

統計調査에 있어서 效率的標本設計와 誤差에 관한 研究*

김종호¹⁾ · 남궁 평²⁾ · 박홍래³⁾ · 이계오⁴⁾ · 이상우⁵⁾ · 함종옥⁶⁾

요 약

본 연구는 현행 작성되는 주요 통계를 분석하여 문제점을 제기하였고 標本調査 실시를 위한 效率的標本設計를 위하여 基本原則을 제안함과 동시에 誤差發生의 原因과 管理를 위한 原則을 제시하였다. 또한 統計報告書形式을 標準化함으로써 統計利用者의 便益과 信賴性을 제고시키려고 하였다.

1. 序 論

政治의 民主發展과 經濟의 持續的인 成長으로 韓國의 地位가 國際 先進隊列에 서게 되었다. 그간의 發展過程이 여하간에 앞으로 永續的으로 대열속에서 동등하게 발전하기 위하여는 具體的이고도 客觀化된 資料에 의하여 올바르게 計劃하고 意思決定을 하는 合理的意識水準을 모든 分野에서 確保해야 할 것이다.

한국에서 生産되는 많은 統計數值들은 대부분의 사람들이 그 正確性이나 信賴性에 대해서 懷疑를 느끼는 실정이며 많은 政策樹立者들이 意思決定에 있어서 統計資料를 利用하려는 傾向을 볼 수 없는 것은 統計의 正確性의 문제와는 별도로 지적되는 사안이기도 하다.

正確한 計劃과 合理的決定이 보장되기 위하여 計劃이나 決定에 있어서 決定權者들이 먼저 統計資料가 基礎로 필요하다는 認識이 先行되어지고 利用되는 統計가 정확해야하며 신뢰가 있어야 함은 再言의 여지가 없는 것이다. 기초자료로 이용되는 통계가 정확하고 신뢰되기 위해서는 統計調査活動 즉 統計를 生産하는 過程이 計劃的이고 合理的으로 잘 진행되어야 한다.

한국의 현실은 統計學의 歷史도 짧고 그리고 統計學에 대한 認識도 올바르게 되어 있지 않은 상태에서 統計調査活動을 안이하게 진행하고 그에 따라 統計가 作成되기 때문에 통계의 正確度가 문제될 뿐만 아니라 그 信賴性의 문제가 상존하고 있다. 이로 인하여 파생되는 個人的損失은 물론 國家的損失과 試行錯誤에 의한 社會的인 非合理的要素의 문제가 항상 발

* 이 논문은 1988년도 문교부 지원 한국 학술 진흥재단의 자유공모 과제 학술연구 조성비에 의하여 연구되었음.

1) 동국대학교 교수 2) 성균관대학교 교수 3) 서울대학교 교수 4) 공군사관학교 부교수
5) 연세대학교 교수 6) 중앙대학교 교수

생하는 것이다. 통계의 정확성이나 신뢰성을 회복하고 파생되는 손실이나 문제점 등을 해결하기 위하여는 統計調査方法에 대한 연구와 실제 조사 활동을 재검하여 질적향상을 기해야 할 것이다.

이상과 같은 현실에 입각해서 본 연구의 목적은 정부나 공공기관 또는 기업체 등에서 통계를 생산하기 위하여 실시하고 있는 統計調査方法의 타당성이나 문제점을 파악하고 效率的 標本設計와 誤差管理를 위한 方法을 研究하고 개선점을 찾아 제시하는데 있다.

2. 現行統計調査의 現況과 問題点

統計法 第3條 規定에 의거하여 승인 작성되는 현행 주요 통계를 작성기관별 및 종류별로 살펴보면 86기관에서 342種의 통계를 생산하고 있다(표 1 참조).

41개 정부기관에서 234종의 통계를, 그리고 45개의 지정기관에서 108종의 통계를 작성하고 있다. 이중에 조사통계가 161종이고 보고에 의한 통계가 181종으로 되어있다.

표 1. 統計作成 現況

(1989년 5월 현재)

| 작성기관 | | 기관수 | 종 류 | | 작성방법 | |
|------|----|-----|-----|-----|------|-----|
| | | | 지정 | 일반 | 조사 | 보고 |
| 정부기관 | 중앙 | 26 | 32 | 202 | 85 | 150 |
| | 지방 | 15 | | | | |
| 지정기관 | | 45 | 4 | 104 | 76 | 31 |
| 계 | | 86 | 36 | 306 | 161 | 181 |

(자료: 경제기획원 조사통계국 통계목록)

2.1 統計分析을 위한 標本設定

342종의 통계를 전수조사하여 분석하고 문제점을 찾으려 하였으나 시간이나 경비의 제약으로 표본을 추출하게 되었으며 가능한 한 많은 표본을 추출하였다.

주요통계 342종으로부터 단순 임의 추출에 의하여 130종의 標本을 추출(추출률 38%)하여 표본자료를 확보하고 이것을 조사방법에 따라 전수조사와 표본조사로 나누어 보았다. 그 결과 전수조사가 55종 이고 표본조사가 57종이며 나머지 18종은 구분이 어려운 통계로 되어 있었다(표 2 참조).

표 2. 調査區分分布

| 조사방법 | 가 지 수 |
|------|-------|
| 전수조사 | 55 |
| 표본조사 | 57 |
| 구분불가 | 18 |
| 계 | 130 |

18종의 통계는 조사계획에 대한 것을 찾아볼 수가 없을 정도의 것이다.

2. 2 전수조사의 문제점

전수조사 55종은 인구통계분야 20종 사회조사분야 17종 공업통계분야 10종 및 농업 통계 분야 8종으로 구성되어 있으며 구체적 통계명칭은 아래와 같다.

| 통 계 명 칭 | 분 야 | 조 사 기 관 | 조 사 기 간 |
|------------------|-----|-----------------|--------------------|
| 상주인구조사 결과보고 | 인 구 | 인 천 직 할 시 | '87. 11. 1~11. 8 |
| 상주인구조사 결과보고 | 인 구 | 제 주 도 | '87. 11. 1~11. 8 |
| 상주인구조사 결과보고 | 인 구 | 부 산 직 할 시 | '87. 11. 1~11. 8 |
| 상주인구조사 결과보고 | 인 구 | 경 기 도 | '87. 11. 1~11. 8 |
| 상주인구조사 결과보고 | 인 구 | 경 상 남 도 | '87. 11. 1~11. 8 |
| 상주인구조사 결과보고 | 인 구 | 전 라 남 도 | '87. 11. 1~11. 8 |
| 상주인구조사 결과보고 | 인 구 | 강 원 도 | '87. 11. 1~11. 8 |
| 상주인구조사 결과보고 | 인 구 | 전 라 북 도 | '87. 11. 1~11. 8 |
| 상주인구조사 결과보고 | 인 구 | 서 울 특 별 시 | '87. 11. 1~11. 8 |
| 상주인구조사 결과보고 | 인 구 | 대 구 직 할 시 | '87. 11. 1~11. 8 |
| 상주인구조사 결과보고 | 인 구 | 광 주 직 할 시 | '87. 11. 1~11. 8 |
| 상주인구조사 결과보고 | 인 구 | 충 청 남 도 | '87. 11. 1~11. 8 |
| 상주인구조사 결과보고 | 인 구 | 경 상 북 도 | '87. 11. 1~11. 8 |
| 인구 및 주택센서스 보고 | 인 구 | 경 제 기 획 원 | '85. 11. 13 |
| 인구이동 통계연보(상) | 인 구 | 경 제 기 획 원 | 1987 |
| 인구이동 통계연보(하) | 인 구 | 경 제 기 획 원 | 1987 |
| 한국도시연감 | 인 구 | 내 무 부 | 1986 |
| 사망원인 통계연보 | 인 구 | 경 제 기 획 원 | 1986 |
| 인구동태 통계연보 | 인 구 | 경 제 기 획 원 | 1986 |
| 상주인구 조사결과 | 인 구 | 충 청 북 도 | '87. 11. 1~11. 8 |
| 한국기업의 부가가치 분석 | 사 회 | 한 국 생 산 성 본 부 | '88. 4. 2~ 5. 10 |
| 소매업 경영동태보고 | 사 회 | 상 공 회 의 소 | '87. 2. 1~ 2. 10 |
| 수출기업의 경영실태보고 | 사 회 | 한 국 무 역 협 회 | '87. 3.~5. |
| 기업 경영분석 | 사 회 | 한 국 은 행 | '88. 3. 2~ 4. 30 |
| 노동생산성 지수 | 사 회 | 한 국 생 산 성 본 부 | 1985 |
| 공무원 통계 | 사 회 | 총 무 처 | '83. 4. 20 |
| 주요 업종의 월간경기 동태 | 사 회 | 한 국 경 제 인 연 합 회 | '86. 10. |
| 사업체 노동실태 조사보고 | 사 회 | 노 동 부 | '86. 5. 1~ 5. 31 |
| 총 사업체 통계조사보고 | 사 회 | 경 제 기 획 원 | '86. 7. 1~ 7. 20 |
| 아파트 주거환경 조사 | 사 회 | 건 실 부 | '86. 10. 29~11. 10 |
| 산업 센서스 보고 | 사 회 | 경 제 기 획 원 | '83. 1. 1~12. 31 |
| 부가가치 노동 생산성 지수 | 사 회 | 한 국 생 산 성 본 부 | '88. 1분기 |
| 서울시 생산 및 시민 분배소득 | 사 회 | 서 울 특 별 시 | 1984 |
| 국민소득 계정 | 사 회 | 한 국 은 행 | 1984 |
| '85년 고속도로 교통량 통계 | 사 회 | 한 국 도 로 공 사 | '86. 11. |
| 경기전망 조사 | 사 회 | 한 국 산 업 은 행 | '88. 4. 1~4. 30 |
| 건설업 통계연보 | 사 회 | 건 실 업 회 | 1987 |
| 차량 교통량 조사보고 | 공 업 | 부 산 직 할 시 | '86. 10. 11 |
| 중소기업 경기전망 조사 | 공 업 | 중 소 기 업 은 행 | '88. 6. 1~6. 10 |

| 통 계 명 칭 | 분 야 | 조 사 기 관 | 조 사 기 간 |
|---------------|-----|-----------------|-----------------------|
| 폐수 배출시설 조사결과 | 공 업 | 환 경 청 | '85. 7. ~10. |
| 건설업 수주연보 | 공 업 | 경 제 기 획 원 | 1984 |
| 전문 건설업 통계연보 | 공 업 | 대 한 건 설 업 회 | 1985 |
| 가입전화 수요조사 보고 | 공 업 | 한 국 전 기 통 신 공 사 | '86. 10. 1~'87. 4. 20 |
| 지가동향 | 농 업 | 건 설 부 | '87. 1 |
| 한국 수문조사 연보 | 농 업 | 건 설 부 | 1988 |
| 농업 기본통계 조사자료 | 농 업 | 농 립 부 | '71. 10. 1 |
| 지적통계 | 농 업 | 내 무 부 | 1988 |
| 에너지 센서스 결과보고 | 공 업 | 동 력 자 원 부 | '86. 1. 1~12. 31 |
| 건설업 통계조사보고 | 공 업 | 경 제 기 획 원 | '87. 2. 14~3. 14 |
| 광공업 통계조사보고 | 공 업 | 경 제 기 획 원 | '87. 4. 1~4. 30 |
| 산업생산연보 | 공 업 | 경 제 기 획 원 | 1987 |
| 자가발전 현황보고 분석 | 공 업 | 한 국 전 력 공 사 | '87. 12. |
| 전국 산림 실태조사 보고 | 농 업 | 산 립 청 | 1980 |
| 농업조사 | 농 업 | 농 수 산 부 | '80. 12. 11~12. 20 |
| 총 어업 조사보고 | 농 업 | 농 수 산 부 | '80. 12. 1~12. 15 |

이상의 통계를 분석해본 결과에 따르면 통계보고서가 일정한 형식을 갖추고 있지 않으므로 참조하기가 용이하지 않았다. 조사목적이나 배경설명이 주어지지 않아서 문제에 대한 정확한 진술이 결여된 통계도 12종이나 되며 조사방법이나 내용을 제대로 명시하지 않아서 조사절차가 명확하지 않은 통계도 13종이나 된다. 조사결과를 명료하게 요약하여 통계에 대한 신뢰를 제고시키는 경우는 19종 뿐이고 참고나 활용하고자 할 경우 제한 요소나 유의사항에 대하여 서술한 경우는 7종 뿐이다. 조사연구 보고서의 기본형식이라고 볼 수 있는 문제의 진술, 조사개요에 있어서 조사방법 및 내용, 조사결과(요약)와 결론(유의사항 포함)의 4가지 요소를 제대로 갖춘 통계는 찾아볼 수 없다.

표 3. 기본개념의 규정여부(규정되어 있는 통계의 수)

| 조사/주요사항 | 조 사 자 | 조사목적 | 조사대상 | 조사기간 | 조사기준시점 | 조사방법 | 모 집 단 |
|---------|-------|------|------|------|--------|------|-------|
| 전 수 조 사 | 54 | 43 | 48 | 35 | 36 | 42 | 39 |

표 3은 예를들면 전수조사 55종 중에 조사자가 나타나 있지 않은 경우가 1종 있으며 모집단 규정이 되어 있지 않은 것도 16종이나 되는 것을 보이고 있다.

통계를 생산하는데 막대한 예산과 노력을 투입하여 실용성이 별로 없는 보고서를 작성한다면 경제적인 손실뿐만 아니라 국가 발전에도 기여하지 못한다는 사실을 인식하고 좀더 세심한 주의와 정성을 기울여야 할 것이다. 정부 부처나 공공기관의 통계작성 실무자들에게 사명감을 심어주고 조사보고서 작성 형식을 교육하여 신뢰받는 통계를 생산하도록 우리 모든 통계인은 더욱 노력하지 않으면 아니될 것으로 보여진다. 조사보고서의 표준화된 형식은 뒤에서 언급할 것이다.

2. 3 표본조사의 문제점

표본조사에 의한 통계 57종은 사회조사 분야가 26종, 공업통계분야 12종, 농업통계분야 9종, 시장조사분야 5종, 보건통계분야 4종과, 인구통계분야 1종으로 구성되어 있으며 실제 통계명은 아래와 같다.

| 통 계 명 칭 | 분 야 | 조 사 기 관 | 조 사 기 간 |
|--------------------|------|----------------|--|
| 임금결정 활용지표 | 사회조사 | 한국생산성본부 | '86. 10. 2~10. 31 |
| 신규인력채용 동태 및 전망조사 | 사회조사 | 한국경영자총협회 | '87. 3. 1~ 4. 20 |
| 임금 실태조사 결과보고 | 사회조사 | 최저임금심의위원회사무국 | '88. 3. 2~ 3. 31 |
| 도시주택임대료및다세대실태조사 | 사회조사 | 대한주택공사 | '85. 10. 23~11. 17 |
| 외래 관광객 여론조사 | 사회조사 | 한국관광공사 | '87. 5. 12~10. 17 |
| 전국민 여행 동태조사 | 사회조사 | 한국관광공사 | '84. 9. 28~10. 12 |
| 24회 기업 금융 실태조사 | 사회조사 | 전국경제인연합회 | '83. 6. 4~ 7. 31 |
| 국민소비형태 및 의식구조조사 | 사회조사 | 한국소비자보호원 | '87. 10. 15 |
| 아파트 입주자 실태조사 | 사회조사 | 대한주택공사 | '85. 6. 10~ 6. 30 |
| 한국주택은행용자주택실태조사 | 사회조사 | 한국주택은행 | '87. 6. 30 |
| 가계금융 이용 실태조사보고서 | 사회조사 | 국민은행 | '86. 7. 2~ 7. 19 |
| 재무분석 | 사회조사 | 한국산업은행 | '88. 3. 1~ 5. 30 |
| 직종별 임금실태조사보고서 | 사회조사 | 노동부 | '87. 7. 1~ 7. 31 |
| 전자 전기공업 경기동향 | 공업통계 | 한국전자공업진흥회 | '88. 7. 5~ 7. 15 |
| 주택건설업체가 본 주택건설경기동향 | 공업통계 | 한국주택은행조사부 | '88. 4. 1~ 4. 9 |
| 정밀측정표준실태조사보고서 | 공업통계 | 과학기술처 | '85. 5. 30~ 8. 31 |
| 중소기업 실태조사보고서 | 공업통계 | 상공부중소기업협동조합중앙회 | '86. 1. 1~12. 31 |
| 중소기업 금융동향조사보고서 | 공업통계 | 국민은행 | '88. 1. 25~ 2. 13 |
| 어업 경영 조사보고 | 농업통계 | 수산업협동조합중앙회 | '86. 1. 1~12. 31 |
| 산림 기본통계 | 농업통계 | 산림청 | 1980 |
| 농가경제, 농산물 생산비 | 농업통계 | 농림수산부 | 1987 |
| 양곡소비량 조사 결과보고서 | 농업통계 | 농림수산부 | |
| 농산물생산비조사결과보고 | 농업통계 | 농림수산부 | '81. 12. 1~'82. 11. 30 '81. 8. 1~'82. 7. 31 |
| 작물통계 | 농업통계 | 농림수산부 | 1986 |
| 기업 경기전망 조사보고 | 시장조사 | 상공회의소 | '88. 2. 27~ 3. 5 |
| 생명보험 성향 전국 조사 | 시장조사 | 생명보험협회 | '85. 10. 22~10. 31 |
| 질병 상해 조사 | 보건 | 보건사회부 | '86. 6. 1~ 6. 30 |
| 국민 영향 조사보고서 | 보건 | 보건사회부 | '85. 9. 16 |
| 기업체 노동비용조사보고서 | 사회조사 | 노동부 | '86. 6. 10~ 6. 30 |
| 임금근로시간제도종합조사 | 사회조사 | 노동부 | '85. 7. 10~ 7. 31 |
| 옥외근로자직종별임금조사 | 사회조사 | 노동부 | '87. 9. 30 |
| 노동력유동실태조사보고서 | 사회조사 | 노동부 | 1987 |
| 노동 생산성 통계조사 | 사회조사 | 노동부 | 1986 매월 월말 |
| 표준자모델 임금조사 | 사회조사 | 상공회의소 | '86. 10. 10~11. 5 |
| 매월 노동통계 조사보고서 | 사회조사 | 노동부 | '88. 4. 30 |
| 노동통계연감 | 사회조사 | 노동부 | '88. 4. 30 |

| 통 계 명 칭 | 분 야 | 조 사 기 관 | 조 사 기 간 |
|----------------------|------|---------------|--------------------|
| 물가연보 | 사회조사 | 경제기획원 조사 통계국 | 1985 |
| 도시 가계연보 | 사회조사 | 경제기획원 조사 통계국 | 1987 |
| 고용구조특별조사 결과보고(시, 도편) | 사회조사 | 경제기획원 조사 통계국 | '86. 11. 17~11. 26 |
| 고용구조특별조사결과보고(전국편) | 사회조사 | 경제기획원 조사 통계국 | '86. 11. 17~11. 26 |
| 국부통계 조사보고 | 사회조사 | 경 제 기 획 원 | '78. 6. 1~ 9. 20 |
| 소규모기업실태조사보고서 | 공업통계 | 국 민 은 행 | '87. 7. 20~ 8. 7 |
| 설비투자 계획조사 | 공업통계 | 한 국 산 업 은 행 | '88. 2. 1~ 2. 28 |
| 중소기업 동향 | 공업통계 | 중 소 기 업 은 행 | '88. 7. 1~ 7. 10 |
| 수출산업 경기 예측조사 | 공업통계 | 한 국 무 역 협 회 | '88. 3. 3~ 3. 19 |
| 과학기술 연구 개발활동 | 공업통계 | 과 학 기 술 처 | 1985 |
| 운수업 통계 조사보고서 | 공업통계 | 경제기획원 조사 통계국 | '87. 4. 1~ 4. 30 |
| '87어가 경제조사 결과보고 | 농업통계 | 농 립 수 산 부 | 1988 |
| 농촌물가 총람 | 농업통계 | 농업협동조합중앙회 | 1985 |
| '87하반기 어업생산량 통계 | 농업통계 | 농 립 수 산 부 | '87. 1. 1~12. 31 |
| 가축통계 조사결과 | 농업통계 | 농 립 수 산 부 | '86. 6. 1 |
| 저축 시장조사 | 시장조사 | 한 국 은 행 저 축 부 | '84. 6. 23~ 7. 10 |
| 고용전망 조사보고서 | 시장조사 | 노 동 부 | '87. 4. 1~ 4. 15 |
| 도소매업 센서스 보고서 | 시장조사 | 경 제 기 획 원 | '82. 7. 1~ 7. 20 |
| 제4차절핵실태조사결과보고 | 보 건 | 보 건 사 회 부 | '80. 3. 10~ 9. 22 |
| 제4차한국장내기생충감염현황 | 보 건 | 보건사회부한국건강관리협회 | '86. 5. 15~ 7. 31 |
| 경제활동연구연보 | 인 구 | 경제기획원 조사 통계국 | 1987 |

이상의 표본조사에 의한 통계 57종의 분석은 먼저 보고서가 기본형식을 갖추고 있는가를 조사하고 표본조사 과정에서는 오차를 빈번하게 야기할 가능성이 있는 과정만을 개괄적으로 분석하였다. 분석결과를 요약하면 다음과 같다. 문제를 정확하게 설정하여 이용자들이 혼돈 없이 활용할 수 있도록 제공된 통계가 있지만 4종의 통계는 조사목적과 학문적 배경을 서술하지 않고 있었다.

항목별 분석결과는 아래와 같다.

표 4. 표본조사에 있어서 기본개념 규정여부

| 조사/항목 | 조사자 | 조사목적 | 조사대상 | 조사기간 | 조사기준시점 | 조사방법 | 모집단 | 프레임 | 추출방법 | 추정방법 | 오차한계 | 표본수 | 비용 |
|-------|-----|------|------|------|--------|------|-----|-----|------|------|------|-----|----|
| 표본조사 | 57 | 53 | 56 | 46 | 31 | 56 | 50 | 34 | 35 | 16 | 10 | 16 | 0 |

(1) 모집단 정의는 명확하며 프레임의 기술은 적절한가?

명색이 표본조사를 통한 통계보고서 인데 모집단 정의를 누락한 통계가 7종이나 되며 형식적으로 모집단의 정의만 주어진 경우는 37종으로 대부분이 프레임을 생략했거나 프레임이

남아서 비표본오차를 많이 발생하게 할 소지를 내포한 경우이다. 완벽하게 모집단 정의과정을 기술한 통계는 14종 뿐이다. 통계보고서 작성 실무자나 관리자들에게 기필코 주의를 환기해야 할 항목이 바로 모집단 정의의 중요성을 인식시켜야 하는 점이다. 이외에 부 모집단에서 표본추출을 실시한 후에 이를 수정하거나 보완함이 없이 전체 모집단에 적용시켜 무리한 추론을 내리는 경우도 있었다.

(2) 표본추출 방법은 적절한가?

표본추출 방법이 전혀 명시되지 않은 통계가 22종으로 표본조사의 기초적인 지식교육의 필요성을 실감하게 한다. 35종의 통계가 추출방법을 설명하고 있으나 완전하게 이론적인 타당성을 갖추고 있는 경우는 10종 뿐이다. 여기서 유의표본추출 방법을 적용할 경우에는 관제의 통계가 보고될 가능성이 크다. 또한 층화추출 방법을 사용할 경우에 대부분이 층화지수에 대한 언급이 없었으며, 프레임이 낡은 경우라든지 또는 다른 이유로 조사원이 추출한 표본을 측정하지 못할 경우에 대한 설명이 결여되어 있었다.

(3) 표본의 크기 산출식이 주어지고 정도나 신뢰도에 대한 언급이 있는가?

표본크기 산출공식이 주어진 경우는 16종 뿐이며 이 중에서 요구정도나 신뢰수준에 대한 설명이 부연된 것은 10종 뿐이다. 산출공식조차 언급이 없는 통계가 무려 41종으로써 통계자료의 신빙성을 더욱 의심하게 한다. 표본조사의 가장 큰 특징중의 하나가 경제성과 시간성이며 경제성의 효율을 높이기 위해서 이론적인 표본크기 산출공식을 제시하는 것이다. 표본의 크기에 따라서는 통계자료 이용상의 제한사항에 대해서도 언급해야 한다.

(4) 모수 추정방법이 명시되었으며 이의 오차에 대한 계산공식이 주어졌는가?

추정방법과 오차의 계산공식을 함께 설명한 통계는 6종 뿐이다. 추정방법만을 제시한 것은 16종이고 아예 추정에 관한 언급조차 없는 통계는 35종이다. 실질적인 표본조사의 목적은 미지의 모수를 추정하는 일이지만 언급조차 없는 통계보고가 65%나 됨은 우리나라 통계의 현주소를 말하고 있는 것이다.

통계학의 전문적인 이론연구나 기발한 표본조사법을 창안하는 것도 중요하지만 보다 더 시급하고 중점적으로 해결해야 할 일은 통계를 작성하는 실무자나 관공부서의 관리자들에게 기초통계학의 기본개념을 파악토록하여 이런 지식이 산지식화되어 곧바로 실용화 될 수 있도록 의식교육도 병행 실시해야 할 것으로 보인다.

관청의 의향대로 통계숫자를 작성하거나 공표하는 경우는 절대로 있을 수 없겠으나 현재의 의식구조나 통계학에 대한 인식결여는 그럴 가능성에 대한 의구심을 버릴 수 없게 하고 있다. 이를 탈피하기 위해서 정부부처에서 작성하고 있는 통계보고서는 표준화된 표본설계과정을 포함해서 총계숫자에 대한 신뢰를 높일 수 있도록 해야 할 것이다.

2. 4 통계조사의 문제점

우리 행정조직상 통계를 담당하는 부서들이 산재해 있어 종합 통계하는 기능이 없기 때문

에 표준화된 통계보고서 양식을 사용하는데 어려움이 있을 것이다. 현재 상황에서 해결해야 할 점은 통계담당 실무요원들에 대한 연수 교육계획을 수립하여 통계의 중요도를 재인식 시키고 새로운 표본조사 방법을 숙지할 기회를 만드는 것이다. 통계보고서의 조사절차 설명에서 조사방법, 조사의 설계 및 관찰방법을 가능한 상세하게 설명해야 한다. 특히 표본조사나 전수조사에서 이 부분이 보완되어야 할 점이다.

전수조사 통계에서는 면접방법이나 설문지 작성에 대한 절차의 설명이 없는데도 가장 이상적인 조사절차를 적용한 것으로 과신하여 통계를 활용함으로써 많은 잘못을 범할 것이다. 특히 전수조사에서 표본오차는 무시할 정도가 될지언정 비표본오차는 상당히 커질 것이다. 비표본오차에 대한 내용은 제4장에서 자세히 다루어 진다. 특히 조사절차에서 오류를 범하는 중요한 사항이 모집단의 정의이다. 조사목적에 꼭 맞는 조사대상과 조사시기를 선정하지 못하고, 많은 경우에 시간이 몇년 지난 목록을 프레임으로 사용했다. 가능한 최근의 목록을 획득하여 프레임으로 삼을 수 있도록 유의해야 한다. 가장 바람직한 방안은 조사통계국 등에서 프레임으로 사용될 수 있는 목록들을 정리 보관하여 전수조사나 표본조사 설계시 이용할 수 있도록 하는 제도가 정착되어야 할 것이다.

조사방법에서는 무응답자에 대한 별도의 대책이 있어야 한다.

무응답률이 적은 경우에는 심각한 오차를 초래하지 않겠지만 만일 무응답 이유가 반대를 위한 행동의 일환이었다면 여기서는 예측하지 못했던 오류를 범하게 될 것이다. 무응답자가 있을 경우를 보완하는 조사방법이나 추정방안을 염두해 두어야 한다. 통계보고서 작성상 표본조사 설계의 내용이 많을 경우에는 별도의 부록으로 첨부 해야만 결과에 대한 감정을 할 수 있다. 추정공식과 추정오차에 대한 언급이 없을 경우에는 이론적 추정값은 무용지물이 되고만다. 정확도와 신뢰계수에 대한 설명이 있어야 함에도 거의 대부분의 통계보고서에는 생략되어 있으므로 감정이나 측정값의 검토가 불가능하다. 대부분의 통계에서 표본의 크기를 계산하는 경우 보다는 미리 정해진 크기에 알맞게 배분 함으로써 통계학의 이론적 배경을 무시하였다. 따라서 가능한한 경제성과 정확성을 기초로 하여 결정할 수 있도록 해야 할 것이다.

3. 効率的標本設計의 主要過程

이상에서 통계조사결과 전반에 관한 내용을 분석하여 문제점을 찾아 보았으나 통계결과에 많은 영향을 미치고 기획단계에서부터 정밀통계이론이 많이 작용하는 과정은 표본조사의 표본설계 과정이다. 제3장에서는 표본조사법을 중심으로 연구하여 오차를 감소시키기 위해서 표본조사를 과정별로 점검해야 할 내용들을 논의하기로 한다.

최적의 표본조사를 통하여 목표로하는 정도를 갖는 추정값을 얻기 위하여 각 주요 과정별로 점검해야할 유의 사항이 있다. 이것을 표준화(점검표화)하여 사용하므로 신뢰성 높은 통계를 생산하는데 활용하도록 해야한다.

(1) 調査者の 明示

調査를 進行하는 기관이 調査者이다. 調査者는 調査主體로서 被調査者인 調査客體로부터 얻고자 하는 정보를 얻어내는 것이 조사자인 것이다. 따라서 조사자가 명시되어야함은 1次的으로 중요하다. 간혹 통계조사에 있어 調査者가 明示되어 있지 않은 경우를 보게되는데 너무도 어이없는 일이다.

(2) 調査對象의 確定

調査者는 被調査者 즉 調査客體가 있어야 조사를 할 수 있다. 이 조사객체가 조사대상이 되는 것이다. 학교당국이 국민학교 학생의 체격조사를 하려고 할 때 학교당국은 조사자이고 학생 개개인이 조사대상이 된다. 통계조사에서 이 개념이 분명하지 않은 경우가 많이 있다.

(3) 調査目的의 明示

調査者는 피조사자로부터 어떠한 정보를 얻어낼 것인가를 분명하게 밝혀야 한다. 피조사자로부터 얻고자하는 정보가 분명하지 않으면 조사자의 조사진행부터 결과가 애매해지고 또한 이용자도 어려운 문제에 봉착한다. 이와같은 조사목적은 통계조사에 있어서 가장 중요하기 때문에 바르고 분명히 정해야 한다.

(4) 母集團의 定義

調査對象 全體로된 집단이 모집단이다. 일견 구체적으로 쉽게 모집단이 규정될 것 같지만 사실상은 그렇지 않다. 따라서 조사자는 조사대상을 정하고 모집단을 잘 규정해야 통계조사가 바르게 진행될 수 있고 조사 결과인 통계가 어떠한 집단에 대한 통계로서 인정되어 이용자가 바르게 이용하게 된다. 앞에서 전수조사나 표본조사가 모집단의 정의가 없이 진행된 경우를 보았다. 정말 어처구니 없는 일이 아닐 수 없다.

(5) 調査期間과 調査基準時點의 設定

調査者는 언제 조사를 진행할 것인가를 일정 기간으로 정해야 한다. 조사하는 과정에서 조사사항들에 대한 답을 어느 시점에서 할 것인가가 문제가 된다. 따라서 조사 시점을 정해야 한다. 조사기간이나 시점이 명시되지 않은 통계는 이용 가치가 없다. 집단현상은 항상 변화하고 그 구성원도 달라지기 때문에 기간이나 시점의 설정없이 조사가 진행될 수 없다.

(6) 프레임과 추출단위 선정

모집단에서 표본을 추출하기 위하여 모집단을 표본추출이 가능하도록 프레임(Frame)으로 정리해야 한다. 그리고 추출시 단위로 쓰여지는 추출단위를 정해야 한다. 프레임은 추출단위가 빠지지 않고 중복되지 않게 리스트 되어야 한다. 이러한 원칙이 지켜져서 프레임이 작성되고 추출단위가 정해질 때 추정의 결과가 분명해진다. 추출단위를 어떻게 정할 것인가는 표본조사의 입장에서 매우 중요하다.

(7) 抽出方法의 선정

모집단을 잘 대표하는 표본이 선정되게 추출방법을 1次的으로 고려해야 하고 2차적으로 표본 오차를 계산할 수 있고 또한 최소화 시키는 추출방법을 생각해야 한다. 또한 추정방식이 가능한가와 주어진 비용이나 정도에 대하여도 고려된 추출법을 강구해야 한다. 57종의 표본조사중 35종(61%)만이 추출방법을 조금은 설명해 놓고 있었다. 표본조사가 어떻게 진행되는 줄도 모르고 표본조사를 한것이 39%라는 현실이 매우 안타까울 뿐이다.

(8) 調査表와 結果表의 計劃

調査表는 調査目的에 따라 調査事項에 대하여 실문을 만들어서 羅列한 것이다. 調査者는 이 調査表에 의거 被調査者를 直·間接으로 상대하여 조사하게 되며 이것을 바탕으로 하여 통계가 작성되고 結果表로서 統計表가 나온다. 따라서 調査表를 꾸미는 과정은 정보를 얻어 내는 과정에 중요한 영향을 미치며 또한 그에 따라 통계표가 작성된다.

(9) 調査方法

統計調査를 全數調査로 할 것인가 아니면 標本調査로 할 것인가가 1차적으로 정해지면 調査者는 直接調査할 것인가 間接調査할 것인가 어느것이 얻고자 하는 정보를 보다 정확하게 얻는 방법이며 예산이나 노력이 적게드는 방법인가 등을 조사목적이나 대상에 따라 잘 정하는 것이 정확한 조사를 진행할 수 있는 것이다.

(10) 要求된 精度

母數에 대한 推定量의 精度가 計量될 수 있어야 한다. 推定量의 오차를 계산할 수 없는 경우는 調査者와 利用者가 협의하여 다른 방안이 강구 되어야 한다. 그리고 주어진 비용으로 요구된 精度를 달성할 수 있는가도 따져야 한다. 이 과정은 다른 과정과도 연관되어 있으므로 무리한 제한조건들은 조사결과에 다른 영향을 끼치므로 주의해야 한다.

(11) 調査業務의 관장

조사기관의 조사 총책임자는 조사목적이나 조사업무에 대해서 구체적으로 충분히 알고 있어야 한다. 統計調査는 전수조사 이거나 표본조사 이거나 정밀하게 그리고 정확하게 진행되어야 한다. 조사기관의 각 부서나 모든 조사과정에서 언제든지 잘못이 발생할 여지는 존재한다. 그 잘못이 곧 통계의 오차로 나타나기 때문에 책임자의 관리가 소홀하면 그에 따라 오차는 가중되어 나타난다. 그리고 조사의 일선에서 활동은 조사원이 담당하고 있으므로 조사원의 역할이 매우 중요하다. 조사원이 조사에 충분히 임할 수 있도록 사전 교육과 넉넉한 보수 그리고 적정한 분량의 책임 등을 책임자가 고려하여 금지를 잃지 않고 조사활동을 할 수 있게 해야한다.

(12) 標本數 n의 決定과 비용

일반적으로 우리는 제한된 예산하에서 움직이고 있다. 통계조사도 주어진 예산하에 진행되기 때문에 예산과 주어진 정도에 적합한 표본수(n)를 정하여 조사를 진행해야 한다. 그런

데 우리의 현실은 정도에 대한 이야기는 좀(35%) 나와 있으나 비용에 대한 이야기는 전혀 나와있지 않다. 특히 표본조사의 경우는 전수조사에서 과다한 비용이나 노력 또는 많은 시간을 줄이기 위하여 계획되는 조사 이므로 비용과 정도에 따라 표본수 결정을 합리적으로 정해야 하는것을 잊어서는 아니된다.

(13) 調査資料의 分析 및 要約

調査된 資料가 잘못된 것은 없는가를 점검할 수 있는 과정이 있어야 한다. 그리고 얻어진 자료를 계량화 하는데 어려움은 없으며 分析 및 要約이 가능한가를 확인해야 한다. 또한 結果表 作成에 만족한 결과를 얻을 수 있는가를 점검해야 한다.

4. 誤差의 發生原因과 管理原則

표본조사나 전수조사의 2가지 경우에서 발생가능한 오차가 비표본오차이다. 표본오차는 모집단 전부를 관찰하지 않고 그중의 일부를 추출하며 관찰한 것을 이용하여 모수를 추정하기 때문에 생긴 오차를 말하지만 비표본오차는 조사의 계획단계에서부터 자료분석이나 보고서 작성의 최종단계의 최종과정에서 부주의나 실수 또는 원인을 알 수 없는 이유 중에서 생기는 오차인 것이다. 비표본 오차의 발생 주요원인은 연구목적의 불명확성, 조사범위의 불분명함, 모집단의 전 영역을 포함하지 못한 조사, 잘못된 개념의 정의, 자료 수집방법의 결함과 분석될 자료의 도표화에서 생기는 오류 등으로 볼 수 있다. 이러한 종류의 원인들을 크게 3가지로 분류하여 비표본 오차의 원인과 종류를 알아 보기로 한다.

4. 1 非標本誤差의 發生原因

4. 1. 1. 조사계획 과정에서 유발되는 오차(무응답오차)

이 오차는 불충분한 준비에서 생기는 것이며 주로 잘못된 표본을 사용하거나, 표본추출단위 및 추출방법의 결함과 부적절한 조사일정 등에서 발생하는 것이다.

만약에 표본틀을 새로 수정하지 않고 낡은 틀을 사용 했다면 표적모집단중 일부는 탈락되어 심각한 편의를 야기시킬 수도 있을 것이다. 어떤 경우에도 추출된 조사대상의 질문에 응답을 거절 하거나 방문시간에 피조사대상자가 자리에 없을 수 있는 경우에 비표본오차는 발생하게 된다. 위와같은 종류의 오차들은 다음과 같은 원인에서 생길 것이다.

- (1) 추출단위의 불명확한 정의로 인한 추출단위의 누락이나 중복, 또는 조사목적에서 벗어나거나 부정확한 개념의 정의
- (2) 정확하지 않은 면담방법이나 잘못된 조사일정 등
- (3) 피조사자들의 질문에 대한 이해부족이나 잘못 구성된 조사표와 자료수집 방법의 결함

4. 1. 2 실제 조사과정에서 발생하는 오차(응답오차)

이 오차는 측정되는 값과 참값의 차이를 말하며 참값을 측정하지 못하는 이유는 다양하다. 예를들어 경지면적조사에서 농장주는 자기소유의 땅이 10ha이라고 응답했지만 실제 지적도

에는 11ha로 기입되어 있다면 1ha는 농장주의 의도적이거나 정확하게 알지 못함에서 오는 오차이다. 이와같은 오차들의 양은 조사의 성격이나 형태 또는 수집하고자하는 정보에 따라 좌우되며 발생원인은 조사자나 응답자 또는 두 사람에게 다 있을 수 있다. 때로는 두가지 원인 사이에 상호작용이 있을 수 있고 이에따라 오차가 커지기도 한다. 측정도구나 기술의 결함에 의해서 응답오차는 생길 수 있지만 응답오차를 유발하는 주요원인들은 다음과 같은 것이 있을 것이다.

- (1) 조사원에 대한 불충분한 감독
- (2) 조사원의 경험미숙과 훈련부족
- (3) 자료수집 과정상의 문제점과 응답자로 인해 발생하는 기타의 오차

4. 1. 3 자료처리 분석과정에서 발생하는 오차(종합표의 오차)

수집된 자료를 처리 분석하는 과정에서 발생하는 오차는 편집방법, 코딩, 도표화 등의 결함에서 올 수 있고 통계적 추정 과정의 오류에서 생길 수 있으며 오차발생의 주된 원인은 다음 사항이 될 것이다.

- (1) 기본자료의 검사미흡
- (2) 코딩 또는 자료입력 등의 자료처리 과정중 발생하는 오차
- (3) 결과 발표나 인쇄과정 중에 범해지는 기타 오차

이상 3가지로 구분하여 나열된 비표본오차가 발생될 원인들은 오차발생 가능성이 있는 원인에 대한 예를 든것 뿐이지 모든 원인을 총 망라한 것은 아니다. 실제로 조사하는 과정 중에는 더 많은 종류의 오차들이 일어날 수 있기 때문에 표본조사 과정에서 여러 형태의 작업이 복합적으로 이루어지고 있을 때 어떤 종류의 오차는 용인될 수 있고 어느것이 잦은 빈도로 발생되는지 또는 결과에 어떤 영향을 미칠 것인지를 판단 하기란 매우 어려운 일이다.

특히 여러 종류의 오차들이 혼합되어 있는 정도를 측정하기란 더 어려울 것이므로 표본조사 설계 당시 통계학자나 통계결과 이용자들은 비표본 오차가 존재할 수 있음에 주의를 기울여야 하고 이것을 가능한 적게 하도록 가능한 조치를 취해야 한다.

4. 2 비표본 오차의 측정과 제어

비표본 오차 측정시 적절한 방법의 적용과 이를 제어하기 위한 적당한 과정의 선정은 과정중의 일부는 표본조사 설계 중에 포함될 수 있으므로 표본 조사 이전에도 세심한 주의가 필요하다. 오차의 측정과 제어를 위해서 정해진 과정과 별도로 마지막 결과에서 출현될지도 모르는 다른 비표본 오차 추정을 위한 독단적인 계획이 필요하게 될 수도 있다. 이런 경우에 대비해서 보다 많은 비용이 들더라도 더욱 정밀한 자료 수집과 처리 방법을 설명하고자 한다.

4. 2. 1 일관성 점검

수집된 자료의 질을 점검하거나 어떤 항목을 포함하고자 할 때는 설문지나 일람표 작성에서 특별히 주의를 해야하며 만약에 질에 대한 정보를 얻기 위한 항목의 조사가 간단하다면 모든 표본 조사에 적용토록하고 그렇지 않은 경우에는 부차표본을 추출하여 그 단위에만 적용함이 좋다. 성별의 비율 같이 안정된 성질을 갖는 항목을 포함시켜 조사자료의 질을 측정하는 것도 유용할 것이다.

또 다른 자료의 질을 측정하는 방법은 분량의 관찰치의 위치를 찾아내기 위해서 어떤 기본 변수의 증가형태로 관찰치를 배열하여 각 표본 단위에 대하여 그림표를 그렸을 때 이 그래프의 형태를 예측할 수 있게 되는데 이 형태에서 이탈되는 값을 갖는 단위를 찾는 것이다. 그리고 이상한 값을 정확하게 조사하면 오차를 줄일 수 있을 것이다.

4. 2. 2 표본점검

오차의 교정과 탐지를 촉진하기 위해서 실행의 다른 과정 중에서 작업을 독립적으로 복제하는 것이 전수조사나 표본조사에서 비표본오차를 제어하거나 측정하는 한 방안이다. 소요되는 비용, 시간과 훈련된 요원의 부족으로 이 방법을 전 실행 과정에서 실시할 수 없으므로 훈련되고 경험있는 비교적 적은 집단을 이용하여 작업과정의 일부를 선정하여 복제 점검을 실시한다. 만약에 표본이 적절하게 설계되고 점검 작업이 효율적으로 실시된다면 비표본 오차의 탐지뿐만 아니라 그의 크기에 대한 윤곽을 얻을 것이다.

만약에 조사업무 전 과정에 대해서 완전 점검이 가능하다면 최종 결과의 질은 상당히 증진될 것이다. 그러나 표본점검에서는 교정작업이 점검된 표본에만 실시 되므로 추출되지 않는 부분의 오차들은 수정되지 않고 있을 것이다. 이런 문제들은 조사의 각 단위에서 나온 결과를 몇개 무더기로 나누어서 각 무더기에서 추출된 표본들만 점검함으로써 부분적으로 해결 할 수 있으며 이때 특정한 무더기에서 오차율이 크면 그 무더기 전체를 점검하여 오차를 교정하여 최종 결과의 질을 향상시킬 수 있다.

4. 2. 3 사후 표본조사(사후 전수조사)점검

비표본오차를 측정하기 위해서 본 조사를 실시한 사람들 보다 더 잘 훈련되고 경험이 많은 조사원을 이용하여 표본 조사에서 조사된 단위들이 부차표본에 대하여 재조사하는 방법을 사후 표본조사 점검이라 한다. 이는 표본 점검의 한 유형으로 생각할 수 있고 이것이 보다 더 큰 효과적인 것이 되기 위해서는 재조사를 본 조사후 즉각 실행하여 가능한 재조사 오차를 피해야 한다.

점검조사는 포함범위(Coverage)와 내용(Content) 오차 두가지 다 측정하기 쉽도록 설계되어야 하며 이를 위해서는 무엇보다도 고단계 지역단위들의 표본에 있는 단위들은 모두 재조사함이 좋고 포함범위 오차의 탐지를 위해서는 단지 최종 단위의 오차만 재조사 한다. 점검 조사가 효과적으로 잘 되었는지 확인하기 위해서 조사자는 표본추출 단위들의 목록과 표본조사에서 얻어진 잘 채워진 설문지나 일람표를 갖고 있어야 하며 점검조사에서 얻어진 자

료와 목록표를 본 조사의 것들과 비교해야 한다. 비교결과에서 점검조사업무의 질을 평가할 수 있게 된다. 인위적으로 삽입된 오차의 존재를 알림으로써 조사자를 보다 더 세심하게 만들 것이다. 조사자들에게 점검표본 단위에 인접한 모든 단위들의 리스트를 작성토록 하여 본 조사에서 이런 단위들이 포함되었는지를 확인하는 것이 바람직하다.

점검조사의 특이한 장점은 점검표본에 포함된 단위들에 대한 두개의 측정값을 대응시켜 관찰된 각개의 차이를 분석할 수 있는 것이고 만약에 상이함이 있다면 비표본 오차의 종류와 성질을 면밀히 분석하여 오차 존재원인을 아는 것이다. 두개의 자료집합을 대응시키는 일은 시간소비와 비용 등의 이유로 어려움이 많지만 경우에 따라서 유용하게 이용되기도 한다.

만일 추출 단위간에 대응을 통한 관찰치 차이의 분석이 불가능 할 때는 점검조사에서 얻어진 자료를 기초로한 모수의 추정치 들과 본 조사의 자료를 근거로 추정치들을 비교함으로써 순수한 오차에 대한 정보를 얻을 수 있는 반면에 각 표본단위의 차이 분석에서는 순수 오차와 총 오차에 대한 두가지 정보를 얻게 된다.

4. 2. 4 외부기록 점검

전수조사에서 비표본오차를 평가하는 한가지 방법은 가능하면 별도의 방법으로 적절한 단위들의 표본을 취해서 모든 단위들이 본 전수조사에서 조사되었는지를 확인하고 서로 대응되면 두값 사이의 상이함이 있는지 알아본다. 이와같은 방법에서 추출한 점검표본들의 모집단의 목록표는 완벽하지 않아도 된다. 예를들면 인구전수조사에서 점검을 실시하는데 이용한 점거표본들은 최근 출생신고 기록부 학생들이나 양로 연금수혜자 명부 등에서 추출될 수 있다. 식별과 위치파악의 어려움 때문에 본조사의 단위들과 대응 시키기가 힘든 것이 이 방안의 단점이다.

4. 2. 5 품질관리 기법

통계적 관리기법이 전수조사나 표본조사 업무에 적용될 수 있다. 왜냐하면 위의 조사업무도 대규모이고 반복적 특성을 가진 작업의 일종이기 때문이다. 관리도나 합격 샘플링 기법들은 대규모 표본조사나 전수조사에서 자료의 질을 평가 하거나 최종 결과의 신뢰도를 향상시키기 위해서 유익하게 이용될 수 있다. 시작부터 올바른 행동을 하도록 규정화된 규칙들의 기법이 사용되는 예를 보자. 전수조사나 표본조사 과정을 각 작업자가 생산품을 객관적으로 검사해 불량률이 나오는 생산과정으로 간주할 수 있다.

각 작업자의 제품이 처음 시작단계에서는 100% 검사되는데 만약에 불량률이 정해진 수준 이하로 떨어지면 그의 제품이 한 표본에 대해서만 검정되며 표본의 크기도 적당하게 결정된다. 만약에 그렇지 않으면 그 제품의 완전한 검정은 계속된다. 검정의 성질에 관한 결정은 각 작업자에게 개별적으로 주어지고 과거에 누락된 불량률을 기초로 하게 된다. 이런 유형의 기법은 큰 규모 업무에 유익하게 이용되어 검정과정의 비용을 감소시킬 뿐만 아니라 최종결과에 대한 명시된 질적 수준이 보장될 것이다.

4. 2. 6 재조사 오차의 연구

재조사 오차 연구의 한 방법은 조사에 포함되어 있는 단위들의 부차표본에서 한번 이상 보고주기와 관련된 자료를 수집해서 분석하는 것이지만 이 연구에서 주된 문제점은 어떤 보고주기에 대한 자료에 다른 보고주기에 보고된 것들의 영향의 크기이다. 서로다른 보고주기를 기초로한 결과들이 일치하지 않을 때 결과들의 차이는 제어표본을 갖거나 가능한 모든 보고주기를 포함하는 시간의 주기에 대한 신뢰할 수 있는 자료를 얻기 위한 특별한 노력을 통하여 어느정도 극복할 수 있을지라도 어떤 보고주기를 인정할 것인가를 결정하기가 어렵다. 위에서 언급한 조건화된 영향을 피하기 위해서는 고려대상의 다른 주기에 대한 자료를 다른 표본들로부터 수집할 수 있으며 이때 유용한 비교를 하기 위해서 독립적인 표본들이 보고주기에 따라 사용된다면 큰 표본이 필요하다. 이 어려움은 다른 보고주기와 상호 관입 부차표본(Interpenetrating sub-sample)과 연관지어 조사한다면 극복할 수 있을 것이다.

재조사 오차 연구의 다른 방법은 다른 보고주기에 대한 추정치를 허용할 수 있는 어떤 추가적인 정보를 수집하는 것이다. 예를들면 인구통계조사에서 작년 한해동안 태어난 사람의 숫자 뿐만 아니라 출생월, 주, 일에 대한 정보를 수집할 수 있으면 사건 발생과 보고날짜 사이의 간격에 따른 출생률을 도표화할 수 있으므로 출생숫자에 대한 보고된 날짜에서 나타날 수 있는 재조사 오차를 탐지할 수 있다.

4. 2. 7 상호관입 부차표본

상호관입 부차표본 기법은 1930년대에 Mahalanobis가 착안 하였고 이는 동일한 방법으로 추출한 두 개 이상의 부차표본으로 이루어진 표본을 추출하는 것이며 각 부차 표본에서 정당한 모수추정을 할 수 있도록 한 것이다. 이 기법은 면담자들에 의한 편의와 정보수집 방법들의 상이함에서 생기는 차이와 같은 비표본 오차를 제어할 수 있는 방법으로 사용 되어 질 수 있다.

부차표본들이 서로 다른 그룹의 조사자에 대해서 관찰되고 처리과정에서 또 다른팀의 조사자들에 의해서 처리한 후에 부차표본들을 기초로한 최종 추정치들을 비교함으로써 조사결과에 대한 질을 점검할 수 있다. 예를들어 서로 다른 그룹의 조사원들에 의해서 조사되고 처리된 네개의 부차표본에서 구한 최종 추정치를 비교하였을 때 세개의 추정값은 비슷하나 다른 하나는 상당히 차이가 큰 추정치라면 일반적으로 상이한 부차표본에 대한 업무의 질에 대한 의문을 갖게 될 것이다. 만일에 자료의 질 평가에 주로 관심을 갖는다면 상호관계가 높은 것들에 기초를 둔 추정량이 되도록 부차표본을 설계해야 한다. 두 개 이상의 동질의 단위의 집락을 표본으로 추출하여 각 집락에서 한 개 단위씩 뽑아 부차표본을 만들면 비표본 오차를 평가할 수 있다. 다른 종류의 비표본 오차를 추정하기 위해서 부차표본의 추정량의 비교는 효과적이 될 것이다. 예를들면 인구센서스에서 두 사람의 조사원에 한집씩 교대로 건너서 조사하도록 배열된 조사단위에서는 두 추정치의 차이가 비표본 오차이다. 비표본 오차의 크기와 성질을 평가하기 위해서 특별하게 훈련되고 경험있는 조사원들에 의해서 한 개의 부차표본이 조사되고 분석되어야 한다.

조사원간의 변동, 도표작성팀간의 변동과 교호작용들과 같은 오차의 여러가지 성분 연구는 여러개의 상호관입 부차표본들을 추출하여 다른 조사원의 그룹과 도표 작성원의 팀들에 의해서 조사업무와 도표작성 업무를 수행토록 함으로써 예를들어 도표작성팀이 두개있고 조사원 집단이 네개 있다면 8개 부차표본으로 된 표본을 추출하여 각 도표작성팀이 부차표본 네개씩을 담당하고 각 조사원 그룹은 각 팀에서 하나씩 뽑아 두개의 부차표본의 조사를 담당하도록 한다. 이 설계에서 총 변동을 도표 작성팀간의 변동, 조사원 그룹간의 변동과 잔차 등의 성분으로 나누어 분석이 가능하다.

5. 통계보고서 형식의 표준화

제2장의 분석에서 가장 큰 애로사항이 보고서들이 형식을 갖추지 않았기 때문에 통계내용을 정확하게 파악하지 못할 뿐 아니라 통계이용을 활성화할 수 없었다. 따라서 제5장에서는 통계보고서의 표준화를 위하여 형식의 골격을 설명하고 130종의 통계보고서 중에서 참고가 될 수 있는것을 전수조사와 표본조사의 경우에서 첨부하여 통계보고서 이용율을 제고할 뿐 만 아니라 신뢰받는 통계생산의 기반 조성에 도움이 되었으면 한다.

5. 1 전수조사 통계보고서

통계조사가 어떤 배경하에서 어떤 목적을 갖고 어떤 절차를 통해서 시행되었고 그 조사결과와는 어떤 것이며 여기서 도출한 결론은 무엇인가를 제시해야 한다. 따라서 반드시 포함되어야 할 4가지 항목을 간단하게 설명하자.

5. 1. 1 문제의 진술

통계조사문제의 진술을 통해서 조사의 목적이 명백하게 주어져야 하고 이 문제를 연구하는데 관련된 학문적 배경에 대한 설명이 주어져야 한다. 첫머리에서는 반드시 통계조사 문제의 이론적 내지는 실제적 중요성이 언급되어야 하며 이 조사와 관련된 다른 조사의 소개와 연혁 등이 필요하다. 그리고 가설이 설정되어 있으면 가설을 밝혀야 하고 통계조사에서 사용된 주요 개념을 정의하여야 한다.

5. 1. 2 조사의 절차

다음에는 조사가 어떻게 시행 되었는가를 밝혀야 한다. 조사의 설계는 어떻게 하였는가? 측정은 어떻게 하였는가? 만약 질문지나 면접의 방법으로 데이터가 수집되었다면, 어떠한 질문을 통하여 얻어진 것인가도 밝혀야 하며, 조사원(면접자)들은 어느정도 훈련을 받았으며 면접중 얻은 경험도 알려 주어야 한다. 관찰에 의한 데이터 수집방법을 사용하였으면 어떤 관찰 지침하에 이루어 졌는가도 밝혀야 한다.

5. 1. 3 조사결과

결과를 제시할 때에는 조사자의 편견없이 문제에 적절한 모든 증거들을 밝혀야 한다. 만

들어진 모든 결과를 제시하기 보다는 조사목적에 타당한 것만 골라서 제시하고 실제로 이용자를 위해서 요약한 결과를 포함하는 것이 좋다. 조사결과가 많은 경우에는 별첨으로 하고 요약된 결과만 간단하게 넣기도 한다.

5. 1. 4 결론

통계사용상의 유의사항이나 제한사항을 명시하여 잘못 이용되는 폐단을 방지한다. 보고서 작성과정이나 통계조사 수행과정 중에서 특기할 만한 사항이 있을 때에는 차후의 통계조사를 위해서 기술함이 좋다.

5. 1. 5 전수조사 통계보고서 예

참고문헌 (2)참조

5. 2 표본조사 통계보고서

전수조사 통계보고서 양식에서와 같이 문제의 진술 조사의 절차, 조사결과와 결론 등의 항목이 필수적으로 포함되어야 한다. 그러나 조사절차는 전수조사와 다르게 설명 되어야 하며 다른 항목들은 거의 같으므로 조사 절차만 상세하게 설명하기로 한다. 조사절차 중에서도 표본설계에서 모집단 설정, 표본추출 방법, 표본크기의 계산, 추정공식과 그의 오차에 대해서 알아보자.

5. 2. 1 모집단 정의

우선 조사에 앞서 조사하고자 하는 대상에 관한 모집단을 명확히 규정하여야 한다. 특히 규정하고자 하는 모집단은 때, 장소, 속성 등 세가지 점에 대해서 명확하게 표현해야 한다. 경우에 따라서는 모집단의 설정이 확정적으로 정확하지 못하여 표본을 모집단에서 추출할 수 없어서 모집단을 추출단위의 리스트로 표현 하도록 만든 프레임을 이용하여 표본을 추출한다. 이때 프레임과 모집단 사이에 겹이 없어야 하며 만일에 있을 경우에는 이를 참조하여 추정값을 수정해야 한다. 절대로 모집단의 일부를 표본대상에서 제외하여 부 모집단 표본추출이 안되도록 해야한다. 프레임 선정시 가능한 최근에 만들어진 목록표를 이용해야 한다.

5. 2. 2 표본추출 방법

모집단을 가장 잘 대표할 수 있는 표본을 추출해야 한다. 여러가지 추출방법 중에서 적합한 것의 선정은 활용가능한 자원과 요구한 정확도의 조건을 충족할 수 있을 만큼의 표본수를 결정하여 표본추출의 편의성, 모수추정의 적법성을 고려하여 결정한다. 구체적인 설명이 있어야만 한다. 예를들면 층화추출법에서는 층화지표는 무엇이며 표본의 층에 대한 배정방법 등의 설명이 되어야 한다.

5. 2. 3 표본크기의 계산

조사비용이나 조사 소요시간의 제한조건 안에서 오차를 최소화할 수 있도록 표본 크기를 결정해야 한다. 실제로 표본조사에서 주어진 요소인 조사비용과 허용오차의 한계를 명시하

여 표본 크기 공식도 제시하고 이에 따라 계산해야 한다. 다목적 표본 조사의 경우에는 어떤 방법으로 표본의 크기를 결정 하였는지 설명이 있어야 한다.

5. 2. 4 모수 추정공식과 오차계산

표본추출 방법이 간단하던지 또는 복잡하던지 전혀 개의치 말고 추정공식을 설명하고 이에 대한 분산공식도 함께 설명이 되어야 한다. 여러목적에서 표본조사가 실시된 경우에는 각각의 목적에 대한 공식이 주어져야 한다.

이상 간단하게 조사절차에 대한 내용만을 설명하였는데 표본조사 시에는 통계이용의 제한 사항에 대한 설명을 첨가함으로써 통계학의 전문지식이 없는 사람들에게 많은 도움을 줄 수도 있다.

5. 2. 5 표본조사 통계보고서의 예

참고문헌 (22)참조

6. 결 론

본 연구는 우리나라 행정부처나 공사기관 또는 기업체 등에서 조사에 의한 통계작성 현황을 파악하고 이들이 작성한 통계보고서를 분석하여 통계활용 분야의 현주소를 알고 또한 이들의 문제점을 찾아 개선 방안을 마련하고자 하였다. 그러나 지금까지 검토한 보고서들의 형식과 내용이 미비하여 통계조사의 본질적인 문제점 파악이 매우 어려운 것을 느꼈다. 따라서 현행 작성되는 중요 통계 342종 중에서 130종을 추출 분석하여 문제점을 제시하고 그 해결방법을 찾기로 하였다.

여기서 몇가지 공통된 문제점들은 통계조사 방법을 올바르게 이해하지 못하고 있으며 실제 조사의 진행도 애매한 경우가 많았다. 또한 비표본 오차에 관한 연구나 이를 감소시킬 대책이 미진하였다.

우리나라 통계조사의 현실적인 문제점을 계량적으로 나타내기 위한 컴퓨터 분석은 실행 불가능 하지만 최적통계 표본조사를 위한 각 과정별 점검사항을 설명하여 어떠한 분야에서 도 표본조사를 하고자 할 때는 참고하여 활용토록 제시되었다.

앞으로 통계숫자의 활용도와 신뢰도를 제고시키기 위해서는 표준화된 통계보고서 형식이 요구되어 지고 또한 통계조사의 발전을 위해서 명료한 통계보고서가 작성되도록 담당부서 요원들의 교육이 절실하게 요구된다.

통계조사 보고서 형식의 표준 구성에 대해서 간단하게 설명함과 아울러 130종의 통계보고서 중에서 참고가 될 만큼 작성된 표본조사 통계보고서(1986 생명보험 성향 전국조사)와 전수조사 통계보고서(1980 농업조사)를 참고문헌 (21), (22)로 두어 참고토록 하였다.

특히 본 연구의 특징은 비표본 오차에 대한 심층있는 연구로 비표본 오차의 본질을 규명하고 발생원인을 분석하였다. 비표본 오차를 줄이기 위한 방안과 이 비표본 오차를 측정 및

제어하는 기법들이 연구되었다.

앞으로 우리나라 통계작성의 현 실태를 명확하게 파악하기 위해서 센서스 한 자료를 획득하여 분석하고 실제 훈련된 조사원을 활용하여 효과적인 표본조사법을 개발하고 실용할 수 있도록 많은 시간과 비용을 투자해야 할 것이다. 특히 통계작성 요원과 관리자에 대한 통계조사법의 교육 프로그램 개발에 관한 연구는 필연적으로 추진되어야 한다.

참 고 문 헌

- (1) 經濟企劃院 調査統計局(1988), 統計目錄
- (2) 經濟企劃院 調査統計局(1988), 非標本誤差의 最小化方案
- (3) 韓國統計學會(1987), 응용통계연구, 제1권 제1호
- (4) 金鍾浩(1987), 標本調査法
- (5) 金海東(1982), 調査方法論
- (6) 金景中(1989), 韓國의 經濟指標
- (7) 金景中(1987), 統計行政의 發展方向
- (8) 金景中(1988), 統計의 現況과 改善方向
- (9) Statistics Canada, Survey Methodology(1975-1985), 25 Volume's.
- (10) National Research Council(1988), The Aging Population in the Twenty-First Century.
- (11) NCHC(U.S.A.), Health Interview Survey Procedure(1957-1974).
- (12) NCHC(U.S.A.), The National Health Interview Survey Design(1975-1983)
- (13) Kruskal, W. H. and Tanur, J. M.(1978), International Encyclopedia of Statistics Vol. I, II, The Free Press.
- (14) Hansen, M.H., Hurwitz, W.N., and Madow, W.G.(1953), Sample Survey Methods and Theory, Vol. I, II, New York.
- (15) Deming, W.E.(1950), Some Theory of Sampling, New York.
- (16) Yamane, T.(1967), Elementary Sampling Theory, Prentice-Hall.
- (17) Cochran, W.G.(1963), Sampling Techniques, New York.
- (18) Kish, L. (1965), Survey Sampling, New York.
- (19) Yates, F.(1960), Methods for Censuses and Surveys, Charles Griffin.
- (20) Bureau of The Census(U.S.A.), The Current Population Survey(Design and Methodology).
- (21) 농수산부, 1980농업조사보고서
- (22) 생명보험협회, 생명보험성향 전국조사 보고서(1986)

A Study on the Efficient Sample Design and the Error in the Statistical Survey

Jong-Ho Kim* · Pyung Namgoong · Hong-Nai Park
Kei-O Lee · Sang-U Lee · Chong-Uk Ham

Abstract

In this paper, we propose many problems by analyzing the various currently statistics, propose the basic principal for the more precise performance of sample survey, and at the same time propose the method(principal) for detecting and manipulating the source of error.

Finally, we are to improving the statistical reliability and increasing the user's convenience by standardizing the statistical reporting form.

*Dept. of Statistics, Dong-Kuk University,
Pil-dong, Chung-gu, Seoul.