geq \beta^5$ 라고 가정한다.

상관 계수는 음이 아니라는 가정하에서(Fellegi, 1964 참조) 추정량의 MSE간의 차이는 정확하게 추정될 수 있으며 실험적인 맥락에서 사용하도록 제안될 수 있다. 부록에 있는 MSE에 대한 공식을 사용하여 추정량을 평가할 수 있으며 광범위한 모수값에 대해서 수로써 비교할 수 있다.

위의 가정 1-4를 사용하여 우리는 추정량의 MSE (16.8), (16.9)를 세개의 모수 ($\beta^2, \beta^4, \sigma_r^2$)를 다양하게 하면서 비교할 수 있다. 추정량 (1 MSE 를 비교하기 위해서, ρ_5는 반드시 고려되어야 한다. 일반성을 잃지 않기 위해서 우리는 $\sigma_s^2 = 1$ 로 설정한다

추정량 (16.8), (16.9), (16.10), (16.11)는 수많은 모수의 조합에 대해서 비교되어 왔다. 지면의 부족으로 우리는 단지 이 연구 결과의 대표적인 몇 가지 예제만을 소개한다. 특히, 다음의 모수 값은 고려된다. 추정량 (16.8)와 (16.9), 및 추정량 (16.10)과 (16.11)에 대해서 ρ_1 = 0.3, 0.8, ρ_2 = 0.05, 0.1, 0.2, σ_r^2 = 0.3, 1, 3, 10이며 우리는 값 ρ_4 = 0.075와 0.15를 추가했다. 이 결과는 그림 16.1부터 그림 16.5까지를 통해 확인할 수 있다.

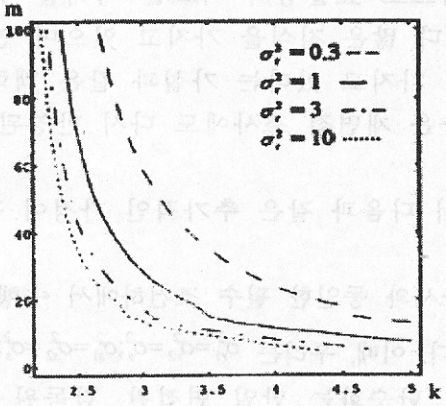


그림 16.1 σ_r^2의 각 값에대해 ρ_2 = 0.1, ρ_4 = 0.075, ρ_1 = 0.3이며 MSE(8) (MSE(9)일때 곡선

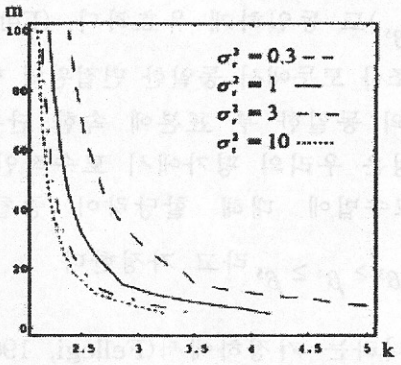


그림 16.2 σ_r^2의 각 값에대해 ρ_2 = 0.2, ρ_4 = 0.15, ρ_1 = 0.3이며 MSE(8) (MSE(9)일때 곡선

16.6 수치 연구의 결과

이장에서 우리는 수많은 모수하에서 면접원 분산의 이중 표본 추정량과 단일 표본 추정량을 비교하는 시뮬레이션 연구 결과를 설명한다. 이 값들은 대부분의 평가 조사 적용에 광범위한 관심을 불러일으켰다. 결과를 논의할때 우리는 연구 결과를 시각적으로 나타낸 그림 16.1-16.5를 참고로 할 것이다.

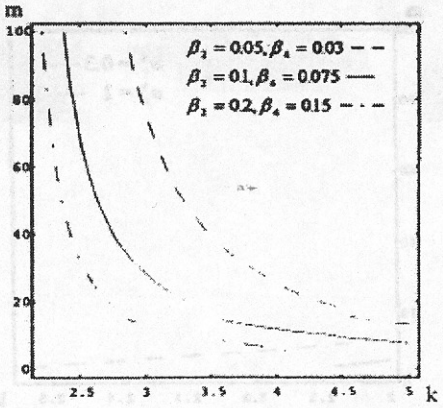


그림 16.3 γ_2, γ_4, $\sigma_r^2 = 1$, $\gamma_1 = 0.3$의 각 값에 대해 MSE(8) (MSE(9)일때 곡선

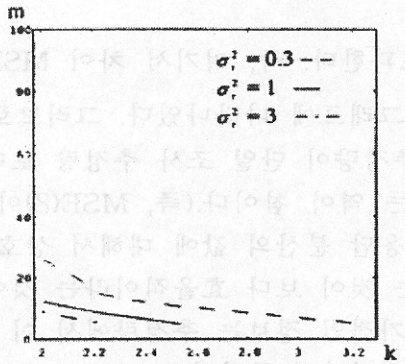


그림 16.4 σ_r^2, $\gamma_2 = 0.1$, $\gamma_4 = 0.075$, $\gamma_5 = 0.05$, $\gamma_1 = 0.3$의 각 값에 대해 MSE(10) (MSE(11)일때 곡선

그림 16.1, 16.2, 16.3에서, 면접 - 재면접 설계를 토대로하고 MSE(8)로 표기된 (16.8)의 면접원 분산 추정량의 MSE는 단일 조사 상호관입 할당량 설계를 토대로 하고 MSE(9)로 표기되는 (16.9)의 대응 추정량의 MSE와 비교된다. 층(k)당 면접원의 수, 면접원의 업무량(m), 무상관 응답 분산 σ_r^2 및 그림에 나타난 γ의 계수의 조합의 대체 값으로 우리는 한 점 MSE(9)를 계산했다.

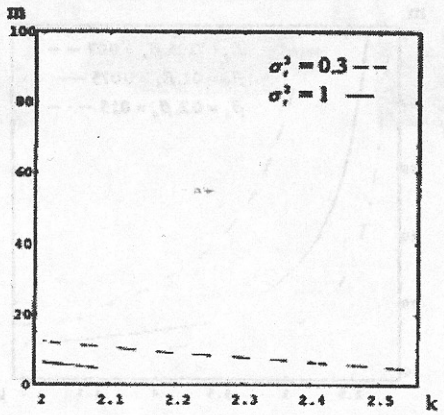


그림 16.5 σ_r^2, $\rho_2 = 0.2$, $\rho_4 = 0.15$, $\rho_5 = 0.1$, $\rho_1 = 0.3$의 각 값에 대해 MSE(10) (MSE(11)일때 곡선

이 값은 MSE(8)를 초과한다. 즉, 여기서 차이 MSE(8) - MSE(9)는 음이며 m과 k의 조합에 대해서 그래프에 나타나있다. 그러므로, 특정한 값 σ_r^2에 대응하는 곡선아래 이중 표본 추정량이 단일 조사 추정량 보다 더 작은 MSE를 가진다. 반면에 곡선 윗부분에서는 역이 참이다.(즉, MSE(8)이 MSE(9)를 초과한다). 이 그림에서 대부분의 단순 응답 분산의 값에 대해서 상호 관입 면접원 할당량을 토대로한 추정량을 사용하는 것이 보다 효율적이라는 것이 명백하다. 즉, 재면접 조사에 의해서 제공되는 추가적인 정보는 추정량에서 이 정보를 병합함으로서 추가되는 추가적인 분산을 상쇄시키지 못한다.

그림 16.4와 16.5는 감독원 효과가 있을때와 없을때 각각에 대한 재면접 조사 자료 추정량 (16.10)과 (16.11)을 비교하기 위한 유사한 곡선이다. 다시, 곡선 아래에서는 MSE(10)이 MSE(11)보다 작다. 즉, 재면접 자료를 사용한 추정량이 보다 효율적이라는 것을 나타낸다. 곡선 위에서는 재면접 조사 자료를 무시하고 상호 관입 할당량 정보만을 사용해서 감독원 효과를 추정하는것이 더 효과적이다.

$k = 2$인 경우에 대해서 평균 면접원의 업무량이 30을 초과할때(또는 더 큰 오차 상관 계수에 대해서는 20), 응답 오차 추정량의 효율성을 증가시키기위해서 재면접 조사를 수행하는 것은 더이상 효율 적이지 않다고 보여진다. 일반적으로 면접원과 감독원의 업무량은 두 가지 종류의 추정량의 상대적인 효율성에 대한 주요한 결정 요소가 된다. 상호관입 할당량을 토대로 하는 추정량 보다 면접 - 재면접 설계를 토대로한 추정량이 더 효율적이기 때문에, 평가 조사에서는 상대적으로 많은 면접원이 필요하고 면접원의 업무량도 많이 요구된다. 이는 그 만큼의 평가조사 비용으로 실제적으로 수행 하기가 어렵기 때문이다.

만일 ρ_1이 표에서 값의 집합으로 부터 다양해 진다면 곡선의 모양도 조금 변경된다. 단지 리콜 효과가 작다면(예를 들어 0.3 이하) 면접원 - 재면접원 설계를 토대로한 추정량의 상대적인 효과가 증가하는 것을 알 수 있다. 이 주요 결과에 대한 추론이 표 16.2에 설명되어 있으며 여기서 MSE(편향과 분산)의 두 성분에 대한 보다 광범위한 분석이 다음과 같은 가정하에서 수행되었다.

Case (1) : $\rho_1 > 0$ 이며 $\rho_i = 0$ (i = 2, 3, 4, 5), 즉, 우리는 응답자 리콜 효과가 양이며 재면접원은 모조사의 면접원과 다르다고 가정한다.

Case (2) : $\rho_i > 0$, (i = 1, 2, 3) 이며 $\rho_i = 0$ (i = 4, 5), Fellegi(1964)에 대해서 고찰된 설계이다. 여기서 다른 조사에서 다른 면접원과 다른 감독원간의 상관관계는 0이라고 가정한다.

Case (3) : $\rho_i > 0$ (i = 1, 2, 3, 4) 이며 $\rho_5 = 0$, 즉, 다른 조사에서 다른 감독원간의 상관관계는 0으로 가정한다.

Case (4) : $\rho_i > 0$ (i = 1, 2, 3, 4, 5), 모든 양의 상관관계의 가장 일반적인 경우

표 16.2 (1)-(4)의 경우의 추정량의 분산과 편향에 대한 연구

	Cases			
	1	2	3	4
(BIAS(8)(- (BIAS(9)(	= 0	≥0 (=0, k=2)	≥0 (=0, k=2)	
VAR(8) - VAR(9) 충분 조건	<0 $\beta^1\sigma^{r^2} \le \sigma^{s^2}$	<0 $\delta^2\sigma^{r^2} \le \sigma^{s^2}$ $2\delta^3(k-1) \ge k\delta^2$	<0 $\delta^2\sigma^{r^2} \le \sigma^{s^2}$	
(BIAS(10)(- (BIAS(11)(	= 0	>0	>0	>0
VAR(10) - VAR(11) 충분 조건	<0 $\beta^1\sigma^{r^2} \le \sigma^{s^2}$	<0 $\delta^2\sigma^{r^2} \le \sigma^{s^2}$ $\delta^3(k-1) \ge k\delta^4$ $k\delta^2 \ge 2(k-1)/\delta^3$	<0 $\delta^2\sigma^{r^2} \le \sigma^{s^2}$ $\delta^3 \ge 2\delta^4$	<0 $\delta^2\sigma^{r^2} \le \sigma^{s^2}$

이 네가지 경우에서 추정량간 비교 결과가 표 16.2에 요약되어 있다. 표 16.2의 첫번째 행에는 추정량 (16.8)의 편향의 절대값간의 차이의 방향 (BIAS(8)(과 추정량 (16.9)의 편향의 절대값(BIAS(9)(가 나타나 있다. 두 번째 행에는 세 번째 행에 주어진 충분 조건이 만족되었을때 두 추정량의 분산간의 차이의 방향을 나타낸다. 7번째 행에서는 충분 조건과 함께 4행과 5행에 (16.10)과 (16.11)에 대해 유사한 결과를 제공한다.

(1)의 경우에는 상관 조건 $\beta_1\sigma_r^2 \le \sigma_s^2$이 만족될때, 즉 리콜 오차가 표본 추출과 무상관 응답 분산사이에 비와 같거나 작을때, 반복 측정을 토대로 한 추정량 모두 상호 관입 모 조사 자료를 토대로 한 추정량보다 효율적이다.

충분 조건하에서 추정량(16.8)는 추정량 (16.9)보다 작은 분산을 갖지만 더 편향되었다. 충분조건은 다음과 같다. (2)의 경우 $\delta_2\sigma_r^2 \le \sigma_s^2, 2\delta_3(k-1) \ge k\delta_2$ 일때, (3)의 경우 $\delta^2\sigma^{r^2} \le \sigma^{s^2}$ 일때, 한 층당 두명의 면접원만 있다면 두 추정량은 동일하게 편향된다.

(1)을 제외한 모든 경우에 있어서, 반복 측정을 통한 감독원 분산 추정량 (16.10)은 반복측정을 하지 않은 대응 추정량 (16.11)보다 작은 분산을 갖는다(충분조건 하에서). 이때 충분 조건은 다음과 같다. (2)의 충분 조건 - $\delta_2\sigma_r^2 \le \sigma_s^2, \delta_3(k-1) \ge k\delta_4$, $k_{\delta^2} \ge 2(k-1)/\delta^3$ 일때 (3)의 충분 조건 - $\delta_2\sigma_r^2 \le \sigma_s^2, \delta_3 \ge 2\delta_4$ (4)의 충분 조건 - $\delta^2\sigma^{r^2} \le \sigma^{s^2}$. 일때. 그러나 이와 동일한 조건하에서 추정량 (16.10)은 추정량 (16.11)보다 편향되었다.

충분 조건이 만족될때, 면접 - 재면접 설계를 토대로한 추정량들은 모두 단일 조사를 토대로한 추정량보다 작은 분산을 갖는다. 게다가 이러한 조건은 현실적으로 만족시키기 쉽다.

추정량 (16.8)과 (16.9)에 대한 표 16.2에 있는 모수 공간의 영역은 모두 $\delta^2\sigma^{r^2} \le \sigma^{s^2}$를 만족하고, (3)의 경우에 $\delta^2\sigma^{r^2} \le \sigma^{s^2}$는 추정량 (16.8)의 분산이 추정량 (16.9)의 분산보다 작다, 충분하다는 것에 주의한다. 이와 마찬가지로, 추정량 (16.10)과 (16.11)에 대한 모수 공간의 영역은 (4)에 대한 충분한 조건을 포함하고 있으며 이는 추정량 (16.10)의 분산이 (16.11)의 분산보다 작다는 것을 의미한다. 그러므로, 단일 조사를 기반으로한 추정량보다 이중 표집을 기반으로한 추정량이 효과적이지 못한 것은 두 조사 모두의 경우에 대해서 단일 응답 분산이 증가하고 조사간 응답오차 상관관계가 증가할때 증가하는 이중 표본 추정량의 편향 때문인 것이 분명하다.

16.7 요약 및 결론

이 연구에서는 응답자, 면접원, 감독원 오차를 평가하기 위한 모형이 제안되었다. 이모형하에서 상관응답 분산의 성분에 대한 수많은 대체 추정량이 고려된다.

특히 감독원을 평가하기 위한 추정량으로 이중 상호 관입 면접-재면접 설계가 제안되었다. 제안된 추정량에 대한 분산에 대한 공식이 유도되었으며 평균 제곱 오차가 다른 저자들에 의해 제안된 추정량과 함께 비교되었다. 모수가 충분히 많은 경우는 면접 - 재면접 추정량이 모조사의 자료만을 사용한 추정량 보다 더 효율적인 경우인데, 분석적으로 그리고 시뮬레이션 연구를 통해서 연구된다.

필수적으로 모조사와 재면접 조사의 응답 편차가 독립일때 면접 - 재면접 추정량이 더 효율적이다. 그러나, 만일 단순 응답 분산이 크고 조사간 응답 오차 상관관계가 양이고 큰경우라면 하나의 조사 추정량이 더 효율적일 수도 있다. 후자의 조건은 면접원과 감독원이 모두 같을 때 일어날 가능성이 크다. 그러므로, 만일 응답 오차가 큰 경향이 있고 조사간 상관관계에 있다면, 모조사에서 감독원 또는 면접원의 효과를 추정하기 위해서 재면접 조사의 결과를 사용하는 것은 효율적이지 않다. 면접원 할당량이 크고 한 층에 두 명 이상의 면접원이 있을때, 단일 상호관입 표본 설계를 토대로 하는 추정량은 이중 상호관입 재면접조사 설계를 토대로 하는 추정량 보다 더 좋은 면접원 효과 추정량을 종종 이끌어 낸다. 마찬가지로, 이 동일한 조건에서 단일 상호 관입 조사는 감독원의 오차를 평가하기 위한 이중 관입 면접 - 재면접 조사 설계보다 효율적일 수 있다. 그러나, 상관 오차를 추정하기 위해서 이중 표본 추정량이 단일 상호관입 설계보다 더 효율적인 몇가지 경우가 있다. 여기서 우리는 면접원 할당량 크기, m의 조합과 감독원당 평균 면접원의 수 k를 확인했다. 정확히 이중 표집 추정량은 k와 m의 낮은 값의 조합에 대해서 보다 효율적이다. 단순 응답 분산이 클수록 k와 m은 낮아진다.

우리의 분석에서 우리는 각 설계하에서 자료 수집 비용은 고려하지 않았다. 물론, 재면접 조사를 수행하기 위해서는 추가적인 비용이 들지만, 단일 조사 추정량이 이중 표본 추정량 보다 작은 MSE를 가지는 경우에는 단일 조사 추정량에 큰 이점을 제공할 수 있다. 그러나, 이중 표본 추정량이 보다 효율적인 경우에 대해서는 효율성과 관련하여 재면접 조사의 비용에 대해 신중한 고려가 보장된다. 우리의 결과는 또한 복합 조사에 대해서 추정량의 효율성에 대한 주제도 더 많은 주목을 받아 마땅하다고 생각한다. Fellegi (1964)에 의해서 제안된 추정량과 유사한 복합 추정량은 편향된 병행 추정량의 조합을 토대로한 추정량은 편향을 작게 만들고 효율적인 추정량을 제공할 수도 있다.

감사의 글

이 절에서의 연구 작업은 두 저자와 함께 수행하였다. 그렇지만, 16.1절은 L. Fabbris가, 그리고 16.2절은 F. Bassi가 집필했다. 우리는 몇 동료들에게서 아이디어와 평가를 받았으며, 그 중 몇 사람은 Claudio Agostinelli, Paul Biemer, Gad Nathan, Lynne Stokes, Nicola Torelli 이다.

<참고 문헌>

Bassi, F.(1991), "Efficienza di stimatori dell'errore extracampionario,"(Efficiency of Estimators of Nonsampling Errors"). Ph. D. dissertation, Statistics Department University of Padua.

Bhat, B.R.(1962), "On the Distribution of Certain Quadraic Forms in Normal Variates," Journal of the Royal Statistical Society, Series B, 24, pp.148-151

Fabbris, L.(1991), "Abbinamento tra fonti di errore nella formazione dei dati e misure dell'effetto degli errori sulle stime,"("The Pairing of Error Sources and Measures of Estimates Variance") SIS Bulletin, 22, pp. 19-54

Fellegi, I.P.(1964). "Response Variance and Its Estimation," Journal of the American Statistical Association, 69, pp.496-501.

Hansen, M.H., Hurwitz, W.N., and Bershad, M.A.(1961), "Measurement Errors in Censuses and Surveys," 38 (2), pp.. 359-374.

Kish, K.(1962), "Studies of Interviewer Variance for Attitudinal Variables," Journal of the American Statistical Association, 57, pp. 92-115.

Mahalanobis, P.C.(1946), "Recent Experiments in Statistical Sampling in the Indian Statistical Institute," Journal of the Royal Statistical Society, 109, 99. 325-370.

<부록>

모형(16.2)하에서의 추정량의 기대값과 분산 및 정규분포 오차 성분에 대한 가정

$$E\left[s_{rh}^2\right]=\frac{\sigma_{rh1}^2+\sigma_{rh2}^2}{2}+\frac{N^*}{N^*}\frac{\sigma_{sh1}^2+\sigma_{sh2}^2-2\sigma_{sh1}\sigma_{sh2}}{2}-\frac{\sigma_{2h1}+\sigma_{rh1}^2-\sigma_{2h2}\sigma_{rh2}^2}{2}-(\beta_{1h}-\beta_{3h})\sigma_{rh1}\sigma_{rh2}$$

$$V\left[s_{rh}^2\right]=\frac{2}{k(m-1)}\left[\frac{N^*}{N^*-1}\frac{\sigma_{sh1}^2+\sigma_{sh2}^2-2\sigma_{sh1}\sigma_{sh2}}{2}+\frac{(1-\sigma_{2h1})\sigma_{rh1}^2+(1-\sigma_{2h2})\sigma_{rh2}^2}{2}-(\beta_{1h}-\beta_{3h})\sigma_{rh1}\sigma_{rh2}\right]^2$$

$$E\left[s_{rh}^2+s_{sh}^2\right]=\frac{N^*}{N^*-1}\sigma_{sh}^2+(1-\delta_{2h})\sigma_{rh}^2$$

$$V\left[s_{rh}^2+s_{sh}^2\right]=\frac{2}{k(m-1)}\left[\frac{N^*}{N^*-1}\sigma_{sh}^2+(1-\delta_{rh}^2\right]^2$$

$$E\left[\hat{\delta}_{2h}s_{rh}^2\right]=\frac{\sigma_{2h1}+\sigma_{rh1}^2-\sigma_{2h2}\sigma_{rh2}^2}{2}-\frac{\sigma_{3h1}+\sigma_{rh1}^2-\sigma_{3h2}\sigma_{rh2}^2}{2}+\sigma_{rh1}\sigma_{rh2}\frac{\beta_{2h}-(k-1)\beta_{3h}+(k-2)\beta_{4h}}{k-1}$$

$$\begin{aligned}V\left[\hat{\delta}_2 s_{rh}^2\right]=\frac{2}{m^2(k-1)}\Bigg[&\frac{N^*}{N^*-1}\frac{\sigma_{sh1}^2+\sigma_{sh2}^2-2\sigma_{sh1sh2}}{2}+\frac{(1-\sigma_{2h1})\sigma_{rh1}^2+(1-\sigma_{2h2})\sigma_{rh2}^2}{2}\\&+m\frac{(\sigma_{2h1}-\sigma_{3h1})\sigma_{rh1}^2+(\delta_{2h2}-\sigma_{3h2})\sigma_{rh2}^2}{2}\\&-\sigma_{rh1}\sigma_{rh2}\frac{(k-1)\beta_{1h}-m\beta_{2h}+(m-1)(k-1)\beta_{3h}-m(k-2)\beta_{4h}}{k-1}\Bigg]^2\\&+\frac{2}{km^2(m-1)}\Bigg[\frac{N^*}{N^*-1}\frac{\sigma_{sh1}^2+\sigma_{sh2}^2-2\sigma_{sh1}\sigma_{sh2}}{2}+\frac{(1-\sigma_{2h1})\sigma_{rh1}^2+(1-\sigma_{2h2})\sigma_{rh2}^2}{2}\\&-(\beta_{1h}-\beta_{3h})\sigma_{rh1}\sigma_{rh2}\Bigg]^2\end{aligned}$$

$$E\left[\hat{\delta}_3 s_r^2\right]=\frac{\delta_{31}+\sigma_{r1}^2+\delta_{41}\sigma_{r1}^2}{2}-\frac{\delta_{41}+\sigma_{r1}^2-\delta_{42}\sigma_{r2}^2}{2}-\sigma_{r1}\sigma_{r2}\frac{(k-1)\beta_2-\beta_4+(k-1)\beta_5}{k-1}$$

$$V\left[{}_1\hat{\delta}_3 s_{r'}^2\right] = \frac{2}{(H-1)m^2k^2}\left[\frac{N^*}{N^*-1}\frac{\sigma_{s1}^2+\sigma_{s2}^2-2\sigma_{s1}\sigma_{h2}}{2}+\frac{(1-\delta_{21})\sigma_{r1}^2+(1-\delta_{22})\sigma_{r2}^2}{2}\right.$$

$$+m\frac{(\delta_{21}-\delta_{31})\sigma_{r1}^2+(\delta_{22}-\sigma_{32})\sigma_{r2}^2}{2}$$

$$+mk\frac{(\delta_{31}-\delta_{41})\sigma_{r1}^2+(\delta_{32}-\sigma_{42})\sigma_{r2}^2}{2}$$

$$\left.-\sigma_{rh1}\sigma_{rh2}(\beta_1+\beta_2+(n-1)\beta_3+m(k-2)\beta_4-mk\beta_5\right]^2$$

$$+\frac{2}{Hk^2m^2(k-1)}\left[\frac{N^*}{N^*-1}\frac{\sigma_{s1}^2+\sigma_{s2}^2-2\sigma_{s1}\sigma_{s2}}{2}+\frac{(1-\delta_{21})\sigma_{r1}^2+(1-\delta_{22})\sigma_{r2}^2}{2}\right.$$

$$+\frac{(1-\delta_{21})\sigma_{r1}^2+(1-\delta_{22})\sigma_{r2}^2}{2}+m\frac{(\delta_{21}-\delta_{31})\sigma_{r1}^2+(\delta_{22}-\sigma_{32})\sigma_{r2}^2}{2}$$

$$\left.-\sigma_{r1}\sigma_{r2}\frac{(k-1)\beta_1-m\beta_2+(n-1)(k-1)\beta_3-m(k-2)\beta_4}{k-1}\right]^2$$

$$E\left[{}_{11}\hat{\delta}_{30}s_r^2\right] = \delta_3\delta_r^2-\delta_4\delta_r^2$$

$$V\left[{}_{11}\hat{\delta}_3 s_r^2\right] = \frac{2}{(H-1)}\left[\frac{N^*}{N^*-1}\sigma_s^2+(1-\delta_2)\sigma_r^2+m(\delta_2-\delta_3)\sigma_r^2+mk(\delta_3-\delta_4)\sigma_r^2\right]^2$$

$$+\frac{2}{Hk^2m^2(k-1)}\times\left[\frac{N^*}{N^*-1}\sigma_s^2+(1-\delta_2)\sigma_r^2+m(\delta_2-\delta_3)\sigma_r^2\right]^2$$