

통계청 학술연구용역

# 자체통계 품질진단 매뉴얼 개발 용역 보고서

수탁기관 : 한국보건사회연구원

통 계 청



통계청 학술연구용역

**자체통계 품질진단 매뉴얼 개발  
용역 보고서**

수탁기관 : 한국보건사회연구원

**통 계 청**

## 제 출 문

통계청장 귀하

이 보고서를 귀청과 용역 계약한 『자체통계 품질진단 매뉴얼개발』 연구  
과제의 최종보고서로 제출합니다.

2007년 12월 6일

주관연구기관명 한국보건사회연구원

연구원 손창균  
장영식  
최성용  
김나연

## 머 리 말

국가의 통계지표는 그 나라의 얼굴이라고 한다. 따라서 선진국일수록 정확하고 체계적인 통계지표를 생산하기 위한 노력에 투자를 아끼지 않고 있다. 우리나라도 국가위상이 높아지고 특히 1995년 OECD에 가입함으로써 선진국에 걸 맞는 통계생산의 체계화와 이를 뒷받침하는 사회조사의 엄밀성 확보가 그 어느 때보다도 중요해지고 있다.

이러한 측면에서 2005년 5월에 경제부총리 산하에 “국가통계심의위원회”가 발족하였고, 2006년 3월에 1차년도 국가통계 품질 향상을 위한 품질 진단 사업을 수행하였고, 2007년 현재 2차년도 품질진단 사업을 진행하고 있으며, 총 3개년에 걸쳐 450여종의 통계를 진단할 계획이다.

품질진단 대상 통계는 각 년도 별로 금융, 주택, 산업, 문화, 교육, 보건복지 등 관련통계별로 구분하여 수행하고 있으며, 이에 조사통계, 보고통계, 가공통계를 모두 포괄하고 있다. 국가통계의 품질의 중요성은 국가 정책수립의 다방면에 걸쳐 영향을 끼치고 있으며, 따라서 통계생산과정에서의 오류는 곧바로 정책 수립의 오류로 나타나기 때문에 통계생산과정의 엄밀성을 아무리 강조해도 지나치지 않을 것이다.

각 통계 작성 기관은 반드시 분야별 통계의 품질진단을 받도록 함으로서 향후 해당 통계의 품질을 제고하고, 관련 기관의 통계작성 부서의 통계에 대한 마인드를 향상시킬 수 있는 좋은 계기가 되고 있다. 또한 각 부서별로 분산되어 작성되고 있는 다양한 통계에 대한 품질 점검을 통해 통계의 중복성을 체크하고, 해당 통계의 개선점을 도출하고, 통계 이용자의 저변을 확대함으로써 국가 통계의 전반적인 품질향상에 기여하도록 하고 있다.

이와 같은 전 국가적인 통계품질 진단 사업을 수행함으로써 전반적인 통계품질 향상에 기여한 측면도 있지만, 연차적으로 수행되는 품질 진단 사업은 매우 방대한 영역에 걸쳐 사업을 수행함으로써 품질진단 관련 대상자들의 업무 폭증과 통계생산

기관의 타율적인 품질진단으로 인한 거부감과 부서 이기주의 등으로 인하여 점차적으로 진단 사업의 노력대비 효율성이 감소하고 있는 실정이다.

따라서 보다 효율적으로 통계생산기관에서 작성되는 통계에 대한 자체 품질 진단을 자율적으로 수행케 하여 예산의 절약과 품질진단의 엄밀화를 이룰 수 있도록 하고, 통계생산부서의 해당 통계에 대한 중요성을 인식시킴으로서 전체적인 국가통계의 품질 향상을 이룰 수 있도록 하였다.

이를 위해 유럽연합(EU)에서 개발하여 운영하고 있는 “자체품질진단서(DESAP)”을 기반으로 국내 통계상황에 맞도록 재구성하여 조사통계, 보고통계, 가공통계에 대한 자체품질진단이 이루어 질 수 있도록 하였다. 유럽연합의 DESAP은 조사통계와 보고통계에 대한 품질진단이 가능하며, 본 매뉴얼에서는 이와 더불어 가공통계도 함께 포괄하고 있다는 측면에서 적용에 의미가 있다고 할 수 있다.

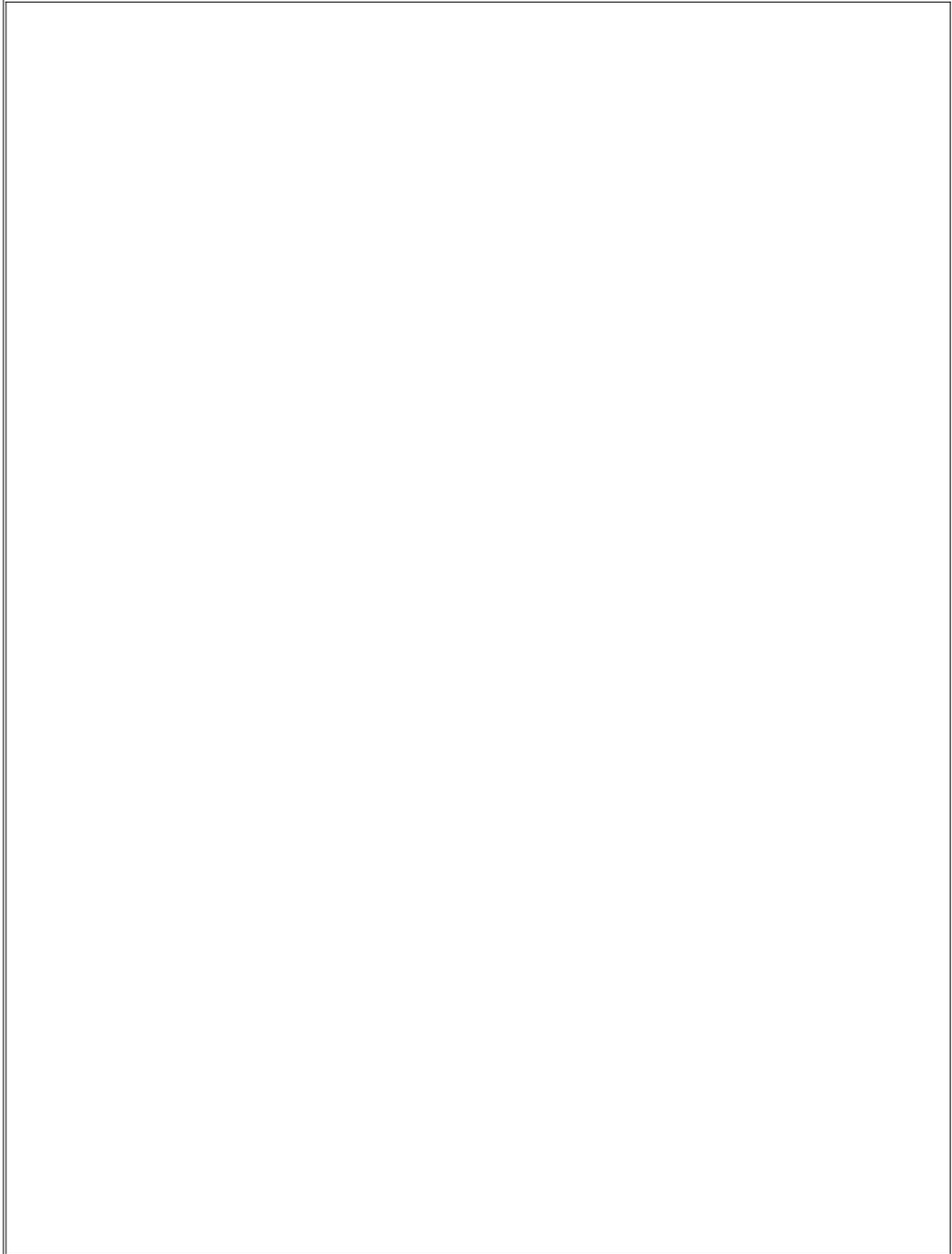
본 연구는 보건사회연구원의 손창균 부연구위원의 책임 하에 장영식 연구위원과 최성용연구원, 김나연 선임연구원이 공동연구로서 국내통계생산 실정에 맞도록 자체 품질진단 매뉴얼을 개발하였다. 끝으로 본 보고서를 검토해준 고경환 연구위원과 도세록 연구위원에게 감사의 마음을 전한다.

2007년 12월  
한국보건사회연구원  
원 장 김 용 문

## 목 차

|                              |    |
|------------------------------|----|
| ABSTRACT .....               | 13 |
| 국 문 요 약 .....                | 17 |
| <br>                         |    |
| 제1장 서 론 .....                | 23 |
| 제1절 연구배경 및 목적 .....          | 23 |
| 제2절 연구 범위 .....              | 25 |
| 제3절 연구 방법 .....              | 28 |
| <br>                         |    |
| 제2장 품질진단 방법의 분류 .....        | 29 |
| 제1절 품질진단 방법의 분류 .....        | 29 |
| 제2절 품질 점검표 .....             | 31 |
| 제3절 품질감사 .....               | 35 |
| 제4절 자체품질진단 .....             | 49 |
| <br>                         |    |
| 제3장 주요 품질진단 차원 및 지표 .....    | 55 |
| 제1절 외국의 주요품질진단 차원 및 지표 ..... | 55 |
| 제2절 관련성 .....                | 56 |
| 제3절 정확성 .....                | 61 |
| 제4절 시의성/정시성 .....            | 84 |
| 제5절 접근성/명확성 .....            | 88 |

|                                     |     |
|-------------------------------------|-----|
| 제6절 비교가능성 .....                     | 92  |
| 제7절 일관성 .....                       | 100 |
| 제8절 비용과 응답부담 .....                  | 109 |
| <br>                                |     |
| 제4장 자체통계 품질진단을 위한 가이드라인 .....       | 113 |
| 제1절 조사실시의 결정 .....                  | 113 |
| 제2절 조사 설계 .....                     | 117 |
| 제3절 자료수집 .....                      | 138 |
| 제4절 자료입력 및 자료처리 .....               | 151 |
| 제5절 자료 분석 및 통계결과의 품질 .....          | 162 |
| 제6절 문서화 및 자료의 공표 .....              | 175 |
| 제7절 개선 사이클 .....                    | 180 |
| <br>                                |     |
| 제5장 자체통계 품질진단서 작성을 위한 이용자 가이드 ..... | 186 |
| 제1절 작성 지침 .....                     | 187 |
| 제2절 자체통계 품질진단 사례 .....              | 311 |
| 제3절 자체통계 품질진단 결과의 활용 .....          | 367 |
| <br>                                |     |
| 제6장 정책적 제언 .....                    | 369 |
| <br>                                |     |
| 참고문헌 .....                          | 373 |
| <br>                                |     |
| [부 록 1] 자체통계 품질진단서 .....            | 375 |
| [부 록 2] 주요 용어 .....                 | 435 |



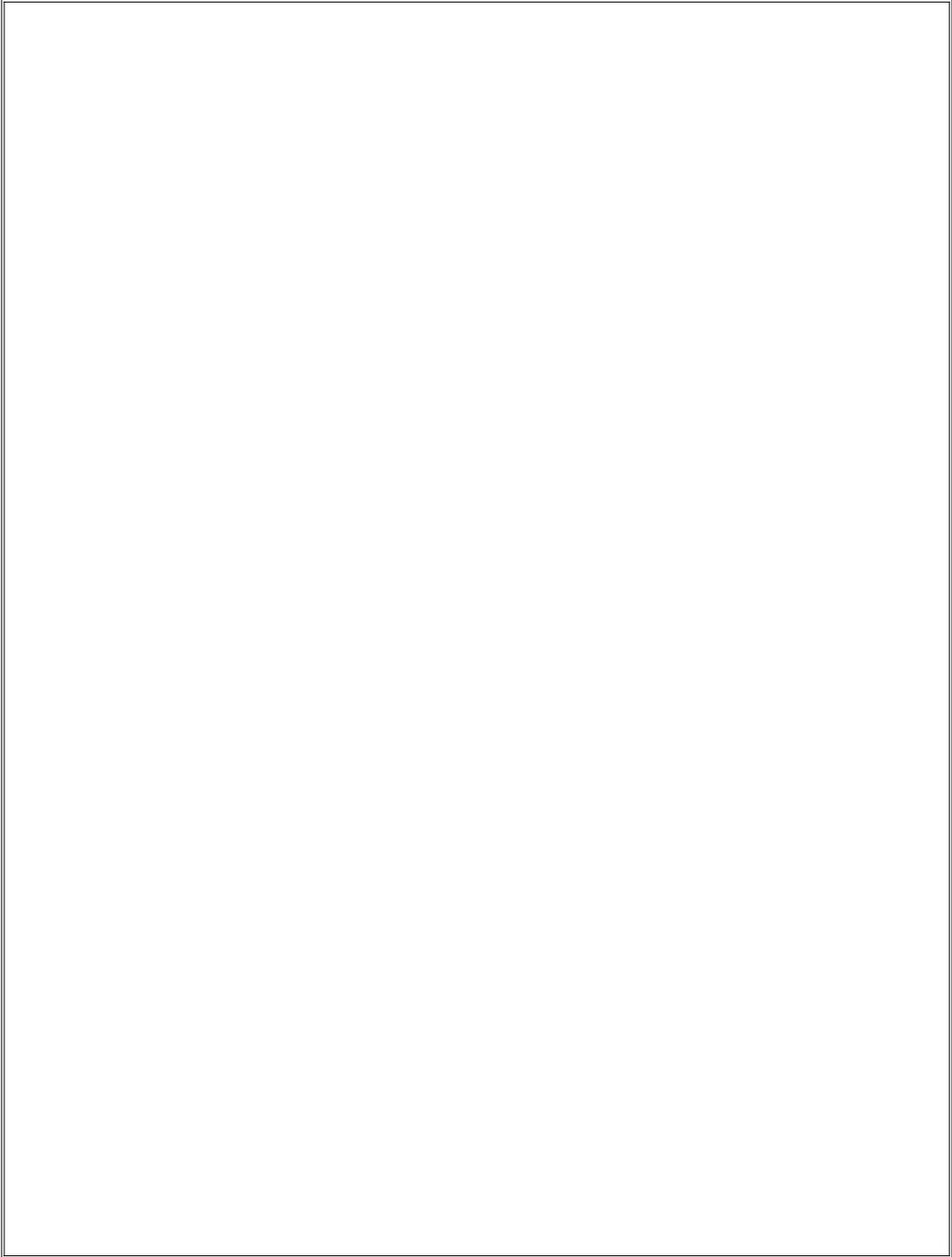
## 표 목 차

|  |     |
|--|-----|
| 〈표 2-3-1〉 감사의 유형 .....   | 37  |
| 〈표 2-4-1〉 통계품질진단의 법적조항 .....   | 49  |
| 〈표 2-4-2〉 자체통계품질진단서의 구성항목 .....  | 52  |
| 〈표 3-1-1〉 외국의 표준 품질관리지표 .....  | 55  |
| 〈표 3-3-1〉 제조업, 가스 수도사업, 건설업에 대한 변동계수의 통계값 .....  | 63  |
| 〈표 3-3-2〉 무역업에 대한 변동계수의 통계값 .....  | 64  |
| 〈표 3-3-3〉 제조업의 중분류 IoP 에 대한 표준오차의 추정치 .....  | 64  |
| 〈표 3-3-4〉 인종별 순수 과소계상의 백분율의 인구통계학적 분석 추정치 .....  | 67  |
| 〈표 3-3-5〉 응답의 일관성 .....  | 73  |
| 〈표 3-3-6〉 센서스와 1982년 취업자조사 .....   | 75  |
| 〈표 3-3-7〉 1997년 센서스 시험조사와 반자동 코딩절차와 수작업 코딩절차 .....   | 76  |
| 〈표 3-3-8〉 과수농가조사에서 무응답 분석 .....  | 79  |
| 〈표 3-4-1〉 주요 통화지수의 이용성 .....   | 87  |
| 〈표 3-6-1〉 비교가능성의 예 .....   | 93  |
| 〈표 3-7-1〉 1960년 보험수가의 총 할증료에 대한 잠정치와 확정치간의 비교 .....  | 102 |
| 〈표 3-7-2〉 고정된 가격에서 IoP 대비 가중 총 생산과 가중되지 않은 총생산<br>그리고 가중 총부가가치와 가중되지 않은 총부가가치의 1995년과 1996<br>년 성장률 비교 ..... | 104 |
| 〈표 3-7-3〉 1993년과 1994년에 ECHP의 1,2차 웨이브와 가계조사(EBF94/95)로부터<br>추정된 구성별 프랑스 총수익의 분포 .....                       | 105 |
| 〈표 3-7-4〉 고정된 가격에서 STPVI와 국민계정(NA)에 따른 1994년부터 1995년 까<br>지 분기별 변동 간 비교 .....                                | 107 |
| 〈표 4-3-1〉 무응답 유형과 처리방법 .....   | 147 |

|  |     |
|--|-----|
| 〈표 5-2-1〉 국민건강영양 조사의 구성 .....          | 313 |
| 〈표 5-2-2〉 국민건강영양조사의 품질차원별 점수 .....     | 332 |
| 〈표 5-2-3〉 부동산 거래 현황 통계의 품질차원별 점수 ..... | 349 |
| 〈표 5-2-4〉 공장등록 현황 통계의 품질차원별 점수 .....   | 365 |

## 그림 목 차

|            |                                   |     |
|------------|-----------------------------------|-----|
| [그림 2-4-1] | 자체진단모형의 구성 .....                  | 51  |
| [그림 3-2-1] | 영국의 공식통계 이용자 분류 .....             | 58  |
| [그림 3-2-2] | ACSI 모형 .....                     | 60  |
| [그림 4-3-1] | 코더의 이해력과 판단의 의무 .....             | 152 |
| [그림 4-4-1] | 종이조사표를 이용한 조사에서 자료처리 절차의 흐름 ..... | 153 |
| [그림 4-4-2] | 컴퓨터를 이용한 조사에서 자료처리 절차의 흐름 .....   | 154 |
| [그림 5-2-1] | 국민건강영양조사의 주요이용자 순위 .....          | 314 |
| [그림 5-2-2] | 국민건강영양조사의 품질진단 다이어그램 .....        | 333 |
| [그림 5-2-3] | 부동산 거래 현황 통계의 주요이용자 순위 .....      | 336 |
| [그림 5-2-4] | 부동산 거래 현황 통계의 품질진단 다이어그램 .....    | 350 |
| [그림 5-2-5] | 공장 등록 현황 통계의 주요이용자 순위 .....       | 353 |
| [그림 5-2-6] | 공장 등록 현황 통계의 품질진단 다이어그램 .....     | 366 |
| [그림 5-3-1] | 진단수행과 결과 도출 및 환류에 대한 흐름도 .....    | 367 |



## ABSTRACT

### 1. Background of Study

□ On basis of European countries, the concern about statistical quality management concentrated during the last 10 years, this endeavor of each country for statistical quality improvement makes better many fields of statistics

— Especially, European countries established LEG(Leadership Expert Group) in 1999, they implemented the work for improvement of statistical quality of ESS(European Statistical System).

□ The quality of statistical data includes the variance and bias as well as accuracy, timeless and relevance.

— As increasing as importance of statistical quality management, Korea revised the rule of statistics including the self assessment of statistical quality in 2006.

□ KNSO make a plan to operate the regular quality assessment for the official statistics 481 during 3 years (2006~2008), and then it operated the quality assessment for the sort of 107 official statistics in 2006, and 180 official statistics in 2007.

This kind of the regular quality assessment spends a lot of cost and time, especially KNSO inputs too much human resources and materials in one time to decrease the effectiveness with respect to resource management. In point of view of the official statistics producer, this work effects negatively.

— This study develops the self assessment program, users manual and guideline of statistical quality for the official statistics agencies based on DESAP of Eurostat,

## 2. Objectives

This study of objectives is the development of the self assessment manual and guideline for the official statistics agencies.

To do this, the manual and guideline based on DESAP of EU reconstructs to adequate for the statistical environment in Korea.

We develop the manual and guideline to make self assessment in statistical agencies, and make to provide the comparable results of quality assessment with the foreign countries, and desire to construct the system of statistics production and quality management.

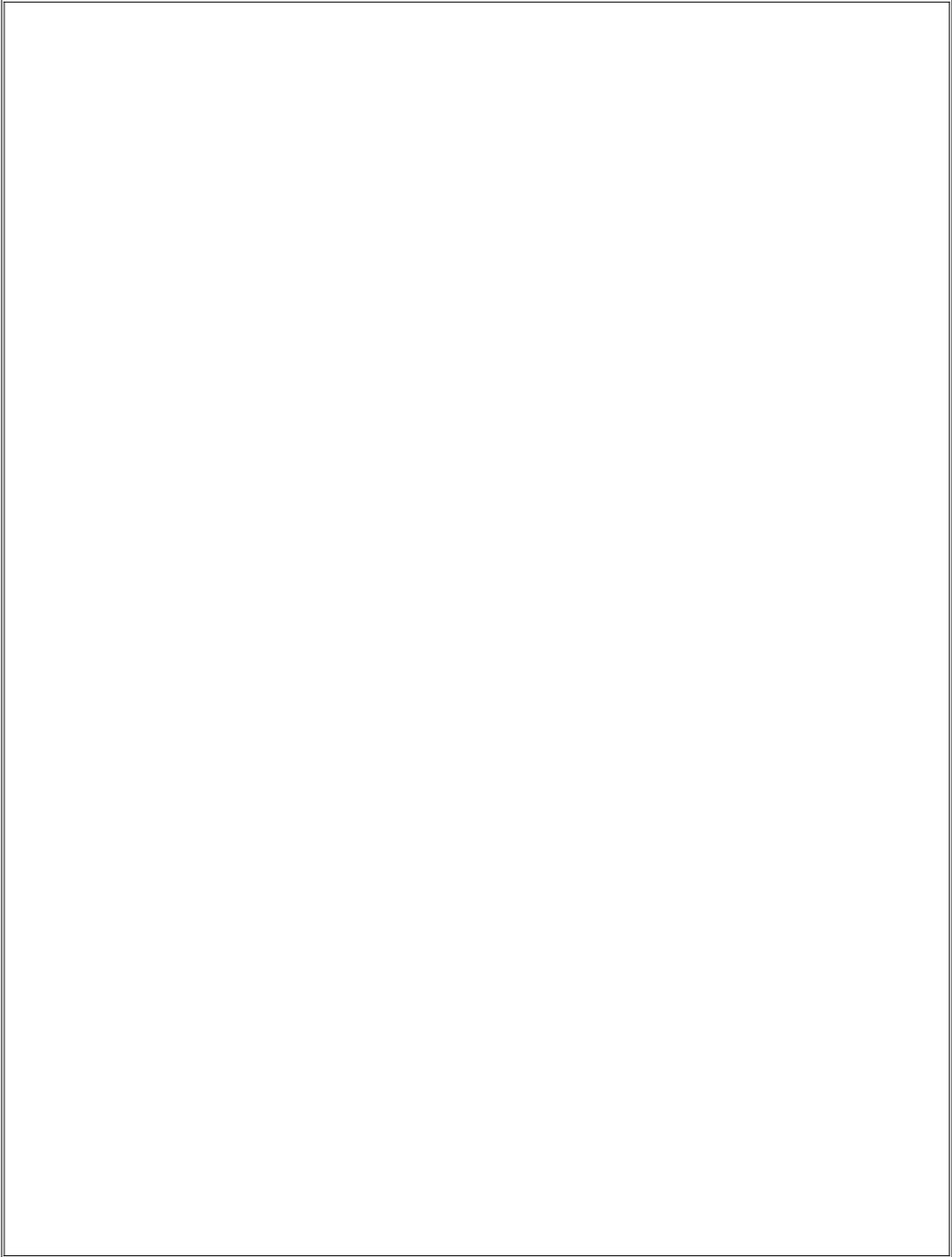
## 3. Methodology of Study

First, we analyze more detail DESAP developed by EU, basically we include the descriptions of each assessment items.

And then we analyze the results of the regular quality assessment in last 2 years conducted by KNSO, we provide the self assessment manual and guideline to help for making the quality report by the survey managers in the statistical agencies.

In case of self assessment for the secondary statistics, it is based on the regular quality assessment items conducted by KNSO and we exam the comparable items from the foreign cases.

We provide the self assessment report for three kinds of the official statistics based on the developed check list, manual and guideline.



## 국문요약

### 1. 연구 배경

□ 지난 10여 년간 유럽 국가를 중심으로 통계의 품질에 대한 관심이 집중되었고, 이러한 국가별 품질개선에 대한 노력으로 많은 부분에 걸쳐 통계의 품질이 개선되어 왔음.

- 특히 1999년도에 품질에 대한 LEG(Leadership Expert Group)을 창설하여 ESS(European Statistical System)의 품질을 개선하기 위한 작업을 수행하였음.
- LEG는 ESS에 대한 권고안을 작성하여 유럽각국의 국가통계에 대한 품질개선을 위한 지침으로 활용하고 있음.

□ 통계자료에 대한 품질은 분산(variance)과 편향(bias)과 같은 통계 품질상의 개념 뿐만 아니라, 정확하게 측정하고 시기적절하게 이용할 수 있는 관련성의 의미도 포괄하는 개념임.

- 보편적으로 이용 적합성과 관련하여 다루어야 할 것으로 캐나다 통계청에서는 관련성(relevance), 정확성(accuracy), 시의성(timeliness), 접근가능성(accessibility), 해석가능성(interpretability) 및 일관성(coherence)으로 요약하고 있음.
- 우리나라에서도 통계품질관리의 중요성이 점차 커지면서 개정된 통계법에 통계 품질 개선을 위하여 자체품질진단의 실시를 법률로서 규정하고 있음.

□ 통계품질관리에 대한 필요성이 점차 높아짐에 따라 통계청에서는 전 국가통계에 대한 총체적인 품질상태 확인을 위해 국가통계전반에 대해 품질진단을 3개년(2006~2008년)에 걸쳐 총 451종의 통계에 대해 통계를 진단하기로 결정하고 2006년도는 전체 대상 중 107종에 대한 품질진단을 실시하였고, 2007년 현재 180여종의 통계에 대한 품질진단을 수행하고 있음.

□ 정기적인 외부 품질진단사업은 비용과 시간이 많이 소요되며, 품질진단을 수행하는 통계청은 많은 인력이 일시에 투입됨으로서 행정력의 과대 투입으로 인해 업무의 효율성이 결여되는 측면이 있고, 진단을 받는 기관은 일시에 많은 종류의 통계에 대한 품질진단을 수행함으로써 업무의 과대로 인한 진단의 필요성에 대한 부정적인 영향을 끼침.

— 보다 개선된 형태의 소관통계에 대한 품질진단을 수행할 수 있고, 자발적인 참여를 유도함으로써 개선효과를 극대화 할 수 있는 방안으로 소관통계에 대한 자체품질진단을 수행할 수 있도록 함.

## 2. 연구목적

□ 본 연구의 목적은 통계생산 기관에서 작성하는 각종 통계에 대한 품질을 점검해 볼 수 있는 진단서 및 매뉴얼을 개발하는데 있음.

□ 이를 위해 유럽연합(EU)에서 개발하여 운용하고 있는 자체품질진단서(DESAP)를 기반으로 국내 통계실정에 부합하도록 자체진단서를 재구성하고, 진단서 작성에 필요한 매뉴얼을 개발하고자 함.

— 통계작성기관에서 소관통계에 대한 자체진단이 이루어질 수 있도록 매뉴얼과 이용자 지침서를 개발하고, EU에서 운용중인 DESAP을 적용함으로써 향후 외국의 진단결과와 비교가 가능하고, 이를 통해 선진국형 통계생산 및 품질관리

시스템 구축에 반영하고자 함.

□ DESAP의 진단 항목은 필수적으로 조사통계와 보고통계 및 가공통계의 점검사항이 포함되어있으며, 6가지 품질차원에 대해 품질 수준을 평가할 수 있는 주요항목이 내포되어 있으므로 정기진단의 결과와 비교가 가능하도록 구성되어 있음.

□ 별도의 진단 체계를 개발하기 보다는 선진국에서 운용하는 포괄적인 시스템을 국내 환경에 적용함으로써 국내외 유사통계 간 통계품질 진단의 결과에 대한 국제 비교가 가능하도록 하였음.

### 3. 연구범위

□ 본 연구는 통계 품질진단에 대한 국내외 선행연구 및 국내 품질관리 현황, 캐나다의 품질관리 동향 등을 통하여 통계 품질관리를 위한 진단서 및 매뉴얼을 제시하고자 하며, 이를 위하여 다각적인 방법이 시도하였음.

□ 또한 관련 전문가의 의견을 수렴하여 연구의 질을 제고하고자 하는 노력을 하였으며, 작성된 진단서를 통계 담당자에게 작성토록 하여 그 난이도와 적합성을 점검하였음.

□ 통계품질을 결정하는 요소는 학자에 따라 여러 가지로 제시하고 있지만, 현재 국내에서 통계품질의 척도로 삼고 있는 7가지 요소로는 다음과 같음.

- 관련성(relevance), 정확성(accuracy), 시의성/정시성(timeliness/punctuality), 접근성/편리성(accessibility/convenience), 비교성/일관성(comparability/coherence), 효율성(efficiency), 서비스성/해석성(serviceability/interpretability)

□ 본 연구에서는 이러한 7가지 품질진단 척도를 감안하여, DESAP의 자체품질진단서의 내용에 포함된 각 항목별 작성 매뉴얼을 개발하고, 이를 위해 관련 용어에 대한 정의와 해설을 부가하며, 진단서의 작성 순서에 따라 내용을 전개하였음.

□ 통계 생산기관에서 작성하는 통계는 조사통계, 보고통계, 가공통계로 구분되며, 가공통계는 조사통계나 보고통계와는 다른 성격을 띠며, 자료의 수집, 처리, 분석에 관한 진단항목이 있기 때문에 별도의 신규 진단 항목의 개발 없이 충분히 포괄할 수 있으므로, 본 매뉴얼에서는 유럽연합(EU)에서 개발한 DESAP를 기반으로 각 분야별로 통계생산자 또는 통계생산부서의 직원이 자체품질진단서를 보다 용이하게 작성할 수 있도록 하는데 주안점을 두어 매뉴얼을 개발하였음.

#### 4. 연구방법

□ 유럽연합(EU)에서 개발한 DESAP을 보다 면밀히 분석하여 기본적으로 국제비교가 가능한 항목별로 상세한 설명을 매뉴얼에 첨가하였고, 국내 통계작성 환경에 맞지 않는 항목은 삭제하여 일관성을 유지하도록 하였음.

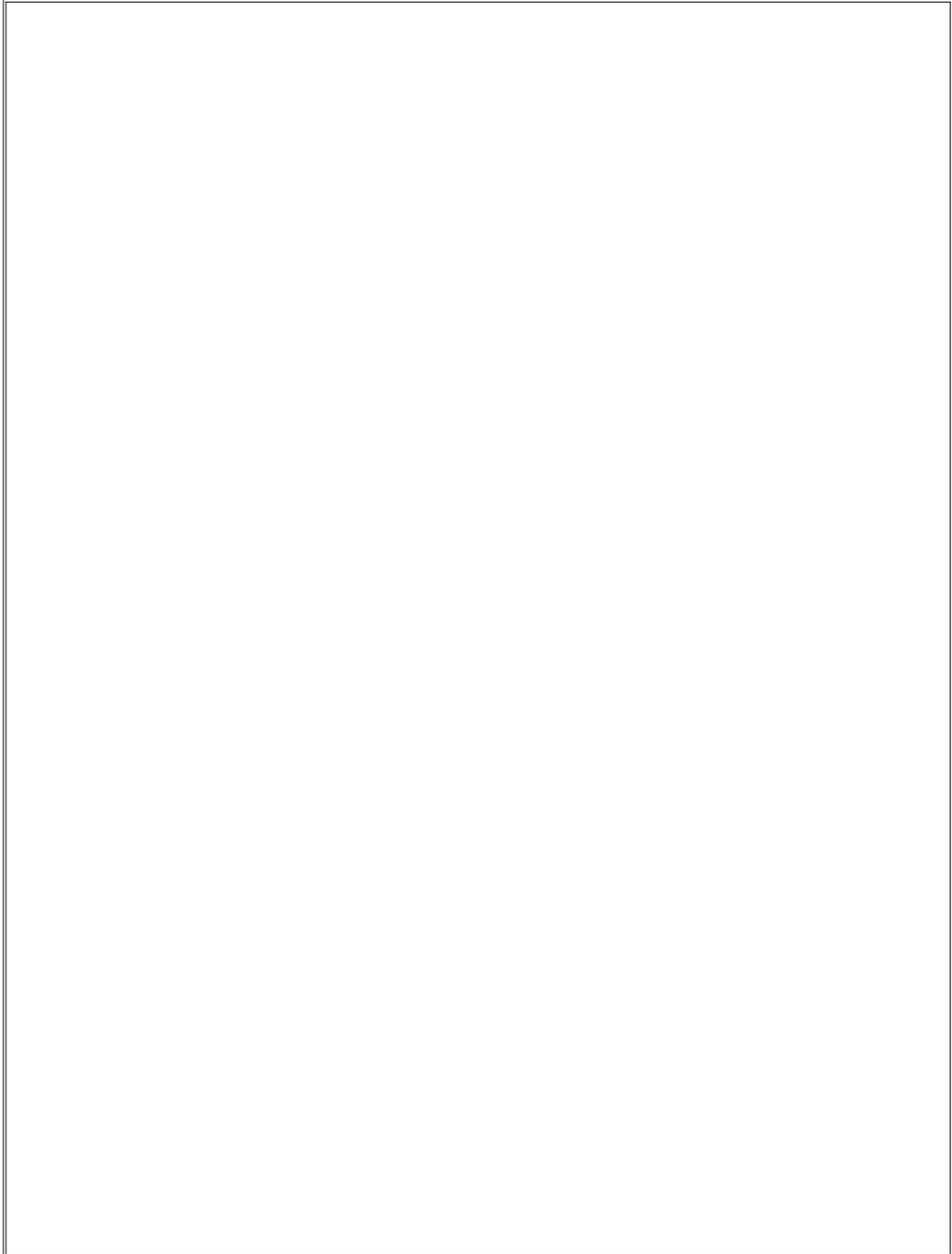
□ 국내에서 2개년 간 수행되어온 품질진단서 및 진단 결과를 분석하여 반드시 점검되어야 할 사항을 자체품질진단서에 포함하였고, 외국의 자체품질진단 사례를 분석하여 이를 품질보고서에 첨가하도록 하였음.

□ 또한 OECD에서 운용한 품질진단의 내용을 분석하여 국내통계작성환경에 맞는 사항을 첨가하도록 함으로서 점검사항의 결과 분석에 비교가능 하도록 하였음.

□ 가공통계에 대한 자체진단서의 경우 국내에서 운용한 외부품질진단서를 기반으로 DESAP의 내용에 부합하는 항목을 준용하였고, 이에 대한 외국의 사례 또한 비

교 가능한 항목을 점검토록 하였음.

이와 더불어 통계작성기관에 개발된 진단서와 매뉴얼을 우편으로 발송하여 개발된 진단서 및 매뉴얼의 활용가능성을 평가함으로서 작성기관의 통계담당자의 관점에서 진단서 작성의 용이성을 향상하도록 하였음.



# 제1장 서론

## 제1절 연구배경 및 목적

### 1. 연구 배경

지난 10여 년간 유럽 국가를 중심으로 통계의 품질에 대한 관심이 집중되었고, 이러한 국가별 품질개선에 대한 노력으로 많은 부분에 걸쳐 통계의 품질이 개선되어 왔다. 특히 1999년도에 품질에 대한 LEG(Leadership Expert Group)을 창설하여 ESS(European Statistical System)의 품질을 개선하기 위한 작업을 수행하였다. LEG는 ESS에 대한 권고안을 작성하여 유럽각국의 국가통계에 대한 품질개선을 위한 지침으로 활용하였다.

통계는 국가의 정책을 수립하는데 있어서 가장 기초가 되는 자료일 뿐 아니라 관련분야의 통계에 대한 수요 또한 점차 증가하고 있으며, 이에 따라 신속하고 정확하게 다양한 형태로 제공이 요구되고 있다. 이와 같은 요구에 부합한 자료를 제공하기 위해서는 이용 적합성이 높아야 한다. 이를 위해 통계품질관리를 위한 노력이 요구되고 있다.

통계자료에 대한 품질은 분산(variance)과 편향(bias)과 같은 통계 품질상의 개념뿐만 아니라, 정확하게 측정하고 시기적절하게 이용할 수 있는 관련성의 의미도 포괄하는 개념이다. 보편적으로 이용 적합성과 관련하여 다루어야 할 것으로 캐나다 통계청에서는 관련성(relevance), 정확성(accuracy), 시의성(timeliness), 접근가능성(accessibility), 해석가능성(interpretability) 및 일관성(coherence)으로 요약하고 있다.

우리나라에서도 통계품질관리의 중요성이 점차 커지면서 개정된 통계법에 통계품질 개선을 위하여 자체품질진단의 실시를 법률로서 규정하고 있다.

통계청에서는 정확한 통계는 국가경영에 필수적인 인프라이며, 통계의 왜곡은 곧 정책의 왜곡이 되고, 정보화 사회는 정확한 통계에서 시작되며, 정확한 통계 데이터 베이스를 구축할 필요성이 대두되고 있고, 통계는 이미 지방자치 시대의 필수 요소가 되고 있기 때문에 통계품질관리의 필요성을 매우 높다고 보고 있다.

이와 같이 통계품질관리에 대한 필요성이 점차 높아짐에 따라 통계청에서는 전국가통계에 대한 총체적인 품질상태 확인을 위해 국가통계전반에 대해 3개년(2006~2008년)에 걸쳐 총 451종의 통계에 대해 통계를 진단하기로 결정하고 2006년도는 전체 대상 중 107종에 대한 품질진단을 실시하였고, 2007년 현재 180여종의 통계에 대한 품질진단을 수행하고 있다.

그러나 정기적인 외부 품질진단사업은 비용과 시간이 많이 소요되며, 품질진단을 수행하는 통계청은 많은 인력이 일시에 투입됨으로서 행정력의 과대 투입으로 인해 업무의 효율성이 결여되는 측면이 있고, 진단을 받는 기관은 일시에 많은 종류의 통계에 대한 품질진단을 수행함으로써 업무의 과대로 인한 진단의 필요성에 대한 부정적인 영향을 미치게 된다. 따라서 보다 개선된 형태의 소관통계에 대한 품질진단을 수행할 수 있고, 자발적인 참여를 유도함으로써 개선효과를 극대화 할 수 있는 방안으로 소관통계에 대한 자체품질진단을 수행할 수 있도록 하였다.

## 2. 연구목적

본 연구의 목적은 통계생산 기관에서 작성하는 각종 통계에 대한 품질을 점검해 볼 수 있는 진단서 및 매뉴얼을 개발하는데 있다. 이를 위해 유럽연합(EU)에서 개발하여 운영하고 있는 자체품질진단서(DESAP)를 기반으로 국내 통계실정에 부합하도록 자체진단서를 재구성하고, 진단서 작성에 필요한 매뉴얼을 개발하고자 한다. 통계작성기관에서 소관통계에 대한 자체진단이 이루어질 수 있도록 매뉴얼과 이용자 지침서를 개발하고, EU에서 운용중인 DESAP을 적용함으로써 향후 외국의 진단결과와 비교가 가능하고, 이를 통해 선진국형 통계생산 및 품질관리 시스템 구축에 반영하고자 한다.

DESAP의 진단 항목은 필수적으로 조사통계와 보고통계 및 가공통계의 점검사항이 포함되어있으며, 6가지 품질차원에 대해 품질 수준을 평가할 수 있는 주요항목이 내포되어 있으므로 정기진단의 결과와 비교가 가능하도록 구성되어 있다.

따라서 별도의 진단 체계를 개발하기 보다는 선진국에서 운용하는 포괄적인 시스템을 국내 환경에 적용함으로써 국내외 유사통계 간 통계품질 진단의 결과에 대한 국제비교가 가능하도록 하고자 한다.

최근 들어 모든 분야에서 품질의 중요성이 점차 강조되고 있으며, 통계분야에 있어서도 예외는 아니어서 통계청에서는 각 부처에서 생산되는 통계에 대한 진단을 실시하였다. 그러나 외부진단에 의한 개선점 모색도 중요하지만 통계생산기관 자체에서 지속적으로 진단을 실시하고 취약한 부문에 대한 대책을 마련하여 개선하여 나간다면 보다 신뢰성 있는 통계생산 및 품질관리 체계가 마련될 것이다.

## 제2절 연구범위

### 1. 연구범위

본 연구는 통계 품질진단에 대한 국내외 선행연구 및 국내 품질관리 현황, 캐나다의 품질관리 동향 등을 통하여 통계 품질관리를 위한 진단서 및 매뉴얼을 제시하고자 하며, 이를 위하여 다각적인 방법이 시도되었다.

또한 관련 전문가의 의견을 수렴하여 연구의 질을 제고하고자 하는 노력을 하였으며, 작성된 진단서를 통계 담당자에게 작성토록 하여 그 난이도와 적합성을 점검하였다.

통계품질을 결정하는 요소는 학자에 따라 여러 가지로 제시하고 있지만, 현재 국내에서 통계품질의 척도로 삼고 있는 것은 다음의 7가지 요소이다.

관련성(relevance)은 데이터가 공헌하는 가치를 질적으로 평가하는 것이다. 이러한 가치는 데이터나 정보가 생산된 목적이나 이용자가 추구하는 목적을 어느 정도 충

족시키는데에 의하여 평가된다. 더 나아가서, 가치는 데이터나 정보 생산의 기회비용, 법적 요구사항, 기관의 수입업무 등의 관점에서 이러한 목적들이 공헌하는 정도에 의해서도 평가된다.

정확성(accuracy)은 통계활동이 측정하고자 하는 양과 특성을 얼마나 정확하게 측정하는가를 의미한다. 정확성은 여러 가지 속성을 가지고 있으며, 실무적인 측면에서 보면, 정확도 측정은 대개 오차 또는 오차 유의도 측면에서 측정되고 기술되며, 오차의 주요 원천들, 예컨대 표본추출, 무응답, 자료처리 등을 통하여 나타난다.

시의성/정시성(timeliness/punctuality)은 특정사건이나 현상의 발생시점과 그 정보를 이용할 수 있는 시점간의 시차를 반영하는 것이다. 하지만, 정보가 가치 있는 것으로 인정되는 기간과 아직 사건이나 현상이 진행 중인 것으로 인정되는 기간 등을 고려하여 판단하여야 한다. 시의성은 일반적으로 신뢰성과 상호 결합되는 관계에 있다.

접근성/편리성(accessibility/convenience)은 통계기관이 소유한 정보의 이용가능성을 반영한다. 또한 이용차원에서 본 정보포맷의 적합성, 공개매체, 메타데이터(meta-data)의 이용가능성, 그리고 이용자가 이용가능성과 입수경로를 알 기회를 가지는지 등이 고려된다. 또한 특정 정보가 그 정보에 가치를 느끼는 이용자에게 제대로 제공되는가도 이러한 특성의 한 측면이 된다.

비교성/일관성(comparability/coherence)은 하나의 통계프로그램에서 작성된 데이터와 정보 간에, 그리고 각기 다른 통계프로그램에서 작성된 데이터나 다른 데이터 세트로부터 수집된 데이터들이 서로 논리적으로 연결되고 완결되는 정도를 나타낸다. 완전한 일관성을 가진 데이터란, 내부적으로, 시계열상으로, 그리고 결과치와 프로그램 간에 논리적인 일관성이 존재하는 것을 의미한다. 이용되는 개념과 모집단은 다른 유사 통계프로그램들의 개념과 모집단 또는 일반적으로 사용되는 개념이나 용어와 논리적으로 구별 가능하여야 한다.

효율성(efficiency)은 통계작성기관이 통계자료를 얻는데 사용된 비용에 비하여 통계자료가 어느 정도 효율적으로 생산되었는가를 평가하는 요소이다.

서비스성/해석성(serviceability/interpretability)은 이용자가 이들을 쉽게 이해하고

이용하고 분석할 수 있는가를 나타낸다. 데이터와 관련된 개념, 모집단, 변수, 용어 등에 대한 적절한 정의, 그리고 데이터가 지니고 있는 한계에 대한 제반정보 등에 의해 해석가능성의 정도가 결정된다.

본 연구에서는 이러한 7가지 품질진단 척도를 감안하여, DESAP의 자체품질진단서의 내용에 포함된 각 항목별 작성 매뉴얼을 개발하고자 한다. 이를 위해 관련 용어에 대한 정의와 해설을 부가하여 진단서의 작성 순서에 따라 내용을 전개하였다.

통계 생산기관에서 작성하는 통계는 조사통계, 보고통계, 가공통계로 구분된다. 조사통계는 전수조사와 표본조사 등으로 구분되는데 다음과 같이 정리할 수 있다.

전수조사통계는 조사대상 전체를 대상으로 하는 통계로 국민 한 사람 한 사람, 사업체 하나하나까지 정확하게 파악해야 할 필요가 있는 조사의 경우에 해당된다.

표본조사통계는 모집단을 대표하는 표본을 추출한 후 적절한 조사방법에 의해 조사를 실시하여 데이터를 수집하고 수집한 데이터를 처리한 후 타당한 통계적 추정 이론에 입각하여 생산하는 통계이다. 이와 관련하여 본 매뉴얼에서는 진단항목을 매우 다양하게 평가할 수 있도록 하였다. 특히 조사방법과 무응답 개선을 위해 매우 자세한 진단항목이 포함되도록 하였다.

보고(행정)통계는 행정계통이나 관련 전산망을 통해 수집되어 작성되거나 다른 행정 행위의 부산물로 얻어진 자료를 수집하여 작성된다. 보고통계 또는 행정통계는 조사통계와는 다른 절차에 의해 생산되는 통계이며, 따라서 관련 진단항목은 별도로 구성하여 진단이 수행되도록 하였으며, 조사통계와 보고통계가 혼합되는 경우에 대해서도 진단할 수 있도록 하였다.

가공통계는 조사통계나 보고통계 등 일차적으로 조사된 통계정보를 2차 가공하여 작성되는 통계이다. 이미 존재하는 데이터를 활용하여 작성되므로 직접 조사를 할 필요가 없다. 앞서의 조사통계나 보고통계와는 다른 성격을 띠며, 자료의 수집, 처리, 분석에 관한 진단항목이 있기 때문에 별도의 신규 진단 항목의 개발 없이 충분히 포괄할 수 있기 때문에 본 매뉴얼에서는 유럽연합(EU)에서 개발한 DESAP를 기반으로 각 분야별로 통계생산자 또는 통계생산부서의 직원이 자체품질진단서를 보다 용이하게 작성할 수 있도록 하는데 주안점을 두어 매뉴얼을 개발하였고, 품질 차

원의 7가지 항목에 대한 상세한 내용은 매뉴얼 내용에 맞도록 설명을 추가하였다.

### 제3절 연구방법

유럽연합(EU)에서 개발한 DESAP을 보다 면밀히 분석하여 기본적으로 국제비교가 가능한 항목별로 상세한 설명을 매뉴얼에 첨가하였고, 국내 통계작성 환경에 맞지 않는 항목은 삭제하여 일관성을 유지하고자 하였다. 국내에서 2개년 간 수행되어 온 품질진단서 및 진단 결과를 분석하여 반드시 점검되어야 할 사항을 자체품질진단서에 첨가하도록 하였고, 외국의 자체품질진단 사례를 분석하여 이를 진단서에 첨가하도록 하였다. 또한 OECD에서 운용한 품질진단의 내용을 분석하여 국내통계작성환경에 맞는 사항을 첨가하도록 함으로서 점검사항의 결과 분석에 비교가능하도록 하였다. 가공통계에 대한 자체진단서의 경우 국내에서 운용한 외부품질진단서를 기반으로 DESAP의 내용에 부합하는 항목을 개발하였고, 이에 대한 외국의 사례 또한 비교 가능한 항목을 점검토록 하였다.

이와 더불어 개발된 진단서와 매뉴얼을 통계작성기관을 대상으로 담당자 면담 또는 우편발송을 통해 개발된 진단서 및 매뉴얼의 활용성을 자체적으로 평가하도록 하여 통계 작성기관의 담당자 관점에서 진단서 작성에 관한 용이성을 점검하여 보다 현실성 있는 진단서가 되도록 하였다. 또한 주제 분야별 통계담당자들 간의 전문가 회의를 통해 진단환경에 맞는 항목들에 대한 검토하는 한편 통계청의 전문가 회의를 통한 매뉴얼내용의 부합성을 평가함으로써 현장에서의 작성단계상의 애로점을 최소화하도록 하였다.

## 제2장 품질진단방법의 분류

### 제1절 품질진단방법의 분류

각국 정부 또는 해당국가의 산업과 비정부기구 등 국가적으로 다양한 목적으로 통계가 요구된다. 이러한 수요는 국가의 경제와 사회여건을 모니터링하며, 정보를 개발하고, 정부와 민간의 프로그램과 투자 정책의 수립과 평가를 수반한다. 이와 같이 국가통계는 제도와 목적, 작성환경이 다르기 때문에 일률적인 기준으로 국가통계를 평가하는 것은 무리가 있다. 한 국가의 경제상태, 인구변동, 투자정책등과 같이 통계의 유용성은 이미 모든 국가에서 인정하고, 이를 기반산업으로 발전시키고 있으나, 국가통계의 품질에 대한 평가의 필요성은 1980년대에 이르러 제기되기 시작하였다.

품질 관리의 기준은 크게 4가지 정도로 분류할 수 있다. 첫째, 품질평가의 기준을 이용자만족도에 중요기준으로 설정한 경우이다. 국가통계의 주요 이용자가 해당통계를 이용하고자 할 때, 접근성, 일관성, 이용자적합성 등을 주요 평가기준으로 채택하고 있는 경우로서 유럽국가의 대부분이 채택하고 있는 평가기준이다. 둘째, 통계작성 절차의 정확성을 중요 평가기준으로 채택한 경우이다. 표준적인 통계작성절차의 준수여부, 자료의 품질상태, 조사절차 및 설계의 정확성 여부 등을 평가기준으로 한다. 셋째, 통계작성기관의 품질개선노력을 중요 평가기준으로 채택하는 경우이다. 조직의 품질개선마인드 및 고객중심, 리더십 등 전사적 품질관리 모형으로 스웨덴, EFQM 등이 대표적인 사례이다. 마지막으로 ISO인증기준에 따라 품질관리, 시험계획 등의 실제 품질관리 수행절차를 점검하는 경우이다. 영국의 경우 조사절차에 대한 ISO인증기준을 적용하고 있으며, 한국의 통계청에서도 2005년 ISO 인증을 획득

하였고, 2006년에 1회 재 인증을 획득한바 있다.

이러한 평가기준과 더불어 평가대상 또는 평가주체에 따라 이용자조사, 담당자조사, 자체평가, 내부감사 및 외부감사 등으로 품질진단 방법을 분류할 수 있다.

이용자조사는 해당통계에 대한 주요 고객 또는 이용자를 대상으로 설문조사를 통해 해당 통계의 품질상태를 파악하는 방법이다. 통계작성기관은 이용자조사를 통해 작성된 통계의 특성이 이용자에게 미치는 영향과 해당기관에서 제공하는 통계의 품질 및 서비스에 대한 인식정도를 파악할 수 있게 된다. 이와 더불어 통계담당자 조사는 해당기관의 통계담당자들에 대해서도 설문을 통해 조직의 통계 품질에 대한 인식정도 및 전문성을 파악하기도 한다. 영국의 ONS와 스웨덴, 핀란드, Eurostat 등에서는 주기적으로 통계담당자를 대상으로 근무의식, 직업관, 소관통계의 품질상태, 통계관련 교육훈련 등에 대한 조사를 통해 이를 업무에 반영하고 있다.

한편, 자체평가 또는 자체진단은 조사관리자(survey manager)가 조사전반에 관한 문항으로 구성된 체크리스트를 작성하도록 함으로서 소관통계의 품질을 평가하는 방법이다. 점검문항은 주요 이용자 및 관계자의 해당통계에 대한 인식정도, 조사표, 정의, 분류 등이 현재 통계수요와 상황을 반영하는지 여부, 문서화, 자료의 일관성, 시계열유지여부, 공표예고, 통계작성절차의 전산화 여부 등으로 구성된다. 이에 대한 대표적인 점검표로는 DESAP(Development for Self Assessment Programme)이며, 이에 대한 상세한 내용은 4장에서 다루기로 한다.

내부감사는 자체적인 감사팀을 구성하여 품질가이드라인과 같은 감사기준을 적용하여 감사결과 미흡한 부분을 보완토록하고 있다. 이와 함께 외부감사는 외부전문가를 외부감사로 의뢰하여 전문가적인 감사를 수행하는 것으로서 통계청의 외부품질 진단이 이에 해당된다고 할 수 있다. 이러한 외부감사 또는 평가는 캐나다에서 각종 통계에 대한 외부감사를 수행하고 있으며, 감사결과를 캐나다통계청 홈페이지에 게재하고 있다(Report of Auditor General of Canada, 1999).

## 제2절 품질 점검표(Quality Checklist)

각국 정부 또는 해당국가의 산업과 비정부기구 등 국가적으로 다양한 목적으로 통계가 요구된다. 이러한 수요는 국가의 경제와 사회여건을 모니터링하며, 정보를 개발하고, 정부와 민간의 프로그램과 투자 정책의 수립과 평가를 수반한다. 이와 같이 국가통계는 제도와 목적, 작성환경이 다르기 때문에 일률적인 기준으로 국가통계를 평가하는 것은 무리가 있다. 한 국가의 경제상태, 인구변동, 투자정책등과 같이 통계의 유용성은 이미 모든 국가에서 인정하고, 이를 기반산업으로 발전시키고 있으나, 국가통계의 품질에 대한 평가의 필요성은 1980년대에 이르러 제기되기 시작하였다.

통계품질평가는 통계작성기관에서 생산된 통계 및 관련 서비스의 내용이 품질평가 기준에 부합하는 지를 판단하고, 신규통계 또는 통계품질개선 계획 및 효율성을 평가하는 중요한 수단이 된다. 이와 관련하여 통계작성과정이 이용자요구사항 및 필요의 우선순위를 충족시키는지에 대한 품질을 평가할 수 있을 것이다. 통계의 품질을 평가하기 위한 과정은 통계작성과정에 따라 복잡하게 구성되며, 자료의 품질평가 목록을 작성하여 평가 틀로 사용이 가능할 것이다. 이와 더불어 통계작성과정을 문서화 하여 평가항목으로 구성하면 품질 점검표로 활용이 가능하다.

각국의 품질평가 모형은 표준화된 방식이라기보다는 해당 국가의 특성에 따라 평가방식이 다르기 때문에 자세한 설명은 생략하기로 하고, 점검표를 작성하여 평가하는 방식에 대해서만 다루기로 한다.

품질 점검표는 통계작성기관 또는 감독기관이 자체 품질평가를 위해 개발하여 사용하고 있는 점검표(check list)로서 통계작성 절차 및 결과자료를 평가하는 데 사용된다. 이러한 품질 점검표를 도입하여 적용하고 있는 대표적인 국가로서는 뉴질랜드 통계청, 영국통계청, 네덜란드통계청 등이다. 작성기관의 통계작성 담당자가 직접 소관통계에 대해 점검표 상의 항목별로 직접 점검하기 때문에 품질문제에 대해 개선

점을 스스로 도출하고, 해결할 수 있기 때문에 자율적 개선에 도움이 된다.

### 1. 뉴질랜드 통계청의 품질 점검표

뉴질랜드 통계청의 품질 점검표의 주요사항을 살펴보면 다음과 같다.

- 통계담당 직원이 소관통계의 주요 이용자와 새롭게 파악된 이해관계자들을 파악하고 있는가?
- 통계작성절차 및 결과에 대한 문서화를 철저히 이행하고 있으며, 문서화된 정보를 이용자들이 쉽게 접근할 수 있는가?
- 사용된 자료의 정의들이 일관성(coherence)을 유지하고 있는가?
- 표본이 주기적으로 개편되고 있는가?
- 계절적 요소에 대해 계절조정 및 분석이 일어나고 있는가?
- 자료의 공표시점이 사전에 홍보되고 있는가?
- 이용자의 요구를 만족시킬 만큼 자료의 생산에 표준적인 시간이 소요되고 있는가?
- 공표된 자료로부터 응답자에 대한 비밀이 보호되고 있는가?
- 품질평가지표에 대해 정기적 평가 및 모니터링이 실시되고 있는가?
- 통계법에 제시된 요구사항이 준수되고 있는가?

### 2. 영국 통계청의 품질 점검표

영국의 통계청에서 개발하여 적용하고 있는 품질 점검표는 5개 부문에 대해 총 43개 문항으로 구성되어 있으며, 매우 상세한 내용으로서 통계품질에 대한 자체점검에 적절한 가이드라인이 될 수 있을 것이다.

#### A. 목적

A1 : 목적

A1.1 조사를 실시하게 된 이유는 무엇인가?

A1.2 어떤 정보를 수립하려고 하는가?

A1.3 다루고자 하는 주제와 주요 자료항목은 무엇인가?

A1.4 자료항목을 정의하는데 적용한 국내외 기준은 무엇인가?

B. 조사범위

B1 : 목표모집단

B1.1 조사의 목표모집단은 무엇인가?

B1.2 모집단의 조사단위를 파악하여 분류하는데 근거가 되는 기준은 마련되어 있는가?

B1.3 목표모집단에 대한 표준분류가 있다면, 이에 쉽게 접근할 수 있는 방법은 무엇인가?

B2: 조사모집단

B2.1 조사의 실제 대상모집단은 무엇인가?

B2.2 목표모집단과 조사모집단이 어느 정도로 근접한가?

B3: 추출틀

B3.1 표본조사의 경우 표본추출에 사용된 추출틀은 무엇인가?

B3.2 시간 경과에 따라 추출틀에 변동이 있는가? 변동이 있다면 어떻게 변동되는가?

B3.3 조사모집단 내부에서 발생, 소멸등과 관련하여 추출틀이 갱신되는가?

B3.4 보고서에 추출틀의 주요변수에 대한 요약된 통계표가 수록되는가?

C. 설계

C1 : 표본조사

C1.1 어떤 형식의 표본설계를 하였는가?

C1.2 목표 표본크기와 조사 완료된 표본크기간의 차이는 어느 정도인가?

C1.3 표본설계에서 층화를 적용시킨 경우 어떤 방법으로 층화하였는가?

C1.4 표본추출방법은 무엇인가?

C1.5 경상조사의 경우 시간의 흐름에 따라 표본설계 방법을 변경한 적이 있는가?

C2 : 자료의 정의

C2.1 보고서에 주된 개념에 대한 설명과 주요항목에 대한 정의를 수록하고 있는가?

C2.2 보고서에 조사표 또는 수집자료 양식을 수록하고 있는가?

C2.3 계속조사에서 시간의 경과에 따라 자료 항목의 정의에 대한 변동여부를 공지하고 있는가?

C3 : 자료수집방법

C3.1 방문조사의 경우 조사원을 훈련시키거나 조사원이 습득해야할 전문지식은 무엇인가?

C3.2 자료가 수집되었을 때 자료의 품질과 관련하여 어떠한 검사를 하고 있는가?

C3.3 조사에서 응답자 오류를 최소화 하기위해 어떤 절차를 거치고 있는가?

C3.4 자료수집방법에 대해 시험 조사를 실시하였는가?

C3.5 수집된 자료 중에 의심스러운 항목이 있는가?

C3.6 무응답을 최소화 하기위해 어떤 절차를 거치고 있는가?

C4 : 자료처리

C4.1 자료처리과정에서 오류를 최소화 하기위해 어떤 절차를 거치고 있는가?

C4.2 이상치를 발견하여 검증하고 발견된 오류를 수정하기위해 체계적인 관리를 하고 있는가?

C4.3 비밀정보 누설 방지를 위해 어떠한 조치를 취하고 있는가?

D. 분석

D1 : 추정

D1.1 표본조사의 경우 보고서에 수록된 추정치를 산출하기위해 어떤 표본가중치 부여방법을 사용하고 있는가?

D1.2 추정결과는 이미 알려진 모집단 값으로 추정되고 있는가?

D1.3 시계열자료인 경우 계절조정 또는 추세조정기법이 사용되었는가?

D2 : 신뢰성

- D2.1 표준편차의 추정치를 제공하고 있는가?
- D2.2 조사결과 추정치의 품질이 비표본오차의 영향을 어느 정도 받고 있는지에 대한 평가결과를 제공하고 있는가?
- D2.3 무응답 상황에 대한 통계표를 제공하고 있는가?
- D2.4 응답자와 무응답자 그룹간의 어떤 차이점이 있는지 알려졌는가?
- D2.5 항목무응답이 어느 정도 존재하는가?
- D2.6 무응답 항목을 채워서 완성된 개별 레코드를 만들기 위해 대체(imputation) 방법이 사용되었는가?

E. 추가사항

E1 : 내재적 비교

- E1.1 보고서에 수록된 추정치간의 효과적인 비교를 위해 가능한 방법은 무엇인가?

E2 : 외적 비교

- E2.1 보고서가 정해진 기일 내에 마감되어야 하는 경우, 즉 잠정치가 갱신예정일 경우 확정치로 갱신되어야할 이유를 밝히고 있는가?
- E2.2 보고서에 수록된 통계분석과 관계가 있거나 관련된 정보를 포함하고 있는 자료출처를 이용자들이 쉽게 찾아볼 수 있도록 하고 있는가?

E3 : 추가정보

- E3.1 조사에 사용된 통계기법과 자료수집방법에 대한 추가정보를 제공하기위한 연락처를 제공하고 있는가?

### 제3절 품질감사

통계품질을 평가하는 방식으로 통계감사(statistical audits)방식을 도입할 수 있다. 통계감사방식을 도입하는 국가로는 영국, 스웨덴, 네덜란드, 캐나다 등 이며, 별도의 통계감사팀을 구성하여 통계작성절차와 결과에 대한 품질감사를 실시하고 있다. 통계감사는 품질가이드라인 또는 통계품질체크리스트의 내용을 근거로 하여 통계작성

의 기획 단계부터 결과자료의 발간 공표까지와 결과물의 품질에 이르기 까지 전 단계에 걸쳐 점검하는 품질평가 방식이다. 통계감사는 회계감사와 같은 개념이라기보다는 통계작성기관의 소관통계에 대해 잘못된 점과 잘된 점을 평가하여 통계작성기관에게 개선사항을 공지하고, 이를 요구하는 수단으로 활용된다. 따라서 통계감사팀은 감사를 수행하는 동안 감사대상 업무와 관련하여 감사업무의 객관성을 보장하기 위해 피감기관의 지시나 영향을 받지 않도록 기관장은 감사팀에게 독립성을 부여해 주어야 한다. 최종 감사보고서는 각 통계작성기관의 기관장에게 제출하여야 하며, 각 부서장 또는 기관장은 소관통계에 대한 감사결과 개선사항에 대한 개선 실천사항을 작성하여 기관장에게 보고하고, 개선계획에 대해 승인을 받아야 한다.

### 1. 감사의 유형

감사의 사전적 정의는 다음과 같다.

“감사는 감사증거를 얻기 위해 체계적이고, 독립적으로 문서화하는 과정이며, 감사기준을 이행하기 위한 범위를 결정하기위해 감사증거를 객관적으로 평가하는 작업이다.”

이때, 감사증거는 기록물, 법률적 사실 또는 그 밖의 정보들로서 이러한 사실들이 감사기준과 증거들의 관련성을 말한다. 또한 감사기준은 일련의 정책, 절차 또는 요구조건 등이 해당된다.

감사의 목적에 따라 내부 또는 외부감사관에 의해 감사가 수행될 수 있으며, 이러한 감사를 수행하는데 요구되는 기간은 수일 에서 몇 개월이 요구될 수 있다. 또한 단순한 절차에 의해 감사가 수행되기도 하고 전체 조직에 확대하여 적용될 수도 있다. 문서검색과 관련하여 감사 자체의 목적에 따라 다양한 형태의 감사방법을 적용할 수 있다. 그러나 감사의 결과는 개선을 위한 점검과 검토로 완전히 동일해야 한다. 감사의 유형을 분류하면 다양한 대안이 가능하다.

〈표 2-3-1〉 감사의 유형

| 감사의 유형                             |   | 국가별 사례                  |
|------------------------------------|---|-------------------------|
| 품질감사                               | 절차감사  | ONS-영국, 호주통계청           |
|                                    | ISO-방식 감사                                   | INE-포르투갈                |
|                                    | 추적감사  | INE-포르투갈                |
|                                    | 우수사례감사                                      | CBS-네덜란드, ONS-영국        |
|                                    | 세부점검  | ONS-영국, 스위스통계청, 덴마크통계청  |
|                                    | EFQM 감사                                     | 호주통계청, DESRATIS-독일연방통계청 |
|                                    | IMF DQAF(Data Quality Assessment Framework) |                         |
| 통계감사                               | CBS-네덜란드, ONS-영국, SCB-스웨덴통계청, 호주통계청         |                         |
| EDP(Electronic Data Processing) 감사 | ONS-영국                                      |                         |
| 긴급감사                               | CBS-네덜란드, ONS-영국                            |                         |
| 경영관리-방식 감사                         | 경영감사  | -                       |
|                                    | 적응감사  | -                       |
|                                    | 조사감사  | -                       |
|                                    | 회계감사  | -                       |

## 2. 감사의 유형별 특징

### 1) 절차감사

#### □ 정의

조직의 표준에 부합하도록 하기 위해 모든 주요한 절차를 전반적으로 검토한다. 이러한 검토는 정기적으로 수행되어야 하며 절차가 해당 조직의 기능에 얼마나 중요한지에 따라 좌우된다. 즉, 절차가 중요한 산출물의 명성이나 전달에 영향을 준다면 검토는 분기별로 수행하는 것이 필요할 수 있다. 그러나 만일 절차가 단지 조직의 하나의 영역에 대해 주요한 영향을 끼친다면 년2회 정도의 검토가 바람직하다.

검토는 투명해야하며, 사용한 방법과 관련된 모든 정보를 이용 가능해야 한다. 검토절차는 생산자, 이용자 그리고 제3자에게 대표성을 가져야 한다.

## □ 범위

- 모든 주요 절차.
- 모든 주요 이용자/제3자를 고려.
- 당면할 모든 주요 이슈를 정의.

## □ 조직적 측면 또는 참고사항

- 절차의 주체를 정의한다.
- 감사회수는 절차에 따른다.
- 산출물/결과는 범위와 목적의 요약, 점검의 이유, 내외부 인사를 포함해야 한다. 권고목록과, 개선에 대한 영역 및 제언목록 등에 가능한 우선순위를 부여한다.
- 참고사항은 최종 점검에 포함될 수 있다.

## □ 적용사례

- ONS-영국

영국의 모든 국가통계의 산출물은 연차적으로 5년간의 프로그램의 부분으로 점검하도록 하고 있다.

- 호주통계청

1999년에 사업체통계(광업, 제조업, 해외무역)에 대해 자료처리과정을 최적화하기 위해 처음으로 평가프로젝트(OPTDA)를 시작하였다. 절차에 대한 세밀한 분석과 개선을 위한 공식적인 제안은 2000년 7월에 종료되었고, 제시된 대부분의 개선안은 2000년 말에 실현되었다.

2000년도에 포괄적 절차분석이 시작되었다. 우선 보다 효율적인 프로젝트 모형 조직으로 전통적인 조직구조를 개편하기 위한 측면으로 특정한 통계적 절차에 관한 기본적인 정보를 살펴보기위해 인구통계국내부에서 사전프로젝트(REORGBEV)로서 출발하였다. 이러한 사전프로젝트의 결과는 통계생산과정과 국 차원의 모든 통계프로젝트의 신규 계획과 관리를 재설계하기 위한 기초가 되었다. 분석은 2001년 9월에

종료되었다. 데이터 에디팅 절차, 발간절차의 신속화 등과 같은 개선점이 실현되었다. 이러한 사전프로젝트의 경험과 결과는 호주통계청의 다른 부서에서 프로젝트와 절차를 재 조직화하는데 이용되었다.

## 2) ISO-방식의 감사

### □ 정의

가장 최근에 감사에 대한 ISO의 정의에 따르면 “감사는 감사증거를 얻기 위해 체계적이고, 독립적으로 문서화하는 과정이며, 감사기준을 이행하기위한 범위를 결정하기위해 감사증거를 객관적으로 평가하는 작업”으로 정의하고 있다. 이러한 국제적인 규범은 감사원칙에 대한 지침을 제시하고 있다. 즉, 감사프로그램을 운영하는 방법, 품질관리와 환경관리 시스템에 대한 감사 수행방법에 대한 지침뿐만 아니라 각 감사관의 능력에 대한 지침을 제공한다. 감사는 조직의 효율성을 개선하기 위해 가능한 중요한 정보를 제공함으로써 정책을 지원하고 경영관리기능을 지원하는 강력한 도구이다.

### □ 범위

이러한 ISO-방식의 감사는 품질경영시스템에서 적용하는 ISO9000:2000 과 환경관리시스템에서 적용하는 ISO14000의 규범을 참고하여 조직에 적용할 수 있다. 이는 해당조직 또는 어떤 유형의 조직에서 운영하는 모든 절차에 적용이 가능하다.

### □ 조직적 측면 또는 참고사항

ISO-방식의 감사는 외적 또는 내부적 감사가 가능하다.

- 내부감사는 품질시스템과 내부적 목적을 점검하기위해 수행되며, 내부품질감사팀에 의해 수행된다.
- 외부감사는 제3자나 조직에 관심이 있는 외부인사에 수행되거나, ISO9001 또는 ISO14001에 대한 인증을 가진 외부평가기관에 의해 품질과 환경 시스템을 독립적으로 감사할 수 있다.

## □ 적용사례

## · INE-포르투갈

2000년 이래로 감사 절차는 통계적 절차에 초점이 맞춰졌다. 감사는 Statistical Production Procedure Handbook에 의해 수행된다. 그러나 내부적인 품질감사의 전체적인 절차는 국제표준규범인 ISO를 고려하여 조직된다. 감사의 목표는 만일 통계 관련자가 조직에 정착된 절차에 따라 업무를 수행하고 있음을 입증하기 위함이며, 이러한 절차에 대한 설명을 개선하기 위한 것이다. 그러므로 감사업무는 지속적인 품질개선을 위한 하나의 도구이다.

위에서 언급한 규범에 따라 내부절차는 내부품질감사 절차로 설정되며, 품질감사 수행을 위한 문서를 제공 받도록 되어있다. 이러한 절차와는 별도로 감사관들의 업무를 위한 문서는 절차에 근거하여 점검표로 제공된다. 감사는 증거조사에 근거하며 조사를 위해 모든 프로젝트에 대한 최소한의 문서에 근거한다. 이러한 최소한의 문서는 Statistical Production Procedure Handbook에 필수적으로 제시된다.

내부감사가 시작되면 이에 대한 훈련이 중요하다. 14명의 감사관이 전체 팀을 구성하며, ISO 규범에 근거하여 감사를 위한 행동양식과 기술을 배우기 위한 훈련프로그램에 참여한다. 전체 절차는 준비를 위해 1년이 소요되며 팀원의 어떤 사람도 전속(full-time) 감사관은 없다. 팀은 서로 다른 영역의 통계학자로 구성되며, 전체과정은 팀 자체적으로 설계된다. 중요한 다른 측면은 각각의 감사는 평균적으로 4~5일 정도 소요되며 각 감사관은 연간 2~3개를 감사하게 된다(각 팀별은 2~3명의 감사관으로 구성됨).

## 3) 추적 감사

## □ 정의

이전의 권고안에 대한 점검 또는 감사로서 승인된 사항과 모든 권고안을 이행했는지에 대한 대규모 점검이며, 이행된 권고안이 바람직한 산출물로 나타나도록 하는 점검으로 사후수행의 개선점으로 점검하는 것이다.

□ 범위

- 이전의 점검으로부터의 권고안을 제시하며, 초기 권고안으로부터의 개선안은 도출하지 않는다.

□ 조직적 측면 또는 참고사항

추적감사는 원래의 점검보다는 소규모이어야 하며, 초기점검으로부터 일정정도의 시간간격을 두고 수행되어야 하며, 권고안을 이행할 수 있는 충분한 시간이 허용되어야 한다.

□ 적용사례

- INE-포르투갈

추적감사는 이전의 점검절차를 따르게 된다. INE-포르투갈의 내부품질감사의 경우 감사결과 비정합성에 대한 수정작업이 필요할 때 추적감사를 수행한다. 추적감사에 필요한 날짜는 감사관과 감사팀이 결정한다.

4) 우수사례 감사

□ 정의

이 감사는 중심영역에서 공유된 정보를 전체조직에서 이용 가능한 조직내부에서 수행된다. 이 절차는 우수사례를 조직전체가 공유할 수 있도록 보장한다.

□ 범위

적용범위는 전체 과정이지만, 감사에 포함된 고위간부는 소규모 또는 단기간의 절차에 투입되는 것이 바람직하다. 우수사례를 확립하기 위해 성공사례뿐만 아니라 실패사례로부터의 교훈을 배우도록 잘 된 것과 그렇지 않은 것을 포함시키는 것이 중요하다.

□ 조직적 측면 또는 참고사항

성공적인 결과물은 사례로 이용되어야 하고 이러한 사례는 전체적으로 선례가 되도록 해야 한다. 대규모 또는 소규모 조직은 이러한 것들을 책임질 수 있지만, 업무는 사례로서 이용할 것이 무엇인지를 정의한 상태에서 수행되어야 한다.

□ 적용사례

· CBS-네덜란드통계청

네덜란드 통계청의 내부 감사팀은 개인과 사업체에 의한 무응답과 관련된 사무실에서 우수사례를 조사한다. 감사결과를 적용하여 응답률을 높이고 있다. 감사팀은 매우 열적으로 업무를 수행한다. 이러한 점이 매우 긍정적으로 작용하고 있다. 최종적으로 감사팀은 일종의 대조 권고안을 마련한다.

5) 세부 점검

□ 정의

세부그룹 평가의 목표는 결과물에 대한 세부적인 생각이 무엇인지에 대한 관점과 논쟁거리를 찾는 것이다. 주요한 결과물을 공표한 후에 또는 다음 단계로 진보하기 전에 어떤 조언을 구했을 때 점검이 가능하다.

□ 범위

세부점검은 모든 절차를 포괄할 수 있지만 모든 주요 세부사항을 포함하여 나열하는 것은 어렵고, 주요절차가 완전함 감사가 아닐지라도 점검이 필요할 때 수행하는 것이 가장 적절할 수 있다.

□ 조직적 측면 또는 참고사항

이러한 점검은 비공식적으로 수행될 수도 있지만, 결과물과 논점들은 향후 참고자료로서 기록되어야 한다. 주요고객, IT 관리자, 이론가 등 모든 주요한 세부점검사항을 규정해야한다.

## 6) EFQM 감사

### 정의

EFQM 모형에 따라 조직의 품질을 평가하는 시스템이다. 이러한 감사에 기초하여 절차와 결과의 개선점을 제안할 수 있다.

### 범위

EFQM 모형은 자체진단의 실행으로 조직에 적용할 수 있고, 내부 팀에 의해 구성될 수 있다. 전체조직에 적용가능하며, 단일 부서와 같은 소규모 조직에도 적용가능하다.

### 조직적 측면 또는 참고사항

자체진단의 실행이나 외부감사는 모형에 대한 숙련된 지식을 가진 감사팀이나 컨설팅 팀에 의해 감독되어야 한다. 모형은 점검표와 평가점수표로 구성되며 모형을 통한 감사팀의 감사로 수행된다. 감사보고서는 최고경영자에게 전달되고 개선안이 최고경영자에 의해 판단하게 된다.

### 적용사례

#### · 호주통계청

호주 통계청은 EFQM 모형에 근거하여 자체진단을 수행하고 전체 조직에 적용하고 있다. 21명의 고위간부와 중간간부들에게 모형에 대한 훈련을 실시하여 2000년에 시작하였다. 자체진단 결과 개선해야할 강점과 영역을 2001년 7월에 제시하였고, 행동계획은 2001년 8월에 전개하였다.

## 7) IMF- DQAF(Data Quality Assessment Framework)

### 정의

IMF로부터 DQAF는 통합적이고 유연한 평가체계로서 자료품질을 연구 환경, 통계적 절차 통계산출물의 특성 등을 평가하는 구조를 이용하여 진단한다. 이러한 체계는 국가통계국, 국제기구, 기금외부의 자료 이용자들 간의 협의로 통계개발기금에 의해 발전되었다. 통계적 관점에서 우수사례와 국제적으로 승인된 개념과 정의를 함께 고려하며, 통합성, 방법론적 순수성, 정확성, 신뢰성, 서비스가능성, 접근성, 상대적인 연구필요성과 같은 자료 품질의 다양한 차원을 포괄한다.

#### □ 범위와 문서

일반적인 DQAF는 자료로 규정된 체계에 대한 포괄성을 제공한다. IMF의 통계부서에서 생산한 통계보고서를 <http://dsbb.imf.org/Applications/web/dqrs/dqrsdqaf/> 사이트에서 제공한다.

### 8) 통계 감사

#### □ 정의

통계산출물의 품질과 다른 통계 결과를 평가하는 시스템이다. 이러한 감사에 기초하여 절차와 결과에 대한 개선점을 제안할 수 있을 뿐만 아니라 통계생산과정에 관련된 조직 전체에 이에 대한 지식을 전파할 수 있다.

#### □ 범위

통계감사는 모든 통계적 절차(조사자료, 연구자료, 행정자료)에 적용 가능하다.

#### □ 조직적 측면 또는 참고사항

통계감사는 통계생산 절차에 관여하고 있는 구성원과 자료를 이용하는 이용자, 문서와 관련된 분석을 하는 관련자들에 대한 면접을 통해 수행된다. 통계적 지식을 가진 내부 감사관에 의해 수행되며, 감사보고서는 관리자에게 제출된다. 감사관은 이러한 유형의 감사를 수행하기 위해 문서들을 참고한다.

## □ 적용사례

## · ONS-영국

통계감사는 모든 주요 산출물에 대해 ONS 내부에서 품질 점검 프로그램(Quality Assurance Review Programme)으로 5년간 순환 식으로 완료된다. 감사관은 UK Prime Minister's Guidance on Survey Control(2001), Reviews of ONS Statistical Surveys(2001), Review Guidance-Set-by-step desk instruction(2003) 등 3가지 지침을 이용하여 감사를 수행한다.

## · CBS-네덜란드통계청

네덜란드 통계청은 체계적으로 서로 다른 통계의 평가를 위해 통계감사 시스템을 이용한다. 이러한 감사에 기초하여 절차와 결과에 대한 개선점을 제안할 수 있다. 이는 Plan-Do-Check-Act 사이클의 부분이다. 이러한 감사의 근거는 통계의 목적에서부터 결과의 공표를 포함하여 결과에 대한 품질을 고려한 전체 통계 절차와 관련된 품질가이드라인(통계품질 점검표)이다. 모든 통계는 5년에 1회씩 감사를 받도록 되어 있다. 감사는 주로 통계의 장단점을 평가하여 통계를 생산하는 부서를 지원하는 수단으로 사용되며, 통계절차와 결과의 개선을 위해 제안을 공식화 하는데 도움을 준다. 실제 감사는 소위 “감사팀”에 의해 수행되며, 감사팀은 감사를 수행하는 동안 해당사항과 관련하여 피검자로부터 직접적인 영향을 받지 않는다. 이러한 사항은 조사의 독립성과 객관성을 담보하는데 매우 중요하다. 현실성과 지속성을 위해 약 40명의 내부감사팀을 운영하고 있다. 이들 감사관들은 외부기관에서 감사수행을 위한 훈련을 받는다. 주기적으로 감사관들을 교체하며 신규 감사관들은 훈련을 받으며, 감사팀 내에서 감사기술 및 감사의 지속성 전문성을 위해 정기적으로 감사에 투입되도록 하고 있다. 내부감사관을 이용하는데 대한 장점은 통계청의 모든 부분을 파악하고 있다는 점이다. 국장, 통계청장, 과장은 최종보고서를 요구할 수 있다. 청장과 과장은 국장에 대해 행동계획을 제시하는 책임을 지닌다. 부서담당자는 청장에게 개선계획에 대한 승인을 요청한다.

· SCB-스웨덴통계청

스웨덴 통계청은 내부감사에 대한 5개년 프로그램을 시작하였다. 체계적인 품질감사는 23개 통계 산출물에 대해 2002년 11월부터 2003년 5월 까지 수행되었다. 14개 통계에 대한 감사가 2003년에 수행되었다. 감사절차를 간단히 설명하면 다음과 같다.

- 특별히 훈련된 한 팀의 감사관들이 감사를 수행하며, 1주일이 걸리지 않도록 한다.
- 감사팀은 개선방안과 우수사례를 포함한 간단한 보고서를 작성한다.
- 감사의 표준화된 절차를 위해 점검표를 사용한다.

· 호주통계청

2003년에 호주통계청은 표준화된 품질보고서에 기초하여 새로운 품질감사 시스템을 시작하였다. 이는 통계위원회의 도움을 받아서 호주통계청의 TQM-위원회에 의해 조직되었다. 월별모임(참여자 : 호주통계청, 통계위원회, 내외부의 주요 통계이용자, 외부 전문가)을 통해 개선에 대한 행동계획을 수립하게 되었다.

9) EDP(Electronic Data Processing) 감사

정의

IT절차의 품질을 평가하기 위한 시스템은 내부 클라이언트에 의존한다. 이러한 감사에 근거하여 절차와 결과에 대한 개선안은 IT절차에 대해 제시될 수 있다.

범위

EDP 감사는 모든 유형의 IT절차에 적용가능하다.

조직적 측면 또는 참고사항

EDP 감사는 IT 지식을 가진 내외부 감사관에 의해 수행되어야 하며, 절차상의 구성원과 이용자, 관련 문서 분석자들의 면접을 통해 수행된다. 보고서는 관리자에게

제공되어야 한다.

적용사례

- ONS-영국

정보관리국(Information Management Division) 내부의 특정 절차에 대한 감사는 외부 컨설팅 업체에 의해 수행되었다. 감사 지침은 Information management Programme(2002) 와 Software management(2002), ABS annual Report(2001-2002), Consultancy review of the CSO(1996) 이다.

10) 긴급감사

정의

긴급감사는 여러 가지 문제점을 안고 있다. 감사결과를 바탕으로 절차와 결과에 대한 개선안을 제안할 수 있다. 주된 목적중의 하나는 매우 빠르게 절차를 개선할 있도록 강제하는데 있다.

범위

긴급감사는 모든 종류의 통계적 또는 비통계적 절차에 적용할 수 있다.

조직적 측면 또는 참고사항

긴급감사는 통계적 절차에 대한 감사의 경우 통계지식을 가진 내부감사팀에 의해 절차에 속한 구성원과 이용자에게 대한 면접을 통해 수행된다. 감사결과는 감사보고서 제출로 완료되며, 개선사항은 TF 팀에 의해 실행된다.

적용사례

- CBS-네덜란드통계청

네덜란드 통계청은 문제가 발생한 곳에 긴급감사를 수행한다. 감사관들은 통계청장에 의해 임명된다. 문제의 중요성에 따라 감사팀은 외부 감사관들로 구성되거나,

내부감사관들 또는 혼합으로 구성된다.

#### 11) 경영관리 방식의 감사

##### □ 경영감사

성과와 관리감사로 알려진 경영감사는 경영의 효율성과 효과성을 평가하는 것을 목적으로 한다. 경영감사는 관리표준-장기계획, 예산, 경영정책과 절차에 대해 행정 조직을 측정하고 평가한다. 비록 재정자료가 지침에 의해 지속되지만 경영을 개선하기 위한 제공되는 수치를 감사한다.

##### □ 적응감사

적응감사는 기관의 정책과 절차에 대한 적응도를 결정하기 위한 감사로서 적용 가능한 법률과 규칙, 계약 등을 점검한다.

##### □ 조사감사

조사감사 : 대학기금의 가능한 부정사용 또는 부적절한 사용을 조사하는 것이다 (University of Rochester). 조사감사는 다른 감사와는 차이가 있는데 이는 다른 감사는 정기적으로 감사결과에 영향을 받을 수 있는 개인에 대해 감사초기에 공지하지 않기 때문이다.

##### □ 회계감사

회계감사: 기금에 대한 책임과 권한, 기금의 수입과 지출을 포함한 회계처리의 결산과 보고에 대한 감사를 말한다. 이 감사의 목적은 현금과 현물자산에 대한 충분한 통제가 되고 있는 지와 자원의 취득과 사용에 대한 적절한 절차적 통제가 되고 있는지를 규명하는 것이다. 그러나 외부 회계감사와는 달리 내부 회계감사는 전문적인 의견을 피력하지 않는다.

## 제4절 자체품질진단

### 1. 자체품질진단의 법적의미

통계품질진단과 관련하여 2007년 10월에 개정된 통계법에 따르면 통계법 제9, 10조, 11조에서 정기통계품질진단, 수시통계품질진단 그리고 자체통계품질진단을 수행하도록 법률로 명시하고 있다. 통계법 9조의 정기통계품질진단의 경우 통계청장은 품질진단 대상통계에 대한 진단계획을 수립하여 진단대상통계, 진단결과의 활용계획, 기타요구사항을 통계작성기관의 장에게 통보하도록 하고 있으며, 진단과정에서의 필요사항 등에 대해서도 통계작성기관에게 요구할 수 있도록 하고 있다.

〈표 2-4-1〉 통계품질진단의 법적 조항

#### 제9조(정기통계품질진단)

- ① 통계청장은 통계의 작성 및 보급의 제반과정에 대하여 10년의 범위 안에서 대통령령으로 정하는 기간마다 통계품질진단(이하 "정기통계품질진단"이라 한다)을 실시하여야 한다. 다만, 작성주기가 10년의 범위 안에서 대통령령으로 정하는 기간을 초과하는 통계의 경우에는 그 통계를 작성하는 해 또는 그 다음 해에 실시할 수 있다.
- ② 통계청장은 다음 각 호의 사항이 포함된 정기통계품질진단계획을 수립하여 정기통계품질진단을 실시하는 해의 2월 말일까지 통계작성기관의 장에게 통보하여야 한다.
  1. 정기통계품질진단의 대상 통계
  2. 정기통계품질진단의 시기 및 방법
  3. 정기통계품질진단 결과의 활용계획
  4. 그 밖에 정기통계품질진단의 실시에 관하여 필요한 사항
- ③ 정기통계품질진단은 제35조에 따른 자료제출요구, 통계작성기관에 대한 방문 확인, 통계응답자에 대한 현지 확인 등의 방법으로 실시한다.
- ④ 통계청장은 제10조에 따른 수시통계품질진단을 실시 중이거나 대통령령으로 정하는 기간 내에 실시하였던 통계와 제11조에 따른 자체통계품질진단을 한 통계 중 그 품질이 우수하다고 판단되는 통계에 대하여는 정기통계품질진단을 면제할 수 있다.
- ⑤ 통계청장은 정기통계품질진단을 완료한 때에는 그 결과를 지체 없이 관계 통계작성기관의 장에게 통보하여야 한다.
- ⑥ 정기통계품질진단계획의 수립, 정기통계품질진단의 방법 및 절차 등에 관하여 필요한 사항은 대통령령으로 정한다.

## 제10조(수시통계품질진단)

- ① 통계청장은 제11조에 따른 자체통계품질진단을 실시하지 아니하거나 품질이 저하되었다고 믿을만한 상당한 이유가 있는 통계에 대하여는 수시로 통계품질진단(이하 "수시통계품질진단"이라 한다)을 실시할 수 있다.
- ② 통계청장은 수시통계품질진단을 실시하는 때에는 해당 통계작성기관에 미리 수시통계품질진단의 사유·시기 및 방법 등을 통보하여야 한다.
- ③ 제9조제3항 및 제5항은 수시통계품질진단에 관하여 준용한다.
- ④ 수시통계품질진단의 방법 및 절차 등에 관하여 필요한 사항은 대통령령으로 정한다.

## 제11조(자체통계품질진단)

- ① 통계작성기관의 장은 소관 통계에 관하여 매년 통계품질진단(이하 "자체통계품질진단"이라 한다)을 실시하여야 한다. 다만, 작성주기가 1년을 초과하는 통계의 경우에는 그 통계를 작성하는 해 또는 그 다음 해에 실시할 수 있다.
- ② 통계작성기관의 장은 소관 통계에 대하여 정기통계품질진단 또는 수시통계품질진단을 받은 때에는 그 연도의 자체통계품질진단을 실시하지 아니할 수 있다.
- ③ 통계작성기관의 장은 자체통계품질진단을 실시한 해의 12월 31일까지 그 결과를 통계청장에게 제출하여야 한다.
- ④ 자체통계품질진단의 방법 및 절차 등에 관하여 필요한 사항은 대통령령으로 정한다.

또한 10조의 수시통계품질진단은 11조의 자체통계품질진단을 실시하지 않은 기관이나, 통계품질이 저하되었다고 판단될 경우 통계작성기관에 대해 수시진단을 수행하도록 하고 있다. 다음으로 12조의 자체진단은 통계작성기관의 장에게 소관통계의 품질진단을 매년 수행하도록 명시하고 있으며, 정기진단과 수시진단, 자체진단이 중복될 경우 자체진단을 면제할 수 있도록 함으로서 진단수행으로 발생하는 예산 및 행정력의 낭비를 최소화 하였다. 자체품질진단 결과는 진단이 실시된 해의 12월 말까지 통계청장에게 제출토록하며, 통계청장은 이에 대한 개선점을 통계작성기관에 환류시킴으로써 품질진단결과를 통계의 품질향상에 반영토록하고 있다.

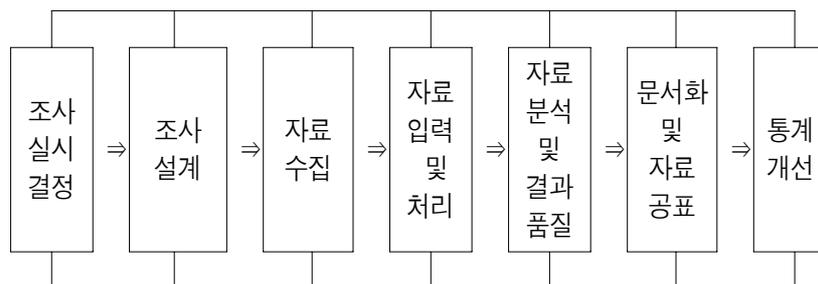
자체진단프로그램(Self Assessment Programme)은 국가통계작성기관에서 소관통계의 품질상태에 대한 자체적인 품질진단을 수행할 수 있도록 개발한 품질진단 프로그램을 말한다. 자체품질진단과 관련하여 2003년 EU 국가들의 각국 통계청이 주축이 되어 2003년도에 진단 프로그램의 개발을 완료하였고, 국내에서도 통계청의 주관으로 자체품질진단 프로그램개발을 위한 절차를 수행하고 있다.

이와 같이 통계진단에 대한 법률적 조항이 개정된 통계법에 추가됨으로서 통계작성기관에서는 정기적, 또는 수시로 소관 통계에 대한 품질진단을 수행해야하며, 통계작성기관의 소관통계에 대해 자체품질진단을 수행하고, 정당한 사유 없이 이를 거부할 수 없도록 하였다. 따라서 자체품질진단의 수행은 법률적 구속력을 가지며 통계작성 기관에서는 반드시 소관통계에 대한 품질진단을 받거나 자체품질진단을 수행하여 소관통계에 대한 품질향상을 위해 노력하여야 한다. 이때, 통계작성 부서에서는 전문적인 지식이 요구되는 사항에 대해서는 외부의 통계전문가를 활용할 수 있도록 하고 있다. 자체품질진단서를 작성함에 있어서 해당 문항에 순서대로 점검사항을 작성토록하고, 이를 분석하여 분석결과를 소관통계의 품질향상에 활용하도록 하였다. 특히 자체품질진단서의 일부 항목은 품질지표 차원의 7가지 사항에 해당되는 항목으로 이를 바탕으로 개선점을 도출할 수 있도록 하였다.

## 2. 자체진단표의 구성

자체진단을 위한 진단서는 EU의 통계품질 자체진단프로그램으로 개발되어 EU의 산하기관인 Eurostat의 주관으로 독일을 비롯한 6개 국가가 2002년 10월부터 2003년 9월 까지 공동으로 개발한 프로젝트이다.

[그림 2-4-1] 자체진단모형의 구성



각 국가별로 통계품질 개선을 위한 자체 점검을 수행하고 있지만, 국가별로 공통

적인 진단서를 적용한 것은 DESAP이 유일하다고 할 수 있다. DESAP의 진단 모형은 총 7장으로 구성되며 각각의 내용은 <표 2-4-2>와 같다.

<표 2-4-2> 자체통계 품질진단서의 구성 항목

| 주요 조사과정              | 하위요소         | 점검 항목수 | 품질수준 |           |
|----------------------|--------------|--------|------|-----------|
|                      |              |        | 항목수  | 품질차원      |
| 1장. 조사수행에 대한 의사결정    | 이용자 요구분석     | 8      |      |           |
|                      | 조사목적         | 1      |      |           |
| 2장. 조사 설계            | 조사개념         | 11     | 3    | 관련성       |
|                      | 표본설계(표본조사)   | 11     |      |           |
|                      | 측정도구개발 및 테스트 | 8      |      |           |
| 3장. 자료수집             | 자료소스         | 27     |      |           |
|                      | 조사원채용 및 교육훈련 | 17     |      |           |
|                      | 무응답 줄이기      | 5      |      |           |
|                      | 실시작업         | 10     |      |           |
| 4장. 자료입력 및 자료처리      | 자료입력         | 3      |      |           |
|                      | 에디팅 절차       | 13     | 1    | 정확성       |
|                      | 대체절차         | 16     |      |           |
| 5장. 자료 분석 및 통계결과의 품질 | 관련성          | 3      | 2    | 관련성       |
|                      | 정확성          | 15     | 3    | 정확성       |
|                      | 시의성 및 정시성    | 5      | 2    | 시의성 및 정시성 |
|                      | 비교가능성        | 2      | 2    | 비교 가능성    |
|                      | 일관성          | 3      | 2    | 일관성       |
|                      | 자료분석         | 3      |      |           |
|                      | 보안관리         | 4      |      |           |
| 6장. 문서화 및 자료의 공표     | 메타자료 문서화     | 10     |      |           |
|                      | 자료공표 전략      | 4      |      |           |
|                      | 자료관리         | 2      |      |           |
| 7장. 개선 사이클           | 적용가능성/유연성    | 3      |      |           |
|                      | 관련분야의 전문성    | 4      |      |           |
|                      | 품질관리         | 5      |      |           |
| 계                    |              | 193    | 15   |           |

<표 2-4-2>에서 보는바와 같이 통계작성 절차에 따라 이용자 요구분석부터 개선

사이클 까지 총 7장에 걸쳐 193개 항목에 대해 점검을 수행하도록 되어 있으며, 각 장에는 별도로 주제별로 절을 구성하여 작성자의 이해를 돕도록 하고 있다. 193개 점검항목을 모두 작성하는 것은 아니며, 소관통계의 점검항목에 해당되는 내용에 대해서만 체크하도록 하고 있다. 한편 7가지 주요 품질진단 항목에 대해서는 모든 통계에 대해 작성하도록 하고 있으며, 만일 소관통계의 내용과 관련이 없을 경우에는 건너뛰기 할 수 있도록 하고 있다. 조사통계의 경우에는 거의 대부분의 항목에 체크되도록 하여 보다 엄밀한 점검이 되도록 하고 있으며, 보고통계나 가공통계의 경우에는 조사통계관련 항목에 대해서는 체크하지 않고 넘어가도록 구성되어 있다. 또한 전문적인 용어의 경우에는 자체진단서 뒤에 용어해설부분을 첨부하여 생소한 통계조사관련 용어를 참조하도록 하였다.

### 3. 국가별 적용사례

#### 1) 영국

영국의 통계청(ONS)은 주요 사업체조사에 대해 주기적으로 품질업무를 위해 내부적인 검토절차를 수행하고 있다. ONS 이외의 별도의 감사팀을 구성하여 품질검토의 목표를 보장하고 있다.

가구조사에 대해서는 데이터수집 활동은 시장조사(market testing)로 수행한다. 품질평가절차의 일부로서 품질목표를 분명하게 규정하고 내외 접촉자에 의한 성과를 점검하고 있다. 이러한 감사는 노동력조사와 소비자물가지수조사에 동일하게 적용하고 있다. 추가적으로 분리된 내부감사절차는 통계결과를 산출하고 협력활동에 대해 작성기관들의 협조와 관련된 일반적인 위험을 규정하는 시스템의 품질을 보장하는데 도움이 된다. 통계감사의 연간 승인계획에서 ONS는 사업체, 또는 개인에 대한 전체적인 응답자 부담의 변동을 보고한다. 동일한 보고서에서 각 조사에 대한 계획된 응답률과 실제 응답률을 보고하도록 하고 있다. 메타자료를 포함하여 통계이용자는 각 조사의 자료품질과 조사방법에 관한 정보를 구할 수 있다.

1998년도에 ONS는 StatBase를 출범시켰고, 여기에는 이용자가 자료 품질과 각각의 조사 방법에 관한 정보를 검색할 수 있도록 하는 메타데이터베이스를 포함하고 있다. 이러한 시발점은 캐나다의 통계문서 작성시스템(Statistical Data Documentation System : SDDS)의 근간이 되기도 하였다.

### 2) 스웨덴 통계청

스웨덴 통계청에서는 각각의 조사는 24개의 자체진단 점검표에 기초하여 매년 “품질보고서(quality declaration)”를 작성토록하고 있다. 이러한 목적은 조사의 품질이 개선되거나, 낙후되거나, 또는 현재상태를 유지하고 있는지를 4가지 특성 - 관련성, 정확성, 시의성, 일관성 - 에 따라 결정하기 위함이다. 부서의 관리자는 각 책임영역에서 자체진단 결과를 검토한다. 이러한 자체진단 결과를 부서장이 요약하게 된다. 이와 같은 연간 보고서는 의회에 보고된다. 스웨덴의 감사원은 보고서를 감사하고 품질에 대한 정보의 신뢰성에 대해 평가한다. 스웨덴 통계청의 자체품질진단과정은 주요 절차변수와 특정한 품질지표에 대해 지속적으로 개선되고 있다.

### 3) 네덜란드 통계청

네덜란드 통계청에서는 1996년에 전체적인 품질관리 프로그램을 도입하였다. 이러한 프로그램을 도입한 목적 중의 하나는 모든 통계 생산 부서에게 품질관리 시스템을 소개하는 것이었다. 임시적인 지침을 1997년도에 발표하였다. 동시에 “통계감사” 시스템을 도입하였다. 이 시스템의 목적은 통계부서에서 어떻게 품질관리 기능을 하고 있는지와 통계 산출물과 절차의 품질을 어떻게 개선할지를 점검하기 위해서였다. 또 다른 목적은 우수사례를 문서화하고 품질관리 시스템에 대한 지침을 구체화하기 위함이다. 모든 감사는 25명의 직원들로 수행되며, 이들은 다양한 부서와 통계감사에 관한 업무를 수행한 직원들에서 뽑힌다. 이들은 감사와 품질관리에 대해 컨설팅 회사로부터 훈련을 받는다.

## 제3장 주요 품질 진단 차원 및 지표

### 제1절 외국의 품질진단 차원 및 지표

대부분의 지표들은 Eurostat에 의해 발간되는 통계와 관련된 것이지만, 이들 대부분의 지표는 유럽연합의 회원국들에 의해 제공되는 메타자료에 기초해야 하며, 지표들은 EU 회원국 자체의 통계품질진단 목적에 따라 적용될 수 있다.

〈표 3-1-1〉 외국의 표준 품질관리 지표

| 품질 구성요소                                   | 지표                             | 1: 주요지표<br>2: 보조지표<br>3: 경험지표 |
|---|--------------------------------|-------------------------------|
| 관련성<br>(Relevance)                        | R1. 이용자만족도 지수                  | 3                             |
|   | R2. 이용가능 통계의 비율                | 1                             |
| 정확성<br>(Accuracy)                         | A1. 변동계수                       | 1                             |
|   | A2. 과대포함 및 오분류율                | 2                             |
|   | A3. 지리적 과소포함 비율                | 1                             |
|   | A4. 개편의 평균 규모                  | 1                             |
|   | A5. 대체율 및 비율                   | 2                             |
|   | A6. 단위응답률(비가중/가중)              | 2                             |
|   | A7. 항목응답률(비가중/가중)              | 2                             |
| 시의성 및 정시성<br>(Timeliness,<br>Punctuality) | T1. 조사대상 기간의 종료와 첫 결과 간의 시차    | 1                             |
|   | T2. 조사대상 기간의 종료와 최종결과 간의 시차    | 1                             |
|   | T3. 효과적인 발간을 위한 시간표의 정시성       | 1                             |
| 접근성 및 명확성<br>(Accessibility,<br>Clarity)  | AC1. 공표거나 판매된 발간물의 수           | 1                             |
|   | AC2. 데이터베이스 접속 건 수             | 1                             |
|   | AC3. 발표된 통계에 대한 메타정보의 완성도 비율   | 3                             |
| 비교가능성<br>(Comparability)                  | C1. 개념과 측정에 대해 유럽 기준과의 차이율     | 3                             |
|   | C2. 유동성에 대한 대조통계의 비대칭성         | 1                             |
|   | C3. 비교가능한 시계열의 수               | 1                             |
|   | C4. 비교가능한 시계열의 길이              | 1                             |
| 일관성<br>(Coherence)                        | CH1. 주요 2차 이용자 요구에 만족하는 통계의 비율 | 3                             |

생산된 모든 통계에 대한 지표를 계산하기란 매우 어렵기 때문에 주요통계 또는 변수는 선택적으로 계산하고 각 통계영역에 대해 일치시켜야 한다. 차원별 자료의 품질을 평가의 점검목적에 대해 연간 지표를 축적하도록 한다. 그러나 대부분의 지표는 주기적으로 지표를 계산해야 한다. 4장에서는 품질진단의 6가지 차원에 대한 정의와 각 품질진단 차원별 지표와 계산 방법들을 고찰하고, 각국의 진단 및 적용사례를 살펴보고자 한다.

## 제 2절 관련성

### 1. 관련성의 정의

**관련성(Relevance)**은 통계가 현재와 잠재적 사용자 요구에 부합하는 정도를 말한다. 이는 요구되는 모든 통계가 생산되고 있는지를 반영하며, 사용된 개념(정의, 분류 등)의 범위가 사용자 요구를 반영하는 지를 나타낸다.

관련성에 관한 보고서를 작성할 때 목적은 통계가 이용자에게 유용한지, 이용자들이 사용하고 있는지의 범위를 기술하는 것이다. 이러한 목적에 대해 통계작성자는 우선 해당통계의 사용자에게 관한 정보(이용자가 누구인지, 얼마나 많은 이용자가 있는지, 얼마나 중요한지 등)를 수집하고, 다음으로 이들의 요구에 관한 정보를 수집하고, 마지막으로 이들의 요구에 부합한지에 대한 정보를 수집해야한다. 추가적으로 품질보고서는 통계산출물의 관련성 차원을 전체적으로 평가하여 이상적인 결론을 도출해야 하며, 관련성이 부족한 부분에 대한 주요한 원인을 기술해야 한다.

이용자의 요구를 적절한 통계적 용어와 산출물로 이들의 요구에 부합하도록 재해석하여 제시해야한다. 되도록이면, 용어와 산출물은 현재의 통계적 개념으로 나타내는 것이 바람직하며 국제적인 비교를 위해 가능하면 국제적으로 통용 가능한 정의와 용어를 사용하는 것이 바람직하다. 이러한 작업은 종합적인 수요를 만족할 수 있

는 점검결과로 나타나는 통계산출물(지표, 지표의 정확한 정의 등)을 정의하도록 한다.

보고서는 이러한 종합적인 측면에서 이용자들 그룹별로 제시하고, 이들의 우선적인 요구사항과 기대치 간의 상충되는 점을 반영하도록 한다. 만일 이용자 그룹이 전략적으로 중요할 경우에는 이들의 요구에 대한 보다 충분한 설명을 기술하도록 한다.

이용자 관련성을 측정하기 위해 정확성, 시의성 등과 같은 여타의 품질요소에 따라 이용자들에게 질문해야 한다. 그러므로 관련성은 여타의 품질 구성 요소와 관련된 필요성과 서로 다른 이용자 분류의 관점과 매우 밀접한 관련이 있다. 그러나 고객의 관점을 항상 고려할 필요는 없다. 서로 다른 이용자나 고객의 상충된 속성이 한 가지 이유이며, 다른 하나는 시간과 공간에 대해 비교될 수 있는 통계의 일치성과 일관성이 필요하기 때문이다. 결과적으로 통계기관은 정상적으로 서로 다른 이용자의 요구를 정의하고 그들의 만족도를 평가할 수 있는 서로 다른 척도를 가진다. 예를 들어 일반적인 “전통적인” 이용자만족도 조사는 통계산출물과 서비스에 대해 알려져 있는 고객 또는 이용자에 대해 직접적으로 표준화된 설문지에 기초한 것이다. 따라서 주요 이용자 요구와 만족도는 면접에 의해 조사되는 반면, 이미지 연구와 같은 다른 측정도구는 대중매체분석 등의 여론조사에 사용될 수 있다. 이용자들의 요구가 만족되었는지에 대한 평가는 모든 효과적인 가능한 수단을 사용하여 이루어져야 한다. 이용자 만족도 조사는 최상의 시나리오이지만, 이러한 대체 측정과 이용자만족도지표로의 대체는 출판물 판매량, 접수받은 질문 수 나 불만사항 등을 보조수단으로 활용해야 한다.

품질보고서에는 우선적으로 이용자 분류사항과 이용자를 기술한 내용이 포함되어야 한다.

#### [사례 3-2-1] 영국의 공식통계에 대한 이용자 분류

Holt 와 Jones(1998)는 영국의 공식통계이용자의 범위를 다음과 같은 트리형태로 분류하였다. 영국의 경우 트리의 상위에는 정부와 의회가 자리하고 있다. 이러한 이

유는 주요 거시경제 통계를 이용하여 주요 경제정책을 수립하기 때문이며, 정책을 수립하고 자원을 분배하는 역할을 트리의 상위그룹에서 담당하고 있기 때문이다.

[그림 3-2-1] 영국의 공식통계 이용자 분류



[사례 3-2-2] Eurostat의 이용자 분류

- 기관분류
  - 유럽 수준 : 위원회, 심의회, 유럽의회, 그 외 유럽기관 들.
  - 회원국 또는 국가 또는 지역 수준 : 경제 또는 재무장관, 그 외 부처별 장관, 국가통계기관과 그 외 통계기관(규범, 훈련 등)
  - 국제기구 : OECD, UN, IMF, ILO 등
- 사회 분야별분류
  - 유럽에서 취업자 협회, 노동조합, 압력단체 등 국가 또는 지역 수준
- 매체별 분류
  - 국제 또는 국가적 매체- 특별한 또는 대중적 - 통계수치나 분석 또는 지시사항에 관심을 가진다. 매체는 일반 대중에게 통계를 전달하는 주요한 경로이다.
- 연구자 또는 학생 분류
  - 연구자 또는 학생들은 통계를 필요로 하며, 이를 분석하거나 특별한 서비스를 필요로 하며 특정 자료에 접근하고자 한다.

□ 기업체 또는 사업체 분류

사업체나 기업체가 소유한 시장의 분석, 시장전략 등에 대해 정보 분야에서 항상 지속적인 서비스를 요구한다.

□ Eurostat 내부 분류

주어진 단위에 의해 산출되는 통계는 그 외 다른 단위에 의해 산출되는 결과에 의존한다(통합계정과 같이 비(ration) 또는 보다 복잡한 지표의 계산에서).

**[사례 3-2-3] 스웨덴 통계청의 이용자 만족도 조사(1997)**

스웨덴 통계청에서는 1997년에 통계청의 주요 통계이용자중 1,000개 이상의 기관에 대해 면접조사를 실시하였고, 다수의 품질 차원에 대해 1점부터 10점까지의 척도를 이용하여 서비스와 산출물의 순위에 대해 질문하였다.

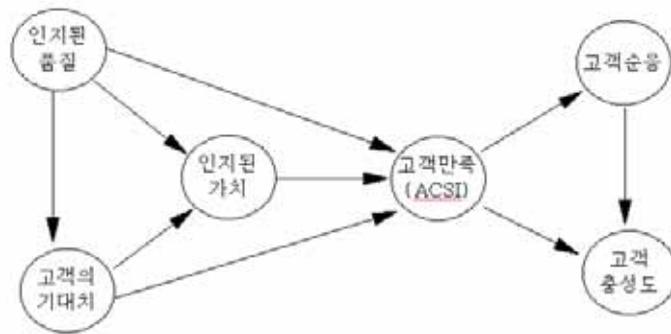
- 이용자는 공정하고 신뢰할만한 통계를 산출할 수 있는 국가통계기관에 대해 높은 수준의 신뢰성을 갖고 있다.
- 국가통계 기관의 직원들에게는 모든 조사에서 높은 점수가 부여되었다. 전화응대, 서비스마인드, 통계와 내용적 경쟁력뿐만 아니라 이해도 등을 반영한 것이다.
- 해당 기관의 자료의 시의성은 다른 품질지표보다 상대적으로 낮게 나타났다.
- 요구, 가격협상, 제시간에 배송, 불만사항의 처리 등이 이용자에 대해 중요한 질문사항이다. 이러한 사항은 해당 기관이 개선할 수 있는 영역이기도 하다.

**[사례 3-2-4] 미국의 고객만족도 지수(American Customer Satisfaction Index; ACSI)**

미국의 고객만족지수는 Claes Fornell 연구소에서 개발된 모형을 근간으로 미국에서 가구고객에 대해 상품과 서비스의 품질에 대한 고객만족도의 국가경제지수이다. ACSI 모형은 고객만족에 대해 고객의 기대치, 인지된 품질과 인지된 가치를 연계한 일종의 인과모형이다. 고객 칭찬과 충성도의 관점에서 결과를 ACSI에 연계한다. 대부분의 회사에 대해 반복되는 고객은 이익을 주는 주요고객이다. 따라서 추정된 고객의 재 구매확률은 재정적 성과를 측정하는 주요지표이다. 이 추정치를 달러로 환산함으로써 ACSI는 시간에 대한 자산으로서 해당 회사의 고객의 순수한 현재 가치

를 계산할 수 있다.

[그림 3-2-2] ACSI 모형



**완비성(Completeness)**은 평가가 용이하다. 이용불가능한 통계의 비율은 특정한 산출물에 대한 접근성으로 완결성 정도의 지표이다. 그러나 한 가지 문제는 모든 통계가 이용자에 대해 동일한 값이 아니라는 것이다. 제공된 통계의 유용성을 나타내는 이용불가능성에 대한 가중비율은 불완비성(incompleteness)을 나타내는 훌륭한 지표가 될 수 있다. 만일 가중비율을 사용할 수 없다면, 이용자들이 불완결성을 올바르게 해석할 수 있도록 원래의 가중비율을 기술통계와 함께 제시하는 것이 바람직하다. 무엇을 이용할 수 없는지와 이유를 밝힐 수 있도록 한다.

완비성은 ESS(European Statistical System) 요구 회의에 대해 분명하게 정의된 목적에 따라 이용해야할 통계와 비교하여 이용 가능한 통계의 범위를 나타낸다.

## 2. 관련성 평가 지표

[관련성 품질지표1]

| 지표             | 유형   | 통계출처     | 주기 |
|----------------|------|----------|----|
| R1. 이용자 만족도 지수 | 경험지표 | Eurostat | 연간 |

이 지표를 계산하기 위해서는 다음과 같은 준비 단계를 제시하고 있다.

- 유의미한 계급수로(최소 5인 이상) 모든 조사에 대한 이용자를 분류한다.
- 이들의 중요도에 따라 이용자 그룹을 서열화 한다.
- 체계적인 문서를 이용자 그룹에 대해 이용자 만족도 측정에 사용한다.

[필요한 메타자료]

- 이용자 그룹의 요구
- 중요 이용자의 서열화
- 동일한 그룹에 속한 이용자의 특성

[관련성 품질지표2]

| 지표              | 유형   | 통계출처                | 주기        |
|-----------------|------|---------------------|-----------|
| R2.이용가능한 통계의 비율 | 주요지표 | Eurostat,<br>국가통계기관 | 월간/분기별/연간 |

[계산 방법]

이용 가능한 통계 비율 = 제공된 값의 수/이용 가능한 필드(변수)의 수

## 제3절 정확성

### 1. 정확성의 정의

각 조사의 목적은 통계를 생산하는 것으로서 목표모집단의 수량화가 가능한 특성인 미지의 값을 추정하는 것이다. 일반적 통계적인 관점에서 **정확성(accuracy)**은 참값에 대한 정확한 계산값 또는 추정치의 엄밀성으로 해석된다. 통계값이 추정치와 차이가 나는데, 이는 변동과 편향 때문이다. 이때 변동성은 표본을 추출할 때 마다

발생하는 임의성 때문이며, 편향은 체계적인 효과로 인하여 통계값의 평균이 미지의 참값과 다를 때 발생한다.

다양한 오차들이 조사과정에서 발생하며, 이는 통계의 오차로 나타난다. 현재까지는 이러한 오차는 통계값에 반영되고 있다. 표본오차(sampling error)는 단지 표본조사에서만 발생한다. 이러한 오차는 모집단의 부분집합인 표본을 임의로 추출하여 계산하기 때문에 발생한다. 비표본오차(nonsampling error)는 표본조사와 전수조사 모두에서 발생한다. 비표본오차에는 다음과 같은 오차들이 있다.

- 포함오차(coverage error)
- 측정오차(measurement error)
- 처리오차(processing error)
- 무응답오차(nonresponse error)
- 모형가정 오차(model assumption error)

통계값과 그 통계값의 기대값 간의 변동은 분산, 표준편차, 변동계수, 신뢰구간 등으로 표현된다. 편향을 계산하기 위해서는 실제 모집단에 관한 정보와 조사과정의 세밀한 정보가 필요하다. 이러한 작업을 통해 실제로 편향값이 음수인지 양수인지에 관해 알 수 있다. 미지의 모집단 값에 대한 통계값의 변동은 분산과 편향의 제곱의 합인 평균제곱오차(MSE)로 표현된다.

## 2. 표본오차

**표본오차(sampling error)**는 표본조사에서만 발생한다. 즉, 모집단으로부터 모든 단위를 사용하지 않기 때문에 발생하는 오차이다. 표본조사로부터 산출되는 통계값은 정확하게 동일한 조사가 전체 모집단에 적용되었다할 지라도 그 값들 간에는 차이가 있을 수 있다. 표본추출에는 두 가지 방법이 있다. 하나는 확률 추출로서 모집단의 각각의 단위가 기지이며, 0이 아닌 확률로 표본으로 선택되는 경우이며, 다른 하나는 비확률 추출이다. 확률 추출에서 표본추출이론과 점근성 이론은 통계값의 기대값과 분산의 추정이 가능하게 한다. 따라서 표본추출에 따르는 변동계수를 추정

할 수 있다. 편향과 MSE는 모집단의 참 값을 모르기 때문에 계산하기 어렵다. 표본 추출이론으로부터 이 값들의 크기와 편향의 부호를 나타내는 지표를 구할 수 있다. 각 통계값의 표본 변동계수는 “표본분산의 제곱근/ 표본평균” 으로 추정된다. 만일 많은 유사통계들이 생산된다면, 보고서에는 간단하게 가장 중요한 통계값과 변동계수값(변동계수의 최소값, 최대값, 평균, 중위수, 4분위수)을 기술할 수 있다. 반면에 비확률 표본추출은 이론적으로 표본오차를 계산할 수 없다. 이러한 표본추출방법으로 추출된 표본은 확률표본추출과 유사하다는 가정으로 확률추출설계에서 사용하는 공식을 표본분산을 계산하는데 이용하고 있다.

만일 비확률 표본추출방법이 사용되었다면, 이론적으로는 표본오차를 계산할 수 없다. 그러나 표본이 확률 추출설계와 유사하게 임의성을 가진다면 표본추출상의 정확성 지표를 사용할 수 있다. 만일 이러한 경우이면, 메타자료는 적용된 접근성을 평가할 수 있도록 필요한 정보를 제공해야 한다.

**[사례 3-3-1] 사업체 통계에서 표본변동에 대한 보고서 내용의 예(Eurostat, 2001c)**

- EU 총계에 대한 변동계수 값들 : 총 변동
  - 변동계수는 무응답, 분류오차, 표본오차를 고려해야 한다. 그러나 아직까지 분류오차에 대해서는 고려되고 있지 않다. 무응답률은 항상 고려되는 것은 아니다.
- EU-총계에 대한 변동 값- 산업 소분류(3-digit)

〈표 3-3-1〉 제조업, 가스 수도사업, 건설업에 대한 변동계수의 통계값  
(단위:%)

| 주요변수 | 평균  | 최대값  | 75백분위수 | 중앙값 | 25백분위수 |
|------|-----|------|--------|-----|--------|
| 업종전환 | 1.1 | 8.3  | 1.2    | 0.8 | 0.4    |
| 부가가치 | 1.2 | 6.0  | 1.6    | 1.0 | 0.5    |
| 인건비  | 0.8 | 5.4  | 0.9    | 0.5 | 0.3    |
| 실물투자 | 3.5 | 41.5 | 3.8    | 1.9 | 0.9    |
| 취업   | 0.9 | 5.4  | 1.3    | 0.6 | 0.4    |

〈표 3-3-2〉 무역업에 대한 변동계수의 통계값

(단위:%)

| 주요변수 | 평균  | 최대값  | 75백분위수 | 중앙값 | 25백분위수 |
|------|-----|------|--------|-----|--------|
| 업종전환 | 1.6 | 3.7  | 2.3    | 1.2 | 1.0    |
| 부가가치 | 1.8 | 7.4  | 2.0    | 1.5 | 0.9    |
| 인건비  | 1.7 | 10.1 | 2.1    | 1.0 | 0.5    |
| 실물투자 | 6.0 | 17.0 | 7.4    | 4.7 | 2.6    |
| 취업   | 2.0 | 6.2  | 3.0    | 1.4 | 0.6    |

## [사례 3-3-2] 영국의 생산성 지수상의 표준오차의 제시(ONS,1996)

영국의 생산성지수(loP)는 복합지수로서 저량 변동에 대해 역전현상을 내리고 조정하여 계산된다. 이 지수는 서로 다른 자료원을 결합하여 얻는다. 여기에는 월별 생산조사(Monthly Production Inquiry:MPI),생산자물가지수(Producer Price Index:PPI), 분기별 저량 조사(Quarterly Stocks Inquiry:QSI)를 포함한다. 이러한 이유로 표준오차의 추정치를 얻기가 어렵다. 다음은 1997년에 추정된 결과이다(Kokic, 1998).

〈표 3-3-3〉 제조업의 중분류 loP에 대한 표준오차의 추정치

| 제조업전체 | DA  | DB  | DC  | DD  | DE  | DG  | DH  | DI  | DJ  | DK  | DL  | DM  | DN  |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.8   | 2.5 | 2.1 | 5.3 | 3.7 | 2.6 | 1.2 | 3.5 | 1.9 | 2.0 | 2.3 | 3.4 | 1.8 | 2.8 |

이 추정치들은 테일러 전개방법과 모수적 붓스트랩 방법을 적용하여 얻은 결과이다. PPI들에 대한 가능한 분산에 관한 가정을 만들어야 하고 저량 조정이나 계절조정의 효과를 포함하지 않았다.

<표 3-3-3>은 loP에 대한 12개월 변동을 나타내는 것으로서 loP 자료가 1990년 기준년도 이후 연간단위로 12개월간 조사된 것처럼 취급하여 얻은 것이다. 붓스트랩 모의실험으로부터 제조업에 대한 표준오차의 추정치는 0.82로서 위의 표에 나타난 값에 거의 근접한다. 전체 제조업 지수에 대한 신뢰구간은 근사적으로 1.6%p 이다.

## (1) 표본오차의 평가지표

## [정확성 품질지표 1]

| 지표      | 유형   | 통계출처                | 주기        |
|---------|------|---------------------|-----------|
| A1.변동계수 | 주요지표 | Eurostat,<br>국가통계기관 | 월간/분기별/연간 |

## [변동계수의 정의]

변동계수는 자료의 퍼짐정도를 나타내는 척도로서 평균에 대한 표준편차의 비로 정의된다. 이 값은 생산된 통계의 주요변수에 대해 추정되어야 한다.

## [추정방법]

$$CV = \text{추정량의 표준편차} / \text{추정량} = s(\hat{\theta}) / \hat{\theta}$$

## [필요한 메타자료]

- 추출설계에 대한 설명
- 추정량
- 분산추정을 위해 사용된 방법
- 만일 지역추정이 층별로 다를 경우 이를 기술

## 2. 비표본 오차

## 1) 포함오차

추출틀은 모집단 단위에 접근하기 위한 하나의 틀이다. 조사모집단은 모집단 단위들의 집합으로 추출틀을 통해 접근이 가능하며 조사의 결론은 이러한 모집단에 실제로 적용이 가능하다. **포함오차(Coverage Error)** 또는 **추출틀오차(Frame Error)**는 목표 모집단과 조사모집단 간의 불일치로 발생한다.

다음과 같은 종류의 포함오차를 구별할 수 있다.

- 과소포함
- 과대포함
- 중복 기재
- 부정확한 보조정보

포함오차는 분산에 편향과 과소 추정의 원인이 된다. 조사에서 모든 접촉된 모집단 단위는 단위의 정보가 정확한지에 관한 프레임 정보를 점검해야 한다. 과대포함의 경우에 부정확한 보조정보와 중복기재는 발견할 수 있다. 표본으로 선택된 단위들 중 이러한 문제의 범위는 전체 프레임에 대해 단위들의 범위를 지정할 수 있다. 다른 한편으로 과소포함은 발견하기 어렵다. 프레임의 품질을 검토하기 위한 특수한 점검조직을 가동하거나, 프레임에 존재하지 않는 조사모집단 단위들을 찾기 위한 특별한 조직을 동원해야 한다.

포함오차에 대해 과대포함률과 오분류율과 같은 지표를 통해 직접적으로 프레임의 품질을 평가할 수 있다. 과대포함률은 잘못된 단위들의 보조정보나 상태 때문이거나, 정확한 분류가 대상 밖의 단위를 포함되게 하는 등의 오분류에 의해 조사대상 범위 밖의 단위를 포함하는 경우 발생한다. 과대포함의 다른 측면은 중복에 의해 발생한다. 오분류율은 이러한 단위를 포괄한다. 프레임 오분류율은 주요 층화변수에 대해 계산되어야 한다.

**[사례 3-3-3] 스웨덴 통계청의 모집단 목록의 과대포함 평가(Greijer, 1996 발췌)**

추정치는 배달되지 않고 반송된 설문지를 바탕으로 계산되었다. 1995년 봄철 동안 143,000명에게 설문지가 발송되었고, 이중 14,000부가 배달되지 않은 설문지이고 32,000부가 응답되지 않은 설문지였다. 1995년에 14,000명중 7,500명이 이민을 간 사람들로써 과대포함으로 처리되지 못했다. 나머지 6,500부의 배달되지 못한 설문지는 다양한 배경변수에 기초하여 조사하였다. 발송되지 못한 설문지의 비율 중 가장 큰 비율을 차지하는 그룹은 스톡홀름에 거주하는 남성, 독신, 젊은이, 개인들로 나타났으며, 이들은 대부분 스웨덴 시민들이 아니었다. 과대포함에 대한 추정값은 이용된

추정량에 따라 15,000명에서 30,000명 사이였다.

**[사례 3-3-4] 미국의 전수조사에서 포함률에 대한 평가(Perwitt, 2000에서 발췌)**

2001년 4월1일 전에 Census Bureau는 포함률 측정조사를 포함하여 초기 인구수의 정확성을 개선하기위해 미국 인구에 대한 집계를 마쳤다. ACE(Accuracy and Coverage Evaluation)이라고 하는 포함률 측정조사는 DSE(Dual System Estimation)로 알려진 통계적 방법에 기초하며, 전통적인 방법에서 누락된 개인이나 오류가 발생한 목록을 수정하기위해 고안된 것이다. 인구통계학적 분석 방법은 전국에 걸쳐 2000년 센서스의 모집단 포함률의 완비성을 평가하기위해 사용되며, 이전 센서스의 변동을 평가하는데 사용된다.

연방의원 수 할당을 위해 이용된 방법은 2000년 4월 1일 당시 미국에 거주하는 모든 개인을 올바르게 파악하고, 또한 연방의 종사자수와 그 당시 해외에 살고 있는 부양가족 수를 파악할 목적으로 고안되었다. 이러한 목적은 완전하고, 정확하게 실현되지는 못했다.

〈표 3-3-4〉 인종별 순수 과소계상의 백분율의 인구통계학적 분석 추정치  
(단위:%)

| 연 도       | 1940       | 1950       | 1960       | 1970       | 1980       | 1990       |
|-----------|------------|------------|------------|------------|------------|------------|
| 계         | 5.4        | 4.1        | 3.1        | 2.7        | 1.2        | 1.8        |
| <b>흑인</b> | <b>8.4</b> | <b>7.5</b> | <b>6.6</b> | <b>6.5</b> | <b>4.5</b> | <b>5.7</b> |
| 비흑인       | 5.0        | 3.8        | 2.7        | 2.2        | 0.8        | 1.3        |
| 차이        | 3.4        | 3.6        | 3.9        | 4.3        | 3.7        | 4.4        |

1970년도부터 1990년까지 매 10년 마다 실시되는 센서스 인구수는 제외되어야 할 사람이 포함된 수와 포함되어야 할 사람이 누락된 수이다. 이러한 오류는 과대계상(overcount)을 유발한다. 두 번째로 과소계상(undercount)을 유발한다. 체계적으로 측정된 이러한 오류의 효과는 매 센서스마다 순수한 과소계상으로 나타난다. 즉, 누락된 거주가의 수가 오류로 기재된 수보다 크게 나타난다.

Census Bureau는 1940년 센서스 이후부터 실질적인 과소계상의 차이를 측정하고, 이를 문서화하고 있다.

1940년 센서스 이후 Census Bureau의 통계학자와 대학의 연구자들은 인구통계학적 분석이라는 통계적 기법을 재정의 하였으며, 이 방법은 포함경향과 나이, 성별, 인종별 포함률의 차이를 측정하는 방법이다. 인구통계학적 분석은 센서스와는 독립적으로 전국에 걸쳐 인구수를 추정하기 위해 출생, 사망, 국내이민자, 해외이민자, 병원등록자 수의 기록과 추정치를 사용한다. 이러한 분석은 비록 오류가 없지는 않지만 흑인과 비흑인 인구 간에 존재하는 차별적 과소계상에 내구력을 나타내고 있다.

### (1) 포함률 평가지표

[과대포함과 오분류율 평가지표]

| 지 표           | 유형   | 통계출처                | 주기        |
|---------------|------|---------------------|-----------|
| A2.과대포함과 오분류율 | 보조지표 | Eurostat,<br>국가통계기관 | 월간/분기별/연간 |

[계산방법]

- 추출률의 과대포괄율 = 과대포괄 단위의 수 / (조사대상 단위의 수 + 조사대상외 단위의 수)
- 추출률의 오분류율 = 조사대상이지만 오분류된 단위의 수 / 조사대상 단위의 수

[필요한 메타자료]

- 추출률에 대한 특성값과 기지의 결여성, 증화변수

[참고]

- 최종 추정치에 조사대상이 아닌(out-of-scope) 단위와 조사대상 단위이지만 잘못 분류된 단위의 영향은 추정과정에서 이러한 단위들의 처리에 좌우된다.
- 조사대상 단위의 수 + 조사대상이 아닌 단위의 수 = 적격한 단위의 수이다.
- 적합한 단위들은 단위응답률에서 고려된다. 적합한 단위와 그렇지 않은 단위간

의 구별은 이용 가능한 정보에 대한 단위들을 고려함으로써 프레임의 품질을 쉽게 평가할 수 있도록 한다.

[지리적과소포함을 평가지표]

| 지 표          | 유형   | 통계출처     | 주기        |
|--------------|------|----------|-----------|
| A3.지리적 과소포함을 | 주요지표 | Eurostat | 월간/분기별/연간 |

[계산방법]

- 지리적 과소포함률=비포괄적인 국가에 대한 총 가중치 / 목표모집단에 있는 모든 국가에 대한 총 가중가치

앞에서 언급한 프레임의 포괄성에 관한 지표 이외에 자료의 신뢰성과 안정성을 평가하기위한 지표로서 “개편의 규모”를 제시할 수 있다.

[개편 규모에 대한 정의]

개편은 자료의 신뢰성 또는 안정성을 나타낸다. 그러나 기본적인 개념은 동일하며 초기에 추정된 값과 최종적으로 추정된 값 간의 접근성으로 정의할 수 있다. 신뢰성의 평가는 시간의 흐름에 따라 추정치를 비교하는 것이다. 즉, 신뢰성의 평가는 개편을 나타낸다.

개편의 규모와 방향의 측정은 다양한 방법으로 가능하다. 여기서는 간단하게 개편의 크기를 평가하는 것으로 한정한다.

[개편규모의 평가 지표]

| 지 표          | 유형   | 통계출처     | 주기        |
|--------------|------|----------|-----------|
| A4.개편의 평균 규모 | 주요지표 | Eurostat | 월간/분기별/연간 |

## [계산방법]

주요변수 ( $X$ ) 가 매년 ( $k$ )번 조사대상 시점과 공표시점간의 시간 간격인 사전에 정의된 시차로서 발표된다. 초기 발표는 초기추정치와 나머지 계속적으로 추정치(개편)를 제공한다.

주어진 조사대상 시점에 대해 동일한  $X$ 변수의  $k$ 개의 다른 값  $\{X_1, X_2, \dots, X_k\}$ 가 이용 가능하다. 이때  $X_k$ 는 최종 공표된 값이라 하자.

$$\square \text{ 개편 오차 : } R(i) = X_i - X_k,$$

$$\text{또는 개편오차율 : } r(i) = \frac{X_i - X_k}{X_k}$$

## [참고 사항]

- 오차백분율은 영역별 비교나 회원국 간의 비교에 유용한다.
- 특정 연도에 대해 지표들은 개편의 특성에 대해 정의할 수 있다. 한 가지 가능한 방법으로 평균절대 오차율(mean absolute percentage error:MAPE)로서 다음과 같다.

$$MAPE = \frac{1}{k-1} \sum_{i=1}^{k-1} |r(i)|$$

- 시계열에 대해  $X$ 는  $(X^1, \dots, X^t, \dots, X^T)$ 인 벡터로 정의할 수 있다. 그러면 지표는 다음과 같이 계산된다.

$$\bar{r} = \frac{1}{T} \sum_{t=1}^T MAPE_t$$

여기서  $MAPE_t$  는  $X^t$ 에 대한 평균 절대오차백분이다.

- 시간에 대한 비교가능성에 대해  $k$  시차와 시간벡터의 길이는 각 연도별로 동일하다.

## [필요한 메타자료]

만일 어떠한 측면에서 개편으로 정확성이 개선되었다면, 개편의 평균 회수와 같은 정보를 포함하는 개편 정책, 개편에 대한 주요한 이유 등.

## 2) 측정오차

**측정오차(Measurement Error)**는 자료를 수집하는 과정에서 발생하는 오차로서 특정 변수값으로 기록된 값과 실제값 간의 차이이다. 이러한 원인을 분류하면 다음과 같다.

- 조사도구 : 자료 수집을 위해 사용한 조사표의 형태 또는 조사표, 측정도구 등이 잘못된 값을 기록하게 할 수 있다.
- 응답자 : 의식적이든 무의식적이든 간에 응답자가 잘못된 응답을 할 수 있다.
- 조사원 : 응답자가 제시하는 응답에 조사원이 영향을 끼칠 수 있다.

측정오차는 산출된 통계치의 편향과 변동성의 원인이 된다. 측정 도구와 조사원의 영향을 측정하기 위해 서로 다른 조사도구나 서로 다른 조사원을 이용하여 반복측정이 가능하다. 대안적으로 실험은 부차적으로 추출된 표본에 대해 조사도구나 조사원을 할당하여 수행한다. 응답자 효과는 동일한 응답자에 관한 독립적인 정보를 필요로 하기 때문에 측정하기가 더욱 어렵다. 측정오차 효과의 평가는 확률적인 측정오차 모형을 가능하며, 이 모형은 관심변수의 추정값과 분산 추정값을 도출하기 위해 표본추출모형과 결합하여 모형화 할 수 있다.

자료에 대한 에디팅은 보통 오류로 나타나는 자료의 불일치성을 제거하는 작업이다. 이러한 오류는 처리오차(코딩과 자료입력)가 된다. 각각의 오류에 대한 에디팅된 레코드의 비율은 원 자료의 품질을 평가하는 지표가 된다. 오류율은 이들의 가중평균을 계산하여 전체 자료에 대한 단순비율과 결합하여 얻는다. 이때, 가중값은 각각

의 에디팅된 레코드의 수와 같다. 조사 자료의 수정이나 대체(imputation)는 자료의 불일치성을 제거하기 위해 사용되는 방법이다. 그러므로 오류율은 자료수집과 처리의 품질을 나타내는 지표이며 최종 자료의 품질을 나타내는 지표는 아니다.

#### [사례 3-3-5] 조사원 효과에 대한 보고사례(Berthier and Neros, 1998)

Berthier 와 Neros(1998)는 조사원 효과를 측정하는 방법을 EHPS(European Household Panel Survey)에서 프랑스조사 결과에 적용하였다. 이들의 기본적인 결론은 다음과 같다.

- 조사원들에게 무응답의 형태(접촉 불가, 장기출타, 응답불능 또는 거절)에 대해 자세히 요구하였다. 그 결과 조사원들로부터 분리 가능한 이유를 발견하였고 면접의 지속성이 무응답에 높은 영향이 있음을 발견하였다.
- 조사원 효과는 생활수준을 평가하는데에는 존재하지 않았다. 이러한 효과는 가구소득에 대해 적은 영향을 끼치며, 가구소득과 생활수준간의 일관성에 대해서는 다소 높은 영향을 끼치는 것으로 나타났다.
- 응답자들에게 정확한 소득을 기입하거나 또는 특정 소득집단에 속하거나 하는 두 가지 선택적 질문을 제시하였다. 이 경우 조사원들이 응답자들의 선택에 영향을 끼치는 것으로 나타났다.

#### [사례 3-3-6] 응답의 일관성에 대한 보고사례(Sarndal, 1992)

미국의 1980년도 인구 및 가구에 대한 센서스의 평가 연구결과 중 일부를 인용하였다. 표본가구를 재조사하였고, 동일한 질문에 대해 서로 다른 응답을 하는 경향이 있음을 발견하였다. 다음의 예시 표는 8596개의 표본가구에 대해 “가구원별로 사용하는 자동차는 몇 대입니까?” 라는 질문에 대해 응답한 내용을 정리한 것이다.

〈표 3-3-5〉 응답의 일관성

(단위: 명)

| 센서스   | 재조사   |       |       |      | 계     |
|-------|-------|-------|-------|------|-------|
|       | 없음    | 1대    | 2대    | 3대이상 |       |
| 없음    | 1,050 | 230   | 49    | 6    | 1,335 |
| 1대    | 119   | 3,308 | 618   | 81   | 4,126 |
| 2대    | 13    | 339   | 1,895 | 248  | 2,495 |
| 3대 이상 | 2     | 32    | 171   | 435  | 640   |
| 계     | 1,184 | 3,909 | 2,733 | 770  | 8,596 |

## 3) 처리오차

자료수집과 통계산출물에 대한 분석을 하기 전에 데이터는 코딩, 자료입력, 자료에디팅, 대체 등과 같은 특정한 처리과정을 거친다. 이러한 과정에서 발생하는 오류를 **처리오차(processing error)**라 한다. 이러한 오류들은 측정오차와 유사한 면이 있다.

처리오차는 측정오차와 마찬가지로 산출된 통계값에 대해 편향과 변동을 유발시킨다. 이 값들에 대한 범위와 영향을 평가하기 위해 통계적 실험을 실시해야 한다. 일정수의 조사표 표본에 대해 재코딩과 재입력과 재대체(reimputation)를 수행한다. 이러한 실험을 하는 동안 실제 표본 조사표는 오류에 대해 수정이 가능해진다. 그러므로 이들 조사표는 오류를 계산하기위한 것이 아니라 나머지 조사 자료의 오류에 대한 지표로 제시하게 된다. 자료를 입력하는 동안 온라인 점검은 입력 오류에 대한 로그 자료를 저장할 수 있다. 마지막으로 다중대체방법(multiple imputation)은 각각의 결측자료나 잘못된 자료에 대한 다중대체값을 명확히 제시하고 이들의 변동성을 추정과정에 반영한다. 측정오차에서 설명하였듯이 에디팅의 오류율은 수정 전 원 자료의 품질을 나타내는 지표로서 유용하게 사용될 수 있다.

처리오차와 측정오차에 대한 구분이 항상 가능한 것이 아니기 때문에 품질보고서에는 측정오차와 처리오차가 혼합된 경우에는 반드시 표시되도록 한다.

처리오차에 대한 지표의 범위는 비율의 형태로서 서로 다른 절차의 점검에 대해 계산된다. 여러 지표 중에서 대체율(imputation rate)만을 선택적으로 다루는 이유는

오차를 처리하는데 있어서 중요하기 때문이다.

**[사례 3-3-7] 스웨덴의 농촌패널 센서스의 에디팅 보고 사례(Medin and Wilson, 1994)**

자료입력 후에 개인 농장에 대한 자료의 논리적 검증과 오류검증을 위해 에디팅 과정을 실시하였다. 소수의 사례에서 완전하게 자동 수정을 실시하였지만, 대부분의 나타난 오류사례에 대해서는 스웨덴 통계청 직원들에 의해 수작업으로 실시되었다. 이 과정에서 상황을 명확하게 판단하기위해 개별 농장으로 전화를 걸어 수정작업을 수행하였다. 확정 통계에 대한 미시적인 에디팅의 영향은 서로 다른 변수 간에 큰 차이가 나타났으며, 동일한 변수에 대해 연간자료에 대해서도 많은 차이가 나타났다.

패널 접근의 결과는 개인농장의 구조적 변화를 조심스럽게 조사하고 기록해야 한다. 따라서 미시적 에디팅은 농장 간 농지의 전용을 검토하기위한 특별한 과정으로 보완된다.

거시적 에디팅은 미시적 에디팅과는 다르며 자료의 수집과 자료처리에서 모든 사전 과정을 마지막으로 검토하는 것이다. 거시적 에디팅은 다양한 종류의 집합된 자료를 비교한다. 이전 년도에 대한 통계치를 비교하는데 특별한 관심을 기울인다. 관측된 주요한 변동을 확인차원에서 개인농장을 추적하였다.

**[사례 3-3-8] 코딩과정에 대한 비교(Chenu 와 Guglielmetti, 2000)**

첫 번째는 센서스 조사 후와 1982년 취업자조사 후에 2회에 걸친 조사에 대해 PCS(Profession et Categorie Socio-professionnelle: Occupation and Socio-Professional Category)의 코딩작업을 비교하였다. 이때 센서스 코딩은 Colibri-COding en Ligne des Bulletin du Recensement Individuel: On-line Coding of the Population census forms 프로세스에 따라 반자동 코딩작업이고, 취업자조사 코딩은 하나의 가이드와 알파벳 색인으로 구성된 문서를 이용한 수작업 코딩이다.

다음으로 PCS 코딩과 1997년 시험센서스기간동안 센서스와 관련된 변수를 비교하였다. 이때 센서스자료는 한편으로는 반자동 방식으로 다른 한편으로는 수작업 검토 과정을 거쳤다. 1982년 비교에서 PCS의 두 코딩 처리간의 차이점은 센서스와 취업

자조사에 응한 응답자들의 문구에 기본적인 차이가 나타났고(신고오차: declaration error), INSEE에 의해 수행된 코딩(분류오차:classification error)에서 차이가 발생하였으며, 이들 두 가지 오차가 결합되어 나타났다. 이러한 오차는 빈번히 발생할 수밖에 없었는데, 1982년도에는 새로운 사회-직업분류가 처음으로 적용된 해이기 때문이다. 더욱이 두 가지 과정은 동일한 형태가 아니다. 취업자조사의 설문지는 조사원에 의해 작성되는 반면에 센서스는 응답자에게 맡겨놓고 나중에 조사원이 수집해가는 방식이었다. 또한 센서스는 응답자의 모든 애타심을 필요로 하는 과정이었고, 취업자조사는 취업자에 초점이 맞춰진 조사내용으로 상세한 노동과 실업 상태를 조사하는 것이었다. 1997년에는 동일한 형태를 사용하여 코딩절차상의 변동만을 비교하였다. 다음의 표는 2가지 사회-직업 코딩간의 불일치 비율을 나타내고 있다.

〈표 3-3-6〉 센서스와 1982년 취업자조사

(단위: %)

| 1982년 센서스의<br>사회-직업 그룹 | 1982년 취업자 조사의 분류 (%) |                  |                   |             | 계   |
|------------------------|----------------------|------------------|-------------------|-------------|-----|
|                        | (a) 다른 그룹            | (b)동일그룹<br>다른 분류 | (c)동일 분류<br>다른 직업 | (d)동일<br>직업 |     |
| 1. 농민                  | 5                    | 24               | 28                | 43          | 100 |
| 2. 자영업                 | 13                   | 7                | 17                | 63          | 100 |
| 3. 관리직                 | 24                   | 10               | 19                | 47          | 100 |
| 4. 일용직                 | 19                   | 7                | 21                | 53          | 100 |
| 5. 비서                  | 13                   | 5                | 19                | 63          | 100 |
| 6. 근로자                 | 9                    | 28               | 16                | 47          | 100 |
| 계                      | 13                   | 15               | 19                | 53          | 100 |

〈표 3-3-7〉 1997년 센서스 시험조사의 반자동 코딩절차와 수작업 코딩절차  
(단위: %)

| 반자동<br>코딩에서의<br>사회-직업 그룹 | 수작업 코딩에 따른 분류(%) |                  |                   |          | 계   |
|--------------------------|------------------|------------------|-------------------|----------|-----|
|                          | (a) 다른 그룹        | (b)동일그룹<br>다른 분류 | (c)동일 분류<br>다른 직업 | (d)동일 직업 |     |
| 1. 농민                    | 3                | 24               | 12                | 61       | 100 |
| 2. 자영업                   | 8                | 6                | 12                | 74       | 100 |
| 3. 관리직                   | 11               | 10               | 18                | 61       | 100 |
| 4. 일용직                   | 12               | 7                | 12                | 69       | 100 |
| 5. 비서                    | 8                | 4                | 11                | 77       | 100 |
| 6. 근로자                   | 5                | 2                | 12                | 71       | 100 |
| 계                        | 8                | 8                | 13                | 71       | 100 |

(1) 처리오차의 평가지표

[대체율 평가 지표]

| 지표         | 유형   | 통계출처                | 주기        |
|------------|------|---------------------|-----------|
| A5.대체율과 비율 | 보조지표 | Eurostat,<br>국가통계기관 | 월간/분기별/연간 |

[대체율에 대한 정의]

대체는 에디팅하지 못한 결측치, 부적절하거나 불일치하는 자료에 대해 대체값을 결정하거나 할당하기위해 이용되는 절차이다. 이 절차는 자료가 결측일때 높은 품질의 추정치를 보장하거나 내부적으로 일치성을 가지는 자료를 생산하기 위해 자료 응답자의 값으로 바꾸거나 에디팅된 값을 할당함으로써 이루어진다.

[계산방법]

□ 비가중된 대체율 = 변수  $X$ 에 대해 대체된 자료의 수/ 총 자료수

□ 가중된 대체율 = 대체된 값에 대해 총 가중값/ 모든 최종 값에 대한 총 가중 값

## [필요한 메타자료]

## □ 사용된 대체방법과 가중치 계산 방법

참고로 이전의 지표를 계산하기위해 대체에 의해 변동된 자료를 식별하기위해 자료에 표식을 한 후 원 자료와 함께 저장하는 것이 바람직하다.

## 4) 무응답오차

**무응답(Nonresponse)**은 모든 조사변수에 대해 자료를 수집을 위한 조사를 못한 경우에 발생한다. 표본으로 할당된 모든 모집단 단위로부터 완전한 자료를 얻지 못할 경우에 발생한다. 수집된 자료로부터 계산된 통계값과 결측치가 없는 자료로부터 계산된 통계값 간의 차이가 무응답에 기인한 무응답 오차이다. 무응답의 유형은 다음과 같이 두 가지이다.

- ① 단위 무응답 : 표본으로 할당된 모집단 단위에 관한 아무런 정보를 얻지 못한 경우
- ② 항목 무응답 : 표본으로 할당된 모집단 단위에 대해 일부 조사변수의 값을 얻지 못한 경우

응답의 크기는 다음과 같이 두 가지의 응답률로 측정된다.

- ① 단위응답률: 자료 수집을 위해 할당된 전체 단위의 수로 적어도 일부변수에 대해 자료가 있는 단위들의 수를 나눈 비율이다.
- ② 항목응답률 : 할당된 단위의 총 수로 주어진 항목 또는 변수에 대해 자료를 가진 단위의 수를 나눈 비율 또는 적어도 일부 변수에 대해 자료가 있는 단위들의 수를 나눈 비율이다.

가중응답률(weighted response rate)은 관측치 수 대신 단위들의 표본가중치의 합으로 계산되며, 값-가중응답률은 추출가중치 대신 보조변수의 값을 합한 것이다.

무응답을 평가하기 위해서는 두 가지 주요 지표를 제시할 수 있다. 단위응답률과 항목응답률이다. 또한 대체율을 무응답 오차와 같이 다룰 수 있다. 실제로 대체는 항목무응답을 수정하기위해 사용된다.

통계값에 대해 무응답의 영향은 변동성을 증가시키고 편향을 유발한다. 변동성의 증가는 무응답이 단순히 이용 가능한 응답자의 수를 축소함으로써 발생된다. 또한 편향의 발생은 무응답자가 응답자에 비해 조사변수 상에서 차이가 나기 때문이다. 무응답자의 행위는 대개의 문헌에서 다음과 같이 3가지 유형으로 분류된다.

① MCAR(Missing completely at random) : 완전임의 결측 자료

특정한 변수에 대해 주어지지 않은 누군가의 확률이 모든 모집단 단위에 대해 동일하다는 가정이다. 이는 확률적으로 무응답자와 응답자간에 확률적으로 차이가 없다는 가정이다.

$$P(Y \text{ is missing} | X, Y) = P(Y \text{ is missing})$$

② MAR(Missing at random) : 임의 결측 자료

특정한 변수에 대해 알려지지 않은 누군가의 확률이 해당 변수의 값에 의존하지 않고, 단지 보조변수에 의존한다는 가정이다.

$$P(Y \text{ is missing} | X, Y) = P(Y \text{ is missing} | X)$$

③ MNAR(Not Missing at random) : 비임의 결측 자료

특정한 변수에 대해 알려지지 않은 누군가의 확률이 해당 변수값에 의존한다는 가정이다. 이 경우 결측자료에 대한 확률은 정보적(informative)이라 한다.

통계값을 산출하고 분산을 추정하기 위해 사용된 추정방법은 나머지 무응답과 결측치를 대체하기 위해 사용한 대체방법에 대해 가능한 한 폭넓게 고려해야 한다.

대체에 기인한 변동성을 추정하기 위해 고안된 방법으로는 다중대체(multiple imputation)방법을 사용할 수 있다. 만일 단일대체(single imputation)방법을 적용하여 분산을 추정하는 경우에는 여러 방법을 고려할 수 있다. (a) 모형-보조방법(Sarndal, 1992) (b) 잭나이프 방법(Rao, 1991; Rao and Shao, 19992) (c) 붓스트랩 방법(Shao and Sitter, 1996) (d) 핫덱 대체(Provost, 1995) (e) 모든 단위에 대한 대체방

### 법(Montaquila and Jurnigan, 1997)

무응답자에 대한 확률부차표본방법이나 확률화응답기법과 제시된 자료가 확률적 기법에 의해 통계값이 추정되도록 추정과정에 포함될 수 있다. 만일 자료가 모든 가능한 노력 후에도 여전히 결측이 있다면, 응답모형을 가정하여 무응답 단위에 대한 적절한 확률을 제시하는 적절한 모형가정을 적용할 수 있다. 이러한 방법으로는 단위에 대한 재가중방법이다. 이때 보조변수를 추정과정에 사용할 수 있다.

### [사례 3-3-9] 과수농가 조사에서 무응답(Sarndal, 1992)

북부 캘리포니아의 과수 농가 중 3,116명에게 우편조사를 실시하였으며, 응답률을 높이기 위해 3회에 걸쳐 무응답자에게 계속적으로 우편으로 설문을 배포하였다. 농가당 과수의 수를 가능하면 응답하도록 하였다. 응답자와 무응답자간의 유사성의 평가와 무응답자에 의한 편향을 평가하기위해 다음과 같은 분석표를 작성하였다.

〈표 3-3-8〉 과수농가 조사에서 무응답 분석

| 분석항목        | 우편발송 회수 |     |     | 무응답 | 계   |
|-------------|---------|-----|-----|-----|-----|
|             | 1       | 2   | 3   |     |     |
| 반송률(%)      | 10      | 17  | 14  | 59  | 100 |
| 농가당 평균과수의 수 | 456     | 386 | 340 | 290 | 329 |

### (1) 단위 무응답 평가지표

#### [단위무응답 평가지표]

| 지표       | 유형   | 통계출처                | 주기        |
|----------|------|---------------------|-----------|
| A6.단위응답률 | 보조지표 | Eurostat,<br>국가통계기관 | 월간/분기별/연간 |

#### [단위응답률의 정의]

단위응답률은 백분율로 표현되는 비(ratio)로서 표본에서 적합한 단위의 수에 대한 면접자 수의 비이다. 가중된 응답률은 표본에서 각 단위의 가중치로 포함확률의 역수를 사용한 비를 계산한다. 어떤 경우에는 이 값은 단위의 중요도를 가중치로 사용한다.

간단히 비-가중단위응답률(un-weighted unit response rate)은 표본에서 자료 수집의 결과를 나타내는 반면, 가중된 단위응답률은 관심모집단에 대해 자료수집이 얼마나 잘 이루어 졌는가를 나타낸다.

대부분의 조사에서 적격성이 알려지지 않은 표본단위가 존재한다. 왜냐하면, 자료수집기간동안 이 단위들의 적격성을 보장할 수 없기 때문이다. 이러한 적격성을 알 수 없는 모든 단위들은 응답률 계산에서 조사대상(in-scope) 단위로 취급한다. 이것은 모든 단위를 비조사대상(out-of-scope)단위로 취급하거나, 적격성이 알려진 단위에 대해 동일한 비율로 존재하는 조사대상 단위의 수를 추정하는 것에 비해 보수적인 응답률 추정치를 제공한다.

[계산방법]

- 비-가중된 단위응답률=추정에 사용된 응답단위의 수/(조사대상 단위의 수+적격성이 미지의 단위의 수)
- 가중된 단위응답률 = 추정에 사용된 가중된 응답단위의 수/(가중된 조사대상 단위의 수 + 적격성이 미지인 단위의 가중된 수)

계층적 자료구조를 갖는 조사에 대한 부분 무응답율(partial nonresponse rate)은 보조적인 지표이다. 예를 들어 가구조사에서 최종 추출단위는 개인이다. 가구의 거의 대부분은 응답을 했음에도 불구하고, 일부가 응답을 거절할 수 있다. 유사한 방법으로 사업체 조사에서 특정한 지역단위의 자료가 결측인 반면, 주요 회사들은 적절한 자료를 제공한다. 이러한 지표의 계산 방법은 응답한 부차 단위의 수를 적절한 부차단위의 총 수로 나누어 계산할 수 있다.

참고로 응답율의 가중치는 (1) 가구조사에서 추출확률의 역수 (설계가중치)가 되

며, (2)사업체조사에서는 가장 관련된 변수를 가중치로 고려한다. 왜냐하면 일부 사업체는 모집단의 다른 사업체 또는 부차모집단(예, 산업)보다 더 큰 비율을 차지하기 때문이다.

[필요한 메타자료]

지표의 해석은 제공된 메타자료에 많이 좌우된다. 유용한 메타자료는 다음과 같다.

- 단위들의 다양한 분류기준에 대한 정의(응답자, 조사대상 단위 등)
- 보조변수를 포함한 적절한 가중
- 자료수집 방법
- 단위교체를 적용했는지 여부
- 재가중을 포함하여 대체절차방법

다음으로 정확성 지표중 하나인 항목응답률에 대해 살펴보기로 한다.

(2) 항목 무응답 평가지표

[항목무응답 평가 지표]

| 지표        | 유형   | 통계출처                | 주기        |
|-----------|------|---------------------|-----------|
| A7. 항목응답률 | 보조지표 | Eurostat,<br>국가통계기관 | 월간/분기별/연간 |

[항목 무응답률의 정의]

항목응답률은 항목에 응답한 적격한 응답단위의 수에 대해 항목에 응답한 적격한 단위들의 수의 비이다. 이 값은 하나의 변수(항목)를 나타내며, 주요변수에 대해 계산되어야 한다.

하나의 질문에 대한 응답은 다음과 같이 분류할 수 있다.

- 값(0을 포함하여 단위가 제공하는 질문에 대한 응답)
- 결측값

어떤 시스템에서는 0값과 결측값 구별이 불가능한 경우도 있으며, 이 경우 항목무응답률을 계산할 수 없다.

[계산방법]

- 비-가중된 항목응답률=항목값이 있는 단위의 수/항목에 대해 조사대상 단위의 수
- 가중된 항목응답률=항목값이 있는 단위의 가중 수/항목에 대해 조사대상 단위의 가중 수

사업체 조사의 경우 다음과 같은 공식을 이용하여 유의미한 값을 계산할 수 있다.

- 가중된 항목응답률=항목값이 있는 단위의 총 가중값/항목에 대해 총 가중 추정치

[필요한 메타자료]

- 단위무응답과 유사한 자료와 더불어 설문지의 내용과 관련된 응답자 부담에 대한 자체진단(민감한 질문의 제시, 길이 등)

#### 5) 모형가정오류

통계적 모형은 추정을 위해 필요할 때가 종종 있으며 조사의 추정단계에서 사용된다. 모든 모형화 작업, 즉, 모형선택은 관련된 자료의 수집과 모형의 모수에 대한 추정은 모형의 모수적 가정으로부터 추정에 필요한 가정까지를 모두 포함한다. 만일 어떤 가정이 위배되면, 통계값의 정확성에 영향을 끼치게 된다.

**모형가정오류(Model assumption error)**는 최종적인 통계값에 편향을 유발하며

더욱이 이들의 적절성에 관한 불확실성뿐만 아니라 모수의 추정량의 변동성은 통계값의 분산증가로 나타난다. 모형화 과정에서 사용된 가정을 전체적으로 검토해야 하며, 민감성 분석을 통해 가정에 대한 통계값의 완고성(robustness)을 점검해야 한다. 모형 추정에 사용된 추가적인 자료가 미지의 정확성을 가진 자료이면 오류-내제 변수모형(error-in-variable model)을 사용하는 것이 바람직할 것이다. 조사 통계값의 추정에는 반드시 사용된 모형에 기인한 변동을 고려해야 하고 다음으로 질적인 평가가 이루어 질 수 있다.

**[사례 3-3-10] NMAR자료를 처리하기 위해 이용한 모형에 대한 평가 사례 (Draper and Bowater, 1999b)**

무시할 수 없는 무응답(non-ignorable nonresponse)에 대해 Forster와 Smith(1998)은 1992년 영국의 일반 선거 패널조사(British general election panel survey)에 참여한 1,242명의 원 표본을 추적 조사한 결과를 발표하였다. 21명이 응답하지 않았고, 86명이 투표하지 않겠다고 하였다. 나머지 1,135명 중 보수당에 투표한 비율이 44.1%, 노동당이 32.2%, 자유민주당이 21.0%, 기타 2.8%로 타나났다. 이들 중 317명이 원래의 조사에서 무응답자였다(원래 조사에서는 41.0%, 25.6%, 30.0%, 3.5%로 나타남).

결과적으로 한 방향으로 투표하였다고 보고된 원래 무응답자들은 자유민주당에 대한 전체 비율이 패턴 혼합모형으로부터 95%신뢰구간 밖에 놓일 것이라고 예측한 전문가들보다 자유민주당에 대해 전혀 예측할 수 없을 정도로 멀리 떨어졌다. 이러한 사실은 전문지식에 근거하여 모형화한 가정이 정당하다 할지라도 모형화에서 전혀 예측하지 못한 사실들이 발생할 수 있음을 나타낸다.

## 제4절 시의성과 정시성

### 1. 시의성과 정시성의 정의

통계 또는 정보의 **시의성(Timeliness)** 정보의 이용성과 그 정보가 나타내는 사건 또는 현상간의 시간간격을 나타낸다. 시의성은 상대적으로 쉽게 그리고 직접적으로 계산된다. 통상적인 측정 방법은 평균산출 시간이다. 최대산출시간은 가장 나쁜 경우로서 나타난다.

**정시성(punctuality)**은 자료의 공표날짜와 자료를 공표해야할 목표날짜 간, 예를 들어 자료를 공표한다고 발표한 공식적인 날짜 또는 규정에 적시되었거나 사전에 이용자와 협의한 날짜 간의 시차를 나타낸다.

정시성과 시의성은 공표되는 통계값의 주기와 관련되어 있다. 예를 들어 월별자료는 조사대상월 이후에 너무 많은 월이 있기 때문에 이용할 수 없고, 그 외에는 관심을 상실하게 된다. 그러므로 주기성에 대해 산출 시간의 비율을 계산함으로써 통계의 주기성으로 산출 시간을 기준으로 삼는데 관심을 가진다. 따라서 이러한 통계는 서로 다른 주기를 가진 조사들 간의 비교가 가능하다. 이 경우 품질 표준을 설정하여 척도로 이용할 수 있다. 실제 산출시간을 표준시간으로 나누거나 표준에 대한 차이의 비율로 실제 산출시간을 나눈 비율로 측정이 가능하다. 실제 산출 시간과 목표 시간 간의 차이를 척도로 이용할 수도 있다.

**자료의 신선도(Data Freshness)**는 조사 주기와 시의성을 결합한 다른 체계가 있다면, 현재 자료가 얼마나 신선한 가를 나타낸다. 대부분의 경우 통계자료는 특정한 개발을 위해 주기적으로 공표한다. 이 경우 이용자를 위해 중요한 점은 가장 최근에 이용가능한 통계의 현재와 조사대상시점간의 시차이다. 만일 고정된 통계생산 시점을  $t$  라 가정하면, 통계의 신선도는 결코  $-t$ 를 초과할 수 없고, 하한 값은  $-(t+T)$  이다. 이때,  $T$ 는 조사주기이다. 유용한 지표로서 연간 임의의 한 시점에서 자료의 신선

도의 평균은  $F=-(t+T/2)$  이다.

### 1) 시의성 평가지표

[시의성 평가지표1]

| 지 표                                | 유형   | 통계출처     | 주기        |
|------------------------------------|------|----------|-----------|
| T1. 조사기간의 최종날짜와<br>잠정치의 발표 날짜간의 시차 | 주요지표 | Eurostat | 월간/분기별/연간 |

[계산방법]

조사기간의 최종날짜와 잠정치의 날짜간의 시차= 잠정치 발표날짜-자료의 조사대상  
날짜

다음으로 시의성에 관한 추가적인 지표로서 조사기간의 최종날짜와 확정치 발표  
날짜간의 시차를 측정하는 방법이 있다.

[시의성 평가지표2]

| 지 표                                | 유형   | 통계출처     | 주기        |
|------------------------------------|------|----------|-----------|
| T2. 조사기간의 최종날짜와<br>확정치의 발표 날짜간의 시차 | 주요지표 | Eurostat | 월간/분기별/연간 |

[계산방법]

조사기간의 최종날짜와 확정치의 날짜간의 시차= 확정치 발표날짜-자료의 조사대상  
날짜

[필요한 메타자료]

시의성과 정시성에 관해 위에서 언급한 지수에 대해 만일 소관통계기관에 자료를

접수한 날짜, 자료처리, 적합성과 품질관리, 참조한 데이터베이스의 이용가능날짜, 발간물의 이용가능날짜, 발간 날짜 등에 관한 정보가 완전하면 도움이 될 것이다.

## 2) 정시성 평가지표

[정시성에 관한 지표]

| 지 표                  | 유형   | 통계출처     | 주기        |
|----------------------|------|----------|-----------|
| T3. 효과적인 발간 시간표의 정시성 | 주요지표 | Eurostat | 월간/분기별/연간 |

[계산방법]

효과적인 발간시간표의 정시성=효과적인 발간물의 실제 날짜-효과적인 발간 계획일

## [사례 3-4-1] 시의성에 대한 국제비교 보고사례 (Deutsche Bank, 2000)

〈표 3-4-1〉 주요 통화지수의 이용성

| 지 수                     | 유럽지역(ECB) |                     | 미국(연방은행)         |                        | 일본(일본은행) |                    |
|-------------------------|-----------|---------------------|------------------|------------------------|----------|--------------------|
|                         | 주기        | 지연 <sup>1)</sup> 일수 | 주기               | 지연 <sup>1),2)</sup> 일수 | 주기       | 지연일수 <sup>1)</sup> |
| <b>물가동향</b>             |           |                     |                  |                        |          |                    |
| · 소비자물가                 | 월간        | 20                  | 월간               | 16                     | 월간       | 28                 |
| · 생산자/도매물가              | 월간        | 36                  | 월간               | 13                     | 월간       | 10                 |
| · 임금 <sup>3)</sup>      | 분기별       | 95                  | 월간               | 5                      | 월간       | 30                 |
| · 수입물가                  | -         | -                   | 월간               | 12                     | 월간       | 10                 |
| <b>1) 자본창출</b>          |           |                     |                  |                        |          |                    |
| · 현금순환                  | 월간        | 30                  | 주간               | 1                      | 월간       | 3                  |
| · 자본조달 M3 <sup>4)</sup> | 월간        | 30                  | 주간               | 10                     | 월간       | 15                 |
| · 은행대출                  | 월간        | 30                  | 주간               | 9                      | 월간       | 11                 |
| <b>2) 경제활동</b>          |           |                     |                  |                        |          |                    |
| · 소비자신용                 | 월간        | 7                   | 월간               | -3                     | 분기별      | 24                 |
| · 산업신용                  | 월간        | 7                   | 월간 <sup>5)</sup> | 2                      | -        | -                  |
| · 주문매상고                 | -         | -                   | 월간               | 26                     | 월간       | 42                 |
| · 건설허기 <sup>6)</sup>    | -         | -                   | 월간               | 18                     | 월간       | 30                 |
| · 실제GDP                 | 분기별       | 73                  | 분기별              | 28                     | 분기별      | 71                 |
| · 개인저축                  |           |                     | 월간               | 28                     | 월간       | 35                 |
| · 산업생산                  | 월간        | 59                  | 월간               | 15                     | 월간       | 28                 |
| · 소매                    | 월간        | 65                  | 월간               | 13                     | 월간       | 26                 |
| · 실업률                   | 월간        | 37                  | 월간               | 5                      | 월간       | 29                 |
| <b>3) 공공금융</b>          |           |                     |                  |                        |          |                    |
| · 총부채수지                 | 연간        | 100                 | 월간               | 20                     | 연간       | >200               |
| · 총 부채                  | 연간        | 100                 | 월간               | 6                      | 연간       | >200               |
| <b>4) 외부경제</b>          |           |                     |                  |                        |          |                    |
| · 무역수지                  | 월간        | 60                  | 월간               | 50                     | 월간       | 24                 |
| · 국제수지                  | 월간        | 60                  | 분기별              | 76                     | 월간       | 42                 |
| · 공적예비비                 | 월간        | 30                  | 주간               | 4                      | 월간       | 4                  |

1) 지연(delay)은 보고기간 종료로부터 날짜수를 나타낸다. 2000년 12월 7일을 기준으로 지연날짜를 나타냄.

2) (days) 날짜음수는 보고기간동안 수집된 자료를 나타내며, 보고기간종료까지 걸린 날짜수.

3) 일본의 임금

4) 일본의 M2+CD

5) NAPM 보고

6) 일본에서 주택건축착공을 나타냄.

## 제5절 접근성과 명확성

접근성과 명확성은 단순성을 나타내며, 기대되는 형식으로 통계를 얻고 허용 가능한 시간 내에 적절한 이용자 정보와 도움을 통해 단순하고 이용자 편의 과정을 통한 이용자가 통계에 쉽게 접근할 수 있는지를 나타낸다. 궁극적으로는 이용자가 통계를 이용할 수 있도록, 통계의 이용을 최대한 허용하는 것이다.

### 1 접근성과 명확성의 정의

**접근성(Accessibility)**은 이용자가 통계에 접근할 수 있는 물리적 조건을 말한다. : 통계의 배포경로, 주문절차, 배급을 위해 소요되는 시간, 가격정책, 시장조건(저작권), 마이크로 또는 매크로 자료의 이용성, 매체(종이, CD-ROM, 인터넷 등).

접근성에 대한 평가는 다양한 형태를 가질 수 있는데, 다양한 측면에서 공표자료에 대해 평가가 가능하기 때문이다. (a) 공표 경로, (b) 이용자가 통계산출물을 얻기 위한 용이성 (c) 이용 가능한 자료형태, (d) 가격정책 등이다.

**명확성(Clarity)**은 통계정보의 환경을 나타낸다. : 통계를 제공하는 적절한 메타자료(문자정보, 설명, 문서 등), 그래프, 지도 등, 통계품질에 관한 정보; 국가통계기관에 의해 사용자에게 제공되는 보조자료 등.

명확성은 더욱 평가하기 어려우며, 명확성은 메타자료가 통계산출물과 함께 공표되기 때문에 통계메타자료의 품질과 관련되어 있다.

캐나다 통계청에서 사용하고 있는 품질체계에서 명확성을 통계메타자료의 관련성으로 나타내며, 메타자료의 이용자요구에 대한 만족도로 평가한다. 이에 대한 평가절차는 평가과정에서 수반되는 정보에 대해 생산자와 향후 이용자를 위한 정보의 적합성과 적절성을 평가하기 위해 이용자로부터 정보를 수집한다. 각각의 산출물에 대한 표준적인 방향은 제공되는 메타자료가 완전한지를 질적인 방법으로 평가하는

데 도움이 된다.

[사례 3-5-1] ECHP(European Community Household Panel)의 평가 보고사례  
(Eurostat 의 ECHP에 대한 국제품질 보고서 발췌)

1. ECHP의 접근성

1.1 사용자 데이터베이스

비밀보호 때문에 사용자 데이터베이스는 다양한 익명성 기준에 부합 하기위해 PDB(Production Data Base)로부터 생산된다. 비록 이러한 익명화된 원자료(micro-data)에 직접 접근은 자료이용에 대한 엄격한 조건을 규정하고 있는 계약에 의해 제한이 필요하지만, 만일 이러한 모든 기준을 만족하면 ECHP 자료는 통계법의 측면에서 비밀해제가 되기 때문에 보다 광범위하게 이용할 수 있다.

1998년 12월에 Eurostat는 매우 자세한 문서를 포함하여 익명화된 사용자 DB를 CD-ROM으로 제작하였다. 자료이용에 대한 엄격한 조건을 규정하고 있는 계약에 응한 관심 있는 개인들과 기관들이 연구목적으로 CD-ROM을 구입할 수 있다. 이러한 제한은 DB에 수록된 개인단위들의 구별 또는 그렇게 하려고 하는 이용자의 요구를 막기 위한 것이다. 공인되지 않은 제3자에 대한 재배포는 허용하지 않는다. 계약조건을 위반할 경우 다양한 벌칙과 불이익을 가하게 된다.

1.2 Eurostat에 의한 발간물

Eurostat에서 발간한 보고서들을 먼저 일반에게 널리 보급하기위한 4~8페이지의 보고서로 1997년 5월, 6월, 9월, 12월과 1998년 6월, 11월, 12월에 각각 보고서를 발간하였다. 이와 동시에 통계연보(Statistical Yearbook, 1998), 수치로 본 유럽(Europe in figures, 1998)등을 발간하였다. 또한 방법론적인 보고서로 "ECHP: volume 1 - Survey methodology and Implementation(1996)" 등을 발간하였다.

결과보고서로서는 1995년 웨이브 결과를 제시한 보고서가 1999년 첫 학기동안 발간 예정이었다. 주요결과에 대한 통계표가 매년발간 된다. ECHP 방법론에 대한 완전한 기술적 보고서는 PAN 문서형식으로 이용가능하다.

## 2. ECHP의 명확성

### 2.1 문서 내용

결과를 제시하는 다양한 문서와는 별도로 Eurostat은 기술보고서를 제공한다. 지금까지 100가지의 PAN 문서가 만들어졌다. 이 문서들은 조사에 대한 정의, 초기 설문지, 조사기관의 지침, 표본추출, 자료수집, 품질관리, 추적원칙, 응답률계산, 패널조사 수행의 다양한 측면, 가중치산정과 대체과정의 기술적 내용 등 다양한 주제로 구성된다.

### 2.2 정보서비스

Eurostat는 이용자에게 다음과 같은 정보서비스를 제공한다.

- ECHP에서 이용 가능한 통계를 잘 알고, 이해하고, 분석할 수 있도록 이용자를 돕는다. 이러한 프로젝트는 내부의 국가통계기관과 외부 위원회 이용자들 간 다수회의를 통해 이루어진다.
- 이용자들의 요구가 무엇인지를 알기위해 통계이용자와의 회의 개최.
- 종이나 디스켓으로서 위원회, National Data Collection Units, 자료상점, OECD 등에 제공되는 통계의 계산.

### 2.3 명확성에 대한 이용자 환류

ECHP에 관한 이용자 만족도 조사가 수행된 적이 없으며, 따라서 이용할 수 없다.

통계영역에 따라 지표로서 “공표된 통계와 함께 완전한 메타자료 정보의 수”를 제안하기도 한다. 이러한 지표를 산출하기 위해서는 연관된 메타자료의 정의가 필요하다.

다음의 지표는 이용자에 대한 메타자료의 이용범위에 관한 정보를 제공하며, 통계적 범주나 주제에 대해 발간된 메타자료의 완비성을 나타낸다.

#### 1) 접근성의 평가지표

## [접근성 평가지표1]

| 지 표             | 유형   | 통계출처     | 주기        |
|-----------------|------|----------|-----------|
| AC1. 발표된 발간물의 수 | 주요지표 | Eurostat | 월간/분기별/연간 |

위의 접근성 지표는 서로 다른 방법으로 쉽게 취합할 수 있다. 다음 중 하나를 선택하여 자료를 공표할 수 있다. (1) 발간물(통계연보, 월간 회보, 소식지, 초점통계(Statistics in Focus), 보도자료, 그 외 주제별 발간물), (2) 디스켓, (3) CD-ROM

발간물과 관련하여 공표된 발간물(종이 또는 인터넷으로부터 내려받기)과 판매된 발간물 둘 다에 대해 지표를 계산하는 것이 바람직하다. 참조 데이터베이스에 대한 접근성과 관련하여 다음과 같은 지표를 제안할 수 있다.

## [접근성 평가지표2]

| 지 표                 | 유형   | 통계출처     | 주기        |
|---------------------|------|----------|-----------|
| AC2. 데이터베이스에 접근한 회수 | 주요지표 | Eurostat | 월간/분기별/연간 |

데이터베이스의 유형에 따라 만일 데이터베이스에 대한 접근은 자료의 일치 회수 또는 자료의 추출회수를 계산할 것인지를 결정해야 한다.

## [필요한 메타자료]

자료공표에 대한 통계적 방법을 다룬 목록을 제공해야 한다.

## 2) 명확성의 평가지표

## [명확성 평가지표]

| 지 표                             | 유형   | 통계출처     | 주기 |
|---------------------------------|------|----------|----|
| AC3. 발표된 통계에 대해 메타자료 정보의 완비성 비율 | 경험지표 | Eurostat | 연간 |

## [계산방법]

메타자료에 관한 평가서에 포함된 서로 다른 측면에 대한 메타자료의 이용가능성을 합하고 (예: 완전하게 이용 가능함=2, 부분적으로 이용 가능함 =1, 이용할 수 없음 =0), 이 값을 제공되어야할 메타자료에 관한 연관성 측면에 관한 총계로 나누어 다음과 같이 계산된다.

$$\text{메타자료의 완비성 비율} = \frac{\sum_{j=1}^M \text{이용 가능한 메타자료}}{\sum_{j=1}^M \text{관련된 메타자료}}$$

여기서  $j$ 는 고려된 메타자료의 양태이다.

## 제 6절 비교가능성

## 1 비교가능성의 정의

**비교가능성(Comparability)**의 목표는 지리적 영역, 비지리적 영역, 또는 시점들 간에 통계를 비교하여 적용된 통계적 개념과 정의의 상이성의 효과를 측정하는 것이다. 통계값들 간의 차이가 통계적 특성에 대한 참값들 간의 차이에 기인한다고 할 수 있다. 두 통계값이 비교가능성이 없다는 사실은 통계를 산출한 조사에 원인이 있다. 이러한 특성은 두 가지 주요한 분류로 그룹화 할 수 있다. (a) 조사의 개념, (b) 측정과 추정의 방법

□ **개념(Concepts)** : 조사를 계획하기 위해 다수의 속성들을 미리 정의하는 것을 말한다. 그러한 속성들은 조사대상 모집단, 특성들, 모집단의 분류 그룹들 등이 포함된다. 만일 두 개 국가에서 동일한 속성의 두 가지 조사가 동일한 정의를 사용하지 않는다면, 이들의 통계 산출물은 비교가능성이 결여된 것이다. 결여의 정도는 정의상의 차이에 직접적으로 연관된다.

예를 들어 A와 B국가에서 실업률을 측정한다고 가정하자. 두 국가 모두 모집단으로 센서스자료를 사용하고, 이중 실업자의 수를 계상한다고 하자. 더욱이 두 국가는 경제활동 상태를 산업 중 에서 몇 개의 항목으로 분류하여 이용한다. 두 국가는 전체 실업률과 분류별 실업률통계를 산출한다고 하자. 만일 연구자가 두 국가의 산업 부문별 실업률을 비교하고자 할 때 두 국가가 다음과 같은 분류기준을 이용한다면, 문제가 발생하게 된다.

〈표 3-6-1〉 비교가능성의 예

| 속성      | A국가                                       | B국가  |
|---------|---|--|
| 경제활동 인구 | 16~65세의 모든 취업자 또는 구직자                     | 18~65세의 모든 취업자 또는 구직자                          |
| 실업자     | 조사기간 동안 직장을 구하거나 취업의사가 있는 일을 하지 않는 경제활동인구 | 조사대상 주간에 2주간 직장을 구하거나 취업의사가 있는 일을 하지 않는 경제활동인구 |
| 산업부문    | 제조업                                       | 제조업, 광업 및 채굴업, 전기, 가스 및 수도업                    |

명확하게 실업률은 완전히 비교불가능하게 된다. A국가는 넓은 의미의 경제활동 인구에 대한 정의를 사용하고 있고, B국가는 상대적으로 협소한 정의를 사용하고 있다. 만일 실업률에서 비교가능하다 할지라도 여전히 산업분에서는 서로 다른 분류를 사용하고 있기 때문에 비교불가능하다.

□ **측정/추정방법(Measurement/Estimation)** : 측정관점은 측정방법과 자료수집방법과 관련된 실질적인 분석방법을 포함하며, 엄격히 표준화 되어야 하기 때문에 비교에서 측정 편향을 통제할 수 있다. 추정의 관점은 다양한 수행적 관점이 포함되며, 가중작업, 추정, 통계적 분석 등이 포함될 수 있다. 일반적으로 이러한 작업은 비교에서 개체모집단이 조건이나 요구에 맞게 유연하게 선택해야 한다. 무엇이 필요한지는 동일한 과정이 아니라 다음과 같은 통상적인 표준과정에 따른다.

다음은 통계값 들의 비교성에 영향을 주는 관점을 보다 세밀하게 분류하였다.

1. 개념
  - 1.1 통계적 특성값
  - 1.2 통계적 측정(지표)
  - 1.3 통계 단위
  - 1.4 목표모집단
  - 1.5 조사모집단
  - 1.6 조사대상 기간과 빈도
  - 1.7 연구영역
  - 1.8 지리적 포함률(시간에 대한 비교가능성을 위한)
  - 1.9 표준
    - 1.10 구조적 효과
    - 1.11 연구영역에 대한 개념적 관점
2. 측정
  - 2.1 표본설계
  - 2.2 자료수집
  - 2.3 자료처리
  - 2.4 추정
  - 2.5 연구영역에 대한 측정관점

다음은 비교가능성의 종류를 분류한 것이다.

**지리적 비교가능성(Geographically Comparability)**

동일한 현상을 측정하는 유사한 조사들은 서로 다른 통계 기관에 의해 수행되고, 서로 다른 지리적 개체로 모집단을 나타내는 유사한 조사들 간의 비교성의 정도를 나타낸다.

**시간에 대한 비교가능성(Comparability over time)**

두 조사 시점간의 비교가능성의 정도를 나타낸다.

**영역 간 비교가능성(Comparability between Domains)**

서로 다른 통계영역에서 유사한 목표 특성을 가진 서로 다른 조사간의 비교가능성의 정도를 나타낸다.

**위의 성질들의 결합(Combination of the above)**

비교가능성의 문제가 쌍으로 발생하는 경우로서 예를 들어 예측치들을 비교하기 위해 여러 국가의 시계열을 사용하고자 할 때 발생한다.

## 2. 지리적 비교가능성

**지리적 비교가능성**의 절대적 측정에 가장 가까운 것은 통계수치를 가장 적절한 표준(golden standard)과 비교하는 것이다. 이는 유럽적 규범이나 단일 국가의 기준적인 조사가 될 수 있다. 만일 다수의 국가의 통계수치가 있다면, 이들 통계값들 간의 모든 쌍을 비교하고 이들을 요약함으로써 비교가능성을 절대적으로 평가할 수 있다. 통계값들은 이 값들이 산출된 조사의 메타자료를 대조함으로써 비교할 수 있다. 첫 번째 단계에서는 각 국가별로 필요한 메타자료를 수집하여 비교가 불가능성의 원인이 될 수 있는 측면들을 포괄적인 리스트를 결정하는 것이다. 메타자료상의 상이함은 양적으로 결정되어야 한다. 한 가지 효과적인 방법은 민감도 분석과 모의 실험을 이용하는 것이다. 다른 방법은 각각의 데이터셀에 대한 일련의 정의를 적용하는 것이다. 개념적으로 간단한 접근 방법은 스코어링 체계를 고안하는 것으로서 비교가능성에 이들의 효과에 따라 각각의 서로 다른 메타자료에 점수를 할당하는 것이다. 결과적으로 이 점수들을 합산하여 두 국가의 메타자료 리스트 중 얼마나 다른지를 알 수 있다. 쌍방비교에 대해 스코어링 체계는 행과 열로 정의되는 국가별 2차원 대칭행렬을 구성하게 된다. 각 셀의 원소는 두 국가의 점수로 비교된다. 한 지

역의 전체점수를 계산하기위해 행 또는 열에 있는 점수들을 더하거나 평균을 구하면 된다. 이 행렬은 향후 분석에서 입력 자료로 활용될 수도 있다. 예를 들어 다차원 척도분석(MDS)을 적용하여 국가별 차이의 상대적 크기가 방법론적 차이에 비례함으로 2차원의 국가별 산점도를 작성할 수 있다. 지도상에서 가장 근접한 두 국가가 다른 국가의 통계값 보다 더 비교가능성이 있게 된다. 지리적비교가능성은 생산된 통계의 주요변수에 대해 개념과 측정에서 유럽 기준과 차이점들을 계상함으로서 평가할 수 있다.

**대조통계(Mirror Statistics)**는 다른 차원에서 통계값의 지리적 비교가능성을 평가하는 또 다른 방법은 대조통계를 이용하는 것이다. 대조통계는 유사한 통계적 특성값들을 서로 다른 소관부서에 의해 측정될 때 얻을 수 있다. 주로 이러한 통계는 외국무역, 관광, 운송통계 와 같은 유동성을 측정할 때 이용된다. 예를 들어 얼마나 많은 수의 관광객이 오스트리아로부터 그리스로 유입되는지를 보고하고, 오스트리아는 그리스로 떠난 관광객 수를 보고하면 서로 대응된 통계가 된다. 전형적으로 이 통계들은 2개의 행렬로 구성되는데 하나는 국내 유동성을 나타내며, 다른 하나는 국외 유동성이다. 국가별 한 쌍에 대해 국내외 유동성의 절대적 차이는 상이성에 기초한 지표를 얻어 각 국가에 대해 합하여 구한다. 이 값은 비교대상 국가와 그 외 국가의 비교가능성의 지표로 해석된다.

대조통계가 있는 영역에서 쌍을 이룬 두 국가에 대해 국내외 간 차이를 측정으로 지리적비교가능성을 평가할 수 있다. 양적인 측정뿐만 아니라 회수로서, 대조통계가 이용 가능한 영역에 대해 국가자료의 품질 수준을 평가하는 중요한 지표이다.

#### (1) 지리적 비교가능성의 평가지표

[지리적 비교가능성 평가지표1]

| 지 표                     | 유형   | 통계출처     | 주기 |
|-------------------------|------|----------|----|
| C1. 개념과 측정에서 유럽기준과의 차이율 | 경험지표 | Eurostat | 연간 |

[계산방법]

$i(i = 1, \dots, N)$ 를 비교대상국가라 하고,  $j(j = 1, \dots, M)$ 을 비교가능성 결여의 원인이라 하자. 변수  $D_{ij}$ 를 다음과 같이 정의하자.

$$D_{ij} = \begin{cases} 1, & j\text{원인에 따른 } i\text{번째 국가에 대해 차이가 있으면,} \\ 0, & \text{차이가 없으면,} \end{cases}$$

특정한 시점에 대해 다음과 같은 표를 유럽기준과의 차이  $D_{ij}$ 를 산출할 수 있고, 각 국가와 원인에 따라 기록된다.

|    |     | 원인       |          |     |          |     |          |
|----|-----|----------|----------|-----|----------|-----|----------|
|    |     | 1        | 2        | ... | j        | ... | M        |
| 국가 | 1   | $D_{11}$ | $D_{12}$ |     | $D_{1j}$ |     | $D_{1M}$ |
|    | 2   | $D_{21}$ | $D_{22}$ |     | $D_{2j}$ |     | $D_{2M}$ |
|    | ... |          |          |     |          |     |          |
|    | i   | $D_{i1}$ | $D_{i2}$ |     | $D_{ij}$ |     | $D_{iM}$ |
|    | ... |          |          |     |          |     |          |
|    | N   | $D_{N1}$ | $D_{N2}$ |     | $D_{Nj}$ |     | $D_{NM}$ |

그러면 주요변수에 대한 차이점 수와 비율을 다음과 같은 공식으로 구할 수 있다.

□ 차이점의 수 :  $\sum_{i=1}^N \sum_{j=1}^M D_{ij}$

□ 차이율 :  $\frac{\sum_{i=1}^N \sum_{j=1}^M D_{ij}}{NM}$

만일  $i$  국가에 대해  $D_{i.} = \sum_{j=1}^M D_{ij} \neq 0$  이면 적어도 하나의 원인이 유럽기준과 차이가 있음을 나타낸다. 상대적으로 큰  $D_{i.}$  는 지리적 비교가능성의 결여가 있을

가능성이 있음을 나타낸다.

이와 유사하게 특정한  $j$ 원인에 대해  $D_{.j} = \sum_{i=1}^M D_{ij} \neq 0$ 이면 적어도 하나의 국가가 유럽기준과 차이가 있음을 나타낸다. 상대적으로 큰  $D_{.j}$ 는 영역간 비교가능성의 결여가 있을 가능성이 있음을 나타낸다.

[필요한 메타자료]

차이에 대한 설명과 차이의 원인은 지표를 해석하는데 유용한 정보가 될 수 있다.

[참조]

비지리적 영역간의 비교가능성 척도(경제활동, 규모에 따른 분류, 생산품, 운송수단, 성별 등)는 지리적 비교가능성에서 사용된 방법과 유사한 방법으로 수행할 수 있다.

[지리적 비교가능성 평가지표2]

| 지 표                | 유형   | 통계출처     | 주기 |
|--------------------|------|----------|----|
| C2. 유동성 대조통계의 비대칭성 | 주요지표 | Eurostat | 연간 |

[계산방법]

유동성 대조통계에 대한 비대칭성은 쌍을 이룬 두 국가의 국내외 유동성과 관련된 자료들 간의 차이로 정의할 수 있다. 특정한 영역의 대조통계로 고려된 국가에 대한 지표값은 국가간 국내외 유동성의 절대 차이의 합과 같다. 따라서 이 지표값은 항상 비음수(nonnegative)이다.

### 3. 시간에 대한 비교가능성

시간에 대한 불일치성(inconsistency)은 특정한 조사대상 기간에 수집된 자료가

특정한 시점에 수많은 특성에 기인하여 이후 시점의 자료와 완전히 비교 불가능할 때 발생한다. 이러한 경우 시계열이 절단되었다고 말한다. 두 조사 대상 기간의 측정의 개념과 방법간의 차이를 조사해야 한다. 동일한 조사의 서로 다른 경우 간의 비교가능성의 척도는 지리적 비교가능성에서 사용된 방법과 유사하다.

#### (1) 시간에 대한 비교가능성의 평가지표

##### [시간에 대한 비교가능성 평가지표1]

| 지 표               | 유형   | 통계출처                | 주기        |
|-------------------|------|---------------------|-----------|
| C3. 비교가능한 시계열의 길이 | 주요지표 | Eurostat,<br>국가통계기관 | 월간/분기별/연간 |

“비교가능한 시계열의 길이” 지표는 시계열상의 최근 절단 시점으로부터의 년도수로 정의할 수 있다. 만일 시계열상의 절단이 없다면 지표는 주요 통계에 대한 시계열의 시작시점으로부터의 년도수와 같게 된다.

다음의 지표는 “비교가능 시계열의 길이”에 대한 보조지표로서 공표된 모든 통계 또는 일련의 통계에 대해 계산이 가능하다.

##### [시간에 대한 비교가능성 평가지표2]

| 지 표              | 유형   | 통계출처     | 주기 |
|------------------|------|----------|----|
| C4. 비교가능한 시계열의 수 | 주요지표 | Eurostat | 연간 |

##### [참고]

시계열의 길이가 길수록 특정한 조사의 시간에 대한 비교가능성을 고려하기 좋다. 시점은 다음의 예와 같이 분류할 수 있다.

- 5년 이하
- 5년에서 10년 사이

- 10년 이상

#### 4. 영역 간 비교가능성

사용자들은 종종 분류에 따라 정의된 서로 다른 영역에 대해 통계를 비교한다. 예컨대 이러한 분류는 경제활동, 규모별 분류, 생산품, 유통방법, 성별 등과 관련되어 있다. 통계값의 추정에 사용된 개념상의 차이는 반드시 보고되어야 한다. 이는 주로 통계적 특성값, 조사대상 기간, 통계단위와 통계 측정의 정의와 관련되어 있다. 그러한 비교가능성의 평가차원은 다양하다. 예를 들어 소기업과 대기업 통계는 비교가능성 평가의 모든 개념과 표준 정의 목표를 포함하여 서로 다른 자료원에 기초한다. 영역 간 비교가능성의 측정은 지리적 비교가능성에서 이용하는 방법과 유사하다.

### 제 7절 일관성

#### 1. 일관성의 정의

통계의 **일관성(coherence)**은 다양한 이용에 대해 서로 다른 방법으로 신뢰성 있게 결합될 수 있는 것에 대한 타당성을 나타낸다. 그러나 일반적으로 일관성을 증명하는 것보다는 비일관성을 증명하는 것이 더 쉽다.

통계가 서로 다른 자료 원으로부터 나왔을 때, 특별히 상이한 속성이나 주기적인 조사로부터 나왔을 때 상이한 접근, 분류, 방법론적 표준에 근거하기 때문에 통계는 완전히 일관성이 없게 된다.

일관성과 비교가능성은 다른 측면에서 데이터셀에 나타난다. 두 개념간의 차이점은 만일 두 자료가 일관성을 가진다고 결정된 근거가 실제자료들 간에 불일치한 경우로서 비교가능성은 단지 메타자료에 근거하여 평가할 수 있다. 이러한 이유는 비교가능성은 연관되지 않은 모집단에 근거하여 작성된 통계간의 비교를 나타내고, 일관성은 동일하거나 상당히 유사한 모집단에 대해 작성된 통계간의 비교를 나타내기

때문이다. 물론 예외가 있을 수 있는데, 유량의 경우 두 기관이 두 국가에 대해 동일한 모집단에 대해 통계를 산출하게 되며, 이때 일관성과 지리적 비교가능성은 동일한 방식으로 평가하게 된다.

다양한 영역으로 일관성을 평가할 수 있다. 여기서는 자료의 1차적인 이용과 2차적인 이용간의 일관성을 고려한다. 실제로 통계기관에서 생산된 자료의 주요한 2차 이용은 통합적인 통계의 입력 자료로 이용된다.

## 2. 잠정 통계와 확정 통계간의 일관성

잠정 통계와 확정통계는 통상적으로 동일한 개념과 자료수집방법에 근거한다. 그러나 잠정치에 대해 이용가능한 정보가 그다지 많지 않으며 절차 또한 매우 신속하게 이루어진다. 잠정통계와 확정통계간의 일관성의 결여는 신뢰성 또는 상세함의 측면에서 구체화 할 수 있다. 이 성질은 정확성과 시의성간의 절충의 산물이기도 하다.

일관성 또는 비일관성 측도는 잠정치와 확정치가 상당히 큰 차이가 났을 때 정의를 선택해야 하며 기준점을 잘 설정해야 한다. 측도에는 (a) 절대오차 백분율 (absolute percentage error)로서  $|(잠정치-확정치)/확정치| \times 100(\%)$  와 (b) 비편향 절대오차 백분율(unbiased absolute percentage error)로서  $|(잠정치-확정치)/(잠정치+확정치)/2| \times 100(\%)$ 으로 표현된다.

<표 3-7-1>은 보험통계를 산출하여 1996년에 보험수가의 총 할증료에 대한 잠정치와 확정치간의 비교를 나타낸 것이다.

〈표 3-7-1〉 1996년에 보험수가의 총 할증료에 대한 잠정치와 확정치간의 비교

| 발표국   | 잠정치     |         | 확정치     |           | 차이 <sup>*주)</sup> |       |
|-------|---------|---------|---------|-----------|-------------------|-------|
|       | 생명보험    | 손해보험    | 생명보험    | 손해보험      | 생명보험              | 손해보험  |
|       | ECU Mio | ECU Mio | ECU Mio | ECU Mio   | %                 | %     |
| 벨기에   | 1,632   | 1,777   | 1,636   | 3,4590.28 | 0.28              | 48.63 |
| 덴마크   | 4,821   | 3,482   | 5,314   | 3,849     | 9.27              | 9.54  |
| 독일    | 48,804  | 74,053  | 49,353  | 73,980    | 1.11              | -0.10 |
| 엘리다   | 791     | 849     | 791     | 849       | 0.00              | 0.00  |
| 스페인   | 5,829   | 5,276   | 5,850   | 5,531     | 0.34              | 4.62  |
| 프랑스   | 51,429  | 42,888  | 61,896  | 42,759    | 17.20             | -0.30 |
| 아일랜드  | 2,479   | 2,009   | 2,478   | 2,033     | -0.05             | 1.18  |
| 이탈리아  | 7,599   | 9,044   | 7,599   | 9,123     | 0.00              | 0.96  |
| 룩셈부르크 | 179     | 365     | 2,382   | 635       | 92.47             | 42.60 |
| 네덜란드  | 15,617  | 13,255  | 15,785  | 13,430    | 1.07              | 1.30  |
| 오스트리아 | 953     | 1,073   | 924     | 1,085     | -3.11             | 1.15  |
| 포르투갈  | 1,400   | 1,178   | 1,400   | 1,178     | 0.00              | 0.00  |
| 핀란드   | 2,038   | 2,156   | 2,038   | 2,146     | -0.01             | -0.46 |
| 영국    | 81,711  | 48,587  | 81,711  | 48,587    | 0.00              | 0.00  |
| EUR15 | 231,503 | 211,419 | 246,217 | 208,653   | 5.98              | -1.33 |
| 아이슬란드 | 7       | 165     | 7       | 165       | 0.34              | 0.00  |
| 노르웨이  | 2,407   | 2,928   | 2,407   | 2,928     | 0.00              | 0.00  |
| 스위스   | 21,265  | 14,236  | 21,540  | 18,671    | 1.27              | 23.76 |

\*주) 여기서 차이는 ((잠정치-확정치)/확정치)×100(%) 로 계산한 값.

### 3. 연간 통계와 단기통계간의 일관성

다양한 특성값에 대해 통계는 연간 또는 주간, 월간, 분기 등의 주기를 갖는다. 이러한 서로 다른 주기를 갖는 통계들은 상이한 방법으로 산출된다. 단순하게 일관성을 측정할 수 있는 방법으로는 두 주기가 적어도 분기, 월간, 주간 중 하나의 통계가 지표로 산출될 때 연간 성장률을 비교할 수 있으며, 수준별 추정치를 제시하는 경우에는 연간 평균 수준이나 총계의 추정치를 비교하는 것으로 가능하다. 이들 두 통계간의 차이가 너무 커서 설명이 필요한 경우 6.1절에서 제시된 차이율 등과 같은 척도를 이용할 수 있다. 통계값이 동일할 필요는 없지만, 차이가 너무 많이 나는 경

우는 곤란하다.

**[사례 3-7-1] 1996년 영국의 연간 생산 및 건설조사 품질보고사례(ONS, 1996)**

생산지수(loP)가 연간 사업체조사(ABI)에 대한 평가기준은 아니다. 그러나 일부 비교작업을 개별 수준에서 실시하였다. ABI와 단기간 조사에 대한 자료수집이 동일한 처리단위에서 이루어지며 개별 단위에 대한 일치성을 검토한다. ABI와 분기별 자본 지출과 저축 조사간의 개별 수준에서 특별히 검토한다. 어떠한 공식적인 비교가 집단적 수준에서는 이루어지지 않는다. 그러나 1995년과 1996년에 초기작업이 loP에서 생산성 측도를 감소한 부가가치와 비교하기위해 산업 중분류 수준(2-digit level)에서 이루어졌다. 다음의 표는 1995년과 1996년 사이에 제조업의 중분류별로 loP성장률 대비 가중되지 않은 총생산(GO)과 가중된 총생산(WGO) 그리고 총 부가가치에 대해 가중되지 않은 값(GVA)과 가중된 값(WGVA)을 비교한 것이다.

<표 3-7-2>로부터 loP에서 성장은 총생산 보다는 총 부가가치의 성장에 더욱 가깝게 나타나고 있다. 그러나 loP자료는 후자에 더욱 근접할 것이라고 기대된다. 가중된 GO(WGO)와 loP 가치간의 차이점은 총합으로 보면 4%이상 나타나며 가중되지 않은 GO는 여전히 약 2% 수준차이로 근접해 있다.

제조업군의 중분류별로는 큰 차이를 나타내고 있다. 명확히 DF 와 DL 간에 큰 차이를 보이고 있다. 전자의 경우 loP의 실제 하락과 ABI의 큰 성장으로 기인한 것으로서 WGO(8.35%)와 loP(-10.86%) 간 차이가 19%이상 나고 있으며, 후자는 DF보다는 아니지만, 차이를 나타내고 있다. 이러한 차이에 대한 이유를 확인하기 위해 산업세 분류별(4-digit)로 검토가 필요하다.

〈표 3-7-2〉 고정된 가격에서 IoP 대비 가중 총생산과 가중되지 않은 총생산 그리고 가중총부가가치와 가중되지 않은 총부가가치의 1995년과 1996년간 성장률 비교

| SIC 92 분류 | IoP 가중치 | 성장률   |        |       |        | IoP    |
|-----------|---------|-------|--------|-------|--------|--------|
|           |         | GO    | GVA    | WGO   | WGVA   |        |
| DA        | 108     | -3.44 | -0.20  | -2.86 | 3.49   | 1.13   |
| DB        | 38      | 0.17  | 3.11   | -3.21 | 0.85   | -1.16  |
| DC        | 5       | 0.75  | 8.40   | 3.62  | 12.44  | -0.11  |
| DD        | 11      | -2.61 | -3.45  | -4.02 | -7.10  | -2.03  |
| DE        | 99      | 2.37  | -0.37  | 1.50  | 0.25   | -1.34  |
| DF        | 18      | 7.74  | -20.61 | 8.35  | -20.22 | -10.86 |
| DG        | 91      | 2.50  | -0.61  | 4.30  | 3.24   | 2.12   |
| DH        | 40      | 4.82  | 5.51   | 4.96  | 5.19   | -1.02  |
| DI        | 30      | -6.08 | -4.31  | -5.42 | -8.45  | -3.84  |
| DJ        | 93      | -0.51 | -7.26  | -0.86 | -10.48 | -0.03  |
| DK        | 72      | -0.24 | -0.68  | -1.39 | -0.90  | -1.76  |
| DL        | 101     | 8.08  | 1.19   | 9.27  | 1.31   | 2.64   |
| DM        | 77      | 6.88  | 5.94   | 8.18  | 6.93   | 3.86   |
| DN        | 29      | 0.41  | 6.14   | 1.95  | 9.56   | 1.69   |
| 제조업 총합    | 812     | 2.02  | -0.38  | 4.55  | -1.03  | 0.27   |

#### 4. 동일한 사회-경제적 영역 통계의 일관성

서로 다른 유형의 통계가 종종 다른 접근 방법으로 측정되지 않고 동일한 현상을 측정한다. 예를 들어 자본의 회전, 가치상승, 변동과 같은 기업에 대한 단기간 통계는 구조적 산업통계 결과로 연간비교가 가능하다. 과거 이용자들의 질의에 올바른 조치를 준비하기위해서 이러한 대표값 들이 너무 많은 변동이 없는지를 점검하는 것이 중요하다. 유량통계(무역, 운송, 국제수지, 관광)에 대해 대조통계는 상이함과 비대칭성을 통해 정확성의 개념을 제시하지만, 사용된 개념에 일부 다른 점이 있다.

#### [사례 3-7-2] 프랑스의 가계조사에 대한 ECHP의 일관성 사례

다음의 <표 3-7-3>은 1993년과 1994년에 ECHP의 1차 웨이브와 2차 웨이브와 가계조사(EBF94/95)로부터 추정된 구성별 프랑스의 총 소득의 분포를 나타낸 것이다.

〈표 3-7-3〉 1993년과 1994년에 ECHP의 1,2차 웨이브와 가계조사(EBF94/95)로부터 추정된 구성별 프랑스의 총 소득의 분포

|                 | ECHP-1993    | ECHP-1994                 | ECHP-1994         | EBF-1994/1995 |
|-----------------|--------------|---------------------------|-------------------|---------------|
|                 | 가구당 소득       | 가구당 소득<br>(1993년에 조사된 가구) | 가구당 소득<br>(모든 가구) | 가구당 소득        |
| 세전 총 소득         | 170,228      | 174,952                   | 171,147           | 168,590       |
| <b>%</b>        | <b>100%</b>  | <b>100%</b>               | <b>100%</b>       | <b>100%</b>   |
| <b>근로소득</b>     | <b>65.2%</b> | <b>64.4%</b>              | <b>64.6%</b>      | <b>63.7%</b>  |
| -월급과 임금         | 55.2%        | 55.2%                     | 55.5%             | 55.8%         |
| -자영소득           | 9.3%         | 8.4%                      | 8.3%              | 7.4%          |
| -부업소득           | 0.7%         | 0.8%                      | 0.8%              | 0.5%          |
| <b>연금</b>       | <b>21.1%</b> | <b>21.6%</b>              | <b>21.3%</b>      | <b>21.6%</b>  |
| <b>그 외 공적이전</b> | <b>7.8%</b>  | <b>8.1%</b>               | <b>8.1%</b>       | <b>8.5%</b>   |
| -실업급여           | 2.4%         | 2.5%                      | 2.5%              | 2.6%          |
| -가족수당           | 3.0%         | 2.8%                      | 2.8%              | 3.1%          |
| -상병급여           | 1.1%         | 1.4%                      | 1.4%              | 1.2%          |
| -공적부조           | 0.2%         | 0.3%                      | 0.3%              | 0.3%          |
| -주거수당           | 1.1%         | 1.1%                      | 1.1%              | 1.3%          |
| <b>자본소득</b>     | <b>4.7%</b>  | <b>4.1%</b>               | <b>4.3%</b>       | <b>5.2%</b>   |
| <b>사적이전</b>     | <b>1.0%</b>  | <b>1.4%</b>               | <b>1.5%</b>       | <b>0.9%</b>   |

<표 3-7-3>으로부터 분명하게 두 자료원은 비록 약간의 차이는 있지만, 매우 근접한 결과를 보여주고 있다. 예를 들어 자영자 소득과 부업소득은 EBF에서 낮게 나타난 반면 사적이전에서는 높게 나타나고 있다. 두 조사간의 소득분포는 일부 분포의 극단값을 제외하고는 거의 유사하게 나타났다. EBF의 경우 10분위 배율이 높게 나타나고 90분위 배율이 약간 낮아 EBF가 ECHP 보다 다소 불균등한 것으로 나타났다. ECHP의 2차 웨이브 원자료와 1차 웨이브, EBF94/95를 비교하면 추정된 세전 총가구소득 간에 유의한 차이는 없다. ECHP에 대해 1차 웨이브에 존재하는 가구에 관련된 2차 웨이브의 열을 직접적으로 비교가능하다.

##### 5. 국민계정 통계의 비교

이용자 요구에 가장 적합한 정보근원으로 이용자 의견에 대해 조사통계와 국민계정을 비교하는 것이 유용하다. 국민계정통계에 사용한 방법은 1차적인 자료원과 조정을 고려한 통계에 대한 설명이 필요하다. 개념상의 변동을 반드시 고려해야 한다.

[사례 3-7-3] 스웨덴의 단기생산량 지수(STPVI)에 대한 일관성 평가 사례  
(ONS,1999)

국민계정은 자본변화에 대해 조정된 STPVI의 특별판 뿐만 아니라 통상적인 비조정된 지수를 수용한다. 출하에 근거한 산업에 대해 특별판에 대한 공식은 다음과 같다.

$$\frac{\sum_{m=1}^3 (L_{m,t}) + S_{q,t}}{\sum_{m=1}^3 (L_{m,t-1}) + S_{q,t-1}}$$

여기서  $L_m$ 은 1990년에  $m$ 월의 출하가격,  $S_q$ 는 1990년의  $q$ 분기동안 자본의 변동 가격이고,  $t$ 는 년도를 나타낸다.

생산 또는 근무시간 수에 근거한 산업에 대해 분기별 변동을 다음과 같이 계산된다. 각 분기마다 국민계정은 현재분기에 대한 지수와 이전년도의 해당 분기에 대한 지수를 수용한다. 이 지수는 기간동안의 변동을 산출하기 위해 사용되며 이 값을 특정 산업의 부가가치에 대한 국민계정 수준에 적용한다. 국민계정은 보통 수용한 자본조정지수를 변화하지 않는다. 생산변동을 사용자 측면에서 변화에 대해 점검한다. 만일 차이가 매우 크면 생산측면을 조정한다. 국민계정은 비록 드물긴 하지만, 자본상의 변동에 대해 조정되지 않은 지수를 사용한다.

국민계정이 수용한 분기별 자료는 공식적으로 공표되지 않지만, 분기별 자료는 STPVI의 월별 지수로부터 계산할 수 있다. 이 값들은 자본의 변화에 대해 조정되지 않은 값이며, 따라서 공식적으로 공표된 STPVI와 분기별 국민계정은 엄격한 의미에서 일관성이 없다. 이러한 일관성의 결여를 설명하기 위해 STPVI와 국민계정에 대해 1994년과 1995년의 분기별 변화를 표로 작성하였다. STPVI의 분기별 변동은 다음 식으로부터 구한다.

$$\frac{\sum_{m=1}^3 I_{m,t}}{\sum_{m=1}^3 I_{m,t-1}}$$

여기서  $I_m$ 은  $m$ 월에 대한 STPVI지수를 나타낸다(1990=100). 또한  $t$ 는 년도를 나타낸다.

실제로 위 식은 출고를 기준으로 할 때에는

$$\frac{\sum_{m=1}^3 L_{m,t}}{\sum_{m=1}^3 L_{m,t-1}}$$

와 같다.

〈표 3-7-4〉 고정된 가격에서 STPVI와 국민계정(NA)에 따른 1994년부터 1995년  
까지 분기별 변동 간 비교

(단위 :%)

| NACE  | NA<br>Q=1 | STPVI<br>Q=1 | NA<br>Q=2 | STPVI<br>Q=2 | NA<br>Q=3 | STPVI<br>Q=3 | NA<br>Q=4 | STPVI<br>Q=4 |
|-------|-----------|--------------|-----------|--------------|-----------|--------------|-----------|--------------|
| C     | 9.9       | 9.2          | 5.4       | 7.7          | 14.2      | 10.5         | 15.5      | 6.9          |
| D     | 13.6      | 12.6         | 12.2      | 12.3         | 11.5      | 11.4         | 5.8       | 4.6          |
| 15+16 | 1.7       | 0.3          | 3.2       | 4.2          | 2.0       | 2.6          | 3.8       | 4.6          |
| 17-19 | 18.0      | 15.3         | 13.1      | 13.1         | 16.6      | 18.4         | 4.4       | 4.7          |
| 20    | 23.6      | 20.5         | 11.0      | 9.7          | 11.3      | 11.1         | 2.7       | 2.0          |
| 21    | 6.9       | 6.1          | 2.0       | 2.1          | -1.0      | -1.2         | -12.8     | -13.4        |
| 21.11 | 12.9      | 9.3          | 7.8       | 4.3          | 4.8       | 1.6          | -11.9     | -14.3        |
| 22    | -3.3      | -7.5         | 2.5       | -2.2         | 0.4       | -5.0         | 3.4       | -2.6         |
| 23    | -6.6      | -6.3         | -8.0      | -7.4         | 17.5      | 17.7         | 2.2       | 2.4          |
| 25    | 4.3       | 5.7          | -1.8      | 0.0          | -6.3      | -4.5         | -9.7      | -9.1         |
| 26    | 19.8      | 24.8         | 5.4       | 6.5          | 5.2       | 5.1          | -3.3      | -4.3         |
| 27    | 0.4       | 0.1          | -0.1      | -0.3         | 0.3       | 0.0          | 3.4       | 3.2          |
| 28    | 28.5      | 26.3         | 16.2      | 17.0         | 16.6      | 15.3         | 8.1       | 7.1          |
| 29    | 20.4      | 20.0         | 23.9      | 25.5         | 18.8      | 16.7         | 7.4       | 11.4         |
| 30    | 17.1      | 8.3          | 7.6       | 13.3         | 10.8      | 12.4         | 10.0      | 10.8         |
| 33    | 4.4       | 6.0          | 11.5      | 8.4          | -7.4      | -13.8        | 3.3       | -2.9         |
| 36+37 | 10.8      | 8.4          | -1.8      | -1.0         | 1.3       | 2.9          | -5.4      | -5.3         |

(1) 일관성의 평가지표

[일관성 대한 평가지표]

| 지 표                                  | 유형   | 통계출처                | 주기 |
|--------------------------------------|------|---------------------|----|
| CH1. 주요한 2차비용에 대한<br>수요를 만족하는 통계의 비율 | 경험지표 | Eurostat,<br>국가통계기관 | 연간 |

특정한 통계에 대한 가장 중요한 2차 이용을 정의한 후 2차이용에 대한 이상적인 성질들을 모집단/통계단위, 변수의 정의, 조사시점, 주기, 분류의 표준화 등과 같은 인자별로 정의할 수 있다. 다음으로 만일 통계가 모든 관련된 인자에 대해 요구를 충족시키는지 또는 그렇지 못한지를 결정한다.

[계산방법]

통계의 주요한 2차이용  $u$   $u=1, \dots, S$ 에 따라 다음의 표와 같이 분류한다.  $k$ 개의 인자 열에서 1은 통계가 주요한 2차이용에 대한 요구를 충족한 경우를 나타내고, 그렇지 않은 경우 0으로 나타낸다. 그러면 인자  $k_{ul}$ 과 관련하여 동일한 2차이용 요구를 만족하는 통계의 비율은 다음과 같이 계산된다.

$$P_{u, k_{ul}} = \frac{\sum_{j=1}^{N_u} I_{k_{ul}, j}}{N_u}$$

여기서  $N_u$ 는 동일한 주요 2차이용의 요구를 만족하는 통계의 수,  $k_{ul}$ 은 2차이용에 대한 1순위 인자.

$$I_{k_{ul}, j} = \begin{cases} 1, & j\text{번째 통계가 } k_{ul}\text{의 요구를 수용한 경우} \\ \text{그외} & \end{cases}$$

따라서 모든  $n_u$ 개의 관련된 인자들과 관련하여 요구를 충족한 동일한 2차이용 통계의 비율은 다음과 같다.

$$P_{u, \cdot} = \frac{\sum_{j=1}^{N_u} I_{\cdot, j}}{N_u}$$

여기서  $N_u$ 는 동일한 주요 2차 이용 통계의 수이다.

$$I_{\cdot, j} = \begin{cases} 1, & j\text{번째 통계가 모든 } n_u \text{ 개의 인자와 관련한 요구를 수용한 경우} \\ \text{그외} & \end{cases}$$

전체 지표는 다음과 같이 계산된다.

$$P_{\cdot, \cdot} = \frac{\sum_{u=1}^S \sum_{j=1}^{N_u} I_{\cdot, j}}{\sum_{u=1}^S N_u}$$

[지표의 해석]

주요 2차이용에 대한 요구를 만족하는 통계의 수와 비율이며, 불충분한 점이 무엇인지에 관한 강력한 지표이다. 만일 지표가 증가하면 해당 통계시스템상의 일관성은 증가하게 된다.

[필요한 메타자료]

각 통계에 대한 문서와 그들의 주요 2차이용에 대한 요구의 정의

## 제 8절 비용과 응답자 부담

비용과 응답자 부담은 통계의 품질을 이들과 분리해서 고려할 수 없다는 측면에서 품질평가 작업의 일부이다. 통계산출물과 연관된 비용의 평가는 다소 복잡한 업무 중의 하나인데, 이는 미리 배분된 비용(예를 들어 사업등록, 공유하고 있는 정보통신기술의 근원 그리고 공표경로)의 부분과 간접비(사무실 공간, 잡비 등)에 대한 체계가 있어야 하며, 자세하고 분명하게 이러한 부분이 나타남으로서 서로 다른 구조를 가진 기관들 간의 국제비교가 가능해지기 때문이다.

우선, 국가통계기관에 대한 직접비용에 대한 조사에는 한계점을 제시한다. 다음 과정으로 이용자에 의해 추가되는 비용을 고려한다. 시간적으로 이 비용들은 현재 통계작성의 대부분에 대해 무시할 수 있을 정도이다.

응답자 부담과 관련하여 재정적 측면에서 현실화 할 수 없는 부분이 있지만, 설문지를 완성하는데 걸리는 시간, 조사원에 응답하는 데 걸리는 시간 등을 고려할 수 있다. 하지만, 영국의 ONS는 응답자 부담을 재정적 측면에서 기업체의 응답부담을 측정할 수 있는 방법을 개발하였다.

<Eurostat에서 제안하고 있는 비용 산출 방법>

① 참여한 직원의 수(참조1)

총 인원수

전문가 또는 관리자 수

② 통계담당부서 비용(참조2)

직원인건비

자료수집비용(인쇄비, 우편비)

무응답처리비용(우편, 전화, 면접)

자료관리비용(오류검토, 에디팅)과

결과처리비용(보간, 표작성, 편집)

기타비용

총 비용 =

| 부서 |
|----|
|    |
|    |

단위: 천원

분배비율(%)

|   |  |   |
|---|--|---|
|   |  |   |
| + |  | % |
| + |  | % |
| + |  | % |
| + |  | % |
|   |  |   |

[참고사항]

만일 SBS(Structural Business Survey) 통계작성 절차에 참여한 직원이 다른 업무에 참여하지 않았다면, 이 문서를 쉽게 작성할 수 있다. 만일 그러한 경우가 아니면 SBS 배분을 추정할 필요가 있기 때문에 복잡해진다.

만일 이용할 정보가 없다면, 참조사항에 언급한 바와 같이 생략, 차이점, 대체적 정의 등에 관한 정보를 참조사항으로 제시하라.

[참조1]

포함된 직원 수는 국가통계기관, 장관, 은행, 지역본부 또는 기관들을 구별해야 한다. 이상적으로 인원수는 직접적으로 고용된 직원으로 한정하며, 일용직 보다는 정규직으로 한정하는 게 바람직하다.

[참조2]

이상적으로 담당직원이 아닌 인건비는 분리하여 제시한다. 만일 불가능하면 총비용을 제시하라. 총 비용은 외부로부터의 구매 또는 조직 내부에서 제공한 것인지에 대한 완전 경상비용, 다시 말해서 직원인건비, 계산, 물질, 도구, 서비스이용 비용 등이 포함되어야 한다. 직원 인건비는 이 업무에 적용한 고용주의 사회보험비용을 포

함해야 한다. 자본지출(투자)을 총비용에 포함해야 한다. 예산은 선택된 영역에서 현재 업무를 모두 포함해야 하지만 EU의 요구사항을 초과하지 않도록 한다.

□ 응답 부담 측정을 위한 ONS 방법

사업체 통계의 비용을 측정하기 위해 Eurostat은 영국의 ONS 방법을 따르도록 제안하였다. 이에 대한 주요한 이유는 다음과 같다.

첫째, 이 방법의 간편성으로 모든 회원국이 쉽게 적용가능하다는 점이다.

둘째, 회원국간의 비교가 가능하며 특별히 상이한 조사체계 간 비교가 가능하다는 점이다.

셋째, 이부분에 대해 이용 가능한 사례가 단지 이 방법뿐이기 때문이다.

이 방법의 기본적인 원리는 표본조사에 근거하여 사업상 재정비용을 측정하는 것으로 구성된다.

R = 응답자수,

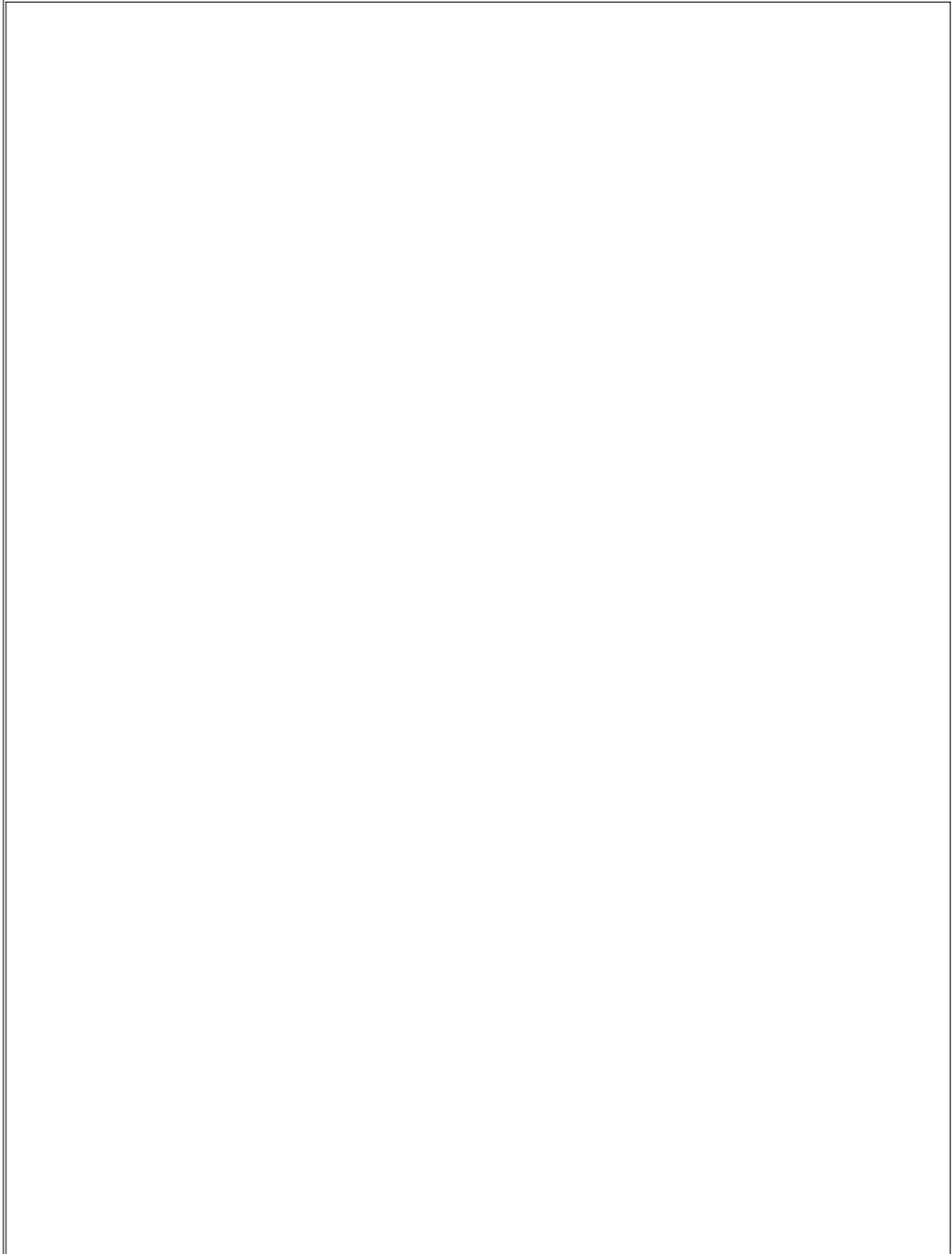
T= 정보를 제공하는 시간, 즉 사전에 완성된 정보를 취합하는데 걸리는 시간, 조사에 참여하는 시간, 조사를 위해 재 접촉에 소요되는 시간을 포함한다.

C= 응답자의 시간에 대한 시간 단위 비용.

ONS는 무응답자에 대해서도 동일한 항목으로 추정해야 한다고 언급하고 있다. 그러나 간단히 하기 위해 비용을 0으로 하거나 응답자와 같은 수준으로 사정할 수 있다. 전체 사업비용은 다음과 같이 계산된다.

$$\text{비용} = R \times T \times C$$

더욱이 Eurostat은 단지 조사통계에 관해서만 비용 산출 모형을 제시하고 있으며, 보고(행정)통계에는 별도로 고려하도록 하고 있다.



## 제4장 자체통계 품질진단을 위한 가이드라인

자체품질진단은 통계를 생산하는 기관, 넓은 의미로서 통계생산자가 소관통계의 품질을 자체적으로 평가하고, 그에 대한 개선점을 수립하여 향후 통계품질을 향상시키고자 유럽연합(EU)에서 도입하여 회원국들의 통계를 자체적으로 평가하는 시스템으로 DESAP을 개발하여 운용하고 있다. DESAP의 품질 차원은 크게 7장으로 구성되며, 각 장별로 하위 항목들로 구성되어 있다. DESAP의 개발목적은 조사관리자들이 자신이 작성하고 있는 통계의 품질을 진단하고, 품질개선 수단을 강구하는데 도움을 주기위한 것으로서, 유럽의 통계시스템(ESS)의 품질기준을 충족하고 있으며, 주제 분야와 특정한 조사방법과 무관하게 원시자료로부터 수집된 개별 통계에 적용시킬 수 있다는 점에서 일반적인 평가 도구이다.

본 장에서는 통계생산자가 소관통계에 대해 자체진단을 수행하기 위한 기본적인 가이드라인을 제시하여, 가이드라인에 부합하지 않는 사항들에 대해서는 개선점을 도출하고 향후 생산되는 다양한 분야의 통계 품질향상에 도움을 주고자 한다.

본 장에서 제시하는 가이드라인은 DESAP의 품질진단차원을 기준으로 각 장별로 주요평가사항을 제시하고 있으며, DESAP의 개별 평가 문항에 제시된 내용을 중심으로 내용을 전개하였다. 따라서 외부진단 절차와는 일부 상이한 측면이 있으며, 가이드라인에 제시된 내용 중에 전문적인 지식을 요구하는 사항에 대해서는 부록에서 다루도록 하였다.

### 제1절 조사실시의 결정

#### 1. 이용자 요구분석

이용자 요구분석은 통계 품질평가의 차원 중 관련성 및 시의성 등과 연관된 지표로 해석될 수 있다. 따라서 이용자 요구분석을 위해서는 통계이용자에게 해당 통계가 유용한지, 이용자가 누구인지, 얼마나 많은 이용자들이 있는지, 해당통계가 얼마나 중요한지 등에 관한 정보가 필요하다. 이를 위해 먼저 소관통계에 대한 이용자 그룹을 분류할 필요가 있다. 이용자 그룹의 분류 기준은 반드시 일정한 기준에 의거하기 보다는 소관통계 이용자들의 중요성에 따라 분류하는 것이 바람직하다. 이러한 측면에서 생산통계의 법적근거 또한 이용자유구분석 중이 하나로 고려할 수 있다.

이용자들에 대한 정보를 수집하고, 해당 통계가 이용자 요구에 적절하게 부합되는지를 분석해야 한다. 만일 이용자 요구분석에서 이용자의 요구에 미충족된 사항이 있으면, 그 원인을 분석하여 향후 추가적으로 이용자들의 요구를 만족시킬 수 있도록 한다.

한편 소관 통계에 대한 이용자 자문그룹을 운영하는 방안도 고려 될 수 있다. 통계 작성자가 통계이용자의 관점을 통계생산과정에 투입함으로써 현재 및 향후 통계이용자의 잠재적 수요를 적절히 포괄할 수 있기 때문이다.

#### [가이드라인]

##### □ 통계생산자는 소관통계 생산의 법적 근거를 명확히 제시하도록 한다.

통계생산자는 이용자 요구분석을 통해 생산되는 통계의 법적 근거를 명시함으로써 향후 생산통계의 합법성을 밝히도록 한다. 통계생산의 법률적 주체가 상위 부서인지 또는 현 부서인지에 대한 자료생산의 책임소재를 명확히 하고, 만일 주기성을 갖는 통계인 경우 그에 대한 법률적 근거를 제시하여 유사통계에 대한 법률적 분쟁 소지를 차단하도록 한다.

##### □ 이용자 요구분석을 위해 필요한 가능한 한 많은 이용자들에 대한 정보를 체계적으로 수집, 관리해야 한다.

이용자 데이터베이스나 또는 파일로 작성된 이용자 리스트를 반드시 관리하도록

한다. 이와 더불어 주요 이용자들에 대한 그룹별로 분류된 DB 목록이 비치되도록 한다. 이때 가능하면 사용자그룹을 기관별, 연구자별로 작성하여 순위를 부여하도록 한다.

**□ 이용자 요구에 대한 가능한 풍부한 정보가 포함되도록 한다.**

이용자들의 통계에 대한 기대치, 향후 이용의 잠재성, 사용자그룹에 대한 정보 제공의 주요전략 등을 상세히 기술함으로써 이용자들의 다양한 내용이 기술되도록 한다.

**□ 추가적인 이용자 그룹에 대한 정보를 얻기 위해 이용자 만족도 조사나 현재 발간된 통계 문서의 독자층에 대한 분석, 관련 통계의 웹 페이지 접속건수, 다운로드 수 등을 분석하여 지표로 활용할 수 있다.**

참고로 이용자 만족도 조사를 위해 기본적으로 포함되어야할 조사 내용을 살펴보면 다음과 같다.

- 응답자의 인구학적 특성
- 통계이용실태
- 통계 및 이용자 서비스에 대한 만족도
- 통계 품질 차원별 만족도
- 대상 통계관련 주요 이슈
- 통계 문제점이나 관련 제안

**□ 통계 이용자 자문그룹을 운영하여 소관통계의 향후 잠재적 이용에 대비하도록 한다.**

통계이용자의 관점에서 보다 포괄적으로 통계수요에 대비하도록 이용자 자문그룹을 운영하는 것이 바람직하다. 자문그룹의 운영과정에서 통계작성절차의 다양한 의견을 청취함으로써 향후 이용자 요구에 적절히 대응할 수 있기 때문이다. 이용자

자문그룹의 범위 또한 가능하면 일반이용자부터 주요 이용자까지를 포괄하도록 한다.

**□ 이용자 자문그룹으로부터 소관통계에 대한 미충족 수요를 파악하고, 그에 대한 적절한 대비계획을 수립한다.**

이용자 그룹간의 토의를 통해 미충족된 사항에 대한 원인과 향후 개선책을 마련하도록 한다. 예컨대 조사항목에 포함되어야 할 일부 항목이 있다면, 향후 조사에서는 지적된 사항이 포함되도록 계획을 수립할 수 있으며, 통계결과표상에 심층적인 분석을 요구할 경우에는 결과분석 항목에 해당사항을 추가하는 등의 계획을 수립하도록 한다.

## 2. 조사목적

자료 수집을 위한 조사수행에 대한 목적을 명문화 하는 것이 무엇보다도 중요하다. 이는 무엇을 위해 조사를 수행하는지에 대한 방향성을 제시할 뿐만 아니라, 누구를 위해 조사를 수행하는가에 대한 이용자 그룹의 정의와도 밀접한 관계가 있다. 즉, 조사목적에 대한 구체적인 기술이 결과적으로 조사로부터 수집된 자료 이용의 대상까지도 포함하기 때문에 반드시 명시할 필요가 있다.

### [가이드라인]

**□ 조사목적은 통계이용자 요구에 맞추어 주어진 비용 하에서 가장 효율적으로 조사가 수행될 수 있도록 명문화 한다.**

통계 이용자의 요구에 100% 만족되도록 조사 목적을 설정하는 것은 주어진 예산과 조사현실을 감안할 때 거의 불가능하다. 따라서 되도록이면 이용자의 요구를 충족시키면서 현실성 있는 조사 목적을 수립하도록 하며, 반드시 이러한 사항들을 명문화 하여 추후 통계작성 환경의 변화에 적절히 대응할 수 있도록 한다.

□ **이용자 자문그룹의 조언을 통해 충분한 검토 후 조사목적으로 결정함으로써 이용자 요구를 충족시킬 수 있도록 한다.**

조사목적을 통계작성기관의 의견만으로 결정할 경우 이용자 요구에 적절히 대응하지 못하는 결과를 초래할 수 있기 때문에 가능하다면 조사목적을 설정할 때 이용자 그룹의 조언을 청취하여 보다 포괄적인 조사목적이 수립되도록 한다.

□ **한번 결정된 조사 목적은 향후 변동되는 통계작성 환경에 따라 주기적으로 검토하여 이용자 요구에 부합되도록 한다. 이를 위해 통계전문가의 조언이 필요한 경우 외부 전문가의 의견을 수렴하는 것도 한 가지 방법이 될 수 있다.**

통계 작성 환경은 시간의 흐름에 따라 반드시 변화한다. 즉, 예산의 변동, 조사 응답자들의 비협조성, 조사원들의 조사경험으로 발생하는 편향, 이용자 요구의 증대 등으로 초기에 수립된 조사목적이 변화하는 통계환경에 부합되지 않을 경우 이러한 내용을 적절히 반영하여 조사목적을 수정하는 것이 바람직하다. 만일 통계작성 기관 내부에 이와 관련된 전문적인 그룹이 없을 경우 외부 전문가의 조언을 통해 적절히 수정하는 방법도 고려할 수 있다.

□ **만일 소관 통계에 대한 품질목표가 있을 경우 이를 조사목적의 항목에 포함시키도록 한다.**

예를 들어 최종 응답률, 표본오차, 추출틀의 포괄성, 시의성 등이 해당 통계 품질의 목표치가 될 수 있을 것이다. 이러한 사항을 조사목적에 포함시킴으로서 이용자에게 참고가 되도록 해야 한다.

## 제2절 조사 설계

조사 설계는 자료수집과 자료처리의 모든 과정에 대해 품질을 보증하기 위한 원칙을 제시하는 단계이다. 이 과정은 특히 조사통계 부문과 보고통계부문의 품질을

평가하는 기본요소가 포함되어 있기 때문에 조사수립단계에서 반드시 고려되어야 할 사항이다. 특히 조사의 개념, 표본설계, 조사표개발 및 실험(테스트)은 조사통계에서 반드시 고려되어야 하며, 이러한 사항들은 품질평가 차원에서 비교가능성, 정확성, 일관성, 시의성 등을 보장하는 개념들이다.

### 1. 조사개념

조사의 개념, 정의, 분류기준 등은 통계의 비교가능성, 일관성, 시의성 등을 결정하는 중요한 항목이다. 조사 초기에 조사의 계획수립단계에서 명확히 기술해야 한다.

개념(concept)은 조사를 계획하기 위해 다수의 속성들을 미리 정의하는 것을 의미한다. 그러한 속성들은 조사대상 모집단, 추출틀, 특성, 모집단의 분류 그룹들이 포함된다. 즉, 이러한 항목들을 불명확하게 기술할 경우 조사의 목적을 달성하지 못할 뿐만 아니라, 개념의 혼선으로 인해 결과적으로 의미 없는 자료가 생산되는 결과를 초래하게 된다.

따라서 조사개념의 불명확성으로 유사한 조사임에도 불구하고 상이성 때문에 서로 비교가 불가능하게 되는 측면이 있게 된다. 이러한 문제점을 해결할 수 있는 방안으로는 조사의 개념, 정의, 분류체계를 수립할 때, 현재 계획하고 있는 조사와 유사한 조사가 있는지를 먼저 검토하는 것이 바람직하다. 만일 기존의 조사에서 이러한 사항이 만족스럽지 않다면, 이를 개선한 보다 구체적이고, 명확한 개념을 도입해야 하며, 통계분류단계에서 오류가 발생했거나, 중복 또는 누락이 발생했다면, 그 원인과 개선방향을 제시하고, 가능한 한 표준분류체계를 사용하도록 한다.

#### [가이드라인]

- 현재 조사와 유사한 개념을 적용한 기존의 조사가 있는지를 검토하여, 표준적인 개념을 적용한 사례가 있을 경우 이를 준용하도록 한다.

현재 계획하고 있는 조사와 유사한 기존의 조사를 검토함으로써, 동일한 모집단을 사용한 유사한 조사임에도 비교가능성이 결여된 통계가 생산되는 것을 방지할 수 있으며, 따라서 생산된 통계의 신뢰성을 확보할 수 있다. 가능하다면 비교가능성을 위해 표준화된 통계분류체계를 사용하는 것이 바람직하다. 이를 통해 출처가 다른 조사로부터 수집된 자료를 비교하거나, 필요에 따라서는 연계(link)가 가능하기 때문이다.

□ 만일 유사한 타 통계의 분류체계와 비교하여 일치하지 않는 부분이 있을 경우, 이를 명시함으로써 이용자들의 혼선을 방지하도록 한다.

□ 통계이용의 확대를 위해 가능하다면, 세 분류 까지 분석이 가능하도록 분류체계를 세분화하는 것이 바람직하다.

이를 위해 표준 분류체계를 적용하도록 한다. 국제비교를 위해 이용 가능한 표준 분류체계로서는 UN, 국제노동기구(ILO), Eurostat, OECD 등의 분류체계가 있으며, 국내적인 분류체계로서는 통계청, 산업자원부, 노동부의 분류체계를 들 수 있다.

□ 만일 통계 이용목적에 따라 세 분류를 상위분류체계로 통합할 경우에는 표준적인 통합체계를 사용하도록 하고, 이러한 사실을 이용자가 명확히 이해할 수 있도록 세부내용을 기술하도록 한다.

제조업의 세분류 또는 세세분류로 분류한 자료를 중분류나 대분류로 통합하고자 할 경우에는 표준산업분류 부여 체계에 따라 적용하고, 이러한 사실을 기술함으로써 이용자가 해당 산업에 대한 분류체계를 명확히 이해할 수 있게 된다.

□ 만일 기존에 적용한 표준분류체계가 없을 경우에는, 가능한 관련 통계에서 적용하고 있는 분류체계를 따르도록 한다.

#### 1) 모집단의 정의

조사대상 전체집합을 모집단이라 한다. 이는 표본추출의 대상이 되며, 추정의 대상이 되기도 한다. 따라서 통계의 정확성, 비교가능성 등을 평가하기 위해서는 모집단을 정확하게 정의할 필요가 있다. 모집단은 목표모집단과 조사모집단 2가지로 정의할 수 있다. 첫째, 목표모집단(target population)은 조사목적에 의해 개념상으로 정의되는 모집단이다. 둘째, 조사모집단(survey population)은 표본추출의 대상이 되는 모집단으로서 추출틀 작성의 대상이 되는 모집단이다. 모집단을 정의할 때 반드시 고려되어야 할 사항으로는 모집단에 포함되는 조사 단위들의 특성과 유형을 명확하게 규정해야 하며, 모집단의 지리적, 공간적, 시간적 범위를 명확하게 정의해야 한다.

#### [가이드라인]

□ **조사의 개념으로 모집단에 대한 조작적 정의가 필요하며, 이를 위해 먼저 목표모집단과 조사모집단을 정확하게 정의해야 한다.**

통상적으로 조사대상 단위전체를 모집단으로 정의하지만, 조사통계의 경우 개념상의 모집단과 실제 조사대상이 되는 모집단은 서로 다른 경우가 있다. 따라서 목표모집단과 조사모집단에 대한 정확한 정의가 필요하다.

□ **목표모집단과 조사모집단이 서로 다른 경우 그 차이가 어느 정도인지를 검토하고, 이에 대한 정보를 제공해야 한다. 조사대상 시점, 조사대상, 조사지역 등 두 모집단에 대한 조작적 정의의 차이 등을 기술함으로써 이용자에 대해 시점에 따른 모집단의 포괄성의 이해정도를 용이하게 할 수 있다.**

두 모집단에 대해 어떠한 측면에서 차이가 나는지, 시간, 공간적으로 얼마나 차이가 나는지 등을 기술함으로써 통계이용자에 대한 통계오용을 방지할 수 있다.

#### 2) 추출틀의 정의

추출틀(sampling frame, 또는 frame)이란 모집단 단위들의 목록이다. 표본조사의 경우 표본을 추출할 대상이 되는 추출틀을 정의해야 한다. 즉, 조사모집단에 속하는 단위들에 대한 리스트를 작성하여, 이를 바탕으로 표본추출작업이 수행되게 된다. 예를 들어 인구주택 총 조사의 경우 전국 단위의 조사구 리스트가 바로 추출틀이 되며, 조사구내의 가구목록이 부차 추출틀의 역할을 한다. 표본설계의 정확성을 담보하기 위해서는 효과적인 추출틀을 마련하는 것이 중요하다. 추출틀은 조사대상에 따라 지역추출틀, 사업체 추출틀 등으로 구분이 가능하다. 표본조사에서 추출틀은 조사특성에 부합되고, 모집단에 포함된 조사 단위들의 중복이나 누락을 최소화 할 수 있는 추출틀을 작성해야 한다. 특히 추출틀은 표본설계, 자료수집, 추정, 품질평가 등의 단계에서 다양하게 이용되기 때문에 추출틀의 정의가 매우 중요하다. 추출틀의 작성단계에서 발생할 수 있는 오류는 크게 두 가지로 고려될 수 있다. 하나는 추출틀의 노후화로 인해 모집단 변동을 적절히 반영하지 못하는 경우로서 과대포함과 과소포함 문제이다. 다른 하나는 추출틀의 주기적인 평가의 미흡으로 발생하는 조사단위의 중복문제이다. 이를 보다 구체적으로 설명하면 다음과 같다.

- 과소포함(under-coverage) : 목표모집단을 조사모집단으로부터 접근할 수 없는 경우로서 전화조사에서 전화가 없는 개인은 전화번호부에 기입되지 않음으로 조사가 불가능한 경우에 발생된다.
- 과대포함(over-coverage) : 목표모집단에 속하지 않은 단위가 조사모집단에 속해 있는 경우로서 사망한 개인이 여전히 전화번호부에 기재되어 있는 경우로 조사 단위 목록에는 존재하지만 실제 모집단에는 존재하지 않는 경우에 발생된다.
- 중복 기재(duplicated list) : 목표모집단의 단위가 한번이상 조사모집단에 존재하는 경우로서 한사람이 여러 대의 전화번호를 등록하여 실제로는 한 사람인데 여러 사람으로 나타나는 경우로서 주기적으로 추출틀을 갱신하지 않음으로서 발생된다.

[가이드라인]

□ **기존의 조사와 비교 가능한 유사한 추출틀을 고려하고, 동일 모집단에 대한 조사의 경우 공통의 추출틀을 사용함으로써 비교가능성과 일관성을 유지하도록 한다.**

계속조사나 유사한 조사의 경우 동일 모집단을 대상으로 조사하는 경우에는 가능하면 공통의 추출틀을 사용하여 조사의 일관성을 유지하고, 기존의 조사자료를 비교하거나 참조함으로써 현 조사의 품질을 평가할 수 있다.

□ **조사목적에 적합하고, 이용 가능한 다양한 추출틀을 확보하여 비교 검토한 후 조사목적 등에 가장 적합한 추출틀을 사용한다.**

조사목적에 적합한 이용 가능한 추출틀은 다양하게 존재할 수 있다. 이러한 여러 추출틀 중에서 경제성, 추출틀의 품질, 이용가능성 등을 종합적으로 판단하여 현재 조사목적에 가장 적합한 추출틀을 사용하도록 한다. 전국 규모 가구조사의 경우 인구주택총조사구 명부가 가장 적절하게 사용가능한 추출틀이 될 수 있다.

□ **사용하고자 하는 추출틀의 품질을 주기적으로 평가하여 과대표함, 과소포함, 중복기재 등을 검토하여 표본추출시 오차를 최소화 하도록 한다.**

가구조사의 경우 표본 조사지역의 재개발 등으로 인해 목표모집단에는 존재하지만 조사모집단에는 존재하지 않는 가구가 발생할 수 있다. 이러한 문제를 검토하여 가능한 한 오차를 줄일 수 있도록 추출틀을 검토하는 것이 바람직하다.

□ **가능한 한 조사단위와 관련된 보조정보가 정확한 추출틀을 확보하도록 한다.**

추출틀에 조사단위와 연관된 보조정보가 많으면 표본설계의 효율을 높일 수 있으며, 추정과정이나 무응답에 대한 조치를 취할 때 해당 조사단위에 대한 보조정보를 활용할 수 있기 때문에 매우 효과적이다. 이러한 측면에서 모집단 자료중 90% 자료 보다는 10% 표본자료가 표본설계 상에서 더 선호된다.

□ **조사주기와 추출틀의 최근 갱신 시기간의 시차를 측정하여 이를 보고서에 반영**

**하도록 한다.**

조사주기 또는 조사대상 기간과 추출틀 간의 시차는 어느 정도 발생하기 마련이다. 그러나 추출틀의 포괄성 유지를 위해 조사대상 기간에 비추어 최근에 갱신된 사항이 있다면, 이를 바탕으로 적절한 품질평가 요소를 적용할 수 있기 때문이다.

**□ 추출틀이 목표모집단과의 점근성을 향상시키기 위한 가능한 방법을 강구하도록 한다.**

이용하고자 하는 추출틀이 목표모집단과 차이가 있을 경우 어떠한 작업을 통해 점근성을 확보할 수 있는가를 검토하도록 한다. 이는 추출틀의 포괄성과 밀접하게 관련되어 있기 때문이다. 만일 조사관리자가 통제할 수 없는 경우 행정담당자 또는 지방사무소등과 접촉하여 이를 보완하도록 한다. 이사, 재개발, 천재지변 등으로 인하여 추출틀에 대한 통제가 불가능하다고 예상되는 경우 행정담당자나 지역사무소 관계자와 긴밀히 협력하여 이를 최종 추출틀 구성에 반영하도록 한다.

**□ 지역 추출틀을 사용하는 경우 추출단위들이 지역적으로 중첩되지 않는 가를 확인하는 방법으로 지역에 대한 지도를 이용하는 것도 고려할 수 있다.**

지역단위 지도(요도)를 이용하여 표본단위들의 중첩성 여부를 확인할 수 있으며, 만일 중첩되는 단위가 발생할 경우에는 추출틀을 변경할 필요가 있다.

**□ 단일 추출틀을 사용하는 것이 여러 측면에서 바람직하지만, 필요에 따라서는 복수의 추출틀을 고려할 수 있다.**

복수의 추출틀을 사용할 경우 표본추출과정, 추정과정에서 이러한 점들을 모두 고려해야 하기 때문에 가능하면 단일 추출틀을 사용하도록 한다. 그러나 추출틀이 불완전 할 경우 상호 보완적인 복수의 추출틀을 고려할 수 있다.

## 2. 표본 설계

표본설계는 조사통계의 경우에만 해당되는 사항으로서 모집단을 잘 대표할 수 있는 표본을 추출하는 과정을 기술한 것이다. 조사통계의 경우 조사대상 모집단으로부터 가능한 객관적인 방법으로 표본을 추출하고, 추출된 표본이 얼마나 정확성을 유지할 수 있는지, 적용 가능한 조사 방법 등을 결정하는데 반드시 필요한 사항이다.

이를 위해 고려되어야 할 사항으로는 다음과 같은 것이 있다.

- 표본추출방법
- 확률추출방법 아니면 비확률 추출방법의 사용여부
- 예산과 정도를 유지하기 위한 적절한 표본규모의 결정
- 층화
- 무응답 대체방법
- 응답부담

#### 1) 표본추출방법

모집단의 특성을 잘 반영한 대표성 있는 표본을 추출하기 위해서는 먼저 조사모집단에 대한 분석이 이루어져야한다. 이러한 이유로 통상적으로 표본설계를 고려할 때 먼저 모집단에 대한 분석이 선행된다. 예를 들어 가구대상 조사를 수행하고자 할 경우에는 해당 연도에 이용 가능한 인구주택총조사 자료를 지역별, 연령별, 조사구 특성별 등으로 분석하여 모집단의 분포를 파악하게 된다. 이러한 선행 작업이 이루어지지 않게 되면, 모집단으로부터 추출된 표본의 대표성 여부를 판단할 수 없기 때문이다.

모집단으로부터 표본을 추출하는 방법은 크게 **확률추출방법(probability sampling)**과 **비확률 추출방법(nonprobability sampling)**으로 분류할 수 있다. 또한 확률추출방법은 **균등확률추출방법(equal probability sampling)**과 **불균등확률추출방법(unequal probability sampling)**으로 분류할 수 있다. 확률 추출방법은 모집단의 단위가 표본으로 선택될 수 있는 0이 아닌 확률이 존재하여, 우연현상에 의해 표본 단위를 선택하는 방법으로서 확률론의 이론에 입각한 추정치와 표준편차 등의 계산이 가능한 방법이다. 통상적으로 표본으로부터 계산된 추정치는 비편향성

(unbiasdness)과 일치성(consistency)등을 만족하게 된다. 추출된 표본에 입각한 모든 추정이론은 확률추출방법에 의해 추출된 경우에만 가능하기 때문에 오늘날 과학적인 표본본조사라 하면 당연히 확률추출법을 근간으로 한다.

확률 추출방법으로는 단순확률추출법(simple random sampling), 계통확률추출법(systematic random sampling), 층화확률추출법(stratified random sampling), 집락추출법(cluster sampling), 이중추출법(two-phase sampling 또는 double sampling), 확률비례추출법(sampling with probability proportional to size ) 등이 있으며, 추출단계에 따라 보다 복잡한 추출방법도 고려할 수 있다.

비확률 추출방법으로는 판단추출법(judgement sampling), 편의추출법(convenience sampling), 할당추출법(quota sampling), 눈덩이추출법(snowball sampling) 등이 있다. 이러한 비확률추출법은 표본의 크기가 크고, 모집단을 적절히 대표할 수 있다면, 확률추출법에 의한 추정치와 표준오차를 점근적인 방법으로 계산할 수 있으나 이는 제한적으로 적용가능한 방법이다.

#### [가이드라인]

□ **적용 가능한 다양한 표본추출법을 고려하여 가장 효율적인 표본추출방법을 선택해야 한다.**

표본추출방법은 최적(optimal)인 경우는 존재하지 않는다. 다만, 주어진 예산 하에서 모집단을 잘 대표할 수 있는 적정규모의 표본을 추출할 수 있는 방법을 고려해야 한다. 이때 가능한 한 확률추출방법을 고려하도록 한다.

□ **기존의 유사한 조사에서 적용한 표본추출방법을 참고하여, 이를 개선할 수 있는 방법이 있으면 개선된 표본추출방법을 적용하는 것도 고려할 수 있다.**

만일 현 조사와 유사한 조사에서 적용한 표본추출방법이 있으면, 비교가능성 측면에서 동일하거나 혹은 유사한 표본추출방법을 고려할 수 있다. 만일, 기존의 표본추출방법에 문제가 있다면, 이를 개선한 새로운 방법을 적용해야 할 것이다.

- **표본추출방법을 결정할 때 조사의 용이성, 조사비용 등을 감안한 추출방법을 고려하도록 한다.**

만일 조사 단위들이 지리적으로 멀리 떨어져 있지 않도록 하려면, 인근 지역들을 일정한 수로 집락화 하여 표본을 추출하는 방법을 고려할 수 있다. 또한 조사단위들 간의 조사비용이 서로 다를 경우 이를 반영한 추출설계가 이루어져야 한다.

- **조사에 적용한 표본추출방법을 가능한 한 상세히 기술해야 한다.**

표본추출과정에서 적용한 추출방법은 추후 추정치의 계산이나, 분산의 계산, 가중치의 계산 등에 적용되어야 하기 때문에 가능한 한 자세히 표본추출방법을 제시하는 것이 바람직하다.

- **가능한 자체가중설계(self-weighting design)가 되도록 표본을 추출하고, 만일 자체가중설계가 아닌 경우에는 표본추출방법에 따른 설계가중값(design weight)을 반드시 명시하여 추정과정에 반영할 수 있도록 한다.**

자체가중설계란 모집단에 속한 최종추출단위(ultimate sampling unit)들의 추출확률을 동일하게 설계하는 방법이다. 자체가중설계인 경우 최종조사단위의 가중값이 동일하게 되기 때문에 자료의 분석에서 별도의 추출설계를 고려하지 않고 범용 프로그램을 적용할 수 있기 때문이다. 그러나 자체가중설계가 용이하지 않을 경우에는 추후 분석에서 설계가중값을 고려해야 하기 때문에 이를 명시하도록 한다.

- **부득이하게 비확률 추출설계를 적용했다면, 그 이유를 상세히 밝히도록 한다.**

확률추출설계를 적용할 수 없는 경우 부득이하게 비확률추출설계를 적용하여 표본을 추출할 수밖에 없다. 이 경우 그에 대한 이유를 명확하게 밝혀둌므로서 조사통계자료 이용의 한계점을 이용자가 인식하게 해야 한다.

- **주기성을 갖는 조사에서는 가급적 유통성 있게 표본설계가 이루어지도록 한다.**

주기적인 조사의 경우 모집단과 표본에 변동이 발생할 수 있기 때문에 표본크기의 변동, 재층화, 추출확률의 조정 등이 가능하도록 한다. 또한 응답자의 부담도 사전에 고려하여 이를 적절히 반영할 수 있는 추출설계가 되도록 한다.

- 만일 비용과 조사효율성을 감안하여 표본설계에 대한 적절한 보완이 필요한 경우에는 보완과정을 명시하여 향후 이용자들이 보완사항을 이해할 수 있도록 한다.

표본설계과정에서 발생한 문제는 추정과정에 영향을 미칠 수 있기 때문에 반드시 전문가의 조언을 통해 문제점을 파악하고 이를 보완해야 한다. 또한 보완된 내용은 보고서에 명시하여 이용자들이 통계를 이용할 때 이를 반영할 수 있도록 해야 한다.

- 응답자의 응답부담을 경감시킬 수 있는 연동표본설계가 가능한지를 검토하도록 한다.

주기성을 갖는 조사에서는 지속적으로 응답자들이 조사에 응답해야 하기 때문에 응답부담이 상대적으로 크다. 이러한 점을 고려하여 연동표본설계(rotation sampling design)를 고려할 수 있다. 연동 표본설계는 보다 정밀한 표본설계와 추정이 필요한 만큼, 전문가의 조언을 통해 설계가 이루어지는 것이 바람직하다.

## 2) 표본규모의 결정

조사목적과 조사여건에 부합하는 표본크기를 결정하는 과정에서 표본규모가 커질수록 조사의 정밀도는 높아진다. 이점 때문에 일반적으로 조사관리자는 가능한 한 표본규모를 크게 하려는 경향이 있다. 그러나 표본의 규모가 커짐에 따라 조사비용이나 조사에 필요한 노력이 많이 들기 때문에 오히려 조사의 질을 떨어뜨리는 요인이 될 수 있다. 그러므로 표본규모는 조사목적에 맞는 목표정도(target precision)를 설정한 후, 이를 만족하는 범위 내에서 가능한 한 작게 규모를 결정하는 것이 바람직하다.

전체적인 추정 이외에 세부영역에 대한 추정을 병행하는 경우 전체 추정값 이외에 세부영역별 추정의 목표정도를 설정한 후 이를 만족시킬 수 있는 표본규모를 결정해야 한다. 표본의 크기는 이용할 추정량이 무엇이나에 따라 다를 수 있기 때문에 활용 가능한 보조정보의 존재여부를 파악한 다음 어떠한 추정방법을 사용할 것인지를 고려하여 표본규모를 결정하는 것이 필요하다.

통상적으로 표본의 규모는 복합표본설계에서 이용가능한 정보가 없을 경우 먼저 단순확률추출법에 근거하여 결정한다. 따라서 층화추출법이나 집락추출법등에 적합한 표본규모를 구하기 위해서는 설계효과(design effect: deff)를 단순확률추출에 근거하여 구한 표본규모에 곱해주어야 한다. 추정량의 정도는 추정량의 표준오차(standard error : SE), 오차의 한계, 변동계수(coefficient of variation:CV) 등으로 나타낼 수 있다.

표본규모를 조정함으로써 표본오차(sampling error)와 무작위무응답(random nonresponse)으로 인한 비표본오차를 관리할 수 있지만, 그 외 비표본오차는 관리가 불가능하다.

#### [가이드라인]

- 조사로부터 얻어지는 추정값에 대해 허용 가능한 불확실성의 정도(허용오차)가 어느 정도인지를 밝혀야 한다.

표본규모를 결정하기 위해서는 목표정도를 설정하고, 그 범위 내에서 표본규모를 결정해야 한다. 따라서 이를 위해 먼저 허용오차의 정도(degree)가 어느 정도인지를 밝히는 것이 필요하다.

- 표본규모를 결정하기에 앞서 모집단의 부분집합인 부차모집단(domain)들에 대한 추정이 필요한지를 결정해야 한다.

모집단 전체를 대상으로 추정량에 요구되는 정도(precision)와 모집단 내 각각의 부차모집단에 대한 추정량의 요구정도는 반드시 구별되어야 한다. 예를 들어 어떤

조사에서 전국적으로 추정량의 정도가 3% 내외 일 수 있지만, 시도별 요구정도가 5%이고, 군 구 별로는 10%가 될 수 있기 때문에 부차모집단의 추정이 필요한지를 결정할 필요가 있다.

□ 표본규모의 결정과정에서 복합표본추출설계를 고려할 경우에는 설계효과(deff)를 구하여 이용하는 것이 바람직하다.

□ 기존의 유사한 조사가 수행된 경우에는 표본규모의 결정과정에 이 정보를 유용하게 사용할 수 있다.

기존의 상대오차와 표본규모를 이용하여 현 조사의 목표오차를 정한 다음, 이 값을 표본규모 산정식에 적용할 수 있다.

□ 주어진 예산 하에서 목표정도를 만족하고, 조사가 가능하도록 현실성 있는 표본규모가 결정되도록 한다.

다양한 방법으로 표본규모를 결정할 수 있지만, 무엇보다도 예산 규모에 알맞게 그리고 목표정도를 만족하는 현실적인 표본규모이어야 한다. 너무 적거나, 너무 큰 규모의 표본규모는 자칫 조사로부터 얻은 자료가 무의미해지거나, 조사현장에서 설계내용을 반영하지 못하게 되는 문제가 발생할 수 있기 때문이다.

□ 표본규모를 산정하는 과정에 가능한 한 범용 프로그램을 이용하여 계산될 수 있는 시스템을 갖추도록 한다.

표본규모의 계산은 굳이 전문가가 아니더라도, 주어진 자료를 이용하여 충분히 계산이 가능하기 때문에 가능하면 범용프로그램을 이용하여 계산이 가능할 수 있도록 하는 것이 바람직하다.

□ 목표 표본규모와 실제 조사완료 후 실현 표본규모간의 차이의 원인을 파악하고, 이에 대한 조치를 강구하도록 한다.

결측 단위의 규모가 상대적으로 클 경우에는 예비표본을 이용하여 대체가 되도록 하는 것이 바람직하다. 예산과 정도를 감안하여 결정한 표본규모가 실제 조사 완료 후 표본규모와 어느 정도 차이가 나는지 파악하고, 적절한 조치를 수행해야 한다. 만일 이들 간의 차이가 일정 수준이상이면, 추가로 표본을 추출하여 조사가 될 수 있도록 해야 한다. 만일 그 차이가 일정 수준 이하인 경우에는 목표정도를 일부 수정하여, 실현 표본규모를 추정과정에 반영토록 한다.

### 3) 층화

표본설계에서 조사목적에 부합하도록 모집단을 층화하는 것이 무엇보다도 중요하다. **층화(stratification)**란 모집단을 단위의 특성에 따라 서로 동질적인 단위들끼리 몇 개의 부분집단으로 나누는 과정이다. 표본추출과정은 모집단을 몇 개의 층으로 나누는 후 각 층별로 독립적인 표본추출이 이루어지게 된다. 효과적인 층화가 이루어질 경우 추정의 효율성을 증가시킬 수 있으며, 층별 부분 통계의 산출이 가능하고, 예산을 절약할 수 있기 때문에 층화를 잘 하기위한 노력이 매우 요구된다.

실제로 가구조사의 경우 지역별, 조사구 특성별, 거처유형별 등으로 층화하여 표본을 추출하며, 사업체 조사의 경우에는 산업분류, 매출액 규모, 종사자수 규모, 지역 등을 층화변수로 고려하여 표본을 추출하고 있다.

모집단에 대한 층화 후, 각 층으로부터 적절한 크기의 표본을 독립적으로 추출해야 한다. 이 과정에서 층 별 표본규모를 결정해야 하며, 층별 표본배분방법으로는 비례배분(proportional allocation), 최적배분(optimal allocation), 네이만 배분(Neyman allocation), 멱배분(power allocation),  $\sqrt{Y}$  배분법 등이 있다.

#### [가이드라인]

□ 모집단을 층화 할 경우에는 층화의 필요성을 먼저 검토해야 한다.

층화추출설계를 적용하기에 앞서 먼저 현재의 조사에 층화 추출설계가 필요한지

를 먼저 검토해야 한다. 추정의 효율성과 비용의 절감 측면에서는 반드시 층화를 고려해야 하지만, 층화가 적절하지 않은 경우에는 굳이 적용할 필요가 없기 때문이다.

**□ 층화 변수를 고려할 때에는 변수의 속성이 지속적으로 유지되는 변수를 선택해야 한다.**

층화변수는 연구변수와 상관관계가 있는 변수를 선택하고, 변수의 속성이 시간의 흐름에 따라 변화하지 않는 변수가 층화변수로 적절하다. 이러한 층화변수는 모집단 내에 존재할 수 있거나, 또는 행정 자료 등에 존재할 수 있기 때문에 가급적 다양한 정보로부터 층화변수를 고려하는 것이 바람직하다.

**□ 층화를 고려할 때, 층의 수를 먼저 결정하기 보다는 다양한 층을 먼저 고려한 다음 이들을 결합해가는 방법으로 층수를 결정하는 것이 바람직하다.**

층의 수를 미리 정하는 것 보다 다양한 보조정보로부터 층을 고려한 다음 동일한 속성을 갖는 그룹들을 결합해 가는 방법으로 층수를 결정하는 것이 좋다. 왜냐하면, 층수를 미리 정할 경우 층에 속한 단위가 일정 규모에 미달하여 추정이 불가능한 경우가 발생하거나, 또는 누락으로 인한 포괄성 오류를 범할 수 있기 때문이다.

**□ 층별 표본크기를 결정할 때에는 가능한 다양한 측면에서 보다 효율적인 표본배분방법을 적용해야 한다.**

단순히 층별 모집단 규모만 제시된 경우에는 비례배분을 적용하고, 층별 변동과 층별 크기가 제시된 경우에는 네이만 배분, 비용까지 고려된 경우에는 최적 배분 까지 고려할 수 있기 때문이다.

**□ 비대칭성이 큰 모집단으로부터의 표본추출은 다른 단위에 비해 월등히 큰 단위가 조사될 수 있도록 반드시 고려해야한다.**

사업체 조사와 같이 모집단에 상대적으로 규모가 큰 단위가 있는 경우에는 가능한 한 표본에 반드시 포함되도록 하는 절사법을 이용한 층화가 바람직하다. 왜냐하

면 규모면에서 일부 사업체가 전체의 약 60~70%를 차지하고 있기 때문에 대표성 측면에서 이러한 단위가 표본에서 제외될 경우 심각한 과소 추정이 발생할 수 있기 때문이다. 따라서 일정규모 이상의 단위는 모두 조사하도록 하고, 나머지 단위에 대해서는 표본조사를 수행하도록 한다.

### 3. 측정도구(조사표)개발 및 테스트

#### 1) 조사표 개발 및 준비

측정도구란 조사목적에 부합하는 자료 수집을 위한 도구로서 통상적으로 조사표를 의미한다. 조사표는 조사대상의 특성을 재는 측정도구가 되기 때문에 잘못 설계된 조사표를 통해 수집된 자료는 통계자료로서의 의미가 없기 때문에 조사표 설계 과정에서 세심한 주의가 필요하다. 조사단계에서 발생하는 오차가 측정오차로서 측정오차의 원인에는 다음과 같은 사항들이 있다.

- 조사도구 : 자료 수집을 위해 사용한 조사표의 형태 또는 조사표, 측정도구 등이 잘못된 값을 기록하게 할 수 있다.
- 응답자 : 의식적이든 무의식적이든 간에 응답자가 잘못된 응답을 할 수 있다.
- 조사원 : 응답자가 제시하는 응답에 조사원이 영향을 끼칠 수 있다.

조사표설계과정에서 먼저 고려되어야 할 사항은 수집되어야 할 자료가 무엇인지를 명확히 규정하는 것이다. 따라서 통계이용자 그룹과 통계생산자 간의 충분한 토의를 거쳐 조사표를 설계하는 것이 바람직하다. 조사도구에 따라 종이와 필기도구를 이용하는 방법과 컴퓨터를 이용한 방법, 전화를 이용한 방법으로 구분할 수 있으며, 조사방식에 따라 조사원이 직접 질문하여 조사표를 작성하는 방식인 면접조사방식과 응답자가 조사표를 직접 작성하는 방식인 자기기입방식이 있다. 앞에서 살펴본 바와 같이 조사표는 자료의 품질과 응답자의 태도, 조사원과 응답자간의 관계에 중

대한 영향을 끼친다. 그러므로 조사항목을 최종적으로 결정하기 전에 시험 조사 (pilot survey)를 통해 조사표의 구성을 결정해야 한다. 이때 고려할 사항으로는 조사표의 형식, 논리적 흐름, 사용된 용어 및 분류체계의 명확성, 응답자 부담, 조사원 및 응답자의 조사항목에 대한 이해도, 면접소요시간, 응답을 위한 자료준비, 조사항목의 민감성 등이 있다.

이러한 측면에서 조사표의 개발과정에서는 통계이용자의 요구사항, 조사기관의 행정적 요구사항, 조사목적과 응답자 그룹의 특성, 조사현장의 상황, 자료처리에 대한 요구사항 까지를 고려하여야 한다.

#### [가이드라인]

- 통계 이용자 그룹의 요구사항이 적절히 반영될 수 있도록 조사표 개발단계에서 충분한 토의과정을 거치도록 한다.

조사를 통해 수집되는 자료는 먼저 자료를 이용하고자 하는 이용자의 요구를 반영해야 한다. 이러한 측면에서 통계생산자는 다양하고 전문적인 이용자그룹의 의견을 충분히 수렴하여 관련전문가의 전문성과 조사 설계자의 전문지식이 결합된 자료의 필요성에 대한 충분하고 구체적인 조사표 설계가 가능해진다.

- 조사표 설계 단계에서 수립된 조사문항에 대해 시험 조사를 통한 사전검증을 반드시 거치도록 한다.

조사표에 대한 검증은 많이 할수록 보다 좋은 조사표가 개발된다. 따라서 예비조사 또는 시험 조사를 통해 조사표에 대한 전반적인 사항을 검토하여 이를 최종안에 반영하도록 한다.

- 향후 자료 분석까지를 고려하여 미리 결과표를 작성하는 것이 유용하다.

결과표를 작성함으로써 조사표에 나타난 문항이 어떠한 형식으로 조사가 되어야 하고, 어떠한 결과분석방법을 적용해야할 지를 판단할 수 있기 때문이다.

□ 조사표에 대한 품질을 검토할 수 있는 체계를 수립하도록 한다.

현재 진행 중인 조사의 조사표에 대해 사후 품질 검토를 위한 체계를 수립하여 지속적으로 보완할 수 있도록 한다.

2) 용어의 정의 및 분류

생산된 통계에 대해 이용자 관점에서 통계자료의 개념과 정의를 명확하게 인식할 수 있도록 해야 한다. 또한 비교가능성, 종합성 등을 평가하기 위해 가능한 한 표준화된 정의와 분류체계를 이용하는 것이 바람직하다. 조사표상에 이용된 용어의 개념이나 정의 및 분류기준은 관련 지침서에 명확히 제시하도록 하며, 통상적으로 사용되는 개념과 다를 경우에는 이용자의 혼선을 방지하기 위해 이러한 사실을 공지하도록 한다. 예를 들어 가구조사의 경우 통계청에서 사용하고 있는 “가구”의 정의는 『1인 또는 2인 이상이 모여서 취사, 취침 등 “생계를 같이하는 생활단위”』로서 이 경우 혈연관계가 없는 사람도 포함될 수 있다는 점에서 “가족”, “가정” 또는 주민등록상의 “세대”와 다를 수 있다.

따라서 여타의 다른 조사에서 가구의 개념을 통계청의 정의에 따르지 않을 경우에는 구체적으로 가구의 개념을 기술할 필요가 있다.

[가이드라인]

□ 조사표에 사용되는 용어의 정의를 명확하게 하도록 하고, 이러한 사항들은 반드시 조사지침서에 수록하여 조사원이 참고할 수 있도록 한다.

동일한 용어를 사용하고 있음에도 조사원과 응답자간의 이해의 정도가 다르게 되면, 응답오차가 발생한다. 따라서 가능한 한 사용하는 용어에 대해 쌍방이 동일한 개념으로 이해할 수 있도록 용어를 명확히 정의하는 것이 필요하다.

□ 국가별 비교를 위해 국제기구 등에서 사용되는 조사표의 경우 한글 번역과정에

**서 가능한 한 의미가 훼손되지 않도록 직역하여 문항을 구성하도록 한다.**

국제기구 등에서 제공하는 조사문항의 경우 이용목적이 국제비교를 위한 경우이므로 가능한 해당 분야 전문가의 도움을 받아 조사문항을 구성하는 것이 바람직하다. 잘못된 의역보다는 직관적인 직역이 질문의 의도에 적합하기 때문이다.

**□ 표준화된 정의와 분류체계를 이용하도록 한다.**

국내의 경우 통계청, 노동부, 산업자원부 등에서 제공되는 표준분류체계와 그밖에 UN, Eurostat, ILO, ISO 등의 국제기구 등에서 정의한 국제표준분류들의 분류체계와 일치시키도록 한다. 정의와 분류체계를 표준화함으로써 서로 다른 출처나 조사에서 수집된 자료나 통계들을 연계시키거나, 비교가 가능해지기 때문이다.

**3) 조사표 설계**

조사표는 응답자로부터 자료 수집을 위한 질문 문항들의 모임이다. 잘못 작성된 조사표는 결과적으로 무의미한 자료를 수집하는 오류를 범하게 된다. 따라서 조사표 작성과정에서 염두에 두어야 할 사항은 조사표의 형식, 논리적 흐름, 사용된 용어 및 분류체계의 명확성, 응답자 부담, 조사원 및 응답자의 조사항목에 대한 이해도, 면접소요시간, 응답을 위한 자료준비, 조사항목의 민감성 등이다.

조사표는 응답자가 조사표를 쉽게 이해하고 답할 수 있도록 작성해야 하며, 용어의 개념이나 정의, 분류 등에 혼선이 없도록 해야 하며, 응답자 측면에서 응답이 곤란한 사항에 대해서는 별도의 항목으로 구분하여 응답이 가능하도록 해야 한다. 또한 질문의 순서, 어귀, 문장 등에 대해서도 심혈을 기울여 작성도록 한다. 질문항목의 구성은 가능한 폐쇄형으로 구성되도록 하고, 필요에 따라서는 개방형 질문을 삽입하도록 한다. 폐쇄형 질문으로 포괄되지 않는 항목에 대해서는 기타항목과 같이 별도의 개방형으로 응답할 수 있는 칸을 두는 것도 좋은 방법이다. 질문항목의 논리적 흐름을 반드시 점검하여 응답오차가 발생하지 않도록 해야 한다.

조사표에 대한 사전 검증 방법으로는 사전조사, 포커스그룹 면접, 심층면접, 조사표에 대한 생각을 여과 없이 대화하는 방법, 부연설명, 행위코딩, 분할 표본검증 방

법 등이 있다. 이러한 검증방법들은 각 방법별로 특징을 가지고 있기 때문에 관련 전문가의 조언을 통해 검증방법을 채택하도록 한다.

#### [가이드라인]

- 조사표상에 표현되는 질문항목들은 응답자의 관점에서 사전 검증된 항목으로 구성되도록 한다. 특히 용어와 응답시간, 논리흐름에 대한 검토가 이루어지도록 한다.

아무리 잘 구성된 조사표라 하더라도 응답자가 이해하기 어려운 문장을 사용하거나, 응답시간이 너무 긴 경우에는 응답과정에서 응답자의 협조를 이끌어내기 어렵다. 따라서 최종적인 조사문항에 대해 응답자의 관점에서 조사관리자 또는 조사표작성자가 응답자의 관점에서 예비적으로 조사표를 작성해보고 그에 대한 문제점을 검토하는 것이 필요하다.

- 조사표 서두에는 조사의 제목 또는 주제, 후원자, 조사목적, 응답자 협조요청, 조사기관, 응답자에 대한 비밀보호정책, 연락처 등에 관한 정보를 제시하도록 한다.

응답자가 조사에 대해 100% 신뢰를 갖는 경우는 거의 없다. 따라서 조사협조를 이끌어 내기 위해서는 조사에 대한 모든 정보를 응답자에게 제공함으로써 상호간 신뢰감을 높일 수 있다. 특히 응답자 개인정보의 비밀보호는 아무리 강조해도 지나치지 않는다. 응답과정에서 질문에 대한 문의사항이나 조사에 대한 전반적인 문의사항이 발생할 경우를 대비하여 연락처를 기입해두는 것도 신뢰성을 높이는 좋은 방법이다.

- 조사표상의 분류를 용이하게 하기 위해 각 장별 제목이나 머리말을 두는 것도 바람직하다.

조사표의 구성은 구조적으로 이루어지는 것이 바람직하다. 따라서 조사 주제별로

구분하여 조사표를 구성하게 되면, 응답자는 대략적으로 조사에 응답할 내용을 미리 염두에 둘 수 있어 응답과정이 보다 수월 할 수 있고, 결과분석의 측면에서는 주제별 분석이 가능하기 때문에 바람직하다.

**□ 응답자의 주의를 환기시키기 위한 다양한 디자인을 채택하고, 그림 또는 삽화 등을 적용하여 흥미를 유발시키도록 한다.**

응답자는 여러 가지 측면에서 가능하면 조사에 응답하기를 꺼리는 경향이 있다. 응답자의 개인 비밀 노출, 지루한 조사문항 등이 한 요인이 될 수 있다. 구성이 뛰어난 조사표는 당연히 응답자의 흥미를 유발시킬 수 있어야 하며, 응답과정에서 지루함을 느끼지 않도록 해야 한다. 이러한 측면에서 다양한 그림이나 삽화를 삽입하여 환기를 시켜주는 것도 좋은 조사표 구성이라 할 수 있다.

**□ 조사대상 기간, 응답단위(조사대상)에 대한 사항이 질문내용에 포함되어 측정오차가 발생하지 않도록 한다.**

통상적으로 조사표를 이용한 조사의 대부분은 현재 상황에 대한 질문보다는 과거 특정 시점에 일어난 사건에 대한 질문으로 구성된다. 이 경우 조사대상 시점을 명확히 제시하여 응답자로 하여금 회상오류를 범하지 않도록 할 필요가 있다.

**□ 종단면 패널 조사와 같이 주기적 조사에서는 질문항목의 일관성 유지가 중요하다. 따라서 이전의 조사표를 검토하여 누락되거나, 새로 추가된 항목이 있는지 여부를 검토하도록 한다.**

주기성을 갖는 조사에서는 조사표상의 일관성이 매우 중요하다. 특히 패널조사와 같은 종단면 조사에서는 더욱 그러하다. 월별, 분기별, 연간 조사에서 사용되는 조사표는 각 항목별로 전월, 전분기, 전년에 작성된 항목과 일관성을 유지하고 있는지를 검토하고, 신규로 삽입된 문항에 대해서는 조사지침에 명기하여 추후 분석과정에서 참고할 수 있도록 해야 한다.

□ 조사원은 조사에 필요한 응답자 가이드 또는 지침서를 지참하여 응답에 참고가 되도록 한다.

조사에 투입되기 전에 조사원은 일정기간 조사에 대한 교육을 받는다. 그러나 교육을 받은 조사원이라 할지라도 응답자가 질문문항에 대한 전문적이고, 구체적인 사항에 대해 문의할 경우 그에 대해 조사원이 쉽게 설명해주기란 어렵다. 이러한 경우를 대비하여 조사원은 항상 조사지침서를 지참하여 응답자의 문의에 현장에서 응답할 수 있도록 하는 것이 필요하다.

□ 조사표의 맨 마지막 부분에는 조사협조에 대한 감사의 표현이 제시되도록 하고, 응답자가 추가적으로 기입할 수 있는 공란을 두는 것도 좋은 방법이다.

조사에 응답하는 응답자는 꼭 조사에 응답해야할 의무가 없다. 따라서 응답자의 중요한 정보를 제공한 데 대한 감사의 표시를 제시하는 것이 조사예절에 부합되며, 만일 응답자가 조사에 대한 특정 소견이나, 개선사항 등에 대한 아이디어가 있을 경우를 대비하여 맨 마지막 장에 공란을 두는 것도 좋은 방법이 될 수 있다.

### 제3절 자료수집

통계생산을 위한 기본적인 절차 중의 하나는 통계산출의 대상이 되는 자료의 수집이다. 자료 수집은 자료를 얻기 위한 수집활동과 더불어 자료수집에 필요한 모든 지원활동을 포함한다. 즉, 협의의 자료 수집은 단순히 조사를 통한 자료의 취합만을 의미하며, 광의의 개념은 이러한 절차 모두를 포괄한다고 할 수 있다. 따라서 자료수집에 필요한 제반요소를 적절히 활용하여 오차 없는 높은 품질의 자료를 얻어야 한다. 자료 수집을 위한 도구로는 종이조사표, 전자적 조사표, 웹조사표 등이 있다. 이러한 자료수집 도구를 이용할 때에는 각 방법별 특징을 파악하고, 현실성 있는 조사표를 이용하는 것이 바람직하다. 최근에는 종이조사표 방식 대신에 IT 기술의 발전으로 컴퓨터를 이용한 전자적 조사표 방식을 도입하여 사용하기도 하며, 일부 특

수한 조사에서는 웹 조사표 방식을 사용하기도 한다. 이러한 전자적 방식의 조사를 도입하고자 할 때에는 조사기관의 역량, 조사원의 컴퓨터활용능력, IT 인프라 등을 감안해야 한다. 단순히 비용, 노력 등만을 고려하여 기반시설을 활용할 수 없는 조사방식을 적용하게 되면, 자료수집이 불가능할 수 있기 때문이다.

자료수집과정에서 발생하는 오차중의 하나가 무응답 오차이다. 따라서 무응답에 대한 적절한 대책을 수립하는 것 또한 자료수집과정에서 중요하게 고려되어야 할 사항이다. 무응답은 조사초기에 수립한 조사대상 표본수를 축소하여 왜곡된 추정값을 산출할 수 있으며, 조사의 신뢰성을 떨어뜨리는 요인이 되기 때문이다.

다음으로 고려할 사항은 조사원에 대한 채용과 훈련이다. 아무리 잘 설계된 조사표라 할지라도 자료를 수집하는 조사원이 업무를 제대로 수행하지 못하면, 양질의 자료를 수집하기 어렵다. 조사원 채용과정에서는 조사의 목적과 조사방법 등을 제대로 이해하고 있는지에 대한 검증이 필요하며, 가능하면 응답자에게 친숙함을 줄 수 있는 조사원을 채용하는 것이 바람직하다. 조사원들은 반드시 조사표에 대한 전반적인 사항, 응답을 얻기 위한 다양한 방법, 응답자에 대한 예절, 자료의 기입 방법 등 제반 과정을 교육과 훈련을 통해 숙지하도록 해야 한다.

조사현장에서 조사원은 다양한 상황을 접하게 된다. 예를 들어, 조사대상 가구의 철거로 인해 응답을 얻을 수 없는 경우, 또는 수해나 천재지변 등으로 인해 조사지역의 접근이 불가능한 경우, 조사대상 사업체의 파산으로 해당 사업체가 사라진 경우 등이 있을 수 있다. 또한 조사팀 내의 조사원의 유고로 인해 조사업무를 수행할 수 없는 경우도 발생할 수 있다. 이러한 사항들을 시의 적절하게 대응할 수 있는 교육과 훈련이 필요하다. 특히 조사지도원(슈퍼바이저)은 다양한 상황에 대한 적절한 대비가 될 수 있도록 조사원들을 적절히 통제하고, 관리하여야 한다.

자료는 반드시 조사를 통해서만 수집되는 것은 아니다. 행정자료와 등록자료, 전산처리과정에서의 자료 등 다양한 자료원이 있으며, 이러한 자료를 수집하기 위해서는 해당 자료에 대한 다양한 정보를 파악하고 있어야 한다. 행정자료의 특성상 조사에 의한 자료가 아님으로 표본오차 등의 제한을 받지 않으며, 모집단 자료로 간주되는 것이 일반적이다. 또한 법률적인 강제성을 갖는 경우가 많기 때문에 특히 행정자

료와 등록 자료를 수집하고자 할 때에는 자료에 대한 보안관리를 철저히 해야 한다. 단순히 통계작성목적으로 이러한 자료를 수집하는 과정에서 자료의 유출이나 보안사항의 누설로 인해 응답자가 피해를 입지 않도록 주의를 기울여야 한다.

### 1. 자료원천(소스)

자료의 원천이란 수집하고자 하는 자료의 형태와 어떠한 방식으로 자료를 수집해야 하는가를 나타낸다. 즉, 조사를 통해 자료를 수집할 경우에는 자료의 원천은 조사자료(survey data)가 될 것이고, 행정절차를 통해 산출되는 자료를 수집할 경우에는 자료의 원천은 행정자료(administrative data)가 된다. 또한 행정자료를 1차 가공한 자료를 수집할 경우에는 자료의 원천은 가공자료(processing data)가 된다. 따라서 자료의 원천에 따라 자료를 수집하는 방식과 기법에 차이가 있다.

조사 자료를 수집하기 위해서는 조사표를 이용해야 한다. 조사표는 작성방식에 따라 면접조사방식(face-to-face interview), 자기기입방식(self-administrative interview)으로 구분할 수 있다. 또한 조사에 사용되는 조사표의 형태에 따라 종이조사표를 이용한 조사방식(PAPI), 컴퓨터를 이용한 조사방식(CAPI), 컴퓨터를 이용한 전화조사방식(CATI)등으로 구분할 수 있다. 한편 조사 자료와 행정자료를 결합하여 자료를 수집하는 방법을 혼합조사방식(Mixed mode)이라 하며, 이 경우 조사 자료와 행정자료간의 연계율과 연계하고자 하는 행정자료의 품질을 고려해야 한다.

#### [가이드라인]

□ 자료 수집을 위한 단계별 효율성을 극대화 하고 자료의 품질을 높일 수 있는 적절한 방법을 활용해야 한다.

조사표가 설계되어 조사를 수행할 경우 예산, 조사기관의 역량과 조사원의 능력과 조사인프라 등을 감안하여 단계별로 가장 효율적인 방법을 선택하도록 한다. 최신의 기법을 적용하여 자료를 수집하는 방법도 고려할 수 있지만, 이는 대내외적인

조사환경에 따라 적절하게 선택하여 사용해야 할 것이다.

**□ 조사 자료를 수집하기위해 사용한 자료수집방법의 근거와 해당 사항을 구체적으로 밝히도록 한다.**

자료를 수집하기위한 기법 즉, 조사표의 유형과 조사방식 등에 관한 사항을 밝힘으로서 자료가 어떠한 절차로 수집되었는가에 대한 정보를 이용자가 파악할 수 있도록 해야 한다. 예를 들어 조사방법은 “면접조사방식”, 조사표 방식은 “종이조사표” 라고 명시함으로써 이용자가 자료를 어떠한 방식으로 수집하였는가를 쉽게 이해할 수 있기 때문이다.

**□ 자료수집 절차에 대한 전반적인 체계를 수립하여, 그에 따른 자료수집이 이루어지도록 해야 한다.**

자료수집 체계로 고려할 사항은 조사일정, 조사비용, 일일보고, 무응답 대책, 에디팅, 자료의 전송방법 등이다. 전반적으로 조사원과 응답자에 대해 무리가 되지 않도록 주어진 예산 하에서 유연하게 일정을 수립해야 하며, 조사과정에서 발생하는 문제점들을 즉시 처리할 수 있도록 일일 보고 또는 주간 보고가 이루어지도록 해야 한다. 또한 조사가 완료된 조사표는 현장에서 슈퍼바이저에 의해 검토되도록 하고, 에디팅 과정에서 품질이 떨어지는 조사표는 반드시 재조사(call-back)를 수행토록 해야 한다. 에디팅 과정을 거쳐 완성된 조사표는 분실되거나 파손될 우려가 있기 때문에 일정기간이 지난 후 신속하게 중앙으로 수거될 수 있도록 해야 한다.

**□ 조사 자료와 행정자료를 연계하여 자료를 수집하는 경우에는 연계의 필요성 여부를 검토하고, 이들 간의 효과를 높일 수 있도록 기관이나 관리자의 협조를 얻어 가능한 많은 정보를 취합하여야 한다.**

조사 자료와 행정자료의 연계는 조사 자료로부터 수집이 어렵거나 기존의 자료를 최대한 이용함으로써 조사비용과 노력을 절감할 수 있다. 이러한 측면에서 두 자료간의 연계가 필요하다고 판단되면, 해당 기관이나 관리자로부터 적극적인 협조를

이끌어 낼 수 있도록 해야 하며, 연계과정에서 필요한 정보를 파악하여 가능한 한 많은 정보를 얻는 것이 중요하다.

**□ 연계에 적용 가능한 방법을 전문가의 조언 통해 검토해야 한다.**

조사 자료와 행정자료 또는 등록 자료를 연계하기 위한 방법은 직접적 결합 방식과 통계적 결합 방식이 있으며 두 방법간의 장단점을 검토하여 최선의 방법을 선택해야 한다. 이를 위해 통계작성기관내에 통계전문가의 조언을 통해 결정되는 것이 바람직하며, 만일 내부 전문가가 없을 경우에는 외부전문가의 조언 통해 결정하도록 한다.

**□ 두 자료의 연계과정에서 발생할 수 있는 문제를 다각도로 검토하여 오류를 최소화해야 한다.**

두 자료간의 연계는 우선 연계에 필요한 키변수(key variable)가 두 자료에 모두 존재해야 한다. 만일 이러한 내용을 충족시키지 못할 경우에는 연계 자체가 불가능하기 때문이다. 또한 연계과정에서 행정자료에 포함된 단위 중 중복 기재된 단위나 동일한 단위가 서로 다른 속성을 갖는 다중단위가 존재할 경우 오류가 발생하게 된다. 이는 행정자료의 포괄성과도 관련이 있기 때문에 반드시 연계되는 자료의 품질을 검토할 필요가 있다. 이러한 측면에서 두 자료간의 연계성공률 등을 기록하여 두 자료간의 일관성을 평가할 수 있다.

**□ 응답자 기입방식의 조사에서는 응답자가 조사지침서를 참고하여 조사표를 작성하기 때문에 조사지침서의 내용이 이해하기 쉽도록 작성되어야 하며, 이를 위한 다양한 방법을 강구해야 한다.**

응답자 기입방식의 조사는 조사원의 도움보다는 응답자 스스로 조사항목에 대한 이해를 통해 조사표가 작성되기 때문에 조사지침서의 내용이 너무 전문적이거나, 이해하기 어렵게 작성되어서는 안 된다. 응답자는 각계각층의 다양한 수준의 사람들이 될 수 있기 때문에 가능한 한 이해가 쉬운 문장을 사용하고, 이해를 돕기 위한 그림

이나 그래픽 등을 이용하여 작성하는 것이 바람직하다. 만일 컴퓨터를 이용하여 응답자 기입방식으로 조사가 되는 경우에는 응답과정에서 실수로 인해 응답할 수 없는 경우에 대한 자세한 방법을 기술해주어야 한다. 웹 조사 방식인 경우에는 응답절차나 흐름을 따라가기 쉽도록 조사지침서를 작성하는 것이 바람직하다.

**□ 컴퓨터를 이용한 조사(CAPI)의 경우 항목무응답에 대한 모듈을 적용해야 한다.**

컴퓨터를 이용한 조사의 경우 각 조사항목에 대해 무응답한 경우 다음 항목으로 넘어가지 않도록 하는 모듈을 적용할 필요가 있으며, 만일 응답을 할지라도 특정 범위를 벗어난 값을 입력할 경우 오류메시지를 나타내주는 자동 에디팅 기능까지 적용된 프로그램을 사용하는 것이 좋다. 이 경우 별도의 에디팅 절차를 거칠 필요가 없기 때문에 비용과 일정을 절약할 수 있기 때문이다.

**□ 컴퓨터를 이용한 전화조사(CATI)의 경우 전화번호의 추출방식과 전화접속 건수 등에 대한 정보를 명시해야 한다.**

전화 조사에 대한 응답률 통계는 전화접속 건수로부터 계산된다. 따라서 전체 전화접속건수로 응답건수를 나눈 값을 최종 응답률로 산정해야한다.

한편 전화조사의 경우에는 가장 최신의 전화번호부를 추출틀로 사용하는 것이 바람직하다. 이 경우 포괄성을 반드시 검토해야 한다. 만일 포괄성이 현격히 떨어지는 경우에는 임의전화걸기(random digit dialing: RDD) 등 포괄성을 개선할 수 있는 방법을 이용하는 것이 바람직하다.

## 2. 조사원 채용 및 교육

조사원은 조사의 품질 나아가서 자료의 품질을 좌우할 만큼 조사에 있어서 중요한 요소 중 하나이다. 조사에 관한 조사원의 이해력, 태도, 사명감 등이 응답자의 응답형태에 밀접하게 작용하기 때문에 조사원의 채용과 교육훈련과정이 매우 중요하다. 조사에 참여하는 인력은 크게 세 부류로 분류한다면 (1) 조사전문가 (2) 조사지

도원 (3) 조사원으로 분류할 수 있다. 조사전문가는 조사의 주제, 관리, 표본이론, 조사표설계, 자료처리 등 조사의 전반적인 부분을 상세히 파악하고, 이해할 수 있을 뿐만 아니라 조사의 다양한 측면을 이해하는 조사통계 전문가라고 할 수 있다. 조사지도원은 조사에 대한 실무경험을 보유하고, 실제조사의 구현과정을 잘 이해하고 감독할 수 있는 사람이다. 조사원 훈련, 현장 사전점검, 현장조사감독, 자료입력 및 에디팅 등의 제반과정을 수행할 수 있는 사람이다. 조사원은 실제조사현장에서 응답자와 밀접하게 관계된 사람으로 조사를 담당하고, 자료를 입력하는 일을 수행하는 사람이다.

조사원에 대한 교육은 단기간에 효율적으로 수행되어야 한다. 특히 최근에 정보통신기술의 발달로 인해 컴퓨터를 이용한 조사가 점차적으로 확대되고 있는 실정에서 조사현장에서 조사원의 컴퓨터 활용능력의 유무가 조사의 성패를 좌우할 수 있다. 따라서 이러한 측면을 고려하여 조사원채용부터 교육훈련에 이르기 까지 철저한 관리를 통해 교육훈련을 극대화 할 수 있도록 해야 한다. 이와 더불어 조사원의 조사참여를 독려하기 위한 다양한 인센티브정책을 수립할 필요가 있다. 통상적으로 전국단위 조사의 경우 조사기간의 1개월 이상 소요되는 경우가 많으므로 조사원의 조사에 대한 부담이 매우 크다. 그러므로 이에 대한 적절한 보상정책을 수립하여야 할 것이다.

조사과정에서 사고나 개인적인 사유로 조사수행이 불가능한 경우가 발생할 수 있기 때문에 일정수의 예비조사원을 보유하여 적시에 유고된 조사원을 대처할 필요가 있다. 이러한 제반 업무는 조사전문가와 조사지도원간의 원활한 대화통로의 확보로 가능하다. 조사원의 관리를 위해 일일 또는 주간, 월간 업무보고 절차를 마련하는 것도 하나의 방법이 될 수 있다. 일일 또는 주간, 월간 접촉 가구 수, 조사완료 가구 수 등을 파악하도록 한다. 이때 조사지도원은 조사원들의 각종 애로사항을 적시하여 이를 조사전문가에게 통보하도록 하여 조사과정에서의 민원을 해결할 수 있도록 한다. 일회성조사의 경우에는 단기간의 교육훈련을 수행하지만, 중단면 조사와 같은 계속조사의 경우에는 조사원에 대한 재교육과 평가기회를 제공하여 조사방법 등을 개선할 수 있도록 해야 하며, 일정한 재충전 기회를 제공하는 것도 필요하다.

[가이드라인]

- 조사원에 대한 채용절차를 중앙 차원에서 체계적으로 수립하여 조사원의 질을 보장할 수 있도록 한다.

조사원은 조사통계 생산의 첨병 역할을 수행하는 중요한 요소 중 하나이다. 조사 과정에서 조사원에 의한 편향을 최소화 하기위한 제반의 노력이 요구되며, 이를 위해 먼저 조사원 채용에 대한 일정, 채용조건, 면접방식, 평가방법, 채용방식 등에 대한 계획을 수립하여 양질의 조사원을 확보할 수 있도록 한다. 이때, 조사지도원의 의견을 반영하는 것도 좋은 방법이 될 수 있다.

- 조사체계를 수립할 때에는 조사원의 평균 면접시간, 평균 면접자수 등을 검토하여 조사일정을 수립하도록 한다.

조사원의 조사부담을 고려하여 조사일정을 수립해야 한다. 1일 동안 조사원이 면접 가능한 회수, 이동시간, 면접시간 등을 반드시 고려하여 일정을 수립하도록 한다. 조사원의 능력을 벗어나는 무리한 일정을 수립하게 되면, 응답의 질을 떨어뜨리게 되며, 따라서 양질의 자료를 수집하기 어렵기 때문이다.

- 조사에 대한 전반적인 이해를 돕기 위해 다양한 교육훈련 프로그램을 개발하도록 한다.

조사에 대한 교육훈련의 성패가 측정오차를 줄이는 한가지 방법이 될 수 있다. 이를 위해 교육훈련 프로그램을 다양하게 수립하여 교육의 성과를 극대화 할 수 있도록 한다. 역할분담, 가상훈련, 조별 토의 등을 통해 조사과정에서 발생할 수 있는 다양한 상황에 대처할 수 있는 능력을 배양하도록 한다. 조사원이 타성에 젖지 않도록 새로운 교육 프로그램을 적용하고, 조사지침서에 대한 숙지정도를 평가할 수 있도록 토의 시간을 할애하는 것 또한 중요하게 고려되어야 한다.

- **조사목적, 조사대상, 조사대상 기간, 무응답처리방법, 추적조사원칙, 응답자 비밀 보호정책 등 조사원이 반드시 숙지해야 할 사항을 검토하여 조사과정에서의 오류를 방지할 수 있도록 한다.**

조사원이 응답자에게 조사전반에 대한 설명을 함으로서 응답자의 신뢰도를 높일 수 있다. 이를 위해 조사원이 반드시 숙지해야 할 사항을 검토하여 응답자로부터 잘못된 응답을 얻지 않도록 해야 한다. 특히 조사대상자 아닌 사람에 대해 조사하거나 조사대상기간을 혼동하여 조사대상 기간 외 자료를 수집하는 일이 없도록 해야 한다.

- **조사지도원에 대한 업무 분담을 통해 효율적인 조사감독이 이루어지도록 한다.**

조사지도원은 조사전문가와 조사원간의 교량역할을 수행할 뿐만 아니라, 조사현장에서 발생하는 다양한 상황에 즉시 대처할 수 있도록 하는 업무를 부여해야 한다. 조사에 대한 전반적인 사항을 이해하고, 조사현장에서 필요한 제반 행정 절차 등을 숙지하여 조사원이 원활한 조사를 진행할 수 있도록 협조해야 한다.

- **조사에 필요한 조사원 이외에 예비조사원을 확보하도록 한다.**

조사원의 조사 과정에서 중도에 탈락하여 원활한 조사수행이 어려운 경우를 대비하여 일정 수의 예비조사원을 확보해야 한다. 이때, 예비조사원은 동일한 수준의 교육훈련에 참여하도록 하여 필요한 경우 적시에 조사에 투입될 수 있도록 한다.

- **조사과정에서 응답자에 대한 가능한 한 많은 정보를 얻도록 한다.**

조사원은 가능한 한 응답자에 대해 조사항목 이외에 많은 정보를 획득하여 향후 자료처리과정에 반영토록 한다. 에디팅 과정에서 응답자에 대한 정보는 매우 유용하게 사용될 수 있기 때문에 조사항목이외에 참고가 될 만한 사항을 별도의 기록장에 기록하여 일일보고 또는 주간보고사항에 기재할 수 있도록 한다.

### 3. 무응답 줄이기

조사의 전 과정에 참여하고 있는 조사관계자들의 노력에도 불구하고 조사 무응답은 발생하기 마련이다. 이러한 조사무응답은 전체조사일정 및 조사 품질에 영향을 끼친다. 그러므로 사전에 조사 무응답에 관한 대비책을 강구하여 조사의 품질을 일정 수준 이상으로 유지, 관리할 수 있어야 한다. 무응답의 원인은 다양한 측면에서 고려될 수 있는데, 추출틀의 포괄성 문제로 조사대상자가 존재하지 않아서 발생하는 무응답, 조사과정에서 조사에 응답하기를 거절하여 발생하는 무응답, 특정 문항에 응답하지 않아서 발생하는 무응답 등 다양한 원인이 있을 수 있다. 무응답의 유형은 크게 2가지로 구분할 수 있으며, 하나는 응답자가 조사자체에 불응하여 발생하는 무응답인 단위무응답(unit nonresponse) 또는 전체무응답(total nonresponse) 이다. 이러한 단위무응답은 조사표 전체문항이 결측이므로 조사표상에서 이용할 만한 정보가 없게 된다. 다음으로 조사표상의 일부 조사항목에 대해 응답하지 않아서 발생하는 무응답인 항목무응답(item nonresponse) 이다. 항목무응답은 조사항목 중 응답한 항목 일부를 이용할 수 있기 때문에 앞에서 언급한 단위무응답과는 다른 처리가 강구된다. 일반적으로 무응답 발생은 설계당시에 설정한 표본크기를 작아지게 함으로서 추정량의 분산을 증가시키는 요인이 되며, 결과적으로 조사의 효율성을 떨어뜨리게 된다. 따라서 가능한 한 무응답이 발생하지 않도록 하는 것이 일차적인 관심이 되어야 하며, 만일 무응답이 발생했다면, 이에 대한 적절한 대책이 마련되어야 한다.

〈표 4-3-1〉 무응답의 유형과 처리방법

| 무응답 유형 | 처리방법                             |
|--------|----------------------------------|
| 단위무응답  | 가중셀조정(Weighting cell adjustment) |
|        | 래킹비 조정(Raking ratio adjustment)  |
|        | 보정(Calibration)                  |
| 항목무응답  | 대체(Imputation)                   |

[가이드라인]

□ **조사과정에서 무응답을 방지하기위한 추적조사에 관한 계획을 수립하도록 한다.**

조사과정에서 응답자의 부재, 미접촉, 장기출타 등의 사유로 무응답이 발생한 경우를 대비하여 조사일정을 수립할 때 추적조사원칙을 함께 고려하도록 한다. 대개의 경우 1차에서 3차까지 추적조사를 수행토록하고 있다. 특히 무응답으로 인한 영향이 큰 단위에 대해 우선 조사원칙으로 정하는 것이 일반적이다.

□ **조사표의 구조적인 결함, 조사도구의 고장에 의해 발생할 수 있는 무응답을 방지하기 위해 컴퓨터를 이용한 조사의 경우 사전에 충분한 검토가 필요하다.**

측정도구의 결함으로 발생하는 무응답은 에디팅 과정에서 무응답의 원인을 규명하기가 매우 어려우며, 특히 컴퓨터를 이용한 조사의 경우 컴퓨터의 결함, 망실 등으로 자료의 일부를 이용하지 못하는 경우가 발생할 수 있기 때문에 조사 수행 전에 반드시 점검하도록 한다.

□ **응답자의 응답을 독려하기위해 가능한 범위 내에서 조사방식을 응답자가 선택할 수 있도록 하는 방법도 고려할 수 있다.**

조사원면접방식의 경우보다는 응답자 기입방식의 조사에서 컴퓨터에 대한 거부감 등으로 응답을 회피할 수 있기 때문에 가능한 범위 내에서 조사방식을 응답자가 스스로 결정할 수 있도록 하는 방법도 바람직할 수 있다.

□ **응답자의 응답내용이 타당하지 않을 경우 응답자에게 재전화 또는 재방문(callback)하는 절차를 수립하는 것이 바람직하다.**

무응답이 발생한 경우 응답을 얻기 위해 추적조사를 수행하고, 만일 응답한 내용이 타당하지 않을 경우에는 조사원이 응답자를 재방문 하거나 재전화를 할 수 있도록 해야 하며, 재방문이나 재전화 절차를 수립하여 체계적으로 응답값에 대한 타당

성을 점검하도록 해야 한다.

**□ 응답자의 정보는 매우 귀중한 자원이다. 따라서 응답자에게 이에 대한 적절한 보상정책을 마련해야 한다.**

조사에 응답하여 응답자의 정보를 제공하게 되면, 이에 대한 적절한 보상이 이루어져야 한다. 일정수준의 현금이나 상품권, 기념품 등으로 응답에 대한 감사의 표시를 전달하는 것이 바람직하다.

**□ 응답률을 공표하도록 한다.**

조사의 신뢰성을 확보하기 위한 방법으로 조사가 완료되면, 응답률을 반드시 공표하도록 함으로서 조사상의 한계 등을 명시하도록 한다.

#### 4. 현장 실사

자료를 수집하는 현장실사단계에서는 조사원과 조사지도원이 조사현장을 방문하여 조사대상 가구나 사업체에 대한 정보를 얻는 과정이다. 이 과정에서 조사지도원은 조사원이 조사업무를 제대로 수행할 수 있도록 지원을 해주어야 한다. 조사의 목적, 방법, 기간, 규모 등이 수록된 공문이나 협조서류를 제공하여 조사에 대한 신뢰감이 갖도록 해야 한다. 한편, 실사현장에서 발생할 수 있는 여러 가지 경우를 대비하여 조사전문가와 긴밀한 연락 체계를 유지하도록 한다. 예를 들어 휴업, 파업 등의 인위적인 사건의 발생과 수해와 같은 천재지변이 발생한 경우에는 조사전문가와 협의를 통해 조사단위의 대체 등과 같은 비상수단을 강구할 수 있도록 해야 한다. 또한 사업체 조사의 경우 업체를 방문하기 전에 먼저 담당자와 면접시간, 조사에 필요한 정보, 면접장소 등에 대해 약속을 하는 것이 바람직하다. 만일 해당 담당자가 없을 경우 유사직급의 대리인으로부터 적절한 응답을 얻어낼 수 있도록 조치되어야 한다. 가구조사의 경우에는 방문날짜와 시간을 미리 공지함으로써 조사대상 가구가 조사에 준비할 수 있도록 해야 한다. 만일 방문한 가구에 대해 응답을 얻지 못한 경

우에는 별도로 재방문 일시를 공지하여 조사가 완료될 수 있도록 해야 한다. 조사표 작성과정에서는 가능한 한 수정을 대비하여 연필과 같은 필기도구를 사용하는 것이 좋은 방법이다. 조사원은 실사현장의 분위기, 응답자의 태도 등에 대한 다양한 정보를 기록하여 추후 보조정보로 활용할 수 있도록 한다. 패널 조사와 같은 종단적 조사의 경우에는 지속적으로 방문하여 조사협조를 독려해야 하기 때문에 가능한 한 친절하고, 예의 있는 자세로 조사에 임하는 것이 중요하며, 응답자와 친밀감을 확보하도록 해야 한다.

#### [가이드라인]

- **조사지도원은 조사현장에서 필요한 각종 자료들을 점검하고, 조사원이 조사현장에서 보조 자료로 활용할 수 있도록 해야 한다.**

조사현장의 위치나 환경을 파악할 수 있는 지도나 관련 행정자료 등을 보조 자료로 활용할 수 있도록 해야 한다. 이렇게 함으로서 조사일정을 효율적으로 관리할 수 있으며, 조사원이 조사이외의 업무를 별도로 하지 않도록 하여 업무부담을 경감시킬 수 있기 때문이다.

- **조사현장에서 가능한 한 합법적인 방법으로 지방행정기관이나 각종 위원회의 도움을 받도록 한다.**

조사현장에서 조사에 대한 이해도를 높이고, 응답자의 조사에 대한 신뢰도를 높이고, 조사 참여를 독려하기 위해 가능한 한 지방행정기관이나 위원회의 도움을 받는 것이 좋다. 지방행정기관의 담당자는 해당 조사지역에 대한 많은 정보를 보유하고 있고, 필요하다면 합법적인 범위 내에서 행정자료를 요청할 수 있기 때문이다.

- **조사현장에서 발생하는 문제점들에 대해 조치를 위한 보고체계를 수립하도록 한다.**

조사현장에서 발생할 수 있는 각종 문제점들을 즉시 해결할 수 있도록 조사원-조

사지도원-조사관리자 간의 연락망을 구축하는 것이 필요하다. 또한 조사팀별 연락망을 수립하는 것도 조사방법상의 문제점, 조사표상의 문제점 등에 대해서는 조사관리자 또는 조사전문가의 조치가 필요하기 때문에 이들 간의 긴밀한 연락체계를 수립해야 한다. 또한 조사관리자는 문제점과 그에 대한 조치내용을 반드시 기록하여 향후 사후분석에 활용할 수 있도록 한다.

□ **조사 완료 후 조사된 조사표는 조사지도원의 사후 점검을 통해 검증하고, 이를 바탕으로 최종 완성된 조사표를 일정한 수집절차를 거쳐 중앙으로 배송하도록 한다.**

조사가 완료된 후 조사지도원은 수거된 조사표에 대한 사후 검토를 거쳐 최종적으로 점검이 끝난 조사표를 중앙으로 배송하도록 한다. 이때 주로 점검할 사항은 조사표상에 기입된 내용 중 인구학적 특성과 주요 분석 변수값을 중심으로 하고, 기록된 내용을 적절히 판독할 수 있는지도 함께 검토한다.

□ **컴퓨터를 이용한 조사에서는 입력된 자료를 출력하여 입력상의 오류를 점검하고, 최종적으로 중앙의 컴퓨터로 전송한다.**

종이조사표 방식과는 달리 컴퓨터를 이용한 조사에서는 조사가 완료되었을 경우 자료가 제대로 저장되었는지를 먼저 확인하고, 저장된 자료를 출력하여 입력자료의 오류발생 여부를 검토하도록 한다. 저장된 자료에 오류가 없을 경우에는 중앙의 컴퓨터로 전송하거나, 파일형태로 발송하도록 한다.

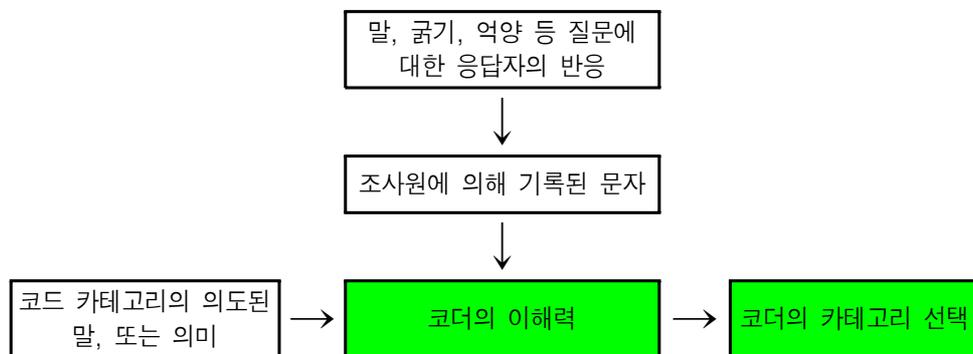
#### 제4절 자료입력 및 처리

조사가 완료된 후 최종적으로 조사지도원의 검토를 거쳐 완성된 조사표는 지역 또는 중앙에서 통계처리를 위한 초기단계로서 자료를 입력하게 된다. 즉 원시자료를 컴퓨터로 해석이 가능한 형식으로 변환하고, 코딩하게 된다. 자료의 입력 방법은 전

자적인 방법을 이용하여 자동 입력하는 방식과 사람의 손으로 수동적으로 입력하는 방식이 있다. 자동입력 방식은 문자인식(MCR), 광학문자인식(OCR), 지능형문자인식(ICR), 음성인식(VRE), 전자교환(EDI) 방식 등이 있다. 각각의 입력방식은 상황에 따라 적절히 사용할 수 있으며, 특히 행정자료나 등록 자료의 입력에 주로 사용하는 방식이 EDI 방식이다. 이와 같은 입력 방식은 주로 종이조사표를 이용할 경우에 사용가능한 방법이고, 만일 컴퓨터를 이용한 조사에서는 조사와 함께 자료가 직접 컴퓨터에 입력되기 때문에 별도의 자료입력절차를 고려하지 않아도 된다. 컴퓨터에 자료를 입력하는 사람을 코더(coder)라 하고, 이러한 작업을 코딩(coding)이라 한다.

코딩작업을 수행하기 위해서 코더는 설문지에 기록된 응답자의 의도를 식별할 수 있어야 한다. 이는 코더의 임무가 응답자의 의도를 하나의 분류항목에 적용하는데 있기 때문이다. 이를 위해 코더는 이상적으로 각 코드 항목을 정의한 연구자의 의도를 완전하게 이해해야 한다.

[그림 4-3-1] 코더의 이해력과 판단 의무



일반적으로 폐쇄형 질문의 경우에는 코더가 바로 코딩작업을 통해 입력이 가능하다. 그러나 개방형 질문항목에 대해서는 응답한 내용을 일정한 분류체계를 이용하여 분류한 다음 코딩작업에 들어가게 된다. 이를 현장코딩(field coding)이라 한다. 현장 코딩은 응답자에게 제시된 개방형 질문에 그들이 응답한 내용을 조사원이 수치형

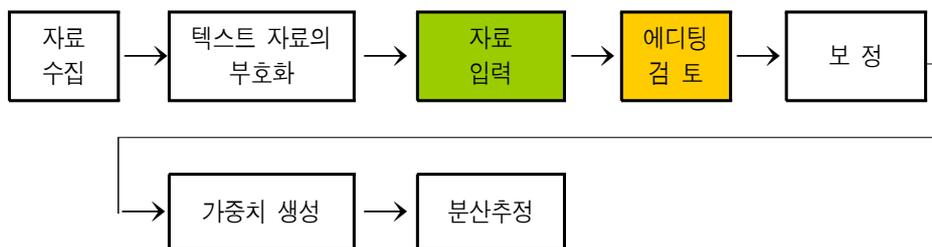
카테고리로 변환하여 개방형 응답을 코드화 하는 과정을 말한다. 그러므로 현장 코딩은 개방형응답과 코딩이 결합된 형태이다. 상대적으로 사무실에서의 코딩은 조사원에 의해 기록된 언어적 응답에 대해 은밀하게 이루어지는 반면에 현장 코딩은 다양한 탐색이 가능하고 풍부한 응답을 청취할 수 있는 장점이 있다.

데이터 수집 후 전형적인 6가지 과정은 다음과 같다.

- ① 부호화(coding) : 문자형 응답을 숫자형 응답으로 변환하는 절차
- ② 자료입력(data entry) : 숫자형 데이터를 데이터 파일에 입력하는 절차
- ③ 에디팅(editing) : 입력된 응답에 대한 에러검토와 불일치 검사
- ④ 대체(imputation) : 항목 결측 데이터에 대한 가공 응답으로의 대체
- ⑤ 가중(weighting) : 비포괄성(noncoverage), 무응답(nonresponse), 불균등 확률 추출에 대한 효과를 줄이기 위한 통계값에 대한 조정 작업
- ⑥ 분산 추정(sampling variance estimation) : 통계치의 불안정성을 평가하기 위한 표본분산의 계산(추출설계 하에서 통계적 오차 계산)

이러한 6가지 단계를 종이조사표 방식과 컴퓨터를 이용한 방식에 대해 작업흐름을 도식화 한 내용이 다음과 같다.

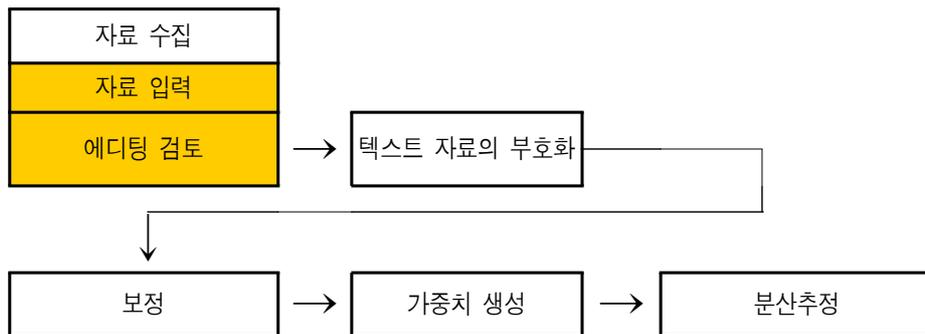
[그림 4-4-1] 종이조사표를 이용한 조사에서 자료처리 절차의 흐름



자료처리 단계에서는 입력된 자료를 검토하는 에디팅작업과 대체작업(imputation)을 거치게 된다. 이 절차는 입력된 자료에 대해 통계처리에 앞서 결측치, 부적절한 자료, 일관성이 없는 자료를 판별하고 오류가능성이 있는 자료를 선별해 내기 위해 반드시 거쳐야 하는 작업이다. 에디팅은 컴퓨터 프로그램에 의해 자동적으로 할 수 있는 부분과 응답자에 대해 추가적으로 조화를 통해 할 수 있는 부분도 있다.

에디팅은 자료를 정제(cleaning)하는 과정이며, 전반적으로 조사의 개선을 위해 필요한 기초정보를 얻을 수 있는 과정이기도 하며, 조사 자료의 품질에 대한 정보를 제공하는 과정이기도 하다.

[그림 4-4-2] 컴퓨터를 이용한 조사에서 자료처리 절차의 흐름



### 1. 자료입력

자료의 입력은 조사가 최종적으로 완료된 후 통계처리를 위해 컴퓨터가 해석가능한 부호로 코딩하여 입력하는 단계를 말한다. 그러므로 코딩작업과 자료입력은 밀접한 관계가 있다. 조사표상에 기재된 응답 내용을 적절한 방법으로 코딩하고, 이를 컴퓨터에 입력함으로써 통계처리를 위한 기본 작업이 완료되는 것이다.

코딩 체계는 모든 응답을 다룰 수 있도록 심지어 응답이 정보를 주지 않는다고 판단되는 경우에도 설계되어야 한다. 그러므로 설문지를 보고 응답을 제시하지 못한

사람들에 대한 코드를 갖도록 하는 것은 매우 중요하다. 질문에 대해 응답할 사항이 없는 사람과 응답하지 않은 사람들을 구분할 수 있는 코드가 필요하게 된다. 즉, 질문에 해당되지 않아 응답을 주지 않은 사람, “비해당” 과 질문에 대해 응답을 주지 않은 사람 “무응답” 은 서로 구분되도록 코드화가 이루어져야 한다. 이러한 예로서 단일 숫자 체계의 경우 “9”는 “무응답”, “0”은 비해당 등으로 코드화 하며, 2자리 숫자 체계에서는 “99”를 무응답으로 “00”을 비해당으로 코드화 할 수 있다.

이와 더불어 특정한 이유로 어떠한 질문에 대해서는 응답을 해주지 않는 경우가 발생하는데, 예를 들어 “응답거절”, “모름” 과 같이 응답을 하는 경우도 있다. 이러한 경우 해당 상황을 구별하기 위해 특정한 코드로 코드화해야 한다.

폐쇄형 질문항목에 대해서는 조사표상에 기입된 내용을 코드화 하여 입력하게 되지만, 개방형 질문항목의 경우에는 응답자료를 몇 개의 항목으로 재분류하여 이를 코드화 하는 작업이 필요하다. 이때, 분류항목수를 적절히 조정할 필요가 있으며, 가능하다면 표준화된 분류체계를 적용하는 것이 바람직하다.

#### [가이드라인]

□ **조사관리자는 자료입력에 적용할 시스템에 대해 검토하여 가장 효율적인 입력시스템을 도입하도록 한다.**

자료 입력방법에는 조사표의 형식에 따라 다양한 방법을 사용할 수 있다. 단기간에 많은 양의 자료를 입력하려면, 자동입력시스템을 도입하는 것이 바람직하여, 만일 이러한 여건을 갖출 수 없는 경우라면, 수동 입력시스템을 적용하고, 입력과정에서 오류를 점검하기위해 이중입력방법(double punching)도 고려하도록 한다.

□ **조사 자료를 입력하기 위한 코딩 작업은 코딩지침서에 따라 수행되어야 하며, 개방형질문에 대한 분류코드로 함께 제공하도록 한다.**

자료 입력을 위한 코딩 지침서는 조사관련 전문가와 조사관리자에 의해 체계적으로 작성되어야 하며, 이를 자료입력 단계에서 이용할 수 있도록 해야 한다. 가능

한 모든 자료는 자료의 성격에 따라 수치화 하거나 부호화 할 수 있도록 하며, 개방형 질문에 대한 응답자료에 대한 분류코드도 마련하도록 한다.

**□ 자료입력단계에서 발생할 수 있는 가능한 오류를 최소화 할 수 있는 입력시스템을 마련하는 것이 바람직하다.**

자료를 입력할 때, 논리적으로 모순되거나 변수의 범위를 초과한 값이 입력될 때, 자동적으로 체크가 가능하도록 하는 입력시스템을 갖추도록 한다. 이러한 절차가 생략되어 자료가 입력될 경우에는 분석과정에서 예상하지 못한 통계적 오류가 발생할 수 있기 때문이다.

**□ 자료의 입력단계에서 발생하는 오류에 대한 원인과 빈도, 해결방안 등을 기록하여 향후에 관련자들에게 환류 시킬 수 있도록 한다.**

입력과정에서 빈번하게 발생하는 오류는 반드시 에디팅 과정에서 점검되도록 하기 위해 이들에 대한 기록을 남겨두는 것이 바람직하다. 또한 단순히 입력과정상에서의 오류인지, 코딩에 의한 오류인지 등 그 원인을 파악함으로써 구조적 오류인지 또는 우연한 오류인지를 밝히고 그에 대한 적절한 해결방안을 마련하도록 한다.

## 2. 에디팅 절차

에디팅 작업은 자료를 정제(cleaning)하는 과정이며, 전반적으로 조사의 개선을 위해 필요한 기초정보를 얻을 수 있는 과정이기도 하며, 조사자료의 품질에 대한 정보를 제공하는 과정이기도 하다. 통상적으로 조사된 자료에 대한 오류수정과 검토가 이루어지는 절차이다. 만일 에디팅 과정에서 자료에 오류가 발생한 경우에는 다음과 같은 규칙에 의해 수정하도록 한다.

- ① 단위의 응답값 전체가 믿을 수 없는 경우에는 데이터를 제거한다.
- ② 다른 보조정보나 다른 항목의 응답값에 의해 오류가 명확하게 고쳐질 수 있을 경우에는 수정한다. 이를 흔히 연역적 대체(deductive imputation)라고 일컫기

도 한다.

- ③ 수정이 적절치 않은 항목의 오류값은 무응답 처리한다.
- ④ 대체(imputation)하는 것이 적절하다고 판단될 때에는 대체 지침에 따라 대체한다.

에디팅 작업을 작업의 성격에 따라 구체적으로 분류해보면 다음과 같다.

□ **형식적 에디팅(format edit)**: 이 과정은 말 그대로 파일에 저장된 조사 자료의 외양에 대해서만 점검하는 과정이다. 각 조사변수에 대해 식별 가능한 코드 체계가 부여되어 있는지, 그 코드 체계에 맞게 제대로 입력되었는지를 점검하는 단계이다.

□ **구조적 에디팅(structure edit)**: 이 과정부터는 입력 데이터의 특성을 감안하여 응답값이 유효한지를 점검하는 단계이다. 특정 조사단위의 응답을 유효한 것으로 인정하고 자료로 활용할 것인지 아니면 자료로서의 가치가 없다고 생각되어 버릴 것인지를 점검하는 것이 포함되며, 다음으로 유효하다고 판단되는 조사단위의 응답 중에서도 각 요소별로 유효한 것으로 인정할 것인지를 점검하는 단계이다.

□ **범위점검(range checks)**: 각 문항별로 유효한 응답값의 범위를 벗어나는 데이터가 있는지를 식별하는 단계이다. 각 문항별로 유효 응답 범위가 무엇인지를 규명해놓은 지침서를 만들어 두는 것이 필요하다. 이 과정에서 비록 유효 응답범위에는 포함될 지라도 이상값(outlier)이 있을 수 있으므로 특정 문항에 대해서는 범위를 좁게 잡아 그 범위를 벗어나는 값에 대해서는 이상 여부를 점검하는 것이 바람직하다.

□ **누락점검(skip checks)**: 조사표를 보면 응답자가 자신이 해당되는 문항에만 답해야 하는 것이 있는 경우가 보통이다. 이럴 때 응답자가 답해야 하는 문항인데 건너뛴 문항이 없는지 아니면 응답해서는 안 되는 항목인데 응답한 것이 없는지를 점검하는 단계이다. 누락 점검을 효과적으로 하기 위해서는 조사표 각 문항들 간의 상호구조와 위계를 잘 나타내는 것이 바람직하다.

□ **일관성 점검(consistency checks):** 서로 관련이 있는 문항들에 대한 응답값의 일관성을 점검하여 응답 데이터의 질을 점검하는 단계이다. 조사표 내에 점검질문(filter question)이 있을 수도 있고, 논리적으로 연결되는 질문이 있을 수도 있기 때문이다.

[가이드라인]

□ **조사관리자는 에디팅 작업에 대한 일정 등을 점검하고, 관련 시스템을 정비하여 체계적인 에디팅 작업이 이루어질 수 있도록 해야 한다.**

조사 자료의 오류점검과 일관성 유지 등을 위해 에디팅은 반드시 필요한 작업이다. 따라서 조사관리자는 이러한 측면을 예산배분과 일정수립과정에서 충분히 고려하여야 하며, 에디팅 방법에 관해 자동시스템을 갖출 것인지, 수동으로 에디팅을 할 것인지를 검토해야 한다. 한번 결정된 방법을 작업중간에 변경하는 것은 자칫 더 많은 오류를 유발할 수 있기 때문이다.

□ **에디팅 과정은 단순한 사항부터 고려하여 점차적으로 복잡한 부분에 대한 작업이 이루어지도록 하는 것이 바람직하다. 또한 이러한 절차에 대한 상세한 지침서를 마련하도록 한다.**

에디팅 작업의 순서는 먼저 단순한 입력오류를 점검하는 절차로부터 출발하여 각 변수값에 대한 일관성 점검까지 순서를 정하여 이루어지도록 하고, 이러한 사항을 상세한 지침서로 관리하도록 한다. 예를 들어 성별변수의 코드값이 “1”인 남성, “2”는 여성으로 부여되어 있다면, 1번과 2번이외의 코드값이 입력되었는지를 점검한다. 다음으로 만일 가구 소득에 대한 변수와 같은 연속적인 값에 대해 일정 범위를 벗어나는 값이 있는지를 점검한다. 가구의 소득이 100원 등으로 상식적으로 받아들이기 힘든 경우의 입력 자료에 대한 점검을 의미한다. 다음으로 논리적 절차에 따라 응답 자료가 입력되어 있는지를 점검한다. 즉, 남성의 경우 여성 질환에 대한 질문

에 답할 수 없도록 되어 있음에도 응답이 되어 있다거나, 15세 이하 아동이 기혼자로 또는 경제활동 인구로 응답되어 있는 경우에 대한 점검 절차이다. 이러한 사항은 일관성 점검과도 동일한 의미를 갖는다.

□ **에디팅 과정에서 무응답한 자료나 일관성이 떨어지는 자료에 대해서는 전화나 재방문을 통해 오류를 점검하도록 한다.**

특정 변수에 대해 무응답한 경우나 응답한 자료의 일관성이 의심되는 경우에는 전화나 재방문을 통해 오류를 점검하도록 한다. 이러한 절차는 중앙에서 보다는 실사현장에서 일차적으로 이루어지도록 하는 것이 바람직하다.

□ **주기적으로 반복되는 조사의 경우 오류에 대한 점검을 에디팅 작업에 너무 의존하는 것은 바람직하지 않다. 왜냐하면 과도한 에디팅은 자칫 조사 자료를 왜곡시킬 수 있기 때문이다.**

조사 자료에 대한 오류점검은 반드시 이루어져야 한다. 그러나 주기성을 갖는 조사에서 전월자료와 비교하여 일부 일관성이 결여될 수 있음을 인정하고, 일정 수준 이상의 에디팅은 삼가야 한다. 실제로 일어난 사실을 이전 조사의 자료와 일관성을 유지하기 위해 수정하는 것은 조사 자료를 조작하는 측면으로 오해할 수 있으며, 전반적인 경향에 맞지 않을 수 있기 때문이다.

□ **에디팅 과정에서 발견된 오류의 수정에 관한 정보를 기록하도록 한다. 이는 조사 자료에 대한 중요한 품질정보가 되기 때문이다.**

조사현장에서의 재 접촉율, 오류변수의 삭제비율, 결측값 또는 오류 변수값에 대한 대체비율 등에 대한 통계를 산출하여 향후 조사 자료의 품질개선폰자로 활용하도록 해야 한다.

□ **에디팅 과정에서 발견되는 오류의 원인이 무엇인지를 명시하는 것이 바람직하다.**

입력오류인지, 또는 조사과정에서 발생한 측정오류인지를 밝혀 자료의 정확성지

표로 활용하도록 한다.

□ 반복조사에서 일관성 있게 잘못 조사되는 내용에 대해서는 에디팅 과정에서 식별하기가 매우 어렵기 때문에 신축성을 가지고 에디팅을 하는 것이 효과적이다.

에디팅 과정을 엄격히 수행하는 것만이 능사가 아니다. 에디팅 과정에서 미처 고려하지 못했던 사항이 발견되면 이를 신속하게 처리할 수 있도록 해야 한다. 반복조사의 경우 일관성 있게 오류가 발생할 경우를 대비하여 보조 자료를 이용한 에디팅 작업을 수행할 수 있도록 신축성을 고려하도록 한다.

### 3. 대체절차

대체작업은 조사 자료의 기본적인 분포를 유지하면서 에디팅 과정에서 식별된 무응답에 대해 응답자와 유사한 값으로 대체하는 과정을 말한다. 대체과정은 통계적으로 엄밀성을 요구하는 단계이므로 전문가의 도움을 받는 것이 바람직하다. 통상적으로 무응답자와 응답자의 성향은 다르게 나타난다. 따라서 조사과정에서 결측된 자료에 대해 일괄적으로 응답자의 값을 대체되어서는 안 된다. 또한 무응답에 대한 대체는 실사현장에서 처리하기 보다는 중앙에서 처리하는 것이 바람직하다.

무응답을 대체한 후 이를 응답값으로 간주하여 처리해서는 안 된다. 왜냐하면, 대체값은 분명히 실제 응답값과 다르며, 대체과정에서 변동성이 증가되기 때문에 추정치의 분산이 증가하기 때문이다. 대체과정은 먼저 무응답변수와 상관이 높은 보조정보들을 이용하여 조사 단위들을 몇 개의 대체 층으로 분류한 다음 각 대체 층 내에서 응답값들을 대체자료제공세트(donor set)로 활용하는 것이 일반적인 방법이다.

대체방법으로는 평균대체(mean imputation), 랜덤핫덱대체(random hot deck imputation), 최근방대체(nearest neighborhood imputation), 비대체(ratio imputation), 회귀대체(regression imputation)등이 있다.

[가이드라인]

- 무응답에 대한 대체작업은 중앙에서 일괄적으로 관리하도록 하며, 조사현장에서 대체가 이루어지지 않도록 한다.

조사 자료에 발생한 무응답은 이론적으로 응답 자료와는 다르다. 따라서 조사원이 임의적으로 무응답을 조사현장에서 대체하도록 해서는 아니 된다. 무응답 대체 모형에 따라 중앙에서 일괄적으로 결측치에 대한 무응답 대체가 이루어 질 수 있도록 철저하게 관리되어야 한다.

- 무응답 대체에 적용 가능한 다양한 이론들을 검토하여 자료의 성질에 적합한 무응답대체방법을 고려하도록 하며, 이를 보고서에 반드시 명시하도록 한다.

무응답 대체방법 중 가장 월등한 방법은 존재하지 않는다. 다만, 자료의 성질에 따라 다양한 방법 중 효율적인 방법을 선택적으로 적용할 수 있으며, 굳이 복잡한 방법을 사용하지 않아도 되는 자료가 있기 때문이다. 통상적으로 결정론적인 방법보다는 확률론적인 대체방법이 선호되며, 단순대체보다는 다중대체가 선호된다.

- 적용하고자 하는 대체방법에 따라 다를 수 있지만, 가능하면 많은 보조정보를 사용하여 무응답 대체를 수행하는 것이 효과적이다.

조사과정에서 수집한 다양한 보조정보와 기존의 유사한 조사로부터 얻은 보조정보 등 활용 가능한 정보들을 대체과정에 투입하는 것이 대체에 의한 오차를 줄일 수 있다. 무응답자에 대한 상세한 정보가 있으며, 보다 정확한 값으로 대체가 가능하기 때문이다.

- 대체된 값은 실제 조사된 값이 아니기 때문에 자료에 일정한 표시를 해주는 것이 좋다.

무응답 대체값은 반드시 조사 자료와 다르게, 예를 들어 조사 자료값은 “1” 대체값은 “2” 와 같은 방식으로 표시를 하도록 한다. 왜냐하면 대체 값을 원시자료로 취급할 경우 추정과정에서 문제가 발생할 수 있기 때문이다. 즉, 대체에 의한 분산의

증가분을 계산 할 수 없기 때문에 분산 추정값을 과소추정하게 된다. 이는 추정과정에서 대체에 의한 변동을 추정하여야 하기 때문이다.

**□ 대체자료의 양은 응답 자료의 양과 비교하여 최소한으로 실시하는 것이 바람직하며, 대체율을 명시하도록 한다.**

실제 조사 자료와 비교하여 대체값은 일정수준 이하로 구성되는 것이 바람직하다. 이는 가공된 자료가 대다수를 차지할 경우 조사 자료의 의미를 상실할 수 있으며, 이론적으로도 허용되지 않기 때문이다. 따라서 무응답대체는 가능한 한 최소한으로 이루어지도록 해야 한다. 또한 이에 대한 정보를 반드시 명시하도록 한다.

**□ 대체를 위한 프로그램이나 S/W는 가능한 한 범용 프로그램으로 작성되도록 하여 수정이 용이하도록 한다.**

대체과정은 자료의 형태에 따라 각각 다르게 적용됨으로 반드시 일률적으로 이루어지지 않는다. 따라서 대체작업이 필요한 경우 이미 작성된 프로그램을 쉽게 수정하여 사용할 수 있도록 범용화 된 프로그램을 사용하는 것이 바람직하다.

## 제5절 자료 분석 및 통계결과의 품질

통계결과에 대한 품질평가는 조사목적이나 측정하고자 하는 현상에 비추어 산출된 통계의 적합성을 평가하는 것이다. 이러한 품질 평가 결과가 제시됨으로서 이용자의 관점에서 통계를 해석하거나 이용수준을 결정하는 데 도움이 되며, 통계작성기관의 관점에서는 향후 통계품질 개선을 위한 중요한 정보로 활용될 수 있다.

자료의 품질평가는 통계생산 활동의 본래 목적에 비추어 자료의 정확도, 신뢰성의 관점에서 최종 산출물을 평가하는 과정이다. 이러한 평가정보를 이용하여 이용자는 조사결과를 보다 유용하게 해석할 수 있게 되며, 통계작성기관은 통계조사 절차를 개선할 수 있다.

지표에 대한 품질평가는 보증 또는 타당성을 의미하는 측면과 오차의 관리에 관한 측면이다. 즉, 전자는 조사결과를 공표하기 전에 자료의 품질한계 등을 검토하는 과정이다. 한편 후자는 자료의 오차요인에 대한 수량화된 정보를 제공한다. 이러한 검토 결과는 대부분 자료가 공표된 후에 가능한 경우가 일반적이다.

통계의 품질평가는 통계작성과정의 일부분으로 간주되어야 한다. 그러므로 조사기획단계에서부터 통계품질평가 절차가 주요업무로 반영되어야 하고, 그에 따른 인력과 예산이 적절히 확보되어야 한다.

자료의 품질 평가의 여건과 목표는 설계과정에서 이미 다룬바가 있지만, 자료수집 및 처리가 완료된 후에도 품질평가 차원에서 자료에 대한 진단이 필요하다. 이러한 사항은 자료 분석 과정에서 고려되어야 한다.

#### [가이드라인]

□ 통계작성기관은 이용자들에게 기본적으로 정확성 관련 요소들에 대한 품질 정보를 제고해야 한다.

통계이용자들은 통계작성기관에서 제공하는 정확성 관련 요소들을 기준으로 작성통계가 얼마나 모집단을 잘 설명하고 있는가를 판단할 수 있다. 상대적으로 오차가 작은 통계결과를 기대하며, 오차의 수준이 기대치를 벗어난 경우 그에 대한 이유를 검토할 수 있도록 해야 한다.

□ 주기적으로 조사가 수행되는 통계에 대해서는 계속적으로 품질평가를 수행한다는 것이 현실적이 않다. 따라서 적절한 평가 주기를 설정하여 수행하도록 한다.

통계결과의 품질평가는 작성기관의 업무부담을 감안하여 가능한 주기적으로 수행되는 것이 바람직하다. 그렇게 함으로서 품질개선을 위한 노력을 기울일 수 있기 때문이다. 그러나 주기적으로 생산되는 통계의 경우 매년 품질평가를 수행한다는 것은 현실적으로 온당치 않기 때문에 적절한 기간을 상정하여 품질평가를 받도록 하는 것이 바람직할 것이다.

□ 통계작성기관 내외부의 전문가와 이용자들을 품질평가과정에 참여토록 한다.

통계품질 평가에 참여하는 인력은 전문성을 갖춘 외부 평가위원과 이용자들로 구성되는 것이 바람직하다. 다면적인 평가를 수행하기 위해 이용자 측면과 전문가적인 관점을 모두 고려해야 하기 때문이다.

□ 통계작성기관은 지속적으로 통계품질관리 활동을 위한 품질관리 조직을 구성하여 운영토록 해야 한다.

통계결과에 대한 품질 평가는 지속적이고, 체계적으로 수행되어야 할 부분임으로 통계작성기관에서는 품질관리를 위한 별도의 조직을 운영할 필요가 있다. 이때, 적절한 인력과 예산배분을 통해 가능한 한 자율적으로 수행되도록 해야 한다.

□ 통계작성기관은 통계품질관리의 체계적이고, 일관성 있는 품질관리 활동을 위한 품질관리 매뉴얼을 작성하여 이를 기준으로 수행되도록 해야 한다.

체계적이고, 일관성 있는 품질관리활동을 위해 품질관리 기준을 마련하고, 전반적인 수행절차 및 각 품질요소별 특징과 주요 품질지표의 도출 등을 위한 품질관리 매뉴얼을 작성하여 활용할 수 있도록 해야 한다.

1. 관련성

관련성(Relevance)은 통계가 현재와 잠재적 사용자 요구에 부합하는 정도를 말한다. 이는 요구되는 모든 통계가 생산되고 있는지를 반영하며, 사용된 개념(정의, 분류 등)의 범위가 사용자 요구를 반영하는 지를 나타낸다.

[가이드라인]

□ 통계작성자는 우선 해당통계의 사용자에게 관한 정보(이용자가 누구인지, 얼마나

많은 이용자가 있는지, 얼마나 중요한지 등)를 수집하고, 다음으로 이들의 요구에 관한 정보를 수집하고, 마지막으로 이들의 요구에 부합한지에 대한 정보를 수집해야한다.

- 이용자의 요구를 적절한 통계적 용어와 산출물로 이들의 요구에 부합하도록 재해석하여 제시해야한다.

용어와 산출물은 현재의 통계적 개념으로 나타내는 것이 바람직하며 국제적인 비교를 위해 가능하다면 국제적으로 통용 가능한 정의와 용어를 사용하는 좋다. 이러한 작업은 종합적인 수요를 만족할 수 있는 점검결과로 나타나는 통계산출물(지표, 지표의 정확한 정의 등)을 정의하도록 한다.

- 종합적인 측면에서 이용자들을 그룹별로 제시하고, 이들의 우선적인 요구사항과 기대치 간의 상충되는 점을 반영하도록 한다. 만일 이용자 그룹이 전략적으로 중요할 경우에는 이들의 요구에 대한 보다 충분한 설명을 기술하도록 한다.

- 이용자 관련성을 측정하기위해 정확성, 시의성등과 같은 여타의 품질요소에 따라 이용자 만족도조사를 통해 이용자들에게 질문하도록 한다.

관련성은 여타의 품질 구성 요소와 관련된 필요성과 서로 다른 이용자 분류의 관점과 매우 밀접한 관련되어 있다. 그러나 고객의 관점을 항상 고려할 필요는 없다. 서로 다른 이용자나 고객의 상충된 속성이 한 가지 이유이며, 다른 하나는 시간과 공간에 대해 비교될 수 있는 통계의 일치성과 일관성이 필요하기 때문이다.

## 2. 정확성

각 조사의 목적은 통계를 생산하는 것으로서 목표모집단의 수량화가 가능한 특성인 미지의 값을 추정하는 것이다. 일반적 통계적인 관점에서 정확성(accuracy)은 참값에 대한 정확한 계산값 또는 추정치의 엄밀성으로 해석된다.

[가이드라인]

□ **주요 분석 변수의 편향의 크기(적어도 부호)를 명시하도록 한다.**

정확성을 평가하기 위해 먼저 주요 분석변수에 대한 편향을 추정하여 실제 목표치와 얼마나 차이가 나는지를 기술해야 한다.

□ **통계값에 대한 변동계수, 신뢰구간, MSE의 추정값, 대안적으로 변동성의 질적 평가가 가능하도록 해야 한다.**

마찬가지로 통계값의 변동을 나타내는 다양한 통계량들을 제공하도록 한다. 특히 변동계수, 신뢰구간, MSE추정치 등을 주요변수들에 대해 제시하고, 만일 이러한 통계량들의 제시가 곤란한 경우에는 변동을 평가할 수 있는 질적인 내용이 기술되도록 한다.

□ **분산 추정과정에서 고려된 오차의 유형에 관한 사항을 기술해야 한다.**

주요변수에 대한 추정치와 분산을 추정하는 과정에서 단순히 분산추정치만을 제시하는 것 보다는 조사과정에서의 오차, 처리과정에서의 오차, 프레임의 불완전성으로 발생하는 오차 등 다양한 오차요인들을 기술하고, 실제로 추정과정에 반영하였는지를 설명해야 한다. 이는 이용자들이 추정치에 대한 자세한 정보를 이용할 수 있기 때문이다.

□ **미리 정해진 규정에 의해 제시된 통계값에 대한 정확성 수준을 평가하고, 오차의 수준과 제안된 개선안과 상충하는 점에 대하여 설명하도록 한다.**

목표오차 또는 목표 상대오차와 같이 설계과정에서 제시된 값과 추정결과로 얻어진 값간의 정확성을 평가하고, 만일 이들 간에 차이가 발생했다면, 어떤 측면에서 차이가 있는지, 또는 차이를 평가했을 때, 거의 무시할 정도인지 등에 대해 설명하도록 한다. 대부분 설계과정에서 설정되는 오차는 완전응답을 가정하여 수립한 것임

으로 대계의 경우 설계과정에서의 오차수준과 추정과정에서 산출된 오차 간에는 차이가 있을 수 있다. 가능하다면, 이러한 차이를 개선할 수 있는 설명을 부가하는 것이 바람직하다.

**□ 주요 통계값의 편향과 변동성을 이전조사와의 비교(성장률, 평균값)를 통해 일관성을 평가하도록 한다.**

정확성은 여타의 다른 품질지표와 관련되어 있다. 따라서 주기적인 조사에서 주요변수에 대해 이전 조사와의 비교를 통해 어떻게 변동하고 있는지를 평가하도록 한다. 이러한 점은 특히 시계열을 유지하는 자료를 평가하는 적절한 방법이 될 수 있다.

**□ 표본오차, 포함오차, 측정, 처리, 무응답, 모형설정오류와 이들의 효과에 관한 정보를 기술하여 이용자가 충분히 활용가능 하도록 해야 한다.**

조사과정, 자료처리과정 등에서 발생한 다양한 오차에 관한 정보를 기술하여 이용자가 충분히 활용할 수 있도록 해야 하며, 통계생산자 측면에서는 향후 통계의 개선을 위한 중요한 정보가 될 수 있기 때문이다.

### 3. 시의성 및 정시성

어떠한 통계든지 적절한 시간 내 결과를 작성하고, 일정을 준수하는 것이 필요하며, 이는 이용자가 결과를 효과적으로 사용하고자 하기 때문이며, 사용자 요구 중에서 가장 중요한 것 중 한가지 이다.

통계 또는 정보의 시의성(Timeliness)은 정보의 이용성과 그 정보가 나타내는 사건 또는 현상간의 시간간격을 나타낸다. 시의성은 상대적으로 쉽게 그리고 직접적으로 계산된다. 통상적인 측정 방법은 평균산출 시간이다. 최대산출시간은 가장 나쁜 경우로서 나타난다.

정시성(Punctuality)은 자료의 공표날짜와 자료를 공표해야할 목표날짜 간, 예를

들어 자료를 공표한다고 발표한 공식적인 날짜 또는 규정에 적시되었거나 사전에 이용자와 협의한 날짜 간의 시차를 나타낸다.

#### [가이드라인]

□ **계획된 통계결과의 공표시점을 준수하고 만일, 차이가 발생했다면 그 원인과 향후 개선 방향을 제시하도록 한다.**

이용자의 측면에서 예상되는 통계결과의 공표나 발간물의 정시성은 매우 중요할 수 있다. 그러나 조사과정의 문제, 자료처리과정에서 예상하지 못한 다양한 원인으로 정시에 자료를 공표하지 못한 경우 이에 대한 원인을 파악하고, 향후 생산 절차상의 개선점을 제시하는 것이 바람직하다.

□ **조사대상 기간과 통계결과의 공표시점간의 시차를 평가하도록 한다.**

조사된 자료의 시점이 너무 오랜 시점이거나, 현재 시점에서 사용하기 곤란한 자료의 경우 시의성에 문제가 발생하기 마련이다. 따라서 가능한 최근시점의 자료를 생산하도록 해야 하며, 이에 대한 평가지표로서 조사대상 기간과 통계결과의 공표시점간의 시차를 제시할 필요가 있다. 이러한 정보는 이용자가 해당 자료의 시의성을 평가하여 자료의 사용여부를 결정할 수 있기 때문이다.

□ **향후 평가지표를 산출하기 위해 평균산출 시간, 최대산출시간 등을 제시하는 것도 고려할 수 있다.**

#### 4. 비교가능성

비교가능성(Comparability)의 목표는 지리적 영역, 비지리적 영역, 또는 시점들 간에 통계를 비교하여 적용된 통계적 개념과 정의의 상이성 효과를 측정하는 것이다. 비교가능성을 분류하면 크게 3가지로 구분된다.

먼저 지리적 비교가능성은 동일한 현상을 측정하는 유사한 조사들은 서로 다른 통계 기관에 의해 수행되고, 서로 다른 지리적 개체로 모집단을 나타내는 유사한 조사들 간의 비교성의 정도를 나타낸다. 다음으로 시간에 대한 비교가능성은 두 조사 시점간의 비교가능성의 정도를 나타낸다. 마지막으로 영역 간 비교가능성은 서로 다른 통계영역에서 유사한 목표 특성을 가진 서로 다른 조사간의 비교가능성의 정도를 나타낸다.

#### [가이드라인]

□ **조사 설계상의 개념과 유사한 다른 통계간의 상이한 개념이 결과물의 비교가능성에 영향을 줄 수 있는 지를 기술한다.**

이는 유사한 통계간의 서로 다른 개념상의 상이성으로 이용자가 혼란을 일으킬 수 있기 때문에 가능한 한 유사통계의 개념과 서로 비교하여 이러한 문제를 방지할 수 있도록 해야 한다. 동일한 가구조사라 하더라도, 경제활동인구의 정의가 다르게 되면 두 통계는 비교가능성이 결여되기 때문이다.

□ **국가별로 사용되는 개념과 국제수준에서 사용되는 개념간의 상이함으로 인한 비교가능성의 결여를 최소화하기 위해 가능하면 국제 비교를 위한 통계는 국제표준을 적용하도록 한다.**

이러한 문제는 종종 선진국과 개발도상국간의 경제규모를 비교하는 과정에서 발생한다. 개발도상국에서 사용하고 있는 통계의 개념이 주요 선진국의 표준과 다르기 때문인데, 이를 방지하기 위해 가능하다면, 국제적으로 통용될 수 있는 개념을 적용하는 것이 바람직하다.

□ **시계열 자료인 경우 조사대상 기간에 단절이 발생하는 경우 시계열 단절 전후에 측정의 개념과 방법상의 차이를 기술한다.**

시계열자료의 경우 시간적인 비교가능성을 평가하기 위해 단절이 발생한 경우

어떤 차이가 있는지를 반드시 명시하도록 한다. 특히 월별, 분기별, 연간 자료의 경우에 시계열 상의 단절은 이용자 측면에서는 매우 중요한 일이 될 수 있다. 따라서 이러한 단절의 원인이 무엇인지, 측정의 개념과 방법간의 차이가 발생했는지를 설명할 필요가 있으며, 그에 대한 영향과 한계점도 밝혀두도록 한다.

□ **자료수집 방법, 조사에 사용된 목표모집단과 조사모집단, 표본추출방법, 추출단위, 분류, 통계적 방법, 통계적 모집단, 자료처리 방법 등의 차이를 기술하도록 한다.**

비교가능성을 확보하기 위해 유사통계와의 상이점을 각각의 차원에서 비교함으로써 이용자가 통계자료를 이용하는데 혼란스럽지 않도록 해야 한다.

## 5. 일관성

통계의 일관성(coherence)은 다양한 이용에 대해 서로 다른 방법으로 신뢰성 있게 결합될 수 있는 타당성을 나타낸다. 그러나 일반적으로 일관성을 증명하는 것보다는 비일관성을 증명하는 방법이 더 쉽다.

통계가 서로 다른 자료 원으로부터 산출될 경우, 상이한 접근, 분류, 방법론적 표준에 근거한 통계와는 완전히 일관성을 가지지 못하게 된다.

다양한 영역으로 일관성을 평가할 수 있다. 여기서는 자료의 1차적인 이용과 2차적인 이용간의 일관성을 고려한다. 실제로 통계기관에서 생산된 자료의 주요한 2차적 이용은 통합적인 통계의 입력 자료로 이용된다.

### [가이드라인]

□ **주요 특성값에 대한 잠정통계와 확정통계를 비교하고, 차이가 얼마나 나는지에 대한 설명과 의견을 제시하도록 한다.**

동일한 통계의 잠정치와 확정치의 차이를 비교하고, 만일 이들 간의 차이가 발생

했다면 그 원인을 밝히도록 한다. 잠정치가 발표되면, 반드시 그 통계에 대한 확정치가 향후에 공표됨을 알려야 하며, 확정치를 발표하면서 어떠한 점이 개선되었는지를 밝혀야 한다. 잠정치를 산출하는 과정에서 고려하지 못했거나, 또는 일부자료의 오류로 인한 것이면, 이러한 오류의 발생원인과 개선점을 명시하도록 한다.

**□ 주기가 다른 연간통계와 단속통계간의 일관성을 비교했을 경우, 이에 대한 차이를 설명하도록 한다.**

조사주기가 다른 두 통계를 결합하는 것은 두 통계가 일관성을 유지한다면 가능한 일이다. 그러나 대개의 경우 조사의 목적과 조사방법, 분류기준 등 상이한 측면이 많이 있기 때문에 일관성이 결여된 두 통계를 결합하고자 할 때에는 반드시 이러한 점을 유념하도록 한다.

**□ 동일영역 통계를 결합하여 사용하고자 할 경우 정확성 요소에 따른 일반적인 특성값에 대한 차이 등에 대해 기술하도록 한다.**

정확성 요소로서는 표본오차, 측정오차, 포함오차, 처리오차, 무응답오차, 모형가정오류 등이며, 동일영역의 통계들을 서로 결합하여 사용할 경우 반드시 두 통계값간의 차이점을 기술하도록 한다. 즉, A통계의 포함오차와 B 통계의 포함오차를 서로 비교함으로써 연계성공률 등을 산출할 수 있다.

## 6. 자료 분석

자료의 분석은 크게 3가지로 구분된다. 첫째는 조사 자료를 근거로 기존통계표를 작성하는 단계이다. 두 번째는 통계이론이 입각하여 추정값을 산출하는 것이다. 세 번째는 앞의 두 단계를 거쳐 이차적인 통계분석을 수행하는 것이다.

통상적으로 자료 분석은 통계산출물에 가치를 부여하기 위해 원시자료를 유용한 정보 형태로 가공하는 절차를 일컫는다. 자료 분석은 통계자료의 품질의 측면에서 보면, 조사의 가장 중요한 과정중의 하나이다. 이와 더불어 자료의 분석은 자료의

품질관련 문제를 규명해주며, 향후 조사 자료의 수집과정을 개선할 수 있게 한다.

여기서는 자료 분석의 모든 단계를 다루기보다는 전반적인 내용을 중심으로 품질 평가와 관련된 필요 사항들만을 다루기로 한다.

#### [가이드라인]

□ **이용자 요구를 충족시키기 위해 가능한 다차원 분석 자료를 제공하도록 한다.**

이용자들의 요구에 대한 맞춤형 분석을 제공하기 위해 가능한 다차원분석결과를 제공하도록 한다. 기초 통계자료의 제공과 함께 이용자가 요구하는 내용으로 분석이 가능할 수 있도록 자료를 제공하는 것이 바람직하다.

□ **2차적인 분석을 위한 고급통계분석을 고려하고, 향후 분석기법의 개선을 위한 노력을 기울여야 한다.**

통계자료에 대한 2차 분석은 대개의 경우 매우 전문적인 분야이며, 이를 위해서는 통계분석 전문가를 두는 것이 바람직하다. 그렇게 함으로서 이용자들의 요구를 충족시킬 수 있기 때문이다. 또한 현재 사용하고 있는 분석기법을 개선할 필요성에 대해 검토해야 한다.

□ **조사단계에서 복합표본설계를 적용했다면, 반드시 분석과정에서 해당 설계를 고려하도록 한다.**

조사 설계과정에서 복합 표본설계를 고려한 경우에는 반드시 이를 분석과정에 반영하도록 한다. 왜냐하면, 분석결과가 왜곡 될 수 있기 때문이다.

□ **통계표 작성과정에서 가로, 세로의 합이 일치하는 지를 반드시 점검하도록 한다.**

특히 조사변수에 결측이 있을 경우 두 합이 일치하지 않는 경우가 발생하기 때문에 이를 반드시 점검하도록 한다.

## 7. 보안관리

보안관리 또는 비밀보호의 개념은 자료가 공표되었을 때, 일반적으로 모르는 단위에 관한 표본단위 정보가 노출됨으로서 모집단의 특정한 단위의 정보가 누출될 수 있는 위험을 줄이기 위한 노력으로 표현할 수 있다. 따라서 통계표나 원시자료를 공표하기 위한 목적에 대해 비밀 누출의 위험도를 평가하여야 한다. 현실적으로 비밀 보호와 정보공개는 정책은 자주 상충되는 경우가 발생한다. 한편으로는 이용자들은 개인, 사업체, 기관 등의 상세한 정보를 요구하고, 다른 한편으로는 조사에서 응답한 응답자들은 자신들의 개인정보가 누출되지 않기를 희망한다. 이는 이용자에게 거시자료를 공개할 때 모집단에 있는 개인에 대한 상세한 정보가 식별되지 않도록 해야 하며, 원시자료의 경우에는 응답자의 이름, 주소 등 개인을 식별할 수 있는 정보의 누출로부터 자유로워야 함을 의미한다.

정보보호 또는 자료의 보안에 대한 내용은 거시자료와 원시자료의 특성이 다르기 때문에 각각에 대해 개별적으로 보안관리를 위한 가이드라인을 제시하도록 한다.

### 1) 거시자료

통계표에 대한 비밀보호를 위한 방법은 셀 은폐(cell suppression), 계급조정(rolling-up), 자료교란(disturbing)과 같이 3가지로 분류할 수 있다. 셀 은폐방법은 특정 기준 이하의 셀값이 존재하는 경우 그 셀에 대해 “\*”와 같은 특수문자로 간단하게 셀 값을 표시해주는 방법으로 비밀 누출이 매우 쉬운 방법이다. 다음으로 계급조정은 통계표의 행 또는 열을 큰 계급구간이나 특성값의 새로운 계급과 결합시키는 방법이다. 이 방법은 개별 항목에 대한 은폐방법보다는 간단하지만 통계표의 기술이나 분석의 의미가 감소되는 단점이 있다. 마지막으로 자료교란은 체계적으로 통계표의 수치를 바꾸는 방법으로서 결과적으로 통계수치들은 개별단위에 대한 정보를 누출할 가능성이 적으면서 통계표의 정보는 왜곡되지 않는 방법이다. 통상적으로 반올림 방법(rounding)이 많이 사용된다. 최근에는 통계수치에 확률오차인 잡음

(noise)를 부과하여 발표하는 방법이 개발되어 원시자료생산 단계에서 정보누설을 차단함으로써 거시자료상의 통계수치들이 근원적으로 보호되도록 하고 있다.

## 2) 원시자료

원시자료에 대한 정보보호는 외부자료와 결합하여 이용하는 경우가 많기 때문에 거시자료와 같이 간단한 문제가 아니다. 더욱이, 원시자료에 대한 정보누설을 평가할 만한 척도도 이용될 만한 것이 충분하지 않기 때문에 정보보호에 적합한 표준화된 방법은 없는 실정이다.

정보유출을 방지하기 위해 통상적으로 모든 공공이용을 위한 원시자료는 단지 표본자료만을 포함하도록 하고, 명확하게 개별단위가 식별되는 정보는 포함하지 않도록 하며, 지역적인 정보를 제한하고, 자료상의 변수의 수를 제한하도록 한다.

### [가이드라인]

**□ 통계자료를 제공하기 전에 응답자의 비밀이 누출되는 지를 충분히 검토하여야 한다.**

통계자료에 대한 이용자의 요구는 비밀보호정책과 상충된다. 그러나 통계생산자는 응답자들의 정보가 누출되지 않도록 반드시 공표 전에 검토해야 한다. 특히 원시자료의 경우 개인 또는 사업체를 구별할 수 있는 정보가 누출되지 않도록 최선을 다해야 한다.

**□ 통계생산자는 거시자료와 원시자료에 대한 비밀보호 정책을 이용자에게 반드시 공지하여 정보공개과정에서의 혼란을 최소화 해에 한다.**

국가통계기관, 통계생산 기관 등에서는 정보공개에 대한 정책을 수립하여 이를 반드시 이용자에게 공개하도록 해야 한다. 통계이용자는 가능한 한 상세한 통계자료를 요구하는 경향이 있기 때문에 합리적으로 이를 제한할 수 있는 정보공개 정책을 마련하도록 해야 한다.

□ 통계결과표상의 정보보호를 위한 보안작업이 적용된 경우에는 반드시 이에 대한 설명을 첨부하여야 한다.

통계이용자들이 해당통계표를 이용할 때, 어떤 기준에 의해 정보보호가 되고 있는지에 대해 이해할 수 있도록 설명할 필요가 있다. 또는 정보보호를 위한 지침서등에 이와 관련된 내용을 기술하는 것도 좋은 방법일 것이다.

## 제6절 문서화 및 자료의 공표

통계자료에 대한 문서화와 자료의 제공은 이용자 측면에서 양질의 자료를 이용할 수 있도록 하며, 이용자 접근성을 고려한 통계생산 전략중의 하나이다. 이용자 관점에서 통계작성기관에서 생산된 자료를 이용하기 위해서는 먼저 접근성이 좋아야 한다. 자료이용이 까다롭거나, 생산된 통계자료로의 접근이 어렵다면, 실제로 통계생산의 주요 목적이 그만큼 상쇄될 수밖에 없다. 따라서 통계작성기관은 소관통계의 수요를 적절히 예측하여 이용자의 요구에 부합하도록 자료제공 정책을 수립해야 한다. 이 과정에서 어떠한 방식으로 자료를 공표할 것인지, 또는 자료의 제공 매체는 무엇으로 할 것인지, 자료이용에 대한 가격정책을 유료 또는 무료화 할 것인지에 대한 다양한 검토가 필요하다.

### 1. 메타자료의 문서화

문서화 작업은 통계작성기관에서 통계 생산을 위한 모든 통계적 활동을 기록하는 과정을 의미한다. 통계자료 이용자에게 조사의 전체 과정과 그 결과를 이해하기 쉽게, 가장 효과적으로 전달하는 것을 목적으로 한다. 따라서 조사 자료를 통계결과로 작성한 문서와 조사과정에 필요한 다양한 정보, 관련 기법 등을 다룬 문서로 구별될 수 있다.

본 절에서는 이러한 문서화 작업 중에서 통계결과의 문서화만을 다루기로 한다.

통계결과의 문서는 통계발간물 또는 보고서 형식으로 작성할 수 있다. 통계발간물의 경우에는 조사부터 결과의 발표까지의 내용을 전체적으로 포괄하여 다루게 된다. 즉, 조사목적, 조사기간, 조사기관, 표본설계, 무응답 처리, 추정, 결과공표, 품질관리에 이르기까지의 내용을 작성하는 것을 의미한다. 한편 보고서 형식은 특정 사안에 대한 상세한 내용을 문서화한 것으로, 표본설계보고서, 품질관리보고서, 무응답처리보고서 등 각 과정별로 중요 항목에 대한 상세히 기술한 것이다. 따라서 문서화란 이러한 전반적인 내용을 이용자가 쉽게 관련 통계의 정보를 파악할 수 있도록 문서로 제공하는 것을 의미한다.

#### [가이드라인]

□ 통계자료를 문서화 할 경우 기본적으로 다음 내용이 포함되도록 한다.

①조사개요, ②통계결과분석 ③통계표 ④통계결과의 품질평가 ⑤부록 등이 수록되도록 한다. 이러한 내용은 이용자가 쉽게 이해할 수 있도록 작성되는 것이 바람직하며, 전문적인 용어의 정의, 조사에 이용된 조사표 등은 부록에서 다루도록 한다.

□ 만일 행정자료를 이용하여 통계자료를 생산한 경우에는 관련 내용을 문서화하도록 한다.

행정자료는 조사 자료와 다르게 표본추출과정이나 추정의 과정이 없기 때문에 이용자가 행정자료 통계를 이용할 경우 매우 당황스럽게 된다. 따라서 자료의 출처, 보고형태, 보고절차, 통계작성 방법 등에 대한 정보를 반드시 기술해야 한다.

□ 제공되는 문서의 형태는 주어진 예산 하에서 가능한 한 다양한 방법으로 제공하도록 한다.

정형화된 문서제공 형태를 탈피하여 가능한 한 다양한 매체를 이용하여 문서를 제공하도록 한다. 이를 위해 먼저 이용자수요를 파악하는 것도 바람직할 것이다. 인터넷, CD-ROM등의 전자 매체를 이용하거나 전통적인 책자형태를 이용할 지를 검

토한다.

## 2. 자료의 제공전략

궁극적으로 모든 국가통계기관과 그 밖의 조사기관은 거시자료와 원시자료를 외부 사용자에게 제공하도록 하고 있다. 이때, 거시자료(macro data)는 통계표, 수치, 빈도 등을 포함한 자료를 말하며, 원시자료(micro data)는 응답자 개인, 가구, 사업체, 또는 그 밖의 단위들에 대한 정보를 제공하는 레코드로 구성된 자료를 일컫는다. 따라서 자료의 제공이란 이러한 형태의 자료를 이용자들에게 제공하는 행위를 말한다. 자료제공자 측면과 이용자 측면은 서로 상충되는 욕구를 가지고 있다. 즉, 이용자는 가능한 한 상세한 자료를 요구하고, 자료제공자는 통계자료의 비밀 보호 등의 이유로 통계자료 이용에 제한을 두려고 하기 때문이다. 따라서 자료제공자는 이러한 측면을 적절히 고려하여 이용자의 요구를 최대한 반영할 수 있도록 해야 한다. 통계작성기관에서는 통계자료 이용에 관한 적절한 자료제공 정책을 수립하여 일관성 있는 자료제공이 되도록 해야 한다. 특히 원시자료의 제공전략은 매우 민감한 부분으로 자료형태에 따라 정책을 수립하는 것이 바람직 할 것이다.

이를 위해 주요이용자 그룹과 통계작성기관간의 공개적인 협의과정도 검토할 수 있다.

### [가이드라인]

□ **이용자들이 자료에 대한 접근성이 용이하도록 배려하는 한편, 다양한 형태의 자료를 생산하여 각 이용자그룹의 수준에 맞는 정보를 제공하는 정책이 필요하다.**

모든 이용자들에게 동일한 수준의 정보를 일률적으로 적용할 경우 자칫 응답자의 정보가 낮은 수준에서 누출될 가능성이 있다. 이를 위해 연구자용, 학생용 등의 자료이용수준을 지정하여 관리함으로써 이용자 수준에 따른 정보관리가 될 수 있다.

- 자료의 형태에 따라 거시자료, 또는 원시자료의 제공기준을 설정하는 것도 고려할 수 있다.

예를 들어 인구자료의 5% 는 원시자료로 제공할 수 있다거나, 사업체 자료는 특정 변수만을 원시자료로 제공할 수 있다는 어떤 자료제공 기준을 설정하여 이를 이용자에게 공지하여 자료에 대한 관리가 이루어질 수 있도록 해야 한다.

- 원시자료나 거시자료 제공시 자료의 품질정보도 함께 제공하도록 한다.

원시자료의 경우 응답률, 대체율, 포괄성 등의 품질정보를 제공하고, 거시자료의 경우 비교가능성, 일관성 등에 대한 정보를 제공하도록 한다. 이는 통계작성기관에서 제공하는 자료의 신뢰성을 향상시킬 수 있는 방안이 된다.

- 통계자료 제공자는 이용자측면에서 통계정보를 쉽게 검색할 수 있도록 다양한 프로그램을 제공하도록 한다.

검색엔진을 사용하거나 통계생산기관의 포털 사이트에서 이용하고자 하는 통계 목록에 쉽게 접근할 수 있도록 해야 한다.

- 원시자료를 이용할 때에는 이용자에게 자세한 자료의 레이아웃 등을 제시하여 자료를 다루기 쉽도록 해야 한다.

자료에 포함된 변수의 수, 각 변수의 형태와 길이, 변수의 레이블과 계산이 필요한 변수는 계산식에 포함된 변수가 어떤 것인지 등에 관한 레이아웃을 제공하여 이용자가 잘못된 자료 분석결과를 오용하지 않도록 해야 한다.

### 3. 자료관리

조사로부터 수집된 자료 또는 행정절차로부터 작성되는 자료는 향후 자료의 재사용, 또는 시계열 분석 등을 위해 적절한 방식으로 저장, 관리되어야 한다. 최근 컴퓨터 기술의 발달로 대용량의 자료도 매우 적은 비용으로 저장 및 관리가 가능하다.

가능한 한 자료는 관리가 용이한 형태로 저장되는 것이 바람직하며, 주로 통계작성 기관에서는 타 시스템과의 연동을 위해 데이터베이스로 관리하도록 하고 있다. 향후 다양한 분석들을 조망하여 보다 효과적으로 활용될 수 있는 관리정책을 수립하도록 한다.

#### [가이드라인]

□ **데이터베이스화하고자 하는 자료를 분류하여 가능한 한 분류목록을 표준화하는 작업이 필요하다.**

도서관에서 책을 열람하듯이 이용자가 통계작성기관의 데이터베이스에 접속했을 때, 저장된 자료가 일목요연하게 목록화 되어야 한다. 산재된 자료는 이용자뿐만 아니라 데이터베이스관리자에게도 혼란을 초래할 수 있기 때문이다.

□ **자료의 저장 및 관리에 적용된 데이터베이스는 적절한 보안관리가 필요하며, 이용자가 데이터베이스에 접근하고자 할 때에는 승인을 받도록 하는 일종의 자료 관리 정책을 수립할 필요가 있다.**

이용자가 아무런 제한 없이 데이터베이스에 접근할 수 있다면, 데이터베이스의 신뢰성과 보안성에 매우 큰 문제가 아닐 수 없다. 따라서 자료를 관리하는 기관의 승인을 받은 이용자만이 접근이 가능할 수 있도록 하는 이용자 관리정책이 필요하며, 이를 공지하도록 해야 한다.

□ **현재 구축된 데이터베이스에 문제가 발생할 수 있는 상황을 대비하여, 주어진 예산 하에서 예비 데이터베이스를 구축하도록 한다.**

예상치 못한 문제로 데이터베이스를 이용할 수 없는 경우를 대비하여 예비적인 자료관리 시스템을 두는 것이 좋다. 현재 구축된 데이터베이스의 자료를 백업하여 예기치 못한 긴급 상황에 대비하도록 한다.

## 제7절 개선 사이클

통계작성 절차 전반에 대한 품질 평가를 통해 개선점을 도출하고, 이를 다시 향후 작성절차에 반영하는 제반과정을 의미한다. 즉, 운영절차를 개선하려는 노력이 생산된 통계의 품질에 지속적으로 영향을 주게 되고 하는 제반과정을 개선 사이클 이라 한다. 따라서 이러한 모든 절차는 통계작성 기관의 일상적인 경영활동과 연계할 필요가 있다. 왜냐하면 해당기관에서 운용하는 모든 통계프로그램에 대해 이용자의 자료 품질, 이용자 수요, 예산제약과 응답부담간의 절충으로 균형을 유지할 책임이 있기 때문이다. 각 품질평가지표에 대한 개선점을 도출하여 이를 문서에 반영하고, 이를 기반으로 통계작성 전반의 품질개선이 기여하도록 해야 한다.

### [가이드라인]

- 통계품질관리는 경영관리의 관점에서 운영되도록 기관장의 통계품질 마인드가 매우 중요하다.

통계작성기관의 장이 어떠한 품질 마인드를 가지고 조직을 운영하는 자가 매우 중요한 사항 중 하나이다. 이는 기관장의 리더십과도 연관되며, 통계품질관리의 중요성을 인지하고, 이를 적극적으로 이행할 수 있도록 장려할 필요가 있다.

- 통계결과를 문서화하는 과정에서 반드시 품질 평가의 내용이 반영되도록 하고, 이에 대한 개선점을 명시할 수 있도록 하여 향후 개선 사이클에 이용되도록 해야 한다.

- 통계작성기관의 통계품질 개선노력을 객관적으로 평가하여 우수기관에 대해서는 인센티브 등을 제공하는 품질관리 유인정책을 마련할 필요가 있다.

통계품질관리결과를 기관별로 평가하여 우수기관에 대해서는 적절한 인센티브를 제공함으로써 자발적인 통계품질 관리가 될 수 있도록 유인정책을 마련하도록 한다. 자의가 아닌 타의에 의한 품질평가는 자칫 거부감을 느낄 수 있으므로 가능한 한 자발적으로 품질개선을 위해 노력할 수 있도록 해야 한다.

**□ 투입비용대비 통계품질의 향상에 기여한 내용은 반드시 향후 품질평가 보고회에서 우수사례로 전파될 수 있도록 한다.**

외부품질진단 결과 보고회와 같이 자체품질진단결과와 그에 대한 개선점을 적절히 제시한 통계품질관리 보고서는 향후 보고회를 통해 우수사례로 각 기관에 전파될 수 있도록 하는 노력이 필요하다.

#### 1. 적용가능성/유연성

통계작성과정에 대한 품질평가 후 새로운 수요나 개선점이 발견되면, 이를 향후 조사과정에 반영해야 한다. 이에 대한 대응력을 유지하고 개선하는 것을 유연성 또는 적용가능성이라 한다. 철저한 관리 감독 하에서 수집된 자료에 대해 조사방법의 개선이나 조사표상의 변수 추가와 같은 새로운 수요가 발생하면 이를 적극적으로 반영할 수 있는 신속적인 대응이 필요하다. 패널조사와 같이 주기적으로 반복되는 조사의 경우 주요조사 이외에 비용절감을 위해 특정 관심사항에 대해 추가로 조사를 수행할 수 있는지를 검토하고, 그에 따라 조사설 계와 자료수집이 가능하도록 보다 유연하게 계획되어야 한다.

#### [가이드라인]

**□ 기존의 통계작성절차에 다른 자료를 수집하기 위한 새로운 방법을 적용할 수 있도록 조사기획 단계부터 검토해야 한다.**

통계작성기관은 이용자 요구를 충족시키기 위한 새로운 작성 절차를 도입할 수

있는지를 조사기회단계에서부터 검토하도록 한다. 즉, 예산, 인력, 조사방법, 조사 설계, 자료수집 등 제반절차를 유연성 있게 설계하는 것이 향후 개선요구에 적절히 대응할 수 있다. 이를 위해 초기단계에서 이용자 그룹에 대한 수요조사를 검토하는 것이 바람직하다.

**□ 통계결과 분석에서 사용되는 통계프로그램에 대한 개선요구에 적절히 대응하기 위해 가능한 프로그램 목록이나 소스코드를 제공함으로써 보다 유연하게 대응할 수 있어야 한다.**

통계작성기관은 통계분석과정에서 이용자가 다양한 통계분석 프로그램을 사용할 수 있도록 가능한 프로그램 목록과 그에 해당하는 소스코드를 제공함으로써 이용자 요구에 대응하도록 한다. 즉, 주로 사용되는 통계패키지인 SPSS 이외에 SAS 나 STATA, SUDDAN, R 등과 같이 다양한 프로그램 소스를 제공하도록 한다.

**□ 새로운 통계수요에 대응하기 위한 제반 조직과 예산, 시간 등을 충분히 감안해야 한다.**

이용자의 요구에 대응하기위해 무리하게 조직을 편성하거나, 예산을 증액하는 것은 매우 어려운 일이다. 그러므로 향후 발생할 이용자 요구를 적절한 수준까지 수용하기위해 예산, 조직, 인력, 시간을 적절하게 예상하여 운영하는 것이 바람직하다.

## 2. 관련분야의 전문성

통계작성기관의 직원은 가능한 범위 내에서 해당 분야의 전문성을 갖도록 해야 한다. 잦은 부서이동은 통계조직의 전문성을 상실하기 쉽다. 현재 정부부처나 지방자치단체의 통계관련부서의 직원들은 이러한 측면에서 전문성이 매우 결여되어 있다고 해도 과언은 아닐 것이다. 통계작성은 일반 행정 업무와는 달리 매우 전문적인 업무영역이다. 따라서 정기적으로 관련 담당자들에 대한 보수교육을 받도록 해야 하며, 새로운 분석이론이나 조사방법에 대한 세미나 또는 워크샵 등을 개최하여 이들

에 대한 전문성을 향상시켜야 한다. 가능하다면, 외부의 우수한 통계인력을 채용하는 것도 고려할 수 있다.

#### [가이드라인]

□ 통계작성기관의 담당 직원의 업무 평가를 통해 통계 관련 지식을 평가하고, 개발할 수 있도록 해야 한다.

일정한 통계관련 주제를 주고 조별 토론 또는 주제별 토론을 통해 담당자들의 통계지식을 평가하고, 이들의 능력을 개발할 수 있는 기회를 부여해야 한다. 통계는 일반 행정업무와는 성격이 다르기 때문이다.

□ 통계담당자들에 대해 정기적인 보수교육을 실시하여 새로운 분석기법이나 조사 방법 등을 접할 수 있도록 해야 한다. 가능하다면, 외부의 전문가를 초빙하여 보다 심층적인 교육이 되도록 한다. 또한 학회나 전문가 협회와의 다양한 교류를 통해 전문영역을 확대할 수 있도록 한다.

□ 기관의 장이나 관리자는 담당직원들이 다양한 교육프로그램에 자발적으로 참여할 수 있도록 업무개선을 추구해야 한다.

좋은 교육프로그램을 수행할 지라도 업무의 과중은 담당자들이 자발적으로 참여할 수 없게 만든다. 따라서 관리자는 이러한 점을 충분히 감안하여 업무를 개선할 필요가 있다. 가능하다면, 이에 대한 평가를 병행하고 인센티브부여 등 각종 프로그램 참여 유인책을 검토해야 한다.

### 3. 품질관리

통계결과에 대한 품질관리는 결과적으로 작성된 통계의 개선과 밀접한 관련이 있다. 즉, 품질관리가 잘 된 통계는 결과적으로 개선점을 충분히 보완하여 새로운 요

구에 적절히 대응 할 수 있으며, 반대로 품질이 개선된 통계는 이용자들의 요구를 충분히 수용할 수 있기 때문이다. 이러한 점은 결과적으로 통계작성기관의 신뢰성을 높이게 되며, 따라서 투입된 비용대비 고품질의 산출물을 얻게 되며, 궁극적으로는 조직경영의 효율성을 추구할 수 있게 된다. 많은 부분에서 아직까지 통계의 품질 관리와 경영관리의 효율성은 별개의 것으로 인식하는 경향이 있다. 그러나 이는 결코 별개의 업무가 아니며, 경영관리의 효율성을 이루기 위해서는 통계작성기관의 통계품질을 관리함으로써 가능하다고 할 수 있다.

#### [가이드라인]

□ 통계품질관리는 경영관리의 관점에서 운영되도록 기관장의 통계품질 마인드가 매우 중요하다. 두 관점이 결코 다른 것이 아님을 인식할 필요가 있다.

□ 품질진단의 주요 사항들을 만족하기 위해 가능한 효율적인 방법을 강구해야 한다.

품질진단의 7개 요소를 모두 만족시키는 통계를 작성하기 위해서는 매우 큰 규모의 예산과 조직이 필요하다. 따라서 주요사항을 점검하여 조직의 상황에 맞춰 성취 가능한 사항들을 중점적인 관리가 필요하다. 이는 통계품질보고서를 근간으로 주요개산사항들을 점검함으로써 가능할 것이다.

□ 통계품질진단은 가능한 범위내에서 주기적으로 수행되도록 하고, 그에 필요한 인력과 예산을 적절히 배분해야 한다.

□ 통계작성의 주요절차에 대한 검토를 통해 예산, 인력등이 적절히 배분되었는지에 대한 검토도 필요하다.

통계의 품질은 전체적인 통계작성과정이 원활히 수행될 때 고품질의 통계가 생산될 수 있다. 따라서 통계작성 절차 전반에 걸쳐 각종 자원이 적절히 배분되었는지

를 점검하고, 자원배분이 치우친 부분에 대해서는 개선을 유도해야 한다.

□ 통계산출은 결코 예산의 낭비가 아님을 인식해야 한다. 주요정책결정의 기초 자료가 되며, 만일 저 품질의 통계로 정책을 결정했을 때 그에 대한 예산의 낭비는 더욱 크며, 통계적 기초자료가 빠진 정책은 결과적으로 탁상공론이 되기 쉽다.

정책의 입안과 결정과정에 반드시 통계는 필요하며, 이를 위해 고급품질의 통계가 생산되어야 한다. 통계결과의 산출에 소요되는 비용은 결코 예산의 낭비가 아니라 마인드가 반드시 필요하다.

## 제5장. 자체통계 품질진단서 작성을 위한 이용자 가이드

이 장에서는 4장에서 제시한 자체통계 품질진단 가이드라인에 근거하여 진단서를 작성하는데 필요한 사항들을 작성자 관점에서 용이하게 작성할 수 있도록 각 문항별로 상세히 제시하고자 한다. 이를 위해 각 점검 문항별로 작성의 근거와 작성에 필요한 사항을 중심으로 내용을 전개하고, 문항에 대한 이해의 폭을 넓히기 위해 중요 용어에 대한 해설을 제시하고, 문항의 세부 항목에 대한 설명을 제시하며, 문항 작성시 전문성이 요구되는 부분에 대해서는 별도의 사례를 제시하도록 하였다.

내용은 자체품질진단의 주요항목을 중심으로 각 장별로 순서에 따라 전개하고, 각 장별 세부 항목에 대해 설명을 다루도록 하였으며, 제시되는 내용은 향후 통계청의 자체품질 진단 관련 홈페이지의 “도움말” 기능으로 활용할 예정이며, 가능한 한 작성자 관점에서 다루도록 하였다. 자체품질진단을 위해 요구되는 제반 사항은 조사통계의 경우 표본설계 내역과 조사품질과 관계된 제반 서류 및 에디팅 작업과 관련된 문서가 반드시 요구된다. 또한 무응답 대체가 이루어진 경우 대체방법 및 수행 결과와 자료 분석 절차에 관한 문서 등이 요구된다. 보고통계와 가공통계의 경우 행정등록 자료의 이용에 관한 행정 절차 및 자료의 이용방법 등에 대한 내용이 수록된 문서와 조사 자료와의 연계작업이 수행된 경우 연계성공률과 같은 지표가 수록된 문서를 준비하도록 한다.

작성 방법은 문항별로 해당하는 항목에 체크하며, 만일 해당 항목이 2가지 이상인 경우 품질특성평가 항목을 제외한 항목에 대해서는 중복체크가 가능하다. 또한 소관 통계의 사항에 부합하지 않는 항목에 대해서는 생략이 가능하며, 점검표상에서 다음 항목으로 이동하여 작성할 수 있도록 하였다.

## 제1절 작성 지침

### 1. 조사수행을 위한 의사결정

#### 1-1. 조사수행의 법적근거

##### [작성 근거]

통계작성기관에서는 직접적으로 관련 통계를 작성하기 위한 법률적인 근거가 마련되어 있다. 예를 들어 환경 통계의 경우 환경 관련 법률에 따라 조사목적이 정해지거나, 보건통계의 경우 보건관련 법률에 의해 조사목적이 정해지기도 한다. 따라서 해당 통계의 조사목적에 따라 법률에 의해 반드시 통계가 생산되어야 하거나, 관련 정책수립의 기초 자료가 되는 통계가 생산되도록 제시하는 경우에 해당 된다. 또한 통계청에 의해 지정통계로 지정된 경우 통계법에 의해 반드시 통계를 생산하도록 하고 있다. 따라서 이러한 법률적 근거가 국내법 또는 국제법(UN, ILO, WHO, 등의 규정 또는 국제해사법과 같은 국제법)에 근거한 경우 관련 항목에 체크하도록 한다.

##### ☞ 작성방법

| 조사명칭         | 법적근거          |
|--------------|---------------|
| 국민건강영양조사     | 국민건강 증진법, 통계법 |
| 가구소비실태조사     | 통계법           |
| 과학기술연구개발활동조사 | 과학기술 기본법 13조  |

#### 1-2. 조사에 대한 주요 이용자 정보

**[작성 근거]**

통계작성기관에서 수행되는 조사에 대해 정부기관, 대학, 연구기관, 학생 등의 각 계 해당 통계의 주요이용자 정보를 파악하고 있는 정도에 따라 관련항목에 체크한다. 이때 주요이용자란 통계작성기관에서 생산한 통계정보를 요구하는 정도에 따라 우선순위를 두고 해당 순위의 상위그룹에 속하는 이용자를 말한다. 또는 정책의 우선순위에 따라 기관별 분류가 가능하면, 정책수립의 상위기관을 주요이용자로 고려할 수 있다(사례 3-2-2 참조).

**☞ 작성방법**

| 조사명칭         | 주요이용자                |
|--------------|----------------------|
| 국민건강영양조사     | 보건복지부, 대학, 연구기관      |
| 가구소비실태조사     | 재정경제부, 통계청, 대학, 연구기관 |
| 과학기술연구개발활동조사 | 과학기술부, 대학, 병원        |

**1-3. 조사 주요 이용자의 순위****[작성 근거]**

통계작성기관에서 생산되는 통계의 주요 이용자에 대한 순위를 중요도에 따라 부여하여 1순위부터 5순위 까지를 체크한다. 1-2 문항에서 파악된 주요이용자 정보를 고려하여 문항을 작성한다. 다음은 해당항목별 구분을 위한 예이다.

- ① 국제기구 : UN 또는 산하기구 , WHO, WTO, World Bank, OECD, EU , ILO 등의 국제기구.
- ② 단체 및 협회 : 시민사회 단체 또는 관련 협회.
- ③ 정부기관 : 정부의 주요 부처 또는 산하 기관.
- ④ 언론기관 : 신문, 방송, 인터넷 매체 등을 운영하는 기관.
- ⑤ 금융기관 : 한국은행, 각종 공공 또는 민간 은행, 보험기관, 기타 금융기관.

⑥ 기업 : 국내외 기업 또는 다국적 기업.

⑦ 연구기관 및 대학 : 국내외 공공 또는 민간 연구기관 및 대학.

☞ 작성방법

| 조사명칭         | 주요이용자 순위                     |
|--------------|------------------------------|
| 국민건강영양조사     | 1.보건복지부, 2.대학, 3.연구기관        |
| 가구소비실태조사     | 1.재정경제부, 2.통계청, 3.대학, 4.연구기관 |
| 과학기술연구개발활동조사 | 1.과학기술부, 2.대학, 3.병원          |

1-4. 주요 이용자의 관리 실태

[작성 근거]

통계작성기관에서 관련 통계의 주요 이용자들에 대한 정보를 어떠한 형태로 관리하고 있는지를 파악하는 문항이다. 따라서 통계작성기관이 발간물이나 자료를 제공하고 있는 기관 또는 개인에 대한 정보를 DB로 작성하여 관리하고 있는지, 이전 조사자료에서 파악된 주요이용자 정보가 있는지, 또는 현재 작성된 통계 이외의 유사 통계에 대한 수요로부터 이용자를 관리하고 있는지를 파악함으로써 가능하다.

☞ 작성 방법

- ① 발간물의 DM발송을 위한 DB 자료가 있는 경우 또는 별도의 이용자 DB를 관리하는 경우
- ② 이전 조사 자료를 파일형태로 관리하여, 조사표 상의 이용자 정보를 이용하는 경우
- ③ 조관통계와 유사한 통계를 사용하여 이용자의 수요에 대응하는 경우
- ④ 이용자 정보를 전혀 관리하지 않는 경우

1-5. 이용자 자문을 구하기 위한 절차

**[작성 근거]**

통계작성기관에서 소관통계와 관련된 사항에 대해 자문을 구할 경우 어떠한 절차로 이루어지는가를 파악하고자 하는 문항이다. 관련 부처와의 협의 또는 연구기관으로부터의 자문에 의해 통계가 생산되는지, 별도의 실무위원회의 결정에 의해 통계가 생산되는지, 마케팅 활동에 의해 통계생산의 필요성을 느껴 조사를 수행하는 지 등 다양한 경로를 통해 이용자자문을 구할 수 있다.

**☞ 작성 방법**

- ① 통계작성 기관 또는 정부관련 기관의 검토과정으로부터 통계작성과 관련된 자문을 구하는 경우
- ② 통계소위원회와 같은 별도의 위원회의 토의과정으로부터 소관통계에 관한 자문과정을 거치는 경우
- ③ 실무이사회 또는 위원회에서 소관통계에 대한 자문과정을 거치는 경우
- ④ 이용자 수요조사와 같은 고객조사를 통해 자문과정을 거치는 경우
- ⑤ 정부기관 또는 공공기관, 연구기관 과의 협의체 구성을 통한 자문과정을 거치는 경우
- ⑥ 통계정보의 판매 등과 관련되어 이용자들로부터 자문을 구하는 경우
- ⑦ 지장조사를 통한 이용자 요구관련 자문을 구하는 경우
- ⑧ 별도의 과정을 거치기보다는 일부 이용자와의 개별적인 접촉을 통해 이용자 자문을 구하는 경우

**1-6. 향후 이용자 수요 예측을 위한 절차****[작성 근거]**

통계작성기관은 이용자 요구에 적합한 통계를 생산해야 한다. 이를 위해 향후 이용자 수요예측을 위한 다방면의 노력이 필요하다. 어떠한 이용자그룹이 소관통계를

이용할 것인지, 얼마나 많은 이용자들이 통계를 이용할 것인지 등에 대한 적절한 수요예측이 필요하며, 이에 대한 절차가 어떠한 방법으로 이루어지는지를 파악하는 문항이다.

#### 작성 방법

- ① 통계작성 기관 또는 관련 정부기관에서 운영하는 통계위원회 또는 소위원회에서 수요예측을 수립하는 경우
- ② 통계작성기관과 정부기관 간의 위원회로부터 수요예측을 수립하는 경우
- ③ 주요 정책 자료나 국회에서 발간되는 기록물을 통해 향후 수요를 예측하는 경우
- ④ 중앙정부 또는 지방자치단체와의 협의를 통한 수요예측이 이루어지는 경우
- ⑤ 주요학술회의 등을 통해 수요예측이 이루어지는 경우
- ⑥ 주요이용자 그룹과의 별도 워크숍등과 같은 모임을 통해 수요예측이 수립되는 경우
- ⑦ 조관통계에 대한 이용자욕구를 파악하여 미 충족욕구를 반영한 수요예측이 이루어지는 경우
- ⑧ 통계작성기관에서 발간하는 통계발간물의 수요를 통한 향후 수요예측이 이루어지는 경우

#### 1-7. 미충족 이용자 수요의 우선순위

##### [작성 근거]

통계작성기관에서 생산되는 통계에 대한 이용자 요구를 파악하기위해 이용자 만족도 조사를 수행하는데, 이때 이용자의 미충족 욕구를 파악하여 이를 분석함으로써 조관통계에 대한 관련성을 향상시킬 수 있다. 이를 위해 미충족 욕구에 대한 우선순위를 부여하여 중요도에 따라 개선사항을 체크할 수 있다.

☞ 작성 방법

예) 1순위) 조사문항의 개선, 2순위) 통계자료 제공의 시의성, 3순위) 자료제공방법의 개선

1-8. 미충족된 주요 이용자 수요의 개선을 위한 계획

[작성 근거]

통계작성기관은 1-7번 항목에서 파악된 미충족 욕구를 개선하기 위한 계획을 기록한다.

☞ 작성 방법

2008년도 통계자료 생산과정에서는 조사표상의 불필요한 조사문항을 제거하고, 현재 제공되고 있는 통계자료 제공 시점을 단축하도록 하며, 원시자료와 메타자료제공 방법을 현재 off-line을 통한 제공에서 웹이나 이메일을 통한 방법으로 개선하고자 함.

1-9. 조사목적의 명문화

[작성 근거]

통계작성기관에서는 조사기획단계에서 조사목적은 반드시 명시하고, 이를 명문화해야 한다. 이는 생산되는 통계자료의 법적, 제도적 근거가 되기 때문이다. 설문지상에 표기되거나, 조사 설계 보고서 상에 조사목적이 명시되도록 한다.

☞ 작성 방법

설문지 표지, 보고서, 조사 설계 등의 내용에 조사목적이 명시되어 있는 경우 명문화된 조사목적으로 간주할 수 있다.

2. 조사설계

## 2-1. 비교 가능한 공식/비공식 조사의 유무

### [작성 근거]

통계작성기관에서는 비교 가능한 조사를 파악하여 통계의 일관성과 시의성 등을 평가할 수 있다. 비교 가능한 통계로서는 전년도에 작성된 통계나 타 기관에서 작성되는 유사통계를 모두 망라할 수 있다. 조사 성격이 유사한 통계를 이용하여 가능한 과학적인 표본설계가 가능하며, 오차의 수준을 예측해 볼 수 있다.

### ☞ 작성 방법

현재 작성하고 있는 통계와 성격이 유사한 통계로서는 동일한 모집단을 사용하거나, 조사방법이 유사한 경우 또는 포괄성이 유사한 통계, 조사시점이 유사한 통계 등이 이에 해당된다.

## 2-2. 조사품질 개선을 위해 활용 가능한 공식/비공식 조사의 유무

### [작성 근거]

현재 작성하고 있는 통계와 비교 가능한 통계를 활용하여 기관에서 작성하고 있는 통계의 개선에 활용가능한 지를 파악한다. 동일한 형태의 가구단위 조사 또는 사업체 단위조사인 경우 작성기관에서 적용하고 있는 표본추출방법이나, 조사방법, 추정방법, 자료처리방법, 공표방법 등에 대한 개선점을 도출할 수 있기 때문이다.

### ☞ 작성 방법

통계작성기관에서 현재 조사와 유사한 성격을 가진 조사가 있는지를 파악하여 해당조사의 품질개선에 활용가능한지를 고려한다. 동일한 프레임을 사용하더라도, 표본의 추출방법이나, 추정방법, 자료처리방법, 공표방법 등에서 개선점이 파악되면, 해당 조사명을 기입한다.

### 2-3. 표준화된 체계, 기준, 개념, 정의의 분류 기준

#### [작성 근거]

통계작성기관에서 조사에 적용하고 있는 각종 분류기준이 표준화되어있는지 여부를 파악한다. 동일한 조사단위를 조사함에도 서로의 분류기준이 상이하여 다른 통계를 생산할 수 있기 때문에 가능한 표준화된 분류기준을 적용하도록 해야 한다. 예를 들어 사업체 조사의 경우 산업분류는 표준산업분류를 적용하고, 사업체의 정의 또한 산업자원부, 노동부 또는 통계청에서 제시하고 있는 정의를 사용하는 것이 바람직하다. 또한 국제비교가 가능한 통계의 경우 가능한 한 국제표준분류를 사용하도록 한다.

#### ☞ 작성 방법

① 국제기준 : ISO, ILO, EU, OECD 기준 등을 고려한 경우

➔ ILO 기준의 실업자 기준을 적용한 경우

② 국내기준 : 산업자원부, 통계청, 재정경제부 등에서 승인된 기준을 고려한 경우

➔ 통계청의 표준산업분류를 적용한 경우

③ 조사특유기준 : 국제기준이나 국내기준에 없거나, 정형화하기 어려운 기준을 적용한 경우

➔ 예를 들어 저소득층을 경상소득의 40%이하 가구와 같이 자체기준에 따라 정의한 경우

### 2-4. 표준화된 분류 기준과 맞지 않는 사항의 유무

#### [작성 근거]

국제기준, 국내기준 등의 표준화된 분류기준에 부합하지 않는 분류기준은 통계의 비교성을 떨어뜨리는 요인이 된다. 유사한 조사임에도 불구하고 분류기준이 달라 상이한 통계가 생산될 위험이 있기 때문에 가능한 한 표준분류를 적용하는 것이 바람

직하다.

☞ 작성 방법

표준분류 기준에 맞지 않는 사항을 기술하도록 한다.

➔ 본 조사에서는 사업장의 규모를 보다 세분화하기 위해 다음과 같이 분류하였다.

- ① 1~4명 ② 5~9명 ③ 10~29명 ④ 30~49명 ⑤ 50~69명 ⑥ 70~99명 ⑦ 100~299명
- ⑧ 300~499명 ⑨ 500~999명 ⑩ 1000명 이상

**2-5. 공통적인 추출틀 사용여부**

[작성 근거]

비교 가능한 통계를 생산하기 위한 조사에서는 공통의 추출틀을 사용해야 한다. 즉, 동일한 모집단에 대해 추정이 가능하도록 함으로서 비교가능성을 개선할 수 있기 때문이다. 국내에서 이용가능한 대표적인 공통 추출틀로서는 사업체 총 조사자료, 인구센서스 자료 등이 있다. 대부분의 가구단위 조사의 경우 인구센서스 자료를 활용함으로써 유사한 조사로부터 비교가능성을 개선할 수 있다.

☞ 작성 방법

- ① 인구센서스 자료, 사업체 조사자료 등 공통의 추출틀을 이용하고 있는 경우
- ② 공통의 추출틀 활용하지 않은 조사의 경우
- ③ 가구단위 조사에서 주민등록 자료나 전화번호부를 활용한 경우

**2-6. 추출틀의 과대포함여부**

[작성 근거]

조사통계의 경우 표본추출의 대상이 되는 추출틀(frame)의 포괄성을 평가해야 한

다. 노후화된 추출틀 경우 과대 또는 과소포함이 내제되어 추정결과가 현실을 반영하지 못하게 되는 문제가 발생할 수 있기 때문이다. 사업체 추출틀의 경우 모집단 사업체들의 업종변경이나, 폐업 등으로 인하여 표본추출 당시 이러한 사업체가 표본으로 추출되는 과대포함이 발생할 수 있으며, 이러한 변동을 적절히 반영하여 표본의 대표성을 유지할 수 있도록 적절한 조치가 필요하게 된다. 가구단위 조사의 경우에는 이사, 재개발 등으로 인해 현재 존재하고 있지 않은 가구가 프레임이 상존하게 되는 과대포함문제가 발생하게 된다. 따라서 이용하고자 하는 프레임이 목표모집단과 어느 정도 부합되는지를 먼저 평가하여 이를 추정에 반영하는 것이 바람직하다.

#### ☞ 작성 방법

##### ① 대량의 과대포함

➡ 목표모집단에 포함되지 않은 허용할 수 없는 수준의 다수의 단위들이 추출틀에 포함되어 있는 경우(20%이상)

##### ② 장당한 과대포함

➡ 목표모집단에 포함되지 않은 다수의 단위들이 추출틀에 포함되어 있는 경우(10~20%이하)

##### ③ 약간의 과대포함

➡ 목표모집단에 포함되지 않은 일부의 단위들이 추출틀에 포함되어 있는 경우(2~5%이하)

##### ④ 아주 작은 과대포함

➡ 목표모집단에 포함되지 않은 극소수의 단위들이 추출틀에 포함되어 있는 경우(2%이하)

##### ⑤ 과대포함이 거의 없음

➡ 추출틀과 목표모집단이 거의 차이가 없음.

## 2-7. 추출틀 과소포함여부

### [작성 근거]

조사통계의 경우 표본추출의 대상이 되는 추출틀(frame)의 포괄성을 평가해야 한다. 노후화된 추출틀의 경우 과대 또는 과소포함이 내제되어 추정결과가 현실을 반영하지 못하게 되는 문제가 발생할 수 있기 때문이다. 사업체 추출틀 경우 모집단의 사업체들의 업종변경이나, 신규사업체의 등록 등으로 인하여 프레임 작성 시 고려되지 않았던 사업체가 현재 존재함으로써 프레임이 이를 반영하지 못하는 문제가 발생하게 된다. 가구단위 조사의 경우에는 재개발, 신규가구의 생성 등으로 인해 인구조사 당시 없었던 가구가 현재 존재함으로써 프레임이 이러한 가구변동 사항을 적절히 반영하지 못함으로써 과소포함문제가 발생하게 된다. 따라서 이용하고자 하는 프레임이 목표모집단과 어느 정도 부합되는지를 먼저 평가하여 이를 추정에 반영하는 것이 바람직하다.

#### ☞ 작성 방법

##### ① 대량의 과소포함

➡ 목표모집단에 포함되지 않은 허용할 수 없는 수준의 다수의 단위들이 추출틀에 포함되지 않은 경우(20%이상)

##### ② 상당한 과소포함

➡ 목표모집단에 포함되지 않은 다수의 단위들이 추출틀에 포함되지 않은 경우(10~20%이하)

##### ③ 약간의 과소포함

➡ 목표모집단에 포함되지 않은 일부의 단위들이 추출틀에 포함되지 않은 경우(2~5%이하)

##### ④ 아주 작은 과소포함

➡ 목표모집단에 포함되지 않은 극소수의 단위들이 추출틀에 포함되지 않은 경우(2%이하)

##### ⑤ 과소포함이 거의 없음

➡ 추출틀 목표모집단이 거의 차이가 없음.

## 2-8. 통계단위의 분류오류 존재여부

### [작성 근거]

통계단위의 분류오류는 비교성과 일관성 등에 문제가 발생하는 원인이 된다. 따라서 가능한 한 이러한 분류오류가 발생하지 않도록 단위를 분류해야 하며, 이를 위해 국제기준이나, 국내기준을 적용하도록 한다. 산업, 직업분류, 연령별 분류, 지역별 분류 등이 포괄적(exhaustive)이어 하며, 상호 배반적(mutually exclusive)이 되도록 분류해야 한다. 그러므로 조사단계에서 단위에 대한 적절한 분류가 이루어 졌는지를 점검해야 하며, 만일 분류오류가 발생했다면, 조사이전 단계에서 반드시 수정하도록 한다.

### ☞ 작성 방법

#### ① 아주 큰 수의 분류오류

- ➡ 통계단위의 분류체계가 거의 맞지 않거나, 포괄적, 상호배반 규칙에 전혀 부합되지 않는 경우

#### ② 상당히 큰 수의 분류오류

- ➡ 통계단위의 분류체계가 상당부분 맞지 않거나, 포괄적, 상호배반 규칙에 일부 항목만 부합되는 경우

#### ③ 일부의 분류오류

- ➡ 통계단위의 분류체계가 일부항목에 대해서만 맞지 않거나, 포괄적, 상호배반 규칙에 상당 부분 항목이 부합되는 경우

#### ④ 아주 적은 분류오류

- ➡ 통계단위의 분류체계가 매우 적은 항목에 대해서만 맞지 않거나, 포괄적, 상호배반 규칙에 거의 대부분 항목이 부합되는 경우

#### ⑤ 잘못된 분류가 거의 없음

- ➡ 적절한 분류체계를 갖추어 분류오류가 거의 없는 경우

## 2-9. 조사주기와 추출틀 갱신주기 간의 시차

### [작성 근거]

조사통계의 경우 추출틀 갱신주기가 포괄성을 평가하는 중요한 기준이 된다. 이러한 점에서 추출틀 갱신 주기가 짧을수록 포괄성이 높게 되며, 조사주기와 추출틀의 갱신주기간의 시차가 짧은 것이 바람직하다. 그러나 현실여건상 그렇지 못한 경우에는 이들 간의 시차를 명시함으로써 통계작성단계에서 이를 고려할 수 있도록 해야 한다. 사업체 총조사, 농어가 기초조사, 인구센서스 등은 5년 주기로 작성되기 때문에 이에 근거한 추출틀은 현실적으로 시차가 발생하게 된다. 조사시점이 2007년 7월 1일 경우 2005년 인구센서스 자료를 추출틀로 이용하게 되면, 2년의 시차가 발생하게 된다.

### ☞ 작성 방법

#### ① 3개월 이내

➡ 2005년 센서스 자료를 이용하여 2006년 3월 이내에 조사를 수행한 경우

#### ② 6개월 이내

➡ 2005년 센서스 자료를 이용하여 2006년 6월 이내에 조사를 수행한 경우

#### ③ 9개월 이내

➡ 2005년 센서스 자료를 이용하여 2006년 9월 이내에 조사를 수행한 경우

#### ④ 12개월 이내

➡ 2005년 센서스 자료를 이용하여 2006년 12월 이내에 조사를 수행한 경우

#### ⑤ 1년-3년 미만

➡ 2005년 센서스 자료를 이용하여 2006년 12월 - 2008년 12월 이내에 조사를 수행한 경우

#### ⑥ 3년 이상

➡ 2005년 센서스 자료를 이용하여 2009년 1월 이후에 조사를 수행한 경우

## 2-10. 조사주기와 추출틀 갱신간의 시차가 짧아지면 추출틀의 과대 또는 과소포함 문제가 해결되는지 여부

### [작성 근거]

조사주기와 추출틀 간의 시차가 짧을수록 어느 정도 포괄성 문제를 해결할 수 있다. 따라서 제반 조사의 추출틀의 최종갱신 시기가 최근 일수록 포괄성은 높게 되며, 조사 추출틀로 이를 사용하는 것이 바람직하다. 그러나 현실적으로 인구센서스 자료의 포괄성 갱신은 이루어지고 있으나, 이를 반영하지는 못하는 상황이다.

### ☞ 작성 방법

- ① 추출틀의 최종 갱신시기와 조사주기간의 시차를 짧게 할수록 포괄성문제를 해결 할 수 있는 경우
- ② 추출틀의 최종 갱신시기와 조사주기간의 시차를 조정할 수 없는 경우

## 2-11. 추출틀과 목표모집단 간의 근접성 측정 방법

### [작성 근거]

추출틀의 포괄성을 개선하기위해 여타의 다른 자료를 서로 연계할 수 있는 경우가 있다. 또는 주기적인 포괄성 검점을 통해 추출틀을 갱신하는 방법도 고려할 수 있다. 사업체 추출틀의 경우 사업체의 휴 폐업 또는 신규등록 자료를 이용하여 사업체 총 조사 자료를 갱신할 수 있으며, 인구센서스자료의 경우 각 지방 사무소의 주기적인 점검을 통해 조사구 변동사항을 반영하여 추출틀의 포괄성을 개선할 수 있다. 그와 다양한 추출틀의 포괄성 개선을 위한 방법을 고려함으로써 과대 또는 과소포함 문제를 평가할 수 있도록 해야 한다.

### ☞ 작성 방법

- ① 목표모집단에 대해 추출틀을 비교 가능한 다른 원 자료간의 결합
  - ➡ 행정자료 또는 등록 자료를 결합하여 추출틀을 갱신한 경우
- ② 중복, 사망, 조사범위 외 단위, 특성변화 파악을 위한 조사결과분석
  - ➡ 추출틀과 조사자료의 분석을 통한 포괄성 변동사항을 점검하는 경우
- ③ 포함정도를 파악하기 위하여 조사표 상에 특수 질문 사용
  - ➡ 조사표 상의 질문을 이용하여 추출틀의 포괄성을 평가하는 경우
- ④ 지역사무소, 실사현장직원, 조사단위 자체를 점검
  - ➡ 조사 실사현장에서 사업체, 또는 조사구를 주기적으로 점검하는 경우
- ⑤ 실사현장에서 추출틀의 일부 자료를 점검(조사범위 외 단위 확인)
  - ➡ 가구 단위조사의 경우 특정 조사구가 가구의 개념에 맞지 않는 단위가 포함되어 있는지 여부를 파악하는 경우
- ⑥ 추출틀을 동일한 지역추출틀의 표본단위들과 비교함
  - ➡ 추출틀로 이용되는 조사구와 지역별 조사구간의 비교를 통한 평가하는 경우
- ⑦ 특정한 기간동안 발생한 변동을 반영하기위해 추출틀을 갱신함
  - ➡ 특정 기간에 조사구의 변동이 발생하여 이를 반영하는 경우- 재개발, 산업지구신규개발
- ⑧ 다른 자료 또는 특별히 설계된 반복조사결과간의 일치성을 점검함
  - ➡ 기존에 조사된 결과나 동일한 조사를 반복하여 조사하는 경우- 패널조사
- ⑨ 동일 표본추출틀을 사용하는 조사에서 산출된 평가정보사용
  - ➡ 같은 추출틀을 이용하는 기존의 조사 자료와의 비교를 통한 평가가 된 경우
- ⑩ 포함범위의 변동을 통제 불가능할 때 행정자료 담당자들을 접촉함
  - ➡ 실사현장에서 추출틀 변동을 파악하기위해 지방자치단체의 행정담당자를 만나 파악하는 경우
- ⑪ 표본설계에서 사용된 지역경계가 분명하고 중첩되지 않는지를 확인하기 위하여 지역추출틀에 대한 지도를 점검함
  - ➡ 조사구 단위 조사에서 지역 추출틀의 지도(조사구 지도)의 경계선등을 파악하는 경우

12 전혀 점검하지 않음

13 기타 방법

## 2-12. 보편적인 표본추출기법의 사용여부

### [작성 근거]

표본조사에서 모집단을 잘 대표할 수 있는 표본을 추출하는 것이 무엇보다도 중요하다. 통상적으로 이러한 표본추출법은 확률추출법을 사용하는 경우가 일반적이다. 따라서 통계작성기관에서 사용하고 있는 표본추출법을 사용하고 있는지 여부를 검토해야 한다. 만일 그렇지 않다면 그 이유를 이용자들에게 명확히 제시하는 것이 바람직하다.

#### ☞ 작성 방법

1 유사한 통계에서 사용하고 있는 표본추출법을 적용한 경우

➡ 층화추출법, 계통추출법, 집락추출법, 확률비례추출법, 이중추출법 등

2 통상적인 표본추출기법을 사용하지 않는 경우

## 2-13. 정확성과 비용을 감안하여 최적의 표본크기를 결정했는지 여부

### [작성 근거]

표본크기를 결정하고자 할 때 고려되어야 할 사항으로 허용오차와 조사비용등이다. 이는 조사 설계 단계에서부터 면밀히 검토하여 최적의 표본설계가 되도록 함으로써 예산의 낭비를 방지하고, 모집단의 특성을 적절히 반영할 수 있는 대표성 있는 표본을 추출할 수 있기 때문이다. 따라서 주어진 비용과 오차의 한계에 부합하는 표본규모를 설정해야 한다.

#### ☞ 작성 방법

- ① 주어진 예산과 오차의 한계를 고려하여 표본규모를 결정한 경우
- ② 예산과 오차의 한계를 고려하지 않고 표본규모를 결정한 경우

## 2-14. 조사에 적용한 표본설계방식

### [작성 근거]

과학적인 표본조사의 경우 확률추출법을 적용하는 것이 일반적이다. 또한 모집단의 분포의 형태에 따라 이를 잘 대표할 수 있는 표본을 추출해야 한다. 확률추출방법에는 단순 확률추출법, 계통추출법, 층화추출법, 집락추출법, 확률비례추출법, 이중추출법 등이 있으며, 비확률 추출법에는 할당추출법, 판단추출법, 눈덩이추출법 등이 있다.

따라서 조사여건과 모집단의 분포에 적합한 표본추출방법을 사용하도록 해야 하며, 가능하면 확률추출법을 이용하도록 한다.

### ☞ 작성 방법

#### ① 미확률 추출설계

- ➡ 할당추출법, 판단추출법, 눈덩이 추출법 등을 사용한 경우

#### ② 절사법

- ➡ 모집단의 분포가 과대하게 기울어진 경우 일정 규모 이상의 단위들에 대해서는 전수조사하고 나머지 단위들에 대해서는 표본조사를 수행하는 방법을 적용한 경우

#### ③ 시간대별 추출법

- ➡ 전화조사의 경우 주어진 조사시간에 따라 표본규모를 달리하여 조사를 수행하도록 하는 방법

#### ④ 단순확률 추출법

- ➡ 등확률 추출설계로서 표본으로 추출되는 단위들이 모두 동일한 확률로 추출되도록 하는 방법

⑤ 계통확률 추출법

- ➡ 모집단 단위들을 순서대로 정렬한 후 일정한 간격을 두어 표본을 추출하는 방법

⑥ 1단계 집락 추출법

- ➡ 모집단의 단위를 일정규모의 집락들로 구성하여 1차 추출단위(PSU)를 표본으로 추출하여 집락내 모든 단위를 조사한 경우- 조사구를 추출하고 추출된 조사구내의 모든 가구를 조사하는 경우임.

⑦ 2단계 집락추출법

- ➡ 모집단의 단위를 일정규모의 집락들로 구성하여 1차 추출단위(PSU)를 표본으로 추출하고, 다시 1차 추출단위내의 2차 추출단위(SSU)를 표본으로 추출하는 경우

⑧ 층화추출

- ➡ 모집단을 상호배반인 층으로 구분하고 이들 모집단 층으로부터 각각 일정규모의 표본을 추출하는 경우

⑨ 이중(층화)추출법

- ➡ 모집단으로부터 대규모 표본을 추출하여 조사한 정보를 이용하여 다시 일정규모의 표본을 추출하여 조사하는 경우

⑩ 확률비례복원추출

- ➡ 모집단의 추출 단위의 크기에 비례하도록 표본을 복원으로 추출하는 경우

⑪ 확률비례비복원추출

- ➡ 모집단의 추출 단위의 크기에 비례하도록 표본을 비복원으로 추출하는 경우

⑫ 기타 방법

- ➡ 앞의 예 이외의 방법을 적용하여 표본을 추출하는 경우

2-15. 비확률 추출설계를 사용한 근거

[작성 근거]

표본조사의 경우 확률추출법을 사용하여 표본을 추출하는 것이 바람직하나, 추출틀 구성이 어렵거나, 또는 존재하지 않는 경우에는 부득이하게 비확률 추출법을 사용해야 한다. 만일 비확률 추출법을 사용했다면, 이에 대한 자세한 이유를 밝힘으로써 이용자들이 통계를 이용할 때 참고가 되도록 해야 한다.

#### 작성 방법

➔ 본 표본설계에서는 모집단 구성이 어렵고, 따라서 추출틀 작성하기 곤란하여 확률추출법을 사용하지 않고, 연령대별로 층화하여 각 연령대별로 일정규모로 표본을 할당 추출하여 조사를 수행하였다.

## 2-16. 층화추출설계시 층별 표본배분 방법

### [작성 근거]

표본조사에서 표본을 추출한 방법이 층화추출인 경우 각 층별로 특성값을 산출할 수 있고, 이를 결합하여 전체적인 추정값을 나타낼 수 있기 때문에 층화추출은 매우 유용한 방법이다. 따라서 층화추출을 사용한 경우에는 각 층별로 일정 규모의 표본을 배분해야 하며, 이를 위해 비례배분, 네이만 배분, 최적배분 방법들이 있다. 이 방법 이외에 떡배분,  $\sqrt{Y}$ 배분 등의 방법도 고려할 수 있다.

#### 작성 방법

##### ① 비례배분

➔ 단순히 모집단의 크기만을 고려하여 표본을 배분하기 위해 다음 공식을 사용한 경우

$$: n_h = n \frac{N_h}{N}$$

##### ② 최적배분

➔ 조사비용과 모집단의 변동, 크기를 함께 고려하여 각 층에 표본을 배분한 경

우로서 다음의 표본배분공식을 적용한 경우

$$: n_h = n \frac{N_h S_h / \sqrt{c_h}}{\sum_{h=1}^L N_h S_h / \sqrt{c_h}}$$

### ③ 데이만 배분

- ▶ 조사비용을 일정하다고 가정하고, 모집단의 변동과 크기만을 고려하여 각 층에 표본을 배분한 경우로서 다음과 같은 표본배분 공식을 사용한 경우

$$: n_h = n \frac{N_h S_h}{\sum_{h=1}^L N_h S_h}$$

### ④ 기타 방법을 적용한 경우

## 2-17. 표본의 연동구조의 이용여부

### [작성 근거]

조사과정에서 조사에 대한 응답자의 응답부담이 클수록 응답을 회피하거나 거짓 응답을 할 가능성이 크게 된다. 또한 주기성을 띤 조사의 경우 한번 표본에 포함되면 주기적으로 응답해야 함으로 이에 따른 응답 부담은 더욱 가중된다. 따라서 이러한 측면에서 일정 기간동안만 표본에 포함되고 그 이후에는 다른 표본으로 대체되는 연동표본(rotation sample)구조를 고려할 수 있다. 대표적인 연동 표본체계로서는 미국의 CPS 조사로서 4-8-4 연동 구조를 가진다.

### ☞ 작성 방법

- ① 연동 표본체계를 따르는 경우에 해당됨.
- ② 그 외 조사방법인 경우임.

## 2-18. 표본추출 및 표본크기를 결정하는데 이용한 소프트웨어

**[작성 근거]**

표본추출을 위한 프로그램을 작성하여 이용하였는지 여부와 그에 따른 표본규모 산출을 위한 프로그램의 사용여부를 검토하는 항목으로 최근 SAS 나 SPSS , STATA, R 등의 범용패키지 프로그램에 표본추출과 관련된 사항들이 개발되어 있으므로 이를 이용하여 표본의 추출과 추정이 가능하다. 이에 대한 설명을 표본설계에 포함시키는 것이 이용자들을 위해 바람직하다.

**☞ 작성 방법**

- ① 자체개발한 프로그램이 있는 경우에 해당함.
- ② 범용 소프트웨어인 SAS, SPSS, EXCEL, STATA, R 등을 이용하여 표본을 추출한 경우

**2-19. 목표표본크기와 실현 표본크기간의 차이여부****[작성 근거]**

일반적으로 조사에서는 무응답과 추출틀의 포괄성 문제 등으로 인하여 일정 수준 차이가 발생할 수 있다. 표본규모를 결정하여 조사를 수행한 다음 실제 표본크기와 어느 정도 차이가 있는지를 점검해야 한다. 이는 표본설계 당시 고려한 오차와 실제 표본크기에 따른 오차의 규모를 판단하여 추정시 반영해야 하기 때문이다. 목표표본크기와 실현표본크기간의 차이가 클수록 조사의 오차의 크기는 증가하기 때문에 이를 반드시 점검할 필요가 있다.

**☞ 작성 방법**

- ① 차이가 없음.
  - ➡ 만일 조사 설계 당시 목표표본크기를 1,000개로 결정하고, 실현표본크기도 1,000개로 조사된 경우
- ② 실현표본크기가 목표표본크기에 비해 큼

➡ 만일 조사 설계 당시 목표표본크기를 1,000개로 결정하고, 실현표본크기가 1,000개 이상으로 조사된 경우

③ 실현표본크기가 목표표본크기에 비해 5%이하로 작음

➡ 만일 조사 설계 당시 목표표본크기를 1,000개로 결정하고, 실현표본크기가 950개 내외로 조사된 경우

④ 실현표본크기가 목표표본크기에 비해 5%이상 작음

➡ 만일 조사 설계 당시 목표표본크기를 1,000개로 결정하고, 실현표본크기가 950개 이하로 조사된 경우

## 2-20. 결측단위에 대한 현장대체 허용여부

### [작성 근거]

조사과정에서 현장 대체의 유무는 조사수행과정에서 반드시 점검해야 한다. 조사원이 임의로 현장대체를 할 경우 표본에 편향이 발생할 가능성이 크기 때문이다. 결측이 발생한 경우 일정 수준의 예비 표본을 제공하여 해당 범위 내에서 대체가 이루어지도록 하고 이를 반드시 체크할 수 있도록 해야 한다.

### ☞ 작성 방법

① 허용함

➡ 조사원이 임의로 현장대체가 가능하도록 하는 경우

② 정해진 한계 내에서 허용함.

➡ 예비표본을 제공하여 그 범위 내에서 현장대체가 가능하도록 하는 경우

③ 허용하지 않음

➡ 만일 조사 설계 당시 목표표본크기를 1,000개로 결정하고, 실현표본크기가 950개 내외로 조사된 경우

## 2-21. 조사 설계의 효율성

**[작성 근거]**

예산, 오차 등을 고려하여 산출한 표본의 규모는 적절한지, 표본추출방법은 현실적인지, 조사과정의 문제를 충분히 고려한 표본설계인지를 판단하여 조사 설계가 효율적인지를 판단해야 한다. 만일 부분적으로 개선할 사항이 있거나, 일부 조사과정에서 고려하지 못한 사항이 발견될 경우 향후 이를 적절히 반영하여 보다 효율적인 조사 설계가 이루어지도록 해야 한다.

**☞ 작성 방법**

- ① 전반적으로 효율적인 경우
- ② 예산과 조사기간 등에서 만족스럽고, 개선할 사항이 거의 없는 경우
- ③ 예산과 조사기간 등에서 만족스럽고, 개선할 사항이 일부 존재하는 경우
- ④ 미효율적인 경우
- ⑤ 조사 설계 전반에 매우 큰 결함이 존재하는 경우

**2-22. 조사 설계의 변동 가능성 여부****[작성 근거]**

조사 설계 전반에 걸쳐 문제점이 발견될 경우 이를 반드시 수정하여 보완토록 해야 한다. 표본추출방법의 문제, 조사과정상의 문제, 표본크기 대비 비용의 과대계상 등의 문제가 발생하게 되면, 조사 자료의 신뢰성과 효율성에 심각한 문제를 야기시킬 수 있기 때문이다. 현재 조사에서 수정이 불가피 한 경우 반드시 문제점을 적시하여 향후 조사에서 이러한 문제를 점검하도록 환류할 필요가 있다.

**☞ 작성 방법**

- ① 지급히 변경해야함.
  - ➡ 조사 설계 전반에 문제가 발생하여 기존의 설계를 변경할 필요가 있는 경우

② 변경해야 하지만, 좀더 테스트해볼 필요가 있음.

➡ 조사 설계 일부분에 문제가 있지만, 검토과정에서 충분히 해결이 가능한 경우

③ 아니오.

## 2-23. 조사표상의 질문 구성 내용

### [작성 근거]

조사의 측정도구인 조사표는 응답자가 응답하기 쉽고, 이해가 용이하도록 작성되어야 하며, 작성과정에서 질문의 흐름을 잘 따라갈 수 있도록 구조적으로 작성되어야 한다. 이를 위해 전문가들의 자문을 통한 질문문항의 구성과 적절한 질문용어를 선택하여 사용하도록 해야 한다. 또한 조사 방법이 종이조사표를 이용하는 경우와 컴퓨터를 이용하는 경우 등에 따라 질문의 구성방법이 달라질 수 있다. 이를 위해 반드시 초기에 작성된 조사표를 사전에 점검하는 과정이 필요하다.

### ☞ 작성 방법

① 사전 검증하지 않은 질문항목들로 구성

➡ 초기 조사표에 대한 사전 검증 없이 최종 문항을 결정하여 조사표를 구성한 경우

② 사전 검증하여 균형 잡힌 질문항목으로 구성

➡ 소규모 표본에 대해 또는 조사원들을 대상으로 하여 조사표에 대한 사전검증을 거친 경우

③ 항목배열에 표준화된 설계로 구성

➡ 조사표 구성과 관련된 이론을 배경으로 표준화된 형태의 질문항목을 구성한 경우

④ 컴퓨터를 이용한 면접을 위해 전자조사표 설계용 프로그램으로 표준화된 소프트웨어로 구성

➡ CAPI, CATI 등과 같은 컴퓨터를 이용한 조사에서 사용하기 위한 프로그램으로 작성된 조사표

⑤ 기관내 전문 조사표 담당자가 점검한 질문항목으로 구성

➡ 통계작성기관내부의 전문가에 의해 조사표를 개발한 경우

⑥ 외부 기관의 전문 조사표 담당자가 점검한 질문항목으로 구성

➡ 통계작성기관 외부의 전문가에 의해 조사표를 개발한 경우

⑦ 국제기구에 의해 제공된 조사표

➡ 국제비교를 위해 외국 또는 국제기관에 의해 개발된 질문항목으로 구성된 경우

## 2-24. 조사표 분량

### 2-24-1. 종단면 조사에서 일반사항의 조사표 분량

### 2-24-2. 종단면 조사에서 추가조사나 일계부 작성시 조사표 분량

#### [작성 근거]

응답자 부담과 관련하여 조사표의 분량이 적절한지를 검토해야 한다. 일반 조사의 경우 20~40분정도가 적절하며, 전화조사의 경우 3~5분 정도가 적절한 것으로 알려져 있다. 한편 일계부 또는 가계부와 같은 매일 또는 매월 정기적으로 작성해야 하는 조사표의 경우 응답자의 응답부담이 상당히 크기 때문에 이에 대한 적절한 보상 정책을 반드시 고려해야 한다.

#### ☞ 작성 방법

① 20분 이내 : 적은 분량으로 응답자가 쉽게 응답할 수 있는 문항으로 구성된 경우

② 45분 이내 : 조금 많은 분량으로 응답자가 조사에 대해 관심을 가지고 응답할 수 있는 문항으로 구성된 경우

③ 45분 이상 : 많은 분량으로 응답자가 응답을 위한 자료제공에 대해 조사원의

설득을 필요로 하는 문항으로 구성된 경우

## 2-25. 조사표 작성시 응답자의 응답 부담정도

### [작성 근거]

통계조사에서 응답자의 조사표 문항에 대한 반응에 따라 조사에 대한 거부감 또는 응답자 부담을 측정할 수 있다. 질문이 평이하고, 조사표의 구조가 단순한 경우에는 그만큼 응답자의 응답부담이 경감되며, 상대적으로 많은 생각이나 관련 자료를 필요로 할 경우에는 매우 큰 응답부담을 가지게 된다.

### ☞ 작성 방법

- ① 낮은 응답부담으로 정보를 쉽게 제공
  - ➡ 주로 만족도나 여론조사와 같이 쉽게 응답이 가능한 경우
- ② 적당한 응답부담으로 일부 정보는 회상이 필요함
  - ➡ 가구원 변동이나 가구원의 현재 나이 또는 교육정도 등과 같이 일부 회상이 필요한 문항으로 구성된 경우
- ③ 응답부담이 있으며, 필요한 정보를 얻기 위해 복잡한 회상이 필요함
  - ➡ 가구의 소득, 지출, 납세기록, 의료비지출내역 등과 같이 회상기간이 긴 문항으로 구성된 경우

## 2-26. 민감한 사항에 대한 응답부담

### [작성 근거]

사회조사의 경우 가구 또는 개인의 소득 또는 지출관련 문항은 매우 민감한 문항으로 응답자로 하여금 매우 큰 응답부담을 느끼게 한다. 따라서 조사표상에서 이러한 문항은 가능한 한 제외하여 응답부담을 경감시킬 필요가 있다. 이를 보충하기 위한 외부 자료와의 연계가 가능하도록 하는 정책적 고려가 필연적으로 고려되어야

한다.

#### ☞ 작성 방법

- ① 낮은 응답부담으로 모든 항목을 쉽게 응답할 수 있음.
  - ➡ 주로 만족도나 여론조사와 같이 쉽게 응답이 가능한 경우
- ② 적당한 응답부담으로 일부 정보는 회상이 필요함
  - ➡ 가구원 변동이나 가구원의 현재 나이 또는 교육정도 등과 같이 일부 회상이 필요한 문항으로 구성된 경우
- ③ 응답부담이 있으며, 필요한 정보를 얻기 위해 복잡한 회상이 필요함
  - ➡ 가구의 소득, 지출, 납세기록, 의료비지출내역 등과 같이 회상기간이 긴 문항으로 구성된 경우

## 2-27. 조사표에 적용된 사항들

### [작성 근거]

조사표의 구성과 디자인, 내용은 응답자에게 얼마나 용이하게 응답을 하도록 할 수 있는가를 결정하는 매우 중요한 요소이다. 따라서 응답자로 하여금 호기심을 유발할 수 있도록 디자인 하거나, 중요 사항에 대해서는 강조표시를 하여 반드시 응답자가 응답할 때 고려할 수 있도록 하는 것이 바람직하다. 또한 조사표 흐름을 잘 이해할 수 있도록 화살표와 같은 지시사항을 조사표에 표시하여 논리적 오류를 범하지 않도록 해야 한다.

#### ☞ 작성 방법

- ① 전문성과 사업성을 나타내는 매력적인 디자인으로 구성됨
- ② 조사표 서두에 조사의 제목 또는 주제, 후원자, 조사목적에 대한 설명, 응답자의 협조요청, 조사기관, 비밀보호에 관한 정보 등을 제시함.
  - ➡ 조사표 표지 부분에 통계법, 조사주체 등에 대한 사항을 제시한 경우

- ③ 장을 구분하기 위한 제목이나 머리말이 있음
  - ➡ 다수의 페이지로 구성된 조사표는 각 주제별로 조사문항을 분류하여 배치한 경우
- ④ 책갈, 삽화, 상징어, 진한 글씨체 사용 등과 같은 그래픽을 추가하여 응답자가 반드시 응답해야 하거나 주의를 환기시키는 표시를 적용함
- ⑤ 개방형 문항이 모든 응답자가 응답하기에 적절하고 용이함
  - ➡ 금액 등과 같이 값을 기록하거나 의견에 대한 기타사항이 있는 경우
- ⑥ 단어와 개념을 일관성 있게 사용하고 있음
- ⑦ 각 문항들이 합당한 이유가 있음
- ⑧ 문항들이 응답자의 논리를 따름
- ⑨ 문항의 배열순서가 건너뛰기 등으로 방향이 적절히 지정됨
  - ➡ 관련문항에만 응답할 수 있도록 화살표와 같은 지시사항이 있는 경우
- ⑩ 조사대상기간과 응답단위 등이 명확함
  - ➡ 조사 지문에 조사대상기간과 가구 또는 사업체 등과 같은 개념이 명시된 경우
- ⑪ 조사표 맨 마지막 페이지에 응답자가 추가로 기입할 수 있는 공란을 마련함.
- ⑫ 조사표 맨 마지막 페이지에 응답에 대한 감사 문구를 제시함.
- ⑬ 지침과 추가 가이드 작성함.
  - ➡ 질문 문항에 대한 이해를 돕기 위한 지침서를 제시하거나 질문에 관련사항을 나타낸 경우
- ⑭ 문의사항이 있을 때 연결 가능한 무료전화 운영
  - ➡ 조사관련 직원 또는 통계작성기관의 전화번호가 제시된 경우

## 2-28. 조사표 품질평가

### [작성 근거]

사용된 조사표의 품질을 2-27에 근거하여 판단함으로써 향후 조사표를 개선할 수

있도록 한다. 응답자 친화성과 체계적인 조사표상의 논리성, 조사예절 등을 포괄하는 조사표를 작성하도록 한다.

#### ☞ 작성 방법

- ① 모든 요구사항을 모두 충족시켜 더 이상 개선이 불필요함.
  - ➡ 2-27문항에서 제시한 항목들에 대해 모든 사항을 만족시키는 경우
- ② 일부 문항에 대해 개선의 여지가 있지만, 대부분의 사항을 충분히 충족시킴
  - ➡ 2-27문항에서 제시한 항목들에 대해 일부사항을 개선할 필요가 있는 경우
- ③ 조수의 문항을 개선이 필요함.
  - ➡ 2-27문항에서 제시한 항목들에 대해 몇몇 사항을 개선할 필요가 있는 경우
- ④ 대부분의 문항에 문제가 있기 때문에 시급한 개선이 필요함
  - ➡ 2-27문항에서 제시한 항목들의 대부분을 만족시키지 못하는 경우

## 2-29. 조사표 검증의 주체

### [작성 근거]

조사 수행에 앞서 조사표에 대한 사전 검증을 통해 구조적 오류를 방지할 수 있으며, 응답자들의 응답을 보다 용이하게 얻어낼 수 있기 때문에 조사표 검증 절차를 거쳐야 한다. 조사표를 완성하는데 걸리는 시간, 응답회피 문항의 조정, 구성개선, 용어의 올바른 정의를 통해 보다 체계적이고 정합성 있는 조사표가 만들어 지도록 해야 한다. 조사표의 검증의 주체는 응답자, 조사원, 내부 위원회, 외부 전문가 등이 될 수 있으며, 각 주체별로 검증이 수행되는 것이 바람직하다.

#### ☞ 작성 방법

- ① 대부적 검증
  - ➡ 통계작성기관 내부의 조사담당자들이 조사표를 검증하는 경우
- ② 통계작성기관 내 별도의 위원회에서 검증

➡ 통계작성기관 내 통계위원회 또는 조사표 검증위원회 등에서 조사표를 검증하는 경우

③~대부에 마련된 조사실험실에서 검증

➡ 체계적인 실험을 위한 면접실 등에서 소수 응답자들이 조사표를 검증하는 경우

④~사전조사를 위한 외부 조사실험실에서 검증

➡ 사전조사를 통한 조사표 검증을 수행하는 경우

⑤~외부전문가가 조사표를 검증

➡ 통계작성 기관 외부의 전문가들에 의해 조사표를 검증하는 경우

**2-30. 조사표 검증방법**

**[작성 근거]**

조사표를 검증하는 방법으로는 사전조사를 통해 응답자의 반응과 조사표 체계 등을 검증할 수 있으며, 이와 더불어 조사원들의 의견을 수렴하여 조사표를 개선할 수도 있다. 한편으로는 조사표의 질적인 검증을 위해 포커스그룹면접, 심층면접, 생각 말하기, 바꿔 말하기, 행태코딩 등이 있으며, 조사대상자를 두 그룹으로 나누어 응답 값을 분석하여 조사문항을 개선하는 방법을 고려할 수 있다.

**☞ 작성 방법**

①~조규모 표본을 선정하여 사전조사

➡ 초기 구성된 조사표에 대한 사전조사를 통해 조사표상의 문제점을 개선하는 경우

②~사전조사를 실시한 후 조사원들로부터 의견을 청취

➡ 초기 구성된 조사표에 대한 사전조사에 참여한 조사원들의 의견을 청취하여 조사표상의 문제점을 개선하는 경우

③-1 포커스그룹 면접(FGI)

- ➡ 소수(5~10명)의 응답자들에 대해 조사목적, 조사문항 등에 관한 유용한 정보를 수집하거나, 그 외 관련 사항을 수집하기 위해 사용된 방법

③-2 심층면접

- ➡ 조사주제에 관한 응답자들의 잠재적 동기, 태도, 신념 등에 관한 정보를 얻기 위해 사용된 방법으로 응답자의 내면적 생각을 파악할 수 있음.

③-3 생각 말하기

- ➡ 응답자의 행동과 의도, 마음을 알기위해 응답자는 다양한 행동을 취하며 동시에 그때그때 떠오르는 생각을 말로 표현하는 방법.

③-4 바꾸어 말하기

- ➡ 응답자의 관점에서 조사문항에 대해 응답자가 가진 생각과 정보를 응답자가 아는 단어로 바꾸어 말함으로써 보다 객관적으로 설문문항을 검토할 수 있는 방법

③-5 행태코딩

- ➡ 조사원과 응답자행위 양태를 용어를 새롭게 정의하거나 조사표를 재설계할 필요가 있는 질문 문항이나 응답자가 응답하기 어려운 질문문항에 대한 정의 수단으로 사용되는 방법.

④ 분리표본검증

- ➡ 동일한 조사표를 서로 다른 두 그룹의 응답자에게 배포하여 두 응답자 그룹 간의 차이를 분석함으로써 조사표의 구성 등을 검증하는 방법.

3. 자료수집

3-1. 조사에 사용된 자료수집 방법의 근거

[작성 근거]

조사에 사용된 자료수집방법이 어떤 근거에 따라 수행된 것인지를 규명하는 것으로 과학적 실험에 의한 검증, 문헌검토 등에 따라 자료수집 방법을 고려할 수 있다.

☞ 작성 방법

- ①~검증에 의한 방법 ➔ 과학적 실험에 의한 검증된 방법에 의해 조사를 수행하는 경우
- ②~문헌검토 ➔ 관련 문헌을 검토하여 자료수집방법을 결정한 경우
- ③~법적인 요구사항 ➔ 관련 법률 또는 규칙에 따라 조사를 수행하도록 하는 경우
- ④~가능한 능력 ➔ 예산이나, 조사인력 등에 따라 자료수집 방법을 결정한 경우
- ⑤~이전조사에 따라 일관적으로 ➔ 주기성을 가진 조사에 따라 이전 조사결과에 따라 자료 수집방법을 결정한 경우
- ⑥~기타방법

3-2. 조사 자료의 수집 방법

[작성 근거]

조사방법은 방식에 따라 면대면 조사와 전화조사, 인터넷조사 등으로 구분될 수 있으며, 면대면 조사는 다시 자기 기입식 조사와 조사원 면접 방식조사로 구분된다. 또한 조사표 형식에 따라 종이조사표 방식, 컴퓨터를 이용한 조사표 방식으로 나눌 수 있다. 이러한 조사방법은 현재 수행하는 조사의 현실에 가장 효율적인 방법을 적용하는 것이 바람직하다.

☞ 작성 방법

- ①~조사원 면접 방식
  - ①-1 전자조사표(CAPI) ➔ 컴퓨터를 이용하여 조사를 수행하는 경우  
(28번 문항으로 이동)
  - ①-2 종이조사표(PAPI) ➔ 전통적인 종이조사표를 이용하여 조사를 수행하는 경우  
(28번 문항으로 이동)

①-3 전자조사표(CATI) ➡ 컴퓨터를 이용한 전화조사를 수행하는 경우  
(24번 문항으로 이동)

② 자기입식 조사방식

②-1 웹조사표

➡ 인터넷상에서 응답자가 직접 조사표를 작성하는 경우(17번 문항으로 이동)

②-2 전자조사표

➡ 컴퓨터를 이용한 조사에서 응답자가 직접 조사표를 작성하는 경우  
(17번 문항으로 이동)

②-3 종이조사표 또는 일계부

➡ 응답자가 종이조사표나 일계부를 직접 작성하는 경우 (17번 문항으로 이동)

③ 행정자료 또는 등록자료

➡ 행정업무에서 생산된 자료 또는 정부 또는 관련기관에 등록된 자료를 사용하여 자료를 수집하는 경우 (4번 문항으로 이동)

④ 혼합방식 ➡ 조사에 의한 자료수집방법과 행정자료 또는 등록자료를 같이 연계하여 사용하는 경우 (3번 문항으로 이동)

⑤ 기타방식

○ 혼합조사 방식

### 3-3. 서로 다른 자료수집 방법에 대한 측정효과 평가여부

[작성 근거]

자료수집방법을 변경하거나, 다른 방법간의 비교연구를 통해 현재 자료수집방법의 효율성을 평가하는 것이 바람직하다. 단순히 자료를 조사만으로 수집하는 것이 자료 변수에 대해 정확성이나 신뢰성에 문제가 있다면, 혼합방식으로 행정자료나 등록자료를 동시에 이용하는 것이 효과적이기 때문이다.

☞ 작성 방법

① 완전히 평가

➡ 모든 품질차원에 대해 두 자료를 모두 평가한 경우

② 충분히 평가

➡ 모든 품질차원에 대해 두 자료를 전반적으로 평가한 경우

③ 평가하였으나, 개선될 수 있음

➡ 모든 품질차원에 대해 두 자료를 평가하였으나, 만족할 만한 수준이 아닌 경우

④ 전혀 평가하지 않음

➡ 두 자료에 대한 효율성을 평가하지 않은 경우

○ 행정자료 또는 등록자료

3-4. 자료의 유지관리 부서

[작성 근거]

행정, 등록자료 등에 대해 별도의 자료관리부서가 있는지를 점검하는 문항이다. 타기관에서 관리하는 지, 아니면 통계작성기관에서 관리하는 지에 따라 자료의 평가가 가능할 수 있기 때문이다.

☞ 작성 방법

① 귀하의 부서 ➡ 통계작성기관의 조사관련 부서

② 귀하의 기관내 다른 부서 ➡ 통계작성기관의 조사관련 부서 이외 타 부서

③ 외부기관의 다른 부서 ➡ 통계작성기관 외부의 타 기관

3-5. 외부로부터의 자료 품질평가 여부

**[작성 근거]**

행정자료 또는 등록자료에 대한 품질평가를 통해 조사 자료와 연계가 가능한지를 점검하는 문항이다. 자료의 연계가능성을 모색하기 위해 요구되는 변수들이 존재하는 지, 연계자료에 포함된 자료는 정확한지 등에 대해 평가를 수행하는 것이 바람직하다.

**작성 방법**

- ① 예 ➡ 외부로부터 자료에 대한 품질평가를 받을 수 있는 경우  
(8번 문항으로 이동)
- ② 아니오 ➡ 그렇지 않은 경우

**3-6. 외부로부터의 품질평가 수용여부****[작성 근거]**

자료에 대한 평가가 가능하다면 외부로부터의 자료에 대한 평가방법을 수용가능 여부를 점검하는 문항으로서 외부 전문가로부터 연계되는 자료에 대한 품질평가를 객관적으로 받음으로서 자료의 신뢰성을 확보할 수 있을 것이다.

**작성 방법**

- ① 예 ➡ 외부로부터 자료에 대한 품질평가를 받을 수 있는 경우
- ② 부분적으로 ➡ 일부 자료에 대해 외부로부터 자료에 대한 품질평가를 받을 수 있는 경우
- ③ 아니오 ➡ 그렇지 않은 경우

**3-7. 자료의 품질에 대한 충분한 정보의 보유여부****[작성 근거]**

연계할 자료의 품질정보를 보유하고 있으면, 자료를 연계할 때 불충분한 부분을 참고할 수 있고, 통계를 생산할 때 자료의 한계점을 명시함으로써 이용자들에게 도움을 줄 수 있다.

#### 작성 방법

- ① 예 ➡ 모든 품질차원에 대한 품질정보를 보유하고 있는 경우
- ② 아니오 ➡ 그렇지 않은 경우

### 3-8. 자료연계의 필요성 여부

#### [작성 근거]

조사 자료와 행정, 등록자료 간의 연계의 필요성을 파악하고 이를 위한 적절한 절차를 수행했는지를 점검하는 문항이다. 불필요한 연계를 통해 자료의 신뢰성을 오히려 떨어뜨릴 수 있으며, 이용자로 하여금 혼란스럽게 할 수 있기 때문에 연계가 반드시 필요한 항목을 점검하여 그에 대한 연계가 이루어지도록 하는 것이 바람직하다.

#### 작성 방법

- ① 예 ➡ 연계가 반드시 필요한 경우
- ② 아니오 ➡ 연계가 불필요한 경우(12번 문항으로 이동)

### 3-9. 레코드 연계방식 형태

#### [작성 근거]

조사 자료와 행정, 등록자료 간의 연계방법에는 두 자료에 동시에 존재하는 key ID를 이용하여 직접적으로 연계하는 방법과 자료의 특성을 파악하여 통계적으로 연계하는 방법을 고려할 수 있다. 전형적인 ID 매치 방법은 직접적인 결합에 해당된

다.

☞ 작성 방법

- ① 직접적인 결합 ➡ key ID를 기준으로 두 자료를 연계하는 경우
- ② 통계적 결합 ➡ 통계적 방법에 의해 두 자료를 연계하는 경우
- ③ 기타

### 3-10. 연계성공률 평가여부

[작성 근거]

두 자료간의 연계시 어느 정도 연계가 성공했는가를 파악하여 연계의 효율성을 평가할 수 있다.

☞ 작성 방법

- ① 예 ➡ 연계성공률을 평가하는 경우
- ② 아니오 ➡ 연계성공률을 평가하지 않는 경우(12번 문항으로 이동)

### 3-11. 연계성공률 평가정도

[작성 근거]

연계성공률을 평가했다면, 어느 정도로 연계가 이루어 졌는지를 평가할 수 있다. 너무 낮은 연계 성공률은 연계의 효율성과 자료의 신뢰성에 문제가 발생할 수 있기 때문이다.

☞ 작성 방법

- ① 거의 오차가 없도록 ➡ 두 자료가 거의 완벽하게 연계된 경우
- ② 만족할 만큼 ➡ 일부 오차가 있지만, 거의 연계가 된 경우

③ 받아 들일만 하지만 개선이 필요

➡ 허용할 수 있는 수준이기는 하지만, 일부 연계자료의 개선이 필요한 경우

④ 받아들이지 못함 ➡ 연계오차가 너무 커서 자료로 이용이 불가능한 경우

### 3-12. 다른 자료 원을 이용할 경우의 문제점

#### [작성 근거]

조사자료 이외에 다른 자료를 사용할 경우 어떠한 문제가 있는 지를 검토하고, 만일 외부 자료를 연계하여 이용함으로써 심각한 문제가 존재하게 되면 연계문제를 재고해야 할 것이다.

#### ☞ 작성 방법

① 포함범위

➡ 이용할 자료의 포함범위(대표성 등)에 문제가 있는 경우

② 분류

➡ 이용할 자료상의 분류오류로 인해 사용이 어려운 경우

③ 정의

➡ 두 자료의 기준이 되는 정의들이 서로 달라서 연계할 경우 문제가 발생하는 경우

④ 갱신 및 시의성

➡ 연계될 자료가 갱신되지 않았거나, 자료의 시의성이 떨어져 연계의 효율성에 문제가 있는 경우

⑤ 중복단위 또는 다중단위

➡ 연계할 자료의 단위가 중복되거나 다중단위가 다수 포함되어 연계성공률이 떨어지는 경우

### 3-13. 보안성 정도

**[작성 근거]**

자료의 수집 및 교환과정에서 개인 또는 사업체와 관련된 중요한 정보가 누설되지 않도록 해야 하며, 이러한 측면에서 자료의 보안성을 평가하는 문항이다.

**작성 방법****① 만족할 만함**

➡ 개인 또는 사업체의 비밀보호에 대한 보안성을 평가할 때 충분히 비밀이 보장되고 있는 경우

**② 받아들일 만 하지만 개선이 필요**

➡ 일부 정보가 누설가능성이 있어 개선이 필요한 경우

**③ 받아들일 수 없음**

➡ 개인 또는 사업체의 정보를 모두 알 수 있어 평가를 받아들일 수 없는 경우

**3-14. 중복단위와 다중단위의 구분****[작성 근거]**

두 자료를 연계함으로써 발생할 수 있는 중복단위와 다중단위의 존재를 파악할 필요가 있고, 만일 이러한 단위들이 존재한다면, 이를 구분하기위해 특별한 표식을 함으로서 표본추출과정에서 이를 반영하는 것이 바람직하다.

**작성 방법**

**① 예** ➡ 중복단위와 다중단위가 존재할 경우 이에 대한 특정한 표식으로 구분이 되는 경우

**② 아니오** ➡ 중복단위와 다중단위에 대한 별도의 표식이 없는 경우(16번 문항으로 이동)

### 3-15. 표본추출 전에 중복단위와 다중단위의 구분

#### [작성 근거]

자료들을 연계하여 표본 추출틀로 이용할 경우 중복단위와 다중단위에 대한 구분이 필요하다. 즉, 추출틀에 존재할 수 있는 동일한 단위가 중복해서 표본으로 추출될 가능성이 있거나, 서로 다른 단위들이 동일한 ID 로 나타난 경우 또한 추출과정에 문제가 발생할 수 있기 때문에 표본추출 전에 이를 확인할 필요가 있다.

#### ☞ 작성 방법

- ① 예 ➡ 표본추출 전에 중복단위나 다중단위에 대해 확인이 가능한 경우
- ② 아니오 ➡ 표본추출 전에 중복단위나 다중단위에 대해 확인하지 않고 사용하는 경우

### 3-16. 조사에 행정자료나 등록 자료의 효율적인 사용여부

#### [작성 근거]

조사자료 뿐만 아니라 행정자료나 등록자료를 연계하여 보다 풍부한 정보를 사용함으로써 효율성을 높일 수 있다. 따라서 자료의 효율적인 이용 측면에서 충분한 고려가 필요하다.

#### ☞ 작성 방법

- ① 예 ➡ 자료의 효율적인 사용 측면에서 행정자료나 등록 자료를 충분히 사용할 수 있는 경우
- ② 아니오 ➡ 자료의 효율적인 사용 측면에서 행정자료나 등록 자료를 충분히 사용할 수 없는 경우

#### ○ 자기기입식 조사

### 3-17. 조사표의 지침이나 안내정도

#### [작성 근거]

자기기입식 조사의 경우 응답자가 조사표를 충분히 이해할 수 있어야 하며, 이와 더불어 조사지침서를 제공하여 응답자에게 조사표 작성시 혼란을 방지해주어야 한다. 따라서 조사원 면접방식의 조사와는 달리 응답자가 충분히 이해할 수 있는 용어를 사용하여 작성된 조사지침서를 제공하는 것이 바람직하다.

#### ☞ 작성 방법

##### ① 매우 탁월함

➡ 조사지침서가 조사표에 근거하여 응답자가 이해하기 쉽도록 다양한 디자인을 적용하여 매우 자세하게 작성된 경우

##### ② 탁월함

➡ 조사지침서가 조사표에 근거하여 응답자가 이해하기 쉽도록 다양한 디자인을 적용하여 충분히 자세하게 작성된 경우

##### ③ 충분한 정보제공

➡ 조사지침서가 조사표를 작성하는데 적절하게 도움이 되는 경우

##### ④ 개선필요 ➡ 일부지침이 누락되거나 지시사항이 어긋나 개선이 필요한 경우

##### ⑤ 부적절함 ➡ 조사지침서가 조사표와 전혀 맞지 않아 지침서의 기능을 하지 못하는 경우

### 3-18. 무료전화의 이용여부

#### [작성 근거]

응답자가 조사표를 직접 작성할 경우 조사지침서로 충분한 정보를 얻을 수 없는 경우에는 응답자가 직접 조사관리팀에 전화를 걸어 필요한 정보를 얻을 수 있도록

무료전화를 이용하는 것이 바람직하다.

☞ 작성 방법

- ①예 ➡ 무료전화를 운영하는 경우(유료전화제외).
- ②아니오 ➡ 무료전화를 운영하지 않는 경우(유료전화만 운영).

**3-19. 문의에 답변할 팀의 구성여부**

[작성 근거]

응답자가 조사표를 직접 작성할 경우 조사지침서로 충분한 정보를 얻을 수 없는 경우에는 응답자의 의문사항을 해결할 수 있도록 전화에 응대할 수 있는 개별 팀을 운영하도록 해야 한다. 만일 현실적으로 개별 팀을 운영하기 어려운 경우에는 조사 관리팀이 직접 응답자와 접촉할 수 있도록 해야 한다.

☞ 작성 방법

- ①예 ➡ 조사관리팀을 제외한 개별적인 팀을 운영하는 경우
- ②아니오 ➡ 개별적인 팀을 운영하지 않은 경우

**3-20. 컴퓨터를 이용한 조사실시 여부**

[작성 근거]

자기기입식 조사에서 웹을 이용한 조사나 컴퓨터를 이용한 자기기입식 조사가 실시되는 지를 점검하는 문항이다. 조사원 면접방식(대면면접)에서 컴퓨터를 이용한 조사의 경우는 해당되지 않는다.

☞ 작성 방법

- ①예 ➡ 자기기입식 조사에서 컴퓨터를 이용한 조사를 실시하는 경우

②~아니오 ➔ 자기기입식 조사에서 컴퓨터를 이용한 조사를 실시하지 않는 경우  
(혼합조사 방식인 경우 45번 문항으로 이동)

### 3-21. 조사표의 일관성 점검여부

#### [작성 근거]

자기기입식 조사에서 프로그램화 된 조사표의 일관성 점검을 필수적으로 수행해야 한다. 종이조사표의 경우 문항별 논리적 순서가 맞는지를 점검하지만, 컴퓨터를 이용한 조사의 경우에는 모듈화된 조사표에서 특정 문항에 대해 논리적 오류가 발생하는지 여부를 점검하도록 한다.

#### ☞ 작성 방법

①~예 ➔ 자기기입식 조사에서 조사문항에 대한 논리적 흐름에 대한 일관성 점검을 수행한 경우

②~아니오 ➔ 일관성점검을 수행하지 않은 경우

### 3-22. 항목무응답 방지를 위한 개별 모드의 개발여부

#### [작성 근거]

자기기입식 조사에서 응답자가 특정문항에 대해 응답을 회피하여 항목무응답이 발생할 가능성이 있는 경우 이를 방지하기 위해 특별히 프로그램화 되어 있는지를 점검해야 한다. 반드시 응답을 해야 다음 문항으로 이동할 수 있도록 함으로서 항목무응답을 방지할 수 있다.

#### ☞ 작성 방법

①~예

➔ 특정 문항에 대해 무응답이 발생할 경우 다음 문항으로 이동할 수 없도록

프로그래밍 된 경우

②~아니오

- ➡ 특정 문항에 대해 무응답이 발생할 경우 이를 방지하기 위한 프로그래밍이 안 된 경우(무응답이 발생해도 다음 문항으로 이동이 가능하도록 한 경우)

### 3-23. 컴퓨터를 이용한 코딩 여부

[작성 근거]

완성된 조사표에 대해 부호화 작업을 컴퓨터를 이용하여 수행하는 경우 코딩오류를 줄일 수 있기 때문에 이에 대한 점검이 필요하다. 컴퓨터를 이용한 조사의 경우 직접적인 코딩작업이 이루어 질 수 있기 때문에 매우 효율적이다.

☞ 작성 방법

- ①~예 ➡ 완성된 조사표에 대해 컴퓨터를 이용한 코딩작업이 수행되는 경우
- ②~아니오 ➡ 수작업에 의해 코딩작업이 수행되는 경우

### ○ 컴퓨터를 이용한 전화조사(CATI)

### 3-24. 전화조사를 실시하는 장소(복수응답가능)

[작성 근거]

전화를 이용한 조사의 경우 조사를 수행하는 장소가 어디인지를 점검하는 문항이다. 통계작성기관내부에서 자체적으로 수행하는지, 별도의 전화조사 센터에서 수행하는지, 아니면, 외주를 통해 조사가 이루어지는지, 조사원의 재택근무를 허용하여 조사원자택에서 수행되는지를 파악한다.

☞ 작성 방법

- ① 회사내 ➡ 통계작성기관 내부의 전화조사팀에서 수행되는 경우
- ② 콜센터 ➡ 통계작성기관 내부의 콜센터에서 수행되는 경우
- ③ 외주 ➡ 통계작성기관에 의해 지정된 별도의 사업체를 통해 수행되는 경우
- ④ 조사원의 자택에서 전화조사를 허용함
  - ➡ 조사원의 재택근무를 허용하여 조사원자택에서 수행되는 경우

### 3-25. 현장조사를 위한 전화걸기 스케줄 프로그램여부

#### [작성 근거]

전화조사의 경우 응답자의 재택률에 의해 응답률이 크게 달라질 수 있기 때문에 재택시간을 검토하여 전화접촉을 위한 시간표가 작성되어 이에 따라 응답자를 접촉하여 조사가 수행될 수 있도록 해야 한다. 따라서 이와 같은 스케줄 프로그램의 여부를 점검하는 문항이다.

#### ☞ 작성 방법

- ① 예 ➡ 전화접촉을 위한 스케줄 프로그램을 운영하는 경우
- ② 아니오 ➡ 전화접촉을 위한 스케줄 프로그램을 운영하지 않고 그때그때 전화걸기를 수행하는 경우

### 3-26. 전화접속건수에 대한 통계산출여부

#### [작성 근거]

전화조사의 경우 최종 응답자 비율뿐만 아니라 응답을 얻기 위해 어느 정도 접촉을 시도했는지에 대한 통계산출을 함으로서 정확성을 판단할 수 있다. 총 접촉건수 중 최종 응답한 빈도수를 체크하여 응답률을 산정할 수 있기 때문이다.

#### ☞ 작성 방법

- ① 예 ➡ 응답률을 계산하기 위해 전체 전화접촉건수에 대한 통계를 산출할 경우
- ② 아니오 ➡ 전체 전화접촉건수에 대한 통계를 산출하지 않는 경우

### 3-27. 전화번호 추출방법

#### [작성 근거]

전화조사를 위해서는 조사대상자의 전화번호를 표본으로 추출해야 한다. 이를 위해 어떠한 추출틀을 사용했는지를 점검하는 문항이다. 전화번호 추출틀에서 추출한 경우, 또는 이전 조사로부터 구축된 전화번호자료로부터 추출한 경우, 또는 전화번호 발생기를 통해 임의로 전화번호를 발생시켜 전화번호를 추출하는 경우 등을 고려할 수 있다.

#### 작성 방법

- ① 추출틀에서 직접 추출
  - ➡ 전화번호부와 같은 전화번호 추출틀로부터 전화번호 표본을 직접 추출한 경우
- ② 이전에 수집된 자료로부터 추출
  - ➡ 이전 조사 자료에 있는 전화번호를 프레임으로 구성하여 이로부터 표본전화번호를 추출한 경우
- ③ 임의전화걸기(RDD)
  - ➡ 전화번호를 난수발생과 같은 방법으로 발생시켜 이를 표본전화번호로 고려한 경우
- ④ 다른 방법
  - ➡ 위의 방법 이외의 전화번호 추출방법을 사용한 경우

### 3-28. 조사원채용을 위한 특별한 절차여부

**[작성 근거]**

조사원은 자료를 수집하는 절차에서 매우 중요한 역할을 담당하며, 조사원에 대한 관리와 교육을 철저히 수행할 필요가 있다. 조사를 적절히 수행할 수 있는 조사원을 채용하고, 조사과정에서 조사원의 탈락 등의 문제가 발생할 경우 적절히 대응할 수 있는 체계를 수립하는 것이 바람직하다. 한편 컴퓨터를 이용한 조사를 수행할 경우 조사원의 컴퓨터 활용 능력을 적절히 평가하여 채용하고, 교육시킴으로서 조사과정에서 발생할 수 있는 문제를 사전에 차단할 필요가 있다.

**☞ 작성 방법****① CAPI, PAPI 또는 CATI실시를 위해 특별히 채용**

- ➡ 컴퓨터를 사용한 조사에 적합한 조사원을 별도의 테스트를 거쳐 채용한 경우

**② 조사원이 갖추어야할 개인적인 요구사항 목록을 보유함**

- ➡ 조사원 리스트를 통해 현 조사에 적합한 조사원을 채용한 경우

**③ 조사원이 갖추어야할 최소한의 자격요건 목록이 있음.**

- ➡ 특별히 현재 조사에 적합한 자격을 갖춘 조사원을 조사원 리스트로부터 채용한 경우

**④ 어떠한 최소한의 자격요건도 고려하지 않음**

- ➡ 조사원의 최소한의 자격만을 평가하여 조사원을 채용한 경우

**3-29. 조사원들의 조사 준비 방법****[작성 근거]**

조사원의 채용과정 뿐만 아니라 조사에 임하는 조사원에 대한 철저한 교육훈련이 반드시 필요하다. 응답자에 대한 조사원의 태도, 조사원의 조사표 이해력, 조사 참여를 유도할 수 있는 능력의 배양 등에 대해 교육을 통해 숙지하도록 해야 한다.

☞ 작성 방법

- ① 다양한 과정으로 구성된 집중교육, 최소의 준비교육(의사소통), 특별 교육과정, 잘 구성된 조사원 매뉴얼 제공함
- ② 한 과정에서 적절한 교육을 받고, 일부 조사원 매뉴얼 제공함
- ③ 응답거부 전환(무응답자 설득)기법과 조사 참여를 유도하는 방법에 대한 특별한 교육을 받음
- ④ 일부 훈련을 시키고 조사원 매뉴얼을 제공하지만 충분하지 않음
- ⑤ 기타(그 외 방법을 기입)

**3-30. 조사 성공에 대한 보너스 지급여부**

[작성 근거]

조사원들에 대한 조사독려를 위해 별도의 인센티브를 지급함으로써 조사의 질을 높일 수 있다. 따라서 조사원들에게 조사 성공에 대해 보너스의 지급여부를 점검하는 문항이다.

☞ 작성 방법

- ① 예 ➡ 일정 수준 이상의 조사를 완료한 조사원에 대해 인센티브를 지급하는 경우
- ② 아니오 ➡ 조사원에 대해 별도의 인센티브를 지급하지 않는 경우

**3-31. 조사원 채용방식**

[작성 근거]

조사에 참여하기 위해 조사원을 어떠한 방식으로 채용하는 지를 점검하는 문항으로 통계작성기관이 직접 조사원을 채용하는 경우와 지역적으로 지점망을 통해 조사

원을 채용하는 경우로 구분할 수 있다. 채용방식에 따라 장단점이 있으므로 적절히 보완적인 방법을 적용하도록 해야 한다.

#### ☞ 작성 방법

- ① 중앙채용방식 ➡ 통계작성 기관이나 조사수행기관이 직접 조사원을 채용하여 운영하는 경우
- ② 중앙채용방식이 아님 ➡ 통계작성기관 또는 조사수행기관의 지역별 지점에서 필요한 조사원을 채용하는 경우

### 3-32. 예비조사원 보유여부

#### [작성 근거]

조사원의 개인적인 사정이나, 사고 등의 유고로 인해 조사를 수행할 수 없는 경우 기존의 조사원을 대체할 예비조사원을 확보하고 있는지를 점검하는 문항이다. 또한 예비조사원이 즉시 조사에 참여할 수 있도록 별도의 교육훈련 과정을 이수토록 해야 할 것이다.

#### ☞ 작성 방법

- ① 조사원 각 그룹별로 특별히 교육받고 접촉하는 사람이 있음
- ② 정기적인 모임을 가짐
- ③ 조사원을 요구할 경우 예비조사원을 지원해줌
- ④ 예비조사원이 전혀 없음

### 3-33. 조사원 관리하는 방법

#### [작성 근거]

조사에 참여하고 있는 조사원들에 대한 적절한 관리체계가 수립되어야 한다. 이를

위해 주기적으로 조사지도원(슈퍼바이저)으로부터 조사원관련 사항을 피드백할 필요가 있으며, 만일 조사과정상에서 조사원들의 성과를 피드백 할 경우에는 전체 조사과정중 일부에 대해 보고형식으로 점검할 수 도 있다.

#### 작성 방법

- ① 주기적으로 피드백을 받음
  - ➡ 주간 또는 월간 단위로 조사관련 사항을 점검하는 경우
- ② 새로 채용된 조사원들은 조사 시 보통 2회 정도 점검을 받음
  - ➡ 기존의 조사원 보다 신규 조사원들에 대한 개별적으로 관리하는 경우
- ③ 의심스러운 사실이 있는 경우 특별 관리함
  - ➡ 일부 조사결과를 분석하여 이상이 있다고 판단되는 조사원에 대해 특별히 관리하는 경우
- ④ 질패한 경우, 마감시간 준수 및 검토결과 등 감안한 성과관리
  - ➡ 조사원의 조사 성공률, 조사일정의 준수, 조사결과의 검토 등을 통해 조사원을 관리하는 경우
- ⑤ 면접의 \_\_\_\_\_ %를 주기적으로 성과관리
  - ➡ 전체 면접량 중에서 주기적으로 일부분의 점검을 통해 조사원을 관리하는 경우
- ⑥ 지침 준수여부, 표본대상, 조사표 기입방법 등에 대한 기본적인 관리
  - ➡ 조사표상의 조사결과를 검토하여 조사원들을 관리하는 경우
- ⑦ 조사원에 대한 관리수단이 없음
  - ➡ 별도의 조사원 관리 체계가 없는 경우
- ⑧ 기타

### 3-34. 조사사실의 공지여부

[작성 근거]

조사를 수행하기에 앞서 응답자들에게 조사실시에 대해 우편, 이메일 또는 전화로 공지하는지를 점검하는 문항이다. 조사 참여를 적극적으로 독려하기 위해 조사지역의 공무원, 관련자 등에게 이를 공지함으로써 조사에 대한 거부감을 일정정도 해소 가능하기 때문이다.

☞ 작성 방법

- ① 예 ➡ 공문서, 이메일, 전화, 방송 등을 통해 조사실시에 관한 내용을 미리 공지한 경우  
 ② 아니오 ➡ 조사실시에 앞서 조사관련 사항을 미리 공지하지 않은 경우

### 3-35. 평균적으로 조사원 일인당 면접회수

[작성 근거]

조사원들의 조사부담과 조사일정, 조사비용등을 평가하기 위해 조사원이 하루 평균 얼마나 면접을 수행하는 가를 점검해야 한다. 너무 많은 면접회수를 부여할 경우 조사 자료에 대한 신뢰성이 문제가 될 수 있으며, 너무 적은 면접회수는 예산, 시간의 낭비가 될 수 있기 때문이다.

☞ 작성 방법

하루 평균 조사원 1인당 면접해야할 회수를 기입한다. 가구단위 조사에서는 1일 평균 몇 가구인지, 사업체조사의 경우 조사원 1인당 평균 몇 개 업체인지를 기록한다.

### 3-36. 조사원 일인이 실시하는 최대 면접회수

[작성 근거]

조사원이 전체적으로 최대 몇 번의 면접이 이루어지는 지를 평가하여 조사원의

업무부담을 평가해야 한다. 3-35 문항과 마찬가지로 조사원 부담을 적절히 평가하여 전체 조사관리가 되도록 하는 것이 바람직하다.

☞ 작성 방법

전체적으로 개별 조사원들이 면접을 완료한 최종 표본수를 계산하여, 이중 가장 큰 값을 기록한다.

### 3-37. 조사원 인적사항 인지여부

[작성 근거]

조사원관리를 위해 조사원의 인적사항을 관리하는 것이 필요하다. 조사원이 사고 또는 개인적인 문제로 인하여 조사가 불가능한 경우 이를 적절히 처리하기 위해 조사원의 인적사항을 파악하는 것이 중요하다.

☞ 작성 방법

- ① 예 ➡ 조사원의 인적사항을 관리하는 경우
- ② 아니오 ➡ 조사원의 인적사항을 관리하지 않는 경우

### 3-38. 조사원의 인구학적 특성이 응답패턴에 미치는 효과에 대한 평가

[작성 근거]

조사원의 인구학적 특성이 응답자의 응답패턴에 영향을 주는 지를 분석하여 만일 응답효과에 영향을 준다면, 이를 적절히 고려하여 조사원을 채용해야 할 것이다. 이는 조사원에 의한 비표본오차 발생의 원인이 되기 때문에 반드시 평가할 필요가 있다.

☞ 작성 방법

① 유의미한 영향을 줌

➡ 통계적 분석에 의해 통계적으로 응답패턴에 영향을 주는 것으로 판단되는 경우로서 조사원의 연령, 성별, 학력 등에 따라 응답자의 응답패턴에 영향을 받는 것으로 분석된 경우

② 약간의 영향을 줌

➡ 조사원의 연령, 성별, 학력 등에 따라 응답자의 응답패턴에 일부 영향을 받는 것으로 분석된 경우

③ 어떠한 영향도 주지 않음

➡ 통계적 분석에 의해 응답패턴에 영향을 주지 않는 것으로 판단되는 경우

④ 이러한 정보가 없음

➡ 조사원의 인구학적 특성에 따른 응답패턴의 영향에 관해 분석을 수행하지 않은 경우

**3-39. 면접가능시간대 파악여부**

[작성 근거]

조사원이 응답자를 방문했을 때 응답자의 부재, 장기출타 등으로 면접이 이루어지지 않을 수 있기 때문에 조사를 수행하기 전에 방문시간대를 미리 알림으로서 면접 성공률을 높일 수 있다. 즉, 조사원이 언제 가구나 사업체를 방문할 것인지를 전화나 이메일로 공지하고, 응답자로부터 면접이 가능한 시간대를 파악하여 방문함으로써 조사원과 응답자간의 신뢰감을 줄 수 있기 때문이다.

☞ 작성 방법

① 예 ➡ 조사원에 의한 면접시간대를 응답자로부터 파악한 경우

② 아니오 ➡ 조사원에 의한 면접시간대를 응답자로부터 파악하지 않은 경우

**3-40. 면접가능시간대에 대한 정보**

**[작성 근거]**

만일 면접가능시간대를 파악했다면, 조사주기 또는 조사기간에 따라 신축성 있게 이에 대한 정보를 적절히 이용할 수 있을 것이다.

**작성 방법**

① 월별 시간대

➡ 월별 조사, 또는 분기(반기)별 조사의 경우 매월 일정한 조사가능시간 대에 대한 정보를 파악한 경우(매월 말 오후1시 등)

② 주별 시간대

➡ 주간별 조사에서 면접가능한 시간대를 주별로 파악한 경우 (매주 화요일 오후 17시 이후, )

③ 지정한 날짜

➡ 일시적 조사에서 면접이 가능한 특정한 날짜에 면접가능 시간대로 파악한 경우(2007년 10월13일)

④ 대략적인 시간대

➡ 특정한 시간대 보다는 오전, 오후 등과 같이 대략적인 시간대로 파악한 경우

⑤ 정확한 시간대

➡ 특정한 시간대를 확실하게 지정하여 파악한 경우 (오후 1시~오후 3시 사이)

⑥ 기타

**3-41. 조사에 컴퓨터 활용여부**

**[작성 근거]**

조사원 면접방식 조사에서 컴퓨터를 이용한 조사가 실시되는 지를 점검하는 문항이다. 자기기입식 조사의 경우는 해당되지 않는다.

☞ 작성 방법

①예 ➡ 조사원 면접조사에서 컴퓨터를 이용한 조사를 실시하는 경우

②아니오 ➡ 조사원 면접 조사에서 컴퓨터를 이용한 조사를 실시하지 않는 경우  
(45번 문항으로 이동)

**3-42. 조사표의 일관성 점검여부**

[작성 근거]

조사원면접방식 조사에서 프로그램화 된 조사표의 일관성 점검을 필수적으로 수행해야 한다. 컴퓨터를 이용한 조사의 경우 모듈화 된 조사표에서 특정 문항에 대해 논리적 오류가 발생하는지 여부를 점검하도록 한다.

☞ 작성 방법

①예 ➡ 조사원면접방식 조사에서 조사문항에 대한 논리적 흐름에 대한 일관성 점검을 수행한 경우

②아니오 ➡ 일관성점검을 수행하지 않은 경우

**3-43. 항목무응답을 방지하기 위한 모듈의 개발여부**

[작성 근거]

조사원면접방식 조사에서 응답자가 특정문항에 대해 응답을 회피하여 항목무응답이 발생할 가능성이 있는 경우, 이를 방지하기 위해 특별히 프로그램화 되어 있는지를 점검해야 한다. 반드시 응답을 해야 다음 문항으로 이동할 수 있도록 함으로서 항목무응답을 방지할 수 있다.

☞ 작성 방법

①예

- ➡ 특정 문항에 대해 무응답이 발생할 경우 다음 문항으로 이동할 수 없도록 프로그래밍 된 경우

② 아니오

- ➡ 특정 문항에 대해 무응답이 발생할 경우 이를 방지하기 위한 프로그래밍이 안 된 경우(무응답이 발생해도 다음 문항으로 이동이 가능하도록 한 경우)

### 3-44. 컴퓨터를 이용한 코딩 여부

#### [작성 근거]

완성된 조사표에 대해 부호화 작업을 컴퓨터를 이용하여 수행하는 경우 코딩오류를 줄일 수 있기 때문에 이에 대한 점검이 필요하다. 컴퓨터를 이용한 조사의 경우 직접적인 코딩작업이 이루어 질 수 있기 때문에 매우 효율적이다.

#### ☞ 작성 방법

- ① 예 ➡ 완성된 조사표에 대해 컴퓨터를 이용한 코딩작업이 수행되는 경우
- ② 아니오 ➡ 수작업에 의해 코딩작업이 수행되는 경우

### 3-45. 무응답 방지를 위한 노력

#### [작성 근거]

조사과정에서 발생하는 무응답을 방지하기 위해 추가적인 조치로서 무응답자에 대한 추적조사를 실시할 수 있다. 이러한 노력을 어느 정도 수행하고 있는 지를 파악하고자 하는 문항이다.

#### ☞ 작성 방법

- ① 1회 추적조사를 특정 일 후에 실시
  - ➡ 특정 일 이후 추적조사를 1회 만 실시하는 경우(1차 방문만 허용)

- ② 2회 이상 추적조사를 몇 일과 몇 일 후에 실시
  - ➡ 정해진 날짜 사이에 2회 이상의 추적조사를 실시하는 경우(2차 이상의 방문을 허용한 경우)
- ③ 추적조사에 우선순위가 있는데 사업체조사 또는 그와 유사한 조사에서는 영향을 크게 미치는 단위를 우선적으로 실시함
  - ➡ 규모가 큰 사업체나 농가 등을 우선적으로 추적 조사하는 경우
- ④ 추적조사에 우선순위가 있는데 무응답 편향의 영향이 큰 지역의 단위를 우선적으로 실시함
  - ➡ 예를 들어 전국적인 조사에서 도시지역의 무응답편향이 추정에 영향을 주기 때문에 도시지역을 추적 조사하는데 우선순위를 부여한 경우
- ⑤ 응답자의 문의사항에 대한 해결을 위하여 표준 절차 및 기반을 갖추고 있음
  - ➡ 응답자의 문의사항에 근거하여 추적조사를 실시하는 경우
- ⑥ 응답과정을 과거조사의 점검 결과에 따라 설계
  - ➡ 이전의 조사로부터 나타난 응답패턴을 고려하여 추적조사체계를 수립한 경우
- ⑦ 기타

### 3-46. 컴퓨터 면접(CAPI, CATI 또는 CASI)에 필요한 사전 타당성 점검시스템이 있는지 여부

#### [작성 근거]

전통적인 조사 방법을 대신하여 컴퓨터를 이용한 조사방법을 적용하는 경우에 전통적인 조사를 대신하여 어느 정도 효율적인지에 대한 타당성을 점검하는 것이 바람직하다. 또한 사전 점검을 통해 조사과정에서 발생할 수 있는 문제점을 보완할 수 있다.

#### 작성 방법

- ① 예 ➔ 컴퓨터를 이용한 조사방식에 요구되는 다양한 필요 요소(인터넷, 노트북, 조사시스템, 프로그램, DB, 예산 등)를 사전에 점검한 경우
- ② 부분적으로 ➔ 모든 측면에서 타당성 점검을 수행하기 보다는 예산이나 조사효율성 측면 등만을 고려한 경우
- ③ 아니오 ➔ 전혀 필요요소를 점검하지 않고 도입한 경우

### 3-47. 컴퓨터에 의한 면접 방법과 전통적인 방법 간에 선택권

#### [작성 근거]

컴퓨터를 이용한 면접과 전통적인 종이조사표를 응답자가 선택하여 조사표를 작성할 수 있는 권한을 얼마나 부여하고 있는지를 파악하는 문항이다.

#### ☞ 작성 방법

- ① 일부 허용한 경우 어느 정도 까지 전통적인 조사표 작성방식을 선택할 수 있는지를 파악 (예: 약 15% 정도 허용함)
- ② 모름

### 3-48. 컴퓨터에 의한 면접(CAPI, CATI 또는 CASI)에 필요한 사전 타당성 점검시스템이 있다면 사무실에서 추가적으로 점검여부

#### [작성 근거]

사전 타당성 점검시스템으로부터 오류가 발생한 경우 추가적으로 시스템을 점검해야 한다. 이를 위해 시스템 전문가와 조사표관리자 등이 협업을 통해 완벽한 조사시스템을 구축하는 것이 필요하다.

#### ☞ 작성 방법

- ① 자료에 오차가 (거의) 없기 때문에 점검하지 않음

- ② 약간 오차가 있지만, 추가점검은 불필요함
- ③ 몇몇 오차를 수정하기 위해 추가점검 필요함
- ④ 오차를 수정하기 위해 추가점검이 필수적임

### 3-49. 타당하지 않은 응답내용을 응답자에게 확인 여부

#### [작성 근거]

에디팅 과정에서 조사원의 실수나 응답자의 거짓응답으로 논리흐름상 타당하지 않은 조사결과에 대해서 추가적으로 응답자에게 응답내용을 전화상으로 확인하는지 여부를 점검하는 항목이다.

#### ☞ 작성 방법

- ① 예 ➡ 에디팅 요원이 조사내용을 점검하여 타당하지 않은 사항에 대해 전화상으로 응답자에게 응답내용을 확인하는 경우
- ② 아니오 ➡ 전화상으로 타당하지 않은 응답내용을 점검하지 않는 경우

### 3-50. 향후 행정기록의 활용 가능성 점검여부

#### [작성 근거]

조사자료 이외에 다양한 측면을 고려했을 때 행정자료이용의 가능성을 점검했는지 여부에 대한 문항으로 조사의 타당성 점검과정에 검토할 필요가 있다.

#### ☞ 작성 방법

- ① 예 ➡ 행정자료 이용의 타당성 점검을 실시한 경우(사업체 등록자료, 주민등록자료이용 등에 대한 검토 여부)
- ② 아니오 ➡ 행정자료 이용의 타당성 점검을 실시하지 않은 경우

### 3-51. 행정자료나 등록 자료의 사용에 대한 제한 정도

#### [작성 근거]

조사자료 이외에 행정자료나 등록자료를 이용할 경우 법률적인 제한점이 있는지를 파악하는 문항으로 이용가능성 여부를 검토할 수 있도록 한다.

#### ☞ 작성 방법

- ① 법적인 제한이 전혀 없음
  - ➡ 행정자료나 등록자료의 사용에 전혀 법률적 제한이 없는 경우
- ② 법적인 제한이 다소 있지만, 사용가능함
  - ➡ 일부 자료이용에 대한 법률적 제한이 있어 관련 기관의 허용여부에 따라 가능한 경우
- ③ 법적인 제한이 매우 심함
  - ➡ 거의 사용이 불가능하지만, 법률적 고려에 의해 자료 관련자만 이용 가능한 경우
- ④ 조사 분야에서 완전히 금지함
  - ➡ 전혀 사용 불가능한 경우

### 3-52. 현장실사 기간이 조사대상 기간을 완전히 포함하고 있는지 여부

#### [작성 근거]

현장실사 기간과 조사대상기간이 같은 경우 응답자가 특정한 사건에 대해 응답하기가 용이하기 때문에 응답오차의 발생가능성이 낮아진다. 만일 그렇지 못하다면 어떤 이유에서 인지를 점검하는 것이 바람직하며, 조사원에게 이러한 사항을 인지하도록 할 필요가 있다.

#### ☞ 작성 방법

①예 ➔ 현장실사 기간과 조사대상기간이 같은 경우 또는 조사대상 기간에 실사 기간이 포함될 경우

②아니오 ➔ 현장실사 기간이 조사대상기간을 포괄하거나 동일하지 않은 경우

### 3-53. 현장실사 기간

#### [작성 근거]

현장 실사에 소요되는 기간은 조사기간이 얼마나 소요되는지를 파악하는 것으로 조사 설계 부문에 명시된 사항과 실제 소요된 기간과의 차이가 어느 정도 인지를 점검할 수 있다.

#### ☞ 작성 방법

➔ 현장실사에 소요되는 기간은 2007년 6월 1일부터 2007년 9월 31일 까지 약 4개월이다.

### 3-54. 조사기간 동안 휴가나 파업 또는 천재지변 등과 같이 예상치 못한 사건의 여부

#### [작성 근거]

조사기간에 휴가나 천재지변등과 같은 변동이 발생할 경우 조사 무응답률이 높게 나타날 가능성이 있으며, 사업체 조사에서 휴가나 파업 등의 기간이 포함될 경우도 응답을 얻어내기 어렵게 된다. 따라서 가능한 조사기간은 이러한 문제가 발생하지 않도록 결정하는 것이 바람직하다.

#### ☞ 작성 방법

①그런 영향은 없음 ➔ 휴가, 파업, 천재지변 등이 전혀 없음

②휴가기간이 있을 경우 얼마나 되는 지를 파악함.

③ 휴가 이외에 파업, 천재지변등과 같은 사건이 발생한 경우에 특정 사건을 기록

### 3-55. 조사원 중 슈퍼바이저와 같이 실사감독을 맡는 사람의 비율

#### [작성 근거]

조사지도원과 같은 조사현장에서 조사원을 관리하고, 조사표에 대한 전반적인 점검을 수행하는 사람의 비율이 어느 정도인 지를 파악하는 문항이다.

#### 작성 방법

➡ 전체 조사원 100명 중 슈퍼바이저는 20명으로 20%에 해당된다.

### 3-56. 조사과정의 문제점들에 대한 조치방법

#### [작성 근거]

조사과정에 다양하게 발생할 수 있는 문제를 조사원들이 어떠한 방법으로 문제점에 대한 조치를 수행하는 지에 대해 점검하는 문항이다. 조사원이 자의적으로 문제점을 해결하지 않도록 철저한 교육이 필요하다.

#### 작성 방법

- ① 즉시 연락할 수 있는 슈퍼바이저가 있음
- ② 슈퍼바이저가 지속적으로 점검
- ③ 오차의 빈도 및 원인에 대한 정보를 피드백하여 보고하고 있음
  - ③-1 조사관리자에게 보고
  - ③-2 주제관련 전문가에게 보고
  - ③-3 조사방법 전문가에게 보고
- ④ 참여자 의견수렴
- ⑤ 조사 시작부터 끝날 때까지 표본단위를 추적하는 전체적인 자료수집과정에 대

한 표본관리 절차가 있음

⑥ 기타(그 외 조치방법을 기술)

### 3-57. 사업체조사에서 확인 사항

#### [작성 근거]

가구단위 조사와는 다르게 사업체 조사에서는 담당자에게 조사실시사항을 공지하고 담당자가 응답에 필요한 자료를 준비할 수 있도록 배려하는 것이 바람직하다.

#### ☞ 작성 방법

① 조사대상 사업체의 관계자에게 연락

➡ 조사원이 사업체의 담당자에게 연락을 한 경우

② 가능한 쉽게 정보를 얻을 수 있는 적당한 시간에 연락

➡ 가능한 바쁜 업무시간이 언제인지를 파악하여 응답자가 여유를 가지고 응답을 할 수 있도록 배려한 경우

③ 응답자에게 편리한 방법 및 형태로 자료를 제공하도록 배려

➡ 반드시 면접에 의한 자료 수집을 가용하지 않고, 필요한 자료를 우편이나 메일 또는 fax로 제공할 수 있도록 배려한 경우

### 3-58. 입력 자료 처리과정 평가

#### [작성 근거]

입력 자료의 처리과정에서 처리오차가 발생하지 않도록 관리되고 있는지를 자체적으로 평가하는 문항으로 개선사항이 발생할 경우 즉각적인 조치를 통해 오차발생원인을 방지하는 것이 바람직하다.

#### ☞ 작성 방법

- ① 최우수 ➔ 입력오차가 거의 없는 경우
- ② 만 족 ➔ 입력오차가 1~5% 사이에서 발생하여 충분히 처리가 가능한 경우
- ③ 받아들일 만 하지만, 개선필요
  - ➔ 입력 자료의 평가시 약 5~10%미만에서 발생하며 에디팅과정에서 처리 가능한 경우
- ④ 받아들일 수 없음
  - ➔ 입력오차가 10%이상 발생하여 전반적으로 입력처리과정을 다시 고려해야할 경우

### 3-59. 입력과 제공되는 마이크로자료의 보안유지 정도

#### [작성 근거]

입력 또는 자료 제공 과정에서 마이크로자료에 대해 보안이 어느 정도 유지되고 있는 지를 점검하는 문항으로 개인정보나 사업체 정보가 누설되지 않도록 관리되어야 한다.

#### ☞ 작성 방법

- ① 전자파일은 암호화
  - ➔ 입력 또는 제공되는 파일에서 중요한 변수 값에 대해 암호화 하여 정보 접근이 안 되도록 하는 경우
- ② 들어오는 조사표에서 식별정보는 제외
  - ➔ 입력조사표에 대해 이름, 생년월일 등과 같은 개인 식별정보를 제외하고 입력하는 경우
- ③ 기타(그 외 내용을 기입)

### 4. 자료입력 및 자료처리

#### 4-1. 자료의 입력 방식

##### [작성 근거]

수집된 자료를 통계적 분석이 가능하도록 처리하는 과정이 반드시 필요하며, 이를 위해 자료를 입력을 위해 어떠한 방법으로 자료를 입력할 지를 검토해야 한다. 현재 컴퓨터 기술의 발달로 자동입력 방법이 보편화되고 있으나, 현실적으로 이를 이용하지 못하는 경우에는 수동 입력방법을 적용해야 한다.

##### ☞ 작성 방법

- ① 수작업으로 입력 ➡ 사람이 직접 컴퓨터에 자료를 입력하는 경우
- ② 전자적으로 입력
  - ②-1 문자인식(MCR)
    - ➡ 조사표에 기입된 문자를 자동으로 인식하는 방식으로 자료를 입력하는 경우
  - ②-2 지능적 문자인식(ICR)
    - ➡ 지능형 문자인식기를 통해 조사표에 기입된 문자를 자동으로 인식하는 방식으로 자료를 입력하는 경우
  - ②-3 광학문자인식(OCR)
    - ➡ OCR 카드에 기록된 자료를 전자적으로 입력하는 경우
  - ②-4 음성인식(VRE) ➡ 음성으로 기록된 자료를 입력하는 경우
  - ②-5 전자교환(EDI) ➡ 전자교환 방법으로 자료를 입력하는 경우
- ③ 자료수집과 동시에 입력됨
  - ➡ CAPI, CASI, CATI 등과 같이 컴퓨터를 이용한 조사의 경우 조사와 동시에 자료가 입력되는 경우
- ④ 기타

#### 4-2. 자료의 코딩 방식

**[작성 근거]**

자료를 입력하기 전에 컴퓨터가 자료를 인식할 수 있도록 조사결과를 부호화하는 작업이 코딩작업이다. 이러한 코딩작업을 어떠한 방법으로 수행하는지를 점검하는 문항이다.

**☞ 작성 방법**

- ① 수작업 ➔ 코딩작업을 사람이 손수 수행하는 경우
- ② 특별히 설계된 소프트웨어에 의한 자동코딩
  - ➔ 특정한 프로그램을 작성하여 컴퓨터에 의해 자동으로 코딩하는 경우
- ③ 기타방법

**4-3. 수작업 코딩시 일관성 점검 수단****[작성 근거]**

코딩작업을 수작업으로 수행할 경우 코드를 관리하거나 일관성 점검을 하는 수단이 있는지를 점검하는 문항으로 처리오차를 줄이기 위한 방법으로 일관성 점검은 필수적이라 할 수 있다. 다른 문항의 조건에 따라 코드에 대해 점검으로 하는 경우에는 종속적인 검증을 수행하는 경우이고, 다른 문항과 무관하게 개별 문항에 대해 각 코드에 대해 일관성 점검을 수할 수도 있다.

**☞ 작성 방법**

- ① 아니오 ➔ 일관성 점검을 위한 수단이 없는 경우
- ② 예 ➔ 관련 문헌을 검토하여 자료수집방법을 결정한 경우
  - ②-1 종속적인 점검 ➔ 다른 문항에 따라 코드에 대한 일관성 점검을 수행하는 경우
  - ②-2 독립적인 검증 ➔ 개별 문항별로 코드에 대해 일관성 점검을 수행하는 경우

#### 4-4. 원시자료 에디팅 필요성 평가

##### [작성 근거]

원시자료에 포함된 결측치, 부적합한 자료, 일관성이 결여된 자료 등에 대해 에디팅을 실시하도록 해야 한다. 이를 위해 원시자료의 품질을 평가하여 에디팅 작업이 필요한지를 검토하도록 해야 한다.

##### ☞ 작성 방법

- ① 원시자료에 오류가 많이 발견되었고 점검은 필수 불가결함
  - ➡ 원시자료의 점검결과 많은 오류가 발견되어 에디팅 작업이 반드시 필요한 경우
- ② 원시자료에서 수정되어야 할 몇 가지 오류가 발견되었음
  - ➡ 원시자료의 점검결과 일부 오류가 발견되어 수정할 필요가 있는 경우
- ③ 원시자료에서 몇 가지 오류가 발견되었으나 수정할 필요 없음
  - ➡ 원시자료에 오류가 있으나 그다지 중요한 사항이 아닌 관계로 에디팅 작업이 불필요한 경우(20번으로 이동)
- ④ 통합된 타당성 체크시스템으로 자료수집과정에서 이미 충분히 점검됨
  - ➡ 컴퓨터를 이용한 조사와 같이 이미 에디팅 점검이 되어 추가적인 에디팅 작업이 불필요한 경우(20번으로 이동)
- ⑤ 아직까지는 원시 자료에 오류가 없고 에디팅 과정도 불필요함
  - ➡ 자료의 입력과 동시에 지속적으로 에디팅 작업을 수행하는 경우(20번으로 이동)

#### 4-5. 에디팅 방법

##### [작성 근거]

자료 점검을 위해 사용되는 에디팅 방법으로는 컴퓨터를 이용한 자동 에디팅 과

정과 수작업으로 에디팅 하는 방법, 두 가지 방법을 혼합하여 수행하는 경우 등으로 구분할 수 있다. 자료의 일관성 점검을 위해 가능하면 자동화된 방법으로 에디팅을 수행하는 것이 바람직하다.

#### ☞ 작성 방법

- ① 자동화된 방법 ➡ 에디팅 프로그램으로 자료에 대해 자동적으로 에디팅 되는 경우
- ② 수작업 ➡ 오류가 발생한 개별 문항에 대해 수작업으로 에디팅을 수행하는 경우
- ③ 두 가지 혼합 ➡ 전체적으로는 자동화 방법을 적용하고, 일부 문항에 대해 수작업으로 에디팅하는 경우

#### 4-6. 채택한 에디팅 방법에 사용하는 자료

##### [작성 근거]

에디팅 작업은 무응답자에 대한 다른 변수들을 함께 고려하여 작업을 실시할 수 있으며, 다른 한편으로는 응답자와 유사한 변수들을 사용하여 결측치나 부적합한 자료에 대해 에디팅 작업을 수행할 수 있다. 또한 무응답자료와 응답자 자료를 모두 사용하여 에디팅 작업을 수행할 수 있다.

#### ☞ 작성 방법

- ① 무응답자 또는 다른 보조 자료에 대해 이용 가능한 자료만 이용함
  - ➡ 무응답자들의 다른 변수들을 비교하거나 그 외 보조 자료를 이용하여 에디팅 하는 경우
- ② 주어진 조사에 대해 다른 응답자들의 관측 자료만을 이용함
  - ➡ 에디팅 대상 문항에 대해 응답자들의 자료를 이용하여 에디팅을 수행하는 경우

③ 두 가지 자료를 혼합

➡ 무응답자와 응답자 자료를 모두 이용하여 에디팅을 수행하는 경우

④ 기타

#### 4-7. 에디팅 과정의 사전점검 여부

##### [작성 근거]

자료처리과정에서 부적합한 자료 또는 결측자료에 대해 에디팅 수행에 앞서 에디팅 작업 전반에 걸쳐 타당한 값으로 대체되는지를 반드시 점검할 필요가 있다.

##### ☞ 작성 방법

① 예 ➡ 에디팅 과정 전반에 대해 점검을 실시하는 경우

② 아니오 ➡ 에디팅 과정 전반에 대해 점검을 실시하지 않고 에디팅하는 경우

#### 4-8. 원시자료에 적용한 에디팅 방법(중복응답가능)

##### [작성 근거]

에디팅 적용 방법에 따라 범위 점검, 논리점검, 통계적 점검, 비교 점검 등을 고려할 수 있으며, 자료의 특성에 따라 다양한 방법을 동시에 고려할 수 있다.

##### ☞ 작성 방법

① 모든 변수 값에 대한 범위 점검

➡ 변수들이 취해야 할 범위를 점검하는 경우

(예: 성별의 경우 남=‘1’, 여=‘2’ 만 존재해야 하며, 점검과정에서 ‘3’의 값은 적합한 값으로 수정하는 경우임)

② 논리적 관계에 근거한 논리적 점검

➡ 문항간의 논리적 관계에 따라 응답여부가 결정되는 경우 논리적으로 종속되

어 있는 문항들 간에 논리흐름을 점검하는 경우

(예: 성별이 남='1'인 경우 여성 질환을 묻는 질문에는 응답할 수 없기 때문에 만일 이 질문에 응답한 남성이 있는 경우 논리적 점검을 통해 수정하는 경우임)

③ 이상치 분석 기법과 같은 통계적 검증

➡ 통계적인 분석을 통해 이상치로 판정되는 값에 대해 수정하는 경우

(예: 소득에 대한 분석을 통해 소득이 '100' 원으로 응답한 경우 이를 이상치로 판정하여 이 값을 수정한 경우임)

④ 논리적 점검 및 통계적 검증의 혼합 형식

⑤ 동일 통계자료의 과거 자료 또는 다른 자료간의 비교

➡ 종단면 조사와 같이 동일한 변수에 대해 과거 자료를 이용하여 점검하는 경우

⑥ 경험적인 점검 절차

➡ 과거 경험으로부터 특정한 변수 값에 대해 부적절한 값이 존재하면 이 값을 수정하는 경우

⑦ 기타(그 외 방법을 기입)

#### 4-9. 수작업 시 에디팅 지침서 제공여부

##### [작성 근거]

에디팅을 어떠한 기준으로 수행할 것인지를 밝혀주는 지침서는 반드시 필요하다. 조사표를 작성한 담당자와 에디팅을 수행하는 직원간의 긴밀한 협조가 필요하다. 수작업으로 에디팅을 수행하는 경우에는 반드시 명문화된 지침서가 구비되어야 처리 오차를 방지할 수 있다. 컴퓨터를 이용한 에디팅의 경우에는 점검을 위한 논리가 이미 프로그램화되어있기 때문에 별도의 문서화된 지침서는 필요하지 않다.

##### ☞ 작성 방법

① 정확하고 최근에 갱신된 지침서가 주어짐

➡ 에디팅 과정에 필요한 내용과 절차 및 에디팅 방법 등에 대해 자세히 수록된 지침서를 제공하는 경우

② 일부지침만 상세하게 제공됨

➡ 전체적인 지침서를 제공하기 보다는 일부 중요항목에 대해 자세한 내용을 지침서로 제공하는 경우

③ 아니오 ➡ 지침서를 제공하지 않음

**4-10. 결측치 또는 0 값의 구분**

**[작성 근거]**

에디팅과정에서 무응답, 비해당, 숫자 '0' 의 값을 구분하는 것은 매우 중요하다. 이는 분석과정에서 해당 값을 제외할 것인지 또는 포함할 것인지에 따라 통계값이 달라질 수 있기 때문에 이용자들을 위해 이들 값에 대해 가능한 명확하게 구분해줄 필요가 있다.

**작성 방법**

① 예 ➡ 결측은 '99' 또는 '999', 비해당 은 '0' 과 같이 명확한 개념의 구분이 되어 있는 경우

② 아니오 ➡ 위와 같은 구분이 전혀 되어 있지 않은 경우

**4-11. 에디팅 과정의 우선순위**

**[작성 근거]**

에디팅 과정은 먼저 변수별로 간단한 범위 점검으로 시작하여 변수간의 논리적 점검으로 전개하는 방법이 바람직하다. 이러한 절차에 의한 에디팅 절차는 처리오차를 줄일 수 있는 보다 효과적인 방법이기 때문이다. 특별히 ID 변수의 오류 점검은

그룹별로 자료를 분석하거나 다른 보조 자료와 결합할 때 반드시 필요한 절차이다.

#### ☞ 작성 방법

- ① 아니오 ➡ 순서와 무관하게 에디팅을 수행하는 경우
- ② 짐작한 오류는 보통 수작업으로 많은 노력을 들여 에디팅을 하고 대부분 자동으로 에디팅 되는 사소한 오류들은 신경을 덜 씀
- ③ ID로 사용되는 변수의 수정작업을 우선적으로 실시(만일 조사 자료가 가중치를 계산하거나 추정을 위한 보조 자료와 결합될 경우)
- ④ 기타

#### 4-12. 과도한 에디팅이나 오류방지를 위한 조치

##### [작성 근거]

조사표의 오류를 점검하기위한 에디팅 과정은 반드시 필요하지만, 필요이상으로 과도한 에디팅은 조사 자료의 왜곡을 가져올 수 있기 때문에 주의해야 한다. 따라서 에디팅을 위한 적절한 기준을 정하고, 필요한 경우 에디팅의 절대 수준 값을 지정하여 기준 수준이상은 에디팅을 하지 않도록 하는 것이 바람직하다.

#### ☞ 작성 방법

- ① 에디팅이 최종결과에 어떤 영향을 주는지를 분석하는 오류 인식 메카니즘이 있어 에디팅 최적점(Optimum of Editing)을 결정함
  - ➡ 에디팅의 영향을 분석할 수 있는 프로그램이 있어 이를 이용하여 최적의 에디팅을 구현할 수 있는 경우
- ② 수정된 자료에 대해서 다시 에디팅하는 과정이 있음
  - ➡ 오류나 결측에 의해 자료를 수정한 후 재차 에디팅 과정을 거치는 경우
- ③ 지나친 에디팅을 방지하는 아무런 절차가 없음
- ④ 기타(그 외 방법을 기술)

#### 4-13. 지속적인 에디팅 여부

##### [작성 근거]

일괄적으로 전체 조사 자료에 대한 에디팅을 수행하는 것이 아니라, 조사 자료의 일부가 취합되었을 때부터 에디팅 작업을 수행하는 지를 점검하는 문항이다. 이 경우 최초 에디팅이 적용된 자료의 나중에 에디팅 되는 자료간의 일관성 점검이 반드시 필요하다.

##### ☞ 작성 방법

- ① 예 ➡ 전체적으로 에디팅 작업을 수행하기 보다는 일부 조사 자료에 대해 지속적인 에디팅 작업을 수행하는 경우
- ② 아니오 ➡ 일괄적으로 에디팅을 실시하는 경우

#### 4-14. 에디팅 과정을 개선하기 위해 과거 오류에 대한 통계보유 여부

##### [작성 근거]

조사 자료의 에디팅 결과에 대한 통계를 산출하여 조사과정의 개선 및 조사표상의 오류발생원인, 에디팅 절차상의 오류 등을 향후 조사에서 반영할 수 있도록 해야 한다. 이를 위해 조사관리자는 과거 조사 자료의 오류발생 원인 등을 파악하여 문서화 하거나 자료를 저장하여 에디팅 과정에서 이용할 수 있도록 하는 것이 바람직하다.

##### ☞ 작성 방법

- ① 예 ➡ 주기적인 조사 또는 유사 통계조사에서 이전 자료의 오류발생 원인과 수정결과를 기록한 문서나 자료가 있는 경우
- ② 아니오 ➡ 과거 오류발생에 관한 통계자료를 보유하고 있지 않은 경우

#### 4-15. 오류에 대한 통계산출여부

##### [작성 근거]

조사 자료의 오류에 대한 통계를 산출하여 그 원인을 규명하고 적절한 조치를 취함으로써 향후 조사의 개선사항에 적극적으로 반영할 수 있도록 해야 한다. 그러한 오류의 원인은 조사표상의 구조적인 오류이거나, 조사원의 이해부족으로 인한 오류, 응답자의 오류 등 다양한 측면에서 발생할 수 있기 때문에 이러한 원인이 규명되면, 향후 조사에서 이러한 내용을 반영하여 오류발생을 사전에 방지할 수 있기 때문이다.

##### ☞ 작성 방법

- ① 예 ➡ 조사변수별 오류 발생 원인과 빈도, 전체적인 에디팅 비율등과 같은 통계를 산출하고 있는 경우
- ② 아니오 ➡ 오류발생과 관련된 통계를 산출하고 있지 않은 경우

#### 4-16. 에디팅 시스템의 신속성 평가

##### [작성 근거]

에디팅 과정을 엄격히 수행하는 것만이 좋은 것은 아니다. 에디팅 과정에서 미처 고려하지 못한 사항이 추후에 발견되면 이를 신속히 처리할 수 있도록 어느 정도 신속성을 가지도록 할 필요가 있다.

##### ☞ 작성 방법

- ① 매우 신속적 : 새로운 에디팅 과정이나 또는 수정된 에디팅 과정을 쉽게 적용가능  
➡ 에디팅 과정에서 새롭게 발견된 오류를 신속히 수정할 수 있도록 하는 경우
- ② 충분히 신속적 : 큰 문제없이 신규 또는 수정과정을 적용가능

- ➡ 에디팅 과정에서 새롭게 발견된 오류를 어느 정도 수정할 수 있도록 하는 경우

③~아님

- ➡ 에디팅 과정이 매우 경직되어 새로운 에디팅 과정을 적용하면 전체적으로 자료를 재점검해야 하는 경우

④~기타

**4-17. 오류원인에 대한 정보**

**[작성 근거]**

조사 자료의 원인을 파악할 수 있는지를 점검하는 문항으로 자료수집과정에서의 측정오류 또는 자료입력과정에서의 코딩오류인지를 파악하여 조사과정이나 자료처리절차를 개선하도록 해야 한다.

**작성 방법**

①~오류원인을 알 수 없음

②~오류의 원인이 일부 알려져 있음

- ➡ 자료의 오류원인을 전체적으로 알 수는 없지만 특정변수의 오류원인을 파악할 수 있는 경우

②-1 측정오류 ➡ 일부 변수가 조사표나 조사원의 면접과정에서 발생한 오류

②-2 코딩오류 ➡ 일부변수가 자료의 입력과정에 발생한 오류

②-3 기타 ➡ 그 밖의 오류 원인

③~오류의 원인을 알 수 있음

- ➡ 자료의 오류원인을 전체적으로 파악 가능한 경우

③-1 측정오류 ➡ 전반적으로 조사표나 조사원의 면접과정에서 발생한 오류

③-2 코딩오류 ➡ 전반적으로 자료의 입력과정에 발생한 오류

③-3 기타 ➡ 그 밖의 오류 원인

#### 4-18. 에디팅 과정에서의 오류 수정방법

##### [작성 근거]

에디팅 과정에서 발견된 오류를 수정하는 방법과 그에 대한 비율이 얼마나 되는지를 점검하는 문항이다. 에디팅 과정에서 발견된 오류는 먼저 응답자나 조사원을 재 접촉하여 응답값을 수정할 수 있으며, 비교적 중요하지 않은 변수의 경우에는 오류가 발생한 변수를 삭제하는 방법을 고려할 수 있다. 또한 통계적인 방법을 사용하여 대체를 하거나 경험적 대체값으로 대치할 수 있다. 점검을 위해 해당 방법과 전체 에디팅 작업 중 차지하는 비율(%)을 기술한다.

##### ☞ 작성 방법

- ① 조사원 또는 응답자를 재 접촉
  - ➡ 오류가 발생한 조사표나 해당 응답자를 파악하여 조사원 또는 응답자에게 전화를 걸어 정확한 값을 얻는 경우
- ② 오류변수를 없앴
  - ➡ 오류가 발생한 경우 중요성이 떨어지는 변수를 삭제하는 경우
- ③ 오류 또는 결측 변수는 좀더 적절한 값으로 대체
  - ➡ 경험상 또는 응답값을 이용하여 적절히 오류나 결측값을 수정하는 경우
- ④ 오류 또는 결측 변수는 대체 값으로 대체
  - ➡ 통계적인 방법으로 오류변수 또는 결측 변수값을 대체하는 경우
- ⑤ 잘 모름
- ⑥ 기타

#### 4-19. 통계값에 대한 대체여부

##### [작성 근거]

에디팅 과정에서 결측이나 오류로 인한 자료를 대체(imputation) 하는 지를 점검하는 문항으로 만일 대체를 실시하면 추정값에도 그에 대한 효과를 산출할 필요가 있다.

☞ 작성 방법

- ① 예 ➡ 다양한 대체 방법을 적용하여 항목 무응답이나 오류를 처리할 경우
- ② 아니오 ➡ 별도의 항목 무응답이나 오류에 대한 처리 방법을 도입하지 않는 경우  
(만일 자료에 대체를 실시하지 않는 경우에는 5장으로 이동)

4-20. 채택하고 있는 대체방법

[작성 근거]

대체방법이 자동화 되어있는지 아니면 수작업으로 대체를 수행하는 지를 점검하는 문항으로 대개의 경우 소규모 무응답에 대한 처리는 간단히 수작업으로 수행하고 일정 규모 이상의 무응답에 대해서는 프로그램을 이용하여 자동적으로 처리하는 것이 합리적이다.

☞ 작성 방법

① 자동화 대체

- ➡ 컴퓨터 프로그램을 이용하여 무응답 대체가 수행되는 경우(예: 미국의 센서스에서는 CONCOR 이라는 대체프로그램을 이용하고 있음1)

② 수작업 대체

- ➡ 조사관리자 또는 전문 에디터에 의해 소규모 항목에 대해 수작으로 대체를 수행하는 경우(예: 엑셀 등과 같은 간단한 소프트웨어를 이용하여 대체를 수

1) 조사통계연구회, 무응답오차, p.133 인용함,

Bureau of the Census, U.S.(1995), "CONCOR user's guide"

행하는 경우)

- ③ 두 방법을 혼합 ➡ 위의 두 가지 방법을 상황에 따라 혼합하여 이용하는 경우

#### 4-21. 채택하고 있는 대체방법에 이용하는 자료

##### [작성 근거]

대체과정에서 이용될 수 있는 자료는 다양하다. 먼저 응답자들의 동일 변수값을 이용할 수 있으며, 외부 자료를 통해 대체를 실시할 수도 있다. 이러한 측면에서 어떤 자료를 대체과정에 이용했는지를 점검하는 문항이다.

##### ☞ 작성 방법

- ① 무응답자에 대해 이용 가능한 자료 또는 다른 보조 자료를 이용함
  - ➡ 일부 항목에 무응답한 단위의 인구학적, 사회학적 특성을 가진 자료나 그 외 보조 자료를 통해 대체를 수행한 경우
- ② 무응답자에 대해 이용 가능한 자료 또는 다른 보조자료, 또는 조사의 다른 응답 단위의 관측 자료를 이용함
  - ➡ 일부 항목에 무응답한 단위의 인구학적, 사회학적 특성을 가진 자료나 그 외 보조자료 또는 해당 변수에 응답한 단위들의 관측 자료를 통해 대체를 수행한 경우
- ③ 주어진 조사의 다른 응답단위의 관측 자료만 이용함
  - ➡ 다른 자료는 이용하지 않고 단지 해당 변수에 응답한 단위들의 관측자료 만을 이용하여 대체를 수행한 경우
- ④ 기타

#### 4-22. 채택하고 있는 대체방법

##### [작성 근거]

단일 값으로 대체하는 방법 중에 특정항목에 응답한 응답자들의 값의 평균을 구하여 이를 대체값으로 사용하거나, 응답자들의 값을 이용하여 회귀분석 후 예측값을 산출하여 이를 대체값으로 사용하는 지를 점검하는 문항이다. 이론적으로 평균대체의 경우 추정값을 과소 추정하는 면이 있어 통상적으로 평균대체보다는 회귀대체를 선호한다.

#### 작성 방법

##### ① 평균대체와 같은 자동생성 방법

- ➡ 전체표본을 몇 개의 층으로 구분한 후 각 층에서 응답자 평균을 그 층에 속한 모든 결측값에 대해 대체하는 방법을 적용하는 경우

##### ② 회귀대체와 같은 예측 방법

- ➡ 무응답이 있는 항목  $y$ 에 응답이 있는  $y$ 의 보조변수  $x_1, \dots, x_k$ 를 회귀모형에 적합 시키는 방법으로  $\hat{y}_i = \hat{\beta}_0 + \sum_{k=1}^K \hat{\beta}_k x_{ki} + e_i$ 인 모형으로 예측하여 대체값으로 사용하는 경우

#### 4-23. 조사에 적용하고 있는 대체방법

##### [작성 근거]

통계작성 기관에서 조사 자료에 항목무응답이 발생한 경우 무응답 자료에 대해 적용한 대체방법을 점검해야 한다. 다양한 통계적인 대체방법을 적용한 경우 이에 대한 구체적인 내용을 명기해야 한다.

#### 작성 방법

##### ① 결정론적인 방법

###### ①-1 논리적 대체

- ➡ 조사표상의 논리적 근거를 이용하여 무응답항목을 대체하는 방법을 수행한

경우

①-2 연혁적 대체

- ➔ 전년도 자료를 이용하여 유사하거나 동일한 무응답자의 값을 대체하는 경우 (콜드택 대체라고도 함)

①-3 평균대체

- ➔ 응답자들의 평균값을 이용하여 무응답 값을 대체하는 경우

①-4 비율 및 회귀대체

- ➔ 회귀모형을 이용하여 무응답 값을 대체하는 경우

①-5 최근방 대체

- ➔ 응답 자료를 순서대로 나열하여 결측값이 있는 경우 그 결측값 바로이전의 응답값으로 대체하는 경우

② 확률적인 방법

②-1 핫택대체

- ➔ 대체층 내에서 대체값을 확률적으로 추출하여 대체하는 경우

②-2 확률잔차회귀 대체

- ➔ 회귀모형의 오차 항을 확률적으로 발생하여 계산된 예측값을 대체값으로 사용하는 경우

②-3 신경망대체

- ➔ 무응답자의 분류를 확률적인 신경망 이론을 적용하여 구성한 후 대체값을 발생하여 이를 사용하는 경우

②-4 확률잔차를 추가한 결정론적인 방법

- ➔ 회귀 대체에 확률잔차를 적용하여 대체값을 발생시킨 후 이를 대체값으로 이용하는 경우

4-24. 대체방법

[작성 근거]

대체 방법에는 크게 하나의 대체되는 값으로 무응답자의 값을 대체하는 단일대체와 여러 개의 대체그룹(donor)을 생성하여 이를 대체값으로 사용하는 다중 대체로 구분된다. 현실적으로 단일대체가 적용하기 쉬운 반면 통계값에 편향이 발생할 수 있기 때문에 다중대체가 선호된다.

#### 작성 방법

##### ① 단순대체

- ➡ 대체값을 하나의 유일한 값으로 대체하는 방법으로 평균대체, 최근방 대체, 콜드텍, 회귀대체, 핫텍 대체, 베이지안 대체 등을 이용한 경우

##### ② 다중대체

- ➡ 무응답자나 응답자가 랜덤하게 발생한다고 가정하여 결측값에 대해 여러 개의 값으로 대체한 경우

#### 4-25. 대체 알고리즘

##### [작성 근거]

결측값이나 오류값을 대체하기 위해 사용되는 대체알고리즘을 보유하고 있는지와 만일 보유하고 있는 알고리즘이 있다면 사용범위는 어느 정도인지를 점검하는 문항이다. 예를 들어 미국의 센서스 국에서 사용하는 CONCOR, 캐나다에서 사용하는 NIM2) 등이 대표적인 대체 프로그램이다.

#### 작성 방법

##### ① 모든 대체과정에 적용한 시스템

- ➡ 결측치에 대한 대체값을 적용하기 위해 모든 대체과정에 사용된 알고리즘을 기술한다.

2) Statistics Canada, 1997, 1996 Canadian Census Demographic Variables Imputation

② 대부분의 대체과정에 적용한 시스템

➡ 전체 대체과정은 아니지만 거의 모든 과정에 대체값을 적용하기 위해 사용된 알고리즘을 기술한다.

③ 부분적으로 대체과정에 적용한 시스템

➡ 일부 대체과정에 대체값을 적용하기 위해 사용된 알고리즘을 기술한다.

④ 어떠한 범용시스템도 적용하고 있지 않음

#### 4-26. 대체방법의 평가결과

##### [작성 근거]

대체방법에 대해 자체적으로 평가할 때 여러 방면에 적용 가능하도록 수립된 방법인지를 점검하는 문항으로 조사관리자의 전문적인 영역에 속하는 분야로서 이를 평가하기 위해서는 다양한 자료를 검토해야 한다.

##### 작성 방법

① 예 ➡ 현재 사용하고 있는 대체 방법이 다른 자료에도 적용 가능하여 비용, 정도 측면에서 다른 방법에 비해 월등히 좋은 경우

② 부분적으로 ➡ 일부 자료에 대해서는 현재 사용하고 있는 대체방법이 효율적이지만 전체적으로는 효율성을 입증할 수 없는 경우

③ 아니오 ➡ 전반적으로 현재 사용하고 있는 대체방법이 효율적이지 못한 경우

#### 4-27. 대체방법의 일관성 평가

##### [작성 근거]

대체된 자료의 분포가 원자료의 분포와 일관성을 유지한다는 측면은 매우 중요한 성질로서 대체에 의해 분포가 왜곡될 경우 대체의 효율성은 매우 낮게 된다. 또한 이러한 결과는 추정치를 매우 편향되게 하는 원인이 되기 때문에 일관성을 평가할

필요가 있다.

☞ 작성 방법

- ① 예 ➡ 대체된 자료와 원시자료 간의 일관성을 유지하고 있는 경우  
(예: 원시자료의 성별 분포와 대체된 자료의 성별 분포가 거의 일치하는 경우)
- ② 아니오 ➡ 대체된 자료와 원시자료 간의 일관성을 유지하지 못하는 경우  
(예: 원시자료의 성별 분포와 대체된 자료의 성별 분포가 서로 다르게 나타나는 경우)
- ③ 모름 ➡ 일관성을 평가할 수 없는 경우

4-28. 대체값에 대한 표식 여부

[작성 근거]

통상적으로 대체된 값은 원시자료와 구별을 위해 일정한 표식을 하는 것이 바람직하다. 대체된 자료는 원시자료와 다른 속성을 가지기 때문이다. 또한 대체에 의한 효과를 계산하기 위해서도 반드시 대체자료에 대한 표식이 필요하다.

☞ 작성 방법

- ① 예 ➡ 예를 들어 대체자료인 경우 “1” 원시자료인 경우 “2” 와 같이 자료상에서 구별을 위한 표식이 있는 경우
- ② 아니오 ➡ 원시자료와 대체자료간의 표식이 없는 경우

4-29. 대체과정에서 원시 자료 값 보존여부

[작성 근거]

대체된 자료와 원시자료는 동시에 보존되어야 한다. 만일 대체의 효과를 분석할

경우 원시자료를 이용해야 하며, 대체된 자료의 일관성 점검을 위해서도 반드시 원시자료는 별도로 보관하는 것이 바람직하다.

#### 작성 방법

① 예 ➡ 대체자료와 별도로 원시자료를 보존하고 있는 경우

② 아니오 ➡ 원시자료 대신 대체자료만 보관하고 있는 경우

### 4-30. 원시자료 대체비율(%)

#### [작성 근거]

원시자료 중에서 얼마나 많은 변수값이 대체되었는지를 평가함으로써 향후 대체에 대한 영향을 평가할 때 유용한 자료로 사용할 수 있다. 이는 품질지표 중에서 정확성을 평가하는 지표로 활용되기 때문에 반드시 대체율을 평가할 수 있도록 해야 한다.

#### 작성 방법

“총 자료수 중에서 특정한 변수에 대해 대체된 자료의 비율”로 계산한다.

### 4-31. 대체과정이 편향을 방지할 수 있다면, 이를 평가하는지 여부

#### [작성 근거]

대체과정으로부터 무응답 편향을 어느 정도 줄일 수 있다면 이에 대한 평가가 가능한지 여부를 점검하는 문항으로 원시자료를 이용하여 추정치를 계산하고, 대체자료를 이용하여 추정치를 계산한 결과를 서로 비교함으로써 가능하다.

#### 작성 방법

① 예 ➡ 편향의 감소가 발생한 경우

② 아니오 ➔ 편향의 감소가 발생하지 않는 경우

#### 4-32. 대체분산의 측정여부

##### [작성 근거]

대체자료는 속성상 원시자료와는 다른 속성을 가지므로 대체자료를 이용한 추정량의 분산에는 반드시 대체자료에 의한 분산을 계산해야 한다. 특별히 다중대체의 경우 SAS에서는 대체에 의한 분산을 계산해주고 있으므로 이에 대한 평가가 가능하다.

##### ☞ 작성 방법

##### ① 대체에 의한 분산

➔ 다중대체 또는 단일 대체에 의한 분산이 어느 정도인지를 계산하여 값을 기술한다. (예 : 전체분산=표본분산 + 대체분산)

##### ② 분산의 분석

➔ 대체 분산값을 계산하기 위해 사용한 분석기법을 기술한다.  
(예: 다중대체에 의한 분석)

##### ③ 대체분산측정 불가 사유

➔ 만일 대체분산을 측정할 수 없다면 그 이유를 기술한다.

##### ④ 대체분산측정을 위한 노력을 기울이지 않음

➔ 대체분산을 전혀 구하지 않은 경우

#### 4-33. 대체과정에서 사용된 소프트웨어의 신축성 평가

##### [작성 근거]

대체과정을 검토한 결과 일정부분 수정이 필요한 경우에 이를 쉽게 수정하여 대체과정을 수행할 수 있는지 여부를 점검하는 문항이다. 프로그램은 항상 신축성을

가져야 한다. 수행과정에 오류가 발생한 경우 수정이 용이해야하며, 다른 모듈 하에서도 실행될 수 있도록 해야 한다.

#### ☞ 작성 방법

- ① 예 ➔ 신축성 있게 프로그램을 구성한 경우 즉, 오류방생에 대해 수정이 용이하도록 프로그램으로 작성된 경우
- ② 부분적 ➔ 프로그램의 일부분에 대해 신축성을 가지도록 작성한 경우
- ③ 아니오 ➔ 전혀 수정이 불가능하도록 프로그램이 작성된 경우

### 5. 자료 분석 및 통계자료의 품질-자료의 품질진단

#### 5-1. 이용자 만족도 평가를 위한 방법

##### [작성 근거]

소관통계에 대한 이용자 만족도 조사를 통해 통계에 대한 이용자 요구사항을 반영하거나 개선할 수 있는 지를 점검한다. 정기적인 만족도 조사를 수행하거나 외부 이용자그룹으로부터의 자문을 통해 요구를 반영할 수 있다.

#### ☞ 작성 방법

- ① 정기적으로 조사에 대한 이용자만족도 조사를 수행
  - ➔ 소관통계의 이용자들에 대한 정기적인 만족도 조사를 수행하는 경우
- ② 이용자 피드백에 의한 체계적인 진단
  - ➔ 소관통계의 이용자들로부터 환류 되는 문제점 또는 개선사항들을 통해 평가하는 경우
- ③ 정기적인 이용자 자문을 통한 피드백
  - ➔ 정기적으로 이용자 그룹의 자문회의를 통한 소관통계에 대한 개선사항 등을 환류 하는 경우

④ 사용자만족도 지수 산출

➡ 소관통계에 대한 사용자 만족도 조사를 통해 만족도 지수를 산출하는 경우

⑤ 사용하지 않음

⑥ 기타(해당 사항을 기술)

**5-2. 주요 사용자 만족도에 대한 이용 가능한 정보의 평가정도**

**[작성 근거]**

주요 사용자 만족도 조사를 위해 이용 가능한 정보의 정도를 평가하는 문항으로서 사용자들에 대한 정보를 적절하게 관리하도록 해야 한다.

**작성 방법**

① 주요 사용자 및 다른 사용자에 대한 정보가 거의 없음

➡ 소관통계의 주요 사용자나 그 외 사용자들의 정보가 없는 경우

② 주요 사용자 정보는 일부 있지만, 다른 사용자 정보는 거의 없음

➡ 사용자 우선순위에 따라 주요 사용자그룹의 정보를 보유하고 있으나, 그 외 사용자 정보는 보유하고 있지 않은 경우

③ 주요 사용자 정보와 다른 사용자 정보가 일부 있음

➡ 사용자 우선순위에 따라 주요 사용자그룹의 정보와 그 외 사용자 정보는 보유하고 있는 경우

④ 주요 사용자 정보는 많이 있으나, 다른 사용자 정보 일부 있음

➡ 사용자 우선순위에 따라 주요 사용자그룹의 정보를 상당 부분 보유하고 있으나, 그 외 사용자 정보는 일부만 보유하고 있는 경우

⑤ 주요 사용자 및 다른 사용자에 대한 완전한 정보가 있음

➡ 사용자 우선순위에 따라 주요 사용자그룹의 정보와 그 외 사용자 정보를 상당부분 보유하고 있는 경우

### 5-3. 조사의 전반적인 관련성 평가정도

#### [작성 근거]

소관통계가 이용자들의 수요를 반영하고 있는 지를 평가하는 문항으로 보유하고 있는 이용자그룹의 의견을 적극적으로 반영할 수 있도록 해야 한다.

#### ☞ 작성 방법

##### ① 거의 관련 없음

➡ 이용자 그룹에 대한 어떠한 의견도 반영하지 않는 경우

##### ② 주요 분야는 약간 관련이 있고, 다른 분야는 거의 관련이 없음

➡ 소관통계 중 주요 통계에 대해서는 이용자 그룹의 수요에 대해 일부 의견을 반영하고 그 외 통계에는 이용자 의견을 반영하지 않는 경우

##### ③ 주요 분야 및 다른 분야 모두 약간은 관련성이 있음

➡ 소관통계 중 주요 통계와 그 외 분야 모두에 대해 이용자 그룹의 수요에 대한 일부 의견을 반영하고 있는 경우

##### ④ 주요 분야는 관련성이 높고, 다른 분야는 약간 관련성이 있음

➡ 소관통계 중 주요 통계분야에서는 이용자그룹의 의견을 적극적으로 반영하고 그 외 분야는 이용자 그룹의 수요에 대해 일부 의견을 반영하고 있는 경우

##### ⑤ 주요 분야 및 다른 분야 모두 동일하게 관련성이 높음

➡ 소관통계 중 주요 통계분야 및 그 외 분야에 대해 이용자 그룹의 수요를 적극적으로 반영하고 있는 경우

### 5-4. 추정값의 정도 측정을 위해 제시하고 있는 값

#### [작성 근거]

통계의 정확성을 측정하는 항목으로서 평균제곱오차, 변동계수, 신뢰구간 등을 적

절히 나타내줌으로서 이용자가 소관통계에 대한 신뢰성을 평가할 수 있도록 해야 한다.

#### 작성 방법

##### ① 평균제곱오차(MSE)

➡ 추정량의 평균제곱오차를 표시하는 경우

##### ② 변동계수(CV) 또는 상대표준오차(RSE)

➡ 추정량의 변동계수나 상대표준오차를 표시하는 경우

##### ③ 신뢰구간(CI)

➡ 추정량의 신뢰구간을 표시하는 경우

##### ④ 정성적인 정확도를 나타내는 문장뿐임

➡ 추정량에 대한 정확도를 단순한 문장으로 표현한 경우

##### ⑤ 어느 것도 제시하고 있지 않음

### 5-5. 평균제곱오차의 이용자욕구 충족여부

#### [작성 근거]

추정량의 평균 제곱오차를 제시한 경우 이용자들의 수요에 적합한 표현인지를 평가하는 항목으로, 만일 이용자들이 다른 측정값을 표시하도록 요구할 경우 이를 적극적으로 반영해야 한다.

#### 작성 방법

##### ① 아니오

➡ 평균제곱오차를 표현하는데 대해 이용자들이 개선사항을 요구한 경우

##### ② 예

➡ 평균제곱오차를 표현하는데 대해 이용자들이 별 다른 개선점을 제시하지 않은 경우

## 5-6. 주요변수에 대한 변동계수의 수준

### [작성 근거]

만일 주요 변수에 대해 정확성을 평가할 수 있도록 변동계수를 제시한 경우 변동계수의 수준에 대해 평가하는 항목으로 조사의 규모나 목적에 따라 다르게 나타날 수 있지만, 대체로 2~3% 이면 매우 만족스러운 수준이며, 3~5% 이면 적절히 받아들일 만 한 수준, 5~10% 이면 중간 수준, 10~20% 이면 변동의 크기가 큰 편이고, 20% 이상이면 매우 큰 편에 속한다.

### ☞ 작성 방법

- ① 모든 통계항목에 대한 변동계수가 매우 크다.
  - ➡ 모든 변수의 변동계수가 20% 이상인 경우
- ② 대부분 많은 항목에 대한 변동계수가 크다.
  - ➡ 대부분 변수의 변동계수가 10~20% 인 경우
- ③ 관련성 있는 항목과 주요 부문에 대한 변동계수는 중간수준이다.
  - ➡ 이용자들의 요구 변수와 주요항목의 변동계수가 5~10% 인 경우
- ④ 관련성 있는 항목 대부분과 주요 분류수준에서 변동계수가 작다.
  - ➡ 이용자들의 요구 변수와 주요항목의 변동계수가 3~5% 인 경우
- ⑤ 모든 주요항목과 모든 주요 분류수준에서 변동계수가 작다.
  - ➡ 모든 변수의 변동계수가 3% 이하인 경우

## 5-7. 변동계수의 수준이 조사목적에 대해 받아들일 만 한지 여부

### [작성 근거]

조사의 목적이 분류항목에 따른 통계의 산출인 경우 각 분류항목별 변동계수의 수준이 적절한지 또는 통합적인 통계의 산출에 대한 변동계수의 수준이 적절한지를

평가하는 항목으로 분류별 변동계수의 수준이 5~10% 이내의 경우 적절하다고 분수 있으며, 20% 이상이면 각 분류별 변동계수의 산출은 고려할 필요가 있다.

#### 작성 방법

- ① 예 ➔ 3~5%의 수준으로 분류별 통계산출이 가능한 경우  
 ② 아니오 ➔ 변동계수의 값이 20% 이상인 경우로서 분류별 통계산출을 재고할 필요가 있음.

### 5-8. 분석될 자료에 대한 추정량의 형태가 Horvitz-Thompson 추정량인지 여부

#### [작성 근거]

통계의 산출에 의한 추정량의 형태가 H-T 추정량인 경우 확률비례추출에 의해 가중치를 산출하고 이를 모수의 추정값에 반영하는 지를 점검하는 문항이다. 단순 집계에 의한 표본평균이나 표본 총합을 제시하는 경우에는 적용하지 않는 것이 일반적이다.

#### 작성 방법

##### ① 예

➔ 추정량의 형태가 가중치를 적용하여 제시된 경우( $\bar{y} = \sum_{k=1}^n w_k y_k / \sum_{k=1}^n w_k$ )

##### ② 아니오

➔ 추정량의 형태가 단순 평균의 형태로 제시된 경우( $\bar{y} = \sum_{k=1}^n y_k / n$ )

### 5-9. 최종적인 정확도를 얻기 위한 표본크기의 충분한지 여부

#### [작성 근거]

추정량의 분산이나 평균제곱오차 또는 변동계수로서 정확도를 평가할 수 있으며, 이러한 값들은 표본의 규모에 따라 크기가 달라지기 때문에 목표 오차에 부합되는 표본크기인지를 평가할 수 있다.

#### 작성 방법

- ① 예 ➡ 변동계수나 상대표준오차의 크기에 대해 표본규모가 적절한 경우로서 상대표준오차의 수준이 3~5% 이내가 되는 표본크기를 얻은 경우
- ② 아니오 ➡ 변동계수나 상대표준오차의 크기에 대해 표본규모가 적은 경우로서 최종표본크기가 상대표준오차 3~5% 이내에 놓일 수 없는 경우

### 5-10. 편향을 조정하기위해 사용된 보조정보

#### [작성 근거]

표본조사를 통해 발생하는 오차는 표본오차와 비표본오차로서 이러한 오차를 줄이기 위해 보조정보를 사용하게 된다. 따라서 보조정보를 사용한 목적이 무엇인지를 나타내는 문항이다.

#### 작성 방법

- ① 무응답편향을 조정하기위해
  - ➡ 단위무응답을 조정하기위해 보조정보를 재가중 방법에 사용한 경우
- ② 표본추출틀 상의 오차를 조정하기 위해
  - ➡ 표본오차를 줄이기 위해 보조정보를 재가중 방법에 사용한 경우
- ③ 측정오차를 조정하기위해
  - ➡ 조사표나 응답자들에 의해 발생한 오차를 조정하기위해 보조정보를 재가중 방법에 사용한 경우
- ④ 재가중 사용안함
  - ➡ 편향 축소를 위해 어떠한 재가중방법도 사용하지 않은 경우

⑤ 다른 형태의 오차를 조정하기 위해

#### 5-11. 재가중 방법을 사용한 경우 사용된 보정방법

##### [작성 근거]

재가중 방법은 추정량의 편향을 축소하기 위한 방법으로서 사후층화, 회귀추정방법, 비추정방법 등을 고려할 수 있다. 재가중 방법에 적용된 방법이 어떤 것인지를 점검하는 문항이다.

##### ☞ 작성 방법

##### ① 사후층화

➡ 모집단에 대한 정보를 이용하여 표본의 셀 값을 조정해주는 방법

##### ② 회귀추정

➡ 관심변수와 상관이 높은 모집단 보조변수( $X$ )를 사용하여 회귀추정에 의해 추정량의 효율성을 증대시키는 경우

##### ③ 비추정

➡ 관심변수와 상관이 높은 모집단 보조변수( $X$ )를 사용하여 비추정에 의해 추정량의 효율성을 증대시키는 경우

##### ④ 기타

#### 5-12. 모형보조추정방법을 사용한 경우 보조변수와 기준변수간의 상관관계측정 방법을 기술

##### [작성 근거]

관심변수와 상관관계가 큰 보조정보를 이용하여 추정치를 구할 경우 추정량의 효율성을 증대시킬 수 있다는 장점을 가진다. 이러한 관점에서 두 변수간의 상관관계가 어느 정도인지를 점검하는 것은 필수적이다.

## ☞ 작성 방법

관심변수와 보조변수간의 다변량 상관관계 또는 단순상관 계수로 표현할 수 있다.

## 5-13. 일반화된 추정 소프트웨어의 사용여부

## [작성 근거]

추정값의 산출에 일반화된 프로그램을 사용했는지 여부를 점검하는 문항으로 대체로 SAS, R, STATA, CALMAR 등 보편화된 프로그램을 사용할 수 있다.

## ☞ 작성 방법

- ① 예 ➔ SAS, R, STATA, CALMAR 등 보편화된 프로그램을 사용한 경우
- ② 아니오 ➔ 자체적으로 개발한 프로그램을 이용한 경우

## 5-14. 표본자료에 가중치의 정확도확보 여부

## [작성 근거]

만일 추정량에 가중치를 적용하여 가중추정량을 산정할 경우에는  $\sum_{i=1}^n w_i = N$  이 되도록 가중치의 정확도를 점검하는 것이 바람직하다. 이와 함께 가중치의 분포를 점검함으로써 과대 또는 과소 가중치의 존재 여부를 판단하여 적절히 처리하도록 해야 한다.

## ☞ 작성 방법

- ① 예 ➔ 가중치의 분포 등을 파악하여 과대 또는 과소 가중치를 적절히 조정하고, 가중치의 합이 모집단 수와 같은지를 점검하는 경우
- ② 아니오 ➔ 별도로 가중치에 대한 점검을 하지 않는 경우

### 5-15. 조사에서 단위 무응답의 범위

#### [작성 근거]

단위 무응답률이 다음과 같은 조사에서 어떤 기준에 속하는 지를 판정하여 평가하도록 한다. 의무적 조사는 법령 등에 의거하여 필수적으로 수행해야 하는 조사를 말하며, 단기 자발적 조사는 횡단면적 조사를 말하고, 수요에 의한 자발적 조사는 통계작성기관 이외의 외부 기관의 의뢰에 의해 실시되는 조사를 의미한다.

|              | 의무적조사  | 단기자발적조사 | 수요에 의한 자발적조사 |
|--------------|--------|---------|--------------|
| 낮은수준의 단위무응답률 | < 5%   | < 20%   | < 40%        |
| 중간수준의 단위무응답률 | 5%~15% | 20%~40% | 40%~60%      |
| 높은수준의 단위무응답률 | > 15%  | > 40%   | > 60%        |

#### ☞ 작성 방법

➔ 위의 기준표를 참조하여 해당하는 문항에 표시하도록 한다.

- ① 단위 무응답률을 모르거나, 허용할 수 없을 만큼 높음
- ② 높은 수준의 단위 무응답률
- ③ 중간수준의 단위 무응답률
- ④ 낮은 수준의 단위 무응답률
- ⑤ 거의 단위무응답이 없음

### 5-16. 무응답의 세부분류가 되어 있는지 여부

#### [작성 근거]

조사를 수행하는 과정에서 무응답의 발생은 필연적이며, 이러한 무응답을 세밀하게 분류함으로써 이용자들에게 보다 질 높은 조사에 관한 정보를 제공할 수 있게 된다. 무응답을 “거절”, “부재”, “접촉 불가능”, “장기출타”, “응답불능” 등과 같이 분류할 수 있다.

☞ 작성 방법

- ①~예 ➡ 무응답을 “거절”, “부재”, “접촉 불가능”, “장기출타” 등과 같이 분류한 경우
- ②~일부 ➡ 무응답을 세분화 하지 않고 특정 사유에 대해서만 분류한 경우
- ③~아니오

**5-17. 단위무응답 편향을 추정하기위해 이용 가능한 자료가 있는지 여부**

[작성 근거]

조사과정에서 발생하는 단위 무응답은 표본의 수를 감소시키기 때문에 추정값을 왜곡시킬 수 있다. 이를 방지하기 위해 이용 가능한 외부 자료나 보조변수 등을 이용하여 보정함으로써 무응답 편향을 줄일 수 있다.

☞ 작성 방법

- ①~외부 자료와 연계(예, 행정자료파일)
  - ➡ 조사 자료와 외부의 행정자료를 연계하여 무응답 단위에 대한 정보를 파악하여 무응답 편향을 추정한 경우
- ②~무응답자중 부차표본을 추출하는 일시적 조사를 통한 자료
  - ➡ 무응답자들 중에서 일정 규모의 부차표본을 추출하여 부차 표본의 값을 이용한 무응답 편향을 추정한 경우
- ③~비교를 위하여 사용된 무응답자의 배경정보 자료
  - ➡ 무응답자들과 응답자들 간의 특정한 패턴을 파악하기 위해 가구 또는 사업체의 배경정보를 이용하여 무응답 편향을 추정한 경우
- ④~무응답에 의한 편향을 추정하기위한 자료가 없음
- ⑤~기타(그 외 자료를 열거)

### 5-18. 조사의 항목무응답 정도

#### [작성 근거]

항목 무응답은 조사변수에 발생하는 무응답으로서 소득, 연령, 세금납부내역 등 개별항목별로 무응답이 존재할 수 있으며, 조사 자료의 신뢰성을 확보하고, 평가하기 위해 이러한 무응답률을 평가하는 것이 바람직하다.

#### ☞ 작성 방법

➔ 조사전체에서 발생한 항목 무응답률을 계산하여 다음의 기준 값에 해당하는 문항에 표시하도록 한다.

- ① 항목무응답을 모르거나 아주 높음(50%이상)
- ② 항목무응답이 높음(15~50%미만)
- ③ 항목무응답이 약간 있음(5~15%미만)
- ④ 항목무응답이 낮음(5%미만)
- ⑤ 항목무응답은 거의 없음

### 5-19. 조사대상기간과 결과의 발간시점 간의 시차를 날짜 수 또는 월수로 기록

#### [작성 근거]

조사대상기간과 조사결과의 발간시점 간의 시차를 검토함으로써 향후 소관통계의 시의성을 개선하기 위한 참고자료로 활용이 가능하다. 두 시점간의 차이가 크게 날수록 통계의 시의성은 낮아지는 결과를 초래하므로 가능한 시점간의 차이가 축소될 수 있도록 해야 한다.

#### ☞ 작성 방법

조사대상 기간과 조사결과의 발간시점간의 차이를 날짜 수 또는 개월 수로 계산하여 명시하도록 한다.

## 5-20. 계획된 발간 일자와 실제 발간일자 간의 시차를 날짜 수 또는 월수로 기록

### [작성 근거]

조사 기획단계에서 통계자료에 대한 예상 발간시점과 실제 발간시점 간의 차이를 명시함으로써 통계결과의 시의성을 개선할 수 있게 된다. 즉, 계획된 발간시점에 비해 너무 긴 시차를 두고 실제로 통계가 발간되었다면, 이를 반드시 개선해야 하며, 적절한 시차로 발간된 경우 시의성과 정시성을 확보하게 된다.

### ☞ 작성 방법

조사 기획단계에서 발간시점과 실제 발간시점간의 시차를 날짜 수 또는 개월 수로 계산하여 명시하도록 한다.

## 5-21. 조사대상 시점과 최종결과의 최초발간시기 간의 시차

### [작성 근거]

통계작성의 기준시점과 최종결과의 발간시점간의 차이를 명시함으로써 이용자들이 시의성 있는 통계결과를 사용할 수 있도록 해야 한다. 이를 위해 조사 대상 기간을 조사표상에 반드시 명시하고, 품질 보고서 작성 시 이를 명기하도록 한다.

### ☞ 작성 방법

#### ① 장당한 시차가 있음

➡ 조사대상 시점과 발간시점 간에 1년(12개월)이상의 차이가 있는 경우

#### ② 다소 긴 시차가 있음

➡ 조사대상 시점과 발간시점 간에 6개월~1년 미만의 차이가 있는 경우

#### ③ 정해진 시차가 있음(합의된 시차)

➡ 부처간의 협의에 의해 조사대상 시점과 발간시점 간 3개월~6개월 미만의 시

차가 있는 경우

④ 약간의 시차가 있음

➡ 조사대상 시점과 발간시점 간에 1개월~3개월 미만의 시차가 있는 경우

⑤ 매우 작은 시차가 있음

➡ 조사대상 시점과 발간시점 간에 1개월 미만의 시차가 있는 경우

## 5-22. 주요발간물의 발간시기 준수여부

### [작성 근거]

통계작성기관의 소관통계에 대한 주요 발간물의 발간시기의 준수여부를 평가하는 항목으로 대부분의 통계작성기관에서는 소관통계에 대해 정기적으로 발간물을 공표하고 있으며 통계의 이용자 만족도를 향상시키기 위해 정시성을 준수할 수 있도록 해야 한다.

### 작성 방법

① 자주 많이 지연됨

➡ 모든 주요발간물의 발간시기를 준수하지 못하는 경우

② 어느 정도 지연으로 자주 맞지 않음

➡ 주요발간물의 발간시기를 1~2개월 미만 지연하며, 발간시기를 자주 지키지 못하는 경우

③ 발간물의 50% 이상 주로 지킴

➡ 주요발간물의 50%이상은 발간 시기를 적절히 준수하고 나머지는 잘 지키지 못하는 경우

④ 발간물의 80% 이상 보통 지킴

➡ 주요발간물의 80%이상은 발간 시기를 적절히 준수하고 나머지는 잘 지키지 못하는 경우

⑤ 거의 언제나 지킴

- ➡ 모든 주요발간물에 대한 발간시기를 준수하는 경우

### 5-23. 시의성 개선을 위한 변화의 필요성

#### [작성 근거]

통계작성기관에서 생산되는 통계를 이용자가 적절한 시점에 이용할 수 있도록 하는 것이 필요하며, 만일 계획된 발간 시점을 준수하지 못할 경우 통계작성 기관 및 협력기관에서는 이를 준수하기 위한 적절한 개선이 요구된다.

#### ☞ 작성 방법

##### ① 자소한 내부변화

- ➡ 통계작성 기관 내부의 업무절차의 개선으로 충분히 시의성을 준수할 수 있는 경우

##### ② 중대한 내부변화

- ➡ 통계작성 기관 내부의 통계생산 과정 전체의 개선이 필요한 경우

##### ③ 외부협력기관의 관련된 변화

- ➡ 통계작성 기관 내부문제로 인한 발간시기가 지연되기 보다는 외부 협력기관의 문제에 의해 발간시기가 지연되어 해당 기관의 개선이 필요한 경우로서 주로 조사기관의 조사 지연 또는 통계작성 지연 등으로 인한 지연이 많은 경우를 차지함.

##### ④ 기타

### 5-24. 다른 통계의 개념과 조사 설계에 대해 지역에 대한 통계결과 간의 비교가능성

#### [작성 근거]

소관통계와 다른 통계간의 연계를 통해 지리적 비교가능성을 평가하는 항목으로

서 만일 비교가능 통계가 없는 경우에는 생략가능하다.

#### 작성 방법

##### ① 연관된 통계와 심각한 차이가 있음

- ➡ 통계작성 기관의 소관통계와 다른 통계간의 연계 결과 매우 다른 결과가 나타나 비교 불가능한 경우(기준시점의 차이나 통계 정의의 차이로 인한 불일치)

##### ② 연관된 통계와 다소 큰 차이가 있음

- ➡ 통계작성 기관의 소관통계와 다른 통계간의 연계 결과 특정 부분에서 큰 차이가 나타나는 경우로 해당 부분의 해석에서 반드시 주의사항을 명기할 필요가 있는 경우(조사방법이나 조사대상의 차이로 인한 불일치)

##### ③ 연관된 통계와 약간의 차이가 있음

- ➡ 통계작성 기관의 소관통계와 다른 통계간의 연계 결과 극히 일부분에서 차이가 발생하는 경우(비교대상통계의 특정 항목에서의 약간의 불일치)

##### ④ 연관된 통계와 다소 차이가 있음

- ➡ 통계작성 기관의 소관통계와 다른 통계간의 연계 결과 특정 항목에 대해 차이가 나타나는 경우(비교대상통계의 특정 항목에서의 다소의 불일치)

##### ⑤ 연관된 통계와 거의 차이가 없음

- ➡ 통계작성 기관의 소관통계와 다른 통계간의 연계 결과 전체적으로 차이가 나타나지 않는 경우

#### 5-25. 조사기간에 대해 시계열적인 통계 결과물의 비가능성

##### [작성 근거]

시계열적인 비가능성은 조사의 개념과 조사 설계상의 변화로 한정하여 비교가능한 지를 평가할 수 있다.

☞ 작성 방법

① 지계열 비교가 안 됨

- ➡ 통계작성 기관의 소관통계와 다른 통계간의 조사개념과 조사 설계상의 차이로 인해 전체적으로 시계열 비교를 할 수 없는 경우

② 지계열적인 비교가능성에 심각하게 제한 적임

- ➡ 통계작성 기관의 소관통계와 다른 통계간의 조사개념과 일부 조사 설계상의 차이로 인해 거의 시계열 비교를 할 수 없는 경우

③ 지계열적인 비교가능성에 어느 정도 제한 적임

- ➡ 통계작성 기관의 소관통계와 다른 통계간의 특정 항목의 조사개념과 조사 설계상의 차이로 인해 특정 항목간의 시계열 비교가 어려운 경우

④ 지계열적인 비교가능성에 약간 제한 적임

- ➡ 통계작성 기관의 소관통계와 다른 통계간의 조사개념과 조사 설계상에의 약간의 차이로 인해 일부 제한이 있는 경우

⑤ 지계열적인 비교가능성에 거의 제한이 없음

- ➡ 통계작성 기관의 소관통계와 다른 통계간의 조사개념과 조사 설계상에 차이가 없어 시계열 비교를 하는데 별 문제가 없는 경우

5-26. 월별 연간 단위로 동시에 다른 주기로 조사하는지 여부

[작성 근거]

정기적인 조사의 경우 월별 또는 연간 단위로 다른 주기로 조사할 경우 일관성 점검이 요구된다. 일관성이 결여된 통계 결과를 사용할 경우 이에 대한 효과를 반드시 명기해야 한다. 만일 다른 주기에 조사하지 않는 경우에는 일관성 점검은 생략한다.

☞ 작성 방법

- ① 예 ➡ 월별 또는 연간 단위로 다른 주기로 조사할 경우

② 아니오 ➔ 월별 또는 연간 단위로 다른 주기로 조사하지 않는 경우

### 5-27. 서로 다른 주기에 대한 결과에 대해 통계의 일관성 평가 및 동일한 조사대상 기간에 대해 다른 조사주기결과와 결합가능성

#### [작성 근거]

만일 월별 단위 통계와 연간 단위 통계를 결합할 경우 이들 통계를 결합할 수 있는지를 평가하는 문항이다. 즉 연간 통계의 경우 월별 통계가 생산되지 않기 때문에 이들 간의 결합은 가능하지만, 두 통계간의 일관성이 결여될 경우 두 통계간의 결합 여부를 신중히 검토해야 한다.

#### ☞ 작성 방법

##### ① 장당한 문제가 있음

➔ 월별 통계와 연간 통계간의 심각한 차이로 인해 연계가 불가능한 경우

##### ② 다수의 상반된 점이 있음

➔ 월별 통계와 연간 통계간의 다수의 연계항목에 대해 상반된 결과가 발생하는 경우

##### ③ 일부 상반된 점이 있음

➔ 월별 통계와 연간 통계간의 일부의 연계항목에 대해 상반된 결과가 발생하는 경우

##### ④ 극히 적은 문제

➔ 월별 통계와 연간 통계간의 연계에 매우 사소한 문제가 있는 경우

##### ⑤ 심각한 문제없음

➔ 월별 통계와 연간 통계간의 연계에 아무런 문제가 없는 경우

### 5-28. 동일한 사회-경제 분야에 대해 통계의 일관성 평가 및 비공식 자료와의 결합 가능성

**[작성 근거]**

서로 다른 유형의 통계가 거의 같은 방법으로 측정되어 동일한 사회현상을 반영하는 측도로 이용되는 경우에 이들 간의 일관성을 평가하는 것이 바람직하다. 즉 통계작성기관에서 생산되는 소관통계와 다른 기관에서 생산되는 동일한 형태의 통계 간에 일관성을 가지는지를 측정하여 두 통계간의 차이점 여부를 파악하는 것이 필요하다. 만일 통계들 간의 차이가 있다면 그러한 차이가 왜 발생했는지를 평가하도록 한다.

**☞ 작성 방법****① 장당한 차이가 있음**

➡ 서로 다른 유형의 통계 간에 매우 큰 차이가 발생한 경우

**② 다수의 차이가 있음**

➡ 서로 다른 유형의 통계 간에 여러 항목에서 차이가 발생한 경우

**③ 일부 차이점이 있음**

➡ 서로 다른 유형의 통계 간에 일부 항목에서 차이가 발생한 경우

**④ 극히 적은 차이**

➡ 서로 다른 유형의 통계 간에 매우 유사한 값을 가지는 경우

**⑤ 짐작한 문제없음**

➡ 서로 다른 유형의 통계간의 일관성 평가에서 별 문제가 없는 경우

**5-29. 이용자 맞춤분석에 대응할 능력의 평가****[작성 근거]**

통계작성기관의 소관통계에 대한 이용자 요구에 얼마나 신속하게 대응할 수 있는지를 평가하는 항목으로 전문성과 유연성, 신속성을 갖도록 해야 한다.

☞ 작성 방법

- ① 요구하는 통계와 관계된 거의 모든 분석에 대한 능력과 전문성을 갖추고 있기 때문에 이용자에게 맞춤분석을 아주 신속하고 유연하게 대응할 수 있다.
- ② 요구하는 통계와 관계된 대부분의 분석에 대한 능력과 전문성을 갖추고 있기 때문에 대부분의 경우 맞춤분석 요구에 대응할 수 있다.
- ③ 요구하는 통계와 관계된 분석능력과 전문성이 없기 때문에 유감스럽게도 대부분 맞춤분석 요구에 대응할 수 없다.
- ④ 자료의 분석에 대한 이용자들의 요구를 만족시킬 의무가 없다.

5-30. 사용하고 있는 추가분석 기법

[작성 근거]

통계자료의 생산과 더불어 통계작성 기관에서는 생산된 자료를 이용하여 추가적인 작업이 이루어지는 지를 평가하는 항목이다. 통계의 부가가치를 높이기 위해서는 2차적인 추가 분석 결과의 생산이 무엇보다도 중요하다.

☞ 작성 방법

- ① 모델링(모형화 작업)
  - ➡ 자료를 이용한 특정 현상을 검증하기 위한 모형을 구축하여 작업하는 경우
- ② 공간분석
  - ➡ 자료를 이용하여 공간 분석을 수행하는 경우
- ③ 지수산출
  - ➡ 자료를 이용하여 두 시점 간의 비교를 용이하게 하기 위한 지수를 산출하는 경우
- ④ 계절 또는 추세 분석과 추정
  - ➡ 자료를 이용하여 계절적인 영향이나 시계열적인 추세를 분석하는 경우
- ⑤ 기타

### 5-31. 작년에 새롭게 도입한 분석기법이나 개편한 적이 있는지 여부

#### [작성 근거]

통계산출의 고급화 또는 문제점을 개선하기 위해 신기술 또는 개편작업이 이루어졌는지를 점검하는 문항으로 이러한 개선 작업으로 소관통계의 품질 향상에 기여할 수 있다.

#### 작성 방법

- ① 예 ➡ 작년에 새롭게 도입한 분석기법이나 개편한 적이 있는 경우
- ② 아니오 ➡ 작년에 새롭게 도입한 분석기법이나 개편한 적이 없는 경우

### 5-32. 이용자들의 자료접근을 위해 보유하고 있는 자료의 형태

#### [작성 근거]

통계이용자들이 접근 가능한 통계자료의 형태를 파악하는 문항으로 원시자료부터 가공된 테이블 형식의 자료까지 해당 되는 문항에 대해 모두 검토하도록 한다.

#### 작성 방법

- ① 빈도표
  - ➡ 자료를 가공한 1차원, 2차원 또는 다차원 빈도표를 제공하는 경우
- ② 다량의 통계표
  - ➡ 자료를 가공한 다양한 통계표를 제공하는 경우
- ③ 마이크로자료-연구자용 파일
  - ➡ 연구자를 위해 마이크로 자료를 제공하는 경우
- ④ 마이크로자료-일반이용자파일
  - ➡ 일반 이용자를 위해 마이크로 자료를 제공하는 경우

### 5-33. 민감한 셀에 대한 정의

#### [작성 근거]

통계자료를 공표할 때 개인 또는 특정 업체의 비밀이 누출되어서는 안되기 때문에 이에 대한 각별한 주의가 필요하다. 이를 위해 특정한 기준을 마련하여 비밀이 누설되지 않도록 해야 하며 만일 민감한 셀에 발견될 경우 이를 보호하도록 하는 조치가 필요하다.

#### ☞ 작성 방법

- ① 한 개의 셀에 해당하는 최소 응답자수는 다음 값을 초과하도록 하고 있음.  
이때, 최소 응답자수는 \_\_\_\_\_ 이다.
- ② 셀에서 응답자의 해당 값의 분포에 대해 현저한 집중도를 측정함.
  - ②-1 C-times 원칙: 2순위 응답자가 가장 큰 단위의 유용한 추정치를 얻지 못하도록 보장함.
  - ②-2 순서 통계들의 선형결합인  $(n, k)$  원칙: 가장 큰  $n$ 명의 응답자가 총계 셀 값의  $k\%$  이상 될 경우 그 셀은 민감하다고 봄.
  - ②-3 0의 빈도수를 가지는 특정한 수준이하 또는 이상의 응답자에 대한 정보를 발표하지 않음, 예를 들어 특정 산업의 사업체는 정해진 값 이상의 수익을 결코 초과하지 못함.

### 5-34. 민감한 셀을 보호하기 위해 보조셀 자료의 삭제 필요성

#### [작성 근거]

민감한 셀이 어떤 보조 자료에 의해 유추되어 나타나는 경우에는 보조 자료의 셀을 삭제하는 것이 바람직하다. 만일 보조셀의 삭제가 필요한 경우에는 이에 대한 사

항을 반드시 기술하여 이용자에게 혼란을 주지 않도록 해야 한다.

☞ 작성 방법

- ① 예 ➡ 보조셀의 삭제가 필요한 경우
- ② 아니오 ➡ 보조셀의 삭제가 필요하지 않은 경우

**5-35. 셀을 보호하기 위한 기존의 소프트웨어 사용여부**

[작성 근거]

셀 보호를 위해 사용하고 있는 프로그램의 보유 여부를 점검하는 문항으로 통계 생산기관에서는 기술적으로 이러한 프로그램을 개발하거나 또는 사용하여 특정 셀을 보호할 필요가 있다.

☞ 작성 방법

- ① 예 ➡ 셀 보호를 위해 사용하고 있는 프로그램을 보유하고 있는 경우
- ② 아니오 ➡ 셀 보호를 위해 사용하고 있는 프로그램을 보유하고 있지 않은 경우

**6. 문서화 및 자료의 공표**

**6-1. 행정자료의 사용 시 관련 문서 보유여부**

[작성 근거]

조사자료 이외에 행정자료를 사용하여 통계를 산출한 경우 해당 자료와 관련된 문서를 보유하고 있는지를 점검하는 문항이다. 문서화 과정에서 이용된 자료의 출처를 명시함으로써 이용자의 편의를 도모할 수 있기 때문이다.

☞ 작성 방법

① 예 ➔ 행정자료의 출처와 자료의 수집방법, 기간 등이 명시된 문서를 보유하고 있는 경우

② 아니오 ➔ 행정자료와 관련된 문서를 보유하고 있지 않은 경우

## 6-2. 통계산출과정의 문서화에 포함된 내용

### [작성 근거]

통계산출과정에 대한 사항은 이용자들의 이해를 돕기 위해 가능한 한 상세한 내용을 문서에 포함시키도록 해야 하며, 이러한 사항들을 점검하기 위한 문항이다. 이용자 요구진단, 조사표 사본, 조사변수에 대한 상세한 사항, 조사방법 및 기법, 조사설계의 변경사항, 원시자료부터 자료의 공표 내용과 이를 저장하고 있는 파일의 내용등과 품질보고서의 내용 등이 포함되는 것이 바람직하다.

### ☞ 작성 방법

통계 발간물에 포함된 각 사항을 검토하여 체크리스트상의 항목에 체크하도록 한다.

## 6-3. 메타자료와 품질지표의 공표수단

### [작성 근거]

생산된 통계의 정보를 제공하기 위해 메타자료와 품질지표를 어떠한 방법으로 제공하는지를 점검하는 문항이다.

### ☞ 작성 방법

① 종이 발간물상의 메타자료

➔ 보고서 형식 또는 책자형식으로 메타자료의 정보를 제공하는 경우

② 인터넷상의 메타자료

➡ 인터넷상으로 메타자료의 정보를 다운받게 하거나 검색이 가능하도록 한 경우

③ 종이발간물 상의 품질지표

➡ 보고서 형식 또는 책자형식으로 품질지표에 관한 정보를 제공하는 경우

④ 인터넷상의 품질지표

➡ 인터넷상으로 품질지표정보를 다운받게 하거나 검색이 가능하도록 한 경우

**6-4. 무응답추적과정, 조사원교육을 위한 문서등과 같이 품질 척도에 대한 문서 보유여부**

**[작성 근거]**

자료수집과정에서 무응답이 발생한 경우 이에 대한 추적절차나 조사원에 대한 교육지침서등을 보유하고 있는지를 점검하는 문항으로 일반적으로 조사원 교육과정에 다른 사항들을 자료의 품질보증을 위한 척도로 이용할 수 있으며, 이러한 사항을 문서화함으로써 자료의 신뢰성을 확보하거나 추후 개선사항을 점검할 수 있다.

**작성 방법**

① 품질보증 척도에 대한 문서는 없음

➡ 무응답 추적, 조사원교육지침과 관련된 내용이 문서화 되지 않은 경우

② 품질보증 척도에 대한 모든 문서를 보유하고 있음

➡ 무응답 추적, 조사원교육지침과 관련된 내용이 문서화 되어 있는 경우

③ 품질보증 척도에 대한 다음과 같은 문서가 있음

➡ 관련 문서의 내용을 기술한다.

➡ 이용가능하지 않은 이유를 기술한다.

**6-5. 통계작성 절차에 필요한 행정처리절차에 관한 문서 보유 여부**

**[작성 근거]**

통계작성에 필요한 제반 절차들이 문서화 되어 있는지를 점검하는 문항으로, 가능한 자세한 사항을 절차별로 기록하거나 관련문서를 보관하도록 함으로서 향후 유사한 조사에 참고자료로 이용할 수 있으며, 또한 현행 조사의 절차상 개선사항을 점검할 때 이용할 수 있다.

**☞ 작성 방법**

- ①예 ➡ 조사원의 채용, 예산의 확보 및 이용, 조사표인쇄, 지방행정기관의 협조공문 및 담당자 명부 등 통계작성 절차상의 필요한 행정처리에 관한 문서가 있거나, 관련 서류를 관리하고 있는 경우
- ②아니오 ➡ 관련내용을 문서화 하지 않았거나, 서류를 관리하고 있지 않은 경우

**6-6. 품질 기준 관련 내용이 문서에 포함되었는지 여부****[작성 근거]**

통계문서의 내용에 포함된 품질지표를 점검하는 문항으로 관련성, 정확성, 시의성 및 정시성, 접근가능성 및 명확성, 비교가능성, 일관성 등 6가지 품질차원 중 문서에 포함된 사항을 점검하여 문서에 포함하도록 함으로서 이용자들의 통계에 대한 이해를 높일 수 있다.

**☞ 작성 방법**

통계품질지표관련내용이 문서에 포함되었는지를 점검하는 문항이다. 관련 사항이 포함되어 있으면 해당항목에 체크한다.

**6-7. 자료 이용이 시계열 단절 등과 같이 제한사항을 제시한 문서가 있는지 여부****[작성 근거]**

이용자를 위하여 조사 자료상의 시계열 단절, 분류의 개선 등과 같이 자료 이용이 유의사항에 대해 설명이 문서화 되어 있는 지를 점검하는 문항이다. 만일 이와 같은 내용이 문화 되지 않으면, 자료의 신뢰성을 떨어뜨리는 결과를 초래할 수 있기 때문에 반드시 이용이 제한 사항을 명시하는 것이 바람직하다.

#### 작성 방법

- ① 예 ➔ 조사 자료의 이용이 별도의 '이용자 유의사항'과 같은 항목으로 자료상의 제한사항을 명시한 경우(예: 2007년 1사분기 자료는 조사정의의 변경으로 일부 자료가 누락되었음)
- ② 아니오 ➔ 조사 자료의 이용시 별도의 자료 이용의 제한사항을 명시하지 않은 경우

### 6-8. 발간물에 대한 외부 이용자의 만족도 정보가 있는지 여부

#### [작성 근거]

통계발간물과 관련하여 이용자 만족도를 평가한 정보가 있는지를 점검하는 문항으로 외부이용자 만족도와 내부이용자 만족도를 평가하여 이를 통계작성절차에 반영하는 것이 바람직하다.

#### 작성 방법

- ① 예 ➔ 이용자 만족도 조사를 통해 통계발간물의 만족도를 평가하고, 이에 대한 정보를 문서화하고 있는 경우
- ② 아니오 ➔ 외부 이용자만족도 조사를 실시하지 않으며, 이에 대한 정보도 없는 경우

### 6-9. 내부직원의 만족도 관련 정보의 보유 여부

**[작성 근거]**

6-8번 문항과 같은 개념으로 외부 이용자 만족도뿐만 아니라 내부이용자 만족도를 평가하여 관리하도록 해야 한다.

**작성 방법**

- ① 예 ➡ 내부 이용자 만족도 조사를 통해 통계발간물의 만족도를 평가하고, 이에 대한 정보를 문서화 하고 있는 경우
- ② 아니오 ➡ 별도로 내부이용자 만족도 조사를 실시하지 않으며, 이에 대한 정보도 없는 경우

**6-10. 발간물 내용의 개선을 위해 포함되어야 할 사항**

**[작성 근거]**

통계발간물을 검토하여 향후 개선할 사항이 있는 지를 점검하는 문항이다. 통계작성 절차의 제반사항으로 자료의 수집, 조사원교육과 훈련, 조사표내용, 조사설계, 자료처리 및 분석, 자료의 관리, 문서의 공표 등 다양한 측면에서 개선사항을 검토하도록 한다.

**작성 방법**

- ① 없음 ➡ 발간물의 개선사항이 없는 경우
- ② 자료수정
  - ➡ 자료입력 단계에서 에디팅 절차를 통한 자료 수정 사항과 관련된 내용을 보완할 필요가 있는 경우
- ③ 자료처리작업
  - ➡ 자료의 입력과 처리 단계와 관련된 사항을 보완할 필요가 있는 경우
- ④ 방법론
  - ➡ 조사 방법론 또는 추정방법과 관련된 내용을 보완할 필요가 있는 경우

⑤ 품질관리

- ➡ 문서의 내용에 포함된 품질관리 사항으로 조사원의 교육과 훈련, 품질지표 등에 대한 내용을 개선할 필요가 있는 경우

⑥ 직원배치

- ➡ 조사관리 업무를 담당할 직원의 변경 등으로 인해 담당자 관련 사항을 보완할 필요가 있는 경우

⑦ 사용하는 소프트웨어

- ➡ 자료의 처리에 이용된 소프트웨어의 변경 및 개선으로 관련 사항을 개선할 필요가 있는 경우

⑧ 기타

**6-11. 이용자측면에서 최근에 생산된 자료목록과 정보서비스의 제공여부**

**[작성 근거]**

통계자료를 이용하고자 하는 이용자의 관점에서 보다 신속하게 자료를 검색할 수 있도록 자료 목록을 제공하거나 관련 내용에 대한 정보서비스를 제공하고 있는지를 점검하는 문항이다.

**☞ 작성 방법**

- ① 예 ➡ 이용자가 관련 통계목록을 쉽게 검색하거나 이용할 수 있도록 관련 정보 서비스를 제공하는 경우
- ② 아니오 ➡ 이용자에게 관련 정보서비스를 제공하지 않는 경우

**6-12. 이용자에게 적합한 자료제공 시스템의 보유여부**

**[작성 근거]**

자료제공을 위해 인터넷, CD, 발간물 등으로 자료를 제공하고 있는 시스템을 갖

추고 있는 지를 점검하는 문항으로 이용자의 접근성을 향상시키기 위해 제반 시스템을 관리할 필요가 있다.

☞ 작성 방법

- ①예 ➡ 인터넷, CD, 발간물 등의 자료제공 시스템을 갖추고 있는 경우
- ②아니오 ➡ 이용자의 자료이용을 위한 적절한 자료제공 시스템을 갖추지 못한 경우

**6-13. 통계표와 그래프 작성 기준의 보유여부**

[작성 근거]

문서화 과정에서 일관성유지를 위해 가능한 한 통계표와 그래프의 작성 지침을 마련하는 것이 바람직하다. 2차원 통계표의 경우 행과 열에 포함될 변수 중 기준변수를 지정함으로써 통계표 해석상의 오해를 방지할 수 있다. 또한 통계표상의 단위를 명시함으로써 이용자가 통계값을 읽는데 불편하지 않도록 해야 한다. 그래프 작성의 경우 2차원, 3차원 그래프의 작성 시 기준 축에 표시될 변수에 대한 기준을 마련하는 것 또한 바람직한 작성 방법이라 할 수 있다.

☞ 작성 방법

- ①예 ➡ 2차원 통계표의 작성에서 기준변수를 행 또는 열에 포함할 것인지, 통계수치의 유효 숫자의 자리 수는 어디까지 표현할 것인지 등과 그래프의 기준 축을 어디로 할 것인지 등에 대한 기준을 문서화 하거나 보유하고 있는 경우
- ②아니오 ➡ 별도의 작성기준이 없는 경우

**6-14. 개편, 잠정치 또는 추계치 등의 표시여부와 제한사항, 잠정자료에 대한 기대 수준에 대한 정보 제공 여부**

**[작성 근거]**

발간물의 공표에 앞서 관련통계의 잠정치 또는 추계치를 발표하거나, 또는 발간물 상에 추계치를 수록할 경우 이에 대한 제한사항과 설명을 수록함으로써 이용자의 이해를 돕도록 해야 한다. 또한 조사대상 표본의 개편 등으로 추정치의 변동이 발생할 수 있으므로 이에 대한 내용도 함께 제공하는 것이 바람직하다.

**☞ 작성 방법**

- ① 예 ➔ 통계 절차의 변경이나 표본 개편, 또는 잠정치의 발표 등에 대한 상세한 설명과 이용이 유의사항을 명시한 경우
- ② 아니오 ➔ 잠정치, 추계치 등에 대한 설명을 제공하지 않는 경우

**6-15. 자료관리와 보관을 위한 체계적인 방법의 보유여부****[작성 근거]**

조사 자료관리 및 보관을 위해 사용하고 있는 시스템이 있는지를 점검하는 문항으로 많은 예산과 시간을 투입하여 얻은 정보의 관리 및 보관은 매우 신중하게 이루어져야 한다. 조사관리자 이외에는 허가를 받은 사람만이 자료의 접근을 허용하는 등의 보안이 필요하며, 만일 DB 와 같은 컴퓨터 시스템으로 자료를 관리할 경우에는 외부의 불법적인 접속을 차단할 수 있는 별도의 보안체계를 수립해야 한다.

**☞ 작성 방법**

- ① 예 ➔ 조사 자료가 DB화 되어 있어 조사관리자 이외에 자료 접근에 대한 인증을 허가한 사람만이 접근이 가능하도록 하고, 외부의 접근에 대한 보안 관리가 체계적으로 수립된 경우
- ② 아니오 ➔ 자료접근에 대한 어떠한 허가도 필요하지 않고, 자료의 관리가 DB화 되어 있지 않은 경우

## 6-16. 자료 보관 시 자료보호를 위한 방법의 보유여부

### [작성 근거]

자료를 관리하거나 보관할 경우 예상하지 못한 만일의 사태에 대비하여 반드시 자료에 대한 사본이나 백업자료를 만들어 관리하는 것이 필요하다. 또한 DB화 된 자료 시스템에 외부의 불법적인 자료접근을 방지하기위해 별도의 보안체계를 수립 하도록 해야 한다.

### ☞ 작성 방법

- ①예 ➡ 원 자료의 사본파일을 보유하고 있거나, 자료가 갱신될 때마다 백업 자료를 관리하고 있는 경우
- ②아니오 ➡ 별도의 사본 파일이나 백업 자료를 관리하지 않는 경우

## 7. 개선 사이클

### 7-1. 새로운 요구에 대응하는 시간의 적절성 평가

### [작성 근거]

소관통계의 품질개선을 위해 통계작성기관의 대응력을 점검하는 항목으로 기존의 조사체계나 분석기법의 변경, 새로운 이용자 요구사항에 대해 신속히 대응할 수 있도록 대비해야 한다.

### ☞ 작성 방법

- ①만족하며 만일 필요한 경우 거의 모든 개선사항을 이행 할 수 있음
  - ➡ 신속하게 개선사항에 대해 대응할 수 있는 경우
- ②만족하지만 특정한 사항에 대해 개선의 여지가 있음

➡ 조사 설계, 자료수집, 자료처리 및 분석, 결과의 공표 등의 절차 중에서 개선할 사항이 있는 경우 이를 구체적으로 기술

③ 만족스럽지 않고, 자료입력시스템의 적용가능성을 개선할 필요가 있음

➡ 통계절차의 변경과 개선에 대해 신속하게 대응할 수 없으며, 특별히 이러한 변경에 대해 자료입력 시스템을 개선해야 하는 경우

## 7-2. 패널조사와 같이 특수한 조사를 수행하기 위한 자료수집기반의 보유 여부

### [작성 근거]

패널조사와 같이 지속적으로 조사를 수행하기 위해서는 숙련된 조사원의 확보 및 패널 표본의 관리를 위한 전문적인 조직이 구성되어야 한다. 또한 과년도 조사 자료의 활용을 통해 조사 자료의 에디팅 점검 등이 이루어져야 함으로 조사 자료의 관리를 위한 별도의 시스템을 운영하도록 해야 한다.

### ☞ 작성 방법

① 예 ➡ 지속적으로 조사업무를 수행할 수 있는 조사원 및 패널표본관리를 위한 조직을 보유하고 있으며, 조사 자료의 체계적인 관리를 위한 시스템을 보유하고 있는 경우

② 아니오 ➡ 지속적으로 조사업무를 수행할 수 있는 조사원 및 패널표본관리를 위한 조직이나, 조사 자료의 체계적인 관리를 위한 시스템을 보유하고 있지 않은 경우

## 7-3. 수요에 신속하게 대응하기 위한 부담과 비용정도

### [작성 근거]

조사표, 자료수집방법, 자료처리 절차, 결과공표 수단 등의 개선에 소요되는 인력과 예산을 적절히 반영할 수 있는 지를 점검하는 문항으로 통계품질 개선을 위해

필요한 전반적인 부담과 비용이 어느 정도 인지를 평가하여 개선사항의 우선순위에 따라 적극적으로 수용할 필요가 있다.

#### 작성 방법

- ① 수용할 수 있으며, 거의 줄일 수 없는 상태
  - ➡ 변화하는 수요에 신속하게 대응할 수 있고, 이에 따른 부담과 비용은 줄일 수 없는 경우
- ② 수용할 수 있지만, 신속한 대응에 소용되는 비용은 줄일 수 없음
  - ➡ 변화하는 수요에 신속하게 대응할 수 있지만, 이에 따른 비용은 줄일 수 없는 경우
- ③ 너무 높음
  - ➡ 변화하는 수요에 신속하게 대응하기 위한 부담과 비용이 너무 큰 경우
- ④ 기타

### 7-4. 조사담당 직원의 업무수행능력

#### [작성 근거]

통계조사는 매우 전문적인 업무 분야로서 조사와 관련된 학문적 배경과 현장 경험 등이 아우러져야 하는 분야이다. 따라서 조사담당 직원이 조사관련 지식과 경험을 충분히 습득하고 있는 지를 점검하는 문항이다.

#### 작성 방법

- ① 조사담당직원은 조사에 관련된 모든 주제에 대해 충분한 전문성을 보유함
  - ➡ 조사 설계, 자료수집, 자료처리 등 조사통계관련 지식과 경험이 풍부한 경우
- ② 조사담당직원은 조사에 관련된 대부분의 사항에 대해 충분한 지식을 보유함
  - ➡ 조사 설계, 자료수집, 자료처리 등 조사통계관련 지식과 경험이 일정부분 보유한 경우

③ 직원 중에는 모든 사항에 대한 충분한 전문성은 없지만 필요한 경우 외주가 가능함

➡ 자체적인 인력 중에는 조사관련 지식과 경험을 가진 직원은 없지만, 이에 대한 예산이 배정되어 외부 인력을 운용할 수 있는 경우

④ 전문성이 충분하지 못함

➡ 조사관련 직원의 전문성이 떨어지며, 외부 인력을 활용할 수 없는 경우

#### 7-5. 직원들의 교육과정의 평가

##### [작성 근거]

통계작성기관의 통계업무 담당자들의 전문성확보를 위해 정기적인 교육과 더불어 관련 기관이나 외부 전문가 협회간의 다양한 교류가 이루어지도록 해야 한다. 또한 통계관련 교육이수자들의 만족도를 평가하여 교육과정의 개선 정책에 반영하도록 해야 한다 .

##### ☞ 작성 방법

① 가능한 한 직원들에 대해 아주 훌륭한 교육과정이 있음

➡ 정기적인 통계교육을 실시하고 있으며, 직원들의 교육에 대한 만족도가 매우 높은 경우

② 직원의 교육은 충분하지만 개선이 필요함

➡ 직원들에 대한 교육을 실시하고 있으나, 이들의 만족도가 보통인 경우

③ 직원교육 상황이 시급히 개선되어야 함

➡ 직원들의 교육과정에 대한 만족도가 매우 낮고, 개선의 필요성에 대한 요구가 매우 높은 경우

#### 7-6. 조사업무 담당자의 공백 없이 업무를 수행할 정도의 전문성이 있는지 여부

**[작성 근거]**

통계관련 담당자들의 전문성은 매우 중요하다. 외국의 경우 통계업무 담당자의 경력이 최소 10년 이상인 직원들이 매우 많으며, 이러한 측면에서 통계업무 담당자의 전문성 재고를 위한 교육훈련이 정기적으로 이루어져야 한다.

**작성 방법**

- ① 주요 담당자의 변동 시 업무공백 없이 전문성을 유지할 수 있음
  - ➡ 충분히 전문성이 있어 업무공유가 가능한 경우
- ② 비록 주요담당자의 변동에 대해 전문성의 유지가 어려울 때도 있지만 대체로 전문성을 유지할 수 있음
  - ➡ 일부 매우 전문적인 업무를 제외하고 대체로 업무공유가 가능한 경우
- ③ 주요 담당자의 변동에 대해 대체 인력의 전문성이 부족함
  - ➡ 업무공백에 대해 대체할 인력의 전문성이 없는 경우

**7-7. 타 기관, 전문가협회등과의 교류정도****[작성 근거]**

통계작성기관의 담당자들의 전문성을 향상시키기 위해 가능한 다양한 분야의 기관과 교류를 확대할 필요가 있으며, 통계의 전문성 증대를 위해 전문학회 또는 협회와의 교류를 강화시킬 필요가 있다. 이와 더불어 해외의 전문기술 습득을 위한 교류 확대 또한 중요하다.

**작성 방법**

- ① 아주 좋음, 다양한 국내외 기관과 협회 간 교류가 있음
  - ➡ 정기적인 세미나, 워크숍, 학술 모임 등으로 교류가 활발한 경우
- ② 만족스러우며 비록 교류를 확대할 수 있지만, 최근 개발동향을 따라갈 만큼 충분한 교류가 있음

- ➡ 전반적으로 정기적인 학술 및 세미나 등의 교류가 이루어지고 있는 경우
- ③ 부족하며 다른 기관이나 전문가 협회와의 교류가 개선되어야 함
- ➡ 연간 1회 정도 또는 그 이하의 교류가 이루어지고 있어 향후 교류확대가 필요한 경우

### 7-8. 조사에 대한 정기적인 점검 수행 여부

#### [작성 근거]

통계작성기관에서 생산되고 있는 소관통계에 대한 점검을 통해 통계생산의 모든 과정에 투입되는 예산, 인력 등의 적절한 배분이 가능하며, 각 절차에서 발생할 수 있는 오류를 방지할 수 있기 때문이다.

#### ☞ 작성 방법

- ① 점검이 이루어지지 않음
  - ➡ 소관통계에 대한 정기적인 점검을 하지 않는 경우
- ② 모든 생산과정을 검토하고 있음
  - ➡ 전체적인 점검절차를 수행하고 있는 경우
- ③ 다음과 같은 과정을 검토함
  - ➡ 통계생산의 전체 과정보다는 특정 부문에 대한 점검이 이루어지고 있는 경우

### 7-9. 효율성, 시의성 또는 품질을 확보하기 위해 최신의 IT 기술의 보유여부

#### [작성 근거]

소관통계의 품질관리 측면에서 효율성과 시의성을 확보하기 위해 최신의 정보통신 기술을 보유하고 있거나, 이를 적용하고 있는지를 점검한다. 국내의 경우 정보통신 기술의 급격한 발전으로 통계작성여건이 상당부분 개선되었으며, 자료의 제공 또는

공표의 측면에서 또한 이러한 IT 기술의 발전에 따라 개선되고 있기 때문에 가능한 다양한 측면에서 이러한 기술을 통계생산과정에 적용하는 것이 바람직할 것이다.

#### 작성 방법

##### ① 대부분의 절차에 대해 IT 기술로 전환해야함

➡ 조사표, 자료입력, 에디팅, 자료의 공표 등 전반적인 절차에 IT 기술을 적용할 필요가 있는 경우

##### ② 거의 모든 과정에 대해 최신의 IT 기술을 사용하고 있음

➡ 자료수집, 자료입력, 에디팅 절차, 자료의 공표 등의 전체적인 과정에 IT 기술을 사용하고 있는 경우

##### ③ 최신의 IT 기술을 사용하고 있지만 다음과 같은 절차에 대한 기술은 개선할 필요가 있음

➡ IT 기술을 전반적으로 사용하고 있지만, 특정한 부분에 대해 개선할 사항이 있는 경우

#### 7-10. 재정, 인력, 사용가능한 기존의 자료 등에 관한 전반적인 검토가 가능한 지 여부

##### [작성 근거]

통계생산기관의 소관통계를 생산하는데 소요되는 자원, 인력등과 함께 사용가능한 기존의 자료가 있는지 여부 등에 대해 광범위한 검토가 가능한지를 점검하는 문항이다. 이에 대해서는 내부 조사통계관련 전문가가 평가하도록 해야 하며, 만일 내부 전문가가 없는 경우 외부 전문가에게 자문을 받도록 해야 한다.

#### 작성 방법

① 예 ➡ 통계예산, 인력, 기존자료의 사용가능 여부 등에 관한 광범위한 검토가 이루어지는 경우

②-아니오 ➔ 위와 같은 구분이 전혀 되어 있지 않은 경우

#### 7-11. 통계생산과정의 효율성을 전문가에게 검토를 받는지 여부

##### [작성 근거]

통계작성기관의 내부에 통계관련 전문가가 있는 경우 통계생산과정이 효과적으로 수행되고 있는 지를 검토하도록 함으로서 불필요한 예산이나 인력의 낭비를 방지할 수 있다. 만일 내부 전문가가 없는 경우에는 외부 전문가로부터 이에 대한 자문을 얻는 것도 고려할 수 있다.

##### 작성 방법

①-예 ➔ 대내외 통계전문가로부터 통계작성과정 전반에 걸쳐 효율성을 점검받는 경우

②-아니오 ➔ 효율성 점검을 받지 않는 경우

#### 7-12. 예산의 20% 증가 또는 20% 감소 시 대응 능력

##### [작성 근거]

통계작성 관련 예산의 증가 또는 감소는 통계를 생산하는 기관 또는 관리자의 관점에서 매우 중대한 사항이다. 따라서 통계작성과 관련된 예산의 기획단계에서 가능한 한 신축성 있게 예산을 편성하는 것이 바람직하며, 예산의 증액 또는 감액 상황에 대비하여 조치사항에 대한 우선순위를 정하여 이에 대응하도록 해야 한다.

##### 작성 방법

가능한 예산 기획단계에서 수립한 우선순위에 따라 대처해야할 사항을 기록하도록 한다.

## 제2절 자체통계 품질진단 사례

### 1. 조사통계 - 국민건강영양조사

#### 1) 개요

##### 가) 조사연혁

『국민건강조사』와 『국민영양조사』를 통합·개편한 『국민건강영양조사』는 국민의 전반적인 건강상태와 영양상태에 대한 전국규모의 대표성과 신뢰성이 있는 통계를 산출하고, 그 추세를 제시함으로써 국가 보건정책기획의 우선순위와 목표를 설정하고 이에 따른 질병관리 및 건강증진 프로그램의 개발에 필요한 기초 자료를 제시하기위한 목적으로 1998년부터 매 3년 주기로 실시하고 있다.

『국민건강조사』는 1962년 보건복지부에서 서울시민 500가구에 대한 건강조사가 효시이며, 1989년부터 실질적인 국민건강조사로 건강면접조사와 보건의식행태조사 두 부분으로 확립되어 1992년, 1995년 3년 주기로 실시되었다.

한편 『국민영양조사』는 1969년 952가구에 대한 국민영양조사가 효시이며, 1969년 이후 1995년 까지 매년 국민영양조사가 실시되었다. 1995년 『국민건강증진법』 제정으로 조사실시의 법률적 근거가 강화되면서 두 조사를 통합하고 건강검진조사를 추가한 새로운 『국민건강영양조사』로 발전되었다.

##### 나) 조사목적

국민건강영양조사는 다음을 목적으로 한다.

- 국민건강증진종합계획(Health Plan 2010)의 목표 중간평가 및 목표수정

- 국민이 건강수준 측정
- 흡연, 음주 등 건강위험요인과 주요 질환의 관련성 분석
- 건강 고위험 집단의 선별과 이들 대상의 건강증진 프로그램 개발
- 한국인의 주요 체위수준등 주요 건강 지표의 분포상태와 추이 파악
- 질병이나 장애의 종류 및 정도에 따른 보건의료서비스 이용현황
- 국가간 비교 가능한 건강지표산출

#### 다) 조사개요

2005년 국민건강영양조사의 목표모집단은 “2005년 4월 1일 현재 대한민국에 거주하는 국민(양로원등 시설, 군복무, 교도소수감자, 외국인은 제외)”으로 한다. 2005년 국민건강영양조사의 조사모집단은 2000년 인구주택총조사의 일반조사구 246,000여개 조사구, 1,480만 가구를 조사모집단으로 한다. 국민건강영양조사의 대상은 가구이므로 기숙사, 보육원, 양로원, 수녀원, 군부대 등과 같은 시설단위 조사구는 조사모집단에서 제외하였다. 이와 함께 2000년 인구주택 총조사 자료를 조사모집단으로 함에 따른 모집단의 변동사항의 일부를 반영하기위해 2000년 11월 1일~2001년 12월 31일 까지 신축된 아파트를 추가적으로 모집단에 포함하였다. 이때 신축아파트 조사구는 아파트의 각 동을 기준으로 각 아파트 단지 내에서 60가구 내지 100가구를 하나의 조사구로 구성하였다. 최종적으로 2,738개의 신축아파트 조사구를 설정하였다.

조사하고자 하는 600개 조사구 중 가구 수에 따른 비례배분 방법으로 인구주택총조사구에서 582개 조사구를 할당하고, 신축아파트 조사구에는 18개 조사구를 할당하였다.

#### 라) 표본규모 및 표본추출방법

표본조사구의 규모는 600개 조사구이며, 표본가구의 규모는 12,000여 가구 약 40,000여명을 목표로 하였다. 이중 보건의식행태조사, 건강검진조사, 영양조사가 건강면접조사와 함께 이루어지는 통합조사를 위해서 200개 조사구 약 4,000가구를 추

출하기로 하였다.

표본추출방법은 1차추출단위인 표본조사구의 추출은 2원층화확률비례계통추출에 의해 추출하였다. 먼저 지역층(7개광역시, 경기, 강원, 충청, 전라, 경상, 제주 등)의 행정구역(동부, 읍면부)과 아파트와 일반조사구를 층화 기준으로 하여 각 층의 조사구수에 비례하도록 배분한 후 가구 수를 기준으로 확률비례계통추출법으로 표본 조사구를 추출하였다.

마) 조사체계

국민건강영양 조사는 다음과 같은 3가지 조사로 구성된다.

〈표 5-2-1〉 국민건강영양조사의 구성

| 조사구분                 | 조사내용  | 조사방법          |
|----------------------|---|---------------|
| 통합조사(200조사구, 400조사구) |   |               |
| 건강조사                 | <ul style="list-style-type: none"> <li>○ 질병이환조사 : 만성질환, 주요질환관리</li> <li>○ 사고중독조사 : 발생원인, 기전, 장소, 손상유형</li> <li>○ 활동제한 및 삶의 질 : 1달간활동제한일수, 활동제한기간, 원인, 시작, 청각, 씹기 기능의 제한, 장애인등록, EQ-5D</li> <li>○ 의료이용조사 : 입원, 외래, 응급실, 치과 약국, 한방 등의 사용량, 서비스만족도, 비용, 치료지연 및 미치료 이유 등</li> </ul> | 개별면접조사        |
|                      | ○ 보건영양조사 : 흡연, 음주, 운동, 비만, 안전의식 등   | 자기기입식         |
| 검진조사                 | <ul style="list-style-type: none"> <li>○ 혈압, 맥박, 체중, 허리둘레 등 신체계측</li> <li>○ 임상검사: 콜레스테롤, 중성지방, HDL-콜레스테롤, 혈당, 간기능검사, B형간염, 신장기능검사, 빈혈검사 등</li> </ul>  | 이동검진센터계측 및 검진 |
| 영양조사                 | <ul style="list-style-type: none"> <li>○ 식품섭취조사 : 식품의 섭취회수, 식품의 재료 등</li> <li>○ 식생활조사 : 규칙적인 식사여부, 식품섭취의 과다여부 등 일반적인 식습관, 외식의 횟수 영유아의 수유기간, 이유보충식의 종류 등</li> </ul>  | 가정방문면접조사      |

2) 자체진단 점검사항

(1) 조사 수행을 위한 의사 결정

■ 이용자 요구분석

- 조사수행의 법적근거
  - 국민건강영양조사는 국내법인 『국민건강증진법』에 근거를 두고 수행되고 있음.
- 조사에 대한 이용자 정보
  - 주요 이용자에 대한 일부 정보만을 가지고 수행되고 있음.
- 주요이용자
  - 1순위) 정부기관, 2순위) 연구기관 3순위) 학교, 4순위) 언론기관, 5순위) 국제기구

[그림 5-2-1] 국민건강영양조사의 주요이용자 순위



- 주요이용자 정보의 관리
  - 주요 이용자들 정보는 이전 조사 자료에 일부가 수록되어 있음.
- 이용자에 대한 자문을 위한 절차
  - 보건복지부의 담당부서인 건강증진 건강정책과에서 계획수립과정부터 관여하여 반영토록하고 있으며, 생산자-이용자 그룹, 다수이용자 그룹의 수요를 파악하는 고객조사를 실시하고 있음.
- 향후 이용자의 수요를 예측을 위한 절차
  - 통계위원회나 소속 소위원회에서 계획수립, 정부부처와 양자간 위원회, 다른 이용자 그룹과 양자간 회의나 워크숍 등을 거치고 있음.
- 미충족 이용자수요
  - 지방자치단체 공무원 및 보건소의 수요에 대해 미충족욕구가 있음.

- 주요 이용자 수요에 적절히 부응하기 위한 계획
  - 지역보건의료계획 수립에 필요한 지역건강조사를 통합하여 운영하고자 하는 계획을 수립함.

- 조사목적의 명문화
  - 조사목적은 조사개요 부분에 명문화되어 있음.

(2) 조사설계

■ 조사개념

- 비교 가능한 공식/비공식 조사
  - 비교 가능한 공식적이거나 비공식적인 통계가 있음.
- 조사품질에 활용할 만한 공식/비공식조사
  - 통계청에서 실시하는 사회통계조사, 보건복지부의 활동계약자실태조사, 출산력 조사 등이 있음.
- 표준화된 체계, 기준, 개념, 정의의 분류기준
  - 건강검진조사의 경우 JIC의 국제적인 기준과 건강면접 조사의 경우 국민건강 조사에 적절한 개념을 조사특유의 기준에 의거하고 있음.
- 비교 가능한 조사에서 공통적인 표본추출의 여부
  - 공동표본추출틀을 사용하고 있지 않음.
- 과대, 과소포함 문제
  - 표본추출틀은 2000년 인구주택총조사구를 사용하고 있기 때문에 시간의 경과에 따라 조사구의 증편 또는 삭제 등으로 인해 과대나 과소 포함이 아주 작게 나타나고 있음.
- 통계단위에 대한 오분류 여부
  - 통계조사 단위에 대한 오분류는 거의 없음.
- 조사주기와 추출틀의 최근 갱신 시기간의 시차
  - 조사주기와 표본추출틀의 최신 갱신시기와 시간차는 1년 초과 3년 미만이며, 이는 표본조사가 5년마다 실시하는 인구주택총조사 조사구를 이용하기 때문으로

사용 가능시점을 고려하면 이론적으로 6년의 차이까지 날수도 있음.

- 조사주기와 표본추출들의 최종 갱신주기가 근접하게 되면 과소나 과대 포함문제의 해결여부
  - 조사주기와 표본추출들의 최종 갱신주기가 근접하게 되면 과소나 과대 포함문제가 해결된다고 볼 수 있음.
- 표본추출들과 목표모집단과 가능한 근접한지를 측정하는지 여부
  - 지방자치단체의 담당 공무원과 업무협조를 통해 점검하거나, 추출들의 일부자료를 조사현장에서 점검하고, 일정기간 동안 발생하는 변동사항을 반영하여 추출들에 갱신하고, 표본설계에서 사용된 지역경계가 분명하고 중첩되지 않는지를 확인하기위하여 지역별 추출들을 지도로 점검함.

**■ 표본설계**

- 보편적인 표본추출기법의 사용여부
  - 표본추출 기법은 국가통계기관의 표본추출 기법을 채택하여 사용하고 있음.
- 최적의 표본크기결정
  - 허용오차의 한계와 조사예산 등을 감안하여 표본크기를 결정함.
- 조사에 적용한 표본설계방식
  - 표본설계는 지역별(특 광역시 및 도지역과 동부와 읍면 부)로 층화하고 각 층의 조사구 크기에 따라 표본을 배분하고, 층별로 표본 조사구를 계통 추출하는 2원 층화집락계통추출로 이루어짐.
- 층화 추출설계에서 표본배분방법
  - 조사단위를 지역, 행정구역을 층화변수로 하고, 각 층의 조사구 크기에 비례배분 하여 확률비례계통 추출에 의해 조사구를 추출함.
- 응답부담을 최소화하기 위하여 연동표본구조를 이용하는지 여부
  - 응답부담을 최소화하기 위하여 향후 조사에서 연동표본구조를 적용하기로 계획함.
- 표본추출 및 크기를 결정하기위해 사용됨 프로그램

- 표본추출을 위해 범용화된 프로그램으로서 SAS를 사용하고, 표본크기를 결정하기 위해서는 엑셀과 같은 프로그램을 이용함.

목표표본크기와 실현표본크기간의 차이

- 실현표본크기(12,005가구)는 목표표본크기(12,000가구) 간의 차이는 거의 없음.

결측 단위에 대한 현장 대체 허용여부

- 결측 단위에 대해서는 계통적으로 순서에 따라 이후 단위로 정해진 한계 내에서 현장대체를 허용함.

조사 설계의 효율성

- 조사 설계는 주어진 예산 및 시간 내에서 매우 만족스럽지만, 적은 부분에서 개선이 필요함.

조사 설계의 변경여부

- 조사방법 측면에서는 조사 설계에 변화의 필요성은 느끼고 있으나 좀 더 테스트 해볼 필요가 있다고 느낌(응답자 부담을 줄일 수 있는 방안을 검토).

**■ 조사표의 개발 및 검증**

조사표의 질문 구성

- 조사표 설계를 위해서 조사항목에 대해 사전에 테스트되어 균형을 이루고 있음.

응답하거나 조사표 작성 소요시간

- 조사표 작성 소요시간은 45분 소요되는 분량임.

응답자가 조사표 작성이나 면접에 응하기 위하여 정보 수집과 관련된 응답

- 응답자의 낮은 응답부담으로 정보를 쉽게 제공함.

민감한 문항에 대한 응답부담

- 조사표 작성 시 민감한 내용에 대한 부담은 적당하나 일부 민감한 항목을 포함하고 있음.

조사표에 적용하고 있는 항목들

- 전문성과 사업성을 나타내는 매력적인 디자인
- 서두에 조사의 제목 또는 주제, 후원자, 조사의 목적, 응답자의 협조요청, 작성

기관, 자료보안 정보 등의 제시

- 장을 구별하기위한 제목과 머리말
- 색깔, 삽화, 상징어, 진하게 쓰기 등과 같은 그래픽을 추가하여 응답자가 반드시 답해야 하거나 주의를 환기시키는 표시를 적용
- 단어와 개념을 일관성 있게 사용하고 있음
- 각 문항들이 합당한 이유가 있음
- 조사문항들이 응답자의 논리에 따름
- 조사문항의 배열순서가 건너뛰기 등으로 방향이 적절히 지정됨
- 조사대상기간과 응답단위 등이 명확함
- 맨 끝 페이지에 응답자에게 감사하는 문구를 제시함
- 지침과 추가 가이드 작성 등 있음

조사에 사용한 조사표의 품질

- 조사표의 품질은 소수의 항목이 개선되어야 한다고 생각함.

조사표 검증의 주체

- 조사표 검증은 내부적 및 사전조사를 위해 외부의 조사시험실을 사용하고 있음.

조사표의 응답부담을 측정하기위해 적용한 조사표 검증 방법

- 사전조사에 의한 조사원에게 설명을 해주는 방법을 택하고 있음.

(3) 자료수집

■ 자료의 원천

조사에 사용된 자료수집 방법

- 문헌검토, 법적인 요구사항, 가능한 능력, 이전 조사에 따라 일관적으로 사용함.

자료수집방법

- 조사원면접방식으로 종이조사표(PAPI)를 이용함.
- 2번으로 이동함

■ 조사원의 채용 및 교육

- 조사원채용을 위한 절차
  - 조사원으로 갖춰야할 최소의 자격요건에 근거하여 선발함.
- 조사를 위한 조사원 준비사항
  - 조사원은 한 과정에서 적절한 교육을 받고 조사원 메뉴 일을 제공함.
- 조사원에 대한 면접성공 보너스 지급여부
  - 조사원이 면접을 성공적으로 마칠 때마다 별도의 보너스를 지급하지는 않고 있음.
- 조사원 채용방식
  - 조사원은 중앙에서 모두 채용함.
- 예비조사원의 여부
  - 조사원을 요구할 경우 예비조사원을 지원해 줌.
- 조사원 관리방법
  - 새로 채용된 조사원들은 조사 시 보통 2회 정도 점검을 받고, 의심스러운 사실이 있을 경우 특별히 관리하며, 지침 준수여부, 표본대상, 조사표 기입방법 등에 대한 기본적 관리를 하고 있음.
- 조사 실시 공지여부
  - 응답 협조가 용이하도록 하기위해 조사가 있다는 사실을 사전에 각 가구에 공지하고 있음.
- 조사원 1인의 평균 면접회수
  - 1인의 조사원이 1일 평균 1.7가구를 면접함.
- 한명의 조사원이 실시하는 최대면접 회수
  - 1인의 조사원이 최대 1일 2가구를 면접함.
- 조사원에 대한 인적사항 파악여부
  - 조사원에 대한 인적 사항은 채용 시 충분히 파악하고 있음.
- 조사원의 인구학적 속성과 목표모집단의 응답패턴의 영향 분석
  - 조사원의 인구학적 속성과 목표그룹의 응답패턴에 영향을 분석하지 않음.
- 면접가능 시간대

— 면접가능시간대 정보를 파악하고 있음.

면접가능시간대 정보의 유형

— 면접에 대해서는 지정한 날짜, 대략적 시간의 정보를 가지고 있음.

조사에 컴퓨터의 활용여부

— 조사에 컴퓨터는 활용하고 있지 않음.

→ ~~수~~번으로 이동함

### ■ 무응답 줄이기

응답을 독려하기위한 방법

— 응답거절 사유에 따라 2번 이상 추적조사를 실시하였고, 응답자의 문의사항을 해결할 수 있는 표준적인 절차와 기반을 가지고 있음.

응답내용에 대한 타당성을 응답자에게 확인하는 절차를 거치는 지 여부

— 응답내용이 타당하지 않을 때 조사지도원이 응답자에게 다시 질의하여 확인함.

### ■ 현장실사

향후 행정기록과 같은 활용 가능한 자료의 점검 여부

— 조사 자료를 사용하는 경우 향후 행정기록과 활용 가능한 자료를 사용가능한지 점검하지 않음.

법률적인 측면에서 행정자료나 등록 자료의 사용에 대한 제한 정도

— 개인의 비밀보호와 관련된 사항이 많기 때문에 법적제한이 매우 심함.

현장실사기간이 조사대상기간을 완전히 포함하고 있는지 여부

— 현장실사기간이 조사대상 기간을 완전히 포함하고 있음.

현장실사기간의 소요기간

— 조사는 3개월 정도가 소요되었음.

조사기간 동안 휴가나 파업 또는 천재지변 등의 사건이 있었는지 여부

— 조사기간 동안 조사에 영향을 미치는 사건은 발생하지 않음.

조사원 중 슈퍼바이저와 같은 실사감독을 맡은 사람의 비율

— 조사지도원(슈퍼바이저)의 비율은 20~25% 정도임.

조사과정에서 발생할 수 있는 문제점들에 대한 조치

— 즉시 연락할 수 있는 조사지도원이 있으며, 슈퍼바이저가 지속적으로 모니터링 하고, 오차의 빈도 및 원인에 대한 정보를 환류 하여 보고하고 있으며, 조사의 시작부터 끝날 때까지 표본단위를 추적하기 위하여 모든 자료수집 과정에 대한 표본 관리 절차가 있음.

조사완료 후 입수되는 자료의 처리과정에 대한 평가

— 입수되는 자료를 처리하는 과정은 받아들이지만 개선이 필요하다고 생각함.

조사완료 후 입수 및 제공되는 원시자료의 보안유지 정도

— 입수와 제공되는 원시자료는 보안유지를 위하여 입수되는 조사표에 식별정보를 제외하고 있음.

#### (4) 자료입력 및 자료처리

##### ■ 자료입력

자료입력방식

— 수작업에 의한 자료입력 방식 채택.

자료코딩방식

— 자료의 코딩은 수작업으로 이루어짐.

수작업 코딩시 코드관리나 일관성 점검

— 일관성은 독립적인 검증을 실시하고 있음.

##### ■ 에디팅 절차

원시자료 에디팅의 필요성

— 원시자료에서 오류가 많이 발견되어 점검은 필수 불가결함.

에디팅 작업방식

— 수작업에 의한 에디팅

채택하고 있는 에디팅 작업에서 이용하고 있는 자료

- 주어진 조사에서 다른 응답자들의 관측된 자료만으로 이루어지고 있음.
- 에디팅 절차에 대한 사전 점검 여부**
  - 에디팅 과정은 사전에 점검이 이루어짐.
- 원시자료에 적용하고 있는 에디팅 절차**
  - 에디팅 과정에서 모든 변수에 대한 값의 범위 점검, 논리적 에디팅, 이상치 분석과 같은 통계적 검증, 논리적 점검 및 통계적 검증의 혼합방식, 동일한 통계 자료의 과거자료 또는 다른 자료간의 비교를 통한 에디팅 방법을 적용하고 있음.
- 수작업을 통한 에디팅 담당자에 문서화된 에디팅 지침서의 제공여부**
  - 정확하고 최근에 갱신된 지침서를 제공하고 있음.
- 결측치 또는 0 값 등의 개념에 대한 구분**
  - 결측치, 0 값 등의 개념에 대한 구분이 있음.
- 에디팅 절차의 우선순위**
  - 에디팅 순서에 우선은 ID로 사용되는 변수의 수정작업을 우선적으로 실시하고 있음.
- 과도한 에디팅이나 자료에 새로운 오류방지를 위한 조치**
  - 과도한 에디팅을 방지하는 대책을 가지고 있지 않음.
- 에디팅 작업의 지속여부**
  - 에디팅 작업은 일부자료가 수집되었을 때부터 지속적으로 이루어짐.
- 에디팅 절차의 개선을 위해 구축해 놓은 과거 조사오류에 대한 이용가능 통계**
  - 에디팅 과정을 개선하기 위하여 구축해 놓은 이전 조사의 오류에 대한 통계는 가지고 있지 않음.
- 조사의 품질을 측정하고 향후 조사개선을 위한 오류에 대한 통계 산출**
  - 조사의 품질을 측정하고 향후 조사개선을 위한 오류에 대한 통계는 산출하지 않음.
- 이용하는 에디팅 시스템의 신축성**
  - 매우 신축적인 에디팅 시스템을 사용하고 있어서 신규 또는 보완된 과정이 쉽

게 적용 가능함.

오류원인에 대한 정보 보유

— 오류의 원인에 대한 정보는 알 수 없음.

#### ■ 대체 절차

에디팅 과정에서 발견된 오류의 수정방법과 에디팅 비율

— 에디팅 과정에서 발견된 오류는 조사원 또는 응답자를 재 접촉하여 수정함.

통계값에 대한 대체여부

— 통계 값에 대한 대체는 적용하지 않음.

→ 5장으로 이동.

#### (5) 자료 분석 및 통계결과의 품질

#### ■ 관련성

조사에 대한 이용자 만족도 측정 방법

— 이용자 피드백에 의한 체계적인 진단 방법을 사용하고 있음

주요이용자 만족도에 대한 이용 가능한 정보의 평가

— 이용자 만족 정도를 알 수 있는 정보는 주요 이용자 정보는 많이 있으며, 다른 이용자 정보는 일부만 있음.

조사의 전반적인 관련성 평가

— 주요 분야는 관련성이 높고, 다른 분야는 약간의 관련성이 있음.

#### ■ 정확성

추정값의 정도를 측정하기위해 제시하는 값

— 변동계수 또는 상대표준오차를 제시함.

주요변수에 대한 변동계수의 수준

— 관련성이 있는 항목과 주요 부문에서 CV는 중간수준임(5~10%사이).

변동계수의 수준이 조사목적에 대해 받아들일 만 한 수준

- 지역별 통계를 산출하기에 적절한 수준임.
- 자료 분석에 H-T 추정량을 사용하는지 여부
  - Horvitz-Thompson 추정량을 사용하고 있음.
- 원하는 정도를 얻기 위해 표본의 크기가 충분히 크게 설계되었는지 여부
  - 최종적으로 바람직한 정도를 얻을 만큼 계획된 표본크기가 충분하지는 못함.
- 추정과정에 어떠한 편향을 조정하기 위해 사용한 보조정보
  - 추정과정에서 보조정보를 사용함으로써 무응답 편향, 표본 추출틀상의 오류를 조정하기 위해 재가중 방법을 사용함.
- 재가중 방법을 실시하기 위해 사용한 보정방법
  - 사후층화 방법으로 재가중을 실시함.
- 표본자료에 대한 가중변수의 합을 계산하여 제시된 가중치의 정확도 측정
  - 표본자료 전체에 대해 가중변수의 합을 산출함으로써 가중치의 정확도를 측정함.
- 조사에서 단위무응답의 수준
  - 낮은 단위 무응답률을 보임.
- 무응답을 세부범주로 분류
  - 무응답을 응답거절, 못 만남, 부재 등으로 세부 범주로 분류하고 있음.
- 단위무응답 편향을 추정하기 위한 이용 가능한 자료
  - 단위 무응답 편향을 추정하기 위한 자료는 없음.
- 항목무응답의 평가
  - 항목 무응답은 매우 낮은 편임.

■ 시의성 및 정시성

- 조사기간과 발간 시점간의 시차
  - 조사기간과 조사결과의 발표 시차는 약 17개월임.
- 계획한 발간일자와 실제 발간일자 간의 시차
  - 계획된 발간일자와 실제 발간일 사이의 시차는 4개월 정도임.

- 조사대상기간과 잠정 또는 최종 결과의 최초발간일 간의 시차
  - 조사대상기간과 잠정 또는 최종 결과의 최초발간과의 시차는 정해진 시간 정도 임.
- 주요발간물에 대해 정해진 발간시기 준수여부
  - 주요 발간물의 계획된 발간 시기는 어느 정도 지연으로 자주 맞추지 않고 있음.
- 시의성을 개선하기 위한 변화의 필요성
  - 중대한 내부의 변화가 있어야 한다고 판단됨.

■ **비교가능성**

- 통계의 중요한 개념과 조사설 계에 연계하여 지역에 대한 비교가능성 평가
  - 연관된 통계 결과와의 비교 가능성은 약간의 차이가 있음.
- 조사대상기간과 관련하여 시계열적인 비교가능성 평가
  - 시계열적인 비교가능성은 어느 정도 제약이 있음.

■ **일관성**

- 월별 또는 연간단위로 동시에 다른 주기로 조사하는지 여부
  - 동시에 다른 주기로 조사하고 있지 않음.
  - ~~월~~번으로 이동.

■ **자료분석**

- 이용자 맞춤분석 제공시 대응능력
  - 자료 분석에 대한 이용자들의 요구를 만족시킬 의무는 없음.
- 사용하고 있는 추가분석
  - 지수산출과 계절 또는 추세 분석과 추정임.
- 작년한해 동안 자료 분석을 위해 새롭게 도입했거나 기존의 방법을 개편한 적이 있는지 여부
  - 작년 한 해 동안 자료 분석을 위해 새롭게 분석방법을 도입하였고, 기존방법의

개편하였음

■ 보안관리

이용자에 대한 자료접근을 위해 보유하고 있는 자료의 형태

- 일반 연구자용 마이크로자료를 보유하고 있음.

(6) 문서화 및 자료의 공표

■ 메타자료의 문서화

통계산출과정에서 문서화하는 내용

- 수요평가, 조사표 사본, 변수에는 변수명.
- 변수의 형태로서는 통계조사표나 행정양식상 항목에 대한 응답을 코딩한 변수, 다른 변수들의 계산으로 구한 변수, 외부자료로부터 가져온 변수, 변수의 통계적 특성(즉, 범주적, 분류적, 계량적), 해당되는 레코드 길이, 측정단위(의미를 가진 경우), 표준분류가 사용되지 않을 경우 코드나 표제 값의 목록, 외부자료에서 가져올 경우 그에 대한 서술, 파생 변수의 경우 계산방법 서술임.
- 분류 서술로서는 표준 또는 일반 분류가 사용될 경우 참조하는 분류, 특정한 분류가 적용된 경우 분류의 출처, 참조된 분류의 일부항목을 처리하는데 특정 주의사항을 설명하는 정보를 포함함.
- 개념과 통계적 단위의 정의에는 정의된 개념의 제목, 정의의 출처, 정의와 그 개념을 측정하는 변수간의 연계사항을 포함함.
- 통계산출과정에서 사용된 조사방법 및 기법을 포함함.
- 조사의 처리과정에서 조사설계의 변경내용을 포함함.
- 원시자료 파일로부터 제공하는 지금까지 생산된 모든 파일을 포함시키는 결과 자료를 낼 때 파일 제목, 자료저장형식, 자료공표상태, 파일의 서술(설계, 레코드 수)등이 포함될 필요가 있음.

메타자료와 품질지표의 공표방식

- 메타자료와 품질지표들을 종이 발간물 내 메타자료나 종이 발간물 내 품질지표

로 제공하고 있음.

- 무응답 추적절차 또는 조사원 교육을 위한 매뉴얼인용과 같은 품질보증 척도에 대한 문서의 보유 여부
  - 무응답자 특성 파악 절차나 조사원 훈련 매뉴얼 등으로 문서화 되어 있음.
- 통계작성 절차에 필요한 행정 및 처리절차에 대한 체계적인 문서의 보유여부
  - 통계작성절차를 지원하는데 필요한 행정과정 및 처리절차, 예를 들면 예산의 확보와 효율적 관리, 조사표 인쇄, 직원채용, 지방행정기관 방문, 언론매체접촉 등에 관한 모든 사항을 체계적으로 문서화하고 있음.
- 품질기준 중 문서에 포함된 내용
  - 정확성 측정 항목 중 변동계수, 단위 응답률, 항목 응답률과 시의성 및 정시성 항목에서 대상기간 완료시점과 최종결과가 산출된 시점간의 평균시차 등을 문서에 포함하고 있음.
- 자료 이용시 예상되는 제한점에 대해 체계적인 문서의 보유여부
  - 자료 이용시 예상되는 한계에 대해 체계적으로 문서화 하고 있음.
- 발간물에 대한 외부이용들의 만족도에 관한 정보
  - 발간물에 대한 외부 이용자들의 만족도 정보는 보유하고 있지 않음.
- 내부직원들의 만족도에 대한 정보
  - 내부 직원들의 만족도 정보를 보유하고 있지 않음.
- 보다 나은 발간물을 위해 필요한 항목
  - 품질관리항목을 보완할 필요가 있음.

**■ 자료제공 전략**

- 생산된 자료에 대한 정보의 검색을 위해 이용자 측면에서 목록시스템과 정보서비스의 제공여부
  - 이용자에게 편리하도록 기관의 발간목록 시스템과 가장 최근에 생산되고 저장된 자료에 대한 정보를 검색할 수 있도록 인터넷 홈페이지에서 정보서비스가 되고 있음.

- 통계작성기관에서 이용자에게 적합한 형식으로 자료를 이용할 수 있도록 하는 자료제공시스템 보유여부
  - 기관수준의 자료제공 시스템은 이용자에게 적합한 형식으로 액세스하게 하고 있음.
- 통계표와 그래프 작성을 위해 사용하고 있는 기준의 여부
  - 테이블과 도표 작성에 사용된 표준화된 기준은 없음.
- 개편이나 잠정치 또는 조기 추계치 등을 설명하고 있는지 여부와 제한사항이나 잠정자료에 대한 기대수준 등의 정보가 제공되고 있는지 여부
  - 개편이나 잠정 또는 속보 등을 분명히 표시하여 설명하는지와 제한사항이나 잠정치에 대한 기대수준 등에 대한 어떠한 정보도 제공되지 않음.

**■ 자료관리**

- 자료의 관리와 보관을 위한 체계적인 방법의 보유여부
  - 자료를 관리하고 축적하기 위해 자료를 DB화 하여 관리하고 있음.
- 자료 보관시 예상치 못한 상황에 대비하여 자료를 보호하는 방법
  - 자료 보관시 예상치 못한 상황에 대비하여 백업 DB를 관리하고 있음.

(7) 개선 사이클

**■ 적용가능성/유연성**

- 다른 정보를 수집하기 위해 사용하고 있는 통계패키지 와 절차, 새로운 수요와 정보요구에 대응하는 시간의 적절성에 대한 평가
  - 만족스럽지 못하고 자료입력시스템 개선이 필요함.
- 패널조사와 같이 특수한 조사를 유연하게 실시하기 위한 자료수집기반이 있는지 여부
  - 영구적인 자발적 조사를 신축성 있게 실시하기 위한 자료수집 틀은 없음.
- 변화하는 수요에 신속히 대응하기 위한 부담과 비용의 정도
  - 변화하는 수요에 신속히 대응하기 위한 부담은 받아들일 수 있으나, 비용은 거

의 줄일 수 없음.

■ **관련분야 전문성**

- 조사담당직원의 지식과 이해력 보유 정도
  - 주제 분야별로 원천자료, 개념 및 분류, 현행 및 향후 이슈, 조사관련 필요한 운용방법 등을 잘 이해하고 우수한 지식을 가진 전문적인 직원이 있음.
- 직원들의 교육과정에 대한 평가
  - 현재 직원의 훈련정도는 충분하다고 생각하지만 개선이 필요함.
- 주요 담당자들의 변동에 업무공백 없이 조사를 담당할 만한 전문성을 유지하고 있는지 여부
  - 주요 담당자들의 변동에 대해 전문성이 부족함.
- 다른 기관이나 전문가 협회와의 교류정도
  - 다른 기관이나 전문가 학회 등과의 교류가 활발함.

■ **품질관리**

- 조사에 대한 정기적인 점검여부
  - 정기적인 모니터링이 모든 통계 생산과정에서 검토되고 있음.
- 효율성, 시의성 또는 품질확보를 위한 이용하거나 사용하고 있는 최신 IT 기술의 보유여부
  - 대부분의 절차에 대해 IT 기술로 전환해야할 필요가 있음.
- 기존의 모든 원천자료를 광범위하게 검토할 필요가 있는지 여부
  - 조사의 효율성을 가능한 제고시키기 위하여 기존의 모든 원천자료를 광범위하게 검토할 필요가 있음.
- 통계생산과정의 효율성을 전문가에게 검토를 받고 있는지 여부
  - 독립적인 전문가에 의해 검토 받고 있음.
- 예산이 20% 많거나 적을 때 조치방법
  - 예산의 확대를 통해 현재 보다 많은 표본수를 유지할 수 있으며, 만일 20%이하

로 작다면 표본수 축소와 같은 조사 설계를 변경해야함.

### 3) 개선아이디어 기입란

#### □ 1장 : 조사수행을 위한 의사결정

현재 주요이용자들에 대한 주요 정보의 관리가 미흡하며, 향후 주요 이용자들에 대한 이용자 만족도 조사를 실시하고, 이들의 요구를 반영하도록 해야함. 이를 위해 자료 이용을 위한 요구가 있을 경우 이용자들의 정보를 체계적으로 관리할 필요가 있음.

#### □ 2장 : 조사 설계

2005년 현재 국민건강 영양 조사는 횡단면적인 조사로서 표본설계의 내용이 다른 통계조사에 비해 체계적인 조사 설계가 이루어짐. 조사 자료에 대한 검증 체계가 적절히 이루어지고 있는 반면에 조사무응답에 따른 조사항목에 대한 결측 자료의 대체과정에 이루어지고 있지 않음. 향후 보다 과학적인 방법을 이용하여 항목무응답에 대한 결측치 대체가 필요할 것으로 봄. 또한 응답자 응답부담을 줄이기 위해 연동 표본설계의 방안을 고려해볼 수 있음.

#### □ 3장: 자료수집

2005년 국민건강영양조사는 “한국보건사회연구원”의 조사원들이 전국의 표본가구를 방문하여 면대면 조사에 의해 자료를 수집하였음. 자료수집과정에서 전문화된 조사원들이 매우 적절하게 조사를 수행하였으나, 전국조사이고 매우 전문적인 분야로서 조사관리의 어려움이 매우 큼. 정기적인 조사원관리를 위한 모니터링 또는 교육이 이루어져야 할 것임. 또한 조사기간이 3개월 정도로 매우 장기간의 조사를 수행하기 때문에 이에 대한 적절한 업무 배분이 이루어져야 할 것임.

#### □ 4장 :자료입력 및 자료처리

2005년 현재 국민건강영양조사의 자료는 수작업에 의한 자료입력방식을 채택하고 있으며, 에디팅 절차 또한 수작업으로 수행되고 있음. 이러한 자료입력 및 처리과정을 보다 개선할 필요가 있음. 즉, 컴퓨터를 이용한 조사 체계를 도입함으로써 입력 및 처리시간을 단축시킬 수 있으며, 에디팅 절차를 보다 과학적인 방법으로 개선할 필요가 있음. 한편 무응답에 대한 대체가 전혀 이루어지지 않고 있으므로 향후 이에 대한 개선이 요구됨.

#### □ 5장 : 자료 분석 및 통계결과의 품질

관련성은 이용자 만족도 조사를 실시하여 적절히 관리가 이루어지고 있음. 정확성은 표본설계당시의 오차를 관리하고 있으며, CV등의 통계를 산출하여 제시하고 있음. 시의성 및 정시성은 자료수집과 처리 및 분석 과정이 매우 장기간 소요됨에 따라 개선이 필요한 부분임. 비교가능성 부분에서는 연간단위 통계와 일부 비교가능하지만 매우 제한적임. 일관성 측면에서는 동시에 다른 주기로 조사하고 있지 않음으로 측정성이 매우 제한적임. 자료의 보안관리측면에서 개인의 비밀이 노출되지 않도록 철저한 관리가 요구됨.

#### □ 6장 : 문서화 및 자료의 공표

보고서 상에 품질관련 지표를 함께 수록하도록 해야함. 특히 국민건강영양조사는 조사수행 보고서를 별도로 작성하여 발간함으로써 이용자들의 편의를 도모하고 있으나, 품질관련 지표를 함께 수록하여 제공하는 것이 바람직함. 한편 원시자료의 제공과정에서 특정 파일형식만을 제공하고 있으나, 보다 이용확대를 위해 다양한 형태의 자료파일을 제공할 필요가 있음.

#### □ 7장 : 개선 사이클

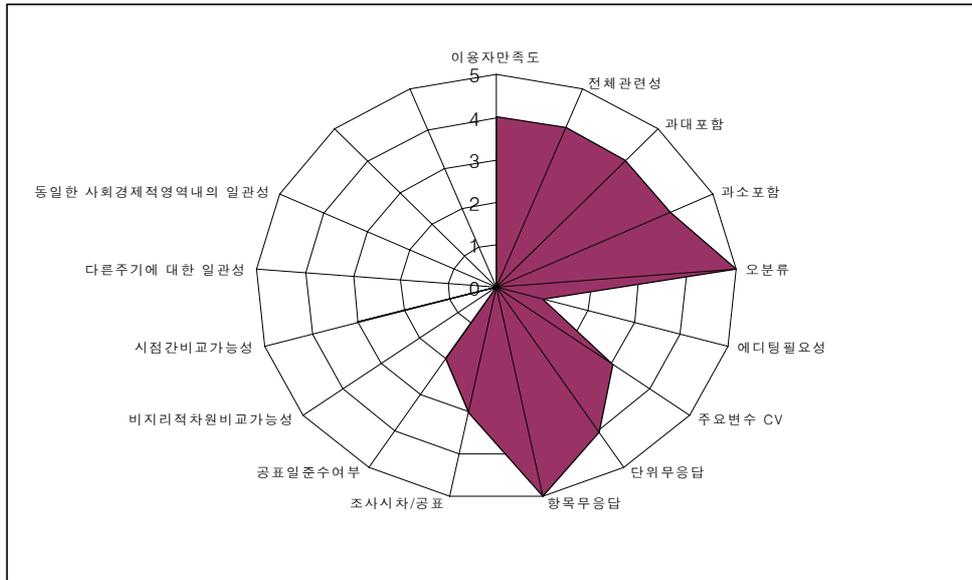
자료처리과정의 대부분이 수작업으로 이루어지고 있으므로 향후 IT 기술을 도입하여 처리기간을 단축함으로써 빠른 시일 내에 자료를 공표할 수 있는 체계를 도입할 필요가 있음. 마지막으로 국민건강영양조사 자료는 보건 분야에서 대표적인 조사통계자료 이므로 관련 직원들에 대한 적절한 교육과 정기적인 모니터링을 통해 조사품질의 개선을 위한 노력이 투입되어야 할 것임.

#### 4) 품질차원별 점수 및 진단 다이어그램

〈표 5-2-2〉 국민건강영양조사의 품질차원별 점수

| 품질차원    | 내 용               | 항 목  | 점 수 |
|---------|-------------------|------|-----|
| 관련성     | 이용자만족도            | 5-2  | 4   |
|         | 전체관련성             | 5-3  | 4   |
| 정확성     | 과대포함              | 2-6  | 4   |
|         | 과소포함              | 2-7  | 4   |
|         | 오분류               | 2-8  | 5   |
|         | 에디팅 필요성           | 4-4  | 1   |
|         | 주요변수 CV           | 5-6  | 3   |
|         | 단위무응답             | 5-15 | 4   |
|         | 항목무응답             | 5-18 | 5   |
| 시의성/정시성 | 조사시차/공표           | 5-21 | 3   |
|         | 공표일 준수여부          | 5-22 | 2   |
| 비교가능성   | 비지리적차원 비교가능성      | 5-24 | -   |
|         | 시점간 비교가능성         | 5-25 | 3   |
| 일관성     | 다른 주기에 대한 일관성     | 5-27 | -   |
|         | 동일한 사회경제적영역내의 일관성 | 5-28 | -   |

[그림 5-2-2] 국민건강영양조사의 품질진단 다이어그램



## 2. 보고통계 - 부동산 거래 현황 통계

### 1) 개요

#### 가) 연혁

부동산거래 현황 통계는 토지와 건축물, 아파트의 거래동향을 파악하여 부동산 시장의 경기상황을 보여주는 중요한 통계이며, 아울러 부동산정책의 기초 자료로 활용하기 위해 통계를 작성한다. 건설교통부에서 1991년부터 토지거래에 관한 정보를 매월 수집하여 토지거래 통계를 발표하여 왔다. 그동안 토지거래통계는 부동산 투기억제 시책 등 토지정책의 효율성을 제고하고 토지시장과 거시경제지표와의 상관분석을 위한 기초 자료로 활용되었으며, 2006년 1월 1일부터 실거래가격을 의무적으로 신고하는 부동산거래 신고제를 공자로 하는 「공인중개사의 업무 및 부동산거래신고에 관한 법률」의 시행으로 토지뿐만 아니라 건축물과 아파트거래 통계 작성이 가능해 졌다. 이와 함께 통계의 명칭 또한 『토지거래통계』에서 건축물과 아파트거래까지를 포함하는 『부동산거래통계』로 변경되었다. 이후 2006년 8월 현재까지 3가지 부동산거래통계 중에서 토지거래통계만 공표되었으며, 건축물과 아파트거래통계는 보도 자료로만 발표되었다.

부동산거래신고제의 시행으로 부동산거래관리시스템(RTMS)이 구축되어 인터넷상으로 거래신고를 할 수 있도록 시스템이 갖추어져 있으며, 실시간 통계처리가 가능해졌다. 현재 까지 시스템 초기단계로서 입력, 거래승인, 통계분석 과정에서 비교적 안정된 것으로 나타나고 있으며, 향후 지속적인 기능 향상이 이루어질 것이다.

#### 나) 통계작성목적

부동산 거래 현황 통계는 다음을 목적으로 한다.

- 토지와 건축물, 아파트의 거래동향을 파악하여 부동산 경기 동향 파악

- 부동산 정책의 기초 자료로 활용
- 토지거래통계의 경우 부동산 투기억제시책 등 토지정책의 효율성 제고
- 토지시장과 거시경제지표와의 상관분석을 위한 기초 자료로 활용

#### 다) 부동산거래 현황 통계의 작성체계

부동산거래 현황 통계의 작성 체계를 살펴보면, 통계생산의 주체는 건설교통부로 되어있으며, 통계생산에 관여하는 기관은 일선 지방자치단체, 건설교통부, 한국토지공사로 분산되어있다. 먼저 일선 지방자치단체에서는 부동산 거래관리시스템을 운영하고, 부동산거래 기초 자료의 입력을 담당하고 있다. 부동산거래 자료를 RTMS에 입력하고, 거래 마감처리를 한후 건설교통부에 자료를 송부하게 된다. 다음으로 건설교통부에서는 부동산거래관리시스템을 관리하고, 시군구에서 마감한 부동산거래 전국자료를 수집한다. 이렇게 집계된 자료를 한국토지공사에서 부동산 거래관련 통계를 분석하고, 부동산동향조사와 부동산거래통계를 건설교통부에 통보하게 되며, 마지막으로 건설교통부에서는 최종적으로 부동산거래통계 자료를 확인하고, 통계자료를 발표하는 절차로 구성된다.

부동산거래통계 중 토지통계와 건축물 통계는 용도지역별, 지목별, 매입자거주지별, 거래처주체별, 거래규모별 등 10가지 유형으로 세분화하여 집계하고 있으며, 아파트통계는 지역별, 단지별, 거래가격별 등 3가지 유형으로 집계할 예정이다.

#### 라) 보고대상

부동산거래 현황 통계의 보고단위는 개인이며, 모집단은 「공인중개사 업무 및 부동산거래신고에 관한법률」 제27조(부동산거래신고의무)에 의한 거래신고대상 및 의무자이다. 대상범위는 국내 부동산거래당사자(매수인, 매도인, 중개업자)를 포괄하며, 대상규모는 부동산거래가 발생한 모든 개인이다. 보고기준시점은 매월말일이며, 보고기간은 매월1일부터 매월 말일까지이다. 부동산거래의 보고방법은 전산망을 이용하며, 보고자는 공무원과 부동산거래당사자이다.

## 2) 자체진단 점검사항

### (1) 조사 수행을 위한 의사 결정

#### ■ 이용자 요구분석

##### 조사수행의 법적근거

- 부동산거래현황 통계는 정부승인 통계로서 국내법인 『공인중개사의 업무 및 부동산거래 신고에 관한 법률』 제 27조에 따라 거래신고대상 및 의무자가 부동산거래관리시스템(RTMS)를 통해 신고하도록 함.

##### 조사에 대한 이용자 정보

- 주요이용자의 일부정보만을 알고 있음.

##### 주요이용자

- 1순위) 정부기관, 2순위) 단체 및 협회, 3순위) 기업, 4순위) 금융기관 5순위) 언론기관

[그림 5-2-3] 부동산 거래 현황 통계의 주요이용자 순위



##### 주요이용자 정보의 관리

- 주요 이용자에 대한 정보는 별도의 파일로 관리하고 있음(DM 발송을 위한 이용자 정보 파일).

##### 이용자에 대한 자문을 위한 절차

- 관련된 이용자와 비공식적인 접촉을 통해 자문을 받음, 즉 시군구의 건의사항 접수후 처리.
- RTMS 구축관련 의견 조회(중개인, 시군구 담당자 교육시 이용자 요구사항 접수)
- 향후 이용자의 수요를 예측을 위한 절차
  - 중앙 또는 지방자치단체간의 양자협의회 또는 미충족 이용자 요구사항을 교육시 파악.
- 미충족 이용자수요
  - 지방자치단체(시군구) 공무원 및 공인중개사에 대한 미충족 이용자 수요가 있음.
- 주요 이용자 수요에 적절히 대응하기 위한 계획
  - 지역별, 주택유형별, 주택가격별, 평형별 주택거래 정보요구에 대응하기위한 적절한 계획을 수립.
- 조사(작성)목적의 명문화
  - 부동산거래현황통계 기본현황에 명문화되어있으며, 홈페이지(<http://www.iklc.co.kr>)에 명시하고 있음.

## (2) 조사 설계

### ■ 조사개념

- 비교 가능한 공식/비공식 조사
  - 비교 가능한 공식적이거나 비공식적인 통계가 있음.
  - 국민은행 부동산 통계 및 각종 부동산 관련 통계
- 표준화된 체계, 기준, 개념, 정의의 분류기준
  - 건설교통부에서 제시하는 개념과 기준에 따라 용도별, 지목별 등으로 분류하고 있음.
- 통계단위에 대한 오분류 여부
  - 통계조사 단위에 대한 오분류는 거의 없음.

### (3) 자료수집

#### ■ 자료의 원천

- 조사에 사용된 자료수집 방법
  - 법적인 요구사항
- 자료수집방법
  - 행정자료나 등록자료
  - ④번으로 이동함
- 자료의 유지관리
  - 기관의 내부부서에서 관리함.
- 자료의 품질평가여부
  - 스스로 평가할 수 있음.
- 다른 자료와 연계필요성
  - 다른 자료와 연계할 필요는 없음
- 다른 자료원을 사용할 경우의 문제점
  - 갱신 및 시의성에 문제가 있음.
- 자료수집 및 교환과정에서 보안성 정도
  - 받아들일 만 하지만 개선할 필요가 있음.
- 중복단위와 다중단위의 구분
  - 중복단위와 다중단위는 자동적으로 구분됨.

#### ■ 현장실사

- 행정자료나 등록 자료의 효율적 사용가능성
  - RTMS시스템에 자동적으로 등록되어 정보를 효율적으로 사용할 수 있음.
- 법률적인 측면에서 행정자료나 등록 자료의 사용제한 범위
  - 법적인 제한이 매우 심함.
- 실사기간이 조사대상기간을 완전히 포함하는지 여부

— 완전히 포함하고 있음.

입력 자료의 평가

— 받아들일 만 하지만, 개선이 필요함.

마이크로 자료의 보안 유지 정도

— 보고되는 자료는 중앙에서 가공하여 처리하고, 마이크로자료는 제공하지 않음.

(4) 자료입력 및 자료처리

■ 자료입력

자료입력방식

— 수작업에 의한 자료입력 방식 채택.

자료코딩방식

— 입력과 동시에 자동 저장됨.

■ 에디팅 절차

원시자료 에디팅의 필요성

— 원시자료에서 많은 오류가 발견되었고, 점검은 필수적임.

에디팅 작업방식

— 수작업에 의한 에디팅

채택하고 있는 에디팅 작업에서 이용하고 있는 자료

— 자료수집 절차에 따라 입력된 DB 자료를 이용함.

에디팅 절차에 대한 사전 점검 여부

— 사전 점검은 이루어지지 않음.

원시자료에 적용하고 있는 에디팅 절차

— 에디팅 과정에서 논리적 점검 및 통계적 검증의 혼합방식을 통한 에디팅 방법을 적용하고 있음.

수작업을 통한 에디팅 담당자에 문서화된 에디팅 지침서의 제공여부

— 에디팅 지침서는 별도로 구비하고 있지 않음.

- 결측치 또는 0 값 등의 개념에 대한 구분
  - 결측치, 0 값 등의 개념에 대한 구분은 없음.
- 에디팅 절차의 우선순위
  - 지자체 마감전 오류검증, 기본사항 미입력 자료 및 토지건축물 면적등 오류자료 검증, 오류자료 지자체에 대한 유선 또는 방문을 통한 오류확인, DB 자료의 수정.
- 과도한 에디팅이나 자료에 새로운 오류방지를 위한 조치
  - 과도한 에디팅을 방지하는 대책을 가지고 있지 않음.
- 에디팅 작업의 지속여부
  - 마감자료에 대한 에디팅 작업을 수행함.
- 에디팅 절차의 개선을 위해 구축해 놓은 과거 조사오류에 대한 이용가능 통계
  - 에디팅 과정을 개선하기 위하여 구축해 놓은 이전 조사의 오류에 대한 통계는 가지고 있지 않음.
- 조사의 품질을 측정하고 향후 조사개선을 위한 오류에 대한 통계 산출
  - 조사의 품질을 측정하고 향후 조사개선을 위한 오류에 대한 통계는 산출하지 않음.
- 이용하는 에디팅 시스템의 신축성
  - 새로운 에디팅 과정이나 수정된 에디팅 과정을 적용할 경우 시간과 비용이 소요됨.
- 오류원인에 대한 정보 보유
  - 오류의 원인에 대한 정보는 보고과정 및 입력과정에서 발생함.

**■ 대체 절차**

- 에디팅 과정에서 발견된 오류의 수정방법과 에디팅 비율
  - 에디팅 과정에서 발견된 오류는 조사원 또는 응답자를 재 접촉하여 수정함.
- 통계값에 대한 대체여부
  - 통계 값에 대한 대체는 적용하지 않음.

→ 5장으로 이동.

#### (5) 자료 분석 및 통계결과의 품질

##### ■ 관련성

- 조사에 대한 이용자 만족도 측정 방법
  - 이용자 피드백에 의한 체계적인 진단 방법을 사용하고 있음
- 주요이용자 만족도에 대한 이용 가능한 정보의 평가
  - 이용자 만족 정도를 알 수 있는 정보는 주요 이용자 정보는 많이 있으며, 다른 이용자 정보는 일부만 있음.
- 조사의 전반적인 관련성 평가
  - 주요 분야는 관련성이 높고, 다른 분야는 약간의 관련성이 있음.

##### ■ 정확성

- 추정값의 정도를 측정하기위해 제시하는 값
  - 어느 것도 제시하고 있지 않음.
- 조사에서 단위무응답의 수준
  - 낮은 단위 무응답률을 보임.
- 무응답을 세부범주로 분류
  - 무응답을 분류하고 있지 않음.
- 항목무응답의 평가
  - 항목 무응답은 낮은 편임(5%이하).

##### ■ 시의성 및 정시성

- 조사기간과 발간 시점간의 시차
  - 조사기간과 조사결과의 발표 시차는 약 20일 임.
- 계획한 발간일자와 실제 발간일자 간의 시차
  - 거의 시차가 없음.

- 조사대상기간과 잠정 또는 최종 결과의 최초발간일 간의 시차
  - 조사대상기간과 잠정 또는 최종 결과의 최초발간과의 시차는 정해진 시간 정도임.
- 주요발간물에 대해 정해진 발간시기 준수여부
  - 발간물의 80%이상을 준수함.
- 지식성을 개선하기 위한 변화의 필요성
  - 외부협력기관이 관련된 변화가 필요함.

■ 비교가능성

- 통계의 중요한 개념과 조사 설계에 연계하여 지역에 대한 비교가능성 평가
  - 연관된 통계 결과와의 비교 가능성은 약간의 차이가 있음.
- 조사대상기간과 관련하여 시계열적인 비교가능성 평가
  - 시계열적인 비교가능성은 거의 제한이 없음.

■ 일관성

- 월별 또는 연간단위로 동시에 다른 주기로 조사하는지 여부
  - 동시에 다른 주기로 조사하고 있지 않음.
  - 6번으로 이동.

■ 자료분석

- 이용자 맞춤분석 제공시 대응능력
  - 자료 분석에 대한 이용자들의 요구를 만족시킬 의무는 없음.
- 사용하고 있는 추가분석
  - 지수산출과 계절 또는 추세 분석과 추정임.
- 작년한해 동안 자료 분석을 위해 새롭게 도입했거나 기존의 방법을 개편한 적이 있는지 여부
  - 개편한 적이 없음.

■ 보안관리

- 이용자에 대한 자료접근을 위해 보유하고 있는 자료의 형태
  - 다량의 통계표를 보유함.

(6) 문서화 및 자료의 공표

■ 메타자료의 문서화

- 행정 등록자료를 사용할 경우 관련된 적절한 문서 보유여부
  - 보유하고 있음.
- 통계산출과정에서 문서화하는 내용
  - 변수의 형태로서는 다른 변수들의 계산으로 구한 변수, 외부자료로부터 가져온 변수, 변수의 통계적 특성(즉, 범주적, 분류적, 계량적), 측정단위(의미를 가진 경우), , 외부자료에서 가져올 경우 그에 대한 서술, 파생 변수의 경우 계산방법 서술임.
  - 분류 서술로서는 표준 또는 일반 분류가 사용될 경우 참조하는 분류, 특정한 분류가 적용된 경우 분류의 출처, 참조된 분류의 일부항목을 처리하는데 특정 주의사항을 설명하는 정보를 포함함.
  - 개념과 통계적 단위의 정의에는 정의된 개념의 제목, 정의의 출처, 정의와 그 개념을 측정하는 변수간의 연계사항을 포함함.
  - 통계산출과정에서 사용된 조사방법 및 기법을 포함함.
- 통계작성 절차에 필요한 행정 및 처리절차에 대한 체계적인 문서의 보유여부
  - 통계작성절차를 지원하는데 필요한 행정과정 및 처리절차, 예를 들면 예산의 확보와 효율적 관리, 조사표 인쇄, 직원채용, 지방행정기관 방문, 언론매체접촉 등에 관한 모든 사항을 체계적으로 문서화하고 있음.
- 품질기준 중 문서에 포함된 내용
  - 시의성 및 정시성 항목에서 대상기간 완료시점과 최종결과가 산출된 시점간의 평균시차 등을 문서에 포함하고 있음.

— 비교가능한 시계열의 수 및 길이.

자료 이용시 예상되는 제한점에 대해 체계적인 문서의 보유여부

— 자료 이용시 예상되는 한계에 대해 체계적으로 문서화 하고 있음.

발간물에 대한 외부이용들의 만족도에 관한 정보

— 발간물에 대한 외부 이용자들의 만족도 정보는 보유하고 있지 않음.

내부직원들의 만족도에 대한 정보

— 내부 직원들의 만족도 정보를 보유하고 있지 않음.

보다 나은 발간물을 위해 필요한 항목

— 자료처리작업 및 품질관리항목을 보완할 필요가 있음.

#### ■ 자료제공 전략

생산된 자료에 대한 정보의 검색을 위해 이용자 측면에서 목록시스템과 정보서비스의 제공여부

— 이용자에게 편리하도록 기관의 발간목록 시스템과 가장 최근에 생산되고 저장된 자료에 대한 정보를 검색할 수 있도록 인터넷 홈페이지에서 정보서비스가 되고 있음.

통계작성기관에서 이용자에게 적합한 형식으로 자료를 이용할 수 있도록 하는 자료제공시스템 보유여부

— 기관수준의 자료제공 시스템은 이용자에게 적합한 형식으로 액세스하게 하고 있음.

통계표와 그래프 작성을 위해 사용하고 있는 기준의 여부

— 테이블과 도표 작성에 사용된 표준화된 기준이 있음.

개편이나 잠정치 또는 조기 추계치 등을 설명하고 있는지 여부와 제한사항이나 잠정자료에 대한 기대수준 등의 정보가 제공되고 있는지 여부

— 개편이나 잠정 또는 조기추계치 등을 분명히 표시하여 설명하는지와 제한사항이나 잠정치에 대한 기대수준 등에 대한 어떠한 정보도 제공되지 않음.

■ 자료관리

- 자료의 관리와 보관을 위한 체계적인 방법의 보유여부
  - 자료를 관리하고 축적하기 위해 자료를 DB화 하여 관리하고 있음.
- 자료 보관시 예상치 못한 상황에 대비하여 자료를 보호하는 방법
  - 자료 보관시 예상치 못한 상황에 대비하여 백업 DB를 관리하고 있음.

(7) 개선 사이클

■ 적용가능성/유연성

- 다른 정보를 수집하기 위해 사용하고 있는 통계패키지 와 절차, 새로운 수요와 정보요구에 대응하는 시간의 적절성에 대한 평가
  - 만족스럽지 못하고 자료입력시스템 적절성을 개선할 필요함.
- 패널조사와 같이 특수한 조사를 유연하게 실시하기 위한 자료수집기반이 있는지 여부
  - 영구적인 자발적 조사를 신축성 있게 실시하기 위한 자료수집 틀이 있음.
- 변화하는 수요에 신속히 대응하기 위한 부담과 비용의 정도
  - 변화하는 수요에 신속히 대응하기 위한 부담은 받아들일 수 있으나, 비용은 거의 줄일 수 없음.

■ 관련분야 전문성

- 조사담당직원의 지식과 이해력 보유 정도
  - 주제 분야별로 원천자료, 개념 및 분류, 현행 및 향후 이슈, 조사관련 필요한 운용방법 등을 잘 이해하고 우수한 지식을 가진 전문적인 직원이 있음.
- 직원들의 교육과정에 대한 평가
  - 현재 직원의 훈련정도는 충분하다고 생각하지만 개선이 필요함.
- 주요 담당자들의 변동에 업무공백 없이 조사를 담당할 만한 전문성을 유지하고 있는지 여부
  - 주요 담당자들의 변동에 대해 전문성이 부족함.

- 다른 기관이나 전문가 협회와의 교류정도
  - 다른 기관이나 전문가 학회 등과의 교류가 활발함.

■ 품질관리

- 조사에 대한 정기적인 점검여부
  - 시군구 담당자와 부동산 중개사에 대한 정기적인 교육이 이루어지고 있음.
- 효율성, 시의성 또는 품질확보를 위한 이용하거나 사용하고 있는 최신 IT 기술의 보유여부
  - 대부분의 절차에 대해 IT 기술을 사용하고 있음.
- 기존의 모든 원천자료를 광범위하게 검토할 필요가 있는지 여부
  - 조사의 효율성을 가능한 제고시키기 위하여 기존의 모든 원천자료를 광범위하게 검토할 필요는 없음.
- 통계생산과정의 효율성을 전문가에게 검토를 받고 있는지 여부
  - 독립적인 전문가에 의해 검토 받고 있지 않음.
- 예산이 20% 많거나 적을 때 조치방법
  - 예산의 확대를 통해 이력 및 분석 시스템을 개선할 수 있음.

3) 개선아이디어 기입란

1장 : 조사수행을 위한 의사결정

주요이용자들에 대한 정보 중 일부만 관리되고 있으며, 이용자 의견수렴 방법이 주로 시군구 담당자들의 건의사항 관리에 치우쳐 있음. 자료의 이용실태를 파악하여 적절한 이용자관리를 위해 미충족 이용자 요구에 대한 반영정도를 높힐 필요가 있음.

2장 : 조사 설계

부동산 거래현황 통계는 보고통계로서 별도의 조사 설계는 이루어지고 있지 않고, 부동산 거래가 발생한 전국적인 통계를 산출함으로써 전수 조사의 성격을 갖추고 있음.

### □ 3장: 자료수집

RTMS시스템이 등록된 자료를 자동적으로 수집하여 통계를 산출하고 있음. 자료 수집과정에서 부동산 중개사의 입력과 시군구 공무원들의 자료 취합에 의해 자료가 수집됨으로 자료의 입력과정에서 발생할 수 있는 입력오류에 대한 철저한 관리가 요구됨.

### □ 4장 :자료입력 및 자료처리

수작업에 의한 자료입력방식을 채택하고 있으며, 에디팅 절차 또한 수작업으로 수행되고 있음. 즉 오류가 발생한 자료에 대해서는 시군구 공무원이 입력자료에 대해 유무선 전화를 통해 오류를 확인하고 보완하는 절차를 밟고 있기 때문에 보다 철저한 관리가 요구됨. 특히 오류별로 통계를 산출하여 특정오류를 감소시킬 수 있는 방안을 모색할 필요가 있음.

### □ 5장 : 자료 분석 및 통계결과의 품질

관련성은 이용자 만족도 조사를 실시하여 적절히 관리가 이루어지고 있음. 정확성은 전수 자료이므로 어떠한 지표도 산출하고 있지 않으며, 무응답을 별도로 분류하고 있지 않음. 시의성 및 정시성은 자료수집과 처리 및 분석 과정이 매우 빠르게 진행되며, 약 20일 정도의 소요시간을 고려할 때 적절하다고 판단됨. 비교가능성 부분에서는 연간단위 통계와 일부 비교가능하며, 타 통계자료와의 비교도 가능함. 일관성 측면에서는 동시에 다른 주기로 조사하고 있지 않음으로 측정이 매우 제한적임. 자료의 보안관리측면에서 개인의 비밀이 노출되지 않도록 철저한 관리가 요구

됨.

**□ 6장 : 문서화 및 자료의 공표**

보고서 상에 품질관련 지표를 함께 수록하도록 해야함. 통계산출과정을 문서화 하여 제공할 필요가 있음.

**□ 7장 : 개선 사이클**

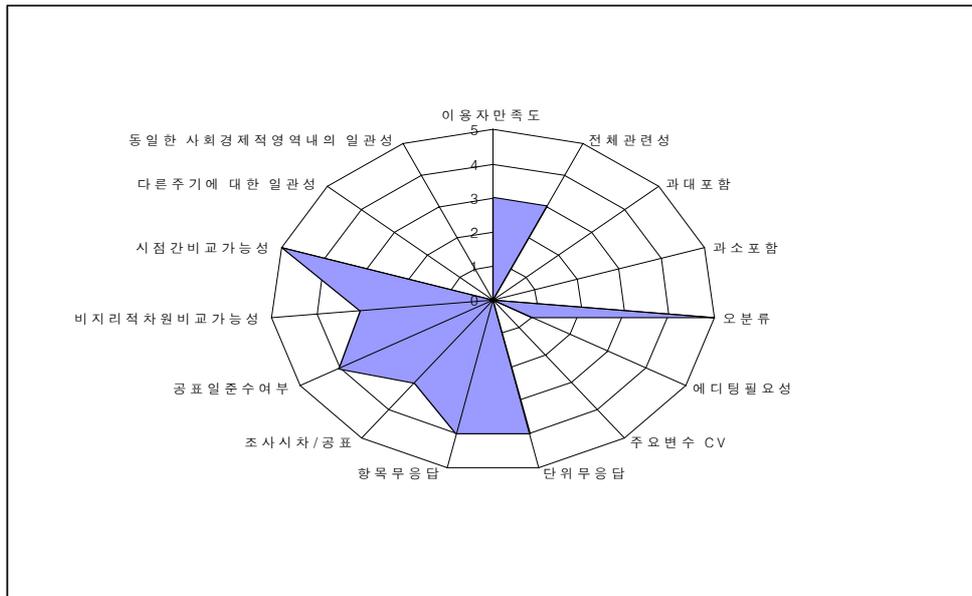
자료입력과정이 RTMS 시스템에 자료를 입력하고 있으며, 전국적인 네트워크로 관리가 가능하며, 향후 개선 및 보안사항이 발생할 경우 보다 적극적인 시스템관리가 요구됨. 즉, 현재는 토지거래만을 입력하도록 하고 있으나, 향후 주택거래관련정보를 함께 처리하도록 하고 있기 때문에 시스템의 안정성을 검토할 수 있는 별도의 관리체계를 마련할 필요가 있음.

4) 품질차원별 점수 및 진단 다이어그램

〈표 5-2-3〉 부동산 거래 현황통계의 품질차원별 점수

| 품질차원    | 내 용               | 항 목  | 점 수 |
|---------|-------------------|------|-----|
| 관련성     | 이용자만족도            | 5-2  | 3   |
|         | 전체관련성             | 5-3  | 3   |
| 정확성     | 과대포함              | 2-6  | -   |
|         | 과소포함              | 2-7  | -   |
|         | 오분류               | 2-8  | 5   |
|         | 에디팅필요성            | 4-4  | 1   |
|         | 주요변수 CV           | 5-6  | -   |
|         | 단위무응답             | 5-15 | 4   |
|         | 항목무응답             | 5-18 | 4   |
|         | 조사시차/공표           | 5-21 | 3   |
| 시의성/정시성 | 공표일준수여부           | 5-22 | 4   |
|         | 비지리적차원비교가능성       | 5-24 | 3   |
| 비교가능성   | 시점간비교가능성          | 5-25 | 5   |
|         | 다른주기에 대한 일관성      | 5-27 | -   |
| 일관성     | 동일한 사회경제적영역내의 일관성 | 5-28 | -   |

[그림 5-2-4] 부동산 거래 현황통계의 품질진단 다이어그램



### 3. 가공통계 - 공장등록현황 통계

#### 1) 개요

##### 가) 주요특징

2003년부터 『산업집적활성화 및 공장설립에 관한법률』 시행령 제8조의 2항에 의거하여 공장설립 관련 승인 등의 절차를 수행할 때는 지체 없이 그 자료를 공장설립정보망에 입력하도록 하고 있다. 2003년 이전에는 법률로 명시되어 있지 않아 승인관련 내용이 누락되거나 잘못 입력된 경우도 발생하고 있다.

공장등록 통계는 한국산업단지공단이 FEMIS 시스템을 통해 입력된 공장등록 자료를 기반으로 가공된 통계를 생산 공표하고 있는 가공통계이다. 공장등록 통계의 작성주기는 반기이고, 작성기준시점은 매년 1월과 7월말이며, 1월은 전년도 12월 31까지 FEMIS에 등록된 사업체에 대한 통계를 작성하고, 7월말에는 익년 6월 30일 까지 등록된 사업체를 대상으로 통계를 작성하고 있다.

공장등록 통계의 조사표인 공장등록현황 조사서는 공장등록의 기본사항과 사업계획내역, 생산시설현황에 대한 항목으로 구성되며, 공장등록기본사항에는 설립일자, 산업단지명, 입주형태 등 36가지의 문항으로 구성되고, 사업계획내역에는 홈페이지주소, 외국인 투자비율, 가동률 등 16가지의 항목으로 구성되어 있다.

공장등록통계현황은 공장설립관리정보망의 등록자료 및 통계내역을 분석하여 산업집적 및 공장설립에 관한 정책 자료로 이용되며, 『FEMIS 공장등록통계현황』으로서 인쇄물과 DB 사이트인 <http://www.femis.go.kr>에서 자료를 제공하고 있다.

##### 나) 통계작성목적

공장등록현황통계는 지역별, 규모별, 업종별 등의 변화추이를 분석함으로써 기업의 동향 및 유형을 파악하여 통계를 작성함으로써 산업집적 및 공장설립에 관한 정

책 자료로 활용하는 것으로 목적으로 한다.

#### 다) 공장등록 현황 통계의 작성체계

공장등록통계현황의 작성과정을 살펴보면 먼저 『산업집적활성화 및 공장설립에 관한법률』에 따라 시장 군수 또는 구청장이나 관리기관은 산업자원부장관이 고시하는 사항에 대한 승인 등을 할 때에는 지체 없이 그 승인에 대한 자료를 제 1항에 의 규정에 의해 공장설립관리 정보망에 입력, 관리하여야 한다. 따라서 235개 지자체의 공장등록관련 담당직원에 의해 사업계획서가 입력되며, 입력과 공표절차는 입주기업체가 지자체에 등록하고, 지자체담당자가 내용을 입력하며, 한국산업단지공단 본부의 공장설립지원팀에서 자료를 가공하고 분석하여 통계를 공표하고 있다.

#### 라) 투입 산출 지표

공장등록통계의 보고단위는 사업체이며, 투입자료는 공장설립관리정보망(FEMIS)의 등록 자료이다. 주요결과지표들로는 신규공장등록현황, 공장설립신규승인현황, 공장등록현황, 고용현황 등이다.

### 2) 자체진단 점검사항

#### (1) 조사 수행을 위한 의사 결정

##### ■ 이용자 요구분석

##### 조사수행의 법적근거

— 국내법으로서 『산업집적활성화 및 공장설립에 관한 법률』에 의해 공장설립관리 정보망에 입력, 관리하여야 함.

##### 조사에 대한 이용자 정보

— 주요이용자의 일부정보만을 알고 있음(해당공무원).

##### 주요이용자

- 1순위) 정부기관, 2순위) 기업, 3순위) 단체 및 협회, 4순위) 금융기관 5순위) 언론기관

[그림 5-2-5] 공장등록 현황 통계의 주요이용자 순위



- 주요이용자 정보의 관리
  - 주요 이용자에 대한 정보는 별도의 파일로 관리하고 있음(DM 발송을 위한 이용자 정보 파일).
- 이용자에 대한 자문을 위한 절차
  - 관련된 이용자와 비공식적인 접촉을 통해 자문을 받음, 즉 시군구 지자체 공무원들의 건의사항 접수후 처리.
  - FEMIS 이용관련 의견 조회.
- 향후 이용자의 수요를 예측을 위한 절차
  - 어떠한 절차도 거치지 않음.
- 미충족 이용자수요
  - 어떠한 미 충족욕구가 있는지 파악하지 못함.
- 주요 이용자 수요에 적절히 대응하기 위한 계획
  - 향후 통계자문위원회를 구성하여 이용자수요에 대응할 계획임.
- 조사(작성)목적의 명문화
  - 공장등록통계현황 발간물에 명문화되어있으며, 관련 홈페이지 (<http://www.e-cluster.net>)에 명시하고 있음.

## (2) 조사 설계

### ■ 조사개념

#### 비교 가능한 공식/비공식 조사

- 비교 가능한 공식적이거나 비공식적인 통계가 있음.
- 통계청의 광공업 동태조사와 비교 가능하지만, 지수산출방식이 서로 다르기 때문에 주의할 필요가 있음.

#### 조사의 품질을 개선에 활용할만한 공식/비공식 통계의 유무

- 통계청의 광공업 동태조사

#### 표준화된 체계, 기준, 개념, 정의의 분류기준

- 국가적인 기준을 적용하고 있음(표준산업분류 )

#### 표준화된 분류기준과 맞지 않는 사항

- 표준산업분류를 사용하고 있기 때문에 분류기준에 맞지 않는 사항은 없음.

#### 통계단위의 오분류 유무

- 잘못된 분류가 거의 없음.

## (3) 자료수집

### ■ 자료의 원천

#### 조사에 사용된 자료수집 방법

- 법적인 요구사항(『산업집적활성화 및 공장설립에 관한 법률』)

#### 자료수집방법

- 행정자료나 등록자료

→ ④번으로 이동함

#### 자료의 유지관리

- 기관의 내부부서에서 관리함.

#### 자료의 품질평가여부

— 스스로 평가할 수 있음.

다른 자료와 연계필요성

— 다른 자료와 연계할 필요는 없음

다른 자료원을 사용할 경우의 문제점

— 정의와 갱신 및 시의성에 문제가 있음.

자료수집 및 교환과정에서 보안성 정도

— 받아들일 만 하지만 개선할 필요가 있음.

중복단위와 다중단위의 구분

— 중복단위와 다중단위는 구분되지 않음.

조사에서 등록 및 행정자료의 효율적 사용여부

— 공장 등록 자료를 가공함으로 현재 등록 자료를 사용하고 있음.

#### ■ 현장실사

행정자료나 등록자료의 효율적 사용가능성

— FEMIS시스템에 시군구 담당자가 공장등록서를 입력함으로 정보를 효율적으로 사용할 수 있음.

법률적인 측면에서 행정자료나 등록자료의 사용제한 범위

— 법률적인 제한이 매우 심함.

실사기간이 조사대상기간을 완전히 포함하는지 여부

— 완전히 포함하고 있음.

실사기간

— 공장설립 계획서의 입력과 동시에 현장 실사가 종료되며, 일정기간동안 입력된 자료를 마감하여 통계를 작성함으로 실사기간은 반기로 고려하는 것이 타당함.

입력 자료의 평가

— 받아들일 만 하지만, 개선이 필요함.

마이크로 자료의 보안 유지 정도

— 보고되는 자료는 중앙에서 가공하여 처리하고, 마이크로자료는 제공하지 않음.

#### (4) 자료입력 및 자료처리

##### ■ 자료입력

###### 자료입력방식

— 수작업에 의한 자료입력 방식 채택(공장설립계획서를 입력함).

###### 자료코딩방식

— 입력과 동시에 자동 저장됨.

###### 수작업코딩시 일관성 점검여부

— 종속적인 검증을 수행함(전년도 자료 또는 신규자료의 일관성 검증).

##### ■ 에디팅 절차

###### 원시자료 에디팅의 필요성

— 원시자료에서 많은 오류가 발견되었고, 점검은 필수적임.

###### 에디팅 작업방식

— 수작업에 의한 에디팅

###### 채택하고 있는 에디팅 작업에서 이용하고 있는 자료

— 자료수집 절차에 따라 입력된 과거DB 자료를 이용함.

###### 에디팅 절차에 대한 사전 점검 여부

— 사전 점검은 이루어지지 않음.

###### 원시자료에 적용하고 있는 에디팅 절차

— 에디팅 과정에서 논리적 점검 및 이상치 분석과 같은 통계적 검증의 혼합방식을 통한 에디팅 방법을 적용하고 있음.

###### 수작업을 통한 에디팅 담당자에 문서화된 에디팅 지침서의 제공여부

— 에디팅 지침서는 별도로 구비하고 있지 않음.

###### 결측치 또는 0 값 등의 개념에 대한 구분

— 결측치, 0 값 등의 개념에 대한 구분은 없음.

###### 에디팅 절차의 우선순위

— 지자체 마감전 오류검증, 기본사항 미입력 자료 검증, 오류자료발생 지자체에 대한 유선 또는 방문을 통한 오류확인, DB 자료의 수정.

과도한 에디팅이나 자료에 새로운 오류방지를 위한 조치

— 과도한 에디팅을 방지하는 대책을 가지고 있지 않음.

에디팅 작업의 지속여부

— 마감자료에 대한 에디팅 작업을 수행함.

에디팅 절차의 개선을 위해 구축해 놓은 과거 조사오류에 대한 이용가능 통계

— 에디팅 과정을 개선하기 위하여 구축해 놓은 이전 조사의 오류에 대한 통계는 가지고 있지 않음.

조사의 품질을 측정하고 향후 조사개선을 위한 오류에 대한 통계 산출

— 조사의 품질을 측정하고 향후 조사개선을 위한 오류에 대한 통계는 산출하지 않음.

이용하는 에디팅 시스템의 신속성

— 새로운 에디팅 과정이나 수정된 에디팅과정을 적용할 경우 시간과 비용이 소요 됨.

오류원인에 대한 정보 보유

— 오류의 원인에 대한 정보는 보고오류 및 입력오류임.

#### ■ 대체 절차

에디팅 과정에서 발견된 오류의 수정방법과 에디팅 비율

— 에디팅 과정에서 발견된 오류는 유무선 통화나 담당자를 재접촉하여 수정함.

통계값에 대한 대체여부

— 통계 값에 대한 대체는 적용하지 않음.

→ 5장으로 이동.

(5) 자료 분석 및 통계결과의 품질

#### ■ 관련성

- 조사에 대한 이용자 만족도 측정 방법
  - 별도의 방법을 사용하지 않음.
- 주요이용자 만족도에 대한 이용 가능한 정보의 평가
  - 이용자 만족 정도를 알 수 있는 정보는 주요 이용자 정보와 그 외 이용자 정보는 일부만 있음.
- 조사의 전반적인 관련성 평가
  - 주요 분야는 약간 관련성이 있고, 다른 분야는 거의 관련성이 없음.

■ 정확성

- 추정값의 정도를 측정하기 위해 제시하는 값
  - 어느 것도 제시하고 있지 않음.
- 조사에서 단위무응답의 수준
  - 낮은 수준의 단위 무응답률을 보임.
- 무응답을 세부범주로 분류
  - 무응답을 분류하고 있지 않음.
- 항목무응답의 평가
  - 항목 무응답은 낮은 편임(5%이하).

■ 시의성 및 정시성

- 조사기간과 발간 시점간의 시차
  - 조사기간과 조사결과 발표 시차는 약 20일 임.
- 계획한 발간일자와 실제 발간일자 간의 시차
  - 거의 시차가 없음.
- 조사대상기간과 잠정 또는 최종 결과의 최초발간일 간의 시차
  - 조사대상기간과 잠정 또는 최종 결과의 최초발간과의 시차는 정해진 시간 정도 임.

주요발간물에 대해 정해진 발간시기 준수여부

— 발간물의 80%이상을 준수함.

시의성을 개선하기 위한 변화의 필요성

— 외부협력기관(지자체공무원)이 관련된 변화가 필요함.

### ■ 비교가능성

통계의 중요한 개념과 조사설계에 연계하여 지역에 대한 비교가능성 평가

— 연관된 통계 결과와의 비교 가능성은 약간의 차이가 있음.

조사대상기간과 관련하여 시계열적인 비교가능성 평가

— 시계열적인 비교가능성은 거의 제한이 없음.

### ■ 일관성

월별 또는 연간단위로 동시에 다른 주기로 조사하는지 여부

— 동시에 다른 주기로 조사하고 있지 않음.

→ 6번으로 이동.

### ■ 자료분석

이용자 맞춤분석 제공시 대응능력

— 자료 분석에 대한 이용자의 요구를 만족시킬 의무는 없음.

사용하고 있는 추가분석

— 지수산출과 계절 또는 추세 분석과 추정임.

작년한해 동안 자료 분석을 위해 새롭게 도입했거나 기존의 방법을 개편한 적이 있는지 여부

— 개편한 적이 없음.

### ■ 보안관리

이용자에 대한 자료접근을 위해 보유하고 있는 자료의 형태

— 다량의 통계표를 보유함.

#### (6) 문서화 및 자료의 공표

##### ■ 메타자료의 문서화

행정등록자료를 사용할 경우 관련된 적절한 문서 보유여부

— 보유하고 있음.

통계산출과정에서 문서화하는 내용

— 변수의 형태로서는 다른 변수들의 계산으로 구한 변수, 외부자료로부터 가져온 변수, 변수의 통계적 특성(즉, 범주적, 분류적, 계량적), 측정단위(의미를 가진 경우), , 외부자료에서 가져올 경우 그에 대한 서술, 파생 변수의 경우 계산방법 서술임.

— 분류 서술로서는 표준 또는 일반 분류가 사용될 경우 참조하는 분류, 특정한 분류가 적용된 경우 분류의 출처, 참조된 분류의 일부항목을 처리하는데 특정 주의사항을 설명하는 정보를 포함함.

— 개념과 통계적 단위의 정의에는 정의된 개념의 제목, 정의의 출처, 정의와 그 개념을 측정하는 변수간의 연계사항을 포함함.

— 통계산출과정에서 사용된 조사방법 및 기법을 포함함.

통계작성 절차에 필요한 행정 및 처리절차에 대한 체계적인 문서의 보유여부

— 통계작성절차를 지원하는데 필요한 행정과정 및 처리절차, 예를 들면 예산의 확보와 효율적 관리, 조사표 인쇄, 직원채용, 지방행정기관 방문, 언론매체접촉 등에 관한 모든 사항을 체계적으로 문서화하고 있음.

자료 이용시 예상되는 제한점에 대해 체계적인 문서의 보유여부

— 자료 이용시 예상되는 한계에 대해 체계적으로 문서화 하고 있음.

발간물에 대한 외부이용들의 만족도에 관한 정보

— 발간물에 대한 외부 이용자들의 만족도 정보는 보유하고 있지 않음.

내부직원들의 만족도에 대한 정보

— 내부 직원들의 만족도 정보를 보유하고 있지 않음.

- 보다 나은 발간물을 위해 필요한 항목
  - 자료처리작업 및 직원배치, 사용소프트웨어

■ 자료제공 전략

- 생산된 자료에 대한 정보의 검색을 위해 이용자 측면에서 목록시스템과 정보서비스의 제공여부
  - 이용자에게 편리하도록 기관의 발간목록 시스템과 가장 최근에 생산되고 저장된 자료에 대한 정보를 검색할 수 있도록 인터넷 홈페이지에서 정보서비스가 되고 있음.
- 통계작성기관에서 이용자에게 적합한 형식으로 자료를 이용할 수 있도록 하는 자료제공시스템 보유여부
  - 기관수준의 자료제공 시스템은 이용자에게 적합한 형식으로 액세스하게 하고 있지 않음.
- 통계표와 그래프 작성을 위해 사용하고 있는 기준의 여부
  - 테이블과 도표 작성에 사용된 표준화된 기준이 있음.
- 개편이나 잠정치 또는 조기 추계치 등을 설명하고 있는지 여부와 제한사항이나 잠정자료에 대한 기대수준 등의 정보가 제공되고 있는지 여부
  - 개편이나 잠정 또는 조기추계치 등을 분명히 표시하여 설명하는지와 제한사항이나 잠정치에 대한 기대수준 등에 대한 어떠한 정보도 제공되지 않음.

■ 자료관리

- 자료의 관리와 보관을 위한 체계적인 방법의 보유여부
  - 자료를 관리하고 축적하기위해 자료를 DB화 하여 관리하고 있음.
- 자료 보관시 예상치 못한 상황에 대비하여 자료를 보호하는 방법
  - 자료 보관시 예상치 못한 상황에 대비하여 백업 DB를 관리하고 있음.

(7) 개선 사이클

■ 적용가능성/유연성

- 다른 정보를 수집하기 위해 사용하고 있는 통계패키지 와 절차, 새로운 수요와 정보요구에 대응하는 시간의 적절성에 대한 평가
  - 만족스럽지 못하고 자료입력시스템 적절성을 개선할 필요함.
- 패널조사와 같이 특수한 조사를 유연하게 실시하기 위한 자료수집기반이 있는지 여부
  - 영구적인 자발적 조사를 신축성 있게 실시하기 위한 자료수집 틀은 없음.
- 변화하는 수요에 신속히 대응하기 위한 부담과 비용의 정도
  - 변화하는 수요에 신속히 대응하기 위한 부담은 받아들일 수 있으나, 비용은 거의 줄일 수 없음.

■ 관련분야 전문성

- 조사담당직원의 지식과 이해력 보유 정도
  - 주제 분야별로 원천자료, 개념 및 분류, 현행 및 향후 이슈, 조사관련 필요한 운용방법 등을 잘 이해하고 충분한 지식을 가진 전문적인 직원이 있음.
- 직원들의 교육과정에 대한 평가
  - 현재 직원의 훈련정도는 충분하다고 생각하지만 개선이 필요함.
- 주요 담당자들의 변동에 업무공백 없이 조사를 담당할 만한 전문성을 유지하고 있는지 여부
  - 주요 담당자들의 변동에 대해 전문성이 부족함.
- 다른 기관이나 전문가 협회와의 교류정도
  - 만족스럽지 못하며, 다른 기관이나 전문가 협회와의 교류가 개선되어야 함.

■ 품질관리

- 조사에 대한 정기적인 점검여부
  - 시군구 담당자에 대한 정기적임 점검이 이루어지고 있음.
- 효율성, 시의성 또는 품질확보를 위한 이용하거나 사용하고 있는 최신 IT 기술의

보유여부

- 입력 절차에 대해 IT 기술을 사용하고 있음.
- 기존의 모든 원천자료를 광범위하게 검토할 필요가 있는지 여부
  - 조사의 효율성을 가능한 제고시키기 위하여 기존의 모든 원천자료를 광범위하게 검토할 필요는 없음.
- 통계생산과정의 효율성을 전문가에게 검토를 받고 있는지 여부
  - 독립적인 전문가에 의해 검토 받고 있지 않음.
- 예산이 20% 많거나 적을 때 조치방법
  - 현재 공장등록통계의 예산은 발간물 제작비용만 포함되어 있기 때문에 예산의 축소는 불가능하며, 예산의 확대를 통해 입력 및 분석 시스템을 개선할 수 있음.

3) 개선아이디어 기입란

1장 : 조사수행을 위한 의사결정

주요 이용자들에 대한 정보 중 DM 발송을 위한 일부 정보만 관리되고 있으며, 이용자 의견을 수렴할 수 있는 체계를 마련할 필요가 있음. 이를 위해 자료의 이용실태를 파악하여 적절한 이용자관리를 위해 미충족 이용자 요구에 대한 반영정도를 높일 필요가 있음.

2장 : 조사 설계

공장등록 현황 통계는 가공통계로서 별도의 조사 설계는 이루어지고 있지 않고, 산업단지에 입주한 공장의 등록현황을 파악하여 통계를 산출하고 있으며, 통계청의 광공업통계와 비교가능한 통계임. 단, 일부 지표는 계산방식 및 정의의 차이로 약간의 차이가 발생하고 있음.

3장: 자료수집

FEMIS시스템에 등록된 자료를 자동적으로 수집, 가공하여 통계를 산출하고 있음. 자료입력과정에서 입력오류의 문제가 발생할 가능성이 있으므로 이에 대한 철저한 관리가 요구됨.

#### □ 4장 : 자료입력 및 자료처리

수작업에 의한 자료입력방식을 채택하고 있으며, 에디팅 절차 또한 수작업으로 수행되고 있음. 즉 오류가 발생한 자료에 대해서는 입력 자료에 대해 유·무선전화를 통해 오류를 확인하고 보완하는 절차를 밟고 있음. 이와 같은 입력오류를 방지하기 위해 오류별로 통계를 산출하여 특정오류를 감소시킬 수 있는 방안을 모색할 필요가 있음.

#### □ 5장 : 자료 분석 및 통계결과의 품질

관련성은 이용자 정보가 일부 있음. 정확성지표산출을 위한 어떠한 지수도 제시하고 있지 않으며, 향후 오류발생통계정도는 산출할 필요가 있음. 시의성 및 정시성은 자료수집과 처리 및 분석 과정이 매우 빠르게 진행되며, 약 20일 정도의 소요시간을 고려할 때 적절하다고 판단됨. 비교가능성 부분에서는 다른 통계와 비교 가능하지만, 통계청의 광공업통계와 보다 비교 가능하도록 보고서에 비교내용을 수록할 필요가 있음. 일관성 측면에서는 동시에 다른 주기로 조사하고 있지 않음으로 측정이 매우 제한적임. 자료의 보안관리측면에서 개인의 비밀이 노출되지 않도록 철저한 관리가 요구됨.

#### □ 6장 : 문서화 및 자료의 공표

보고서 상에 품질관련 지표를 함께 수록하도록 해야함. 통계산출과정을 문서화 하여 제공할 필요가 있음.

#### □ 7장 : 개선 사이클

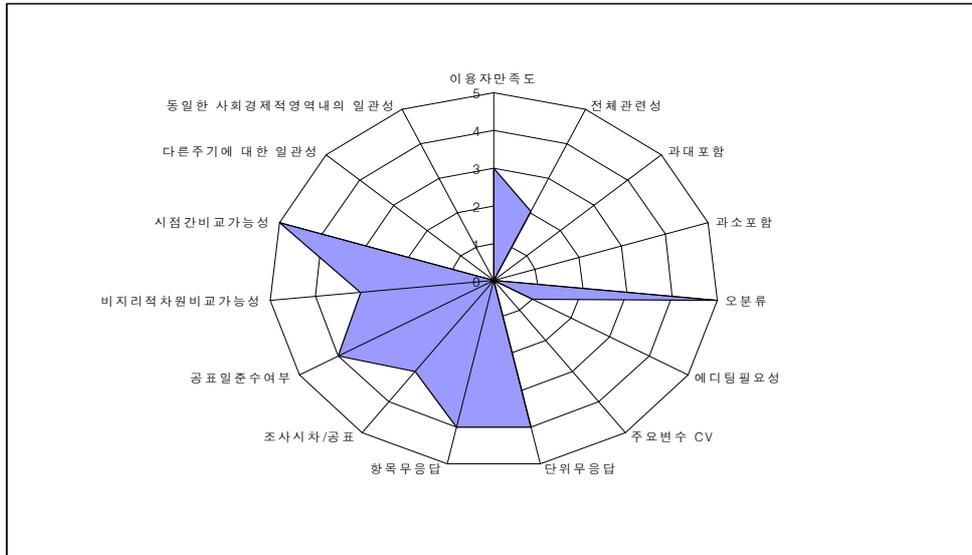
FEMIS 시스템에 자료를 입력하여 이를 가공하여 통계를 산출하고 있으므로 전국적인 네트워크로 관리가 가능하며, 향후 개선 및 보안사항이 발생할 경우 보다 적극적인 시스템관리가 요구됨.

#### 4) 품질 차원별 점수 및 진단 다이어그램

〈표 5-2-4〉 공장등록 현황 통계의 품질차원별 점수

| 품질차원    | 내 용               | 항 목  | 점 수 |
|---------|-------------------|------|-----|
| 관련성     | 이용자만족도            | 5-2  | 3   |
|         | 전체관련성             | 5-3  | 2   |
| 정확성     | 과대포함              | 2-6  | -   |
|         | 과소포함              | 2-7  | -   |
|         | 오분류               | 2-8  | 5   |
|         | 에디팅필요성            | 4-4  | 1   |
|         | 주요변수 CV           | 5-6  | -   |
|         | 단위무응답             | 5-15 | 4   |
|         | 항목무응답             | 5-18 | 4   |
| 시의성/정시성 | 조사시차/공표           | 5-21 | 3   |
|         | 공표일준수여부           | 5-22 | 4   |
| 비교가능성   | 비지리적차원비교가능성       | 5-24 | 3   |
|         | 시점간비교가능성          | 5-25 | 5   |
| 일관성     | 다른주기에 대한 일관성      | 5-27 | -   |
|         | 동일한 사회경제적영역내의 일관성 | 5-28 | -   |

[그림 5-2-6] 공장등록통계의 품질진단 다이어그램

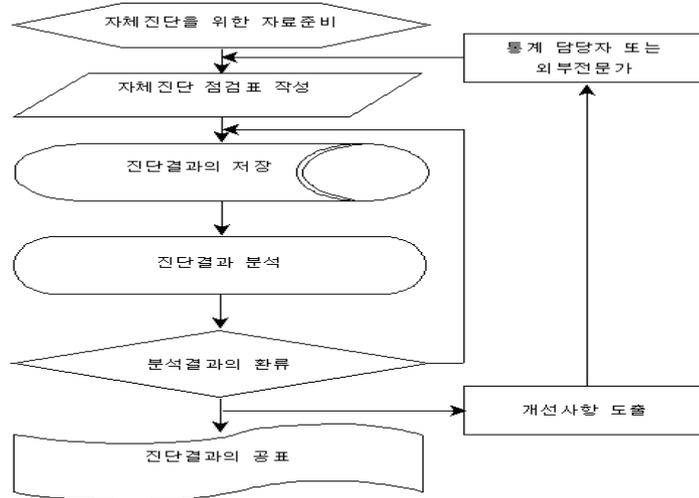


### 제3절 자체통계 품질진단 결과의 활용

정기품질진단과는 다르게 통계작성기관에서 실시하는 자체품질진단의 수행절차는 먼저 수립된 자체품질진단서를 통계청의 품질관리 사이트에서 통계작성기관의 조사 통계담당자가 개별항목을 작성하도록 하였다. 따라서 본 보고서에서는 수기작성의 예를 들어 자체품질진단의 사례를 기술하였지만, 실제로 자체품질진단을 수행할 때에는 인터넷상으로 이루어지게 된다. 그러므로 통계담당자는 자체품질진단에 필요한 각종 문서나 정보를 취합하여 관리할 필요가 있으며, 정기품질진단 평가를 받은 기관의 경우에는 관련 보고서 및 개선사항에 대한 문서를 보관함으로써 자체품질진단이 순조롭게 수행될 수 있다.

자체통계 품질진단서작성과 결과의 활용을 위한 절차를 흐름도를 통해 살펴보면 다음과 같다.

[그림 5-3-1] 진단수행과 결과 도출 및 환류에 대한 흐름도



먼저 통계작성기관의 통계담당자나 또는 조사통계전문가에 의해 통계청의 품질관리사이트(<http://www.codi.go.kr>)에서 자체품질단서를 항목별로 완료하여 저장하게 되면, 점검표 자료는 통계청의 DB에 저장되며, 통계청에서는 저장된 결과를 체계적으로 분석하여 소관통계의 품질점검결과에 대한 각종 통계표와 그래프를 출력할 수 있도록 지원하게 되며, 결과를 종합하여 소관통계의 개선사항을 자체진단 점검표 순서에 따라 작성하여 이를 통계작성기관에게 전달하고, 이를 웹사이트에 공시하도록 한다. 또한 지적된 개선사항에 대해 통계작성기관별로 별도의 의견을 청취할 수 있도록 함으로서 통계청과 통계작성기관 간의 소관통계의 개선을 위한 의사소통이 이루어지게 한다.

한편 통계이용자들은 해당 통계의 자체진단 결과를 참고하여 이용하고 하는 통계의 품질정보를 쉽게 검색할 수 있게 함으로서 이용자들이 해당통계에 대한 이해를 돕도록 할 것이다. 또한 통계이용자들의 품질진단에 관한 의견을 제시할 수 있도록 별도의 이용자 코너를 설치하여 통계작성기관과 통계청, 그리고 통계이용자간의 의견교환이 이루어질 수 있도록 할 것이다.

이와 같은 절차에 의해 수행된 통계작성기관의 자체품질진단 결과를 통계청으로부터 환류 받아 통계작성기관에서는 소관통계의 품질진단 결과 개선사항에 대해서는 단기 개선과제와 중장기 개선과제로 분류하여 해당 과제의 이행 계획을 수립하여 통계청에 보고하고, 통계청에서는 지속적으로 관련 통계의 개선이행 사항을 점검토록 함으로서 자체품질진단의 결과가 지속적으로 환류 될 수 있도록 해야 한다. 통계작성 관련 예산의 확충과 인력의 수급 계획 등 물리적인 부분과 통계작성 과정에서 나타나는 품질차원별 개선 사항 등의 논리적인 부분 등으로 구분하여 개선계획을 수립하는 것이 바람직하며, 향후 개선과제의 수행에 필요한 로드맵을 구체적으로 작성하여 제출토록 해야 한다.

## 제6장. 정책적 제언

### 1. 통계 작성기관

소관통계에 대한 자체품질 진단을 수행한 통계작성 기관은 진단 결과를 바탕으로 소관통계에 대한 개선점을 취합하여 향후 통계작성 시에 이를 적극적으로 반영할 수 있도록 해야 한다. 이러한 개선사항의 반영은 신설된 통계법에도 명시된 만큼 선택사항이 아닌 필수사항임을 인식해야 한다. 특히 통계작성 기관의 장은 통계품질의 개선이 예산과 인력의 효율적 활용이며, 결코 예산의 낭비가 아님을 제고할 필요가 있다.

먼저 통계작성기관의 장은 소관통계의 품질개선을 위한 적극적인 품질관리 활동이 이루어 질 수 있도록 예산과 조직을 운영할 필요가 있다. 소극적인 품질관리는 결과적으로 소관통계의 신뢰성을 떨어뜨리는 결과를 낳을 수 있으며, 소관 통계의 지속여부까지 위협하게 된다. 한편 통계청 등 외부 유관기관과 전문가들 간의 긴밀한 협의체를 구성할 필요가 있다. 조사통계의 경우 소관통계의 조사 설계, 자료 분석, 결과의 공표 등 모든 절차에 대해 일관적으로 자문기구 등에 의한 점검을 받는 것도 고려해 볼 수 있다.

다음으로 통계작성 기관의 소관 통계 책임자는 작성되는 통계의 중요성을 깊이 인식하고, 진단 결과 나타난 개선점 특히 전문성이 요구되는 부분에 대해서는 반드시 자체검토와 함께 외부 전문가들에 의한 점검이 이루어지도록 해야 한다. 내부의 문제는 내부적으로 해결할 수 없는 경우가 많기 때문이다. 한편 조사를 직접 수행하지 않고 외부 조사기관에 의뢰하여 조사가 수행되는 경우 통계 담당자는 반드시 조사기관에서 수행되고 있는 모든 조사업무를 총괄적으로 관리해야 한다. 통계결과의 책임은 조사기관이 아니라 통계작성기관에 있기 때문이다. 예산의 편성, 조사원의

조사수행 절차, 표본대상가구나 사업체에 대한 응답성향 파악과 더불어 통계담당자들이 수시로 조사결과를 보고받도록 함으로서 조사의 품질을 향상시킬 수 있다. 조사과정에 대한 소극적인 관리는 결과적으로 잘못된 조사결과를 얻기 쉽기 때문이다.

결과적으로 통계작성 절차의 모든 과정에서 발생한 문제점 또는 개선점, 우수사례 등을 단계별로 점검하여 소관통계의 품질 향상을 위해 최선을 다해야 할 것이다.

## 2. 통계 이용자

통계이용자는 통계작성 기관에서 자체품질진단 결과를 참고하여 이용자 관점에서 보다 개선된 통계를 작성할 수 있다. 이용하고자 하는 통계의 품질 평가결과를 토대로 해당 통계의 이용실태를 파악 할 수 있고, 이용자 서비스가 얼마나 잘 되고 있는지를 평가할 수 있다. 또한 품질진단 차원별 분석을 통해 해당 통계의 품질이 각 차원별로 어떠한 상태인지를 알아볼 수 있게 된다. 세부적으로 이용자 관점에서 살펴보면, 조사통계의 경우 조사 기획단계, 조사목적, 표본설계, 자료처리, 추정과정, 오차의 발생원인 등을 파악하여 해당통계의 이용의 제한점 또는 효율성 등을 평가할 수 있을 것이다. 또한 주기적으로 생산되는 통계의 경우 이전 통계의 품질진단 결과로부터 향후 개선점을 통계작성기관에 제공함으로써 보다 향상된 품질의 통계를 생산하는데 조언을 줄 수 있을 것이다. 통계이용자의 통계에 대한 관심도에 따라 해당 통계의 품질은 더욱 향상될 수 있거나, 저하될 수 있기 때문이다.

또한 자체진단 결과 품질정도가 떨어지는 통계의 경우 이용자들의 조언에 따라 개선점을 반영하여 추후 통계작성 과정에 반영할 수 있도록 하는 경로를 개발하는 것이 필요하다. 특별히 조사항목의 개선이나 표본추출방법의 개선, 추정과정의 개선 등은 조사통계에서 반드시 고려할 수 있는 사항이다.

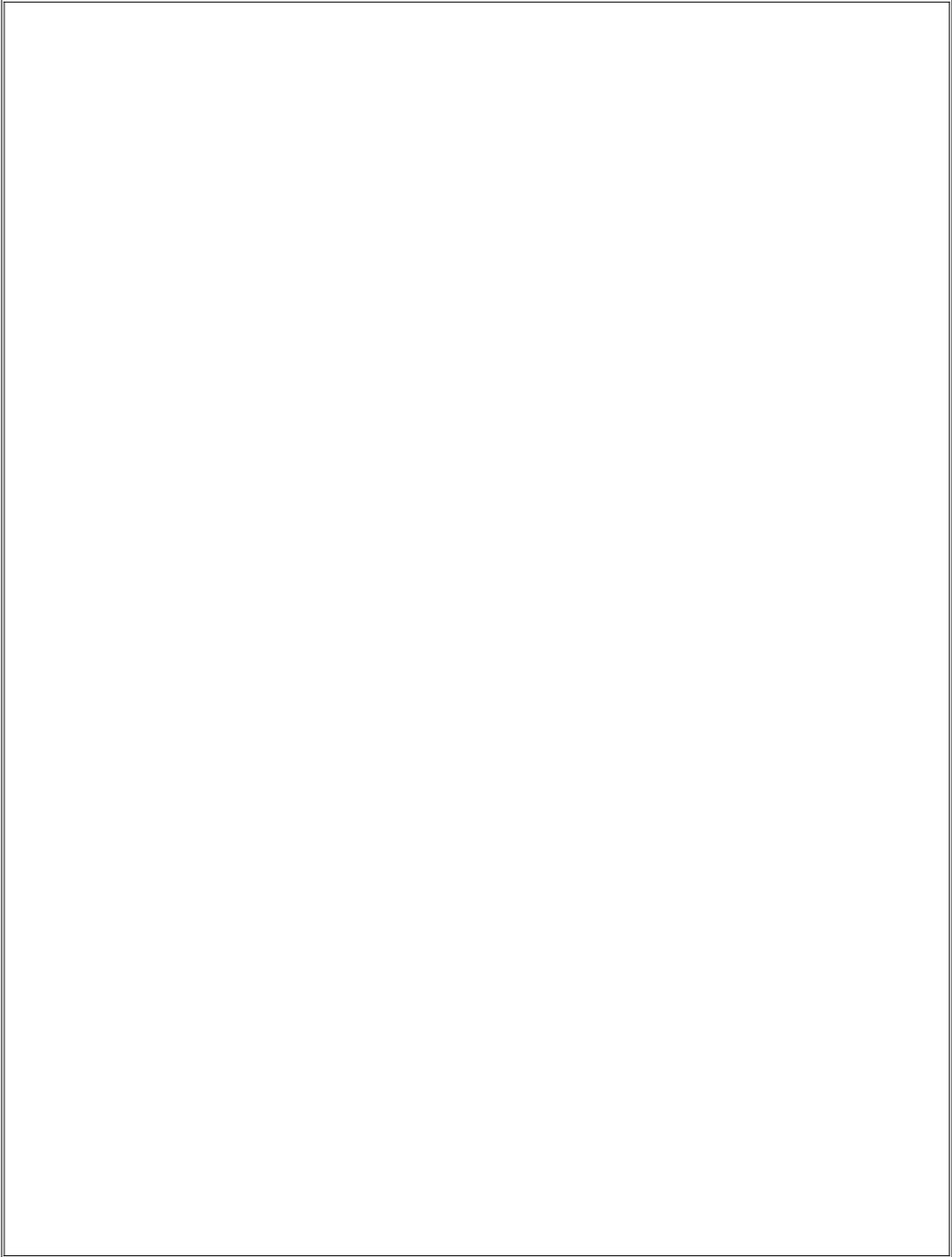
전반적으로 통계이용자들의 요구사항을 적극적으로 반영하여 통계작성기관의 소관통계의 품질이 개선되도록 해야 하며, 이러한 절차는 국가의 정책 수립에 기여하는 바가 크기 때문에 매우 신중하게 검토하여야 한다.

### 3. 통계청

통계작성 기관이면서 국가 통계의 전반적인 품질을 관리하고 있는 통계청의 역할은 매우 중요하다고 할 수 있다. 최근 통계법이 개정되어 국가 통계작성기관에서 생산되는 각종 통계의 품질을 관리 감독하게 됨에 따라 우선 통계작성기관에서 생산되는 소관통계에 대한 자체품질진단이 적절히 수행될 수 있도록 협력할 필요가 있다. 각각의 통계작성기관의 통계적 기반이나 인적, 물적 상황이 모두 다르기 때문에 자체진단의 수행능력이 상대적으로 부족한 기관에 대해서는 적절한 훈련과 지원이 되도록 해야 하며, 상대적으로 자체품질진단 결과가 우수한 기관에 대해서는 해당 통계의 우수사례를 취합하여 이를 전파하고, 또한 그에 따른 각종 인센티브를 부여하는 방안을 모색할 필요가 있다. 자체품질진단은 궁극적으로 국가통계의 품질 개선을 위한 정책임을 홍보하고, 각 기관의 협조를 이끌어 낼 수 있도록 해야 할 것이다.

한편 자체품질진단 결과를 이용하여 통계작성기관에게 각종 지원 프로그램을 개발하여 운영함으로써 해당기관의 통계품질을 향상시킬 수 있도록 해야 한다. 즉, 자체진단결과에서 나타난 개선점을 세밀히 파악하여 부족한 사항에 대해서는 개선 아이디어를 제공할 수도 있고, 우수한 항목에 대해서는 자료를 취합하여 타 기관의 진단에 벤치마킹 할 수 있도록 정책적으로 지원할 수도 있다. 이와는 별도로 각 통계작성 기관에 대한 자차진단 프로그램의 수행에 대한 의견을 수렴하기 위해 정기적으로 학회 또는 협회와 공동으로 세미나 또는 워크숍을 개최함으로써 외부전문가와 통계작성 기관 그리고 통계청간의 협력관계를 구축할 필요가 있다. 이러한 전략은 결과적으로 통계이용자를 확보하고, 통계를 홍보하는 좋은 기회가 될 수 있기 때문이다.

마지막으로 통계청에서는 통계작성기관이 용이하게 자체품질진단을 수행할 수 있는 표준화된 시스템을 제공하고, 진단 결과를 다양한 형태로 가공하여 일반 이용자와 전문가들에게 제공함으로써 통계의 품질향상에 도움이 되도록 해야 할 것이다.



## 참고문헌

- 김설희, 박은영 (2006). 국가통계 자체진단 프로그램개발, 2006 한국조사연구학회 춘계학술대회 발표논문집.
- 김수택, 정기호, 김설희 (2006). 가공통계품질평가지표개발, 2006 한국조사연구학회 추계학술발표대회 발표논문집.
- 손창균, 홍기학, 이기성 (2006). 표본추출 및 관리매뉴얼, 한국보건사회연구원.
- 장영식, 고경환 도세록, 이수현, 임재연 (2006). 정책통계품질향상을 위한 진단기준 및 품질관리 매뉴얼 마련, 한국보건사회연구원.
- 조사통계연구회, 무응답 오차(2000). 자유아카데미.
- 통계청, 통계에도 품질관리가 필요합니다., 2004.4.
- 통계청, 통계품질관리 이렇게 합니다., 2007.4.
- 통계청, 복지분야 국가통계품질진단 최종결과보고서, 2006.9.
- 한국통계학회, 통계품질관리 표준매뉴얼 작성 학술연구-국가통계작성기관의 자체적인 품질관리를 위한 지침서, 2003, 11.
- Aitken, A., Horngren, J, Jones, N., Lewis, D., and Zilhao, M. J. (2003). Handbook on improving quality by analysis of process variables, Final Report, Eurostat.
- Biemer, P. and Lyberge, L. (2003). Introduction to Survey Quality, Hoboken, N.J.: Wiley.
- Bureau of Census, U.S.(1995), CONCOR user's guide.
- Eurostat (2002). Quality in European Statistical System - The way Forward, Office for Official Publications of the European Communities, Luxembourg.
- Eurostat (2003). Assessment of Quality in Statistics : Definitions of quality in statistics, WG, Luxembourg, October, 2003.

- Eurostat (2003). Assessment of Quality in Statistics : Methodological Documents Standard Report, WG, Luxembourg, October, 2003.
- Eurostat (2005). Quality in Statistics: Standard quality Indicators, WG, Luxembourg, May, 2005.
- Eurostat (2005). European Statistics Code of Practice- for the National and Community Statistical Authorities, <http://europa.eu.int/comm/eurostat/quality>
- Hustoft, A. G., Linnerud, J, and Saebo, H.,V. (2004). Quality and metadata in Statistics Norway, Proceedings of Q2004, European Conference on Quality in Survey Statistics, Mainz.
- Korner, T. (2003). Development of a Self Assessment Programme, Final Report, Eurostat.
- Korner, T.(2006). Data Quality Management in the European Statistical System, Proceedings of Q2006, European Conference on Quality in Survey Statistics, London.
- OECD, Quality framework and Guidelines for OECD Statistical Activities, 2003
- Öberg, S.(2006). Quality Issues in the European Statistical System, Proceedings of Q2006, European Conference on Quality in Survey Statistics, London.
- Saebo, H., V.(2006). Systematic Quality Work in Official Statistics-Theory and Practice, Proceedings of Q2006, European Conference on Quality in Survey Statistics, London.
- Statistics Canada, 1997, 1996 Canadian Census Demographic Variables Imputation.
- Statistics Canada, Managing the Quality of Statistics, 1999 Report of the Auditor General, April, <http://www.oag-bvg.gc.ca/domino/reports.nsf/html/9903ce.html>
- Statistics Canada, Statistics Canada's Quality Assurance Framework, 2002.
- Zilhao, M. J., Santos, A. C., Davies, J., Howell, S., and Booleman, M. (2003) State of the art concerning the auditing activity in NSI's, Final Report, Eurostat.

## 자체통계 품질진단서

|       |  |             |               |  |
|-------|--|-------------|---------------|--|
| 통 계 명 |  | 작<br>성<br>자 | 소 속 부 서       |  |
| 승인번호  |  |             | 성 명           |  |
| 작성기관  |  |             | 전 화 번 호 ( ) - |  |
| 작성일자  |  |             | E-mail        |  |

## 진단서 사용을 위한 가이드라인

### ■ DESAP이란?

DESAP은 유럽통계시스템(European Statistical System:ESS)하의 통계조사의 체계적인 품질평가를 위한 포괄적인 체크리스트이다. DESAP은 조사관리자들 자신이 작성하는 통계의 품질을 진단하고, 품질개선 수단을 강구하는데 도움을 주기 위한 도구로 고안되었다. DESAP의 개발과정에서 다양한 분야의 시험 조사를 통한 테스트를 거친 바 있다. DESAP은 ESS 품질기준을 충분히 충족하고 있으며 통계자료의 품질과 관련된 주요한 부분을 포괄하고 있다. DESAP은 주제 분야와 특정한 조사방법과 무관하게 마이크로 데이터로 수집된 개별적인 통계에 적용시킬 수 있다는 관점에서 일반적인 도구이다.

### ■ 점검표의 작성 주체

점검서는 조사관리자 즉, 해당 조사(혹은 조사의 주요 부분)에 책임이 있는 사람이 별도의 준비 없이 작성할 수 있도록 고안되었다. 그러나 경우에 따라서는 평가과정에 다른 전문가(즉 조사방법론 전문가)를 포함하는 것이 유용할 수도 있다. 자체진단을 수행하는 또 다른 대안으로는 조사부문별로 책임 있는 팀과 함께 작성하는 방법도 있다.

### ■ DESAP의 목적

☞ DESAP은 다양한 기능을 수행하도록 고안되었다.

무엇보다도 우선 DESAP은 진단도구로서, 전체적인 조사의 품질을 평가하고 ESS 품질평가요소를 포괄하는 품질평가 프로파일을 수립하기 위하여 사용되어야 한다. 다음으로 체크리스트는 품질개선 수단의 관점에서 지침을 제공하고 품질에 문제가 있는 통계의 잠재적 위험성의 평가가 이루어질 수 있도록 한다. 또한 시간과 분야별(통계기관에서 집중적

으로 자료를 축적할 경우)로 품질수준의 단순 비교를 위한 수단을 제공해 준다. 통계를 담당하는 부서에 대한 자원의 배분의 근거를 제공할 수 있으며, 해당업무에 대한 교육에 도움을 줄 수도 있다. 점검서를 작성하기에 앞서 해당 조사에 대한 DESAP의 모든 잠재적인 활용방안을 검토해야한다.

#### ■ DESAP의 구조

DESAP은 조사의 초기 단계부터 최종 단계까지 모든 과정에 따라 점검되도록 하는 과정 중심(process-oriented)방식으로 구성되어 있다. DESAP은 관리 가능한 구조를 제공하기 위해 7개 주요 조사과정과 24개 하부 요소를 정의하고 있다. 이들 절차와 하부요소는 다음 페이지에 제시한다.

## ■ DESAP의 과정 및 과정별 하부요소

### 1장. 조사 수행을 위한 의사결정

이용자 요구 분석  
조사목적

### 2장. 조사 설계

조사개념  
표본설계(표본조사만 해당)  
측정도구(조사표 또는 양식 등) 개발 및 테스트

### 3장. 자료 수집

자료소스  
조사원 채용 및 교육  
무응답 줄이기  
현장실사

### 4장. 자료입력 및 자료처리

자료입력  
에디팅 절차  
대체(imputation)절차

### 5장. 자료분석 및 통계결과의 품질

관련성(Relevance)  
정확성(Accuracy)  
시의성 및 정시성(Timeliness and Punctuality)  
비교가능성(Comparability)  
일관성(Coherence)  
자료분석(Data analysis)  
보안관리(Disclosure control)

### 6장. 문서화 및 자료의 공표

메타자료 문서화  
자료제공전략  
자료관리

### 7장. 개선 사이클

적용가능성/유연성  
관련분야의 전문성  
품질관리

## ■ 점검서 사용방법

처음 1장부터 완성해나가되, 한 단계씩 차근차근 작성하도록 한다. 조사관리자는 질문에 직접 응답하거나, 각 조사를 책임지고 있는 팀과 함께 작성할 수도 있다. 다른 방법으로 팀원들에게 점검서를 복사하여 나누어 주고, 모두에게 작성하게 한 후 결과에 대한 토론도 가능하다. 또한 조사 기관의 형태에 따라 특정한 질문에 대해서는 그 기관 내 다른 부서의 전문가에게 문의할 수도 있다.(예를 들면 표본설계나 정확성 척도에 관한 질문과 관련하여 통계전문가에게 문의가능 함)

점검서는 앞에서 언급한 바대로 목적에 따라 서로 다른 형태의 질문으로 표현하고 있다. 첫째 품질진단을 위한 질문은 남색배경과 옆에 감청색 막대모양으로 표시하였다. 이러한 질문들은 다이어그램으로 피드백 하여 요약된 형태로 제공한다.

주기적으로 점검서를 작성할 경우 즉, 시점에 따라 품질수준을 비교해 보고자 할 경우에는 진단용 질문만 별도로 모아놓은 집중 형식을 사용할 수 있다. 하지만 처음으로 자체진단을 수행하는 경우에는 전체 점검서를 작성해야 한다.

다음으로 여러 가지 문항으로 구성된 품질개선과 관련된 질문이 있으며, 각 통계작성기관들이 자료 품질진단 척도를 종합한 것이다.

일반적으로 이러한 응답문항들은 해당 조사에 대해 점검표에 표시할 있기 때문에 시간을 많이 소요하지 않고 작성이 가능하다. 그러나 다른 관점에서 보면 조사과정에 대한 개선 가능한 아이디어를 얻을 수도 있다.

작성과정에서 수집된 진단관련 아이디어를 적절히 문서화 하기위해, 모든 아이디어, 즉 점검서를 작성하는 동안 가능한 방법론적인 변화를 기록하는 것이 바람직하다. 이러한 아이디어를 각 장별로 기록하여 모아둘 수 있도록 기록 란을 제시하고 있다. 이 페이지는 점검서를 작성할 동안 참조하기 위하여 이 페이지를 뜯어낼 수 있도록 하였다.

마지막으로 주관식 질문을 제시하였고, 이들 중 일부 응답문항을 제시하지 않았지만, 일부는 품질개선 아이디어 개발하는데 유용한 질문을 하기위한 목적이 있다. (예를 들어, 우선적으로 이용자 수요에 주요하게 부적합한 사항은 무엇입니까?)

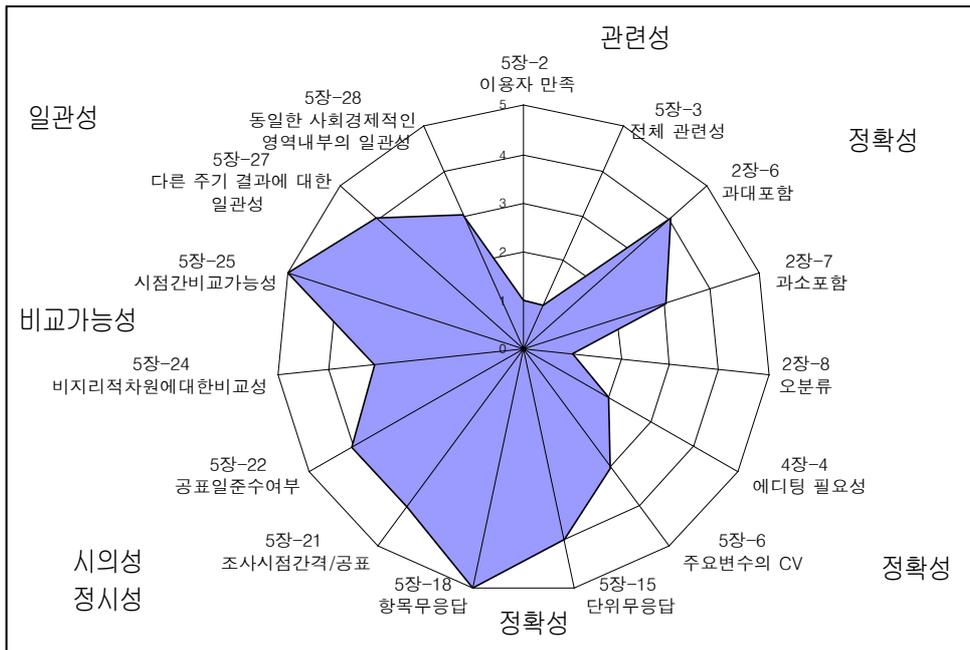
■ 주요 용어

점검서를 작성하는 과정에서 친숙하지 않은 전문적인 용어가 있을 수 있다. 이 경우 점검서의 맨 뒷부분에 용어설명을 제시하고 있다.

■ 진단 다이어그램을 작성하는 방법

1. 체크리스트를 작성할 때 다이어그램이 있는 페이지를 오려둔다.
2. 한쪽이 진한 남색 막대가 있는 진단 질문에 응답할 때 응답한 값을 다이어그램에 그대로 기입한다.
3. 진단질문 문항의 응답 값은 수치이므로, 그 값과 같은 값을 다이어그램에서 찾는다.
4. 다이어그램에서 각 질문에 응답한 값들을 서로 연결한다. 만일 어떤 질문이 해당조사와 연관되지 않았으면, 다이어그램에서 해당문항에 체크하지 않는다.
5. 모든 진단 질문과 다이어그램상의 연결선을 완성하면 아래와 같은 그림이 완성된다.
6. 다이어그램에 나타난 영역이 해당 통계의 품질 프로파일을 나타낸다.

〈다이어그램 1〉 진단 다이어그램



**■ 개선 아이디어 기입란:**

이 페이지에는 점검표를 작성하는 동안 담당 통계에 대한 어떤 개선 아이디어를 기입하는데 사용된다. 작성하는 동안 이 페이지를 오려내어 보관하고 있다가 점검표 안에 기입하는 대신 이 페이지에 아이디어들을 모을 수가 있습니다. 예를 들면, 다음에 다시 검토해 보기를 원할 경우 도움이 됩니다.

1장

2장

3장

4장

5장

6장

7장

## 1장. 조사 수행을 위한 의사결정

### ■ 이용자 요구 분석

통계산출물이 이용자 요구사항을 충족시킬 경우 관련성(Relevance)이 있다고 진단한다. 이용자의 요구분석은 조사초기에 수립되어야 한다. 주요이용자들과 관련된 다양한 사안에 대하여 다음 각 항목별로 응답하시오.

#### 1. 조사수행의 법적근거는 무엇입니까?

- 1) 국내법
- 2) 국제법
- 3) 법적근거 없음
- 4) 기타(다음 란에 합법적 근거를 명시하시오) ↴

#### 2. 조사에 대한 주요 이용자를 모두 알고 계십니까?

- 1) 모든 주요 이용자에 대한 정보가 있음
- 2) 주요 이용자의 일부 정보만 알고 있음
- 3) 아니요, 주요 이용자에 대한 정보가 부족함

#### 3. 조사(통계)의 주요 이용자는 누구입니까? - 중요도에 따라 1순위부터 5순위 까지 순위를 부여하시오.

- 1) 국제기구
- 2) 단체 및 협회
- 3) 정부기관
- 4) 언론기관
- 5) 금융기관(중앙은행 포함)
- 6) 기업
- 7) 연구기관 및 대학
- 8) 모름
- 9) 기타(다음 란에 주요 이용자를 기술하시오) ↴

**4. 주요 이용자들의 정보가 어떻게 관리되고 있습니까?**

- 1) 이용자 관련 정보 관리를 위해 이용자DB에 수록됨 (예, 고충처리)
- 2) 이전 조사 자료에 일부 정보가 수록되어 있음
- 3) 부족한 이용자수요를 유사통계에서 관리함
- 4) 어떤 정보도 없음
- 5) 기타(다음 란에 정보를 기술하시오) ↴

**5. 이용자의 자문을 구하기 위하여 어떠한 절차를 거치십니까?**

- 1) 관련된 정부부처를 계획수립 과정에 포함시킴
- 2) 생산자-이용자그룹(예, 통계소위원회)을 통해 주제별 특정사항에 대한 토론절차를 거침
- 3) 실무이사회 또는 위원회
- 4) 다수 이용자 그룹의 수요를 파악하는 고객조사
- 5) 우선적인 주요 이용자와 공식적인 협약체결(서비스수준협약)
- 6) 통계정보 이용과 관계된 마케팅 활동
- 7) 시장조사와 같은 사회-경제학 분야 파트너와 협력
- 8) 관련 이용자와 비공식적인 접촉
- 9) 기타(절차를 기입하시오) ↴

**6. 향후 이용자 수요를 예측하기 위하여 어떠한 절차를 거치십니까?**

- 1) 통계위원회나 소위원회에서 계획 수립
- 2) 정부부처와 양자 간 위원회
- 3) 정부의 정책안건 또는 국회기록물 검색
- 4) 중앙 또는 지방자치단체간의 양자간 위원회
- 5) 이용자 그룹의 학술그룹에 임원으로 참여
- 5) 다른 이용자 그룹들과 양자간 회의나 워크숍
- 6) 미 충족 이용자 요구사항 파악
- 7) 발간물 검토
- 8) 어떤 절차도 거치지 않음
- 9) 모름
- 10) 기타(절차를 기입하시오) ↴

7. 미 충족시킨 이용자 수요를 우선순위로 기입하십시오.

모름

8. 미 충족된 주요 이용자 수요에 적절히 대응하기 위한 계획은 무엇입니까?

■ 조사목적

실행계획수립 단계에 앞서, 조사의 내용, 특성, 기본적인 정의 및 분류 등을 구체화하는 원칙을 설정하는 것이 중요합니다. 이 원칙을 설정함에 있어, 질문에 조사목적을 명확히 나타내는 것이 매우 중요합니다.

9. 조사의 목적이 명문화되어 설정되어 있습니까?

1) 예

2) 아니오

♣ “1장. 조사수행을 위한 의사결정” 절차에 대한 질문 중에서 고려해야 할 사항이 있으면 해당 내용을 기입 하십시오.

## 2장. 조사 설계

조사 설계 단계는 자료수집 및 처리의 모든 과정에 대한 품질을 보장하기 위한 원칙을 제시하는 단계이다.

### ■ 조사 개념

조사의 개념들을 신중히 선택하고 설계하는 것은 비교가능성, 일관성, 시의성 등을 확보하기 위하여 특히 중요한 개념이다. 다음의 질문들은 초기 계획수립단계에서 명확히 하여야 할 주요한 사항들이다.

#### 1. 해당 조사와 비교할 만한 공식/비공식 조사가 있습니까?

- 1) 예  
 2) 아니오 → 3번으로 이동

#### 2. 조사의 품질을 개선하기 위해 활용할 만한 공식/비공식 조사가 있습니까?

- 1) 아니오  
 2) 예 (관련 조사를 기입하십시오) ↴

#### 3. 조사에서 다음 중 적용하고 있는 표준화된 체계, 기준, 개념, 정의 및 분류는 어떠한 기준입니까?

- 1) 국제적인 기준  
 2) 국가적인 기준  
 3) 조사특유의 기준

#### 4. 조사에서 표준화된 분류기준과 맞지 않는 사항이 있습니까?

- 1) 아니오  
 2) 예(구체적으로 설명) ↴

**5. 조사의 추출틀은 비교 가능한 조사에서 사용하고 있는 공통적인 표본추출틀 인가?**

- 1) 예
- 2) 아니오
- 3) 공통 표본추출틀이 있으나 사용안함

**6 사용하고 있는 추출틀은 어느 정도의 과대포함 문제가 있다고 판단하십니까?**

- 1) 대량의 과대포함  
☞ 목표모집단에 포함되지 않은 허용할 수 없는 수준의 다수의 단위들이 추출틀에 포함되어 있거나, 과대포함률을 모름.
- 2) 상당한 과대포함  
☞ 목표모집단에 포함되지 않은 다수의 단위들이 추출틀에 포함됨.
- 3) 약간의 과대포함  
☞ 목표모집단에 포함되지 않은 일부의 단위들이 추출틀에 포함됨.
- 4) 아주 작은 과대포함  
☞ 목표모집단에 포함되지 않은 극소수의 단위들이 추출틀에 포함됨.
- 5) 과대포함이 거의 없음  
☞ 추출틀과 목표모집단이 거의 차이가 없음.

**7 사용하고 있는 추출틀은 어느 정도의 과소포함 문제가 있다고 판단하십니까?**

- 1) 대량의 과소포함  
☞ 목표모집단에 포함된 허용할 수 없는 수준의 다수의 단위들이 추출틀에 포함되지 않거나, 과소포함률을 모름.
- 2) 상당한 과소포함  
☞ 목표모집단에 포함된 다수의 단위들이 추출틀에 포함되지 않음.
- 3) 약간의 과소포함  
☞ 목표모집단에 포함된 일부의 단위들이 추출틀에 포함되지 않음.
- 4) 아주 작은 과소포함  
☞ 목표모집단에 포함된 극소수의 단위들이 추출틀에 포함되지 않음.
- 5) 과소포함이 거의 없음  
☞ 추출틀과 목표모집단이 거의 차이가 없음.

**8 통계 단위에 대한 분류오류가 있습니까?**

- 1) 아주 큰 수의 분류오류
- 2) 상당히 큰 수의 분류오류
- 3) 일부의 분류오류
- 4) 아주 적은 분류오류
- 5) 잘못된 분류가 거의 없음

**9. 조사주기와 추출틀의 최근 갱신 시기간의 시차가 어느 정도입니까?**

- 1) 3개월 이내
- 2) 6개월 이내
- 3) 9개월 이내
- 4) 12개월 이내
- 5) 1년 - 3년
- 6) 3년 이상

**10. 조사주기와 추출틀의 최종 갱신 시기 간의 시차가 짧아지면 과대 또는 과소 포함문제가 감소됩니까?**

- 1) 예
- 2) 아니오

**11. 다음 중 어떠한 작업을 통해 표본추출틀이 목표모집단과 가능한 근접한가를 측정하는가?**

- 1) 목표모집단에 대해 추출틀을 비교 가능한 다른 원 자료간의 결합
- 2) 중복, 사망, 조사범위 외 단위, 특성변화 파악을 위한 조사결과분석
- 3) 포함정도를 파악하기 위하여 조사표 상에 특수 질문 사용
- 4) 지역사무소, 실시현장직원, 조사단위 자체를 점검
- 5) 실시현장에서 추출틀의 일부 자료를 점검(조사범위 외 단위 확인)
- 6) 추출틀을 동일한 지역추출틀의 표본단위들과 비교함
- 7) 특정한 기간동안 발생한 변동을 반영하기 위해 추출틀을 갱신함
- 8) 다른 자료 또는 특별히 설계된 반복조사결과간의 일치성을 점검함
- 9) 동일 표본추출틀을 사용하는 조사에서 산출된 평가정보사용
- 10) 포함범위의 변동을 통제 불가능할 때 행정자료 담당자들을 접촉함
- 11) 표본설계에서 사용된 지역경계가 분명하고 중첩되지 않는지를 확인하기 위하여 지역추출틀에 대한 지도를 점검함
- 12) 전혀 점검하지 않음
- 13) 기타(다음 란에 방법을 기입) ↴

○ 해당 조사가 표본조사인 경우에만 “표본설계” 부분을 응답하시오. 그 외에는 23번 문항으로 가시오.

■ 표본설계(표본조사인 경우에만 응답)

최고수준 또는 요구수준의 정확성은 표본설계의 수준에 좌우됨으로 표본설계는 매우 신중하게 고려되어야 한다.

12. 조사에 보편적으로 통계 작성기관에서 채택하고 있는 표본추출 기법을 사용하고 있습니까?

- 1) 예
- 2) 아니오

13. 정확성과 비용을 감안하여 최적의 표본크기를 결정하였습니까?

- 1) 예
- 2) 아니오

14. 다음 중 조사에 적용한 표본설계방식은 무엇입니까?

- 1) 비확률 추출설계
- 2) 절시법
- 3) 시간대별 추출(Flow-sampling)
- 4) 단순확률추출
- 5) 계통확률추출
- 6) 1단계 집락추출
- 7) 2단계 집락추출
- 8) 층화추출
- 9) 이중(층화)추출
- 10) 확률비례추출복원추출
- 11) 확률비례추출비복원추출
- 12) 기타방법 (자세한 방법을 기입) ↴

15. 비확률추출설계를 사용했다면, 그 이유를 설명하시오

16. 총화추출 설계를 적용한 경우 다음 중 어떠한 방법에 의해 총별로 표본수를 배분하였는가?

- 1) 비례배분
- 2) 최적배분
- 3) 네이만배분
- 4) 기타방법 (자세한 방법을 기입) ↓

17. 응답부담을 최소화하기 위하여 연동표본 구조를 이용합니까?

- 1) 예
- 2) 아니오

18. 표본추출 및 표본크기를 결정하기 위하여 사용하는 소프트웨어는?

- 1) 자체개발 시스템
- 2) 범용화된 표본추출 소프트웨어 또는 표본추출 프로그램 등  
소프트웨어에 대해 설명하십시오. ↓

19. 목표 표본크기와 실현 표본크기 간에 차이가 있습니까?

- 1) 차이가 없음
- 2) 실현 표본크기가 목표 표본크기보다 큼
- 3) 실현 표본크기가 목표 표본크기보다 5% 이하로 작음
- 4) 실현 표본크기가 목표 표본크기보다 5% 이상 작음

20. 결측된 단위에 대해 현장대체를 허용하고 있는가?

- 1) 허용함
- 2) 정해진 한계 내에서만 허용함
- 3) 허용하지 않음

**21. 지금까지 모든 사항을 고려했을 때, 조사를 위한 조사 설계의 효율성 어떻습니까?**

- 1) 효율적임
- 2) 주어진 예산 및 시간 내에서 매우 만족스럽고, 적은 부분 개선될 수 있음
- 3) 주어진 예산 및 시간 내에서 만족스럽지만, 개선이 이루어져야 함
- 4) 부적절함
- 5) 큰 결함이 있음

**22. 순수하게 방법론적인 관점에서 볼 때, 조사 설계를 변경해야 한다고 생각하는가?**

- 1) 시급히 변경해야 함
- 2) 변경해야하지만, 좀 더 테스트 해볼 필요가 있음
- 3) 아니오.

○ **조사표를 사용할 경우에만 다음 부분인 “측정도구(조사표)의 개발 및 검증”에 답하십시오. 아니면 3장으로 이동하십시오.**

**■ 측정도구(조사표)의 개발 및 검증**

적절하게 개발된 측정도구(조사표류)를 사용하면 측정오차를 최소화시킬 수 있으며 응답자와의 관계도 최적화할 수 있다. 다음 질문들은 조사표 또는 면접의 품질을 과 관련한 주요 측면들에 초점을 맞춘 것이다.

**23. 조사표 설계과정에 대해 조사표는 다음 중 어떠한 질문들로 이루어 졌습니까?**

- 1) 사전 검증하지 않은 질문항목들로 구성됨
- 2) 사전 검증하여 균형 잡힌 질문항목들로 구성됨
- 3) 항목배열에 표준화된 설계로 구성됨
- 4) 컴퓨터를 사용한 면접을 위한 전자조사표 설계용 프로그램으로 표준화된 소프트웨어로 구성됨
- 5) 기관 내 전문 조사표 담당자가 점검한 질문항목으로 구성됨
- 6) 외부기관의 전문 조사표 담당자가 점검한 질문항목으로 구성됨
- 7) 국제기구에 의해 제공된 조사표임

**24. 응답하거나 조사표를 작성하는데 소요되는 시간을 고려할 때 조사표분량은 어느 정도입니까?**

- 1) 20분 이내 쉽게 작성할 수 있는 분량임
- 2) 45분 이내 작성할 수 있는 분량임
- 3) 45분 이상이 소요되는 분량임

☞ **종단면 패널조사의 경우에 조사표 분량은 어느 정도입니까?**

**24-1. 일반사항 조사의 경우**

- 1) 20분 이내 쉽게 작성할 수 있는 분량임
- 2) 45분 이내 작성할 수 있는 분량임
- 3) 45분 이상이 소요되는 분량임

**24-2. 추가 조사 시 조사표나 일계부 작성의 경우**

- 1) 20분 이내 쉽게 작성할 수 있는 분량임
- 2) 45분 이내 작성할 수 있는 분량임
- 3) 45분 이상이 소요되는 분량임

**25. 응답자가 조사표 작성이나 면접에 응하기 위하여 필요한 정보 수집과 관련하여 응답 부담은 어느 정도입니까?**

- 1) 낮은 부담으로 정보를 쉽게 제공할 수 있음
- 2) 적당한 부담으로 일부 정보는 회상이 필요함
- 3) 응답부담을 있으며, 필요한 정보를 얻기 위해 복잡한 회상이 필요함

**26. 응답자가 조사표를 작성하거나 면접에 응할 때 민감한 문항에 대한 응답 부담은 어느 정도입니까?**

- 1) 낮은 부담으로 모든 항목을 쉽게 응답할 수 있음
- 2) 적당한 부담으로 일부 민감한 문항이 있음
- 3) 응답부담이 있으며, 매우 민감한 문항이 다수 있음

**27. 다음에 열거된 항목들은 조사표 설계의 품질을 결정하는 항목들입니다. 귀하의 조사 표에 적용하고 있는 항목에 체크해주시기 바랍니다.**

- 1) 전문적이고 세련된 디자인으로 구성됨
- 2) 조사표 서두에 조사의 제목 또는 주제, 후원자, 조사목적에 대한 설명, 응답자의 협조요청, 조사기관, 비밀보호에 관한 정보 등을 제시함
- 3) 장을 구분하기위한 제목이나 머리말이 있음
- 4) 색깔, 삽화, 상징어, 진한 글씨체 사용 등과 같은 그래픽을 추가하여 응답자가 반드시 응답해야 하거나 주의를 환기시키는 표시를 적용함
- 5) 개방형 문항이 모든 응답자가 응답하기에 적절하고 용이함
- 6) 단어와 개념을 일관성 있게 사용하고 있음
- 7) 각 문항들이 합당한 이유가 있음
- 8) 문항들이 응답자의 논리를 따름
- 9) 문항의 배열순서가 건너뛰기 등으로 방향이 적절히 지정됨
- 10) 조사대상기간과 응답단위 등이 명확함
- 11) 조사표 맨 마지막 페이지에 응답자가 추가로 기입할 수 있는 공란 을 마련함.
- 12) 조사표 맨 마지막 페이지에 응답에 대한 감사 문구를 제시함
- 13) 지침과 추가 가이드 작성함
- 14) 문의사항이 있을 때 연결 가능한 무료전화 운영

**28. 27번에서 제시된 사항에 대하여 조사에 사용한 조사표의 품질은 어느 정도라고 판단 합니까?**

- 1) 모든 요구사항을 모두 충족시켜 더 이상 개선이 불필요함
- 2) 일부 문항에 대해 개선의 여지가 있지만, 대부분의 사항이 충분히 충족시킴
- 3) 소수의 문항은 개선이 필요함
- 4) 대부분의 문항에 문제가 있기 때문에 시급히 개선이 필요함

**29. 조사표는 누가 검증 합니까?**

- 1) 내부적으로 검증함
- 2) 통계작성기관 내 별도의 위원회가 검증함
- 3) 내부에 마련된 조사 시험실(survey laboratory)에서 검증함
- 4) 사전조사(pretest)를 위한 외부의 조사시험실에서 검증함
- 5) 외부전문가가 조사표를 검증함

**30. 조사표의 길이, 요구사항, 디자인 측면에서 조사표의 응답부담을 측정하는데 조사표 검증에 사용된 방법은 무엇입니까?**

- 1) 소규모 표본을 선정하여 사전조사(pretest)
- 2) 사전조사(pretest)를 실시한 후 조사원으로부터 보고를 듣는 방법
  - 3) 문항에 대한 응답자 행태/이해에 대한 질적 검증
    - 3)-1 포커스그룹 면접
    - 3)-2 심층(in-depth)면접
    - 3)-3 생각말하기(think-aloud)면접
    - 3)-4 바꾸어 말하기
    - 3)-5 행태 코딩
- 4) 분리표본 검증(split sample testing)
- 5) 기타 (그 외 방법을 기술하십시오) ↴

**♣ “2장. 조사 설계” 절차에 대한 질문 중에서 고려해야 할 사항이 있으면 해당 내용을 기입 하십시오.**

### 3장. 자료 수집

#### ■ 자료 원천

가능한 한 가장 정확한 자료를 얻기 위하여 여러 가지 다른 자료수집 방법을 고려해야 한다. 조사 관리자는 측정오차와 응답부담을 줄이고 비용을 최소화하기 위한 자료수집 방법을 선택해야 한다.

#### 1. 조사에 사용된 자료수집 방법은 다음 중 무엇에 근거하고 있습니까?

- 1) 검증에 의해
- 2) 문헌검토를 통해
- 3) 법적인 요구사항
- 4) 가능한 능력
- 5) 이전 조사에 따라 일관적으로
- 6) 기타(채택 근거를 기입) ↴

#### 2. 다음 어떤 자료수집 방법으로 조사 자료를 수집합니까?

##### 1) 조사원 면접조사방식

###### ☞ 대면 면접

- 1)-1 전자조사표(CAPI)
- 1)-2 종이조사표(PAPI)

→ 28번으로 이동

###### ☞ 전화

- 1)-3 전자조사표(CATI)

→ 24번으로 이동

##### 2) 자기기입 조사방식

- 2)-1 웹 조사표
- 2)-2 전자조사표(CASI)
- 2)-3 종이조사표 또는 일계부

→ 17번으로 이동

- 3) 행정자료나 등록자료

→ 4번으로 이동

- 4) 혼합방식(이 경우 위의 문항 중 필요한 문항에 체크하고, 해당하는 각 문항으로 이동하고, 다음 3번 문항에 답하시오)

- 5) 기타(방법을 기입) ↴

## ○ 혼합방식 조사

## 3. 서로 다른 자료수집방식에 따른 측정효과를 평가 했습니까?

- 1) 완전히 평가  
 2) 충분히 평가  
 3) 개선 여지 있으나 평가함  
 4) 전혀 평가 안 됨

## ○ 행정자료 또는 등록자료

## 4. 사용하는 자료를 어디에서 유지 관리합니까?

- 1) 귀하의 부서  
 2) 귀하의 기관 내 다른 부서  
 3) 외부기관의 다른 부서

## 5. 귀하 스스로 자료의 품질을 평가할 수 있습니까?

- 1) 예 → 8번 문항으로 이동  
 2) 아니오

## 6. 외부로부터 품질평가를 받을 수 있습니까?

- 1) 예  
 2) 부분적으로(아래에 자세히 설명) ↴  
 3) 아니오(아래에 자세히 설명) ↴

## 7. 자료의 품질에 대한 충분한 정보를 갖추고 있습니까?

- 1) 예  
 2) 아니오

## 8. 자료를 다른 자료와 연계할 필요가 있습니까?

- 1) 예  
 2) 아니오 → 12번 문항으로 이동

9. 어떤 형태의 레코드 연계방식을 사용합니까?

- 1) 직접 결합(결정적 또는 확률적 결합)
- 2) 통계적 결합
- 3) 기타(그 외 방법을 기입) ↓

10. 연계성공률을 평가합니까?

- 1) 예
- 2) 아니오 → 12번 문항으로 이동

11. 연계성공률을 어느 정도로 평가합니까?

- 1) 거의 오차 없도록
- 2) 만족할 만큼
- 3) 받아 들일만 하지만 개선이 필요
- 4) 받아들일 수 없음

12. 다른 자료 원을 사용할 경우 어떤 문제가 있습니까?

- 1) 포함범위
- 2) 분류
- 3) 정의
- 4) 갱신 및 시의성
- 5) 중복포함 또는 다중단위 존재

13. 자료 수집 및 교환과정에서 보안성을 어느 정도로 평가합니까?

- 1) 만족할 만큼
- 2) 받아들일 만 하지만 개선 필요
- 3) 받아들일 수 없음

14. 중복된(duplicate)경우와 다중단위(multiple unit)를 구분하고 있습니까?

- 1) 예
- 2) 아니오 → 16번 문항으로 이동

15. 중복단위나 다중단위를 구분할 수 있는 경우 표본추출 전에 확인토록 합니까?

- 1) 예  
 2) 아니오

16. 조사에 등록 및 행정 자료를 효율적으로 사용할 수 있습니까?

- 1) 아니오  
 2) 예 (내용을 설명) ↓

→ 혼합방식이 아닌 경우 50번 문항으로 가시오.

○ 자기기입식 조사

17. 조사표의 지침이나 안내가 어느 정도라고 평가됩니까?

- 1) 매우 탁월함  
 2) 탁월함  
 3) 충분한 정보를 제공함  
 4) 개선이 필요함  
 5) 부적합함

18. 응답자가 문의사항이 있을 때 연결 가능한 무료전화를 운영하고 있습니까?

- 1) 예  
 2) 아니오

19. 문의하는 질의에 답변을 위한 팀이 구성되어 있습니까?

- 1) 예  
 2) 아니오

20. 컴퓨터를 이용한 조사를 실시합니까?

- 1) 예  
 2) 아니오 → 혼합방식이 아닌 경우 45번 문항으로 이동

21. 조사표에 대한 일관성을 점검합니까?

- 1) 예
- 2) 아니오

22. 항목 무응답을 방지하기 위해 개발된 모듈이 있습니까?

- 1) 예
- 2) 아니오

23. 컴퓨터를 이용하여 코딩을 하고 있습니까?

- 1) 예
- 2) 아니오

➔ 혼합방식이 아닌 경우 45번 항목으로 가시오.

○ 컴퓨터를 이용한 전화조사(CATI)

24. 전화조사를 어디에서 실시합니까?(중복선택가능)

- 1) 회사 내
- 2) 콜센터
- 3) 외주
- 4) 조사원의 자택에서의 전화조사 허용

25. 현장작업을 위하여 전화걸기 스케줄 프로그램이 있습니까?

- 1) 예
- 2) 아니오

26. 전화접촉건수에 대한 통계를 산출하고 있습니까?

- 1) 예
- 2) 아니오

**27. 어떤 방법으로 전화번호를 추출하고 있습니까?**

- 1) 표본추출틀에서 직접 추출함
- 2) 이전 수집한 자료에서 추출함
- 3) 임의전화걸기(RDD)
- 4) 다른 자료 (그 외 추출방법을 기입) ↴

**■ 조사원 채용 및 교육**

면접방법은 응답자와 관계형성에 막대한 영향을 끼친다. 이런 관계를 최적화하고 측정 오차를 최소화하려면 조사원의 채용과 교육방법에 주의를 기울여야 한다.

**28. 조사원 채용을 위한 특별한 절차가 있습니까?**

- 1) CAPI, PAPI 또는 CATI 실시를 위해 특별하게 채용
- 2) 조사원이 갖추어야 할 개인적인 요구사항 목록이 있음
- 3) 조사원이 갖추어야 할 최소의 자격요건 목록이 있음
- 4) 어떠한 최소의 자격요건도 고려하지 않음

**29. 조사를 위해 어떻게 조사원들을 준비시킵니까?**

- 1) 다양한 과정으로 구성된 집중교육, 최소의 준비교육(의사소통), 특별 교육과정, 잘 구성된 조사원 매뉴얼 제공함
- 2) 한 과정에서 적절한 교육을 받고, 일부 조사원 매뉴얼 제공함
- 3) 응답거부 전환(무응답자 설득)기법과 조사 참여를 유도하는 방법에 대한 특별한 교육을 받음
- 4) 일부 훈련을 시키고 조사원 매뉴얼을 제공하지만 충분하지 않음
- 5) 기타(그 외 방법을 기입) ↴

**30. 조사원들이 면접을 성공적으로 마칠 때마다 보너스를 받으니까?**

- 1) 예
- 2) 아니오

**31. 조사원은 어떠한 방식으로 채용합니까?**

- 1) 중앙채용방식
- 2) 중앙채용방식이 아님

**32. 예비 조사원들이 있습니까?**

- 1) 조사원 각 그룹별로 특별히 교육받고 접촉하는 사람이 있음
- 2) 정기적인 모임을 가짐
- 3) 조사원을 요구할 경우 예비조사원을 지원해줌
- 4) 예비조사원이 전혀 없음

**33. 조사원을 관리하는 방법이 있습니까?**

- 1) 주기적으로 피드백을 받음
- 2) 새로 채용된 조사원들은 조사 시 보통 2회 정도 점검을 받음
- 3) 의심스러운 사실이 있는 경우 특별 관리함
- 4) 실패한 경우, 마감시간 준수 및 검토결과 등 감안한 성과관리
- 5) 면접의 \_\_\_\_\_ %를 주기적으로 성과관리
- 6) 지침 준수여부, 표본대상, 조사표 기입방법 등에 대한 기본적인 관리
- 7) 조사원에 대한 관리수단이 없음
- 8) 기타(그 외 조사원관리에 이용한 방법을 서술하십시오) ↴

**34. 조사가 있다는 사실을 미리 각 가구(또는 사업체)에 우편, 이메일 또는 전화로 공지합니까?**

- 1) 예
- 2) 아니오

**35. 평균적으로 한명의 조사원이 몇 회 면접을 실시합니까?**

**36. 한명의 조사원이 실시하는 최대 면접 회수 어느 정도입니까?**

37. 조사원에 대한 인적사항을 알고 있습니까?

- 1) 예  
 2) 아니오

38. 조사원의 인구학적 속성과 목표 집단의 응답 패턴 간에 어떠한 영향을 주는 것으로 분석됩니까?

- 1) 조사원의 인구학적 사항이 조사 참여행태에 유의미한 영향을 줌  
 2) 조사원의 인구학적 사항이 조사 참여행태에 약간의 영향을 줌  
 3) 조사원의 인구학적 사항이 조사 참여행태에 어떤 영향도 주지 않음  
 4) 이와 같은 정보를 갖고 있지 않음

39. 면접가능 시간대(timing)에 대한 정보를 파악하고 있습니까?

- 1) 예  
 2) 아니오

40. 면접가능 시간대 대해 어떤 정보를 가지고 있습니까?

- 1) 월별 시간대  
 2) 주별 시간대  
 3) 지정한 날짜  
 4) 대략적 시간(오전, 오후, 저녁)대  
 5) 정확한 시간대  
 6) 기타(시간대 유형을 기입) ↴

41. 조사에 컴퓨터를 활용합니까?

- 1) 예  
 2) 아니오 → 45번 문항으로 이동

42. 조사표에 대한 일관성을 점검합니까?

- 1) 예  
 2) 아니오

43. 항목무응답을 방지하기 위한 모드를 개발하여 사용합니까?

- 1) 예
- 2) 아니오

44. 컴퓨터를 이용하여 코딩을 합니까?

- 1) 예
- 2) 아니오

■ 무응답 줄이기

무응답을 줄이면 표본크기가 축소되지 않도록 하고 편향을 줄일 수 있다.

45. 무응답을 방지하기 위하여 즉, 응답을 독려하기 위하여 다음 어떤 방법을 사용합니까?

- 1) 1회의 추적조사를 \_\_\_\_\_일 후에 실시
- 2) 2회 이상 추적조사를 \_\_\_\_\_일 과 \_\_\_\_\_일 후에 실시
- 3) 추적조사에 우선순위가 있는데 사업체조사 또는 그와 유사한 조사에서는 영향을 크게 미치는 단위를 우선적으로 실시함
- 4) 추적조사에 우선순위가 있는데 무응답 편향의 영향이 큰 지역의 단위를 우선적으로 실시함
- 5) 응답자의 문의사항에 대한 해결을 위하여 표준 절차 및 기반을 갖추고 있음
- 6) 응답과정을 과거조사의 점검 결과에 따라 설계
- 7) 기타(그 외 방법을 기입) ↓

46. 컴퓨터 면접(CAPI, CATI 또는 CASI)에 필요한 사전 타당성 점검시스템이 있습니까?

- 1) 예
- 2) 부분적으로
- 3) 아니오

47. 응답자 또는 자료제공자가 컴퓨터에 의한 면접 방법과 전통적인 방법 간에 선택권이 있다면 몇 % 정도가 컴퓨터에 의한 면접 방법을 선택합니까?

- 1) \_\_\_\_\_%
- 2) 모름

48. 컴퓨터에 의한 면접(CAPI, CATI 또는 CASI)에 필요한 사전 타당성 점검시스템이 있다면 사무실에서 추가적으로 점검을 합니까?

- 1) 자료에 오차가 (거의) 없기 때문에 점검하지 않음
- 2) 약간 오차가 있지만, 추가점검은 불필요함
- 3) 몇몇 오차를 수정하기 위해 추가점검 필요함
- 4) 오차를 수정하기 위해 추가점검이 필수적임

49. 응답내용이 타당하지 않을 때 (전화상으로) 응답자에게 다시 응답내용을 확인하는 절차를 거칩니까?

- 1) 예
- 2) 아니오

#### ■ 현장실사

실사작업 계획을 수립하는 것은 최소의 비용으로 자료를 수집하고 최적의 조사기간을 설정하기 위해 필요합니다.

50. 조사 자료를 사용하는 경우, 향후 행정기록과 같이 활용 가능한 자료를 사용가능한지 점검했습니까?

- 1) 예
- 2) 아니오

51. 법적인 측면에서 행정자료나 등록 자료를 사용하는 것이 어느 정도로 제한되고 있습니까?

- 1) 법적인 제한이 전혀 없음
- 2) 법적인 제한이 다소 있지만, 사용가능함
- 3) 법적인 제한이 매우 심함
- 4) 조사 분야에서 완전히 금지함

52. 현장실사 기간이 조사대상 기간을 완전히 포함하고 있습니까?

- 1) 예
- 2) 아니오

53. 현장실사 기간은 어느 정도 소요됩니까?

54. 조사기간 동안 조사결과에 영향을 미칠 수 있는 휴가나 파업 또는 천재지변 등과 같이 예상치 못한 사건이 있었는가?

- 1) 그런 영향은 없음
- 2) 자료수집기간 중 \_\_\_\_\_주 동안 휴가기간이 있었음
- 3) 자료수집기간 중 \_\_\_\_\_이 있었음

55. 조사원 중 슈퍼바이저와 같이 실사감독을 맡는 사람의 비율은?

56. 조사과정에서 발생할 수 있는 문제점들에 대해 어떠한 조치합니까?

- 1) 즉시 연락할 수 있는 슈퍼바이저가 있음
- 2) 슈퍼바이저가 지속적으로 점검
  - 3) 오차의 빈도 및 원인에 대한 정보를 피드백 하여 보고하고 있음
- 3)-1 조사관리자에게 보고
- 3)-2 주제관련 전문가에게 보고
- 3)-3 조사방법 전문가에게 보고
- 4) 참여자 의견수렴
- 5) 조사 시작부터 끝날 때까지 표본단위를 추적하는 전체적인 자료수집과정에 대한 표본관리 절차가 있음
- 6) 기타(그 외 조치방법을 기술) ↴

**57. 사업체조사의 경우 다음 사항을 확인합니까?**

- 1) 조사대상 사업체의 관계자에게 연락
- 2) 가능한 쉽게 정보를 얻을 수 있는 적당한 시간에 연락
- 3) 응답자에게 편리한 방법 및 형태로 자료를 제공하도록 배려

**58. 조사완료 후 입력되는 자료를 처리하는 과정은 어느 정도로 평가할 수 있습니까?**

- 1) 매우 만족스러움
- 2) 만족스러움
- 3) 받아들일 만 하지만, 개선필요
- 4) 받아들일 수 없음

**59. 조사완료 후 입력 및 제공되는 마이크로자료의 보안유지는 어느 정도입니까?**

- 1) 전자파일은 암호화
- 2) 들어오는 조사표에서 식별정보는 제외
- 3) 기타(그 외 내용을 기입) ↴

♣ “3장. 자료수집” 절차에 대한 질문 중에서 고려해야 할 사항이 있으면 해당 내용을 기입 하십시오.

## 4장. 자료 입력 및 자료 처리

### ■ 자료입력

원시자료를 컴퓨터로 해독 가능한 형식으로 변환하고 코딩하는 과정에서 오차가 발생할 수 있다. 이러한 오차를 최소화하기 위하여 다양한 방법을 고려해야한다.

#### 1. 자료를 어떻게 입력합니까?

- 1) 수작업으로 입력
  - 2) 전자적 작업으로 입력
    - 2-1) 문자인식(MCR)
    - 2-2) 지능적 문자인식(ICR)
    - 2-3) 광학적 문자인식(OCR)
    - 2-4) 음성인식(VRE)
    - 2-5) 전자 교환(EDI)
- 3) 자료 수집 시(컴퓨터에 의한 면접)
- 4) 기타(그 외 입력방식을 기입) ↓

#### 2. 자료를 어떻게 코딩합니까?

- 1) 수작업
- 2) 특별히 설계된 소프트웨어에 의한 자동코딩
- 3) 기타(그 외 방법을 기입) ↓

#### 3. 수작업으로 코딩할 경우 코드를 관리하거나 일관성을 점검하는 어떤 수단이 있습니까?

- 1) 아니오
- 2) 예
  - 2-1) 종속적인 검증
  - 2-2) 독립적인 검증

## ■ 에디팅 절차

에디팅 절차는 결측치, 부적절한 자료, 일관성 없는 자료 등을 식별하고 오류가능성이 있는 자료를 선별하기 위하여 반드시 실시하여야 한다. 이때, 에디팅(editing)이란 단순히 오류 가능성이 있는 자료를 식별해 내는 과정을 의미한다.

### 4 원시자료 에디팅의 필요성에 대해 어떻게 평가하십니까?

- 1) 원시자료에서 오류가 많이 발견되었고 점검은 필수 불가결함
- 2) 원시자료에서 수정되어야 할 몇 가지 오류가 발견되었음
- 3) 원시자료에서 몇 가지 오류가 발견되었으나 수정할 필요는 없음  
→ 20번으로 이동
- 4) 통합된 타당성 체크시스템으로 자료수집과정에서 이미 충분히 점검됨  
→ 20번으로 이동
- 5) 아직까지는 원 자료에 오류가 없고 에디팅 과정도 불필요함  
→ 20번으로 이동

### 5. 어떤 에디팅 방법을 사용하십니까?

- 1) 자동화된 방법
- 2) 수작업
- 3) 두 가지 혼합

### 6. 채택하고 있는 에디팅 방법은 다음 중 어떤 자료를 이용하는가?

- 1) 무응답자 또는 다른 보조 자료에 대해 이용 가능한 자료만을 이용함
- 2) 주어진 조사에 대해 다른 응답자들의 관측 자료만을 이용함
- 3) 위의 두 가지를 혼합
- 4) 기타(그 외 방법을 기입) ↓

### 7. 에디팅 절차를 사전점검 하였는가?

- 1) 예
- 2) 아니오

**8. 다음 중 어떤 에디팅 절차를 원시자료에 적용하고 있습니까?(중복체크가능)**

- 1) 모든 변수에 대한 범위 점검
- 2) 논리적 관계에 근거한 논리적 점검
- 3) 이상치 분석 기법과 같은 통계적 검증
- 4) 논리적 점검 및 통계적 검증의 혼합 형식
- 5) 동일 통계자료의 과거 자료 또는 다른 자료간의 비교
- 6) 경험적인 점검 절차
- 7) 기타(그 외 방법을 기입) ↓

**9. 수작업을 통한 에디팅 담당자에게 문서화된 명백하고 상세한 에디팅 지침서를 제공합니까?**

- 1) 정확하고 최근 갱신된 지침서가 주어짐
- 2) 일부 지침만 상세함
- 3) 아니오

**10. 결측치 또는 0값 등의 개념에 대한 구분이 있습니까?**

- 1) 예
- 2) 아니오

**11. 에디팅 과정에 우선순위가 정해져 있습니까?**

- 1) 아니오
- 2) 심각한 오류는 보통 수작업으로 많은 노력을 들여 에디팅을 하고 대부분 자동으로 에디팅 되는 사소한 오류들은 덜 신경을 씀
- 3) ID로 사용되는 변수의 수정작업을 우선적으로 실시(만일 조사 자료가 가중치를 계산하거나, 추정을 위해 보조 자료와 결합될 경우)
- 4) 기타(그 외 방법을 기입) ↓

**12. 과도한 에디팅이나 자료에 새로운 오류를 방지하기 위해 어떤 조치를 취합니까?**

- 1) 에디팅이 최종결과에 어떤 영향을 주는지를 분석하는 오류 인식 메커니즘이 있어 에디팅 최적점(Optimum of Editing)을 결정함
- 2) 수정된 자료에 대해서 다시 에디팅하는 과정이 있음
- 3) 지나친 에디팅을 방지하는 아무런 절차가 없음
- 4) 기타(그 외 방법을 기술) ↴

**13. 에디팅 작업이 지속적으로 진행될 수 있습니까? 예를 들면 일부자료를 수집했을 때부터 가능합니까?**

- 1) 예
- 2) 아니오

**14. 에디팅 과정을 개선하기 위하여 구축해 놓은 과거 조사의 오류에 대한 이용가능한 통계가 있습니까?**

- 1) 예
- 2) 아니오

**15. 조사의 품질측정을 하고 향후 조사의 개선사항을 제시하기 위한 오류에 대한 통계를 산출합니까?**

- 1) 예
- 2) 아니오

**16. 귀하가 사용하는 에디팅 시스템의 신축성은 어느 정도라고 생각하십니까?**

- 1) 매우 신축적 - 새로운 에디팅 과정이나 수정된 에디팅 과정을 쉽게 적용가능
- 2) 충분히 신축적 - 큰 문제없이 새로운 에디팅 과정이나 수정된 에디팅 과정을 적용가능
- 3) 아님 - 새로운 에디팅 과정이나 수정된 에디팅 과정을 적용하는데 시간과 비용이 소요됨
- 4) 기타(그 외 내용을 기술) ↴

**17. 오류의 원인에 대한 정보가 있습니까?**

- 1) 오류 원인을 알 수 없음
  - 2) 오류의 원인을 부분적으로 알 수 있는 경우
    - 2)-1 측정오류
    - 2)-2 코딩오류
    - 2)-3 기타(그 외 원인을 기술) ↴

- 3) 오류의 원인을 알 수 있는 경우
  - 3)-1 측정오류
  - 3)-2 코딩오류
  - 3)-3 기타(그 외 원인을 기술) ↴

**■ 대체과정(Imputation Procedure)**

대체과정은 자료의 기본적인 분포를 유지하면서 오류를 수정하고 결측된 자료를 대체하기 위해 이용되는 절차이다.

**18. 에디팅 과정에 의해 발견된 오류를 어떻게 수정합니까? 적용하고 있는 방법에 따라 얼마나 에디팅 과정에 포함하고 있는지를 %로 나타내 주십시오.**

- 1) 조사원 또는 응답자를 재 접촉 \_\_\_\_%
- 2) 오류변수를 없앴 \_\_\_\_%
- 3) 오류 또는 결측 변수는 좀 더 적절한 값으로 대체 \_\_\_\_%
- 4) 오류 또는 결측 변수는 대체값(Imputed Value)으로 대체 \_\_\_\_%
- 5) 잘 모르겠음
- 6) 기타 ↴

**19. 통계 값에 대한 대체를 합니까?**

- 1) 예
- 2) 아니오 ➔ 5장으로 이동

**20. 채택하고 있는 대체방법은 무엇입니까?**

- 1) 자동화 대체
- 2) 수작업 대체
- 3) 위의 두 가지 방법을 혼합

**21. 채택하고 있는 대체방법은 다음 어떤 자료를 이용합니까?**

- 1) 무응답자에 대해 이용 가능한 자료 또는 다른 보조자료 (논리적이며 연혁적인)를 이용함
- 2) 무응답자에 대해 이용 가능한 자료, 다른 보조자료, 조사의 다른 응답단위의 관측 자료를 이용함
- 3) 주어진 조사의 다른 응답 단위의 관측 자료만 이용함
- 4) 기타(그 외 이용 자료를 기술) ↓

**22. 채택하고 있는 대체방법은 다음 중 어떠한 방법으로 대체하는가?**

- 1) 평균대체와 같은 자동생성 방법을 적용
- 2) 회귀대체와 같은 예측 방법을 적용

**23. 조사에 적용하고 있는 대체방법은 다음 중 무엇입니까?****☞ 1) 결정론적인 방법**

- 1-1) 논리적 대체(Logical imputation)
- 1-2) 연혁적 대체(Historical imputation)
- 1-3) 평균 대체(Mean imputation)
- 1-4) 비율 및 회귀 대체(Ratio and Regression imputation)
- 1-5) 최근방 대체(Nearest-neighbour imputation)

**☞ 2) 확률적인 방법**

- 2-1) 핫덱 대체(Hot deck imputation)
- 2-2) 확률잔차회귀 대체(Regression with random residual imputation)
- 2-3) 신경망 대체(Neural network imputation)
- 2-4) 확률 잔차를 추가한 결정론적 방법

**24. 다음 대체 방법 중 어떠한 대체 방법을 사용합니까?**

- 1) 단순대체(결측치가 하나의 대체값으로 사용됨)
- 2) 다중대체(여러 개의 값이 대체값으로 사용됨)

**25. 어떤 범용 시스템을 대체 알고리즘에 적용하고 있습니까?**

- 1) \_\_\_\_\_ 시스템을 모든 대체과정에 적용하고 있음
- 2) \_\_\_\_\_ 시스템을 대부분의 대체과정에 적용하고 있음
- 3) \_\_\_\_\_ 시스템을 부분적으로 대체과정에 적용하고 있음
- 4) 어떠한 범용시스템도 적용하고 있지 않음

**26. 여러 가지 방법 또는 제공자(donors)를 적용할 수 있다면, 적용결과들이 다른 대체 방법에 비해 최선의 방법으로 평가되어 있습니까?**

- 1) 예
- 2) 부분적으로
- 3) 아니오

**27. 변수와 기본적인 분포간의 관계가 유지된다는 측면에서 대체방법은 내부적으로 일관성이 있고 받아들일 수 있습니까?**

- 1) 예
- 2) 아니오
- 3) 모름

**28. 대체값에 어떠한 표식(flag)이 있습니까?**

- 1) 예
- 2) 아니오

**29. 대체과정에서 원시 자료값은 보존됩니까?**

- 1) 예
- 2) 아니오

**30. 원시자료의 몇 %가 대체 됩니까?**

\_\_\_\_\_ %의 레코드가 대체됨

31. 대체과정이 모든 값을 관측하지 못하기 때문에 발생하는 편향을 방지할 수 있다면, 실제로 이를 평가합니까?

- 1) 예
- 2) 아니오

32. 대체에 의한 분산을 측정합니까?

- 1) 대체에 의한 분산은 \_\_\_\_\_ 임
- 2) 이러한 분산은 다음과 같은 방법으로 분석함(방법을 기술) ↴

- 3) 여러 방법을 시도했으나, 대체에 의한 분산은 측정 불가능함.  
그 이유는 다음과 같음(이유를 기술) ↴

- 4) 대체에 의한 분산을 측정하려는 아무런 노력도 하지 않음.  
그 이유는 다음과 같음(이유를 기술) ↴

33. 대체과정에서 수정을 쉽게 하기위하여 이용 가능한 소프트웨어 패키지나 충분히 신속성 있는 절차를 사용합니까?

- 1) 예
- 2) 부분적
- 3) 아니오

♣ “4장. 자료입력 및 처리” 절차에 대한 질문 중에서 고려해야 할 사항이 있으면 해당 내용을 기입 하십시오.

## 5장. 자료 분석 및 통계결과의 품질 - 자료의 품질진단

각 통계작성기관에서는 자료의 품질을 진단할 필요가 있다. 비록 품질차원에 대한 기본적인 여건과 목표가 설계과정(2장)에서 이미 정해졌지만, 자료수집 및 처리가 완료된 후에도 품질차원인 관련성, 정확성, 시의성, 접근가능성 및 명확성, 비교가능성 및 일관성 등의 관점에서 자료에 대한 진단이 필요하다. 이러한 의미에서 5장은 주요부분이 자료의 품질을 진단을 목적으로 함과 동시에 향후 자료 분석과 관련된 많은 질문항목으로 구성되어 있다.

### ■ 관련성(Relevance)

**관련성**은 작성된 통계가 현재와 미래의 이용자들의 수요를 충족하는 정도를 나타낸다. 따라서 필요로 하는 모든 통계가 생산되는지와 정의, 분류 등과 같은 개념들이 이용자 수요를 반영하는 정도를 의미한다.

#### 1. 조사에 대한 이용자 만족도를 측정하기 위하여 다음 어떤 방법을 사용합니까?

- 1) 정기적으로 조사에 대한 이용자만족도 조사를 수행
- 2) 이용자 피드백에 의한 체계적인 진단
- 3) 정기적인 이용자 자문을 통한 피드백
- 4) 이용자만족도 지수 산출
- 5) 사용하지 않음
- 6) 기타(해당 사항을 기술) ↴

#### 2 주요 이용자 만족도에 대한 이용 가능한 정보가 어느 정도로 평가합니까?

- 1) 주요이용자 및 다른 이용자에 대한 정보가 거의 없음
- 2) 주요이용자 정보는 일부 있지만, 다른 이용자 정보는 거의 없음
- 3) 주요이용자 정보와 다른 이용자 정보가 일부 있음
- 4) 주요이용자 정보는 많이 있으나, 다른 이용자 정보 일부 있음
- 5) 주요이용자 및 다른 이용자에 대한 완전한 정보가 있음

**3** 조사의 전반적인 관련성을 귀하께서는 어느 정도로 평가하십니까?

- 1) 거의 관련 없음
- 2) 주요 분야는 약간 관련이 있고, 다른 분야는 거의 관련이 없음
- 3) 주요 분야 및 다른 분야 모두 약간은 관련성이 있음
- 4) 주요 분야는 관련성이 높고, 다른 분야는 약간 관련성이 있음
- 5) 주요 분야 및 다른 분야 모두 동일하게 관련성이 높음

**■ 정확성(Accuracy)**

일반적인 통계적 관점에서 **정확성**이란 계산 값 또는 추정 값이 미지(unknown)의 정확한 값 또는 참값에 얼마나 근접한가를 나타낸다. 정확성을 확보하고 평가하는 것은 기본적으로 표본오차와 포함률 오차, 측정오차, 처리오차, 모형가정오차, 무응답편향 등에 따른 비표본오차 등을 평가함으로써 가능하다.

**4. 추정 값의 정도를 측정하기 위하여 다음 어떤 값을 제시하고 있습니까?**

- 1) 평균제곱오차(MSE) **⇒ 5번 문항으로 이동**
- 2) 변동계수(CV) 또는 상대표준오차(RSE) **⇒ 6번 문항으로 이동**
- 3) 신뢰구간 **⇒ 8번 문항으로 이동**
- 4) 정성적인 정확도를 나타내는 문장뿐임 **⇒ 8번 문항으로 이동**
- 5) 어느 것도 제시하지 않음 **⇒ 8번 문항으로 이동**

**5. 만일 평균제곱오차를 산출하고 있다면, 이 값이 이용자의 요구를 충족합니까?**

- 1) 아니오
- 2) 예

**6** 주요 변수에 대한 변동계수는 어떠합니까?

- 1) 모든 통계항목에 대한 변동계수가 매우 크다.
- 2) 대부분 많은 항목에 대한 변동계수가 크다.
- 3) 관련성 있는 항목과 주요 부문에 대한 변동계수는 중간수준이다.
- 4) 관련성 있는 항목 대부분과 주요 분류수준에서 변동계수가 작다.
- 5) 모든 주요항목과 모든 주요 분류수준에서 변동계수가 작다.

7. 변동계수의 수준이 조사목적에 대해 받아들일 만한 수준인가?

- 1) 예
- 2) 아니오
- 3) 부분적으로 받아들일 만함(이유를 설명) ↴

8. 분석될 자료(data set)에 대한 추정량의 형태가 Horvitz-Thompson(가중변수의 표본추출가중치가 표본으로 추출된 단위의 포함확률의 역수와 같다)추정량 인니까?

- 1) 예
- 2) 아니오

9. 최종적으로는 원하는 정확도를 얻기 위해 표본크기가 충분히 크게 설계되었습니까?

- 1) 예
- 2) 아니오

10. 다음 중 어떠한 편향을 조정하기위해 추정과정(보정 또는 재가중을 포함)에 보조정보를 사용하였습니까?

- 1) 무응답 편향을 조정하기위해
- 2) 표본 추출률 상의 오차(비포함률 조정)를 조정하기위해
- 3) 측정오차를 조정하기위해
- 4) 재가중 사용안함 → 12번 문항으로 이동
- 5) 다른 형태의 오차(그 외 편향을 기술) ↴

11. 만일 재가중 방법을 실시한다면 어떤 보정 과정을 사용하였습니까?

- 1) 사후층화
- 2) 회귀추정
- 3) 비 추정
- 4) 기타(그 외 방법을 기술) ↴

12. 만일 모형-보조(model-assisted) 추정방법을 사용하였다면, 보조변수나 기준변수가 추정 변수와 상관이 있는지를 어떻게 확인하였는지 방법을 기술하시오.

13. 일반화된 추정 소프트웨어를 사용합니까?

- 1) 아니오, 자체개발 시스템  
 2) 예(구체적으로 기술)

14. 표본자료에 대해 가중변수의 합을 계산함으로써 제시된 가중치의 정확도는 확보되어 있습니까?

- 1) 예  
 2) 아니오

#### 15 조사에서 단위무응답의 범위는 어느 정도입니까?

서로 다른 조사유형에 따라 서로 다른 전제조건이 주어지기 때문에, 의무적 조사, 단기 자발적 조사, 자발적 조사에 대한 단위무응답을 판단하는 기준은 구별되어야 한다. 단위무응답에 대해 다음 범주를 활용하시오.

|              | 의무적조사    | 단기자발적조사 | 자발적조사   |
|--------------|----------|---------|---------|
| 낮은수준의 단위무응답율 | < 5%     | < 20%   | < 40%   |
| 중간수준의 단위무응답율 | 5% - 15% | 20%-40% | 40%-60% |
| 높은수준의 단위무응답율 | >15%     | > 40%   | > 60%   |

- 1) 단위무응답율을 모르거나, 허용할 수 없을 만큼 높음  
 2) 높은 수준의 단위무응답율  
 3) 중간수준의 단위무응답율  
 4) 낮은 수준의 단위무응답율  
 5) 거의 단위무응답이 없음

16. 무응답을 응답거절, 못 만남, 부재등과 같은 세부 범주로 분류되어 있습니까?

- 1) 예  
 2) 일부  
 3) 아니오

**17. 단위 무응답 편향을 추정하기 위한 이용 가능한 자료가 있습니까?**

- 1) 외부 자료와 연계(예, 행정자료파일)
- 2) 무응답자중 부차표본을 추출하는 일시적 조사를 통한 자료
- 3) 비교를 위하여 사용된 무응답자의 배경정보 자료
- 4) 무응답에 의한 편향을 추정하기위한 자료가 없음
- 5) 기타(그 외 자료를 열거) ↴

**18 조사의 항목무응답은 어느 정도라고 평가하고 있습니까?**

- 1) 항목무응답을 모르거나 아주 높음(50%이상)
- 2) 항목무응답이 높음(15%이상 50%미만)
- 3) 항목무응답이 약간 있음(5%이상 15%미만)
- 4) 항목무응답이 낮음(5%미만)
- 5) 항목무응답은 거의 없음

**■ 시의성(Timeliness) 및 정시성(Punctuality)**

어떤 통계든지 적절한 시간 내 결과를 작성하고 일정을 준수하는 것은 중요하다. 이러한 사실은 사용자 요구 중에서 가장 중요한 것 중 하나일 것이다. 왜냐하면, 이용자가 결과를 효과적으로 사용할 수 있기 때문이다.

정보의 **시의성**은 결과를 나타내고자 하는 상황이나 사건과 결과자료를 이용할 수 있는 시간과의 차이 이다. **정시성**은 자료의 공표일자와 공표목표일자간(공식적으로 데이터를 공표를 표명한 날짜, 규칙으로 정해놓거나, 또는 사전에 부처간에 합의된 날짜)의 시차를 의미한다.

**19. 조사대상기간과 결과를 발간하는 시점사이의 시차를 일수 또는 월수로 나타내시오.**

20. 계획한 발간일자와 실제 발간일자 간의 시차를 일수 또는 월수로 나타내시오.

21 조사대상기간과 잠정 또는 최종 결과의 최초 발간일 간의 시차는 어느 정도입니까?

- 1) 상당한 시차가 있음
- 2) 다소 긴 시차가 있음
- 3) 정해진 시차가 있음(합의된 시차)
- 4) 약간의 시차가 있음
- 5) 매우 작은 시차가 있음

22 주요 발간물에 대해 정해진 발간시기를 보통 준수합니까?

- 1) 자주 많이 지연됨
- 2) 어느 정도 지연으로 자주 맞지 않음
- 3) 발간물의 50%이상 주로 지킴
- 4) 발간물의 80%이상 보통 지킴
- 5) 거의 언제나 지킴

23. 시의성을 개선하기 어떠한 변화가 필요합니까?

- 1) 사소한 내부 변화
- 2) 중대한 내부 변화
- 3) 외부 협력기관이 관련된 변화
- 4) 기타(내용을 기술) ↴

#### ■ 비교가능성(Comparability)

**비교가능성**은 통계가 지역들 간의 비교, 비지역적으로 또는 시계열적으로 비교될 경우 통계적 개념, 측정도구, 측정과정의 적용에서의 서로 다른 영향을 측정하고자 하는 것이 목적이다. 통계들 간의 차이는 통계적 특성치의 참값 간의 차이에 기인한다고 할 수 있다.

**24 다른 통계의 중요한 개념과 조사 설계에 연계하여 지역에 대한 통계적 결과물 비교가능성은 어느 정도인 것으로 평가되니까?**

- 1) 연관된 통계와 심각한 차이가 있음
- 2) 연관된 통계와 다소 큰 차이가 있음
- 3) 연관된 통계와 약간의 차이가 있음
- 4) 연관된 통계와 다소 차이가 있음
- 5) 연관된 통계와 거의 차이가 없음

**25 조사 대상기간과 관련하여 시계열적인 통계 결과물의 비교가능성은 어느 정도인 것으로 평가되니까?**

☞ 시계열적 비교가능성은 개념과 조사설계상의 변화에 따른 것으로 제한될 수 있다.

- 1) 기본적인 변화나 문제 때문에 시계열 비교가 안 됨
- 2) 시계열적인 비교가능성이 심각하게 제약됨
- 3) 시계열적인 비교가능성이 어느 정도 제약됨
- 4) 시계열적인 비교가능성이 약간 제한됨
- 5) 시계열적인 비교가능성에 거의 제한 없음

■ **일관성(Coherence)**

통계의 **일관성**이란 다른 방법으로 확실하게 결합되거나 다양한 용도에 쓰일 때의 적절성을 말한다.

통계가 단일 원시자료로부터 산출되었을 때, 관련된 조사로부터 구한 기본적인 결과를 보다 복잡한 결과 산출하기위해 다양한 방법으로 신뢰성 있게 결합될 수 있다는 측면에서 통계는 정상적인 일관성을 가진다고 한다.

통계가 다른 원시자료 특히 다른 속성 또는 주기로 작성되는 통계조사로부터 산출되었을 때, 다른 접근방법, 분류 및 방법론적 기준에 근거할 수 있다는 측면에서 완전히 일관적이지 못할 수 있다. 관련된 결과로부터 산출된 통계는 완전히 일관되지 못한 결과가 나올 수 있으며, 이 경우 사용자는 나타날 수 있는 가능한 효과를 명확하게 밝혀야 한다.

26. 월별 또는 연간 단위로 동시에 다른 주기로 조사하고 있습니까?

- 1) 예
- 2) 아니오 → 29번 문항으로 이동

27 서로 다른 주기에 대한 결과와 관련하여, 귀하가 생산하는 통계의 일관성은 어느 정도라고 평가하십니까? 다시 말하면, 동일한 조사대상기간에 대해 다른 주기의 결과를 확실하게 결합할 수 있습니까?

- 1) 상당한 문제가 있음
- 2) 다수의 상반된 점이 있음
- 3) 일부 상반된 점이 있음
- 4) 극히 적은 문제
- 5) 심각한 문제없음

28 동일한 사회-경제 분야에서 귀하가 작성하고 있는 통계의 일관성을 어느 정도라고 평가하십니까? 중요한 비공식 통계를 포함한 다른 원시자료로부터 작성된 통계를 확실하게 결합시킬 수 있습니까?

- 1) 상당한 차이가 있음
- 2) 다수의 차이가 있음
- 3) 일부 차이점이 있음
- 4) 극히 적은 차이
- 5) 심각한 문제없음

#### ■ 자료 분석

자료 분석의 목적은 자료의 주요 결과를 탐색하고, 목표 집단에 적합한 방향으로 공표를 위한 데이터를 준비하는 것이다.

**29. 이용자에게 맞춤분석을 제공하고자 할 때 신속하고 유연하게 대응할 수 있는 능력이 어느 정도라고 평가하십니까?**

- 1) 요구하는 통계와 관계된 거의 모든 분석에 대한 능력과 전문성을 갖추고 있기 때문에 이용자에게 맞춤분석을 아주 신속하고 유연하게 대응할 수 있다.
- 2) 요구하는 통계와 관계된 대부분의 분석에 대한 능력과 전문성을 갖추고 있기 때문에 대부분의 경우 맞춤분석 요구에 대응할 수 있다.
- 3) 요구하는 통계와 관계된 분석능력과 전문성이 없기 때문에 유감스럽게도 대부분 맞춤분석 요구에 대응할 수 없다.
- 4) 자료의 분석에 대한 이용자들의 요구를 만족시킬 의무가 없다.

**30. 사용하고 있는 추가분석 방법은 다음 중 어떤 것입니까?**

- 1) 모델링(모형화 작업)
- 2) 공간분석
- 3) 지수산출
- 4) 계절 또는 추세 분석과 추정
- 5) 기타(구체적으로 기술) ↴

**31. 작년 한 해 동안 자료 분석기법을 새롭게 도입했거나 기존방법을 개편한 적이 있습니까?**

- 1) 예
- 2) 아니오

**■ 보안관리**

응답자 정보에 대한 기밀은 가능한 자료의 유용성을 유지하면서 최대한 준수되어야 한다.

**32. 발간물 또는 (준)일반 이용자에 대한 자료접근에 대해 어떤 형태의 자료를 보유하고 있습니까?**

- 1) 빈도 표
- 2) 다량의 통계표
- 3) 마이크로자료-연구자용 파일
- 4) 마이크로자료-일반이용자용 파일

**33. 어떤 기준에 의하여 통계표에서 제거해야 할 민감한 셀을 정의하고 있습니까?**

- 1) 한 개의 셀에 해당하는 최소 응답자수는 다음 값을 초과하도록 하고 있음.  
이때, 최소 응답자수는 \_\_\_\_\_ 이다.
- 2) 셀에서 응답자의 해당 값의 분포에 대해 현저한 집중도를 측정함.
- 2-1) C-times원칙: 2순위 응답자가 가장 큰 단위의 유용한 추정치를 얻지 못하도록 보장함.
- 2-2) 순서 통계들의 선형결합인  $(n, k)$  원칙: 가장 큰  $n$ 명의 응답자가 총계 셀 값의  $k\%$  이상 될 경우 그 셀은 민감하다고 봄.
- 2-3) 0의 빈도수를 가지는 특정한 수준이하 또는 이상의 응답자에 대한 정보를 발표하지 않음, 예를 들어 특정 산업의 사업체는 정해진 값 이상의 수익을 결코 초과하지 못함.

**34. 민감한 셀을 보호하기 위하여 보조 셀의 자료를 삭제할 필요가 있었습니까?**

- 1) 예  
 2) 아니오

**35. 셀들을 보호하기 위하여 기존의 소프트웨어를 사용합니까?**

- 1) 예  
 2) 아니오

♣ “5장. 자료 분석 및 통계결과물의 품질” 절차에 대한 질문 중에서 고려해야 할 사항이 있으면 해당 내용을 기입 하십시오.

## 6장. 문서화 및 자료의 제공

### ■ 메타자료의 문서화

자료원천, 개념, 정의와 방법 등을 포함한 산출과정에 대해 완전하고 명확하게 기술해야 한다. 이러한 과정은 이용자들이 자료를 보다 잘 이해하고 활용할 수 있게 할 뿐만 아니라 통계작성기관내부에서 품질의 보장과 개선에 이용하고, 다른 사람들이 통계산출 절차의 인계에 중요하게 이용할 수 있기 때문이다.

#### 1. 행정자료를 사용하는 경우 행정 등록과 관련된 적절한 문서가 있습니까?

- 1) 예
- 2) 아니오

#### 2. 통계산출과정을 문서화할 때 다음 중 어떤 사항들이 포함되어 있습니까?

- 1) 이용자 요구진단
- 2) 조사표 사본
  - 3) 변수에 대한 사항
    - 3)-1 변수명
    - 3)-2 변수의 형태
      - 3)-2-1 조사표나 행정양식상 항목에 대한 응답을 코딩한 변수
      - 3)-2-2 다른 변수(파생변수)들의 계산으로부터 구한 변수
      - 3)-2-3 외부원천자료로부터 가져온 변수
      - 3)-2-4 변수의 통계적 특성(즉, 범주형, 순서형, 계량형)
      - 3)-2-5 해당되는 레코드 길이
      - 3)-2-6 관찰변수의 상세한 수준
      - 3)-2-7 측정단위(의미를 가진 경우)
      - 3)-2-8 표준분류가 사용되지 않을 경우 코드나 표제값의 목록
      - 3)-2-9 외부자료에서 가져올 경우 그에 대한 출처의 서술
      - 3)-2-10 파생 변수의 경우 계산방법의 서술
    - 3)-3 분류 서술
      - 3)-3-1 표준 또는 일반 분류를 사용할 경우 참조된 분류
      - 3)-3-2 특정한 분류가 적용된 경우 분류의 출처
      - 3)-3-3 참조된 분류의 일부항목을 처리하는데 특정 사항을 설명하는 정보
  - 3)-4 개념(예, 실업)과 통계적 단위(기업 또는 지역단위)의 정의

- 3)-4-1 정의된 개념의 제목
- 3)-4-2 정의의 출처
- 3)-4-3 정의와 그 개념을 측정하는 변수간의 연계사항
- 4) 통계 산출 과정에서 사용된 조사방법 및 기법
- 5) 조사의 처리과정에서 조사 설계의 변경
  - 6) 원시자료 파일로부터 공표자료 파일까지 산출된 모든 파일을 포함한 결과들은 다음사항을 문서화 할 필요가 있음.
    - 6)-1 파일제목
    - 6)-2 자료저장형식
    - 6)-3 자료공표상태
    - 6)-4 파일의 서술(파일설계, 레코드 수)
- 7) 품질보고서

**3. 메타자료와 품질지표들을 어떻게 공표하고 있습니까?**

- 1) 종이 발간물상의 메타자료
- 2) 인터넷상의 메타자료
- 3) 종이 발간물상의 품질지표
- 4) 인터넷상에의 품질지표

**4. 무응답 추적절차 또는 조사원을 위한 교육 매뉴얼 인용과 같은 품질보증 척도에 대한 문서를 보유하고 있습니까?**

- 1) 품질보증 척도에 대한 문서는 없음
- 2) 품질보증 척도에 대한 모든 관련된 문서가 있음
- 3) 품질보증 척도에 대한 다음과 같은 문서가 있음 ↴

하지만, 문서가 다음과 같은 이유로 적절하지 않음(이유를 서술) ↴

**5. 경제적인 기존조사의 검토를 포함하여 조사에 대한 예산의 확보와 사용, 조사표 인쇄, 직원채용, 지방행정기관 또는 언론매체 접촉 등과 같은 통계작성절차에 필요한 행정 및 처리절차를 체계적으로 문서가 있습니까?**

- 1) 예
- 2) 아니오

**6. 아래 품질기준 가운데 해당조사의 문서에 포함시키고 있는 경우는 어느 것입니까?**

**1) 관련성:**

- 1)-1 이용자 만족 지수
- 1)-2 공표된 발간물 수 또는 데이터베이스 접속회수

**2) 정확성:**

- 2)-1 변동계수(CV)
- 2)-2 단위 응답률
- 2)-3 항목 응답률
- 2)-4 에디팅 비율 및 비
- 2)-5 대체 비율 및 비
- 2)-6 과대포함 및 오분류율
- 2)-7 평균적인 개편규모

**3) 시의성 및 정시성:**

- 3)-1 효과적인 발간 일정의 정시성
- 3)-2 조사 대상기간 완료시점과 최초 결과 시점간의 평균시차
- 3)-3 대상기간 완료시점과 최종결과 시점간의 평균시차

**4) 접근가능성 및 명확성:**

- 4)-1 통계를 공표하기위해 사용된 방법 가지 수 및 형태

**5) 비교가능성:**

- 5)-1 품질관리 측면에서 개념과 측정상 불일치의 수 및 비율
- 5)-2 비교가능한 시계열의 수 및 길이

**6) 일관성:**

- 6)-1 주요한 2차적인 자료 활용에 대한 요구사항을 만족시키는 산출물의 수와 비율

**7. 자료이용에서 시계열의 단절과 같이 예상되는 제한점에 대해 체계적인 문서를 보유하고 있습니까?**

- 1) 예
- 2) 아니오

8. 발간물에 외부 이용자들의 만족도에 관한 정보가 있습니까?

- 1) 예  
 2) 아니오

9. 내부 직원들의 만족도에 관한 신뢰할 만한 정보가 있습니까?

- 1) 예  
 2) 아니오

10. 보다 나은 발간물을 만들기 위해 다음 중 어느 항목이 포함되어야 하는가?

- 1) 없음  
 2) 자료수정  
 3) 자료처리 작업  
 4) 방법론  
 5) 품질관리  
 6) 직원배치  
 7) 사용하는 소프트웨어  
 8) 기타(그 외 항목을 기술) ↓

#### ■ 자료제공전략

이용자들이 사용하기 좋은 형태로 접근가능하고 사용가능하면 조사결과는 높은 가치가 있다. 또한 자료제공자는 이용자들이 결과를 이해하는데 도움을 주어야 한다.

11. 가장 최근 생산, 저장된 자료에 대한 정보를 검색할 수 있도록 이용자 측면에서 목록 시스템과 정보서비스를 제공하고 있습니까?

- 1) 예  
 2) 아니오

12. 통계작성 기관에서 이용자에게 적합한 형식으로 자료를 이용할 수 있도록 하는 자료 제공 시스템이 있습니까?

- 1) 예  
 2) 아니오

13. 통계표와 그래프 작성을 위해 사용하고 있는 기준이 있습니까?

- 1) 예
- 2) 아니오

14. 개편이나 잠정 또는 조기 추계치 등을 분명히 표시하고 설명하는지와 제한 사항이나 잠정 자료에 대한 개편의 기대수준 등의 제공된 정보가 있습니까?

- 1) 예
- 2) 아니오

■ 자료관리

향후 자료의 재이용, 시계열분석 등을 위하여 많은 자료들을 저장할 필요가 있으며, 이러한 자료들은 체계적이고 정교한 방법으로 관리하여야 한다.

15. 자료를 관리하고 보관하기 위한 체계적인 방법을 보유하고 있습니까?

- 1) 예
- 2) 아니오

16. 자료 보관 시 예상치 못한 상황에 대비하여 자료를 보호하는 방법이 있습니까?

- 1) 예
- 2) 아니오

♣ “6장. 문서화 및 자료제공” 절차에 대한 질문 중에서 고려해야 할 사항이 있으면 해당 내용을 기입 하십시오.

## 7장. 개선 사이클

품질관리는 통계작성기관의 경영과정과 통합할 필요가 있다. 왜냐하면 그 기관은 해당 기관에서 운용하는 모든 통계프로그램에 대해 이용자의 자료 품질, 이용자 수요, 예산제약과 응답부담간의 절충으로 균형을 유지할 책임이 있기 때문이다. 운영절차를 평가하고 개선하려는 노력은 생산된 통계품질에 지속적으로 영향을 주는데 이를 **개선 사이클** 이라고 한다. 다음 항목은 개별 조사의 품질의 관점에서 개선 사이클과 특히 관련이 있다.

### ■ 적용가능성/유연성

조사를 갱신하는 것은 정보를 수집하고 새로운 수요에 유연하게 대응하는 능력을 유지하고 개선하는 것을 의미한다.

#### 1. 다른 정보 수집에 대해 통계소프트웨어 패키지와 절차의 적용가능성과 새로운 수요와 정보요구에 대응하는데 소요되는 시간은 어느 정도라고 평가하십니까?

- 1) 만족하며 만일 필요한 경우, 거의 모든 개선사항을 이행할 수 있음  
 2) 만족하지만, 다음과 같이 통계적 절차에 대해 개선의 여지가 있음  
 ➔ 개선해야할 절차를 기술 ↓

- 3) 만족스럽지 않고, 자료입력시스템의 적절성을 개선할 필요가 있음

#### 2. 영구적인 자발적 응답자 표본으로 구성된 패널조사와 같이 특수한 조사를 유연하게 실시하기 위한 자료수집의 기반구조를 보유하고 있습니까?

- 1) 예  
 2) 아니오

**3. 변화하는 수요에 신속한 대응에 대한 부담과 비용은 어느 정도로 평가하고 있습니까?**

- 1) 수용가능하며, 거의 줄일 수 없는 상태임
- 2) 수용할 수 있지만, 신속한 대응에 드는 비용은 줄일 수 있음
- 3) 너무 높음
- 4) 기타(그 외 사항을 기술) ↴

**■ 관련분야의 전문성**

통계작성기관은 우수한 품질의 통계정보를 생산하기 위해 이용 가능한 능력의 유지와 개선을 목표로 해야 한다.

**4. 조사담당직원은 주제 분야, 원천자료, 개념 및 분류, 현행 및 신규 이슈, 조사와 관련하여 필요한 운용방법 등의 우수한 지식과 이해력을 보유하고 있습니까?**

- 1) 조사담당 직원은 조사에 관련된 모든 주제에 대해 충분한 전문성을 보유함
- 2) 조사담당 직원은 조사에 관련된 대부분 사항에 대한 충분한 지식을 보유함
- 3) 직원 중에 모든 사항에 대한 충분한 전문성이 없지만 필요한 경우 외주 가능함
- 4) 전문성이 충분하지 못함

**5. 직원들의 교육과정은 어느 정도라고 평가하십니까?**

- 1) 가능한 한 직원들에 대해 아주 훌륭한 교육과정이 있음
- 2) 직원의 교육은 충분하지만 개선이 필요함
- 3) 직원교육 상황이 시급히 개선되어야 함

**6. 주요 담당자들의 변동에 대해 업무공백 없이 조사업무를 담당할 만한 전문성을 유지할 수 있습니까?**

- 1) 담당자의 변동 시 업무 공백 없이 전문성을 유지할 수 있음
- 2) 비록 주요 담당자의 변동에 대해 전문성의 유지가 어려울 때도 있지만 대체적으로 유지할 수 있음
- 3) 주요 담당자의 변동에 대해 전문성이 부족함

**7. 다른 기관이나 전문가 협회 등과의 교류는 어느 정도라고 평가하십니까?**

- 1) 아주 좋음, 다양한 국내외 기관과 협회 간 교류가 있음
- 2) 만족스러우며, 비록 교류가 증가하고 있지만 최근 개발동향을 따라갈 만큼 충분한 교류가 있음
- 3) 만족스럽지 못하며, 다른 기관이나 전문가 협회와의 교류가 개선되어야 함

**■ 품질관리**

품질관리의 전반적 목표는 합당한 수요를 만족시키는 비용 효과적인 산출물을 작성하여 내는데 있다.

**8. 투입에 대한 품질과 비용에 대비하여 산출물의 품질을 평가한다는 측면에서 모든 통계 생산과정의 효율성을 고려하여 조사에 대한 정기적인 점검이 이루어지고 있습니까?**

- 1) 점검이 이루어지지 않음
- 2) 모든 생산과정을 검토하고 있음
- 3) 다음과 같은 과정이 검토됨(과정을 기술) ↓

**9. 효율성, 시의성 또는 품질을 확보하기 위하여 이용가능하거나 사용하고 있는 최신의 IT 기술이 있습니까?**

- 1) 대부분의 절차에 대해 IT기술로 전환해야 함
- 2) 거의 모든 과정에 최신의 IT기술을 사용하고 있음
- 3) 최신의 IT 기술을 사용하고 있지만, 다음과 같은 절차에 대한 기술은 개선해야 함(상세히 서술) ↓

**10. 조사의 효율성을 가능한 제고시키기 위하여 재정자원, 인적자원, 사용할 수 있는 기존자료 등과 같은 기존의 모든 원천자료를 광범위하게 검토할 수 있습니까?**

- 1) 예
- 2) 아니오

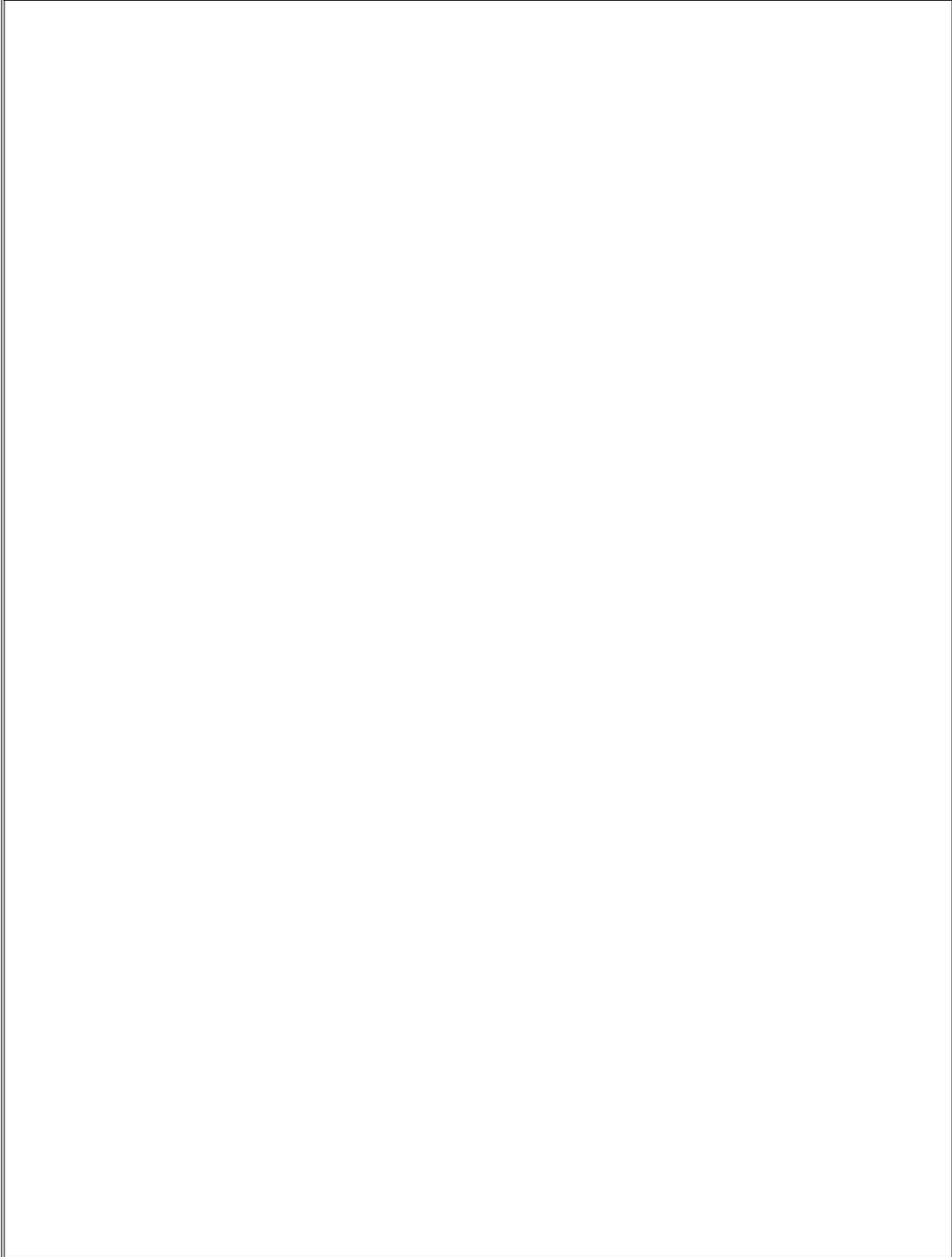
11. 통계생산과정의 효율성을 독립적인 전문가에 의해 검토 받고 있습니까?

1) 예

2) 아니오

12. 만일 귀하의 예산이 현재와 비교하여 20% 정도 많거나, 20% 정도 적다면 어떻게 하겠는가?

♣ “7장 개선 사이클” 절차에 대한 질문 중에서 고려해야 할 사항이 있으면 해당 내용을 기입 하십시오.



## 주요용어

### □ 거부율(Refusal rate)

가구나 사업체 모집단에서 성공적으로 접촉한 사람들의 일부가, 필요한 정보 제공을 거절하는 경우. 이 비율은 거절하는 수를 원래 얻고자 하는 표본의 전체수로 나누어 산출 가능함.

### □ 과대 에디팅(Over-editing)

확실한 품질개선을 할 수 없는 광범위한 에디팅을 말함.

### □ 과대포함(Overcoverage)

과대포함은 목표 모집단에 속하지 않은 단위가 있거나 동일한 단위가 한번이상 추출틀에 속한 경우에 발생하는 경우로서 사업체 조사에서 명부 작성 시 존재한 사업체가 조사 당시에는 폐업하여 존재하지 않는 경우임.

### □ 과소포함(Undercoverage)

과소포함은 목표 모집단에 속하는 틀에서 일부 단위가 누락된 경우에 발생하며, 사업체 조사의 경우 조사당시 신규사업체는 사업체 명부 작성 시 누락됨으로서 실제 모집단에는 포함되지만 추출틀에는 존재하지 않는 경우임.

### □ 관련성(Relevance)

관련성은 현재 및 미래의 이용자 수요에 부응하는 정도를 말하며, 필요로 하는 모든 통계가 생산되고 있는지와 이용자 수요를 반영한 개념(정의, 분류 등)을 사용하는지를 의미함.

### □ 계통임의 표본추출(Systematic random sampling)

모든 표본단위가 추출간격 ( $r, r+k, r+2k, \dots$ ) 에 따라 표본으로 추출되도록 하는 표본추출방법으로 가구단위 조사에서 표본조사구내에서 가구명부에 기재된 순서에 따라 일정간

격으로 표본가구를 선정하는 경우에 적용할 수 있음.

□ **기대값 (Expected Value, Expectation value, expectation)**

동일한 실험조건하에서 반복적으로 계산된 가설적인 평균을 말함.

□ **단위 무응답(Unit non response)**

단위 무응답은 표본단위로부터 자료를 하나도 얻지 못할 경우 발생하며, 가중치 조정이나 사후층화 등의 방법으로 단위무응답을 처리함.

□ **단위 응답률(Unit response rate)**

표본의 유효한 단위수 중에서 면접을 완료 단위수의 백분율로 표시된 비율로서 어떤 경우에는 단위의 중요성을 반영하는 값이 가중요소로도 이용됨(사업체의 종업원 규모 등)

□ **연구영역(Study domains )**

통계는 모집단의 서로 다른 부차그룹, 소위 연구영역으로 표현되며, 이들 연구영역은 보통 분류에 따라 정의됨.

□ **대체(Imputation)**

대체는 에디팅 동안에 파악된 결측, 무효값, 불일치 응답의 문제를 해결하기 위해 사용됨. 대체는 응답값의 일부 또는 결측치를 그럴듯한 값이나 내부적으로 일관성 있는 레코드를 생성하여 그 값으로 변경시키는 작업을 말함.

□ **레코드 연계(Record linkage)**

레코드 연계는 개인, 가구 또는 기업의 두 가지 이상 다른 자료 원으로부터 하나의 개별 원시자료를 만들기 위하여 통계 또는 연구목적으로 연결하는 것. 레코드 연계는 귀중한 통계정보의 잠재적으로 중요한 원천이 됨. 예를 들면 레코드 연계를 통해 특정한 종류의 암 검사 방법의 효과나 심장진료의 장기적인 효과를 확인하는데 도움이 됨.

□ **레코드 연계 성공률(Success rate of the record linkage)**

레코드 연계 성공률은 레코드들이 연결될 때 얼마나 매치되는지를 백분율로 나타냄. 레코드 연계(record linkage)참조.

#### □ 마이크로 에디팅(Micro-editing)

개별적인 측정 자료를 검사함으로써 오류를 찾아내는 작업으로 레코드나 질문수준에서 수행하는 에디팅 작업임.

#### □ 모집단(Population)

모집단은 하나의 정의된 사람, 사물, 또는 사건으로 분류된 전체모임이나 전 인구 또는 전체영역(universe)을 의미함. 모집단에는 목표 모집단과 조사모집단의 두 가지 형태가 있음. 목표 모집단은 정보를 찾기 위한 조사대상 전체를 말하며, 조사 모집단은 조사에서 정보를 얻을 수 있는 모집단을 말함. 목표 모집단은 조사의 범주(scope of the survey)로도 알려져 있으며 조사 모집단은 조사의 포괄성으로 알려져 있음. 행정기록에 대한 모집단의 개념은 목표 모집단으로 관련 법령 및 규칙이 되며, 실제로는 고객 모집단임.

#### □ 목표 모집단(Target population)

추정 대상이 되는 모집단으로 추정 값이 요구되는 단위들의 집합으로 가구단위 조사 또는 인구단위 조사의 경우 목표모집단은 '인구센서스 당시 전 국민' 등으로 정의할 수 있음.

#### □ 명부(Register)

(행정)명부는 행정기록(administrative records)의 부분 그룹. 어떤 행정레코드가 단위 급 자료라면 이를 명부라고 부를 수 있음. 행정 명부는 행정자료에서 유래하지만, 통계적 목적에 맞추기 위하여 통계적인 처리를 거친 후에는 통계명부가 됨.

#### □ 무응답(Nonresponse)

무응답이란 대부분 조사에서 관측되지 않는 형태를 말함. 조사를 위해 선정된 한 개 또는 그 이상의 요소에 대해 한 개 또는 그 이상의 연구대상 변수를 관측하지 못한 경우를 말함. 이러한 무응답의 원인은 접촉 불가능, 부재중, 답변 불가능, 여유부족, 강력한 거절, 접근 불가능, 회답하지 않은 조사표 등이 있음. 처음 두 가지 경우에는 조사대상으로 선정된 단위와의 연락이 전혀 이루어지지 않은 상태를 말함.

□ **무응답 편향(Nonresponse bias)**

무응답 오차 설명을 참조

□ **무응답 오차(Nonresponse error)**

무응답 오차는 조사에서 응답을 받아내는데 한 개 또는 모든 항목을 얻지 못한 경우를 의미. 무응답이 발생하면 유효한 표본 크기가 줄고, 대체를 사용함에 따라 분산이 증가하게 되며, 무응답자들과 응답자들의 응답 패턴이 서로 다르다면 이에 따른 편향의 원인이 됨.

□ **조사원 오차(Interviewer error)**

조사원 오차는 조사원들이 동일한 조사를 다른 방식으로 수행함으로써 응답자의 답변에 영향을 주는 경우 오차임. 이러한 오차의 예는 질문을 잘못 읽거나(응답자에게 응답오차 유도), 응답자가 답을 선택하는데 영향을 줄만한 억양을 사용하여 질문을 하고, 조사원이 응답자의 답변을 잘못 기입할 경우 등이 해당됨.

□ **모형가정오차(Model Assumption Error)**

모형가정은 통계나 지수를 산출하기 위한 보정, 일반화된 회귀추정량, 전체 또는 상수범주, 벤치마킹, 계절조정 및 선행하는 정확성 요소에 포함되지 않는 다른 모형 등의 방법을 사용하는 데 기인한 오차를 말함.

□ **매크로에디팅(Macro-editing)**

매크로 에디팅은 수집된 자료나 레코드 전체에 대해 점검함으로써 개별적인 오류를 찾아내는 과정을 말함. 점검은 보통 모형, 즉 도표나 수식을 토대로 총량추정에서 개별적인 레코드의 특정필드에 영향을 주는지 결정함.

□ **비확률표본(Non probability sample)**

비확률표본은 확률적 기회보다는 다른 사실을 근거로 단위를 선정한 표본. 예를 들면 편의표본, 경험표본, 또는 연구자의 판단에 의한 표본 등이 비확률 표본에 해당됨.

□ **바꾸어 말하기(Prephrasing)**

응답자에게 그들의 말로 질문 내용을 반복하게 함으로서 질문에 대한 이해력 등을 판단할 수 있는 방법으로 조사표를 점검하는 방법 중의 하나임.

#### □ 범주 밖 단위(Out of scope units)

조사대상 기간 동안 목표 모집단에 속하지 않기 때문에 추출틀에 포함되어서는 안 되는 단위로서 만일 이러한 단위가 조사되게 되면 추정량의 과대포함을 초래하게 됨.

#### □ 변동계수(Coefficient of variation)

자료집합의 흩어진 정도로서 보통은 '상대표준오차'로 사용됨. 평균의 표준편차에 대한 비율로 정의되며, 조사 자료의 정도(precision)를 평가하거나, 추정의 정도, 표본수의 산출 등에 이용됨.

#### □ 분산(Variance)

분산추정을 통해 추정값의 질을 평가할 수 있으며, 신뢰구간을 계산하고 정확한 결론을 끌어내는데 도움이 되는 통계값임. 표본분산은 표본조사와 추정의 품질에 대한 주요 지표 중 하나이며, 표본분산은 이용자들이 생산된 통계에 대한 보다 나은 결론을 끌어내는데 도움을 주며, 이는 조사의 설계 및 추정단계에서 주요한 정보로도 사용할 수 있음.

#### □ 비교성(Comparability)

비교성은 서로 다른 지리적 지역, 비 지역적인 영역, 시간경과에 따른 다른 통계들 간의 차이가 통계의 참값에의 차이에서 기인하는 정도를 말함.

#### □ 비표본오차(Non sampling error)

표본추출의 변동으로 인해 일어나지 않은 표본추정상의 오차. 이러한 오차는 추출틀의 결함, 표본단위에 대한 오분류, 표본단위 선정오류, 개인적인 편차, 잘못된 이해, 편향, 조사원 또는 응답자의 부주의, 거짓응답으로 인한 자료수집과정의 실수, 또는 자료처리 단계에서의 실수 등에 기인한 오차임.

#### □ 상대표준오차(Relative standard error)

상대표준오차(RSE)는 추정값의 신뢰성에 대한 척도임. 추정값의 RSE는 추정값의 표준오차( $SE(r)$ )를 추정값 자체로 나누어서 산출함. 이 값은 추정값의 백분율로 표시되며  $RSE = 100 \times (SE(r)/r)$  으로 계산됨.

□ **생각 말하기 면접(Think-aloud interviews)**

응답자가 한 질문 항목에 대해 어떻게 해석하고, 정보를 검색하고, 답변을 형성하고, 응답할 대안을 선택할지 생각하는 바를 크게 말함으로써 조사표를 평가할 수 있는 방법임.

□ **시간대 표본추출(Flow-sampling)**

시간 단위별로 표본단위를 추출하는 것. 예를 들어 어떤 경계를 지나가는 사람의 행렬에서 매 15분 간격으로 지나가는 사람에게 조사를 수행하는 방법임.

□ **시의성(Timeliness)**

정보의 시의성은 정보가 나타내는 사건이나 현상과 정보사용 가능까지 걸리는 시간을 반영함.

□ **신뢰구간(Confidence Interval)**

미지의 모집단 모수(parameter)  $\theta$ 에 대한  $\alpha\%$ 의 신뢰구간으로 여러 개의 독립된 표본들이 추출될 경우  $\alpha\%$ 의 구간에 모수  $\theta$ 가 포함되도록 하는 하나의 구간으로, 표본을 통해 계산됨.

□ **심층면접(In-depth interviews)**

모집단에 대해 많은 것이 알려지지 않았을 때, 참석자들에게서 예비적인 아이디어를 얻기 위해, 훈련받은 직원에 의해서 실시되는 면접방식을 말함.

□ **오분류(Misclassification)**

오분류란 주어진 분류와 관련하여 통계단위를 잘못된 단위로 분류한 것을 의미. 예를 들면 기업체를 사업체로 분류한 경우임.

□ **오차(Error)**

일반적으로 하나의 실수 또는 일상적인 오류로서 예를 들면 총 오차 또는 피할 수 있는 실수; 하나의 현상에 대한 자료가 다른 것에 오류가 있다고 귀착되는 참조오류, 복사오류, 해석오류 등. 보다 제한적인 의미에서 오차는 통계학에서 사용되는 경우로서 발생된 값과 실제 또는 기댓값과의 차이를 말함. 편차는 기회효과임. 이러한 견지에서 예를 들면 관측 오차, 방정식에서의 오차, 가설검증에서 1종오류, 2종오류, 추정치를 둘러싸는 오차대

(band), 오차의 정규곡선 등을 일컬음.

**□ 완비성(Completeness)**

완비성이란 필요로 하는 모든 통계가 사용가능한지를 나타내는 척도임. 보통 한 종의 통계(시스템)에서 얻는 통계의 양에 대한 사용가능한 자료의 양의 비율로 나타냄.

**□ 에디팅, 에디트(Editing, Edits)**

자료 에디팅은 결측, 무효 또는 일치하지 않는 자료입력을 파악하거나 잠정적으로 오류 자료 레코드를 지적하기 위하여 점검하는 작업임.

**□ 이용자 만족도 조사(User Satisfaction survey)**

통계 이용자들의 만족 정도를 평가하기 위한 통계조사임.

**□ 인라이어(Inlier)**

어떤 통계적 분포 안쪽에 있는 자료값. 인라이어는 양호한 자료값(good data value)과 구분이 잘 되지 않기 때문에 찾아서 수정하기 어려운 단위임. 단순한 예로 인라이어에는 잘못된 단위로 보고된 레코드 내 값 또는 십씨를 화씨로 기록한 것 등이 있음.

**□ 2단계집락표본추출(Two-stage cluster sampling)**

1단계에서 선택된 표본 집락에서 부분적으로 표본을 추출하는 것을 의미함.

**□ 이상치(Outlier)**

n번 측정한 한 표본에서 정해진 숫자가 나머지 숫자들과 너무 멀리 떨어져 있어서 다른 모집단에서 오지 않았는지 표본추출기술에 문제가 있었는지 의문을 일으키는 경우에 이러한 값을 이상치라고 함

**□ 이중(층화)표본추출(Two-phase (stratified) sampling)**

추출틀에서 층화를 위한 정보를 얻을 수 없는 경우 1차적으로 대규모 표본을 추출하여 층화정보를 얻을 목적으로 추출설계를 하는 것을 말함.

#### □ 일관성(Coherence)

통계의 일관성이란 다른 방법, 다른 용도를 위해 신뢰성 있게 조합된 자료의 적합성을 의미하지만 보통 일관성을 증명하기 보다는 모순된 경우를 예로 드는 것이 용이함. 유일한 자료 원으로부터 자료를 수집했을 때 보통 일관성이 있다고 하는데 관련조사로부터 나온 1차적인 결과를 좀 더 복잡한 결과를 산출하기 위해 여러 가지로 조합하더라도 신뢰할 수 있다는 것을 의미함. 다른 원천, 특히 특성이나 주기가 다른 조사들로부터 자료를 얻을 때 통계자료는 완전히 일관되지 못하다고 하는데 이는 자료들이 각기 다른 접근, 분류 및 방법론적 기준에 근거한다고 판단하기 때문. 근접한 결과를 보여준다고 해도, 이용자들이 알아야 하며 나을법한 효과 등 일관적인 메시지를 완전히 전달하지 못함.

#### □ 1단계집락표본추출(One-stage cluster sampling)

추출틀을 여러 개의 집락으로 나누고, 이러한 집락으로부터 표본을 추출하는 방법을 의미함. 이때 집락은 인위적으로 구성하거나 자연스럽게 구성될 수 있음.

#### □ 일치추정량(Consistent estimator)

표본크기가 커짐에 따라 확률적으로 추정량이 모수로 수렴하는 추정량의 성질을 의미함.

#### □ 임의표본추출(Random sampling)

모집단 내 모든 단위가 표본에 포함될 수 있는 0이 아닌 확률을 가지는 경우를 말함.

#### □ 접근성 (Accessibility)

접근성이란 이용자가 자료를 얻게 되는 물리적인 조건들을 의미함. 장소, 주문방법, 받는 시간, 투명한 가격정책, 편리한 마케팅 조건(저작권 등), 마이크로 또는 매크로 자료 이용가능성, 다양한 형식(종이, 파일, CD롬, 인터넷 등)이 접근성을 평가하는 기준이 됨.

#### □ 정확성(Accuracy)

일반적인 통계적 관점에서 정확성은 산출 결과 또는 추정값이 정확한 또는 참된 값에 근접한 정도를 의미함.

□ **절사분계점(Cut-off threshold)**

주로 비용이나 부담 때문에 요구하는 통계에 거의 기여하지 않는 대상을 목표 모집단 (따라서 표본 틀)으로부터 제외시키기 위해 사용함(예 소규모 기업). 분계점 이하의 모집단의 기여도를 무시할 정도이거나 어떤 모형에 의해 추정될 수 있을 경우임.

□ **전자적인 자료교환(Electronic Data Interchange(EDI) 또는 Electronic Document Interchange)**

EDI란 대중적으로 출판되고, 세계적으로 표준화된 형식에 의한 컴퓨터 간의 자료교환 형식을 말함.

□ **정시성(Punctuality)**

정시성은 실제 자료배포시점과 목표로 하는 시점 즉, 반드시 배포되어야 할 시점 간의 시간 경과를 의미함. 예를 들면 공식적인 공표일정, 법령에 제정되어 있거나 전에 이용자와 약속한 시점 등을 의미함.

□ **조사기준시기(Reference period, reference time)**

자료가 수집되는 시간의 기간을 의미함.

□ **조사 설계(표본추출 계획, 표본추출 설계)(Survey design, Sampling plan sample design)**

조사 설계에서는 어떤 고정된 특정 조건 내에서, 모든 실현가능한 자료수집의 정해진 성질이 정의됨. 이것의 정확한 의미처럼 용도는 '표본계획', '조사설계', '표본추출계획' 또는 '표본추출 설계' 등으로 다양함. 이들은 자료처리 등을 포함하여 한 (표본)조사의 전체 계획수립을 구성하는 한 개 이상의 부분을 포함함. '표본추출 계획'은 표본을 선정하는 모든 과정을 한정적으로 의미. '표본설계'라는 용어는 추정방법까지 포함. '조사 설계'는 조사의 여러 측면, 예를 들면 조사원 선정 및 훈련, 표 작성 계획 등을 포함함. '표본설계'는 명백히 정의된 의미로 사용되는데, 주어진 추출틀에서 명확한 방법으로 표본을 추출하는 규칙이나 명부리스트로 정의됨.

□ **재가중(Reweighting)**

재가중은 추정값이 산출될 때 응답값에 원래 가중치를 확대시켜 주는 작업 또는 그 값

을 의미함. 재가중은 주로 단위 무응답에 대한 것. 또한 보조정보를 사용함으로써 정밀도를 향상시키는데 사용됨. 표준적인 방법으로는 사후층화(post-stratification), 보정(calibration), 응답성향 모형(response propensity modelling) 등이 있음.

□ **집중그룹면접(Focus Group Interviews)**

조사원들이 조사의 문제점이나 잘된 점을 파악하고 잠재적인 문제를 어떻게 해결할지 제안하기 위하여 적당히 맞춘 그룹과 면접하는 방법임.

□ **참값(True value)**

원시자료를 수집하거나 수학적 연산을 할 경우, 완벽한 조사표를 사용하여 어떤 형태의 오류를 범하지 않고 얻을 수 있는 실제 모집단 값을 의미함.

□ **처리 오차(Processing error)**

자료가 수집되면 추정량이 산출되기 전에 일련의 과정을 거침: 코딩, 입력, 에디팅, 가중치, 제표 등의 과정에서 발생하는 오차를 자료처리 오차라고 함.

□ **추출틀(Frame)**

추출틀은 표본단위가 선정된 지도, 목록, 성명목록 등 물리적인 분야와 관계있는 대상, 그리고 자료에 대해 상세히 수록함.

□ **추정값(Estimate)**

엄격히 구분하면, 하나의 추정은 일련의 환경에서 하나의 추정량에 의해 산출되는 특정한 값을 의미함.

□ **추정량(Estimator)**

추정량은 모집단의 어떤 상수를 추정하는 규칙 또는 방법. 보통은 표본값의 함수로 표현되며 따라서 이에 따른 추정의 신뢰성을 평가하는데 매우 중요하게 기여하는 변수임.

□ **측정오차(Measurement error)**

측정오차는 자료수집방법, 응답자, 조사표(또는 다른 도구)등으로부터 발생하는 조사응답의 오차임. 이에는 무응답, 혼동, 무지, 부주의 또는 거짓된 결과로서 나타나는 오차가 있

으며, 조사원에 기인한 오차로서 부실하거나 부적절한 교육을 받은 조사원, 응답자의 답변에 대한 선입견, 또는 고의적인 오차; 조사표 질문의 어휘, 질문이 표현된 문장의 순서 또는 문맥, 응답을 얻어내는 데 사용된 방법 등에 기인함

□ **층화표본추출(Stratified sampling)**

모집단은 상호 배타적인 동질의 부차모집단(strata)으로 처음에 나누어지고 각 층으로부터 전체분산을 줄이거나, 각층별 통계를 산출 위하여 확률표본을 추출하는 방법임.

□ **코딩(Coding)**

코딩은 구두로 말한 정보를 수치나 다른 기호로 전환하는 기술적인 절차로서 이렇게 함으로써 계산과 통계표 작성이 용이해짐.

□ **코딩오차(Coding error)**

조사 응답에 대해 잘못된 코드가 부여된 경우에 발생하는 오차를 말함.

□ **크기비례 확률표본추출(Probabilities Proportional to Size(PPS) sampling)**

각 간위에 대해 단위의 크기에 비례하도록 확률을 부여하여 그 크기에 따라 표본을 추출할 때 이를 크기비례확률 표본추출이라고 함.

□ **통계적 특성치(Statistical characteristics)**

특정한 그룹의 모든 통계적 단위에 대하여, 상세한 양적인 변수값(거래액, 평균소득)을 요약하는데 사용되는 통계적 척도에 의해 정의되는 수치값(거래액, 가처분소득).

□ **통계적 측도(Statistical measure)**

특정 그룹(연구영역)에 있는 통계 단위에 대해 개별적인 수량적 변수 값의 요약한 것.

□ **통계 단위(Statistical unit)**

통계조사의 대상으로 통계적 특성을 가진 존재를 의미함.

□ **투명성(Clarity)**

투명성이란 자료제공 시 적절한 메타자료, 즉 도표 및 지도, 삽화 등의 첨부 여부, 자료

품질에 대한 정보 이용 가능 여부, 통계기관에서의 어떤 추가적인 지원 정도 등 자료의 정보환경을 의미함.

□ **포함오차(Coverage error)**

포함오차는 표본추출에 사용된 틀 내에 목표 모집단의 일부단위가 포함되지 않았을 때 발생하는 오차(과소포함)와 한 번 이상 포함되어 실제 대표하는 단위를 파악할 수 없을 때 발생하는 오차(과대포함). 포함오차의 근원은 추출틀에서 발생함.

□ **표본오차(Sampling error)**

모수와 추정값 간의 차이로서 임의로 표본을 추출하고 추출된 표본만을 관측하기 때문에 발생하는 오차: 불완전한 추출 때문에 발생한 오차, 응답이나 추정값 내 편향, 또 측정 및 기록 등 오차와는 다름.

□ **표준오차(Standard error)**

통계의 표본분포의 분산의 양의 제곱근을 말함.

□ **품질관리조사(Quality control survey)**

반복된 조사를 아주 경험이 풍부한 직원들에게 수행하도록 하는 것. 완벽한 자료(zero-default)를 얻기 위한 것이며 조사의 결과와 비교 가능한 조사함

□ **품질지수(Quality Index)**

품질에 대한 단일차원의 종합정보, 모든 가능한 품질지수들의 가중평균을 의미함.

□ **편향(Bias)**

일반적으로 어느 경우든지 왜곡시킬 수 있는 확률오차(random error)와 달리, 대표성을 가진 통계적 결과의 체계적(systematically)인 왜곡을 초래하는 어떤 효과를 의미함.

□ **평균제곱오차(Mean Squared Error)**

추정을 위한 총 평균제곱오차는 편향의 제곱에 분산을 더한 것임. 총 MSE의 편향과 분산에 영향을 주는 오차의 몇 가지 요인이 있음. 특정화(specification), 추출틀(frame), 무응답(nonresponse), 측정(measurement), 자료처리(data processing), 표본(sampling)오차 등

입.

□ 항목 무응답(Item non response)

항목 무응답은 응답자가 전부가 아닌 일부만 응답하거나 받은 정보를 사용할 수 없을 경우에 발생하며, 이러한 무응답 처리 방법은 대체를 통해 가능함.

□ 항목 응답률(Item response rate)

항목응답률은 한 항목에 대하여 응답한 적절한 단위 수에 대한, 그 항목에 대해 적절히 답변한 단위 수의 비를 말함.

□ 확률표본(Probability sample)

모집단으로부터 직접 단위를 추출하는 경우에 확률표본추의 특정 조건은 다음과 같음.  
 ① 표본 집합  $\{S_1, S_2, \dots, S_m\}$ 을 정의할 수 있으며, 이들은 표본 추출과정을 통하여 얻을 수 있음. ② 기지의 추출확률  $p(s)$ 는 각 가능한 표본  $s$ 와 관련됨. ③ 모집단의 모든 단위가 추출확률이 0이 되지 않게 함(nonzero). ④ 각 가능한 표본  $s$ 는  $p(s)$ 라는 확률을 부여하여, 임의로 추출함. 이와 같은 네 가지 조건을 만족하면 확률표본이라고 함.

□ 행정레코드(Administrative record)

공식적인 통계생산보다는 원래 다른 목적으로 수집된 자료를 의미함. 의사결정 또는 기록된 개별적인 문제에 대한 조치를 취하는 데 사용된 정보를 포함함.

□ 행태코딩(Behaviour coding)

면접이 진행될 동안 조사원이나 응답자가 나타낸 행태의 정해진 유형별 빈도를 체계적으로 코딩하는 것을 말함.