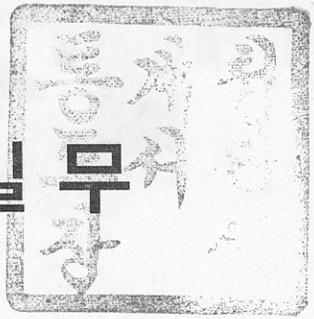


310.11
통140
22

번역자료
91-06-024

통계 조사의 실무

— 집계(제표)편 —



1991. 12.

통 계 청
자료처리국 자료처리과



42083

우리 청에서는 통계자료처리 업무에 많은 노력과 시간을 투입하는데 비하여 업무내용이 명확히 정의되어 있지 않아 관련 기관과의 협조에 어려움이 많고 특히 통계조사를 기획하고 조사하는 부서에서도 자료처리 과정을 올바르게 이해하지 못하여 시행착오를 거치는 사례가 있어 왔습니다.

이에따라 통계조사 실무중에서 자료처리 업무에 대한 내용을 상세히 설명한 자료가 필요하던중 일본에서 최근에 간행된 「통계조사의 실무(집계편)」 책자의 내용이 비교적 우리의 자료처리 환경과 유사한 내용으로 되어있어 이를 번역 발간하게 되었습니다.

이 책자를 통하여 통계조사를 기획하거나 실제 조사를 담당하는 직원들에게 자료처리를 이해시킴은 물론 자료처리를 직접 담당하는 직원들에게도 우리의 자료처리 업무와 비교하여 개선 발전시킬 수 있는 계기가 되기를 바랍니다.

아울러 본 책자를 번역한 전산개발과의 권경환 사무관, 편집, 교정, 전산 입력을 위해 수고한 자료처리과 직원들에게 감사드리며, 본책자 내용에 대한 문의사항이나 자료처리 업무에 대한 좋은 의견은 자료처리과 방운화 사무관에게 연락해 주시기 바랍니다.

1991년 12월

자료처리과장 林東升

머 리 말

통계조사는 조사의 기획·실사부분과 실사후의 제표(집계) 부분의 두가지로 이루어져 있다.

최근에 통계를 둘러싸는 환경이 반드시 좋다고 만은 할 수 없으므로 통계 조사의 실무에 관한 저서가 거의 없다고 해도 좋을 정도이다.

다행히 조사의 기획·실사부분은 (株) 교세이에서 출판되어 (藤田峯三 저 : 통계조사의 실무) 호평을 얻고 있지만, 후자인 제표의 실무는 찾아보기가 힘들다. 이유는 「간단히 순서대로만 계산하면 된다」는 식으로 가볍게 생각했었기 때문이 아닐까 생각한다.

현대는 컴퓨터 시대이다. 제표실무도 인간으로부터 컴퓨터로 옮겨왔다. 이와 같은 시기에 藤田씨의 추천도 있어서 「통계」 잡지에 연재한 글을 정리한 것이 본서이다.

컴퓨터와 관련하여 제표의 실무를 기술한 셈이지만 충분한 설명을 다했다라고는 말하기 어렵다. 독자 여러분의 꾸짖음이 있기를 바란다.

마지막으로 본서의 집필, 간행에 있어서 협력을 아끼지 않으신 많은 관계자 여러분에게 깊이 감사드리는 바이다.

鈴木芳雄

목 차

제 1 장 통계조사의 제표	3
1. 제표부문의 위치	3
2. 수작업부문의 제표개요	4
3. 컴퓨터부문의 제표개요	8
제 2 장 제표업무의 실제	11
1. 제표준비	11
2. 접수·정리	15
3. 내용검사	20
4. 부호기입	24
5. 체크리스트 심사.....	28
6. 결과표 심사	37
7. 제표계획	44
8. 진도관리	58
제 3 장 컴퓨터에 의한 제표	61
1. 컴퓨터 입문	61
2. 컴퓨터에 의한 통계표 작성방법	70
3. 통계표를 위한 기능	74
4. 입력데이터의 작성	76
5. OMR 조사표의 취급	80
6. 데이터 체크	86

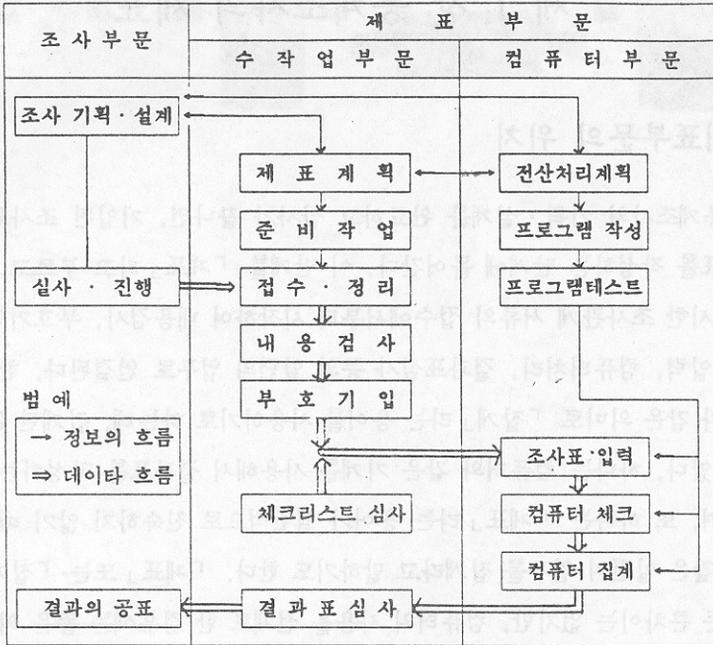
7. 에라데이타의 정정	101
8. 체크요령서 작성	110
9. 집계 프로그램의 구성	120
10. 기타 프로그램의 구성	140
11. 프로그램 개발	147

제 1 장 통계조사의 제표

1. 제표부문의 위치

통계조사의 기획·설계를 완료하고 실사가 끝나면, 기입된 조사표로부터 통계표를 작성하는 단계에 들어간다. 이 단계를 「제표」라고 부르고, 조사표를 위시한 조사관계 서류의 접수에서부터 시작하여 내용검사, 부호기입, 조사표의 입력, 컴퓨터처리, 결과표심사 등의 일련의 업무로 연결된다. 한편 「제표」와 같은 의미로 「집계」라는 용어를 사용하기로 하는데, 집계란 2가지의 의미가 있다. 하나는 컴퓨터와 같은 기계를 사용해서 결과표를 작성하는 과정을 말하며, 또 하나는 「제표」라는 용어가 일반적으로 친숙하지 않기 때문에 제표와 같은 일련의 업무를 집계라고 말하기도 한다. 「제표」또는 「집계」 어느 쪽이든 큰차이는 없지만, 컴퓨터의 사용을 전제로 한 경우에는 좁은 의미의 「집계」인지 넓은 의미의 「집계」인지 혼동하기 쉬우므로, 넓은 의미일 때는 「제표」, 컴퓨터의 처리과정만을 가리킬 때는 「집계」라 한다. 제표 전체과정에서 조사표가 처리되는 순서대로 요약하면 [그림 1]과 같다. 통계조사가 기획되고 실사를 실시한 후, 결과를 공표하기 위해서는 2개의 부문을 연계시켜 진행한다. 즉 「조사부문」과 「제표부문」이다. 제표부문은 다시 수작업 부문과 컴퓨터 부문으로 나누어진다. 예를들면 총무청의 경우에 조사부문은 통계국이, 제표부문은 통계센터가 중심이 되어 추진한다. 한편 조사부문은 기획, 설계, 시험조사, 都道府縣에서의 실사 및 심사 등 많은 업무를 실시하고 있지만, 본 교재의 주제가 아니므로 생략한다.

그림1 부문간 관련도



2. 수작업부문의 제표개요

(1) 제표계획

일반적으로 제표는 컴퓨터 관계부문을 제외하면 간단하고 단순한 작업처럼 생각되지만 제표의 각 업무로 조사표가 지체없이 이동되어 처리되기 위해서는 면밀한 계획과 철저한 진행·작업상황의 관리가 필요하다. 통계조사가 기획·설계부터 시작되듯이 제표도 제표계획으로부터 시작된다. 제표계획은 통계조사의 결과공표가 조사부문에서 희망하는 기간내에 이루어지도록 작성되어야 한다. [그림 1]의 수작업 부문에서는 집계방법, 조사표 매수 등으로 부터 필요한 소요인원을 산출하고, 동시에 제표기간(예정)동안 동원 가능한 인원을 산출하여 가용인원(실제 노동하는 인원)에 대하여 세밀하게 검토하여야 한다. 소요인

원과 가용인원을 비교해서 가용인원이 많으면 그 조사의 제표는 가능하다. 실 제로 제표기간내에 동원한 인원수의 변동에 따라서 컴퓨터 부문의 집계계획을 연계시키는 것은 그렇게 단순한 것은 아니다. 게다가 최근 조사환경의 악화로 조사표 기입의 질이 저하되고 있어 소요인원이 증가하고 있다. 가용인원은 직 원수에 제한되어 있으므로 제표방법을 연구하거나 OA화를 진척시키지 않으면 희망하는 기간내에 제표를 완료할 수 없는 상황이다. 제표계획은 가장 기본적인 인원계획(소요인원, 가용인원)을 중심으로 하여 「제표과정」의 작성, 업무 교육, 제표용품의 조달계획 등을 포함하여 작성한다.

(2) 준비 및 접수정리

조사표류는 조사가 끝나면 都道府縣에서 邦계국으로 송부된다. 송부의 기 일, 방법 등은 조사 전에 지시되며, 그것에 의거하여 제표부문에서 접수가 시 작된다. 여기서부터 조사표를 취급하는 제표실무에 들어가는데, 그전에 조사표 는 취급하지 않더라도 준비업무를 해야 한다. 준비업무는 조사표 접수이후에 제표업무에 필요한 여러가지 재료의 준비나 제표과정의 작성, 관련직원에 대한 교육실시 등의 업무이다. 여기에서의 준비정도가 그후의 제표업무 전반에 많은 영향을 미치며 조사표가 접수된 후에야 제반준비를 하는 것은 시기적으로 너무 늦다. 접수정리는 3가지 분야로 나누어 진다. 첫째는 조사표를 정확히 수령하 는 것이다.(관계서류의 종류와 수량을 포함) 둘째는 제표업무가 원활히 진행되 도록 「묶음」을 만드는 것이다. 「묶음」이라고 하는 것은 처리단위(1인 1일 의 처리량이라든가, 운반의 편리성 등으로 정한다)라고 생각해도 좋다. 셋째는 접수·정리된 조사표 등을 창고에 보관·관리하는 것이다.

(3) 내용심사 및 부호기입

앞 단계의 준비 및 접수정리는 조사관계 서류의 물리적인 정리이다. 여기서는 조사표의 질적인 정리를 실시한다. 조사표의 기입내용이 올바르다면 내용심사를 실시할 필요가 없지만, 실제로 정도의 차는 있을지언정 어느정도 오류가 있다. 이 내용심사는 조사표를 1매씩 수작업으로 점검하는 것이며, 점검의 요점은

- ① 기입을 잊어버리고 있지 않는가?(무기입)
- ② 여분에 기입하고 있지 않는가?(과잉기입)
- ③ 틀린 회답란을 고르고 있지 않은가?
- ④ 조사항목간에 논리적 모순을 일으키고 있지 않는가?
- ⑤ 조사표의 파손, 더러움이 없는가?

등이다. 조사의 종류에 따라 특히 ④의 점검범위가 다르고, 또한 ⑤는 마크시트 방식의 조사표를 채택한 경우에 중요하다. 컴퓨터 집계를 위해서는 수작업에 의한 내용검사와 컴퓨터의 데이터체크를 구분할 필요가 있다. 컴퓨터가 체크할 경우에 ④는 인간보다 정확히 판정할 수 있지만 ①, ②는 종합적으로 판정하는 것이 어렵다. 컴퓨터 체크와 병용할 경우, 사람이 정확한 판단을 할 수 있는 항목이나 기본적으로 중요한 조사항목은 수작업으로 실시하며, 논리적인 체크는 컴퓨터에 맡기는 편이 좋다.

부호기입은 내용검사와 동시에 실시되기도 한다. 부호기입은 컴퓨터 집계를 위해서 빠뜨려서는 안된다. 사전에 조사표를 설계할 때에 제표부문에서 새삼스럽게 부호기입을 하지 않아도 좋도록 회답란을 만들면 동시에 부호화가 되는 것도 있다. 그러나 문자응답 조사항목의 산업란, 직업란은 부호의 종류도 많고 또 그 정의를 이해해서 응답자가 부호를 기입하는 것도 어려우며, 특히 오랜세월의 경험과 광범위한 전문지식이 있는 사람이 실시하지 않으면 잘못된

부호로 기입될 우려가 있다.

이와 같은 문자응답 항목의 부호기입은 제표업무 중에서 많은 인원과 시간을 필요로 하는 부분으로서 제표부문의 취약점이기도 하다.

(4) 심사업무

심사에는 2가지 종류가 있는데, 첫째는 체크리스트의 심사이다. 컴퓨터 체크에 의해 조사표 기입내용의 오류가 지적되면 수작업으로 오류를 정정한다. 오류가 있는 항목에 대해서는 컴퓨터 처리에 의해 오류의 종류가 체크리스트에 기재되며, 그것에 의거하여 해당되는 조사표를 찾아 정정한다. 여기에서 컴퓨터로부터 오류의 지적이 올바른 것인지 주의해야 한다. 마치 눈사태가 작은 원인으로부터 발생하여 점점 크게 되는 것과 같이 오류의 근본을 컴퓨터가 빠뜨림으로서 그 파급결과로 다른 항목에서 오류가 발견되는 일이 종종 있기 때문이다. 따라서 컴퓨터가 지적한 오류부분만을 점검하는 것이 아니라 조사표 전체를 신중히 재검토할 필요가 있다. 체크리스트의 심사에서는 조사표를 정정한 후 재입력을 위하여 정정된 조사표를 컴퓨터부문으로 이동한다. 이동에 필요한 수속, 필요요원, 시간적 지체 등을 고려해 볼때 온라인 방식으로 정정하는 것도 좋은 방법이다. 컴퓨터체크와 오류정정은 수회 반복하여 더이상 정정할 필요가 없으면 컴퓨터 집계에 들어간다.

둘째는 결과표의 심사이다. 컴퓨터 집계가 끝나면 결과표를 출력하여 수작업부문으로 이동한다. 결과표는 한표씩 결과수치 및 양식(체제)이 심사된다. 수치의 심사는 기본적인 「총수」와 「내역」의 합계와 일치관계, 다른 표의 수치와 일치관계, 혹은 대소관계를 우선 실시한다. 이 부분은 논리적인 심사이므로 컴퓨터 심사가 효과적일 수 있다. 그리고 수치를 시계열로 비교, 다른 관련 조사와의 비교 등을 통하여 완벽을 기해야 한다.

체제의 심사는 인쇄문자의 잘림, 굵힘, 자리수 정렬, 행간격의 미비 등 특히 한자 프린터로 인쇄하여 그것을 인쇄용으로 사용하는 경우에는 충분한 심사가 필요하다. 직할시가 있는懸은 없는懸과는 양식이 다르므로 주의를 기울여야 한다. 체제의 심사는 수작업이 제일 좋은 방법이다.

이와 같은 심사가 완료되면 조사부문의 결과표가 송부되고, 그 조사의 제표는 종료된다. 물론懸별로 보고서가 간행되는 경우에는 전국계의 결과자료를 송부하고 난 뒤에 종료된다.

3. 컴퓨터 부문의 제표개요

(1) 전산처리 계획

컴퓨터 부문의 제표도 전산처리 계획으로부터 시작된다. 이것은 제표계획과 병행해서 진행되며, 대개 2가지 계획으로 이루어진다. 즉 컴퓨터의 가동계획과 프로그래밍 계획이다. 컴퓨터의 가동계획은 해당조사의 집계시간을 산출하여 수작업부문의 진행예정에 맞추어서 집계기간내에 그 시간을 배분하는 것으로서 이 계획에 따라서 컴퓨터를 가동시킨다. 프로그래밍 계획은 집계방식의 검토, 프로그래밍 팀의 편성, 프로그래밍 기간, 최종점검의 실시시기 등을 조사표 매수, 결과표 수, 결과표의 복잡함(프로그래밍의 난이도에 관계한다)을 고려하여 작성한다.

(2) 프로그램 작성과 테스트

프로그래밍 작업이 계획대로 진행되는지의 여부는 프로그램의 작성이 순조롭게 이루어 지는가에 달려있다. 일반적으로 프로그램 작성과 같은 지적노동은 작업시간이 길다고 해서 업무가 순조롭게 진행된다고는 볼 수 없다. 특히 알고

리즘을 생각해야 하는 업무에서는 작업자가 갖추고 있는 소질, 능력이 크게 영향을 미친다. 수작업을 포함해서 체포업무가 지체되는 원인은 프로그램의 작성이 늦어졌기 때문인 경우가 많다. 프로그램의 작성은 치밀한 분석과 정리된 코딩이 필요하고, 성격적으로는 「돌다리를 두드리고 건너는 사람」 또는 「남을 보면 도둑이라고 생각한다」는 식의 신중함이나 의심이 많은 사람이 필요하다. 다시 말하면 조사부문과 수작업부문에서의 자료를 신중히 분석하여 그러한 자료를 무조건 받아들이지 않는 것이 필요하다는 의미이다.

프로그램의 완성까지에는 두번의 테스트를 거친다. 하나는 각각의 프로그램을 완성시키기 위한 원시프로그램의 점검이고, 또다른 하나는 본 집계를 가상해서 소량의 실제 조사표를 사용해서 각 프로그램을 연결하여 실행시키는 최종점검이다. 프로그램의 테스트는 숫자의 정리증명과는 다르다. 프로그래머는 여러가지 경우를 가상하고, 테스트 데이터를 통하여 처리를 확인하는 방법으로 테스트한다. 이때 프로그래머는 잘못된 테스트의 추진, 틀린 데이터의 준비, 사실과 다른 결과에도 불구하고 정당이라고 믿어버리는 오류를 범할 수 있다. 이것을 피하는 방법으로 프로그램 작성자 이외의 사람이 테스트를 실시하면 좋다. 그러나, 다른 사람이 작성한 프로그램은 이해하기 어려우므로 특히 중요한 프로그램이외에는 이 방법이 실시되지 않는 경향이 있다.

이와 같은 이유 때문에 최종점검은 프로그래머 개인의 이해가 올바른지를 확인하기 위해서도 필요한 것이다. 그러나 최종점검에서는 가끔 오류를 야기시키는 데이터가 포함될 수 있으므로 이것도 충분하지는 않다.

(3) 집 계

집계를 위한 프로그램은 봉계표의 작성을 수작업으로 작성할 때와 거의 유사한데, 기본적으로는 5종류가 된다. 「Garbage in Garbage out」 즉 「입력이

쓰레기이면 출력도 쓰레기이다」라는 격언대로 바른 데이터를 사용하여 집계하지 않으면 바른 결과표는 얻을 수 없다. 따라서 컴퓨터집계에서는 먼저 데이터의 오류점검을 수행하는 프로그램이 필요하다. 올바른 데이터가 준비되면 도수분포를 실시하는 프로그램, 지역계 등을 산출하는 프로그램, 구성비·백분율 등을 계산하는 프로그램, 집계된 수치를 결과표의 체제로 조정하는 프로그램 등을 단계적으로 연계시키면 된다.

집계는 인간의 개입을 배제하여 자동화되고는 있지만, 체크프로그램은 예외이다. 데이터에 오류가 있으면 전술한 것과 같이 체크리스트가 수작업 부문으로 송부되어 오류가 정정될 때까지 중단된다. 여기에서의 중단이라는 것은 집계 프로그램의 단계에서 일시 중단되는 것이며, 체크프로그램의 연산 도중에 일시 중단되는 것은 아니다. 오류의 정정과 체크프로그램의 연산이 여러번 반복됨으로서 오류 데이터는 통계이용에 지장이 없을 정도로 정정되고 또한 컴퓨터에 의해 자동정정된 후 집계 프로그램으로 이동한다. 집계 프로그램에 의해 작성된 결과표를 출력하여 수작업부문에 송부함으로써 컴퓨터 부문의 제표업무는 끝난다.

제 2 장 제표업무의 실제

1. 제표준비

우선 제표계획에 대하여 먼저 설명해야겠지만 제표 각 단계에 대한 업무내용을 이해한 후에 제표계획의 내용을 포착하는 것이 쉬으므로 본 장의 마지막에 설명하기로 하고, 먼저 실무를 위한 준비과정에 대하여 언급하겠다. 앞서서도 언급한 바와 같이 조사관계 서류가 송부되고 나서 급히 필요한 서류를 작성하는 것은 늦다. 송부되기 전에 각종 사무용품류나 서류를 준비하여 미리 기입을 할 수 있도록 하여야 한다. 더우기 「제표업무처리절차」의 작성, 「참고서류」의 수집, 「업무교육」의 실시 등도 이 준비과정의 범주에 속하는 것이지만, 이들은 나중에 언급하기로 하고 여기서는 다음단계인 접수·정리에 필요한 사항에 대해서만 설명한다. 조사규모에 따라 결과표의 작성 규모도 달라진다. 하나의 예로서 조사표에 마크시트 방식을 채택한 중규모 정도의 조사에서 준비과정 내용을 소개하면 다음과 같다.

우선 작성할 서류는 이하의 (1)~(22)이다.

- (1) 都道府縣 요계표 묶음표지
- (2) 市區町村 요계표 묶음표지
- (3) 조사구 요계표 묶음표지
- (4) 세대명부 묶음표지
- (5) 조사구요도 묶음표지
- (6) 영문조사표·조사표(부계 세대용) 묶음표지
- (7) 조사표 표지(무인 조사구용)
- (8) 조사구 아이덴트 케이스 대장

- (9) 조사구 아이덴트 케이스
- (10) 조사구 아이덴트
- (11) 조사 관계서류 제출상황 일람표
- (12) 조사 관계서류 접수 기록표
- (13) 무인조사구 기록표
- (14) 추출을 불일치 조사구 일람표
- (15) 접수정리 질의 처리표
- (16) 영문조사표·부재 세대용조사표등 사용매수 점검
- (17) 불량조사표·미사용 조사표 매수점검
- (18) 관계서류 가(임시) 수령서
- (19) 관계서류 수령서
- (20) 불량조사표 매수기록표
- (21) 세대수 및 세대인원 증감표
- (22) 질의표

이들 서류의 내용은 대강 다음과 같은 4종류로 대별된다.

- ㉠ 조사관계 서류는 조사기획시에 양식이 정해지고 이들 서류가 접수 되면 각각 정리·보관되어 참조되는 것으로서, 그를 위한 표지가 필요하다. 예를 들면 都道府縣 요계표, 세대명부 등은 기획시에 소정양식을 정하고 접수시에 따로따로 정리하여 서류마다 별도의 표지가 만들어진다. (1)~(7)
- ㉡ 결과표 작성과정에서 필요한 서류지만, 都道府縣의 실사단계에서는 사용되지 않는것으로서, 예를 들면 (10)조사구 아이덴트는 조사표를 수령하고 나서 필요한 것으로 이것은 조사표를 조사구 단위로 구분하기 위한 특별한 마크시트이다.

㉔ 결과표 작성과정에서 반드시 필요한 것은 아니지만, 장래에 참고하기 위하여 작성하는 것으로서, 예를 들면 (17) 불량조사표, 미사용 조사표의 매수 점검은 집계에 필요가 없다. 그러나 불량조사표의 양은 그 이유와 함께 현상의 분석과 개선을 위하여 필요하다.

㉕ 조사관계 서류의 종류와 수량 및 수령을 증명하는 것으로서, 서류가 발견되지 않을 때 혹은 서류의 매수, 상자수에 의문이 생길때에 송부 상태를 확인하기 위해서 수령하고 본실한 것인가를 명확히 하기 위해 필요하다. ((11, 12, 18, 19) 등)

이상의 서류에 미리 기입할 수 있는 항목은 미리 기입을 하는데 일반적으로 기입할 수 있는 항목은 都道府縣의 명칭과 그 번호 혹은 묶음번호 정도이다. 기입에는 검은색 볼펜을 사용하는 것으로 통일하는 것이 좋다. 무슨 필기구를 쓰든지 상관없을 것 같지만 제표업무의 각 단계에서는 서류에의 기입도 많고,

표1 묶음작성

묶음표지의 종류	마분지의 규격	매수 (묶음수)	기 입 사 항
都道府縣 요계표	B4	1	都道府縣번호란에 「01-47」로 기입한다
市區町村 요계표 (접수정리부)	B4	47	都道府縣별로 都道府縣명의 고무인을 찍는다. 지역별 표본조사구수 일람표에 의해 조사구수를 기입한다.
조사구 요계표 세대 명부 조사구 요도	B5 B4 B4	58 58 58	都道府縣별로 都道府縣명의 고무인을 찍고, 조사구입력번호 및 총 묶음수에 대한 묶음번호를 기입한다. (주) 북해도, 기옥, 천엽, 동경, 신금천, 정강, 애지, 대관, 병고, 광도 및 북강의 11都道府縣에 대해서 각각 2묶음씩 작성한다.
영문조사표 조사표 (부재세대용)	B4	47	都道府縣별로 都道府縣명의 고무인을 찍는다.

또 기입의 오류를 정정하는 일도 많다. 이때 필기구나 잉크의 색이 각기 다른 면 쓸데없는 판단을 하는 경우도 있고 또한 문제가 발생했을때 원인규명의 장

애가 되기도 한다. 都道府縣 요계표나 세대명부 등의 서류를 철하는 경우 종이
 가 얇으므로 그 서류와 같은 치수의 마분지 2매로 「표지」, 「속표지」를 만
 들고, 샌드위치와 같이 서류를 끼워서 만든다. 이때 표지에는 (1)~(5)와 같은
 묶음표지를 풀로 붙이고, 묶음을 묶는 것에 면 테이프를 사용하면 좋다. 묶음
 수는 각각의 서류의 매수와 1일의 처리량 혹은 운반의 용이성으로 결정되지만
 묶음수의 예를 들면 [표 1]과 같다.

都道府縣 요계표는 각 縣마다 1매 정도이므로 전체 縣을 하나의 묶음으로
 한다. 따라서 묶음표지는 1매만 만들면 된다. 조사구 요계표, 세대명부, 조사
 구 요도는 각 縣을 하나의 묶음을 원칙으로 하지만 (注)에 있는 11 都道府縣은
 이들 서류의 매수가 많으므로 2개의 묶음으로 나누어서 합계 58매의 묶음표지
 를 만든다. 조사구 요계표의 묶음표지의 예가 [그림 2]에 있다. 준비업무에
 서의 요점을 정리하면 다음과 같다.

- ① 조사관계 서류의 종류별로 묶음표지 등의 서류를 작성한다.
- ② 묶음수는 각 조사관계 서류의 매수, 제표업무에서의 처리량, 운반의 용
 이성 등으로 결정된다.
- ③ 미리 기입할 수 있는 항목은 기입해 둔다.

그림2 조사구요계표 묶음표지

○ ○ 기본조사		
조 사 구 요 계 표		
都道府縣명		
조 사 구 일련번호	~	
총 매 수		매
묶음번호	묶음중의	묶음

2. 접수·정리

都道府縣에서의 조사관계 서류의 정리방법은 조사시에 결정이 되어 있다. 그 내용은 대략 다음과 같다.

- ① 조사를 종료한 조사표는 「세대번호」 순서대로 정돈하여 조사구마다 「표지」, 「속표지」를 붙이고, 조사표정리 주머니(비닐 주머니)에 넣어 「조사구 일련번호」 순서대로 배열한다.
- ② 조사구 요계표, 세대번호, 조사구요도는 각각 「조사구 일련번호」 순서대로 쌓아올려 50조사구마다 소정의 부분을 철끈으로 맨다.
- ③ 市區町村 요계표는 市區町村 번호순서대로 정돈하여 그위에 都道府縣 요계표를 위에 놓고 소정의 부분을 철끈으로 맨다.
- ④ 조사관계 서류를 「수송상자」에 넣을 때는 「조사표 등 조사관계 서류제출명세서」(제출명세서라고 생략함)를 최상부로 하고 이하 「都道府縣 요계표」, 「市區町村 요계표」, 「조사구 요계표」, 「세대명부」 및 「조사구요도」의 순서대로 조사표와는 별도로 일괄해서 첫번째 「수송상자」에 넣는다.
- ⑤ 조사표는 「조사표 상자」의 일련번호의 순서대로 「수송상자」에 넣는다.
- ⑥ 조사표(부계세대용), 조사원·지도원·市區町村·都道府縣 또는 세대에서 재기입하여 불필요한 조사표(불량조사표), 영문조사표 및 남은 조사표(미사용 조사표) 또 市區町村에서 개봉한 조사표, 제출용 봉투 및 조사표 봉투(부계 세대용)는 조사표 등 조사관계 서류를 제출할때 합해서 「수송상자」에 넣는다.
- ⑦ 「수송상자」는 철도편 또는 트럭편으로 수송한다.

이와 같이 都道府縣측에서 수송상자에 넣기까지의 순서가 정해져 있고 이

것을 제표부문에서 포장을 푸는 것에서부터 접수·정리에 들어가는 셈이다. 수송의 수단은 전기⑦이외로 항공편이나 휴대편 등도 사용되고 있지만 이들은 특수한 경우에 해당한다. 경상조사 혹은 소규모 조사에서 조사표가 적고, 또한 마감이 임박해 있는 경우에 지리적으로 먼 지역에서는 항공편을 가까운 지역에서는 직원이 지참하는 방법을 사용한다.

그림3 양식에

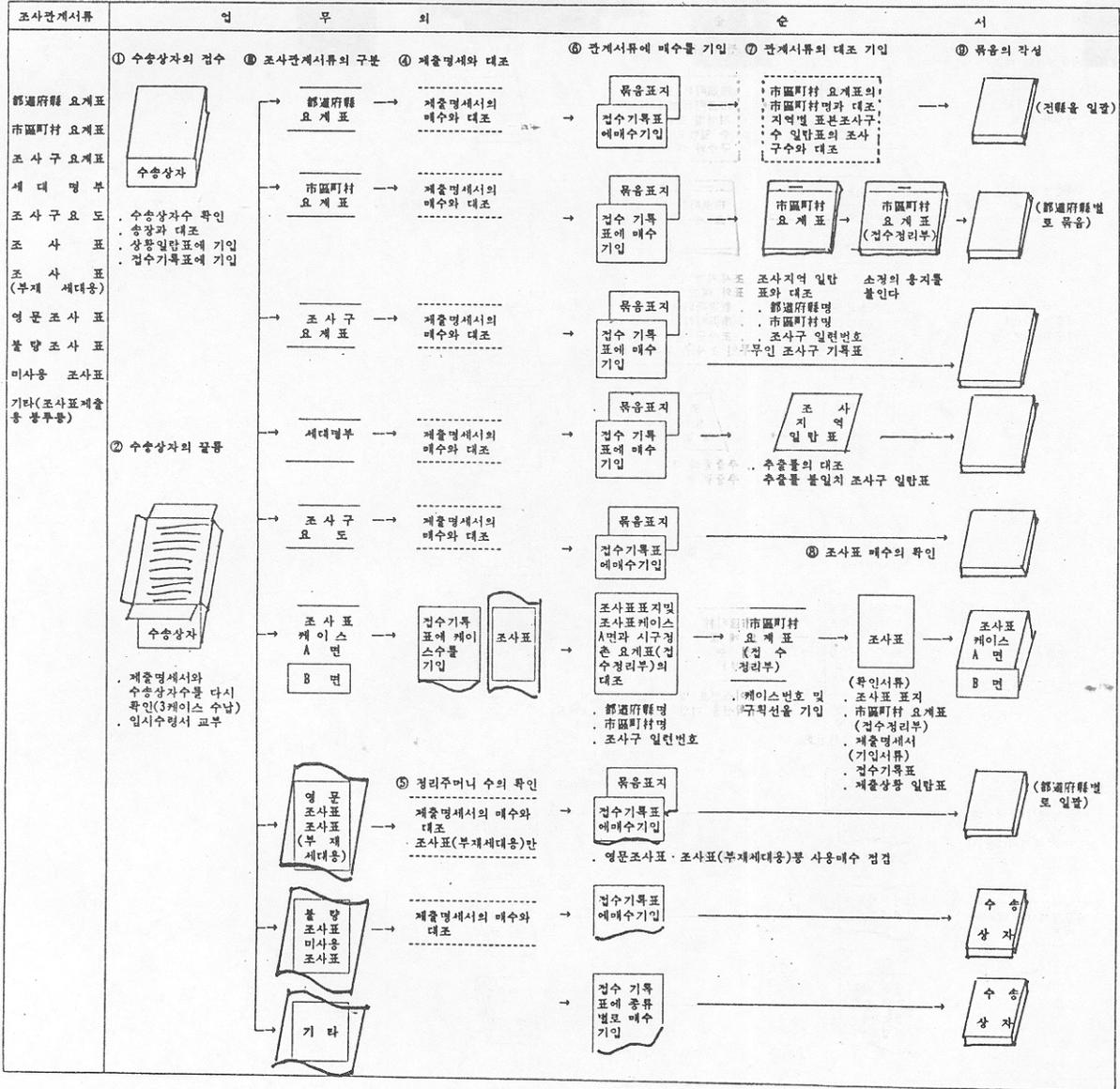
○ ○ 기 본 조 사

조사관계서류 접수기록표

	조사구수	운송상자 都道府縣	요계표 市區町村	요계표 조사구요계표	세대명부 조사구요도	조사표케이스	조사표매수	조사표대용 조사표대용V	봉투 부재세대용	봉투 조사표제출용	영문조사표	불량조사표	미사용조사표	비고
전 국														
01	북해도													
02	청삼현													
03	암수현													
04	궁성현													
05	추전현													
· · ·														
47	충니현													

都道府縣으로부터 온 「송장」과 수송상자의 수량을 점검한 후 첫번째 상자의 수송상자를 풀고, 제출명세서와 수송상자에 들어 있는 각 서류, 조사표 상자수와 대조한다. 그리고 기재된 대로 수령하였으면, 준비업무에서 준비한 서류 (11), (12)에 소정사항을 기입한다. 서류 (11), (12)의 소정사항에는 (11)에 都道府縣별로 「도착 월, 일」, 「수송방법」, 「조사표 매수」, (12)에 都道府縣별로 「수송상자수」, 관계서류마다의 「매수」를 기입하게 되

그림4 접수·정리업무의 프로우 차트



어 있다. 참고로 (12)의 양식은 [그림 3]과 같다. 매수를 세는데는 시간이 걸리므로 수송상자수를 확인한 후 임시수령서(서류 (18))을 해당 都道府縣에 교부한다.

조사표는 마크시트 방식의 경우 「표지」, 「속표지」(이들은 조사표를 보호하기 위해)를 붙여서 비닐주머니속에 넣어진 채로 조사표 상자에 있기 때문에 조사표 상자수 및 비닐주머니수의 대조를 실시한 후(都道府縣, 市區町村, 조사구 번호의 체크는 실시한다), 조사표를 비닐주머니에서 꺼내서 매수를 확인한다. 조사표는 「표지」, 「속표지」를 붙여서 본래의 조사표 상자로 되돌린다.

이와 같이하여 모든 관계서류의 대조, 수량의 파악이 끝나면 본수령서(서류 (19))를 해당 都道府縣에 교부한다. 접수·정리업무를 플로우차트로 나타내면 [그림 4]와 같이 되고, 각각의 서류마다 분류된 매수의 대조, 묶음의 작성으로 진행하여 정리가 끝난다. 정리가 끝난 조사표 및 관계서류는 창고에 수납되며 수송상자등 짐꾸러미 재료는 폐기된다. 이때 짐꾸러미를 풀고, 마무리 정리하는 작업은 제한된 장소에서 행하여져야 하며, 소량의 서류라도 폐기물 쪽으로 들어가는 일이 없도록 주의해야 한다.

접수·정리의 요점은

- ㉠ 접수시에 수송상자수를 확인하여 기록한다.
- ㉡ 각 조사단계 서류마다 매수가 맞는지 확인하여 수량을 기록한다.
- ㉢ 각 조사관계 서류를 묶음으로 분할하여 정리한다.
- ㉣ 창고에 보관한다.

와 같이 되고, 특히 여기에서 파악된 수량이 이하의 제표업무의 기본이 되므로

엄밀히 실시할 필요가 있다.

3. 내용 검사

제표업무 가운데 내용검사와 다음에 나오는 부호기입은 2가지의 의미에서 중요한 업무이다. 하나는 완성품인 결과표의 품질을 좌우하며 또 하나는 이 업무에 많은 인원이 필요하며 처리기간이 길다는 것이다.

첫째, 품질의 저하를 초래하는 문제인데 품질저하 요인은 3단계로 나누어 볼 수 있다. 제 1단계는 조사의 기획·설계단계에서, 예를 들면 모집단의 선정, 조사객체의 추출 등에서 발생하는 요인, 제 2단계는 실시시점에서 객체의 파악 누락, 기입자의 거부 혹은 오기입과 같은 요인, 제 3단계는 제표에서의 부호기입의 오류나 키엔트리시의 오류와 같은 요인이다. 각각의 마이너스 요인에 대해서는 그 나름대로의 대책을 강구하고 있지만 전부 제거하는 것은 어렵기 때문에, 제표단계 조사표에는 실사의 오류가 항상 포함되어 있다고 보아야 한다. 이러한 오류의 발견과 정정으로 실시시의 마이너스 요인을 조금이라도 줄이고, 정확한 부호기입을 할 수 있도록 하는 것이 내용검사의 목적이다.

둘째는 요원수의 문제에 대해서는 통계국 조사의 경우에는 일반적으로 조사표의 매수가 많은 점이다. 내용검사 및 부호기입은 조사표 1매를 대상으로 실시하기 때문에 조사표의 매수에 비례하여 작업자 수와 기간이 필요하다. 게다가 이 작업에는 전문지식이 요구되므로 누구나 할 수 있는 것은 아니다. 예를 들면 산업분류, 직업분류의 부호기입 및 검사는 최근과 같이 첨단기술산업과 같은 새로운 직업이 증가하여 그 방면의 지식이 없다면 실태 파악을 할 수 없고, 내용검사 등의 업무를 수행할 수 없다. 내용검사와 부호기입이 과거에는 독립된 업무였다. 물론 업무내용에는 차이가 있다. 전자는 조사표의 기입에 오

류가 있는지를 조사하는 것이고, 후자는 올바른 조사표를 전제로 부호를 기입하는 것이다. 여기에서 독립이라고 하는 것은 모든 조사표에 대해서 내용검사를 완료한 후 부호기입에 들어간다는 의미이다. 이 경우에는 내용검사를 해야 할 조사표를 꺼내 검사를 끝낸 후 본래대로 되돌리고, 부호기입시에 다시 조사표를 꺼내 부호를 기입한 후 본래대로 되돌리는 식으로 조사표를 「꺼내 본래대로 되돌린다」는 작업을 중복하고 있다. 과거에 이와같은 중복을 알면서 독립해서 작업을 처리한 것은, 컴퓨터 이용 이전의 내용검사는 조사표의 전체 기입사항에 대해서 기입이 누락되었는지의 간단한 검사에서부터 다른 조사사항과의 복잡한 관련검사에 이르기까지 한 사람이 수행하기 위해 기억해야 할 검사의 순서, 규칙이 많았기 때문에 작업량이 많아 한 사람이 부호기입까지 실시할 여유가 없었던 것 같다.

현재는 모든 조사에 컴퓨터가 사용되어 내용검사도 인간과 컴퓨터로 분업하고 있다. 즉 타 서류와의 대조 등과 같은 기본적인 사항의 검사는 인간이 하고, 전 항목의 논리적·관련적 검사는 컴퓨터에 의해 이루어진다. 그렇게 함으로써 내용검사 부담이 경감되어 부호기입까지 할 수 있게 되었다고 필자는 생각한다. [그림 5]는 조사표로 마크시트를 사용한 경우에 내용검사와 부호기입을 동시에 실시한 경우의 플로우차트이다. 조사표의 확인은 접수·정리업무에서 이미 실시되어 있지만, 조사구내의 개개의 조사표에 대해서는 다음과 같은 내용을 확인한다.

- ① 조사표 상자 A면과, 조사표 표지에 기재한 都道府縣명, 市區町村명 및 조사구 일련번호가 일치하는지 확인한다.
- ② 조사구별로 조사표 표지와 조사표의 조사구 일련번호가 일치하는지 확인한다.
- ③ 개개 조사표의 조사구 일련번호 기입누락, 기입오류가 있었던 경우는

市區町村 요계표, 세대명부, 전후의 조사표 및 조사표 표지의 조사구 일련번호 등을 참조하여 정정한다.

- ④ 세대주 조사표에 기재되어 있는 세대번호를 세대명부와 대조하고, 다른 세대원에 대해서는 세대주 세대번호와 대조한다.

확인이 끝나면 조사표 각란의 내용검사를 실시한다. 이를 위하여 내용검사를 위한 「내용검사 과정」이 작성되고, 그것에 따라서 작업이 실시되지만 그 「과정」에는 조사표 각란의 검사에 대한 원칙이 기술되고 또 각란의 검사순서도 포함되어 있다. 예를 들면 어느 조사에 대한 검사원칙의 예로서 「조사표의 각 사항에 대해서 마크나 문자, 숫자의 기입누락, 불량마크(마크가 테두리에서 비어져 나오는 등) 또는 마크가 2개소 이상 있는 경우에는 수정보완을 중점적으로 실시한다. 더구나 미기입 등으로 관련된 항목 및 관계자료를 참조하더라도 처리하기 어려운 경우에는 상사의 지시를 받는것으로 되어 있는데, 마크시트가 아닌 일반 조사표의 경우는 거의 같고, 다른점이라면 마크시트를 채택하고 있는 경우에 마크에 대한 수정보완이 첨가된다는 것이다.

조사표 각란의 내용검사는 각 조사항목이 바르게 기입되어 있는지, 관련된 조사항목간에 모순은 없는지, 오류가 있는 경우의 정정하는 것 등 3가지의 작업을 실시하지만 컴퓨터에 의한 체크와 분담하여 기본적인 항목에 한정해서 실시하는 일이다. 인구·세대에 관한 조사를 예로 들면 기본적 항목으로서 「성명」, 「남녀별」, 「세대주와의 관계」, 「출생년월」, 「배우자의 유무」등이지만 그외에 세대번호와 하나의 세대가 2매 이상의 조사표로 나누어질 경우 연결번호도 내용검사의 대상으로 첨가해 두는 편이 좋다. 그 이유로서 컴퓨터체크에서 세대구별은 세대번호의 차이를 중심으로 점검되는데, 그 이유는 번호의 차이가 세대를 구분하기 때문이다. 즉 세대번호가 같으면 별도 세대라도

같은 세대로 판단한다. 그렇기 때문에 이 두 경우에 대해서도 2중 3중의 점검을 실시하며, 예를 들면 조사구내의 세대수를 알고 있는 경우는 그 수와 일치하는지 또 세대주와의 관계에서 동일세대에 세대주가 2인 경우 혹은 세대주가 없는 경우에 의문 조사표로서 컴퓨터에 경고할 수는 있지만 오류라고 확정할 수는 없다. 그러므로 수작업에 의한 내용검사에서 확인해 두는 편이 확실하다. 같은 이유로 앞에서 예를 든 기본적 항목은 컴퓨터도 체크하지만 정정하는 것은 곤란하므로 인간이 검사하는 편이 좋다. 기본적 항목의 기입누락, 기입오류는 그 상황에 따라 정정해야 할 내용이 다르므로 내용검사 과정에서는 "종합적으로 판단"하여 수정보완해야 한다. 예를 들면 「남녀별」이 기입누락 될 경우, 우선 「성명」을 본다. 분명히 남녀를 판단할 수 있을 것 같아도 「배우자의 유무」의 정보에서 다른 세대원 사이에서 모순이 없는가를 확인한다. 「배우자의 유무」에서 정보가 얻어지지 않을 때에는 본인의 다른 조사항목(산업, 직업 등) 혹은 세대를 구성하는 각인의 조사항목에서 남, 녀 어느쪽으로 하여야 가장 모순이 적은가를 고려해서 결정한다. 「성명」에서 남녀별을 결정할 수 없는 경우도 위와 같은 순서로 된다. 조금이라도 많은 정보를 얻기 위하여 관련이 있는 항목을 점검, 수정보완하는 작업은 조사의 구조정의의 이해뿐만 아니라 경험에서 얻어지는 노우하우와 조사관련의 전문지식을 필요로 하고 이것을 "종합적으로 판단"이라고 표현하고 있다.

4. 부호기입

이상과 같이 내용검사가 끝나면 부호기입에 들어간다. 부호기입에는 3종류가 있다. 첫번째는 조사표의 회답선택란 중 하나에 동그라미를 치거나 마크하는 곳에 기입자가 표시하는것, 두번째는 근무지, 통학지의 縣名, 市區町村명과

같이 기입된 회답과 같은것을 부호표에서 찾아 소정란에 기입하는것, 세번째는 산업분류, 직업분류와 같이 문자기입된 회답을 다른 조사항목을 참조하면서 분류부호표에 있는 부호로 기입하는 것이다.

첫번째의 기입법에서 데이터 엔트리 방식으로 입력데이터를 작성하는 경우에 약간의 문제점이 있다. 그것은 조사표의 인쇄색과 필기구의 색이 같으면 키오퍼레이타가 회답선택란을 찾기 어렵다는 점과, 회답지를 뒤쫓아가 시선이 끊임없이 움직여 키엔트리하기 어렵다는 점이다. 그래서 과거에는 회답란을 별도로 마련하여 거기에 기입시키는 것도 있었지만 현재로는 조사표의 인쇄색을 필기구의 색과 다른색으로 바꾸고, 조사표의 설계시에 키오퍼레이타의 시선이 크게 움직이지 않도록, 즉 위에서 아래로 좌에서 우로 움직이도록 고려함으로써 해결하고 있다.

마크시트방식 조사표의 경우는 조사표의 인쇄색과 필기구의 색과는 당연히 다르게 되어있고 키오퍼레이타가 회답선택란을 찾아서 엔트리하는 의미가 아니기 때문에 오히려 적극적으로 채용해야할 방법이다.

두번째의 기입법은 단순한 이기에 해당하는 것과 많은 자료를 참조하는 것으로 나누어진다. 근무지·통학지란은 전국의 市區町村명이 기입되지만 부호표에 의해 해당하는 코드를 찾을 뿐이므로 수가 많아도 문제는 없다. 문제가 된다고 하면 많은 자료를 참조하는 소비관계 조사의 품목에서 지역에 따라 같은 물건의 이름이 다른 경우와 신제품의 등장이다. 전자의 예로서 관동에서 「사쓰마아게」(어육을 갈아서 당근 우영 등을 섞어 기름에 튀긴음식)가 관서에서는 「텐뿌라」로 불리며, 또 생선도 지역에 따라 여러가지 이름으로 불리고 있다. 이것도 부호표에 각지에서 불리고 있는 이름이나 신제품을 첨가하면 기본적으로 해결 할 수 있지만 각지에서 불리고 있는 이름이나 신제품을 얼마만큼 포함할 수 있을 지가 문제가 된다.

세번째의 기입법은 전문적 지식을 필요로 한다. 산업분류, 직업분류가 대표적인 것이다. 산업분류의 기입은 사업체의 경제활동에 의해 실시된다. 그 사업체가 복수의 경제활동을 하고 있는 경우에는 주요한 경제활동에 의해 결정된다. 직업분류의 기입은 개인이 종사하고 있는 업무에 의해 결정되지만, 이것도 복수의 업무를 하고 있는 경우에는 주된 업무에 의해 결정된다. 이상은 산업, 직업의 결정원칙이다.

산업분류, 직업분류는 대분류, 중분류, 소분류의 3단계로 구성되며, 각 단계의 항목수는 [표 2], [표 3]과 같다. 실제의 기입은 조사표의

표2 산업분류(국세조사)

대 분 류	중 분 류	소 분 류
A 농업	1	2
B 임업	1	1
C 광업	2	2
D 석유업	1	4
E 건설업	1	1
F 제조업	22	83
G 전기·기계·화학·금속·정유업	1	4
H 섬유·정수업	8	9
I 도자기·식품업	9	41
J 목재·가구업	1	3
K 유리·도자기·금속·기계·화학·정유업	1	3
L 기타 제조업	24	53
M 건설업	2	2
N 운수업	1	1
(계) 14	75	210

표3 직업분류(국세조사)

대 분 류	중 분 류	소 분 류
A 전관리직	11	53
B 사무직	3	7
C 사자	4	9
D 판매직	2	13
E 농·축·수산업자	3	14
F 운수·통신·서비스업자	2	8
G 기술직	5	16
H 장·공정·작업자 및 노무직업자	22	148
I 농·림·수산업자	1	5
J 서비스업자	3	19
K 기타직	1	1
(계) 11	57	296

「근무처·업주 등의 명칭」, 「근무처·업주 등의 사업의 종류」, 「본인 업무의 종류」, 「종사상의 지위」 등으로 산업분류, 직업분류를 기입한다.

근무처의 명칭, 사업의 종류, 업무의 종류는 문자 회답란이므로 분류기호표에 게재되어 있는 용어가 기입된다는 보장은 없다. 오히려 「사무원」이라는 식으로 어떤 종류의 사무에 종사하고 있는지 이해할 수 없는 기입이 많다.

또 최근에는 하이테크(High Technology), 바이테크(Bitechnology) 서비스 경제화의 진전, FA화(공장자동화) OA(사무자동화)화 등에 수반되어 새로운 산업, 직업이 증가하고 있다. 조사표의 기입도 가다까나(일본문자) 알파벳이 많이 사용되고, 다양한 의미를 포함하고 있다. 예를 들면 피트니스(Fitness)사업이라든가 코디네이터(Coordinator)만 하더라도 관련된 부분은 인테리어(실내장식), 패션, 아트, 쿠킹(요리), 식탁, 음식, 야채, 식물(Botanical) 컬러(색채) 등으로 다양하다.

FA화의 진전은 제조공정을 자동화하고 산뜻하게 함과 동시에 「00공」이라고 하는 이미지를 불식해 버렸다. 한사람이 몇개의 업무를 수행하여 분류기입이 더욱 어렵게 되고 있다. 이때문에 필요한 「분류의 기준」, 「분류부호표」의 기본적 자료이외에 「상품사진」, 호칭을 수록한 「분류 예시집」, 기업활동의 참고가 되는 「회사총람」 등의 참고자료도 정비해 두지 않으면 안된다. 참고자료를 어디까지 망라할 수 있는지 여부에 따라 기입 업무가 쉬워진다.

분류기입된 조사표는 그 기입이 올바른지 어떤지 검사를 한다. 이 경우는 모든 조사표에 대해서 실시하는 것이 기본이지만, 집계기간 등의 관계에 비추어 일정한 비율로 추출하여 검사를 하고, 그 결과를 기입단계로 Feedback하여 기입의 정도 유지에 노력하여야 한다. 이와 같이 세번째의 기입법은 전문지식과 많은 자료를 필요로 하기 때문에 평소부터 그 지식의 향상을 위해 노력할

필요가 있다. 부호기입이 끝나면 조사표 상자에 넣어서 컴퓨터처리시까지 창고에 보관한다. 마크시트 방식의 조사표의 경우에는 조사구 단위로 조사표를 구분하기 위해 조사구 아이덴트(마크시트)에 그 조사구의 세대수 조사표 매수를 이기(마크한다)하고 조사표 상자에 넣어서 창고에 보관한다.

5. 체크리스트 심사

(1) 체크의 종류

수작업으로 집계한다고 가정해 보자. 예를 들면 「남녀별 인구수」를 구하고 싶을때 조사표의 「남녀별」의 회답에 따라 남자의 조사표와 여자의 조사표 2개로 분류, 각각의 매수를 세면 「남녀별 인구수」가 구해진다.

여기에서 중요한 것은 「남녀별」란은 "남"인가 "여"의 회답밖에 없다고 하는 것이다. 만약 남녀불명(기입누락 등으로)이 있으면 아마 2개로 분류하고 있는 과정에서 발견되어 분류 작업을 일시중단하고 적정한 정정을 한후에 분류 작업을 재개할 것이다. 다음에 컴퓨터로 「남녀별 인구수」를 집계한다면 「남녀별」이란 부호에 의해 "남", "여"를 판단하고 그 매수를 카운트해서 조사표 전부를 카운트하면 인구수가 산출된다. 만약 처리 도중에 남녀 불명이 있으면 컴퓨터를 정지시켜 잘못된 조사표를 찾아내어 정정하는 어리석은 처리를 하게 된다. 그러므로 컴퓨터로 집계하는 경우에는 수작업에 의한 내용 검사가 충분히 실시되었다고 하지만 만일의 간과나 키엔트리의 오류, 마크시트 독취기의 오류 등이 일어날 수 있으므로 데이터를 정비하는 데이터체크가 최초의 처리과정이다. 최근에는 데이터체크의 기능을 적극적으로 이용하여 수작업으로 기본적인 심사로하고, 컴퓨터는 복잡한 체크나 관련항목의 체크를 하도록 한다.

데이터체크는 조사표의 내용을 기록한 자기 테이프의 데이터를 「00조사 체크요령」, (체크해야 할 항목, 순서 등을 나타낸 것)에 따라 컴퓨터로 체크를 실시하며, 오류가 있는 데이터는 오류의 종류를 나타내는 기호를 붙여 프린트 아웃(체크리스트) 하고, 올바른 데이터는 체크리스트에 인쇄하지 않고 자기 테이프에 기록한다. 이 데이터체크를 여러번 반복하고 체크리스트 상에 여러 데이터가 프린트되지 않을때 데이터체크가 종료된다. 체크요령의 예가 [그림 6]에 있다. 이와같은 체크요령은 조사의 구조, 조사표의 기입방법, 결과표에서 조사항목의 크로스 상황, 정의 등을 정리하여 작성한다.

그림6 체크 요령의 일부

번호	체크 항목	체크의 내용			기호
32	경영조직 × 본소·지소별 × 회사에 대하여 (1) 자본금액	○ 경영조직은, 「2 주식회사」 「3 유한회사」 「4 합명·합자·상호회사」 상기이외	○ 본소·지소별은, 「3 지소·지사·지점」 상기이외	○ 자본금액은, 블랙 상기이외 01 ~ 10 상기이외 블랙 상기이외	J1 J2 J3

번호

- 체크의 결과를 나타낸다. ○ 오류아님
× 오류이다 △ 확인할 필요없다
- 경영조직이 「2 주식회사」, 「3 유한회사」 및 「4 합명·합자·상호회사」의 경우, 이 흐름에 따라 다음 항목의 체크(이 경우 「본소·지소별」로 나간다.
- 경영조직이 「2 주식회사」, 「3 유한회사」 및 「4 합명·합자·상호회사」 이외의 경우, 이 흐름에 따라 다음 항목 체크(이 경우 「자본금액」)로 나간다.

체크의 종류는 다음과 같다.

a) 오프코드 체크

조사표의 내용을 항목단위로 그 항목의 코드가 규정코드인가 검사하는 체크이다. 예를 들어 「남녀별」란은 봉사 남은 "1" 여는 "2"라고 코드가 붙여 있으므로 "1", "2" 이외의 코드는 규정외의 코드로 예러가 된다. 규정외의 코드를 오프코드라고 부른다.

b) 관련항목 체크

관련된 항목간에 내용의 모순이나 불합리한 점은 없는가를 검사하기 위한 체크이다. 예를 들면 「출생년월」과 「교육정도」라고 하는 2개의 항목간에서 교육정도가 "대학졸업"이라고 코드되어 있어도 연령이 너무 어린 경우에는 불합리하다는 것이다.

여기에서 주의해야할 것은 각각의 항목이 규정코드로 예러는 아니지만, 조합해서 그 관계를 보면 불합리적일 수도 있다. 그러므로 어느쪽의 항목이 잘못인가는 조사표에서 조금이라도 관계있는 다른 항목을 점검함으로써 알 수 있다.

c) 범위(Range) 체크

오프코드 체크의 변형이다. 오프코드 체크가 규정코드를 가진 항목에 적용되는 것이라면, 범위체크는 금액, 수량 등과 같이 어떤 폭을 가진 항목에 대해서 적용하는 것이며, 그 값이 맞는가 예러인가를 명확히 판정할 수 없는 것이 많다. 예를 들면 생선의 가격도 지역에 따라 계절에 따라 불균형이 있다. 따라서 상식적인 고가, 염가 범위의 적당한 값으로 하고, 그것보다 너무 비싸거나 너무 싼 값의 경우를 예러로 하

고 있다. 고가, 염가 쪽의 가운데서 체크하는 것이므로 범위체크라고 한다. 이 범위체크에서의 문제점은 한계를 나타내는 고가, 염가의 결정 방법과 이 한계를 넘는 데이터가 반드시 에러라고 판정할 수 없다는 것이다.

d) 확인체크

반드시 에러라고는 할 수 없지만, 조사표의 확인을 필요로 하는 데이터를 검출하기 위한 체크이다. 체크라고해도 a, b의 체크기법과도 다르고 범위체크와 같이 에러라고 판정하지 못하는 경우 등이 이것에 해당한다.

이상의 네 종류 이외에도 데이터의 배열을 체크하는 순서체크, 합계와 내역이 일치하는 체크들이 있으며, 조사항목의 성질을 조합해서 체크를 하기도 한다. 데이터체크의 기능은 이상과 같지만, 또 하나의 편리한 기능은 자동수정 기능이다. 이것은 오류를 범한 항목의 코드를 컴퓨터에 의해 올바른(혹은 모순을 일으키지 않는다고 해야 할지) 코드로 "자동수정"하는 기능이다. 물론 이 기능에도 한계가 있어 컴퓨터로 무엇이든 자동수정이 되는 것은 아니고, 잘못된(Error) 데이터의 양이 통계집계상 거의 미미한 경우로 제한된다. 일반적으로 데이터체크는 2~3회 반복되는 것이 보통이다. 1회째의 데이터체크를 1차 체크, 2회째를 2차 체크 등으로 부르지만, 회수를 거듭 할때마다 에러수는 감소된다. 그리고 마지막에 남은 에러를 자동수정 한다고 한다. 자동수정한 코드는 프린트하지 않는 한 확인할 수 없다. 따라서 자동수정한 코드가 집계상 "좋지 않다" 라고 하면 곤란하므로 1차 자동수정 한 것을 체크리스크상에 프린트하여 적절하지 못하면 정정할 수 있게 한다. 이것을 「

임시 자동수정」이라고 부른다.

그림7 체크리스트

블럭 : ★★ 체크구분 : 1차 체크

	데이터 일련번호	시구정촌코드			조사표간 케이스	기본조사구 번호	분할 번호	사업소 번호	정정 코드	3 경영조직	4 본소·지소 별	5 개설시기	6 의류태 사업	
		도도부현	시구정촌	점사숫자									1 사업	2 업태
항목번호		01				02	03	04	05	06	07	08	09	10
정정지시 데이터	1121	13	102	4	002	0351		0019	034002	1	1	10	554	15
체크기호														G4
항목번호		01				02	03	04	05	06	07	08	09	10
정정지시 데이터	1125	13	102	4	002	0051		0023	034002	3	1	01	122	12
체크기호														
항목번호		01				02	03	04	05	06	07	08	09	10
정정지시 데이터	1136	13	102	4	002	0051		0031						
체크기호									★	★	★	★	★	

[그림 7] 체크리스트상의 정보에는 조사표의 순번을 나타내는 데이터의 일련번호, 都道府縣, 市區町村, 조사구 등의 지역을 나타내는 번호, 개별의 데이터를 나타내는 객체번호(세대번호, 사업체 번호 등), 데이터의 내용, 에러의 종류를 나타내는 체크기호, 자동수정코드 등이 있고, 에러가 있는 조사표의 전내용이 프린트 된다.

(2) 정정 데이터의 작성

사업체 통계조사를 예로 들면 그 순서는 [그림 8]에 나타난 것과 같다. 우선 에러가 있는 조사표를 골라낸다. 이를 위해서는 체크리스트 상의 都道府縣,

市區町村, 기본조사구, 사업체번호에 해당되는 조사표를 꺼내어 프린트된 데이터와 대조한 뒤에 각 체크기호 마다 처리 절차에 의해 정정을 실시한다 [그림 9].

조사표를 재심사하면 간단히 원인이 판명되는 경우와 각종의 참고자료를 참조하지 않으면 판명할 수 없는 경우가 있다. 전자는 오프코드 체크와 같이 규정코드가 아닌 경우이고, 후자는 관련항목체크와 같이 조합한 경우의 모순이다. 어느쪽의 항목을 정정하면 좋은가는 해당항목 뿐만 아니라 다른 항목까지도 점검하고 난뒤가 아니면 결정할 수 없다. 정정은 에러가 있는 조사표의 해당항목을 바른 코드로 부호기입하고, 체크리스트 데이터의 내용부분에 해당하는 부분을 빨간 볼펜으로 부호를 정정한다. 그리고 나서 데이터 일련번호를 등글게 줄친다. 정정 데이터의 작성방법은 전에는 에러가 있는 조사표를 새로이 입력했기 때문에 키엔트리(데이터를 키보드로 치는것)량이 많고, 올바른 항목이 다시 입력되어 에러가 되는 등의 결점이 있었다. 지금은 체크리스트를 사용하여 데이터 일련번호, 정정항목번호, 정정부호를 키엔트리하기 때문에 키엔트리량이 줄고, 목적인 항목만 정정된다. 그래서 키 오퍼레이타가 체크리스트를 보고 어느것을 키엔트리하면 좋은지 눈으로 판단할 수 있도록 빨간 볼펜을 사용하고 또 데이터 일련번호를 등글게 줄치고 있다. 임시 자동수정된 코드에 대해서는 다른 에러데이터와 같이 조사표를 재심사하고, 그결과 임시 자동수정 코드와 같게 정정하면 정정 데이터는 만들지 않아도 된다.

즉 체크리스트의 일련번호를 등글게 줄칠 필요는 없다. 정정 데이터가 만들어 지지 않으면 임시 자동수정을 한 자동수정 코드가 올바른 것으로 간주한다. 임시 자동수정이 좋지 않다면 올바른 코드로 정정하기 위해, 앞서 기술한 순서에 따라서 체크리스트의 일련번호를 등글 줄로 그어 정정 데이터를 만든다. 확인체크도 같은 방법으로 조사표를 점검하고, 올바르면 정정 데이터의 작성은

그림8 체크리스트 심사업무의 순서

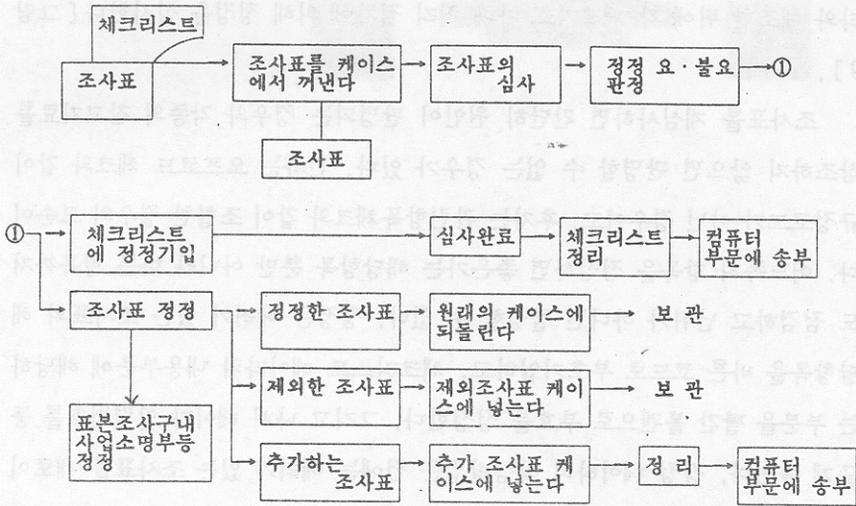


그림9 심사와 처리의 방법

체크기호	심 사 와 처 리
J1	
J2	
J3	<p>「1 사업체의 명칭 및 전화번호」 및 「8 회사에 대해서」란의 기입을 참조해서 「경영조직」 및 「본소·지소별」에 오류가 없는지 확인하고, 오류가 있으면 정정한다.</p> <p>「경영조직」 및 「본소·지소별」에 오류가 없으면 다음에 의해 처리한다.</p> <p>J1 : 「자본금액」을 「브랭크」로 정정한다.</p> <p>J2 : 「자본금액」에 대해서 해당 사업소에 조회한다.</p> <p>J3 : 「자본금액」을 「브랭크」로 정정한다.</p> <p>기본조사구내 사업소 명칭 해당부분의 정정도 행한다.</p>

필요없다. 자기테이프에 기록된 데이터의 정정을 컴퓨터로 실시하는 정정기능은 3가지가 있다.

첫번째 기능은 「갱신」이다. 데이터 일련번호로 지정된 데이터를 자기테이프에서 찾아 정정 항목번호에 따라서 해당 조사항목을 바꾸어 놓는 기능이다.

두번째 기능은 「삭제」이다. 조사표가 중복 되거나 다른 조사구의 데이터가 잘못 섞여 있을때에 사용된다. 체크리스트 상에도 「중복 데이터」는 여러로서 표시하는 것이 보통이고, 삭제하고 싶은 데이터 일련번호의 앞에 「X印」을 빨간 볼펜으로 기입하고 데이터 일련번호만을 키엔트리한 정정 데이터를 만든다. 「갱신」과 같이 자기테이프 상에서 일련번호로 지정된 데이터를 찾아서 삭제한다.

세번째 기능은 「추가」이다. 조사표가 부족할때 추가하는 기능이다. 체크리스트상에 부족한 데이터는 당연히 프린트되지 않는다. 조사구 단위의 매수 등에 의해 부족한 것이 분명할때는 어느 조사표가 자기테이프에 없는가를 확인하기 위해 체크리스트에 조사구내의 조사표를 모두 프린트한다. 특히 마크시트 방식의 조사표에서는 마크시트 독취기의 조사표 단위에서 리젝트가 발생하므로 감독수에 의한 체크를 빠뜨려서는 안된다. 조사표의 추가를 하는 경우는 「갱신」, 「삭제」와는 별도로 बै치(BATCH)로 처리한다. (조사표를 키엔트리하는 것은 체크리스트의 키엔트리와는 별개이다).

(3) 온라인 방식에 의한 정정

컴퓨터 기술의 진보로 디스플레이 장치를 사용하여 컴퓨터와 대화하면서 여러가지 처리를 할 수 있게 되었다. 체크리스트의 정정데이터를 키엔트리작업으로 작성하고, 또한 컴퓨터용으로 입력을 한후, 실제의 여러 데이터를 정정하는 방식이, 제표재료의 이동이라는 관점에서 보면 바람직한 것은 아니다. 키엔

트리를 위하여 체크리스트를 키엔트리 부문으로 보내고, 키엔트리 후의 데이터를 컴퓨터부문에 보내는 식이다. 게다가 체크리스트를 심사해서 정정한 쪽이라고 한다면 정정한 결과를 일초라도 빨리 알고 싶지만 재료의 이동에 수반하는 시간적 지연 때문에 즉시 알 수는 없다. 데이터 정정의 처리 형태로서는 조사표를 꺼내는 것이 아니라 체크리스상에서만 정정할 수 있고(조사표의 추가는 별도로서) 게다가 컴퓨터상의 데이터 정정도 체크리스트의 정정 부호를 키엔트리 하므로써 가능하게 되는 것이 바람직하다. 이런 처리 형태가 되기 위하여 온라인 방식을 이용하여 체크리스트의 정정 데이터를 디스플레이 장치에서 컴퓨터로 지시하는 것에 따라 데이터를 정정할 수 있는 방식이 사용되고 있다. 이 방법의 좋은점은 제표재료의 이동이 적고, 정정 결과를 알기 위한 시간적 지연이 적다는 것이다.

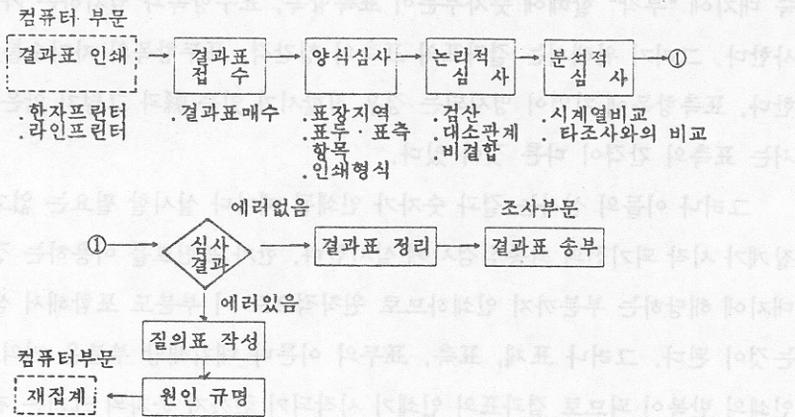
이상적인 것은 조사표를 꺼내는 일이 없이 체크 리스트에 정정을 하면 디스플레이 장치로 실제의 데이터를 찾아 해당 항목을 정정하는 것이다.

컴퓨터는 정정된 부호에 의거하여 데이터 체크를 행하는 디스플레이 장치로 그 결과(정정이 올바른가 어떤가의)를 표시한다. 이와 같이 되면 데이터 체크를 몇회나 반복하는 것이 없어지고 제표기간도 적지 않게 단축할 수 있다. 이상이 체크 리스트 심사업무의 내용이다. 이 사무는 컴퓨터 처리와 밀접한 관계가 있고 체크리스트의 양식, 정정의 방법, 정정 데이터의 열거방법, 추가조사표의 취급 등 모든 프로그램의 작성 방법에 관계하고 있기 때문에 이해하기 어려운 경우가 많다.

6. 결과표 심사

체크리스트의 심사단계에서 조사표 오류 정정을 완료하면 컴퓨터에 의해 통계표(결과표)가 집계되고, 그 결과는 인쇄되어 수작업에 의한 심사를 거친다. 이것을 「결과표 심사업무」라고 하고, 그 내용은 [그림 10]에 있다.

그림10 결과표 심사업무의 플로우 차트



컴퓨터로 집계된 결과표를 통계보고서로 사용하기 위해서 판하(판목을 뜨기 위한 밑글씨)를 작성하는데 2가지 방법이 자주 사용된다. 그 하나는 표제, 표측항목, 표두항목, 각주(본문의 아래에 따로 둔 보충설명) 등이 미리 인쇄되어 있는 대지(사진이나 그림을 붙이는 두꺼운 종이)위에 컴퓨터로 인쇄한 결과표의 숫자부분을 "부착"(종이의 일부분을 도려내고 새 종이를 바름)하여 사진촬영하는 방법이고, 다른 하나는 표제, 표측항목, 표두항목, 결과숫자 등을 모두 한자 프린트로 인쇄하여 사진촬영하는 방법이다. 판하를 작성할때 어느쪽의 방법을 채택하는가에 따라 결과표 심사업무의 내용이 약간 달라진다. 심사는

「결과표 양식의 심사」, 「결과의 논리관계 심사」, 「결과의 분석관계의 심사」로 나누어 실시된다.

(1) 양식심사

대지에 의한 판하작성의 방법에서 대지 심사는 통계보고서를 간행하는 조사부문에서 실시하고, 제표부문에서는 컴퓨터에서의 인쇄형식이 대지의 양식 즉 대지에 "부착" 할때에 숫자부분이 표측항목, 표두항목과 일치하는가를 심사한다. 그러기 위해서는 결과표의 표측의 행간격, 표두항목의 자리수를 확인한다. 표측항목에 지역이 명시되는 경우 직할시가 있는 縣과 그렇지 않은 縣에서는 표측의 간격이 다른 것이 있다.

그러나 이들의 심사는 결과 숫자가 인쇄될 때마다 실시할 필요는 없고, 본 집계기 시작 되기전의 최종점검시에 실시한다. 한자 프린트를 이용하는 경우는 대지에 해당하는 부분까지 인쇄하므로 원칙적으로 이 부분도 포함해서 심사하는 것이 된다. 그러나 표제, 표측, 표두의 이른바 대지해당 부분은 거의 같은 인쇄의 반복이 되므로 결과표의 인쇄가 시작되기 전까지 문자의 크기는 괜찮은지, 인쇄위치는 올바른지, 표측, 표두의 테두리내에 문자가 균형있게 배치되어 있는지, 오자 탈자는 없는지 등을 미리 심사해 놓는다. 마치 대지의 교정을 먼저 끝내 놓는 것과 같다.

여기까지는 결과표의 양식(대지 해당분)의 심사이다. 이것이 끝나면 다음과 같은 결과표의 숫자부분의 형식적 심사에 들어간다.

- ① 난외 항목의 확인을 한다.
- ② 결과 숫자의 체계를 확인한다. 예를 들면 3자리마다 "공백" 또는 "컴마"가 찍혀지는지, 소수점이 올바른 위치에 있는지 등의 체크를 한다.

- ③ 결과 숫자가 명시상의 약속이외의 기호로 인쇄되어 있지 않은지 확인한다. 예를 들면 「...」라든가 「 - 」(결과 숫자가 논리적으로 존재하지 않는다)는 명시상의 약속한 기호이다.
- ④ 명시단위를 확인한다. 한사람 단위인지, 천사람 단위인지, 혹은 소수점 이하 1자리인지, 2자리인지 등이다.
- ⑤ 인쇄의 농도가 일정한지 더러움이 없는지 확인한다.

(2) 논리적 심사

통계표에는 해당 통계표내에 있어서 또는 다른 통계표와의 관계에 있어서 논리적 관계를 가진 항목이 있다. 이들을 심사하여 집계수치의 올바름을 보증한다. 그러기 위해서는 아래와 같은 관계를 체크한다.

(7) 검산(산술관계)

내역을 가감산하여 총수의 일치여부를 확인한다. 이것은 결과표의 심사에서 가장 기본적인 것이다. 여기서 주의 할것이 2가지가 있다. 하나는 미상(분류불능의 경우도 같다)의 취급이다. 조사에 따라 내역의 하나로 「미상」 항목이 있는 경우와 「미상」이 없이 「총수」에 편의적으로 포함되어 있는 경우가 있다. 전자에서는 총수와 내역의 합계는 일치하므로 검산 하기가 쉽지만, 후자에서는 일치하는것과 일치하지 않는 것이 있다. 미상의 크기는 명시적으로 판명할 수 없으므로 일치하지 않는 경우에서의 검산의 보증정도는 미상의 크기에 좌우된다. 또 하나는 샘플조사나 구성비 등과 같이 집계수치나 계산수치가 「사사오입」 되어있는 경우이다. 이것도 「미상」과 같이 총수와 내역의 합계가 일치하기도 하고, 일치하지 않기도 한다. 다만 이 경우 「미상」

과는 달리 사사오입으로 생기는 불일치이므로 그 차는 ± 1 정도로 근소하다.

큰 경우는 오류라고 생각하는 편이 좋다. 난외의 명시항목에서 총수와 내역의 관계가 있는 경우에도 검산은 한다. 그러나 난외항목에서의 검산은 결과표의 한페이지의 모든 셀(요소)에 대해서 성립하므로 계산량이 많다. 따라서 주요한 셀에 한정해서 확인하면 충분하다.

(L) 등치관계

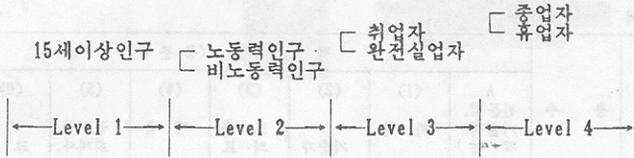
해당 통계표내 또는 다른 통계표와의 관계를, 불문하고 같은 값이 되는 항목이 있다. 등치관계는 이를 위한 체크이다. 예를 들면 국세조사에서 배우관계, 연령, 남녀별, 15세이상 인구의 통계표와 연령, 남녀별 인구의 통계표의 15세이상의 수치는 일치한다. 표측항목, 표두항목에 같은 명시가 있다고 해서 일치한다고는 할 수 없다. 어떠한 분류레벨의 항목인가를 이해하고 분류레벨을 맞춘뒤에 비교하지 않으면 안된다. 등치관계와 산술관계와는 유사하지만 등치관계는 원칙적으로 계산할 필요없이 동일한 항목이 있는지 어떤지 확인하는 것이다.

(C) 대소관계

산술관계, 등치관계는 같은 값으로 있는지 어떤지의 체크였었지만 대소관계는 「항목 A는 항목 B보다 반드시 크(적)다」라고하는 관계를 점검하는 것이다.

예를 들면 국세조사(1% 추출집계)에서 연령, 남녀별 인구의 통계표는 총인구(외국인을 포함)와 일본인으로 구성되어 있고, 수치관계는 "총인구는 일본인 인구보다 크다"로 되어 있다. 이 경우에도 앞에서 기술한 미상과 같이 어느 연령층에 외국인이 없으면 같은값이 된다.

노동력 조사에서 대상객체의 구분(취업상태)이



로 되어 있으므로 레벨1>레벨2>레벨3>레벨4로 된다.

일반적으로 표측항목, 표두항목은 계층구조로 되어 있으므로 대소 관계가 성립한다.

(근) 정치관계

이것은 집계하기전에 수치가 분명하게 정해져 있는것을 체크하는 것이다. 구성비의 경우 분모에 해당하는 셀(요소)에 대하여 100으로서 나타내기도 하고, 국세조사의 1% 추출집계에서 숫자의 아래 2자리를 00으로 하고 있는 것도 정치관계의 범위에 속한다. 일정한 값(집계하지 않아도 알 수 있다)이라고 한정하면 이 정치관계를 적용할 수 있는 경우는 많지 않다. 그러나 감독수(집계 기본수)와의 확인은 집계수치를 예정할 수 있는 것이므로 여기에 포함 시켜도 좋다.

(□) 비결합관계

표측항목, 표두항목으로부터 이루어지는 매트릭스의 셀(요소)에는 집계치가 존재해도 좋은 셀과 존재해서는 안되는 셀이 있다. 존재해서 안된다는 의미에는 두가지가 있는데, 하나는 표측, 표두의 조합(짜맞춤)이 되는 조사 데이터가 조사시점에서 발생할 수 없는 경우로 [표 4]의 상단의 통계표에서 실선으로 둘러싼 부분이 여기에 해당한다. 또 하나는 표측, 표두의 조합이 논리적으로 불합리한 경우로 같은 [

표4 비결합관계의 예

(ㄱ) 비종재의 예

남여, 산업 (대분류)	총 수	직업 (중분류)							
		A 전문적, 기술적 직업 종사자	(1) 과학 연구자	(2) 기술자	(3) 보건 의료 종사자	(4) 법률 종사자	(5) 공인 회계사, 세리사	(6) 교원	(7) 종교가
전국	58,217,500	6,095,100	69,400	1,444,900	1,588,600	39,900	52,800	1,393,900	133,900
A 농업	4,845,200	3,900	-	3,400	500	-	-	-	-
B 임업	139,200	1,700	-	1,600	100	-	-	-	-
C 어업	434,200	500	-	300	200	-	-	-	-
K 부동산업	484,900	8,000	-	5,800	100	-	-	-	-
L 서비스업	11,923,500	4,818,300	62,000	518,000	1,480,900	28,700	52,800	1,333,900	133,900
M 공무(타에 분류되지 않는 것)	2,014,500	119,500	1,000	55,700	17,300	11,200	-	-	-
N 분류불능산업	105,200	-	-	-	-	-	-	-	-

(ㄴ) 불합리의 예

	총 수	1인	2	3	4	5	6	7인 이상	1세대당 인원
전국									
일반세대(B)	36,640,084	6,615,346	6,955,945	6,801,455	8,977,067	4,196,376	1,982,789	1,111,108	3.22
6 다다미 미만	522,650	509,607	11,011	1,483	468	79	1	1	1.03
6~8 다다미	1,656,739	1,503,290	107,886	28,527	13,584	2,857	464	131	1.13
9~11	2,497,938	1,297,618	523,230	328,705	276,176	61,812	7,947	2,450	1.93
...									
48~59	2,996,301	72,074	389,768	507,688	692,414	599,589	454,697	280,071	4.31
60 다다미 이상	2,224,037	39,013	215,470	301,561	400,834	450,959	445,382	370,818	4.79
1인당 다다미수	9.2	14.8	13.6	10.0	7.8	7.6	7.9	7.3	4.79

표 4]의 하단 통계표의 실선으로 둘러싼 부분이다. 이와같이 숫자가 있으면 안되는 셀은 통계표의 구조를 결정할때에 알고 있는 사실이며, 이것을 「비결합」이라고 부르고 있다. 비결합 셀에 수치가 있는 경우, 잘못되어 있는 것은 당연하고, 데이터 체크의 단계에서 모순된 데이터를 빠뜨릴 공산이 크므로 데이터 체크 요령의 재검토를 필요로 한다.

(3) 분석적 심사

결과표가 양식적으로도 논리적으로도 오류가 없는 경우 다음으로 실시하는 것은 집계치의 질(타당성)의 심사이다. 대표적인 것으로 시계열 비교와 다른 유사조사의 결과와의 비교가 있다.

(ㄱ) 시계열 비교

전회, 전전회 조사와 비교하는 단기간의 것과 30년 전후와 비교하는 장기간의 것이 있지만 후자의 경우 데이터를 수집하기 어렵고, 조사항목이 변화되어 왔고 정의가 바뀌는 등 비교하기 어려운 면이 있지만 단기간의 비교에 있어서는 데이터의 수집도 용이하고 조사항목이나 정의의 변경도 적으므로 반드시 실시한다.

시계열 비교에 있어서는 주요한 항목에 대해서 전회와의 증감수, 증감율을 수치로 혹은 그래프 등에 의해 심사한다. 예를 들면 국제조사에 있어서는 인구에 대해서 배우관계별, 노동력 상태별, 종사상의 지위별, 산업별, 직업별 등의 인구수를, 또 세대에 대해서 가족 유형별, 경제구성별, 주택소유의 관계별 등의 시계열 비교를 하고 있다. 더우기 코호트(Cohort) 분석도 실시하고 있다.

(ㄴ) 타 조사와의 비교

관련하는 다른 조사와의 비교를 실시한다. 예를 들면 국세조사에서 주택에 관련되는 결과수치는 주택 통계조사와, 가족유형별 보봉세대는 후생행정 기초조사와, 취업상태에 대해서는 취업구조 기본조사와 비교한다. 시계열 비교에 있어서 타조사와 비교하는 경우는 조사대상이나 항목의 정의를 잘 이해함과 동시에 조사시점의 불일치에 따른 수치 차이의 해석을 신중히 실시하지 않으면 안된다.

이상과 같은 심사를 실시하고 의심이 있으면 그 내용을 기술한 「질의표」를 작성하여 원인을 구명하고, 재집계 등의 조치를 취한다. 의의가 없으면 결과표를 정리하여 공표를 하는 조사부문에 송부한다. 결과표를 통상 3부를 인쇄하고 있다. 심사용(연필로 체크 표를 기입하므로 더러워져 있다), 판하용, 보관용 이다. 한자 프린터의 경우는 최초로 심사용을 인쇄하고, 심사를 완료하고나서 2부 인쇄한다. 판하가 되는 결과표는 인쇄일록이 없는지 확인한다.

결과표 심사업무는 컴퓨터를 이용해서 산술관계, 논리관계, 비결합관계의 체크를 할 수가 있어, 결과표의 매수가 대량인 경우 대단히 효율적이다. 또 시계열 비교, 다른조사와의 비교도 컴퓨터를 이용해서 보기쉬운 형으로 편집하고 그것을 심사하는 것에 의해 스피트업(속력 증가)을 도모할 수 있다. 이들 컴퓨터 이용에 대해서는 나중에 기술하는 것으로 한다.

7. 제표계획

전술한 바와 같이 조사표 접수에서 결과표 심사까지 수작업부문의 제표사무의 구조를 보아 왔지만 이들은 어디까지나 이론에서 본 "사무의 흐름"이었고

실제로 업무를 하게 되면 몇사람으로서 실시하는지 언제까지 완료하는것인지 업무의 내용은 어느 정도인지 등이 분명해야 한다. 이것을 분명하게 하는것이 「제표계획」이다.

제표계획으로서 특히 중요한 사항은 다음과 같다.

- ① 기본수의 파악
- ② 집계수단의 결정
- ③ 제표방법의 결정
- ④ 작업량의 견적
- ⑤ 제표기간의 결정
- ⑥ 요원의 할당
- ⑦ 기타

- 제표수속(절차)의 작성
- 연수계획의 작성
- 제표재료의 조달
- 작업장소의 확보 외

①~⑥은 어떠한 업무의 흐름을 몇사람으로 실시하며 언제까지 완료하는지를 나타내는 이른바 계획의 골조부분에 해당하고 ⑦의 각항은 그것의 살붙이기 부분에 해당한다. 제표계획을 작성할때 ①~⑥의 순으로 검토해 가지만 이들은 상호 관련이 있으므로 제표계획의 확정까지에는 ②~④의 諸量을 변경시키면 ⑤, ⑥을 다시 산출 해야 하는 반복작업이 된다. 몇 회 반복하는기는 한정된 자원인 요원의 수와 사회의 요청인 조기공표(제표기간의 단축)와의 균형의 문제로 된다.

(1) 기본수의 파악

얼마만큼의 양을 처리해야 하는가를 파악하는 것이 「제표계획」 작성의 제 1보이다. 제표사무의 각 단계에서 처리하는 모든 기본수가 산출의 대상으로 된다. 예를 들면 조사표 매수, 조사구수, 접수시의 조사표 운송상자수, 혹은 각종 묶음표지 매수 등을 기본수로서 들 수 있다. 이 기본수가 판명되고 1인 1일의 처리량을 추정할 수 있으면 제표기간을 산출할 수 있다. 또 제표기간을 결정하면 소요인원을 산출할 수 있다(이들의 산출방법은 나중에 설명한다). 기본수는 정확하지는 않지만 미리 판명하는 것과 추계계산으로 도출하는 것이 있다. 예를 들면 조사구수는 전자의 예이고 사업체수라든가 인구수는 전회조사에서의 증감율에 의해 혹은 최근의 통계자료에서 산출하는 것이 많으므로 후자의 예에 해당한다. 기본수 중에서 조사표 매수의 파악은 가장 중요하다. 조사표에는 하나의 조사객체에 1매의 조사표가 사용되는 경우와 복수의 조사객체를 연기식으로 한 조사표가 사용되는 경우가 있다. 예를 들면 사업체 통계조사에서는 조사객체는 사업체이다. 그리고 한 사업체의 각 사항은 1매의 조사표에 기입한다. 또 취업구조 기본조사는 세대원 가운데 15세 이상의 사람을 조사 대상으로 하고 있고 각 인마다 1매의 조사표를 사용한다. 조사대상과 조사표가 1대 1로 대응하는 경우는 조사표의 매수를 비교적 포착하기 쉽다.

한편 국세조사에서는 연기식의 조사표를 사용한다. 조사대상은 세대원 한 사람 한사람이다. 따라서 한 세대의 세대인원수에 의해 조사표는 1 매 이거나, 2 매 이상이기도 하므로 조사표의 매수는 세대인원별 세대수에서 산출해내는 것이 된다. 또한 연기식을 확대한 것으로 책자형식의 조사표가 있다. 가계부형식이 이것에 해당한다. 가계조사의 경우 매월 8,000세대를 대상으로 조사를 실시하고 있기 때문에 세대수 즉 가계부 권수의 파악은 용이하지만 1권의 가계

부에 얼마만큼의 기입행수가 있는지 추계하는것은 상당히 어렵다. 매일의 물건 사는것이 조사대상이고, 기입된 행수 모두에 대해서 내용검사, 부호기입의 대상이 되므로 그 기본수로서의 행수를 실제에 가까운 상태로 파악하지 않으면 안된다. 어떤 종류의 조사표가 사용되고 있어도 최종적으로는 보다 가까운 기본수를 추계하는것이 필요하고, 그러기 위해서는 평상시부터 통계자료나 과거의 실적치 등을 연구해 두는것이 중요하다.

(2) 집계수단의 결정

수작업에 의해 집계할 것인지 컴퓨터에 의해 집계할 것인가를 결정한다. 컴퓨터 시대인 오늘날 수작업집계는 시대에 역행한다고 생각될지도 모르지만 조사표가 소량이고 결과표가 1~2 표로 작고 내용도 간단할 경우 프로그램의 작성기간, 조사표의 키엔트리 및 컴퓨터 체크에서의 에러 데이터의 정정 등을 고려하면 전체의 기간에 있어서는 수작업집계편이 빠르다. 그러나 일반적으로는 컴퓨터 집계가 주류이다. 컴퓨터 집계로 할 경우 수작업부문의 제표업무에서 크게 영향을 받는것은 내용검사, 에러 리스트 심사, 결과표 심사의 각 업무이다. 즉 컴퓨터에서의 처리형태, 처리내용의 정도에 따라 혹은 다른말로 하면 수작업부문에서 실시하고 있는 작업을 컴퓨터에 어느정도 부담시킬 수 있는가에 따라 제표기간, 소요인원이 바뀐다. 되도록 고도의 일을 컴퓨터에 부담시키는 편이 효율적인 것은 알지만 컴퓨터 시스템의 능력, 프로그램작성 기간 및 능력 등으로부터 제약을 받는다.

(3) 제표방법의 결정

제표업무의 각 단계에서의 처리방법을 결정한다. 즉 내용검사, 부호기입, 에러 리스트 심사, 결과표 심사 등의 제표방법을 컴퓨터로 처리한다고 생각하

면서 결정한다.

예를 들어 내용검사 업무에 있어서는 검사를 조사항목의 전부에 대해서 실시하는지, 기본 항목만으로 한정시키는지, 내용적으로는 오프코드 체크 만으로 하는 것인지, 타 항목과의 관련 체크를 실시하는 것인지, 또 에러 리스트 심사를 어느정도의 에러 데이터량이 되면 중단하는 것인지 등 정도의 유지를 고려해서 결정한다. 컴퓨터는 대량의 데이터를 고속으로 처리한다고 하는 특징을 가지고 있지만 양뿐 아니라 질의 문제로서 지금까지 당연한것 같이 수작업으로 처리하고 있던 것도 컴퓨터로 처리할 수 없을까 다시 생각하는 것도 필요하다. 제조업무에서 가장 많은 인원과 긴 기간을 필요로 하는 산업, 직업분류기호의 기업은 수작업이 아니면 될 수 없지만, 인공지능의 이용과 동시에 조사항목을 브레이크 다운(Break-Down) 하는것 등에 의해 컴퓨터를 이용하는 것을 장래의 문제로서 검토하지 않으면 안될 것이다.

(4) 작업량의 산출

제조업무에서의 작업량은 「인/일(M/D)」라고 하는 단위로 산출한다. 즉 어느 업무를 완료하기까지 투입할 인원과 완료까지 필요로 하는 일수와의 곱을 작업량으로 한다는 의미이다. 예를 들면 작업 A의 작업량이 120인/일이라고 산출된 경우 이것을 3일로 완료시키기 위해서는 $120\text{인/일} \div 3\text{일} = 40\text{인}$ 이고, 1일 당 40인이 일을하면 된다고 하는 것이 된다. 만약 120인의 요원이 있고 A의 작업을 위한 자료가 120인으로 균등히 배분된다고 하면 계산상은 1일($120\text{인일} \div 120\text{인}$)로 완료하는 것이 된다. 그러므로 작업량이 적산되어 일당의 종사가능 한 요원수가 판명되면 완료까지의 일수가 계산되고, 또 기간(일수)이 정해지면 1일 당의 요원수를 계산할 수 있다. 작업량의 산출은 그 작업의 기본수와(작업) 능률로부터 산출한다. 능률이라고 하는 것은 단위시간(봉상은 1일)당의 1인의

작업량이다. 요컨대 1인으로 1일 일하면 얼마만큼의 일을 소화시킬가하는 것이다. 이것을 능률이라고 부르기 보다는 오히려 「단위 업무량」이라고 해야 할 지도 모르지만 여기에서는 제표부분에서의 관용적 사용에 따라 「능률」을 사용하는 것으로 하였다. 따라서 작업 A의 기본수와 그 능률이 판명되면 A의 작업량은 기본수 ÷ 능률로 계산할 수 있다. 실제로는 작업 A는 몇개의 작업단위로 나누어지기도 하므로 능률을 추계하는 것은 기본수의 추계보다도 어렵다. 예를 들면 작업 A로서 내용검사 업무의 작업량을 산출한다고 하자. 우선 내용검사 업무를 몇 개의 작업단위로 나누면 다음과 같다.

- ① 都道府縣 명, 市區町村명, 조사구번호의 확인
- ② 세대번호의 확인
- ③ 내용검사
- ④ 부호기입
- ⑤ 조사표 배열순서의 확인
- ⑥ 조사구 아이덴트 시트에의 소요사항의 이기
- ⑦ 조사표 상자에의 수납

①의 작업 내용은 조사표 상자 표면에 쓰여져 있는 都道府縣 명, 市區町村 명, 조사구 번호와 상자내에 들어있는 조사표 표지(조사구마다 붙어 있다)에 기입되어 있는 3항목 내용을 비교해 보고 수납잘못이 없는가를 점검하는 것이다. 이런 경우의 기본수는 확인하는 회수이므로 조사표 표지의 출현회수 즉 조사구수의 3배이다. 3배라고 하는것은 都道府縣, 市區町村, 조사구 번호가 3개소 있기 때문이고, 都道府縣이 47이므로 47회만 확인해서는 안되며 市區町村도 市區町村의 수만큼만 확인할 수 있다. 어디까지나 조사표 표지가 출현할때마다 都道府縣도 市區町村도 확인하지 않으면 수납 오류의 발견에는 불충분하게 된

다. 다음에 능률을 추계한다. 이것은 조사표 상자와 조사표 표지의 해당 내용을 눈으로 확인하는 작업을 1인이 1일중 실시 하려면 몇 회 할 수 있을가를 추계하는 것이다. 물론 인간은 기계가 아니므로 일정한 속도로 하루종일 확인할 수 없고, 또 확인한 결과 오류가 있는 경우 중단되는 일도 있다. 따라서 일정 속도로 확인 할 수 있다고하는 능률을 안전률로 보고 최종의 능률이라고 한다. 실무상에 있어서는 이 능률의 추계에는 과거의 실적치가 도움이 된다. 실적치가 없거나 혹은 OA화 등에 의해 작업형태가 현저하게 바뀐 경우는 시험제표를 하여 능률의 측정을 하지만 때로는 시험제표를 하면 능률이 높아지기 때문에 적용에 대해서는 충분히 주의한다.

이상과 같이 하여 ①의 기본수와 능률이 산출 되었으므로 ①의 작업량은 계산할 수 있다. 마찬가지로 ②~⑦의 기본수, 능률로 각 작업량을 계산하고 그들을 합계한 것이 내용검사 업무의 작업량이 된다. 더구나 ⑩의 확인부분은 조사표 표지 1매에 대해서 3회로 하였지만 3개소를 하나로 합쳐서 조사표 표지 마다 1회 행하여도 좋다. 이 경우의 기본수는 3배할 필요 없이 조사구수 자체이며, 능률은 3회를 하나로 합치므로 전보다 낮게 된다. 작업단위를 세분하면 정확한것 같은 인상을 주지만, 인간이 하는 작업은 작업단위가 세세하게 됨과 더불어 그 경계가 모호하게 되고 또 기본수, 능률을 추계에 의뢰하고 있기 때문에 너무 세분하는 것은 바람직하지 못하다.

(5) 제표기간의 결정

각 제표업무의 작업량을 (4)의 방법으로 계산하여 제표기간을 구한다. 간단한 예로서 제표업무와 그 작업량을

접수·정리 : 100인/일

내용검사 2,500인/일

에러리스트 심사 750인/일

결과표 심사 800인/일

로 가정한 경우에 제표기간을 매일 23인의 요원이 종사할 수 있는 것으로 하고 또한 각각의 제표단계의 작업량을 데이터 엔트리, 컴퓨터 처리(일수는 가정) 단계도 포함시켜 순서대로 처리해 가는 방법으로 계산을 한다. 즉 앞의 작업이 끝나고 나서 다음 작업으로 넘어가는 진행방식이다. [표 5]에 각 제표단계의 제표기간을 정리해 두었다. 계산식의 결과는 0.5일 단위로 나타내고 있고, 계산의 결과 총 193.5일의 기간을 필요로 한다. 이 193.5일은 실질적 일수이므로 제표의 개시월일부터 제사날 등의 휴일을 제외하고 달력에 의해 실제 노동일수를 구하고 실질적으로 193.5일에 해당하게 되는 월일이 제표기간의 마지막이 된다.

표5 제표단계별 제표기간

제 표 단 계	작업부문	작업량	요원수	계 산 식		제표기간
				인/일	인 일	
1. 접수 · 정리	수작업	100	23	$100 \div 23 =$	4.3	4.5
2. 내 용 검 사	"	2500	23	$2500 \div 23 =$	108.7	109
3. 데이터 엔트리	컴퓨터	-	-	-	-	7(가정)
4. 데이터 체크	"	-	-	-	-	2(가정)
5. 에러리스트 심사	수작업	750	23	$750 \div 23 =$	32.6	33
6. 결과표 집계	컴퓨터	-	-	-	-	3(가정)
7. 결과표 심사	수작업	800	23	$800 \div 23 =$	34.8	35
합 계	-	4150	23	-	-	193.5

위 예와 같이 각 작업을 순서대로 처리한다고 가정한 제표기간 산출의 문제점은 「재료 기다림」이 발생하는 것이다. 왜냐하면 내용검사가 전부 완료하고 나서 데이터 엔트리, 컴퓨터 처리를 하고 에러 리스트 심사에 들어가므로

데이터 엔트리, 컴퓨터 처리를 하고 있는 9일간 23인은 재료(에러리스트)가 없어 일을 하고 싶어도 못하고 있는 것이 된다. 이것은 「결과표 심사」의 단계에서도 마찬가지다. 이와 같은 상태를 피하기 위해 일반적으로는 각 작업의 종반 부분과 다음 작업 전반 부분을 병행시키기도 하고 여러 縣의 처리를 병행적으로 처리시키기도 하고 다른 조사의 재표와 서로 협조 시키면서 「재료 기다림」의 해소에 노력하고 있다.

(6) 요원의 할당

통계조사의 빠른 공표를 각 방면에서 요구하고 있지만 작업량의 크기, 소요인원의 제약으로 인한 물리적 한계가 있다. 여기에서는 재표기간을 주어 소요인원을 구해보자. 재표업무의 종류와 그 작업량은 (5)와 같다고 하고 예를 간단히 하기 위해 재표기간을 월 단위로 다음과 같이 준다.

접수·정리	4월 ~ 5월
내용검사	5월 ~ 8월
데이터엔트리	5월중순 ~ 9월중순
데이터 체크	
에러리스트 심사	6월 ~ 9월
결과표 집계	7월 ~ 10월중순
결과표 심사	8월 ~ 10월

여기에서 각 재표 업무가 일부 중복, 특히 결과표 심사가 결과표 집계(컴퓨터 처리)가 끝나지 않는 가운데 시작하고 있는 것은 국세조사와 같이 都道府縣별로 순차 공표하는 초사를 가정한 때문이고, 지역별로 공표하지

않는 조사에 있어서는 결과표 집계가 끝나지 않으면 결과표 심사에 들어갈 수 없다. 에러 리스트 심사이전의 제표업무는 이론상은 병행할 수 있다. 요원수의 산출에 앞서서 4월~10월의 일요일, 제사날을 제외한 실제 노동일수를 계산한다. 토요일은 반일 근무로서 0.5일로 한다. 그 결과는 [표 6]의 표두의 월의 아래에 나타내고 있다.

접수·정리의 작업량 100인/일을 4월, 5월의 양 월로 할당하는 것은 날짜수에 비례하여 할당한다.

즉 4월분은

$$\begin{aligned}
 \text{4월의 작업량} &= 100\text{인일} \times \frac{\text{(4월의 일수)}}{\text{(4월, 5월 일수 합계)}} \\
 &= 100 \times \frac{23}{(23+21.5)} \\
 &= 51.7 \rightarrow 52
 \end{aligned}$$

같이하여 4월의 작업량을 산출하지만, 날짜수 비례로 계산하면 사사오입의 관계에서 단수(끝수)가 나오므로 그 조정은 최종월로 한다. 그러기 위해서는 최종월의 값은 총작업량에서 최종월 직전까지의 각월의 작업량을 공제한 나머지로 하면 좋다. 따라서 접수·정리의 5월분은

$$100\text{인/일} - 52\text{인/일} = 48\text{인/일로 된다.}$$

이와 같이하여 각 제표업무의 각 월의 작업량을 계산한다. 그 결과는 [표 6]의 기간을 나타내는 바(Bar)의 상에 가르키고 있다. 4월에서 10월까지의 각 작업량을 월마다 합계한 것이 "월간 작업량"이고 이것을 각 월의 일수로 나눈 값(앞당겨서)이 1일당의 소요인원이다. 따라서 1일당 실제 노동 인원과의 차가 과부족인원(1일당)으로 된다. [표 6]의 예에 있어서는 5월~8월에 인원의 부족이 있고 많은때는 2,3인씩이나 부족하다. 인원

이 부족한 월은 당연히 업무가 지연되기 때문에 이 요원 계획에 있어서는 기한내에 완료하지 않는것이 된다. 이것을 해결하기 위한 대책으로는 3가지 방법이 있다. 첫째는 제표기간을 연장해서 월당의 작업량을 줄이고, 둘째는 아르바이트 또는 타 부문에서의 지원에 의해 실제 노동인원을 늘리고, 셋째는 제표방법을 재검토하고 개선을 해서 작업량을 줄인다. 이 순서대로 따르는 것은 어려움이 있지만 바람직한 방향이다.

표6 제표요원계획

		소 요	4월	5월	6월	7월	8월	9월	10월
		작업량	23일	21.5일	24일	25일	23.5일	22일	24일
1. 접수 · 정리	인복 100		52	43					
2. 내용 검사	2500			572	638	665	625		
3. 데이터엔트리	—								
4. 데이터체크	—								
5. 에러리스트심사	750				190	198	187	175	
6. 결과표집계	—								
7. 결과표심사	800						271	253	276
월 간 작업량		4150	52	620	828	863	1083	428	276
1 개 월 당	실근무 작업량 (연실근무인원)	3749	529	494.5	552	575	540.5	506	552
	과부족작업량 (연과부족인원)	-401	477	-125.5	-276	-288	-542.5	78	276
1 일 당	A. 소요인원	—	3	29	35	35	46	20	12
	B. 실근무인원	—	23	23	23	23	23	23	23
	C. 과부족인원	—	20	-6	-12	-12	-23	3	11

실제로는 세가지의 방법을 혼합해서 요원수를 산출한다. 더우기 [표 6]에서 1일당의 인원 (A,B,C)를 구하고 있지만 이것은 이해하기 쉽게 구한것이고 사실 정확하지는 않다. 예를 들면 4월의 1일당 소요인원을 3인 (52인/일 ÷ 23일, 절상)으로 계산하였지만 3인이 23일간 일을 하면 69인/일로 되는 17인/일(69인/일 - 52인/일)을 상회한다. 따라서 5월의 작업량

48인/일 가운데 17인/일분을 먼저 소화해 버리는 계산이 된다. 그렇다면 4월은 23인이 일하는 것이므로 접수·정리에 2개월씩이나 걸리게 하지 않고 4월중에 완료시키는 안도 생각할 수 있지만 조사표의 송부에 4월~5월의 2개월을 필요로 하게 되면 4월은 사람은 있어도 조사표를 기다리는 상태가 된다. 또 4월의 작업량을 52인/일로 억제하면 4월의 시작부터 3인으로 일할 필요는 없고 1인은 6일간(52인/일 - 2인×23일) 일하면 적당한 것이 된다. 이와같은 오차가 나오는 것은 1일당 인원수 산출로 절상을 하고 있기 때문이다. 그래서 엄밀히 하려고 하면 23인분의 각월의 작업량(4월의 경우는 23인×23일)을 산출하여 그것과 월간 작업량과의 과부족을 구하는 편이 정확하고 또한 실제적 이기도 하다. 예를 들면 5월의 작업 부족량은 $620\text{인/일} - 494.5\text{인/일} = 125.5\text{인/일}$ 로 이것을 5월 시작부터 인원의 보충을 하려고 하면 $125.5\text{인/일} \div 21.5\text{인} = 5.8\text{인}$ 약 6인이 되고 5월 후반의 10일간에서 보충하려고 하면 $125.5\text{인/일} \div 10\text{일} = 12.6$ 약 13인이 된다. 보충되는 일수에 따라 보충인원을 계산하고 인원이 정해지면 일수를 간단히 계산할 수 있지만 반면 125.5인/일의 부족이라고 해도 양적인 크기가 1일당 6인의 부족이라는 표현 방법에 비해 파악하기 어려운 결점이 있다.

(7) 기 타

지금까지 제표계획의 기초부분에 대해서 설명했고 이하는 주제별로 고려해야 할점을 기술한다.

(가) 제표 수속의 작성

1인이 모든 제표업무를 처리하는 경우는 그 사람의 생각에 의해 문서없이 처리하는것도 허용 되겠지만 일반적으로 많은사람이 처리를

- 분담해서 실시하므로 표준화된 처리 순서를 기술한 입문서가 필요하다. 이것을 「제표수속」이라고 총칭하여 제표의 각 단계마다 분책으로 작성하는 것이 보통이다. 그러나 조사대상이 적거나, 혹은 처리절차가 간단한 조사에 있어서 분책으로 할 정도가 아닌 경우는 한권으로 합쳐도 좋다. 분책의 경우나 한권의 경우 「접수·정리」에서 「결과표 심사」에 이르기까지의 각 단계마다 다음의 사항을 고려하면서 작성한다.
- 해당 업무의 개요를 기술한다. 가능하면 플로우차트를 그려놓는 것도 필요하다.
 - 해당 업무에 필요한 일반적인 주의사항을 기술한다. 예를 들면 필기구의 종류, 정정의 방법, 조사포 오손시의 처리 등이다.
 - 해당 업무 처리에 필요한 서류의 양식 견본을 부록으로 붙여둔다.
 - 각 업무처리의記述은 예시를 붙여 이해하기 쉽게 한다. 가능하면 2색 이상의 색채인쇄가 좋다.

(L) 연수계획의 작성

각 제표업무의 내용을 이해하기 위하여 「제표수속」을 텍스트로 해서 연수를 실시한다. 연수는 해당업무의 실시전이면 어느때라도 좋지만 연수에 이어 곧 실무에 들어가는 것이 가장 좋다. 연수는 강의(설명)와 연습(실습)으로 구성한다. 제표 계획면에서는 연수기간 중에 2가지 점에 주의할 필요가 있다. 첫째는 연수기간중에는 업무를 수행할 수 없으므로 가동인원은 "제로"가 되지만, 제표업무의 총 작업량은 연수에 필요한 총 작업량 = 연수인원 × 연수기간을 계산한다. 두번째는 업무가 「제표수속」대로 행하여지고 있는지, 또 연수의 성과가 제대로 나오고 있는지를 체크하여야 한다. 예를 들면 「내용검사·부호기

입업무」의 경우 단기간에 조사표를 모두 재검사를 하든가, 국세조사
와 같이 장기간을 필요로 하는 경우는 중간에서 재 연수를 실시하고
새로이 발생한 문제의 처리나 새로운 사례를 수집하여 「제표수속」의
부족을 보충하는 것이 필요하다. 적은 조사의 경우에는 담당자 전원이
「접수·정리」부터 「결과표 심사」까지의 처리를 실시하는 일이 많
으므로 그 수속을 한번에 연수하여도 좋지만 큰 조사의 경우에는 제표 각
단계의 업무를 분담하는 사람수가 다르게 되어 있고 [표 6]에서 본
것과 같이 업무가 시간적 제약을 수반하여 시작되므로 전원이 「접수
·정리」나 「결과표 심사」를 연수할 필요는 없다.

(ㄷ) 제표재료의 조달

조사표를 시작으로 요계표, 명부류, 요도류 등의 제표에 필요한
용품을 「제표재료」라고 총칭하지만 이들 제표재료가 수작업부문의
담당자와 함께 컴퓨터 부문에 적당하게 이송되어야 한다. 특히 수작업
부문에 있어서는 1일의 처리량에 따라서 제표재료가 반입되지만 그 중
에서도 조사표는 엄중히 관리·보관 되어야 할 것이므로 1일 업무가
끝난뒤 사무실에 방치되는 일이 없이 반드시 창고나 보관고에 보관한
다. 따라서 업무를 시작할때 1일의 처리량에 알맞는 조사표를 담당자
에게 송부하고 1일의 마지막에 처리가 끝난 조사표를 창고에 반납하게
해야한다. 큰 조사일수록 많은 사람이 작업을 하고 각 제표 업무도 경
합해서 실시되므로 필요로 하는 제표 재료를 필요한 부서에 매일 반입
·반출하는 계획은 요원사용 계획과 함께 중요한 것이다.

(ㄹ) 작업장소의 확보

제표업무는 책상위의 업무이긴 해도 의외로 넓은 작업장소를 필요

로 한다. 예를 들면 「내용검사·부호기입 업무」를 보면 책상 위에는 조사표 상자, 조사표, 세대명부 등이 놓여지며 산업, 직업의 부호기입 시에는 산업분류, 직업분류의 코드북 색인집이라고 하는 참고서류가 놓여진다. 그밖에 부호를 기입할때 의문이 있으면 각종의 참고서적을 참조하기 때문에 책상위에는 이들의 재료가 문혀버리는 일도 많다. 작업실은 앞서 언급하였지만 1일분의조사표가 반입되어 오프로 그것을 놓을 장소가 필요하고 통상은 반입·반출을 용이하게 하기위해서 짐수레(Bogie)를 사무실에 대기시키므로 그 부분의 공간을 확보해야 한다. 대량의 조사관계 서류가 송부되어 가는 경우 접수·정리를 하기전에 일시적인 체화(운반치 못해 밀린화물)가 생긴다. 또, 수송상자의 포장을 푸는것, 관계서류의 분류, 폐기물 등 「접수·정리사무」를 실시하는 장소가 필요하게 된다. 이 경우 작업장소의 확보는 송부기간 동안의 일시적인 것이므로 회의실, 체육관 등을 이용하는 일도 많다.

8. 진도관리

실제로 작업을 시작해 보면 제표계획대로 진행되지 않는 일이 많다. 특히 각 제표업무는 초기단계에서 늦어지기 쉽다. 즉 예정한 처리량에 도달하지 않는 것이다. 그 이유로서 다음의 요인을 들 수 있다. 첫째는 기본수의 파악을 잘못된 경우이다. 특히 실제보다 과소로 견적한 경우 인원의 부족을 초래해서 기간내에 집계를 할 수 없다. 기본수의 파악은 앞서도 기술했지만 추계에서 구하는 경우가 있는데, 그 대표적인 예가 「체크리스트 심사업무」에서의 「체크리스트」의 매수 혹은 「에러 데이터」의 양이다. 에러 데이터의 양은 조사표의 기입상황, 체크요령서의 상세한 지시에 의해 달라지는 것은 당연하고, 전회

조사로 추계를 하기는 어렵다. 둘째는 능력의 산출에 문제가 있는 경우로 2가지의 문제를 안고 있다.

첫번째 문제는 숙련률(그 업무에 익숙해 지기까지 능력을 낮게 한다)과 숙련기간(예정능률에 이르기까지의 기간)을 어느 정도로 계산했는지이고, 두번째 문제는 대도시를 가진 縣에서 처리의 곤란함(예를 들면 대도시에는 다채로운 직업, 산업이 있어 부호기업의 능률이 저하된다)을 능력산출에 어느 정도 반영시켰는지이다. 그러나 이것은 이론상으로 확실히 고려해야 하지만 의의 처리에 있어서 부호기업 작업의 중단 등을 생각하면 현실로는 어떻게 반영시켜야 좋은지 알 수 없으므로 기업능률의 저하나 중단이라고하는 리스크(Risk)를 모두 평균화한 전회조사의 능률을 참고로 하게 된다. 숙련률과 숙련기간은 많은 사람으로 장기간을 필요로 하는 제표업무에 적용한다. 대표적인 것은 「부호기업 업무」이다. 이것은 요구되는 전문 지식이 많으므로 기억하기 까지의 능력은 낮다. 그래서 숙련률과 숙련기간을 써서, 예를 들면 최초의 1개월은 숙련률을 80%, 다음의 1개월은 90%, 3개월째에는 예정능률(100%) 이라는 식으로 보정을 하는 것이다. 숙련률, 숙련기간은 업무내용의 난이도에 따라 바뀐다.

이상과 같이 제표계획에는 약간의 불확정 요소가 내포되어 있으므로 계획대로 업무가 진행되고 있는가를 체크할 필요가 있다. 체크 항목은 품질, 처리량, 능력의 3가지이다. 여기에서는 품질은 생략하고 「처리량」, 「능력」은 각 제표업무마다 [그림 11], [그림 12]와 같이 그래프로 나타내면 전체를 포착하기 쉽다. 그래프화 하기 위해서는 1일의 「처리량」, 그 처리량을 만들어 낸 「투입인원」이라는 기초적인 수치의 수집이 필요하다. 초기단계에서는 처리량, 능력도 예정보다 낮은 것이 보통이지만 제표계획 작성자가 각 요소에서 이와 같은 추이를 분석하면서 각각의 제표업무가 「언제」 끝나는지 또 각 제표업무에서 「재료 기다림」이 발생하지 않는지를 검토하고 제표계획의 미세

한 조정을 실시해야 한다.

그림11 처리량 그래프

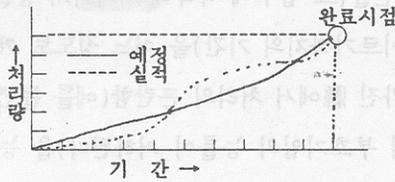
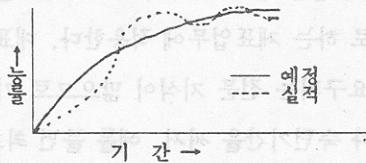


그림12 능률 그래프



제 3장 컴퓨터에 의한 제표

본장부터는 컴퓨터 부문에서의 제표에 대해서 설명할 것이나, 그전에 컴퓨터에 대해 필요한 최소한의 지식을 이해해 두고자 한다.

구체적인 예가 없이는 이해하기 어려운 부분은 관련되는 부분에서 해설을 하였다. 특히 여기에서의 컴퓨터 지식은 제표관계 업무에 있어서 컴퓨터가 할 수 있는 일과 할 수 없는 일의 구별정도의 것이지 프로그래머가 되기 위한 전문적 지식이 아니므로 컴퓨터를 알고 있는 사람은 읽지 않아도 상관이 없다.

1. 컴퓨터 입문

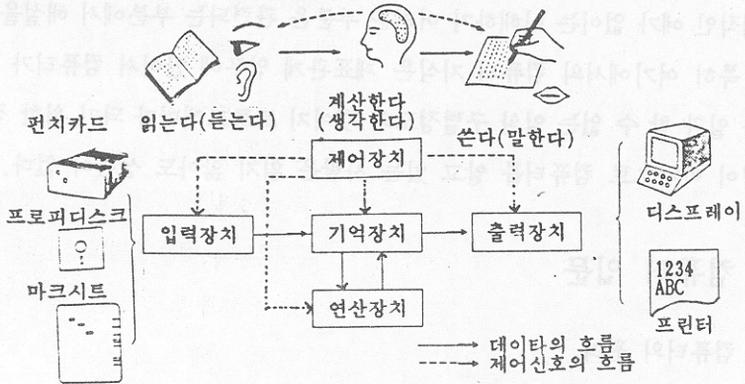
(1) 컴퓨터의 원리

컴퓨터는 [그림 13]과 같이 입력장치, 기억장치, 연산장치, 출력장치, 제어 장치의 5개의 장치로 이루어져 있다. 입력장치는 컴퓨터에 데이터를 전달하는 장치이다. 전달된 데이터는 기억장치에 보관되고 계산할 때에는 기억장치에서 연산장치로 옮겨져서 실행되며, 결과는 다시 기억장치에 되돌려 진다. 한편 그 결과를 알리기 위하여 출력장치가 사용된다. 제어장치는 입력, 기억, 연산, 출력의 각 장치가 정연하게 작동할 수 있도록 제어하는 장치이다. [그림 13]에 의거하여 추상적으로 컴퓨터의 기본요소와 활동에 대해서 기술하였지만, 컴퓨터가 전자탁상계산기와 달리 거의 인간의 개입없이 자동적으로 계산을 할 수 있는 것은 설명한 5개의 장치와 그것을 움직이는 프로그램이 있기 때문이다.

자동적으로 계산하는 기계의 원리는 130여년 전에 찰스 바베지(1792-1876)에 의해 고안되었다. 즉 전술한 5개의 장치와 프로그램이 있으면, 자

동적으로 계산을 할 수 있다는 개념이다.

그림13 컴퓨터의 5요소



전자의 활동을 이용한 컴퓨터는 1945년에 개발된 진공관식의 ENIAC이 최초이다. 그 이후 현재와 같이 우수한 기능을 발휘하는 컴퓨터에 이르기까지 여러종류의 개량, 발전이 있었고, 특히 진공관에서 트랜지스터, IC, LSI, VLSI에 이르기까지 경이적인 진전은 있었지만, 한가지 변하지 않은 것은 컴퓨터를 움직이기 위한 명령을 순차적으로 처리한다는 것이다. 이 방법은 폰 노이만 (1903-1957)에 의해 고안 되었으며, 명령을 순차적으로 처리하는 컴퓨터를 「노이만형 컴퓨터」라고 하며, 현재 사용중인 대부분의 컴퓨터는 이 형에 속한다. 또한 컴퓨터의 진보과정을 「세대」라고 부르는데, 초기의 제 1세대에서 현재의 제 4세대까지 이르렀지만 노이만형 컴퓨터의 한계, 예를 들면 프로그램의 작성에 많은 노력이 필요하기 때문에 프로그램 완성이 지연되는 것이 큰 문제로 되어 오고 있다. 그래서 다음의

제 5세대의 컴퓨터에 있어서는 인간의 사고에 가까운 추론을 하고 또한 명령을 병렬적으로 처리하는 컴퓨터를 만드는 것이 연구되고 있으며, 몇몇 성과가 나오기 시작하고 있다.

현재 직장에서는 퍼스컴, 워드프로세서, 팩시밀리 등 OA化(사무자동화)가 한창이며, 많은 사람들이 기기를 사용하여 업무처리를 하고 있다. 과거 처럼 컴퓨터 알레르기나 컴퓨터에 대한 잘못된 이해, 즉 「컴퓨터는 무엇이든 할 수 있다」, 「컴퓨터는 올바르다」라고 하는 맹신형은 적어지고 있다. 그러나, 아직도 컴퓨터라고 하면 별세계다 라고 생각하는 사람이 많은 것도 사실이다. 그러면 「컴퓨터」는 무엇일까?. 정보처리 기계라고 알고 있지만 本質은 「계산기」이다. 그러므로 계산에 필요한 데이터를 주어야만 한다. 계산하려고 하는 데이터를 컴퓨터에 주는 것을 「입력한다」라고 한다. 탁상용 전자계산기에 있어서는 계산에 필요한 데이터는 숫자키를 눌러서 데이터를 주므로 「입력한다」는 숫자키를 누르는 것으로 행하여진다. 「입력」된 데이터는 컴퓨터의 기억장치(메모리라고 약칭 한다)에 보관된다. 메모리라는 것은 극장의 지정석과 같이 좌석이 늘어서 있다고 생각하면 된다. 하나의 좌석에는 어떠한 이름의 사람이라도 한사람만 앉을 수 있다. 메모리의 한자리 (좌석에 해당)에는 하나의 「문자」가 저장(기억하는것) 된다. 여기에서의 문자는 컴퓨터가 인식할 수 있는 것으로 제한되며, 통상 영자, 숫자, 쉼이나 콤마라고 하는 기호 등 256 종류이다. 최근의 컴퓨터는 한자(漢字)도 취급하고, JIS 제 1수준은 2,965종류나 된다. 메모리의 한자리 수에는 약속된 256종의 문자밖에 저장될 수 없으므로 약속되어 있지 않는 한자의 처리에는 특별한 연구를 하고 있고, 한자 문자를 메모리 에서는 2자리 문자의 조합으로 표현하고 있다. 그리하면 1자리로 256종류이므로 2자리로 $256 \times 256 = 65,536$ 종류의 문자를 나타내는 것이

된다. 메모리의 크기(좌석 총수와 같음)를 나타내는데 흔히 「64K 바이트」라든가 「1M (메가) 바이트」라고 하는데, 이것은 메모리의 자리수가 64,000자리 및 100만 자리라는 것이며, 바이트는 8비트(비트는 0과 1의 조합)로 1문자를 표현하는 체계의 1자리를 의미하는 명칭이다.

계산은 메모리에 기억되어 있는 숫자들이, 연산장치에 이송되어 행하여지며, 그 답은 메모리에 되돌려진다. 연산장치는 주판이라고 생각하면 좋다. 주판으로 계산할 답을 종이(메모리)에 써두는 것과 같다. 컴퓨터의 계산결과는 메모리에 있으므로 인간으로서는 알 수 없다. 인간에게 이해할 수 있도록 하는 것은 「인쇄한다」라든가 「화면에 비치게 한다」든가 하지 않으면 안된다. 메모리 내의 문자를 컴퓨터로부터 꺼내는 것을 「출력한다」라고 한다. 주변장치는 차치하고, 컴퓨터는 계산 데이터를 입력해서 우선 메모리에 기억시키고, 메모리내의 데이터를 사용해서 계산하고 그 결과를 출력하는 기계이다. 또한 「계산」이라고 하는 기능 대신에 「찾는다」라고 하는 기능을 부가하면 메모리에 기억되어 있는 특정한 문자를 찾을 수가 있다. 이것은 입력된 데이터(메모리에 있다)에서 특정한 문자를 가진 데이터를 고를 수가 있고, 컴퓨터가 「계산」 이외의 처리도 가능하다는 것을 의미하고 있다. 예를 들면 데이터베이스라든가 정보검색 등이 이것에 해당하며, 이것이 바로 컴퓨터가 정보처리기계라고 일컬어지는 연유이다.

(2) 명령

컴퓨터를 움직이기 위해서는 「명령」이 필요하다. 명령은 「○○을 해라」고 하는 것과 같이 컴퓨터를 마치 사람으로 생각하고 어떻게 할 것인가를 전하는 것이다. 예를 들면 입력장치를 사용해서 데이터를 메모리에 기억하기 위해서는 「읽어라」라고 하는 명령을 사용한다. 명령은 컴퓨터

의 종류에 따라 고유의 약속이 있어 표현형식의 통일은 되어있지 않지만 사고방식은 같다. 「읽어라」하는 정도로는 불완전하다. 메모리의 어떤 위치에 데이터를 기억시키면 좋은 것인지 또 어느 입력장치부터 읽는 것이냐가 지시되어 있지 않으므로 「읽어라」라고 명령된 정도로서는 컴퓨터는 움직이지 않는다. 인간의 경우는 그 경우의 상황을 쉽게 판단하여 행동을 하지만 컴퓨터는 사람과 같이 능숙하게 판단하여 행동할 수 없으므로, 자세한 부분까지 지시해 주어야 한다. 컴퓨터는 불명한 점이 있으면 작동하지 않기 때문에 명령에는 목적에 해당하는 「어디에」라든가 「무엇을」이라는 보조정보가 필요하다.

上記 例에서 보듯이 입력장치의 종류와 메모리의 위치를 나타내지 않으면 명령은 완전하지 않다. 메모리에는 지정석의 번호와 같이 0부터 시작하는 번호가 붙여있는데, 이것을 「어드레스(번지)」라고 한다. 명령에는 다음과 같은 종류가 있다.

(7) 입력, 출력에 관한 명령

입력장치에서 데이터를 메모리에 기억시키는 명령은 「읽어라」라든가 「입수해라」라고 하는 것과 같은 동사로 나타낸다. 입력장치에는 카드 독해장치, 플로피 디스크 장치, 자기테이프장치, 광학식 독해 장치 등 여러가지의 종류가 있어 데이터가 어느장치의 매체에서 만 들어지고 있는가에 따라 이들의 장치를 구분하여 사용한다.

메모리의 내용을 출력장치로 꺼내는 명령은 「써라」라든가 「인쇄해라」라고 하는 등의 표현을 한다. 출력장치에도 여러 종류가 있는데, 예를 들면 인쇄장치, 자기테이프 장치, 디스플레이 장치 등이 있으며, 인쇄해 보고 싶은 때에는 인쇄장치, 비디오 화면에서 보고싶을

때에는 디스플레이 장치를 지정한다. 입력장치, 출력장치 중에는 입력, 출력 겸용형의 장치가 있다. 자기매체를 사용한 테이프, 플로피 디스크, 디스플레이 등은 겸용형이다.

(L) 계산에 관한 명령

컴퓨터의 계산은 「가감승제」의 4종이 기본이다. 계산의 방법은 인간과 같으며 메모리에 있는 2개의 수치사이에서 가감승제가 행하여진다. 그리고 답을 기억 시키기 위한 메모리의 장소(어드레스)를 나타낼 필요가 있다.

어드레스라고 하는 개념은 다소 이해하기 어렵지만 컴퓨터는 「반복계산」이 가능한 기계이다. 예를 들면 3+4를 생각해 보자. 명령을 「3과4를 더해라」라고 한다면, 8+5의 경우에는 「8과 5를 더해라」가 된다. 이것은 「수치 M과 수치 N를 더해라」라고 하는 형식으로 M을 급여액, N을 봉근수당으로서 100인분의 급여계산(단순한)을 한 경우 M, N이 100인 각각으로 다르다면 100인분의 명령을 만들고 그 사용 구분을 잘 분간해야 한다. 100인이 1000인이 되면 1000인분의 명령을 만들어야 하고 또한 급여액, 봉근수당이 달라진 경우에도 명령의 변경을 해야한다. 그래서 어드레스의 개념을 도입하면 앞의 3+4는 3이 기억되어 있는 메모리의 어드레스 X와 4가 기억되어 있는 메모리의 어드레스 Y를 사용해서 「X와 Y를 더해라」라는 형식이 된다. 3이 기억되어 있는 어드레스(좌석번호를 가정해라)가 1200번지이고 4가 기억되어 있는 어드레스가 1250번지인 경우 「1200번지의 내용과 1250번지의 내용을 더해라」라고 하는 것으로 1200번지의 내용을 8로, 1250번지의 내용을 5로 하면 같은 명령으로 8+5를 실행할 수 있다. 즉 X, Y는 대수

식의 변수와 같이 그 내용을 바꾸므로해서 몇가지의 값을 계산할 수 있게 된다. 어드레스 X, Y의 내용을 바꾸는 데에는 예를 들면 입력 명령을 사용해서 어드레스 X, Y에 데이터를 읽어 들이면 되므로 조금전의 100인분, 1000인분의 급여 계산은 첫번째 사람과 같은 메모리 어드레스에 읽어 들여 데이터를 「바꾸어 놓아」서 마치 첫번째사람 처럼 생각하여 계산한다. 이것을 100인분, 1000인분 반복시키면 좋은 의미로 가령 급여액, 봉근 수당이 달라져도 입력하는 데이터를 변경하면 되므로 명령을 바꿀 필요는 없다. 이러한 것을 할 수 있는 것도 메모리의 내용을 직접명령으로 사용하지 않고 어드레스라고 하는 간접적인 지정으로 명령을 조립하고 있기 때문이며, 이로써 컴퓨터의 융통성을 높일 수 있다.

(C) 비교에 관한 명령

「비교」라고 하는것은 메모리 어드레스 X에 있는 문자와 어드레스 Y에 있는 문자와의 관계로 보통은 「크다」, 「작다」, 「동등하다」라고 하는 상태를 점검하는 것이다. 비교명령은 「메모리 어드레스 X에 있는 문자는 어드레스 Y에 있는 문자보다 크느냐」라든가 「동등하냐」 등으로 표현되어 비교의 결과(에스인지 노우인지)에 따라 다음에 실행할 명령이 선택된다.

정확히 철도의 포인트(레일의 전철기)와 같은 분리기의 기능을 가지고 있다. 그런데 어드레스 X, Y의 문자가 수치라면 「크다」, 「작다」, 「동등하다」의 관계는 이해하기 쉽지만 문자가 영문자나 기호의 경우는 「동등하다」의 관계 이외는 해석하기 곤란한 경우가 있다. 예를 들면 「컴마」와 「A」는 어느 쪽이 큰가라고 하더라도 인간의 세계에

서는 대소관계가 숫자에 한정되어 있기 때문에 답을 낼 수 있는 방법이 없다. 컴퓨터의 세계에서는 우리들이 가, 나, 다 순서라든가 A, B, C 순서라고 하는 것과 같이 편의적으로 순서를 정하고 있는 것처럼 비트순으로 대소관계를 정하고 있다.

이 비교 명령은 컴퓨터의 각종의 명령 가운데서도 특히 중요하고, 컴퓨터가 여러가지 경우로 대응하여 처리할 수 있는것은 이 「비교, 교환」의 기능이 있기 때문이며, 앞의 100인분의 반복계산을 「비교하여 그 결과에 따라 다음의 명령을 선택하기」 때문에 가능하게 된 것이다. 데이터 체크에 있어서 조사표 각 항목의 정당성의 판정은 이 비교 명령의 대표적인 예이다.

(ㄷ) 이동에 관한 명령

메모리 X에 있는 문자를 메모리 Y로 이동시키는 명령이다. 형식적으로는 「메모리 X의 문자를 메모리 Y에 옮겨라」라는 표현이 된다. 컴퓨터의 경우 이동은 「복사」와 같은 것이다. 즉 이동 후의 메모리의 상태를 보면, 메모리 X에 있는 문자는 이동 전과 변하지 않지만 메모리 Y쪽의 문자는 메모리 X의 문자와 같게 된다. 예를 들면 입력 데이터인 조사표의 조사항목 순서를 메모리 안에서 바꾸어 넣거나 또는 컴퓨터 처리의 결과를 추가해서 출력하고 싶을때에 이동 명령을 사용한다. 이것은 「계산 명령」, 「비교 명령」과 함께 중심을 이루는 명령이다.

(ㄹ) 편집에 관한 명령

여기에 말하는 편집은 계산결과에 3자리수마다 쉼표를 넣기도 하고 소수점을 넣어 숫자를 보기 쉽게 하는것을 의미한다. 컴퓨터의 처

리에 있어서 제로와 블랭크는 다른 문자이다. 우리들 인간세계의 계산에 있어서는 예를 들면 13에 5을 더하는 것은 $13+5=18$ 이거나 $00013+05=00018$ 이라도 답은 18로 인식하지만, 컴퓨터의 계산에 있어서는 유효숫자 좌측 좀더 정확히 말하면 계산결과에 필요한 자리수는 블랭크가 아니라 제로로 해 두어야 하고, 따라서 계산결과에도 좌측에는 제로가 남아있기 때문에 그대로 답을 출력하면 00018로 인쇄 되어 버린다. 그밖에도 수치가 마이너스인 경우에 "-" 기호의 표시, 혹은 \$나 ¥ 등의 기호 표시도 「편집명령」으로 행한다.

(3) 프로그램

컴퓨터에 어떤 동작을 요구하는 것을 명령이라고 하며, 명령에 대한 대략적인 활동을 살펴 보았다. 이들의 명령을 순서대로(처리하고 싶은 순서대로) 열거한 것을 「프로그램」이라고 한다. 따라서 인간은 프로그램에 의해 컴퓨터와 커뮤니케이션이 이루어지며, 프로그램은 의사소통을 위한 언어로 쓰여져 있다. 그래서 프로그램을 구성하는 언어를 「프로그램언어」라고 한다. 프로그램언어에는 몇개 레벨이 있지만 컴퓨터가 명령을 이해해서 움직이는 것은 「기계어」라고 하는 레벨 만이 가능하다. 이 레벨은 암호에 가까운 문자로 늘어져 있으므로 도저히 인간으로는 이해할 수 없고 프로그램의 작성도 곤란하다. 더구나 컴퓨터의 종류가 다르면 기계어도 다르므로 프로그램의 호환성(공통이용)이 없다.

다음의 레벨로서 프로그램의 작성이 기계어보다 용이하며 인간이 이해하도록 기호로 쓰는 「어셈블러 언어」라든가, 다른 종류의 컴퓨터라도 프로그램을 공통이용 할 수 있는 문장형식으로 쓰는 「컴파일러 언어」가 출현 하였다. 이런 언어는 인간이 이해할 수 있는 반면에 컴퓨터는 이해할

수 없으므로(결국 작동되지 않음) 어셈블러 언어나 컴파일러 언어를 기계어로 「번역」할 필요가 생기며 어셈블러 언어나 컴파일 언어를 다른 기계어로 바꾸어 주는 번역 프로그램을 어셈블러 또는 컴파일러 라고 한다.

앞서 데이터는 각종의 명령에 의해 메모리에 기억되기도 하고 메모리에 있는 내용으로 계산 되기도 하는 것을 설명하였다. 데이터의 길이(자리수)가 10자리수일때 메모리에 기억시키면 10개분의 테두리를 점령한다. 또 계산한 답이 15자리수이면 15개분의 테두리를 점령한다. 프로그램은 명령의 모임이고 번역해서 기계어가 되면 암호와 같은 문자열이 되어서 데이터와 같이 메모리에 기억시킨다. 탁상용 전자계산기의 경우는 명령(+키 라든가 ÷ 키 등)을 인간이 눌러주기 때문에 탁상용 전자계산기 스스로의 판단에 따라 다음의 명령을 고를 수는 없다. 한편, 컴퓨터에 있어서는 명령이 메모리 안에 저장되어 있기 때문에 비교명령에 따른 판단에 의거해서 다음의 명령을 선택하기도 하고 명령 그것을 다른 명령으로 바꾸어 놓기도 하고 혹은 명령의 보조정보(어드레스 부분)를 변경하기도 하는 것을 할 수 있기 때문에 여러가지 조건에 대응하는 계산을 할 수 있어 인간의 지령(이것은 인간의 개입을 의미한다)을 필요로 하지 않는 것이다.

2. 컴퓨터에 의한 통계표 작성방법

간단한 예로 작성한 프로세스를 설명하자. 「표 7」의 都道府縣別, 연령(각세)別, 남녀별 인구수의 통계표를 만들기로 한다. 조사표의 조사항목은 「都道府縣」, 「연령」, 「남녀별」로서 먼저 수작업으로 작성하는 경우의 순서를 생각해 보면 대강 다음과 같이 된다.

표7 통계표의 예

도도부현별연령(각세)별 남녀별인구 전국			
	총수	남	여
총수			
0			
1			
2			
3			
·			
100세 이상			

중현
북해도

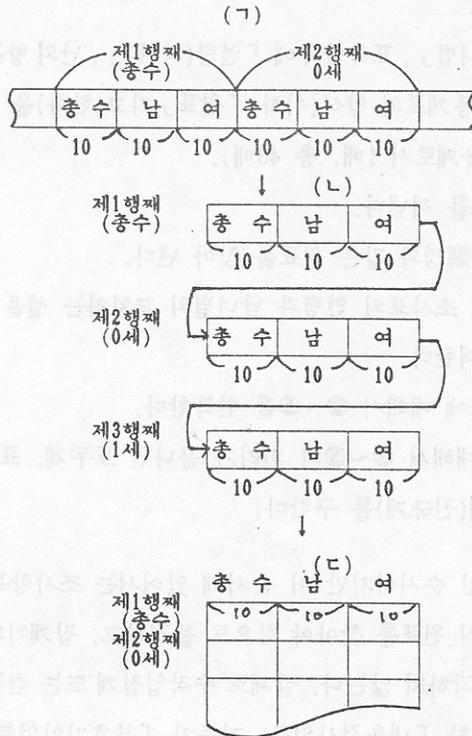
- ① 표두항목에 「남녀별」, 표측항목에 「연령(각세)」, 난의 항목에 「都道府縣명」을 기입한 통계표의 양식(이하 「원표」라고 한다)을 만든다(1縣 1매로서 47매, 전국계로서 1매, 총 48매).
- ② 첫번째의 조사표를 꺼낸다.
- ③ 조사표의 都道府縣명과 같은 원표를 찾아 낸다.
- ④ 찾아낸 원표에서 조사표의 연령과 남녀별의 교차하는 셀을 찾는다.
- ⑤ 그 셀에 "1"을 더한다.
- ⑥ 두번째의 조사표에 대해서 ③~⑤를 반복한다.
- ⑦ 모든 조사표에 대해서 ③~⑤의 처리가 끝나면 표두계, 표측계를 구한다.
- ⑧ 都道府縣의 합계(전국계)를 구한다.

이상이 기본적인 순서이지만 이 순서에 있어서는 조사항목의 오류라든가 해당하는 都道府縣의 원표를 찾아야 함으로 불편하고, 집계치의 크기 즉 셀의 크기(자리수)를 생각하지 않는다. 실제로 수작업집계 또는 컴퓨터로 집계하는 경우에는 이미 기술한 「내용검사업무」라든가 「부호기입업무」를 거치고 나서 통계표 작성작업에 들어가는 것이 되기 때문에 조사항목의 오류라든가 셀의

자리수 등은 고려되어 있다.

그런데 컴퓨터에 의한 통계표의 작성은 어떻게 하는 것일까. ①의 단계는 수작업과 같은 "원표"를 메모리내에 만든다. 메모리는 앞에도 기술한 것과 같이 지정성이므로 원표의 크기에서 그 점유한 자리수를 계산해야 한다(명령과 원표가 점유하는 메모리가 컴퓨터의 메모리 용량을 넘고 있는가 어떤가를 확인하기 위함). 「표 7」은 일본의 인구를 구하고 있으므로 「억」의 단위까지 필요한 것이므로 셀의 크기는 9자리수이다.

그림14 원표 그리는 방법



① 都道府縣의 원표의 크기는 표측이 102행, 표두가 3구분, 1셀을 10자리수(9자리수 이상이면 좋다)로서 $102\text{행} \times 3\text{구분} \times 10\text{자리수} = 3060$ 자리수가 된다. 48매분의 원표를 메모리내에 그리려면 $48 \times 3060 = 146880$ 자리수 필요하게 된다. 여기서 메모리 내에 원표를 그린다 것은 「그림 14(7)」과 같이 메모리는 띠 모양을 이루고 있고 그 연속하는 장소에 제 1행째의 표두 3구분 30 자리를 차지하고, 이어서 제 2행째의 3구분 30자리를 차지하는 식으로 테두리를 잡아간다. 실재는 띠처럼 그려져 있어도 인간이 이해하기 어려우므로 이것을 그림 14(L) → (C)와 같이 변형해서 원표와 같은 형으로 한다. 따라서 최초로 계산한 것 같이 48매분의 원표가 차례차례 늘어서 있는 것이 된다.

②의 단계의 첫번째 조사표를 꺼내는 것은 입력장치로 부터 조사표 데이터(예를 들면 플로피 디스크)를 메모리에 「읽는다」라고 하는 것이다. 컴퓨터는 메모리에만 있는 데이터를 가지고 비교라든가 계산이라든가의 처리를 할 수 있으므로 「읽음」으로서 데이터를 메모리에 넣어 처리할 수 있도록 해야 한다. 읽는 장소(메모리)는 48매의 원표가 있는 메모리와 겹쳐서는 안된다. 그러면 원표의 계산된 데이터를 파손해 버리기 때문이다.

③의 단계의 원표를 찾는 것은 2가지의 방법이 있다. 하나는 메모리내의 원표에 어떠한 눈금, 예를 들면 都道府縣별로 번호를 붙여 놓고 읽은 조사표의 번호와 그것과 일치하는지 「비교명령」을 사용해서 찾는 방법과 또 하나는 하나의 원표가 3060 자리수의 크기를 가지고 있으므로 조사표의 都道府縣 번호와 그 단위수를 「곱셈」해서 첫번째의 원표에서 얼마만큼(메모리의 테두리에서) 떨어져 있는가를 계산하는 방법이다. 북해도(01)의 경우는 $01 \times 3060 = 3060$ 정도 떨어진 메모리에 북해도의 원표가 있고, 암수현(03)에 있어서는 $03 \times 3060 = 9180$ 정도 떨어진 장소에 암수현의 원표가 있다.

④의 단계에서도 교차하는 셀을 찾아내는데 「비교명령」을 사용하는 경우

와 조사표의 숫자코드의 「곱셈」으로 편위를 산출하는 경우의 2종류가 있고, 비교명령은 성별과 같이 코드의 종류가 적은 것에 적용되고 계산에 의한 방법은 都道府縣 연령 등과 같이 코드의 종류가 많은 것에 적용된다.

⑤의 단계에 있어서는 가산해야 할 최종의 메모리 위치가 판명되어 있으므로 「가산명령」을 사용하여 "1"을 더하는 것이다. 다만 모든 원표의 각 셀은 최초의 "1"을 더하기 전에 제로로 해 두어야 한다. 그 이유는 ①의 단계에서 원표의 크기를 계산하고 메모리의 테두리를 확보하였지만 그 테두리의 내용이 제로인지를 보장할 수 없기 때문이다.

⑥의 단계는 입력장치에서 「읽어야」 할 데이터가 없을 때 「읽어라」라는 명령을 받으면 「읽을 데이터는 없다」라는 상태를 표시하고 컴퓨터의 제어장치에도 신호를 보낸다. 그리고 어떤 종류의 명령에 따라 ①의 단계를 실행한다.

⑧의 단계도 ⑦과 같이 합계를 구하는 정도이므로 「가산명령」을 사용해서 셀 일련번호의 계산을 행한다. 이상으로 48개의 원표에 도수분포가 행하여진 것이 된다.

3. 통계표를 위한 기능

간단한 도수분포표를 만드는 예를 설명하였지만 프로그램의 전문가용이 아니라고 하면서 왜 이와같이 자세하게 썼는가 하면 도수분포라고 하는 가장 기본적인 통계표를 구하는데도 모두 명령을 사용하지 않으면 안되기 때문으로 「해당하는 원표를 찾는다」라고 하는 인간의 행위가, 컴퓨터에서는 어떠한 원리로 어떠한 명령의 조합으로 실현할 수 있는가를 나타내고 싶었기 때문이다. 결국 컴퓨터는 지시된 명령을 충실히 실행할 뿐이고 컴퓨터 스스로가 일의 " 좋고 나쁨"을 판단할 수 있는 것은 아니다. 이 점이 인간과 다른 부분이고 인간의 세계에 있어서는 「...적당히 처리한다」라든가 오랜세월의 경험에서 「...인

듯하다」라고 하는 감각으로 판단하고 행동하는 것은 흔히 있는 일이지만 이 행동을 컴퓨터화 할 수는 없다. 흔히 「컴퓨터로 적당히 처리해 다오」라고 하는 것과 같은 터무니 없는 요구도 있지만 「적당」이라든가 「인 듯한다」라는 가 등은 누가 행하였어도 같은 결과가 얻어지는 행동을 발견할 수 없고, 더구나 그것은 제한된 명령으로 논리적으로 표현하는 것이 불가능하기 때문에 컴퓨터로 「모호」한 현상의 처리는 할 수 없다.

따라서 컴퓨터로 어떤 처리를 시키고 싶다면 그 처리의 내용에 대해서 논리적으로 표현 (문장이나 플로우차트 라도)하고, 또한 인간측의 행동을 합해서 전체적인 고찰을 하지 않으면 안된다. 예를 들면 사무의 OA화, 컴퓨터화라고 막연히 말하는 것이 아니라 그를 위해서는 현재 인간이 행하고 있는 사무를 누가 그 사무를 담당하여도 같은 결과가 얻어지도록 절차(수속)를 알기 쉽게 정리하는 일이 필요한 것이다.

한편 통계조사를 행한 후 통계보고서를 간행하기까지 필요한 컴퓨터의 기능을 생각하면 기본적으로는 다음의 5가지로 요약된다.

- ① 데이터 체크
- ② 원표의 작성
- ③ 합계의 산출
- ④ 통계수치의 가공
- ⑤ 보고서 체제(형식) 편집

이 가운데 ②~⑤는 수작업부문과의 관계는 없고, 전문 프로그래머에 의해 프로그램이 작성되며, 데이터를 정돈하면 컴퓨터로 처리시킬 뿐이다. 수작업 부문에 관련되는 것은 ①의 전 단계로 「입력 데이터의 작성」, ①의 작업후 「정정 데이터의 작성」; ⑤의 작업후 「결과표의 심사」이다.

4. 입력데이터의 작성

(1) 입력기기의 종류

데이터 작성을 위하여는 여러가지 원리의 입력기기가 사용되고 있지만 대량의 조사표에서의 데이터 작성을 위해서는 다음의 2가지 장치가 대표적인 것이다.

(가) 키보드를 이용한 장치

타이프 라이터 혹은 요즘의 유행으로 말하면 워드프로세서에 붙어 있는 키보드 상의 문자를 키엔트리하는 것에 의해 데이터를 작성하는 장치이다. 대표적인 것으로서 데이터 엔트리 장치가 있고 이것을 비치한 플로피 디스크에 키엔트리된 데이터가 자기기록되어 진다. 플로피 디스크가 출현하기 전에는 자기테이프나 자기디스크가 있었고 그전에는 종이카드가 사용되었고, 명칭도 「카드 펀치기」라고 불리어져 지 금도 펀치라고 하는 말은 남아 있다. 대량의 데이터를 효율적으로 작성하는 데 적당하다.

같은 모양의 장치로 VDT 장치(Video Display Terminal)가 있다. 데이터 엔트리 장치와 거의 같은 키 보드를 가지고 있고 키엔트리해서 데이터를 작성하는 것은 변함이 없지만, VDT 장치가 컴퓨터의 제어하에 있는 「온라인 방식」으로 가동되는 것이 다르다. 데이터 엔트리 장치는 컴퓨터와는 관계없이 독립해서 가동한다. 이와 같은 가동형태를 「오프라인 방식」이라고 부르고 있다.

(L) 빛을 이용한 장치

조사표의 각 항목을 인간이 순차 키엔트리하는 방법은 단순하지만 장기간을 필요로 한다. 그대문에 대량의 조사표를 사용하는 조사에서는 조사표를 직접 기계가 독해해서 자기 테이프나 플로피 디스크에 기록하는 방식을 채택하고 있다. 이런 경우 독해에 빛의 반사를 이용하므로 「광학식 마크 독해 장치(OMR)」라든가 「광학식 문자 독해장치(OCR)」라고 부르며, 전자는 "-"와 같은 마크를 독해하는 장치이고, 후자는 인쇄한 문자(영자, 가다가나, 숫자) 또는 손으로 쓴 문자를 독해하는 장치이다. 특히 OCR의 손으로 쓴 문자는 기종에 따라 문자의 형이 정해져 있고 그것에 가까운 형으로 쓰지 않으면 독해를 할 수 없는 것이 많다.

OMR, OCR에는 각각 일장일단이 있는데 통계조사라고 하는 범위로 생각해 보면 OMR은 질문의 답을 하나 고르는 택일식 회답인 경우에 적당하며, OCR은 금액, 사람수라고 하는 수치를 기입 하는 경우에 적당하다. 그러나 통계조사의 조사표는 일반적으로는 택일회답과 수치회답이 섞여 있는 것이 보통이므로 OMR, OCR의 양방의 기능을 가진 독해장치가 필요하다. 양기능을 겸비한 장치는 대개 OCR이 기본이고 택일회답 부분의 "-"마크는 마이너스 기호로 대응해서 문자로서 읽지만, OMR기능의 일부(농담비교 - 후술)를 이용할 수 없는것, 독해 속도도 극단으로 저하하는 것, 그리고 가장 큰 문제는 조사표의 수치 회답 부분의 손으로 쓴 문자에 제약이 너무 많고 독해 정도가 떨어지는 것이다. 따라서 누구라도 용이하게 기입할 수 있고 독해정도가 높은 OMR이 이용되고 있다.

(2) 정도의 확보

내용검사 업무로 일정한 정도를 보유한 조사표를 컴퓨터로 처리할 수 있는 데이터로 전환하는 단계 즉 입력 데이터 작성 단계에서 오류를 범하면 아무것도 되지 않는다. 그러나 데이터 엔트리 방식에서는 인간이 그 작업에 종사함으로써 오류를 제로로 할 수는 없지만 최소로 하는 노력, 연구는 필요하다. 그러기 위해서 대개 다음과 같은 점에 유의하는 것이 필요하다.

(7) 조사표의 설계

설계의 기본은

- ① 조사의 목적을 어떠한 조사항목으로 구체화 하는가
- ② 조사항목의 표현, 배치를 어떻게 하는가
- ③ 제표시의 편의함을 어떻게 하는가

이다. 특히 내용검사, 부호기입, 데이터 엔트리 등의 제표업무에 밀접히 관련 있는 것은 ②이다. 조사객체가 통계에 숙지하고 있는 일은 드물고, 대부분은 모르는 사람들이라고 해도 좋다.

조사표에 기입이 용이하다는 것은 조사객체의 심리적 부담을 경감하고, 「기입오류」, 「기입 누락」 등을 적게 할 수 있다. 또 내용검사, 부호기입, 데이터 엔트리 등에도 같은 효과를 기대할 수 있으므로 조사항목의 배치가 가장 중요하다.

기입을 용이하게 하기 위해서는

- ① 관련되는 조사항목을 모아서 배치하고 분산시키지 않는다.
- ② 조사구 번호, 세대번호와 같은 식별항목은 조사표의 상부에 모아서 배치한다.

㉔ 조사항목은 조사표의 좌에서 우로, 상에서 하의 순으로 배치한다.
와 같은 기본원칙을 생각하면서 행한다.

데이터 엔트리를 행하는 경우라도 조사항목이 복잡하게 배치되어 있으면 엔트리 해야할 항목을 빠뜨리기도 하고 또 키 오퍼레이타의 눈의 움직임(좌에서 우, 상에서 하가 자연스러움)에 역행하게 될지도 모른다. 또, 조사표의 인쇄도 중요하다. 통상, 기입은 흑색 연필로 행하여 지므로 그 기입이 질문과 명백히 구별되는 색이 좋다. 일반적으로는 차색이든지 녹색 계통이 된다.

(L) 오류의 발견

데이터 엔트리 방식에서는 베리화이(검공) 하는것이 일반적이다. 베리화이는 엔트리된 데이터를 한번더 반복 키엔트리해서 오류가 있는지 어떤지를 점검하는 방법이므로 키 오퍼레이타는 처음에 키엔트리 했던 사람이 아닌 다른 사람이어야 한다. 왜냐하면 조사표에 혼동하기 쉬운 문자가 있는 경우에 (예를 들면 7과 9 등) 동일 오퍼레이터이면 최초의 판단 이상의 것은 얻을 수 없기 때문이다. 사람에 따라 혼동하기 쉬운 문자의 판단이 다르므로(베리화이 할때에 데이터 엔트리 장치는 이미 엔트리 된 문자와 다른것을 오퍼레이터에게 알려준다) 그래서 어느쪽의 판단이 올바른 것인가 조사표를 잘 보게 되므로 오류를 발견할 수 있는 것이다. 따라서 이처럼 혼동하기 쉬운문자가 없는 조사표에서는 베리화이가 키 오퍼레이터의 단순한 키엔트리 실수를 발견하는데 불과해, 조사표의 기입오류 자체는 키 오퍼레이터가 발견할 수 없다. 그렇지만 데이터 엔트리 방식에서 조사항목마다 프로그램에 의해 체크를 짜넣을 수도 있지만 키 오퍼레이터의 키엔트리 속도와 체크의 속도가 조화를 이루지 않으면 안된다. 예를 들어 체크에

시간이 걸리면 다음의 항목을 키 오퍼레이터가 키엔트리 하고 싶어도 「기다림」 상태가 되어 흐르는 리듬이 중단되어 버리게 된다. 그 때문에 간단한 체크 밖에 짜 넣을 수 없다.

온라인으로 접속되어 있는 VDT 장치에서 데이터를 입력하는 경우는 베리화이 방식을 사용하지 않는다. VDT 장치의 조작은 다른 업무를 하면서 행하는 것이 보통으로, 예를 들면 경리업무에서 전표 처리가 필요할때 VDT 장치를 사용하는 식이다. VDT 장치에서의 입력은 컴퓨터로 부터의 입력지시 화면에 따라서 대화식으로 행하여져 하나의 조사항목을 입력할때마다 입력된 데이터가 올바른지 오퍼레이터의 화면 확인으로 오류를 발견한다.

OMR 방식의 경우는 조사표 자체가 컴퓨터 가독형이므로 바로 독해 장치에 걸어 데이터 입력을 행한다. 거기에는 인간이 데이터를 변환시키는 프로세스는 없기 때문에 독해 장치만의 잘못이 된다. 자기기록된 데이터나 구멍이 뚫린 펀치카드의 기계에 의한 독해는 거의 100%에 가까이 정확하지만 OMR 조사표와 같이 사람이 연필로 기입한 것의 독해는 연필의 농도가 일정하지 않다든가, 마크의 굵기가 가늘게 표시된 것 등 독해조건으로서는 나쁜 것이 많다. 그래서 독해상 가장 많은 경우로서 2개의 마크를 비교해서 진한편의 마크를 채택하는 「농담비교」의 기능을 독해장치에 부가하고 있다. 이것에 의해 지우개의 지우는 방법이 불완전한 경우라도 정확히 읽을 수 있도록 하고 있다. 그러나 이러한 기계적 기능을 부가해도 여러가지 원인으로 정확히 읽지 않는 일도 많으므로 프로그램에 의한 데이터 체크는 반드시 실시한다.

5. OMR 조사표의 취급

OMR 방식은 대량의 조사표를 비교적 단기간에 입력 데이터화 할 수 있는 것으로 입력 방식 가운데에서 가장 실용적인 것이다. 그러나 독해의 원리가 빛의 반사의 변화를 이용하고 있으며, 인간의 시각과는 약간 다르므로 OMR 조사표의 취급에 망설임이 있는 것도 사실이다. 그래서 OMR의 기본적인 기능이나 조사표의 취급에 대해서 보충적으로 설명해 두고 싶다.

(1) OMR 조사표의 물리적 성질

[그림 15]는 OMR 조사표의 예이다. 크기는 최소가 B5판, 최대가 B4판으로 그 사이라면 임의의 크기라도 좋다. 통계국에서의 조사표에 B4판이 많은 것은 조사항목의 수나 그 배치로 인해서 그만큼 크기로 되었을 따름이다. 조사표를 OMR 장치에 이송해서 마크를 정확히 읽기위하여 고려되어야 할 사항 2가지가 있다.

하나는 조사표의 신축의 정도이다. 타이밍 마크와 마크기입 테두리와 의 위치관계는 mm단위 이하의 정도로 결정되어 있으므로 습기나 기계적인 끌어당김에 의해서도 조사표의 신축이 적은 펄프로 만들고 있다. 이 신축이 크면 독해가 부정확하게 된다. 마크기입 테두리는 타이밍 마크에 가까운쪽에 두는 편이 신축의 영향에서 조금이라도 벗어날 수가 있다. 또한 하나는 코너 카트이다. 조사표의 네 모퉁이 모두 둥글게 카트해 두는 편이 각(모난 귀퉁이)이 접혀 지지 않아서 좋다. 그러나 네 모퉁이라도 둥글게 카트하면 조사표의 상하의 역전이나 뒤집음의 구별을 할 수 없으므로 하단의 하나는 경사지게 카트해 둔다. 조사표를 보내주는 방향은 [그림 15]에서 나타내는 것 같이 긴쪽의 방향이다. 긴쪽 방향으로 보내는 것은 주행중의 조사표의 접촉을 적게할 수가 있기 때문이다. OMR 조사표의 인쇄색이 보기

그림15 OMR 조사표의 예

1990년 인구 주택 총조사
(진수조사표)

* 조사 대상 가구 : 1990년 12월 31일 현재 주민등록이 되어 있는 가구
* 조사 대상 인구는 1990년 12월 31일 현재 주민등록이 되어 있는 인구를 기준으로 하며, 외국인 주민은 제외한다.
* 조사 대상 가구 및 인구는 1990년 12월 31일 현재 주민등록이 되어 있는 가구 및 인구를 기준으로 하며, 외국인 주민은 제외한다.

1. 성	2. 성	3. 성	4. 성	5. 성	6. 성
()	()	()	()	()	()
7. 성	8. 성	9. 성	10. 성	11. 성	12. 성
()	()	()	()	()	()
13. 성	14. 성	15. 성	16. 성	17. 성	18. 성
()	()	()	()	()	()
19. 성	20. 성	21. 성	22. 성	23. 성	24. 성
()	()	()	()	()	()
25. 성	26. 성	27. 성	28. 성	29. 성	30. 성
()	()	()	()	()	()

↑ 보내는 방향

마크테두리
타이밍마크

경사진 카트

동근 카트

좋다고 할 수 없는것은 빛에 대한 반응을 둔하게 하기 위해서이며, 인쇄색을 거뭇하게 하거나 진하게 하면 마크라고 오독될 수 있다.

(2) 마크

OMR은 조사표의 백지 부분의 반사광과 흑색 연필의 마크 부분의 반사광이 그 강도가 다른것에 의해 마크의 유무를 판단하고 있다. 그 때문에 마크는 면적이 큰편이 좋으므로 O, V, X는 -에 비해 독해가 현격히 떨어져 사용하지 않는다. 마크의 독해는 OMR 장치에 마크기입행임을 알리는 「타이밍 마크」를 감지하고 나서 행하므로 타이밍 마크가 없는 부분은 독해하지 않는다. 마크하기 위한 필기구로 연필을 사용하는 것은 오류를 고치기 쉽기 때문이며 볼펜, 싸인펜이 불가능하다는 의미는 아니다. 그러나 볼펜, 싸인펜의 경우는 잉크색이 메이커에 따라 다르므로 품목을 지정해야 하고, OMR의 빛의 파장을 조정할 필요가 있으며 또한 기입자가 실수한 때의 정정을 할 수 없다는 것이 치명적이므로 주로 연필을 사용한다.

마크는 정해진 테두리내에 쓰지만 면적이 큰 편이 좋으므로 가는것보다 굵은편이 좋고 또 연한것 보다 진한편이 좋다. 다시 말하면 끝이 둥글게 되어 있는 연필로 슬쩍 한자쓰는것 같이 쓰는 편이 좋다. 색칠 놀이(색칠을 하도록 윤곽만 그려놓은 그림)와 같이 몇번이나 겹치게 써서 검고 윤이 나는것 같은 마크는 인간의 눈에는 검게 보여도 빛의 반사가 잘되기 때문인지 오히려 독해가 나쁘게 된다. 기입시에 테두리를 벗어나는 일은 흔히 있다. 실수한 때에는 지우개로 잘 지운다. 상식적으로 보아서 지운 상태라면 좋다. OMR은 올바르게 기입된 마크와 지운 마크와의 「농담비교」를 해서 진한편의 마크를 우선 해서 독해하는 기능을 가지고 있지만 이

기능은 어디까지나 같은 행에 한해서이다. 따라서 기입행을 벗어난 경우, 동일행에 없다면 지운편의 불충분한 마크를 잃어버리는 일이 있다. 특히 지운뒤에 연필의 흔적이 움푹패여 남아 있으면 이것을 마크로서 읽으므로 주의한다.

연필 흔적의 움푹패임을 남기지 않으려면 HB정도의 연필로 끝을 뾰족하지 않게 또한 책받침을 깔고 쓰는 것이 제일 나으며, 0.5mm의 샤프펜은 가느다르므로 피하는 편이 좋다. 또 정규(자)를 사용해서 제도와 같이 쓰여진 마크도 있지만 기입 테두리내의 마크는 깨끗한 직선이 아니라도 굵기, 진함만 정돈하여 있으면 된다.

(3) 독해 불능

OMR의 어려움은 인간의 시각과 기계의 독해가 다른것이다. 인간측에서 판독한 마크라도 OMR은 독해불능이 되는 것이 있다. 또 거꾸로 마크가 없는데 「마크있다」로 판정 되는 것도 있다. OMR 조사표에는 「더럽히지 말라」, 「구기지 말라」, 「접혀지게 하지 말라」 라는 주의가 반드시 기재되어 있고, 또한 지시대로 표시되면 100% 정확히 읽는다. 수많은 조사표를 보면 독해가 부정확한 원인으로서

- ① 물리적 변형이 있다. 예를 들면 「접힘」, 「주름」, 「뒤블림」 등
- ② 오염되어 있다. 예를 들면 커피, 엽차, 간장, 스탬프 등으로 더러워져 있다.
- ③ 부착물이 있다. 예를 들면 밥알, 호치키스, 클립, 셀로판 테이프 등이 붙어 있다.

등으로 분류할 수 있다. ①에 대해서는 OMR 장치에 손조로이 조사표를 보내주지 않는 상태가 되어 챔(조사표가 OMR 장치의 가운데서 형태가 뭉그러

진것)이 발생한다. 그러나 곤란한 것은 변형의 정도이다. 어느정도 「접혀」 있으면, 또 「뒤블러」 있으면 잼을 발생하는가 하는 것은 개인차가 있어 일정한 기준을 만들기 어렵다. ②에 대해서는 다시 작성하는 것을 행한다. 특히 타이밍 마크 부분의 더러워짐은 OMR의 독해가 잘못될 수 있으므로 다시 작성한다. ③은 잼의 원인이 되므로 다시 작성한다.

(4) 리젝트

OMR 장치에 보내진 조사표는, 엑셉트를 받는 접시와 리젝트를 받는 접시로 갈라져 나온다. 엑셉트 편은 조사표는 독해가 된 것이며 입력 데이터는 작성된다. 다만, 마크를 올바르게 읽었는지 어떤지는 알 수 없다. 한편 리젝트한 편은 조사표는 기계적 원인으로 독해를 할 수 없었던 것이며 입력 데이터는 만들어지지 않는다. 기계적 원인이라고 하는 것은 조사표가 2매 겹쳐 보내졌든가 타이밍 마크가 규정수보다 많거나 적고 그리고 또 (3)에서의 재 원인이다. 리젝트된 조사표의 전매수에 대한 비율은 극히 적지만 데이터가 작성되지 않으므로 원인을 규명후 재 입력이 필요하다.

이상, OMR 조사표의 취급은 일반 조사표에 비해 세세한 신경을 쓸 필요가 있지만 물리적인 변형, 더러움, 부착물의 3가지점에 주의하면 리젝트를 좀더 적게할 수가 있다. 리젝트가 적으면 그만큼 재입력의 조사표가 적게 되므로 업무의 진행이 원활하다. 예상외로 깨닫지 못하는 경우로서 조사표를 마분지 상자의 조사표 상자에 수납할 때에 혹은 비닐주머니에 넣을 때에 조사표의 네귀퉁이를 접어버리는 일이 있다. OMR 조사표의 취급은 그 기준을 정하기 어려운데 있고 따라서 여기에서는 OMR 장치의 특징을 중심으로 기록해 보았다. 다소나마 참고가 된다면 다행이다.

6. 데이터 체크

컴퓨터로 집계를 시작하기전에 반드시 데이터 체크를 행한다. 그 이유는 다음과 같은 오류가 있기 때문이다.

- ① 기입자측의 오류
- ② 데이터 엔트리시의 오류, 광학식 마크 독취기에서의 오류
- ③ 내용검사시의 누락, 부호기입시의 오류

입력데이터(조사표)에는, 가령 내용검사나 엔트리시의 베리화이를 충분히 행하였다고 하지만 이와 같은 각종의 오류가 남겨져 있을 가능성이 있으므로 남아 있는 것을 전제로 컴퓨터를 사용하여 마지막의 「오류의 검출」을 행한다. 이것을 데이터 체크 라고 하지만 컴퓨터에서의 오류의 발견은 체크 조건을 엄하게해서 될 수 있는한 많은 오류를 발견하도록 하고 있다. 그러나 체크의 조건을 너무 엄하게 하면 「과잉 체크」의 상태가 되어 올바른 데이터까지 오류로 간주하여 버린다. 따라서 오류라고 판정되어도 조사표를 점검하면 올바르다고 하는 일도 있다. 이와 같이 결과를 확인만 하는것도 있어, 이것을 「확인체크」라고 부르고 조사표를 점검하는 편이 안전하다고 하는 "회색"의 데이터에 자주 쓰인다.

(1) 데이터 체크의 흐름

데이터 체크에 의해 검출된 여러 데이터는 조사표의 기입내용과 대조되어 올바르게 정정되는 것이 전제로 되기 때문에 여러 데이터의 정정처리를 생각하면 데이터 체크의 흐름은 [그림 16], [그림 17]의 2종류가

있다. [그림 16]은 올바른 데이터와 에러 데이터를 나누어 기록(화일)해 두는 방법이고, [그림 17]은 올바른 데이터와 에러 데이터를 합쳐서 기록(화일)하는 방법이다.

그림16 올바른데이터와 에러데이터를 분리한 플로우

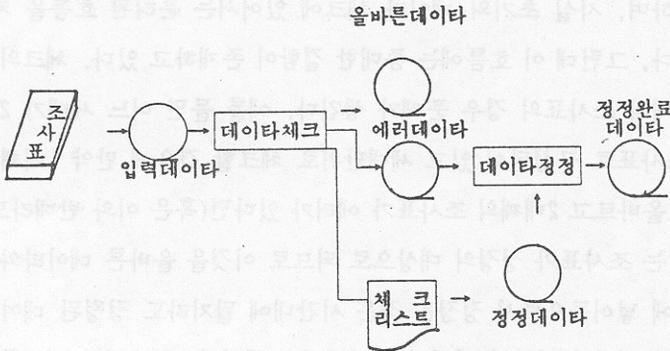
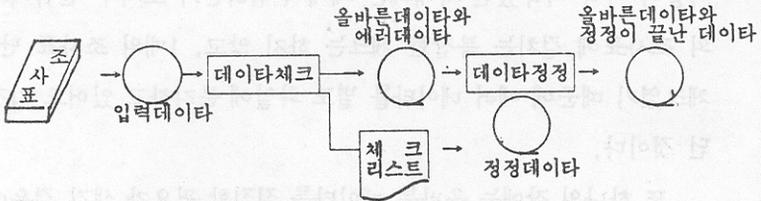


그림17 올바른데이터와 에러데이터를 일괄한 프로우

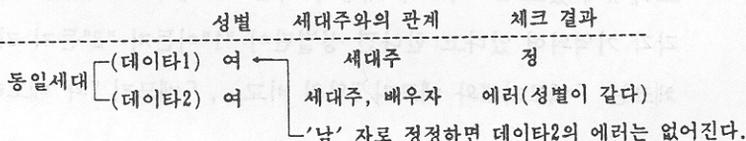


에러 데이터는 올바른 데이터에 비해 양이 매우 적기 때문에 에러 데이터의 정정에 필요로 하는 컴퓨터의 처리시간은 올바른 데이터와 에러데이터가 분리되어 있는 화일편이 합쳐있는 화일편 보다 적다. 그것은 올바른 데이터는 정정할 필요가 없으므로 처리의 대상에서 제외해도 되기 때문이다. 그러므로 올바른 데이터와 에러 데이터를 나누어서 화일해 두는편이 유리하며, 사실 초기의 데이터 체크에 있어서는 분리된 흐름을 채용하고 있었다. 그런데 이 흐름에는 중대한 결함이 존재하고 있다. 체크의 범위가 복수매의 조사표의 경우 문제가 생긴다. 예를 들면 어느 세대가 2매 이상의 조사표로 구성되어 있고 세대단위로 체크할 경우에 만약 1매째의 조사표는 올바르고 2매째의 조사표가 에러가 있다면(혹은 이와 반대라도) 에러가 있는 조사표가 정정의 대상으로 되므로 이것을 올바른 데이터와 분리한 화일에 넣어둠으로서 정정은 적은 시간내에 될지라도 정정된 데이터에 더 이상의 에러가 있는지 어떤지를 컴퓨터로 확인할 때 본래의 세대를 구성하여 (결국 올바른 데이터와 조합시키므로) 체크를 해야 하기 때문에 의미가 없다. 그러므로 에러 데이터가 별도의 화일이 되어 있으면 본래의 세대로 되돌리기 위해서 2개의 화일을 1개로 합쳐야 하므로 불편하다. 컴퓨터의 이용이 아직 미숙했던 시대에는 세대 단위라든가 조사구 단위에서의 복수의 조사표에 걸치는 복잡한 체크는 하지 않고, 1매의 조사표 단위에서의 체크였기 때문에 에러 데이터를 별도 화일에 분리하고 있어도 지장이 없었던 것이다.

또 하나의 장애는 올바른 데이터를 정정할 필요가 생긴 경우에 분리되어 있으면 올바른 데이터를 정정한 뒤에 체크가 필요하고, 에러 데이터를 정정한 뒤에도 체크가 필요하기 때문에 각각의 체크에 대해서 정오 2개의 화일이 작성되어 정정의 순서가 번잡하게 되어 버린다. 여기서 올바른 데

이타를 정정하는 것에 의문을 가질지도 모르지만, 예를 들면 올바르다고 판정된 데이터가 있는 항목을 정정함에 따라 에러가 소멸해 버리는 일이 있기 때문이다. 일반적으로 데이터 체크는 조사표의 순으로 또 조사항목의 순으로 행하여져 정오가 확정이 된다. [그림 18]에서 첫번째 데이터의 내용은 세대주가 여자였어도 이상한 것은 아니므로 「올바르다」로 판정되어 있다. 두번째의 데이터는 세대주의 배우자이면서 성별은 여자로 되어 있기 때문에 컴퓨터는 「에러」로 판정했다. 그래서 에러 데이터의 정정을 위하여 조사표를 점검하면 이름은 여성 이름이며, 일은 가사를 주로하고 있는 등으로 오히려 여성으로 판단한 편이 좋은 경우에는 올바른 편 데이터의 조사표에서 이름, 업무, 혹은 연령으로부터 남성이라고 유추할 수 있다면 올바른 데이터를 남자로 정정한 편이 합리적이다. 여기에서 말하는 올바른 데이터는 데이터 체크에서 에러가 발견되지 않았던 데이터를 말하는 것이지 절대로 오류 없는 데이터라고 하는 것은 아니다. 체크가 불충분하면 에러가 검출 되지않고 올바른 데이터 취급이 된다. 그렇지만 올바른 데이터를 여러번 정정한다는 의미가 아니라 그 필요가 생긴때는 귀찮은 순서를 거치지 않아도 처리를 할 수 있어야 한다는 것이다. 따라서 최근의 데이터 체크에 있어서는 올바른 데이터와 에러 데이터를 하나로 화일하는 방법을 채택하고 있다. 정정 처리할 때에 정정을 필요로 하지 않는 올바른 데이터까지 입력 데이터로서 취급하므로 데이터량은 많게 되지만 최근의 컴퓨터의 성능은 획기적으로 진보한 것이므로 처리속도의 느리고 빠름보다 처리 순서의 간소화에 중점을 두고 있다.

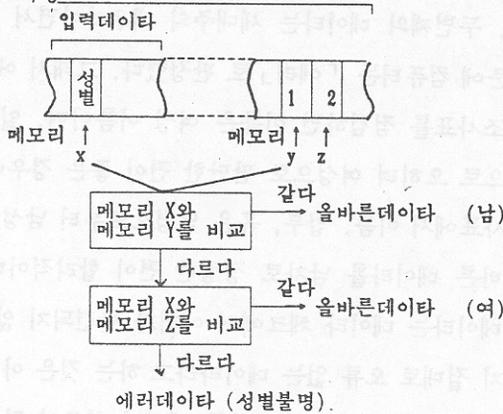
그림18 올바른데이터의 정정 예



(2) 체크의 종류

컴퓨터로 데이터 체크를 행하는 데는 「비교명령」과 「계산명령」으로 구체화하므로 이 두개의 명령에서 체크의 종류는 저절로 결정된다.

그림19 오프코드 체크과정 기억장치내부



(7) 오프코드 체크

입력데이터에는 성별, 국적, 배우관계라고하는 코드에 관한 「코드항목」과 금액, 종업원수 라고하는 수량에 관한 「계량항목」이 있다. 오프코드 체크는 코드항목에 관해서 약속대로의 코드가 붙여져 있는지 어떤지를 점검하는 것이다. 예를 들면 성별란의 남은 "1", 여는 "2"로 약속되어 있다면 "1", "2" 이외는 무기입 (블랭크라고 하는 영문으로 간주한다)을 포함해서 모든것이 오류이다. 이것을 컴퓨터로 행하면 [그림 19]와 같이 입력 데이터가 메모리에 읽어져 성별란이 메모리 X에 있고 또 비교의 대상이 되는 "1"이나 "2"가 메모리 Y, Z에 각각 기억되어 있다고 한다면 성별란이 "1"이든지 "2"든지 라고 하는 체크는 「메모리 X와 메모리 Y와의 비교」, 「메모리 X와 메모리 Z와의

비교」의 2회로 행하여 진다. 따라서 메모리 X의 란이 n개의 코드로 성립되어 있다면 n회의 비교를 하는것이 된다. 예를 들면 산업란에 소분류 코드(3자리)가 올바르게 붙여져 있는지 어떤지라고 하는 체크에 있어서는 소분류 코드는 국세조사에서는 210종류 있으므로 210회의 비교명령을 기억시켜야 한다. 그러나 이렇게 하는것은 매우 비효율적이다. 그래서 앞의 성별의 예를 보면 "2"는 "1"이 있는 메모리 Y를 1자리 오른쪽으로 비켜놓은 장소에 있다. 즉 메모리 Z는 메모리(Y+1)로 표현 된다. 산업소분류 코드 3자리가 메모리 Y에서 오른쪽에 순서대로 기억시키면 1번째의 소분류 코드는 메모리 Y, 2번째의 코드는 Y보다 3자리 앞이므로 메모리(Y+3), 3번째의 코드는 메모리(Y+3×2), 대체로 n번째의 코드는 메모리(Y+3×(n-1))에 기억되어 있다. 따라서 210회의 비교명령을 기억시키는 대신에 입력 데이터의 산업란의 메모리를 X로서, 「메모리 X와 메모리 (Y+3×(n-1))을 비교해라」라고 하는 명령을 하나 기억시켜, n을 1부터 210까지 변경시키면서 이명령을 210회 반복 사용하는 것이다. 그 반복하는 사이에 동등한 상태에 이르면 올바른 데이터이며, 210회 반복해도 동등한 코드가 없는것은 에러 데이터이다. 다만 이 방법은 오프코드 체크의 원리를 나타내기 위하여 예시한것 뿐이며, 실제로 컴퓨터의 프로그램을 기억시킬때에는 「2진 탐색법」이라고 하는 좀더 고도한 방법을 사용한다. 이 방법을 사용하면 210종류의 코드는 최대 약 8회의 비교로 판정을 얻을 수가 있다.

(L) 관련 체크

각 항목의 오프코드 체크에서 에러가 없어도 몇개의 관련하는 항목을 조합시키면 모순을 일으키는 데이터가 있다. 이것을 발견하는 것

이 관련 체크이다. 예를 들면 「연령」과 「학력」과의 관계는 각각 단독항목의 체크에 있어서는 어려가 없어도 그들을 조합시켜 보면 학력을 얻기위한 최단 코스에서 하한의 연령이 판정되어 오프코드 체크에서는 발견되지 않았던 조사항목간의 모순을 발견할 수 있다. 그러나 여기에서 주의할 것은 가령 모순이 발견 되었다해도 모순을 일으킨 조사 항목이 어느쪽에 어려가 있는것인지는 알수 없다.

예를 들면 「연령이 15세」로 「대학졸업」으로 되어있는 경우 연령의 편이 오류가 있는지, 대학졸업의 편이 오류가 있는지는 알지 못한다. 조금더 실마리를 얻기 위하여 관계하는 다른 조사항목 즉, 「배우관계」와 「연령」의 관계를 점검하는 것도 필요하다. 될수 있는 한 많이 관계하는 조사항목을 체크해서 어느 항목이 에러인지 혹은 어려가 농후한지 판정할 수 있게 하고 있다.

관련 체크에서 또하나 주의해야할 점은 에러인지 어떤지 알지 못하는 경우가 있다는 것이다. 조금전의 「연령」과 「학력」과의 관계는 분명히 어느쪽인가의 오류가 있지만, 예를 들면 「취학별(재학생, 졸업, 미취학)」, 「학교구분(유치원, 소중학교, 고교, 대학교 등)」, 「연령」의 관련체크로 「대학에 재학하고 연령 50세」라고 하는 데이터는 현실의 사회를 보면 있을 수 있는 것이므로 에러라고 말할 수는 없다. 그렇다고 해서 무조건으로 올바른 데이터로 간주해 버리는 것은 위험하다. 어쩌면 「연령 50세」든지 「재학」이 틀리게 코딩되어 있는지도 모르는 때문에 이런 종류의 에러 데이터는 확인이 필요한 데이터가 된다. 에러라고 특정할 수 있는 경우라도, 혹은 에러라고는 말할 수 없지만 확인을 필요로 하는 경우라도 최종적으로는 조사표를 점검해서 판단해야 한다.

컴퓨터에 있어서의 처리는 오프코드 체크와 같이 「비교명령」의 조합으로 행한다. 데이터 체크에 있어서는 흔히 「콘스탄트」라고 하는 용어가 사용된다. 이것은 「상수」라고 하는 의미이다. 상수가 있으면 당연히 변수가 있다. 예를 들면 산업 소분류의 체크를 생각하자. 입력 데이터의 산업란은 입력 데이터에 의해 그 코드는 변할 수 있으므로 「변수」이다. 그것에 대해 메모리에 기억되어 있는 210종류의 산업분류의 코드는 입력 데이터의 처리중에는 변하지 않기 때문에 「상수 = 콘스탄트」이다. 산업분류의 코드는 각 都道府縣의 입력 데이터에 대해 같은 콘스탄트를 사용할 수 있다. 즉 산업분류 코드는 미리 프로그램 가운데에 작성해 두는 편이 좋다.

한편 「상주지」와 「근무지(통학지)」의 관련 체크에 있어서는 통근 가능한 지역은 상주지에 의해 바뀌는 것은 보통이다. 따라서 각 都道府縣(상주지)마다 근무지(종업지)를 바꾸어야 한다. 메모리 내에 있는 근무지 코드를 바꾸는 데는 프로그램을 각각의 都道府縣별로 준비하든지, 근무지 코드를 일종의 입력 데이터와 같이 간주하여 메모리에 읽어드려 내용을 바꾸는 수 밖에 없다.

전자의 방법은 프로그램의 변경은 쉽지만, 같은 종류의 프로그램을 47본 준비하는 것이 되므로 사용되지 않는다. 후자의 방법이 일반적이며 프로그램으로 입력 데이터(조사표)를 읽기전에 근무지 코드(콘스탄트)의 데이터를 읽어서 메모리의 내용을 변경하는 것이다. 이때 근무지 코드를 카드에 펀치하는 일이 많으므로 "콘스탄트 카드"라고도 불리어지고 있다.

(C) 범위 체크

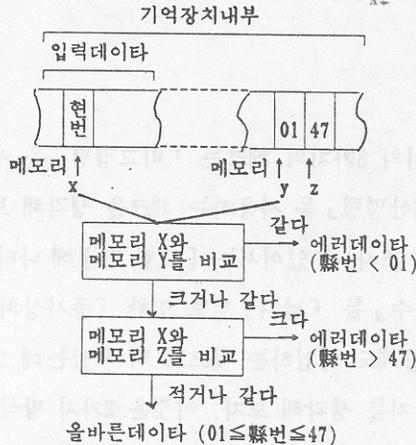
계량항목을 체크할 때에 사용하는 방법이다. 어느 물품의 가격이 전국 통일가격이라면 그 체크는 일종의 오프코드 체크와 같다. 예를 들면 담배의 가격은 품목만 정해지면 전국 같은 가격이므로 코드 항목처럼 체크를 할 수 있다. 게다가 담배의 종류마다 일정한 값이므로 오류인지 아닌지를 알 수 있다. 문제는 가격에 폭이 있는 경우의 체크이다. 지역에 따라 가격에 차가 있고 또 날짜에 따라서도 특매일에 사면 싼것 같이 어떤 폭을 가지고 있는 일은 많다. 가게조사의 가게부의 기입은 그와 같은 상황에 따라 달라진다.

이 경우의 체크는 가격의 상한가와 하한가를 주어서 그 범위내에 있는지 어떤지 점검하는 것이 보통이다. 어느 범위내에 있는지 어떤지 체크한다는 의미에서 「범위체크」라고 부르고 있다. 상한가 하한가에 각각 확정치가 주어지는 경우 어림수치 밖에 주어지지 않는 경우, 혹은 확정치와 어림수치의 조합의 경우라고하는 것처럼 여러 경우가 있고 특히 어림수치가 사용되고 있는 경우에는 그 범위를 벗어난다고 (범위 이외) 에러라고 단정하는 것은 적당하지 않으며, 조사표의 확인을 필요로 하는 데이터라고 판단해야 한다.

범위체크에 있어서 체크하는 값은 상한가와 하한가의 사이에서 연속하고 있는 것이 필요하고 빠지는 것이 예상되는 경우는 올바른 체크가 되지 않는다. 대체적으로 계량 항목은 코드항목과 다르며 올바른 데이터의 범위가 정해져 있지 않는것도 많고 또 그 범위(상한가, 하한가)도 「모호」한 것도 있으므로 체크로 「결정적인 방법」을 빼뜨리지 않고 조금이라도 에러를 찾으려면, 예를 들면 그 계량항목의 단위의 자리수는 「0」이나 「5」이라든지 또는 아래 3자리수는 반드시 「제로」라고 하는 조사에 따른 특징을 적극적으로 이용해야 한다. 오프

코드 체크에서도 하나의 수법으로서 범위체크를 사용할 수 있다. 코드 항목인 산업분류(소분류)는 상당한 부분이 코드가 연속해 있으며, 都道府縣 번호도 01-47까지 연속하고 있다.

그림20 범위 체크과정



컴퓨터로 범위체크를 하는 경우 약간의 주의가 필요하다. 하나의 예로서 都道府縣 번호가 01-47의 범위로 있는지 어떤지 체크할 경우를 생각한다. [그림 20] 과 같이 「비교명령」을 2회 사용한다. 최초는 都道府縣 번호가 「01」보다 큰 (같은것을 포함해서)가를 점검한다. 다음에 「47」보다 적은 (같은것을 포함해서)가를 점검해서 에러 데이터를 찾지만 2자리수 이상의 범위 체크 특히 숫자인 때에는 데이터가 「숫자」인 것을 확인하고나서 「비교명령」을 2회 사용해야 한다. 그 이유는 앞에 기술했지만 컴퓨터의 메모리의 1자리(테두리)는 256종류의 문자를 수용할 수 있으므로 256진법의 표기와 아주 같은 표현이 되기 때문이다. 즉 만약 2자리수의 경우를 생각하면 십자리가 「0」일 때 일자리에는 256종의 문자가 있을 수 있다. 십자리가 「3」일 때

「4」일 때도 같다. 따라서 01-47의 범위에는 「1A」라든가 「2※」라고 하는 조합도 포함된다. 이것은 십자리의 자리수의 대소로 01-47에 포함되는지 아닌지가 정해져 버리기 때문이며 「1A」라든가 「2※」를 에러로 하기 위해서는 都道府縣 코드는 숫자라는 특징을 이용해야 한다.

(근) 합계체크

지금까지의 3가지의 체크는 「비교명령」을 사용한 예이었다. 이번에는 「계산명령」을 사용하는 체크를 생각해 보자, 사업체 통계조사(총사업체 조사)에 있어서는 [그림 21]에 나타내는 것 같이 「사업체의 종업원수」를 「남녀」별로 또한 「종사상의 지위」별로 조사하여 각각의 합계도 기입하는 것으로 되어 있는데 그 합계가 입력 데이터에 필요한지를 생각해 보자. 이것은 2가지 방식으로 나누어 생각해 볼 수 있다. 첫번째는 컴퓨터의 내부 연산이 만들어 내는 숫자를 일부러 입력 데이터로 줄 필요는 없다고 하는 생각이 근저(根底)에 있고 그 때문에 데이터 엔트리에서의 터치수(키를 누르는 회수)를 줄이고, 엔트리하는 문자수를 되도록 적게 한다는 사고방식이다. 터치수가 줄면 엔트리의 능률이 오르고 또 베리화어도 충분히 되어 내역부분의 에러는 거의 없다고 하는 것이 전제로 된다.

두번째는 내역부분의 숫자체크는 오프코드 체크에 비하여 올바른지의 판정이 불충분하기 때문에 합계치를 일종의 감독수로서 사용하고 하는 사고방식이다. 이 2가지의 사고방식 가운데 물론 후자의 편이 비교적 더 좋다. 체크를 위한 정보는 많으면 많을수록 좋기 때문이다. 컴퓨터로 합계와 내역의 수치를 맞춰보는 것을 「합계 체크」라고 부

트며, 「계산명령」을 내역의 개수만큼 사용해서 합계치를 구하고, 다음에 「비교명령」으로 입력데이터의 합계치와 같은지 어떤지를 점검하는 것이다. 계산은 숫자에 대해 행하는 것이므로 비슷자에 대해서 「계산명령」을 사용하면 컴퓨터는 일반적으로 프로그램의 실행을 중단한다.

따라서 숫자라는 보증이 없는 데이터에 대해서는 숫자인 것을 확인하고 나서 「계산명령」을 사용한다. 숫자인것의 확인은 1자리씩 0에서 9까지의 숫자인가를 「비교명령」을 사용해서 점검할 수 밖에 없다. 이와 같이 합계와 내역이 일치하는지 어떤지 체크한다고 하는 간단한 것이라도 컴퓨터로 처리하려면 프로그램 언어상의 제약을 생각하면서 명령을 기억시켜야 하므로 프로그램 작성은 간단하지 않다.

그림21 합계체크의 예

11. 출근근수 출근지 및 퇴근시간		12. 연건출근액 출근액, 퇴근액 포함		13. 연건 주요성인의 출근시간, 퇴근 시간 및 일반관리비중에서 비용성인인 소요된 비용.	
		(1) 12월 한 해 총합인수		출근지, 퇴근지, 연건출근액, 연건퇴근액	
		계	남	계	남
1. 퇴근지	출근지				
2. 퇴근지(출근지 제외)	출근지				
3. 퇴근지(출근지 제외)	출근지				
4. 퇴근지(출근지 제외)	출근지				
5. 퇴근지(출근지 제외)	출근지				
6. 퇴근지(출근지 제외)	출근지				
합 계					
7. 퇴근지(출근지 제외)					
8. 퇴근지(출근지 제외)					
9. 퇴근지(출근지 제외)					
10. 퇴근지(출근지 제외)					
11. 퇴근지(출근지 제외)					
12. 퇴근지(출근지 제외)					
13. 퇴근지(출근지 제외)					
14. 퇴근지(출근지 제외)					
15. 퇴근지(출근지 제외)					
16. 퇴근지(출근지 제외)					
17. 퇴근지(출근지 제외)					
18. 퇴근지(출근지 제외)					
19. 퇴근지(출근지 제외)					
20. 퇴근지(출근지 제외)					
21. 퇴근지(출근지 제외)					
22. 퇴근지(출근지 제외)					
23. 퇴근지(출근지 제외)					
24. 퇴근지(출근지 제외)					
25. 퇴근지(출근지 제외)					
26. 퇴근지(출근지 제외)					
27. 퇴근지(출근지 제외)					
28. 퇴근지(출근지 제외)					
29. 퇴근지(출근지 제외)					
30. 퇴근지(출근지 제외)					

(□) 검사 숫자에 의한 체크

데이터 엔트리시의 오류를 발견하는 유력한 방법이다. 이것은 코드(숫자)로 어떤 종류의 계산을 행해서 「검사숫자」라고 불리는 1자리수의 숫자를 끌어내어, 그것을 본래의 코드의 말미(끝)에 더해서 새로운 코드로 만드는 것이다.

계산의 방법에 따라 Modules(모듈) 10, 모듈 11의 2종류가 있지만 모듈 11로 그 계산법을 설명한다. 예를 들면 코드 「1987」의 검사숫

자는 [그림 22]의 ①~④의 순서에 의해 「9」이다. 이 9를 최초의 코드 「1987」의 말미에 붙여 「19879」를 새로운 코드로 해서 사용한다. 따라서 부호기입업무에 있어서 4자리수 코드 대신에 검사숫자를 더한 5자리수 코드로 기입을 하여 데이터 엔트리를 행한다. 컴퓨터에 의한 체크로는 앞에 ①~④의 계산을 행하고 말미의 검사숫자와 같은지 어떤지 비교해서 에러를 찾는 것이다.

그림22 검사 숫자 계산 예

$$\begin{array}{lcl}
 \text{①} & \left[\begin{array}{cccc} 1 & 9 & 8 & 7 \\ \times & \times & \times & \times \\ 5 & 4 & 3 & 2 \end{array} \right. & \begin{array}{l} \leftarrow \text{원래의 코드} \\ \\ \leftarrow \text{승수} \end{array} \\
 \text{②} & \begin{array}{c} \downarrow \downarrow \downarrow \downarrow \\ 5+36+24+14=79 \end{array} & \leftarrow \text{곱의 합} \\
 \text{③} & \overline{79 \div 11 = 7 \text{ 나머지 } 2} & \\
 \text{④} & 11 - 2 = 9 & \leftarrow \text{검사 숫자} \\
 \text{⑤} & 1987 \boxed{9} & \leftarrow \text{검사 숫자를 붙여} \\
 & & \text{새로운 코드를 만듦(5자리)}
 \end{array}$$

오프코드 체크에 있어서는 약속된 코드에서 벗어난 것이 에러인 것은 자명하지만 약속된 코드라고 해서 「올바르다」라고하는 보증은 없다. 예를 들면 올바른 코드 내에서 데이터엔트리를 잘못하는 경우로 성별란에 「1」로 엔트리 해야 할것을 「2」로 엔트리 할 수 있기 때문이다. 그러나 검사숫자 기입코드의 경우는 다르다. 검사숫자를 포함해서 어느 자리수를 엔트리 오류라 해도 컴퓨터의 계산과 맞지 않게 된다. 상당한 우연이 겹치지 않는 한은 「올바른」 엔트리를 보증할 수 있는 것이다. 검사숫자는 1자리수인데 「나머지」의 수치에 따라서는 2자리

수의 값이 얻어질 수 있다. 그때는 아래의 1자리수를 검사 숫자로 하면 좋다. 예를 들면 「나머지」가 「1」의 경우는 $11-1=10$ 으로 검사 숫자는 「0」으로 하고 나머지 없이 딱 나누어 떨어진 경우는 나머지는 「0」이므로 $11-0=11$ 로 검사숫자는 「1」로 한다.

엔트리된 숫자만의 체크로 정오의 판정을 할 수 있는 이 방법은 우수하지만 이것을 「계량항목」의 금액이나 수량에 적용하려고 하면 잘되지 않는다. 그 이유는 코드 항목에서는 코드의 종류를 알고 있기 때문에 검사숫자를 미리 계산해 둘 수가 있고, 그것을 부여한 코드로 기입을 하면 되지만, 계량항목에서 출현하는 숫자는 여러가지 값을 가질 수 있기 때문에 미리 검사숫자를 계산해 둘 수 없는 것과 또 기입은 객체측이 행하고 검사숫자의 산출까지는 의뢰할 수 없기 때문이다.

(H) 순서 체크

입력 데이터가 어느항목에 관해서 일정한 순서로 되어 있는지 체크하는 것이다. 대체로 조사표는 市區町村번호, 조사구번호, 세대번호(조사표 번호)의 순으로 정돈하여 송부(통계국)하지만, 입력 데이터가 작성되기까지의 각종 수작업에 의한 작업을 거치므로 이 순서가 올바르게 보증을 할 수 없다. 특히 OMR을 이용한 조사에 있어서 OMR 개개 조사표에는 조사구 번호의 기입란은 있어도 마크 기입란은 없다. 어느 조사구에 속하는가는 조사구 번호를 마크한 「조사구 아이덴트(시트)」를 맨앞에 두는 것에 따라 행하고 있으므로 이 아이덴트의 마크 오류나 열거순의 오류는 개개 조사표의 체크에 영향을 주는 것이다. 그래서 개개 조사표의 (7)~(10)의 체크에 앞서서 순서 체크만을 행하고, 세대번호 레벨까지는 올바르게 하는 보증을 해두는 것이다.

더우기 데이터 체크를 행하기 전에 소트(열거 순을 정돈하는 것)를 해서 순번을 정하는 것은 좋지 않다. 예를 들어 조사구번호 순으로 소트해 버리면 조사구 번호가 오류 마크 되었을때, 터무니 없는 경우로 열거되어 그것이 어느 조사구 번호의 오류였는지 또 그 조사구를 찾는 것이 곤란하게 되어 버린다. 반드시 소트해야만 하는 때에는(예를 들면 사업체 통계조사) 최초의 입력 데이터 순의 번호를 컴퓨터로 붙이고 나서 행한다. 그렇게 하면 입력 데이터 순이 실마리가 되어 해당하는 조사구를 찾을 수가 있다.

(스) 대상구분 체크

입력 데이터의 기본적인 속성인 데이터의 종류, 조사년월, 都道府縣 번호 등을 컴퓨터에 주어 체크하는 방법이다. 대상구분은 「콘스탄트」로 주든지 컴퓨터 오퍼레이터로 부터 입력을 받든지 상관없다. 예를 들면 都道府縣 단위로 처리하는 경우 하나의 체크 프로그램으로 47의 縣을 처리하므로 현 번호가 01 ~ 47에 있는가라는 체크는 「소쿠리」와 같이 구멍이 뚫여 있으므로 불완전하다. 지금 처리하는 縣이 「02」이면 「02」 이외는 오류가 되기 때문이다. 01~47의 체크에서는 이 종류의 오류는 발견할 수 없다. 이것은 또 매월의 조사에서도 마찬가지이다. 처리하는 월을 컴퓨터에 주는 체크를 하지 않으면 매월분의 입력 데이터를 처리해도 그 오류를 발견할 수 없다.

이상 몇개의 체크방법에 대해서 설명 하였지만 최적의 방법은 없다. 조금이라도 효과가 있다고 생각되는 방법을 조합하거나 데이터가 가진 특징을 이용해서 여러 데이터를 포착하려는 노력이 필요하다.

7. 에러 데이터의 정정

에러 데이터는 에러의 종류가 발생하고 있는 조사항목에 기호로 붙여져 인 쇄되며 이것을 「체크리스트」라고 부르는 것은 이미 기술하였다. 에러 데이터 의 정정은 2가지의 단계로 나누어져 있다.

첫번째 단계는 수작업에 의한 정정으로 정정데이터를 작성해서 배치방식으 로 행하는 방법과 단말장치를 사용해서 온라인 방식으로 행하는 방법이 있고, 둘째단계는 컴퓨터에 의한 정정(자동정정 혹은 자동수정)이다.

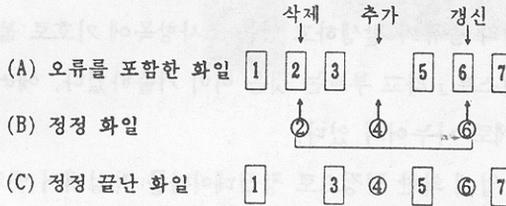
(1) 수작업에 의한 정정

(ㄱ) 배치 방식에 의한 정정

원칙적으로 조사표를 거슬러 올라가 정정 데이터를 작성한다. 이것에 대해서는 역시 앞에서 기술했으므로 반복하지는 않지만, 컴퓨터가 정정 데 이터를 어떤 구조로 에러 데이터에 작용해 가는가를 생각해 보자. 정정 데 이터의 종류에는 「갱신」, 「삭제」, 「추가」의 3가지가 있고, 그것을 받아서 컴퓨터의 정정 기능도 「갱신」, 「삭제」, 「추가」의 3가지가 있 다. 간단한 예를 [그림 23]에 의거해서 설명한다.

(A)는 데이터 체크에 의해 에러로 지적된 데이터와 올바른 데이터를 하나로 정리한 화일이다. 화일 (A)에 대해서 「2의 데이터는 불필요하므로 삭제하고 싶다」, 「4의 데이터를 추가하고 싶다」, 「6의 데이터는 6으로 바꿔 놓고 싶다」라고 하는 정정을 하기 위해서, (B)의 화일을 작성 하였다. 컴퓨터는 이 (A), (B)의 화일로 부터 스텝 1~스텝 8의 처리를 거 쳐서(C)의 정정 완료 화일을 작성한다.

그림23 정정처리 방법



	스텝1	스텝2	스텝3	스텝4	스텝5	스텝6	스텝7	스텝8
화 일	(A) (B) 7							
	6	6	6	6	6	6	6	6
	5	5	5	5	5	5	5	5
	3 ⑥	3 ⑥	3 ⑥	3 ⑥	3 ⑥	3 ⑥	3 ⑥	3 ⑥
	2 ④	2 ④	2 ④	2 ④	2 ④	2 ④	2 ④	2 ④
메 모	1 ↔ 2	2 ↔ 2	3 ↔ 4	5 ↔ 4	5 ↔ 6	6 ↔ 6	7 ↔ X	X ↔ X
정정 화 일	(C) 1	삭제	3	④ 추가	5	⑥ 갱신	7	완료

먼저 컴퓨터는 스텝 1의 앞에 (A), (B)의 각 화일로 부터 첫번째의 데이터를 메모리에 기억 시킨다. 메모리의 가운데서 (A)의 데이터와 (B)의 데이터를 비교해서

- ① (A) < (B)라면 (A)의 데이터를 (C)에 기록하고, (A) 화일에서 다음의 데이터를 메모리에 읽어 드린다.
- ② (A) > (B)라면 (B)의 데이터를 (C)에 기록하고 (B) 화일에서 다음의 데이터를 메모리에 읽어 드린다.

③ $(A) = (B)$ 일때 (B) 데이터의 삭제를 지시하고 있다면 (A), (B)의 양 화일에서 다음의 데이터를 메모리에 읽어 드린다.

(B)의 데이터로 「갱신」을 지시하고 있다면 (B)의 데이터를 (C)에 기록하고 (A), (B)의 양 화일에서 다음의 데이터를 메모리에 읽어 드린다.

④ 어느쪽이든지 한편의 화일이 끝이났다면 또 한편의 화일이 끝날때 까지 (C)에 기록한다.

의 규칙을 적용해서 스텝 1~스텝 8을 처리한다. 우선 스텝 1에서 $(A) < (B)$ 라면 1이 (C)에 기록된다. 스텝 2에서 $(A) = (B)$ 로 삭제라면 각 화일에서 다음의 데이터가 메모리에 읽어드려 진다. 스텝 3에서는 $(A) < (B)$ 이므로 3이 기록된다. 스텝 4에서는 $(A) > (B)$ 이므로 ④가 기록된다. 스텝 5는 스텝 3과 같다. 스텝 6은 $(A) = (B)$ 이지만, 「갱신」이므로 ⑥의 편이 기록된다. 스텝 7에서 (B)가 끝나 있음으로 (A)의 나머지의 데이터를 모두 (C)에 기록한다. 그런데 이 처리에 있어서 고려해야 할 문제가 3가지 있다. 첫번째 열거순이다. (A), (B)의 각 화일의 순서는 「승순(번호의 작은 것에서 큰 순으로)」으로 정렬되어 있어야 한다. 그렇지 않으면 바라는 것과 같은 정정을 할 수 없다.

예를 들면 화일 (A)가 1, 3, 2 ... 7 로 늘어서 있으면 ②의 삭제는 움직일 수 없게 된다. 이것은 3과 ②를 비교하면 ②의 편이 작기때문에 규칙에 의하면 ②가 (C)에 기록된다. 일반적으로 순서는 市區町村번호, 조사구번호, 조사표 번호의 번호순으로 열거하는 것이 보통이지만(조사표의 승부가 이 순서로 되어 있으므로) 이들의 항목이 올바

르게 입력 (엔트리의 오류, 조사표의 취급의 오류 등)되었다는 보증을 없으므로 이것을 키 항목(2개의 데이터를 비교할 때에 사용하는 항목)으로 할 수는 없다. 또 소트를 해서 순번으로 정돈할 수도 없기(이유는 전술) 때문에 입력 데이터에 일련번호를 컴퓨터에 의해 부가해서 이것을 키 항목으로 하는 것이다.

두번째는 「추가(삽입)」 조사표의 취급이다. [그림 23]의 정정의 로직에 의하면 추가(삽입) 데이터는 정정 데이터 화일의 도중에 위치해 있어야 한다. 그러나 실제의 정정 데이터의 작성은 체크리스트를 보면서 행하며 체크리스트 상에 정정해야할 부호를 기입해서 그것을 데이터 엔트리 하고 있다.

따라서 조사표가 탈락해 있으면 체크리스트에는 인쇄되어 있지 않음으로 탈락한 조사표를 찾아서 그 내용을 체크리스트의 추가하고 싶은 위치에 이기해서 그것을 다른 정정 데이터와 같이 엔트리 하든지 혹은 추가조사표 만을 정리해서 데이터 엔트리하고 별도의 처리절차에 의해 데이터 체크를 해야한다. 이와 같이 조사표의 추가는 논리적으로는 간단해도 실제의 작업을 생각하면 귀찮은 일이 많으므로 정정처리를 설계할때 유의 해야 한다.

세번째 문제는 정정처리에서 조사표의 내용에 대해 보증하고 있지 않는 것이다. 「정정」이라고 하는 말로 조사표가 「올바른」 상태로 고쳐지고 있다고 생각하는 것은 착각이다. 단지 데이터를 바꾸었을 뿐이므로, 정정 데이터나 추가조사표의 내용에 대해서는 다시 컴퓨터로 데이터 체크를 해야 한다. 배치 방식에서의 정정은 엔트리 방식이건 마크시트 방식이건 정정 데이터를 입력한 화일 ((B)에 해당)로 작성해야 한다. 예러가 있는 조사표를 점검하면 오류있는 조사항목은 하나이

든지 기껏 돌이며, 그밖에 조사항목은 올바른 경우가 대부분이다. 정정의 종류 가운데 「갱신」은 이전에는 조사표 단위로 「갱신」을 하고 있었다. 이것은 정정하지 않아도 좋은 올바른 조사항목까지 엔트리하는 것을 의미하며, 터치수의 증가나 엔트리 오류에 의한 새로운 에러의 발생을 초래하고 있었다. 그래서 बै치 방식에 의한 에러 데이터의 정정을 표준화해서, 현재로는

- ① 오류가 있는 조사표를 포함해서 입력 데이터에 컴퓨터로 일련번호를 붙인다.
- ② 각 조사항목을 번호로 식별한다.
- ③ 일련번호와 정정하고 싶은 조사항목의 번호에 계속해서 올바른 부호를 엔트리 한다.
- ④ 컴퓨터는 일련번호나 조사항목의 번호에 의해 올바른 부호로 바뀌 놓는다.

처리방법으로 하고 있다. 이것에 의해 정정하고 싶은 조사항목만을 엔트리하면 되므로, 터치수는 격감하고 또 안해도 되는 오류를 방지할 수 있게 된다.

(L) 온라인 방식에 의한 정정

체크리스트상에 올바른 부호를 기입하는 곳까지는 बै치방식과 같다. 다른 체크리스트를 엔트리부문으로 송부해서 정정데이터를 작성하는 것이 बै치방식, 거기에 대해 체크리스트에 의거하여 단말장치를 조작해서 오류의 데이터를 직접 화면상에 불러내어 오류 있는 항목을 정정하는 것이 온라인 방식이다.

백치 방식의 데이터 정정 프로세스에서는 2가지 문제점이 있다.

첫째 문제점은 오류의 내용을 파악하기 위하여 조사표에 거슬러 올라가는 것이 필요하다. 조사표의 기입 상황을 보는것은 단순한 오류(예를 들면 엔트리에서의 오류, 부호기입의 오류 등)을 대부분 정정할 수 있고, 또 기입자 측의 이해부족으로 생긴 오류(예를 들면 가사를 업무라고 해석 하고 있다)도 다른 항목의 기입상황으로 부터 판단할 수 있으며, 체크리스트는 입력 데이터 리스트이므로 조사표의 기입상황을 정확히 반영하지 않기 때문이다. 그래서 조사표의 창고로부터의 반출, 또한 해당하는 조사표의 꺼냄 등의 노력이 필요한 것이다.

두번째 문제점은 데이터가 정정이 되어도 오류가 있는 조사표에서 에러가 없어졌는지의 여부를 모른다는 것이다(이것은 앞에 기술했다). 이 문제는 데이터 정정과 데이터 체크가 나누어져 있기 때문에 발생하는 것이고, 데이터 정정 고유의 문제라고 하는 의미는 아니다.

백치방식에서는 프로그램의 실행을 효율적으로 하기 위하여 데이터 정정과 데이터 체크를 나누어 처리하고 있기 때문이다.

한편, 온라인 방식에서는 전술한것과 같이 단말장치의 화면상에 오류가 있는 조사표 데이터를 불러내어 정정하는데, 이때의 정정은 배치방식에서의 표준화한 방법과 마찬가지로 조사항목 단위로 행하여진다. 게다가 정정데이터는 인간이 눈으로 확인하고 있으므로 신뢰할 수 있다. 온라인 방식에서는 그 처리 형태에서 오류가 있는 항목만을 정정할 뿐만아니라, 또한 진일보해서 정정된 데이터(조사표)를 즉시 체크해서 새로운 에러의 발생여부를 정정시점에서 알 수 있고, 새롭게 발생한 에러의 해결을 위하여 조사표의 기입상황이 화면상에 나오게

되므로 조사표의 반입·반출을 일체 하지 않아도 된다.

온라인 방식의 데이터 정정에 있어서는 여러 조사표를 화면에 불러내어 데이터를 정정 체크해서 결과를 안다고 하는 프로세스가 되고, 화면과 대화하면서 여러 수정을 추구해 간다는 점에서 감각의 처리가 된다. 그러나 현상에 있어서는 아직 실현하고 있지 않지만 최근의 기류의 진보를 생각하면 실현은 시간의 문제가 아닐까라고 생각한다.

(2) 컴퓨터에 의한 정정

수작업에 의한 정정을 수회 반복해서 여러 데이터가 극히 소량이 된 시점에서 컴퓨터에 의해 정정하는 것을 「자동수정」이라고 부르고 있고, 다음과 같은 4종류의 「자동수정」 방법이 있다.

(7) 정치(定値)에 의한 자동수정

조사의 내용, 결과의 경향등에서 일정한 값으로 자동수정하는 방법이다. 간단한 예로서 각 조사항목에 「미상」을 만들어 놓고, 여러 데이터를 전부 「미상」으로 처리하는 것이다. 일정치로서 「최빈수」, 「평균치」, 「중위수」 등이 있는데 어느것을 채택하는 가는 각각의 조사성격에 따른다. 또 「자동수정」의 정도를 높이기 위하여 조사 데이터를 유사한 그룹으로 정리해서 여러 데이터를 찾으면 그것과 유사한 그룹에서 정치를 꺼내 자동수정하는 방법도 병용한다.

(L) 택일치(擇一値)에 의한 자동수정

예를 들면 몇개의 코드 가운데서 하나를 상황에 따라 선택해서 자동수정하는 방법이다. 자동수정하는 코드가 일정하지 않는 경우가 (7)과 다른 점이다. 간단한 예로서 「남녀별」의 부호는 1(남), 2(여)

의 2개가 있고, 남녀 비율은 거의 반반이므로 「남녀별」의 예러는 번갈아 1,2식으로 붙이는 것이다. 무엇보다도 이 방법은 최후의 수단이고 관련하는 다른 조사항목이나 수작업에 의한 정정을 할때에 조사표를 보고 성명에서 정정되는 것이 대부분이다. 또, 이 방법을 확장해서 N개의 부호에서 이루어지는 조사항목에 예러가 있는 경우에는 그 항목의 과거의 집계자료에서 N개의 부호의 구성 비율을 계산해서 그 비율에 따라서 부호를 할당하는 것도 있다.

(ㄷ) 추계치에 의한 자동수정

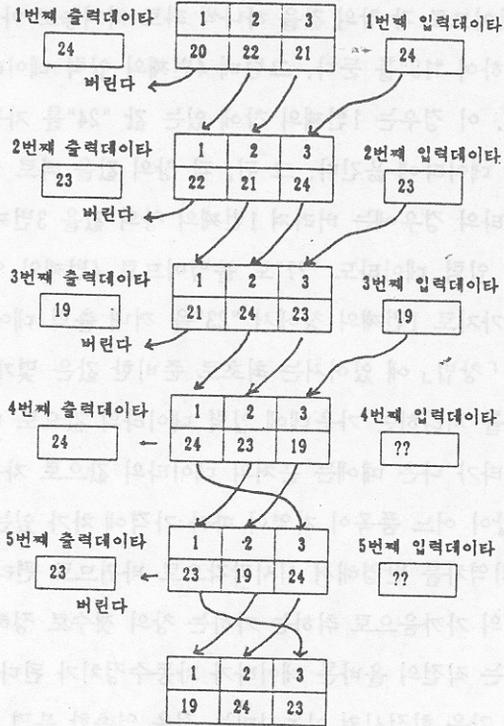
올바른 데이터에서 조사항목간의 관계를 점검하여 추계식으로 만들고, 거기에 적용시켜 자동수정치를 구하는 방법이다. 추계식은 전회 조사의 올바른 데이터를 사용해서 미리 만들어 컴퓨터에 입력해 들 필요가 있고 또 정도를 높이기 위해서도 유사 그룹별로 적용하는 등 고려해야 한다.

(ㄹ) 창법(窓法)에 의한 자동수정

자동수정치를 과거의 조사결과에서의 경향에 따라 결정하는게 아니라 지금 현재 처리중인 데이터에서 경향을 알고 자동수정치를 구하는 것이다. 시시각각으로 값이 변화하는 것에서 「핫 데크법」이라고도 불린다(이것에 대해 (ㄱ)의 정치 자동수정은 「콜드 데크법」이라고 한다). 처리중의 올바른 값을 보존하는 장소를 「창」이라고 부르고 있고(통계국·통계센터에서의 부르는 방법), 예러 데이터를 발견하면 그 「창」에서 자동수정치를 꺼내므로 「창법에 의한 자동수정」이라고 한다. 이 방법도 정도를 높이기 위하여 조사항목의 여러가지 속성에서 유사한 그룹을 만들고, 예러 데이터의 자동수정은 그 그룹의

창에서 꺼내도록 하고 있다.

그림 24 「창법」의 예



「창법」의 사고 방식을 나타내는 예를 들면 [그림 24]와 같다. 가령, 어느 품목의 가격을 자동수정하는 경우에 만약 창을 3개 준비한다면, 그 창에는 가격이 불명한 때에 자동수정 해야할 값을 넣어 둔다. 우선 최초의 입력 데이터 "24"는 올바르기 때문에 그대로 출력하고 각 창의 값을 좌로 하나 비켜 놓고 3번째의 창에 올바른 값을 "24"를 둔

다. 결국 올바른 데이터라면 창의 값을 좌로 하나씩 비켜놓고 마지막의 창을 올바른 값으로 채워준다.

2번째의 입력 데이터의 "23", 3번째의 입력 데이터의 "19"는 올바른 값이므로 각 창의 값을 하나씩 좌로 비켜놓고 마지막의 창에 "23"에 계속하여 "19"를 둔다. 그런데 4번째의 입력 데이터는 "??"으로 불명하다. 이 경우는 1번째의 창에 있는 값 "24"를 자동수정 값으로 해서 출력 데이터에 옮긴다. 그 뒤, 각 창의 값을 좌로 비켜 놓지만 올바른 데이터의 경우에는 버려져 1번째의 창의 값을 3번째의 창에 둔다. 5번째의 입력 데이터도 "??"로 불명이므로 4번째의 입력 데이터의 때와 마찬가지로 1번째의 창에서 "23"을 꺼내 출력 데이터에 옮긴다.

「창법」에 있어서는 최초로 준비한 값은 몇개의 올바른 입력 데이터를 처리하는 가운데에 입력 데이터의 값으로 바뀌어 지므로 불명 데이터가 나온 때에는 근처의 데이터의 값으로 자동수정 된다. 이 예와 같이 어느 품목이 지역이 따라 가격에 차가 있는 경우에는 창의 값이 지역차를 반영해서 시시각각으로 바뀌므로 편리하다. 근처를 어느 정도의 가까움으로 취하는 가하는 창의 갯수로 정해진다. 창이 하나인 경우는 직전의 올바른 데이터가 자동수정치가 된다. 불명 데이터일 때 창의 값을 회전시켜 이동시키는 것은 연속한 불명 데이터에 같은 값이 자동수정되는 것을 막기 위함이다.

8. 체크 요령서 작성

「체크요령」이라는 것은 조사항목의 에러를 컴퓨터에 의해 논리적으로 발견하는 순서를 기술한 것이다. 체크요령에는 분명한 에러뿐만 아니라 「에러일

지도 모른다」라고 하는 판정이 곤란한것도, 또 에러의 경우에 컴퓨터에 의해 정정하는것 까지 포함하고 있는 것이 보통이다. 체크요령을 쓸때는 조사의 내용을 잘 알고 있는 것은 물론, 수작업에 의한 내용검사 업무 가운데 논리적으로 검사를 할 수 있는 부분을 컴퓨터로 대행하려고 하는 의미이므로 내용검사 업무에 정통한 사람이 적당하다. 체크요령이라고 하면 바로 컴퓨터와 결부시켜서 컴퓨터 부문의 사람(시스템 아날리스트, 프로그래머 등)이 작성하는 것이 좋다고 생각하기 쉽지만, 컴퓨터를 잘 아는지의 여부보다는 컴퓨터에서의 순서는 생각하지 않고 사람이 각 조사항목의 체크를 어떠한 순서로 행하며 바른 결과를 무엇인가를 명확하게 제시할 수 있어야 한다. 시스템 아날리스트, 프로그래머라는 컴퓨터 부문의 사람들이 인간세계의 체크순서를 컴퓨터 세계의 체크순서로 잘 번역한다고 해서 수작업부문의 제표 업무에 반드시 밝다고는 할 수 없기 때문이다.

예를 들면 조사구내의 조사표에 중복(조사표 번호가 같은것)이 있는지 여부를 체크하는 경우 인간세계의 체크요령으로는 「조사구내의 조사표에 중복이 있는지. 만약 있으면 그 조사구내의 모든 조사표를 인쇄한다」라고 쓸것이다.

이것을 컴퓨터로 실현하기 위해서는 체크 프로그램의 구조(만드는 방법)에도 의존하지만 조사구 내의 조사표를 전부 메모리에 저장하고 나서 조사표 번호의 중복을 점검하든지 만약 저장할 만큼의 메모리가 없다면 조사표 번호만을 저장해서 중복의 체크를 한다. 이것은 당연히 중복이 있으면 인쇄에 영향을 준다. 여기에서 중요한 체크 요령은 무엇을 바라는 것인지를 명확히 기술하는 것이고, 컴퓨터로 가능한지의 여부를 생각하는 것은 필요없다. 그것은 프로그래머가 요구된 순서를 컴퓨터로 실현하기 위해 연구를 하고, 그 결과 컴퓨터 자원(메모리, 자기 디스크 등)의 부족으로 불가능한 때에 체크 요령을 변경하면 된다. 그러면 체크 요령을 좀더 정확히 기술하기 위해서는 아래의 표현방법을 적당히

조합해서 사용한다.

(1) 수식에 의한 표현

수식은 가장 정확하지만 체크요령에서의 수식이용은 수학에서 말하는 경우의 수식과는 약간 형식을 다르게 하고 있어 엄밀성이 모자라는 면이 있다. 예를 들면 都道府縣 번호의 범위를 「都道府縣=01~47」이라든가 남녀별을 「남녀별=1, 2」라고 나타내기도 하는데 "~"라든가 ", " 등의 기호를 사용하는 것은 상식적으로 정해져 있는 것이지 엄밀한 정의에 따르고 있는 의미는 아니다. 만약 체크 요령과 같은 논리를 엄밀히 수식으로 정의 할 경우 그것을 기억하고 있지 않으면 체크요령을 쓸 수 없기 때문에 그렇게 복잡한 약속은 할 수 없다.

(2) 문장에 의한 표현

일반적으로는 문장으로 표현된다. 그러나 AND 조건(문장에서 "및")이라든가 OR조건 (문장에서는 "또는")이 하나의 문장중에 많이 나오는데 어디까지가 AND/OR의 범위인가가 명확히 포착할 수 없다는 결점이 있다. 또 문장을 쓰는 사람이 상세하게 썼다고 생각하더라도 읽는 사람의 지식의 정도에 따라 해석에 차가 생길 수 있으며, 더구나 쓰는 사람도 장문이나 조건이 복잡하게 얽히게 되면 논리의 모순을 깨닫지 못하는 일이 종종 있는 것처럼 요점에 있어서는 그 의도하는 비를 전달할 수 있어도 세부에 대해서는 논리의 엄밀성이 결여, 모호해지기 쉽다. 예를 들면 「A가 1이고 B가 2라면 C를 3으로 한다. 그렇지 않으면 C를 4로 한다」라고 하는 문장을 어떻게 해석하는가 이다: 전반의 문장은 누구라도 이해한다. 문제는

후반으로 「그렇지 않으면」의 어구에 걸리는 범위를 「B가 2」로 하는지, 전체로 확장시켜 「A가 1」까지로 하는가에 따라 결과는 달라진다. 「B가 2」까지로 하면 「A가 1이고 B가 2가 아닌데 C를 4로 한다」로 되어, 「A가 1」이다라는 조건을 지켜야만 한다. 또 전체가 관련된다고 한다면 「A가 1, B가 2의 둘다 동시에 성립하는 경우를 제외하고는 4로 한다」라는 의미가 되고 여기에는 「A가 1로서 B가 2가 아닌」 때도 「A가 1이 아니라면 B의 값에는 관계하지 않는다」 때도 포함된다.

이와 같은 문장에서는 상이한 해석이 나올 수 있기 때문에, 어느쪽의 해석이 올바른지는 쓰는 사람에게 질문하는 수밖에 없다.

(3) 표에 의한 표현

크로스표를 이용해서 표두, 표측에 쓰여진 항목간의 관계를 나타낸 것이며, 표를 보는 방법을 약속해 두면 잘못된 해석을 할 염려는 없다. 예를 들면 [표 8]은 「산업 소분류」와 「경영조직」 「사업소의 형태」 「종업자 총수」의 관련을 표 형식으로 나타낸 것이다. 이 표에서의 약속 사항은 「X표는 있을 수 없는 조합을 나타낸다」, 「△표는 있을 수 있지만 조사표를 확인할 필요가 있다」, 「공란은 있을 수 있는 조합을 나타낸다」의 3가지로 X표는 여러 데이터, △표는 여러인지 여부가 의심스러운 데이터, 공란은 올바른 데이터라고 하는것이 된다. 따라서 체크리스트상에 인쇄하는 데이터는 앞의 양자이다. 이와 같은 표형식으로 나타내면 논리는 명쾌하게 되지만, AND/OR를 포함한 관련하는 체크에서는 표두, 표측이 계층구조가 되어 보기 힘들어 진다.

표8 표형식에 의한 표현

산업소분류	경영조직												사업체 형태							종업원 총수							
	산업(사업소)						기업산업						사업체 형태							1인	5인	10인	20인	30인	50인	100인	
	1	2	3	4	5	6	7	2	3	4	1	2	3	4	5	6	7	4인	9인	19인	29인	49인	99인	259인			
925 정신박약·신체장애자 복지사업																											
926 갱생보호사업	×	×	×	×	×			×	×	×	×									△	△	△	△	△			
929 기타의 사회보법·사회복지																											
93 학술연구기관																											
931 자연과학연구소																											
932 인문과학연구소																						△	△	△			
94 정치·경제·문화단체																											
941 경제단체	×	×	×	×	×			×	×	×	×																
942 노동단체	×	×	×	×	×			×	×	×	×												△	△			
943 학술·문화단체	×	×	×	×	×			×	×	×	×													△			
944 정치단체	×	×	×	×	×			×	×	×	×												△	△			
949 달리 분류되지 않는 영리적 단체	×	×	×	×	×			×	×	×																	
95 기타의 서비스업																											
951 집회장																								△			
952 도축장																								△			
959 달리 분류되지 않는 서비스업																						△	△	△			

(4) 디씨전 테이블에 의한 표현

표형식을 더욱더 발전시킨 것으로 1960년 전후로 아메리카에서 각광을 받은 방법이다. 앞의 표형식에서는 예러가 되는 코드의 조합은 명시 할 수 있어도 그 결과 어떻게 하는 것인가 하는 행동부분의 기술은 예를 들면 「체크리스트에 인쇄한다」 등과 같이 문장으로 보충하고 있다.

이것을 표의 일부로 거두어 들인 것이 디씨전 테이블이라고 일컬어 진다.

디씨전 테이블에는 몇개의 약속이 있다. 우선 표의 형식은 [표 9]와 같은 요소로 구성된다. 테이블 헤더(Header)는 디씨전 테이블이 여러개 있는 경우에 구별하기 위한 것이기 때문에 체크요령에 응용하는 것이라면 체크항목이라도 기입해 두면 된다.

표9 디씨전 테이블의 요소

테이블 헤더	규칙 헤더
조건 스템브	조건 엔트리
행동 스템브	행동 엔트리

표10 디씨전 테이블에 의한 표현

	규칙1	규칙2	규칙3	규칙4
A는 1이나	Y	Y	N	N
B는 2이나	Y	N	Y	N
C는 3으로 한다	X			
C는 4로 한다		X	X	X

조건 스템브(STUB)에는 대상이 되는 문제의 조건을 질문형식으로 기입한다. 예를 들면 「성별은 남자인가」라는 식이지만, 반드시 문장이 아니라도 좋은 조건이 확실하기만 하다면 수식을 사용해도 좋다. 행동 스템브는 조건 스템브에 쓰여진 조건의 결과로서 취급해야 할 행동, 예를 들면 「조사표의 내용을 프린트 아웃한다」 등과 같은 문장으로 쓴다. 조건 엔트리는 조건 스템브의 각 질문에 대한 답을 Y(예스), N(노)로 기입하든지, 또는 Y, N으로의 기입대신에 답 그것을 기입한다. 행동 엔트리에는 조건 엔트리의 답에 대해서 취급해야 할 행동에 X가 기입된다. 마지막으로 규칙 헤

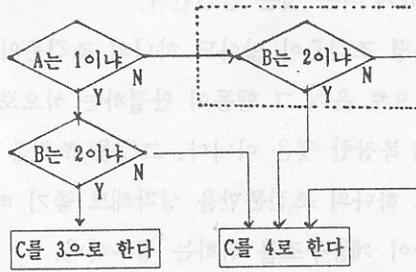
더는 조건 엔트리에의 기입상황에 따라 발생하는 그들의 행동을 구별하기 위한 번호를 기입한다. [표10]은 「A가 1이고 B가 2라면 C를 3으로 한다. 그렇지 않으면 C를 4로 한다」 라고하는 문장을 다씨전 테이블로 나타낸 것이다. 조건 스테브에는 「A는 1인가」라고 질문형식으로 기입한다. B도 같이 「B는 2인가」라고 기입한다. 이것에 대해 행동 스테브에는 「C를 3으로 한다」 경우와 「C를 4로 한다」 경우를 기입한다. 그런데 전반의 문장은 규칙 1의 경우에 해당하는 것이다. 후반의 「그렇지 않으면」라고 하는 것은 규칙 2~규칙 4에 해당하고 「그렇지 않으면」이 「B가 2」에 걸린다면 규칙 2, 「A가 1」에까지 걸리는 것이라면 규칙 3, 규칙 4이기 때문에 문장의 경우와 같은 모호한 경우가 없이 명쾌히 나타내지고 있다. 이와 같이 다씨전 테이블은 어떻게든지 체크요령의 기술로 향하고 있는 것 같이 보이지만 그것은 [표 10]과 같은 간단한 경우이며, 각각의 조건은 스테브가 Y,N라고 하는 것의 모든 조합을 열거하는 것을 의미하므로 조건 스테브의 수가 많게 되면 규칙의 수는 가속도 적으로 증가한다(조건 스테브 수를 N으로 하면 2^n 이 된다). 예를 들면 [표 10]에서는 조건 스테브가 2가지로 규칙수는 4이지만, 조건 스테브가 4가 되면 규칙수는 4배의 16이 되어 버린다. 일반적으로 수많은 조건 스테브의 Y,N의 조합 가운데에는 무의미한 것도 생기기도 하고 그것을 삭제해야 하는 것도 있다. 중요한 것은 Y,N의 조합에 의한 규칙수의 증대는 다씨전 테이블을 장대화, 장황화해 버리며, 이해하기 쉽다라는 장점을 잃어 버리게 된다.

(5) 플로우 차트(Flow Chart)에 의한 표현

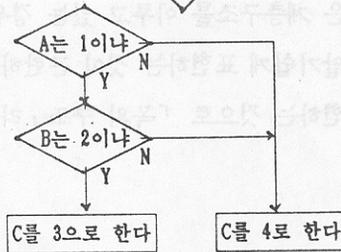
문장이나 표형식에 적당치 못한 곳이 있다면 그림에 의해 알기쉽게 하는 방법으로 프로그램을 작성할 때에 사용하는 플로우 차트가 있다. 프로그램은 엄밀한 논리에 따라 작성되는 것이므로 플로우 차트는 논리적으로 조립되는 것이기 때문에 논리를 표현하는데 적당하다.

[그림 25]는 [표 10]의 디씨전 테이블을 플로우 차트로 나타낸 것이다. 이 그림은 상단의 (7)의 디씨전 테이블을 충실히 베낀 것으로 점선으로 둘러싸고 있는것 같은 분명히 의미 없는 질문문이 그려져 있다. 디씨전 테이블에서 이것에 해당하는 부분은 규칙 3과 규칙 4이다. 디씨전 테이블에서는 알기어려운 조건 스테브가 플로우 차트에서는 일목요연 하기 때문에 복잡한 경우에는 디씨전 테이블보다 논리적으로 표현할 수 있다.

그림25 플로우 차트에 의한 표현



(7)



(L)

다만, 프로우차트법에 의해서도 조건문(마름모 꼴의 도형)이 많아지고, 게다가 조건문 하나의 결과에 또한 별도의 조건문이 연결, 그 결과에 또 조건문이 계속 하는것 같이 마치 옥상옥을 쌓아올리는 것같이 되어 하나의 행동에 겨우 다다른것 같은 계층적 구조로 되어 논리의 전망이 급속히 나빠게 된다. 또 최근에는 플로우차트가 컴퓨터 모임에서 그 사용이 감소되고 있음을 부정할 수 없다.

(6) 목 구조도에 의한 표현

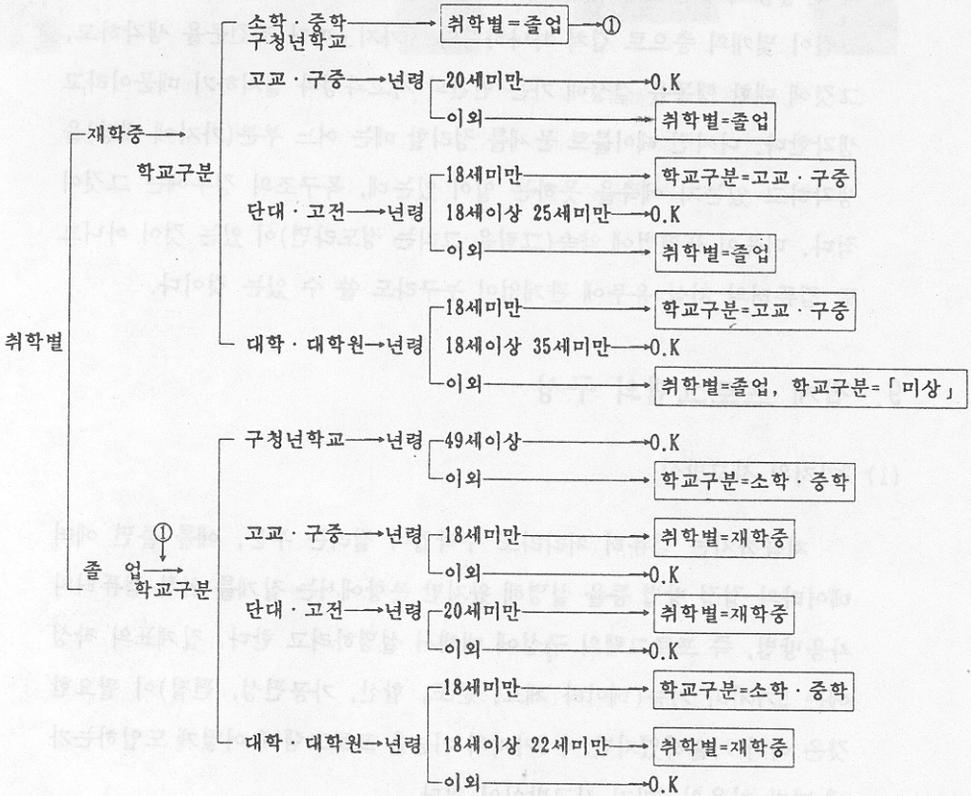
지금까지 논리의 전개가 좋든가 나쁘든가의 표현법을 해 왔지만 논리전개가 나쁘다 혹은 체크요령이 복잡하다 라고 하는 상태는 정도의 차는 있겠지만 필자는 다음과 같은 2가지의 요인에 의해 발생한다고 본다. 첫째는 조건문이 많은 것이다. 조건문의 결과에 따라 취해야 할 행동이 많게 되면 복잡하게 되는 것은 당연한다.

그러나 가령 조건문이 많아도 하나의 조건문의 행동이 완결되고나서 다음의 조건문으로 옮겨 그 행동이 완결하는 식으로 직렬적으로 연결되어 있다면 그렇게 복잡한 것은 아니다. 그것은 조건문 상호간의 관계하는 정도가 낮으므로 하나의 조건문만을 생각해도 좋기 때문이다.

둘째는 조건문이 계층구조를 취하는 경우이다.

계층의 쌓아올림을 「깊이」라고 하는데 그 깊이가 깊어 질수록 복잡하게 된다. 이와 같은 계층구조를 이루고 있는 경우에는 지금까지 기술한 방법으로는 논리를 알기쉽게 표현하는 것이 곤란하다. 그래서 계층구조를 되도록 알기쉽게 표현하는 것으로 「목의 구조」라고 불리어지는 것이 있다.

그림26 목구조에 의한 표현



계층구조의 대표적인 예는 「조직도」이다. 이것은 예를 들면 부, 과, 계가 어떻게 위치(상하의 관계, 가로와의 관계)하고 있는지를 「목의 구조」를 사용해서 나타낼 수 있다. 종서(내리쓰기) 조직도에서는 상층부가 위에 쓰여져 있고 몇개의 가지로 갈라져 부, 과 등이 연결되어 있으므로 이것의 상하를 거꾸로 해 보면 정확히 「나무」가 하늘로 향하여 뻗고 있는 것 같이 보이므로 이러한 도형을 「목구조」라고 부르고 있다. 이 목구조

를 체크 요령에 응용한 예가 [그림 26]이다. 이것과 비슷한 도형은 일상의 문장중에서 무의식적으로 쓰는 일이 있지만, 목구조가 알기 쉬운것은 조건이 몇개의 층으로 겹쳐 하나하나의 「가지」마다 조건문을 생각하고, 그것에 대한 행동을 결정해 가는 인간의 사고과정과 일치하기 때문이라고 생각한다. 디씨진 테이블로 문제를 정리할 때는 어느 부분(가지에 해당)을 생각하고 있는지 예측을 못하는 일이 있는데, 목구조의 경우에는 그것이 적다. 더욱이 표현법에 약속(그림을 그리는 정도라면)이 있는 것이 아니므로 컴퓨터의 지식 유무에 관계없이 누구라도 쓸 수 있는 것이다.

9. 집계 프로그램의 구성

(1) 2가지의 사고방식

지금까지는 컴퓨터 처리라도 수작업이 걸리는 부분, 예를 들면 여러 데이터의 정정 방법 등을 설명해 왔지만 본항에서는 집계를 위한 컴퓨터의 사용방법, 즉 프로그램의 구성에 대해서 설명하려고 한다. 집계표의 작성에는 5가지의 기능(데이터 체크, 분포, 합산, 가공편성, 편집)이 필요한 것은 이미 기술하였지만 이 5가지의 기능을 프로그램에 어떻게 도입하는가에 따라 다음의 2가지 사고방식이 있다.

첫째는 5가지의 기능을 하나의 프로그램에 작성하는 사고방식이며, 둘째는 5가지의 기능을 각각의 프로그램에 분산시키는 사고방식이다. 지금 기능을 중심으로 2가지의 사고방식으로 나누었지만 이것을 통계표로 착안해보면, 첫번째 사고방식은 체크에서 편집까지의 기능을 가진 프로그램을 각 통계표마다 작성하는 것을 의미하고, 두번째 사고방식은 모든 통계표에 대해서 각 기능마다의 프로그램을 작성하는 것을 의미하고 있다. [그림

27]은 양 사고방식을 개념적으로 나타낸 것이다.

즉 그림에 있어서는 굵은선 화살표를 수직방향으로 더듬어 찾을지, 수평방향으로 더듬어 찾을지에 따라 첫번째 사고방식, 두번째 사고방식이 얻어지는 것이다. 2가지의 사고방식에는 각각 일장 일단이 있고, 어느쪽이 우수하다라고는 단정할 수 없지만 통계표를 집계하는 "집계에서라면

- ① 처리시간은 적게하고 싶다.
- ② 프로그램 작성의 부담을 가볍게 하고 싶다.
- ③ 다시 고쳐 하기를 최소시간으로 억제하고 싶다.

라고하는 기본자세에서 집계의 방법을 고르게 된다.

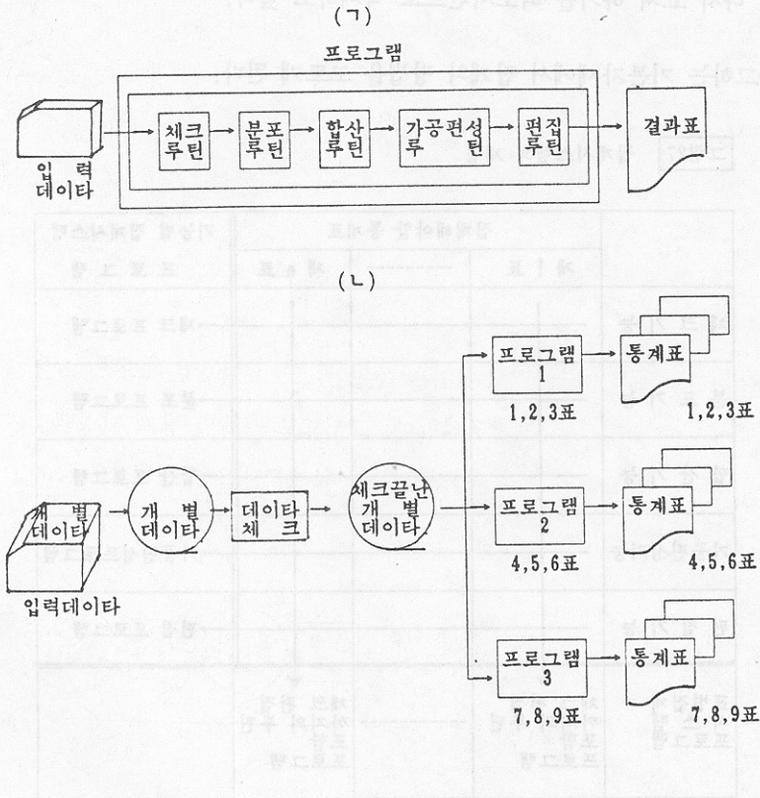
그림27 집계시스템의 개념

	집계해야할 통계표			기능별 집계시스템 프로그램
	제 1 표	-----	제 n 표	
체 크 기 능				→체크 프로그램
분 포 기 능				→분포 프로그램
합 산 기 능				→합산 프로그램
가공편성기능				→가공편성프로그램
편 집 기 능				→편집 프로그램
표별집계시스템 프로그램	↓ 체크 편집 까지의 포함 프로그램	-----	↓ 체크 편집 까지의 후턴 포함 프로그램	

우선 첫번째 사고방식에 대해서 보자.

이것은 통계표마다 프로그램을 작성해서 집계하기 때문에 「표별 집계 시스템」으로 불러도 좋다. 이 경우 프로그램은 [그림 28(7)]과 같이 입력 데이터(조사표)의 체크에서 통계표의 편집까지를 루틴(명령의 모임)화해서 작성되지만 통계표의 종류도 많고, 하나의 프로그램으로 처리할 수 없는 경우에는 5가지의 기능을 가진 프로그램을 여러개 만들어야 한다.

그림28 표별집계시스템의 처리순서



이와 같은 예가 [그림 28(L)]이다. 이 그림에서 데이터 체크가 독립해서 있는 것은 여러 데이터를 정정할 때에 조사표의 개념, 원래자리로 되돌린다고 하는 업무를 생각해 볼때 각각 개별의 프로그램이 데이터 체크 기능을 갖는 것은 유리한 방법은 아니다.

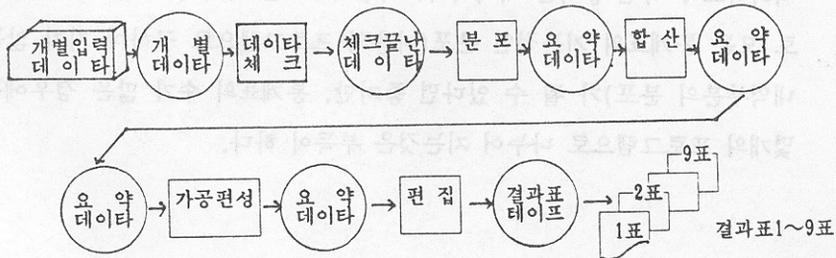
오히려 단숨에 체크를 행하여 정정하는 편이 효율적(조사표의 개념이 1회로 끝남으로)이라고 말할 수 있다.

「표별 집계 시스템」의 특징은 입력 데이터를 읽고, 즉시 결과표가 출력되는 (외관상의) 일괄처리에 있다. 따라서 각각의 프로그램으로 가장 양이 많은 개별 데이터를 반복처리하는 것은 「처리 시간을 적게하고 싶다」라고 하는 조건을 충족 시키지 못한다. 또 프로그램은 각 기능을 루틴으로서 작성하고 있으므로 당연히 복잡하다. 더구나 프로그램 에러가 발생한 경우 에러가 어느것인지 하나의 기능으로 한정되어도 그 기능만을 다시 고쳐 사용하기가 곤란하다. 설령 가능하여도 프로그램의 상당한 수정을 필요로 하기 때문에 통상은 프로그램 전체를 다시 수정하게 된다.

다음은 두번째 사고방식에 대해서 알아보자.

이것은 [그림 29]와 같이 각 기능을 독립한 프로그램으로 작성해서 그것들을 조합시켜 집계 시스템을 구성한다. 그러므로 「표별 집계 시스템」에 대해서 「기능별 집계 시스템」이라고 부른다. 이 시스템의 특징은 첫째는

그림29 기능별집계 시스템의 처리순서



하나의 프로그램은 하나의 기능밖에 가지고 있지 않기 때문에 「표별집계 시스템」의 한 프로그램 다기능에 비교하면 프로그램의 구성은 단순하다. 두번째는 가장 양이 많은 개별 데이터 화일의 처리회수를 적게 할 수 있다.

개별 데이터를 입력으로하는 프로그램은 「데이터 체크」와 「분포」이다. 이 가운데 「데이터 체크」에 있어서의 출력은 개별 데이터이지만 「분포」에 있어서의 출력은 개별데이터에 비해 양적으로 적게 집약한 「요약 데이터」이다. 「데이터 체크」에 있어서는 「표별 집계 시스템」이거나 「기능별 집계 시스템」이라도 프로그램의 구성은 아무런 변함이 없다. 오히려 독립된 프로그램으로 하는 편이 처리상도 바람직하다. 2가지의 시스템으로 바꾸는 것은 「분포」 이후의 각 기능의 프로그램이다. 「표별 집계 시스템」에서는 하나의 프로그램 중에서 개별 데이터의 분포가 전부 끝나기까지는 다음의 기능, 예를 들면 「합산」은 실행되지 않는 것에도 불구하고 루틴으로 메모리를 점령하고 있다.

루틴의 작성방법에는 각종의 기법이 있고, 반드시 메모리를 점령하고 있다고는 할 수 없지만 여기에서는 양 시스템의 기본적인 사고방식의 차이를 비교하는 것이 목적이므로 「점령」하고 있는 것으로서 설명을 행한다. 개별 데이터의 처리 회수를 줄이는데는 1회의 처리로 될 수 있는 한 많은 통계표의 분포를 만들어 버리는 것이다. 이러한 처리를 하기 위해서는 「분포기능」 이외의 기능을 메모리에서 빼내 그 분포만 통계표를 여분으로 처리하도록 하면 좋다는 의미이다. 이상적으로는 하나의 분포 프로그램으로 모든 통계표의 기본적인 분포(다음의 프로그램으로 구하여 지지 않는 내역부분의 분포)가 될 수 있다면 좋지만, 통계표의 수가 많은 경우에는 몇개의 프로그램으로 나누어 지는것은 부득이 하다.

세번째 특징은 에러수정에 대해 유연성을 가지고 있다. 프로그램의 에러는 예상외로 많다. 에러를 범하지 않는 것이 물론 중요하지만, 에러가 발생하면 재빠르게 그 원인을 찾아 대책을 세워야 한다.

통상은 프로그램을 정정하고 컴퓨터로 재처리를 한다. 그대에 프로그램이 단순한 구성으로 되어 있으면 에러 원인의 발견도 그 정정도 용이하며, 재작성에 대해서도 원래 기능마다 프로그램이 작성되어 있기 때문에 에러 원인으로 된 프로그램 이후를 부분작성만 하면 끝난다. 예를 들면 「합산기능」의 프로그램에 에러가 있는 경우, 그것보다 앞의 「분포기능」의 프로그램은 다시 고쳐야할 필요는 없다. 이것에 대해 「표별 집계 시스템」에서는 각 기능은 루틴으로서 독립해 있어도 하나의 프로그램으로 연속해서 처리되기 때문에 에러가 있는 루틴을 분리해서 재처리 하는것은 곤란하다. 따라서 프로그램 전체를 재작성하는 것으로 되어 개별 데이터의 처리부터 시작해야 한다.

네번째 특징은 프로그래머의 경험 년수, 기량에 따라 프로그램의 분업을 할 수 있는 것이다. 「표별 집계 시스템」이라도 [그림 28(L)]과 같이 3가지의 프로그램을 분업할 수 있지만 이 분업은 5가지 기능의 프로그램 기술을 습득한 프로그래머를 전제로 하고 있다. 이것에 대해 「기능별 집계 시스템」에서는 개별 데이터의 체크나 분포처럼 처리시간이 많이 걸리는 프로그램은 에러 때의 피해도가 크므로 상급 프로그래머가 담당하고 인쇄를 위한 편집 프로그램은 초급 프로그래머가 담당하는 방식으로 프로그래머의 경험 정도, 기술 레벨 등의 요소를 생각해서 분업체제를 취하는 것이다. 이 일은 또 프로그래머의 양성에도 관계된다. 프로그램을 쓰는 방법(작성방법) 등 논리면의 연수를 받고, 쉬운 업무를 실습해서 숙달 시킨다는 것이 「기능별 집계 시스템」에서는 하기 쉽다.

다섯번째는 개개의 프로그램의 표준화, 범용화(공통 이용하는 것)를 하기 쉽다는 것이다. 하나의 프로그램은 하나의 기능만을 가지므로 필연적으로 간소화하여 그 결과 표준화, 범용화가 하기 쉽다. 「기능별 집계 시스템」은 질적으로는 쉬운 프로그램이긴 하지만 양적으로는 프로그램의 갯수가 증가하는 경향이 있기 때문에 범용화 하지 않으면 도리어 프로그래머의 부담이 되어 버리는 단점이 있다.

이상과 같이 양 시스템을 비교해 보면 「기능별 집계 시스템」이 우수하다는 인상을 받지만 소량의 개별 데이터의 처리나 통계표의 수가 적은 경우에는 「표별집계 시스템」편이 적당하다.

(2) 체크 프로그램의 구성

체크 프로그램의 구성은 다음의 4가지로 집약된다.

첫번째는 데이터 체크 기능이다. 이를 위한 기법으로서 오프코드 체크나 관련 체크 등이 있다는 것은 이미 기술하였으므로 반복하지 않는다.

두번째는 에러 데이터의 정정이다. 특히 컴퓨터에 의한 자동수정을 작성해 두어야 한다. 만에 하나 수작업에 의한 정정을 태만한 때에 에러의 데이터가 출력되는 것을 막기 위하여 필요하다. 그러나 이 대책은 예상이 빗나가는 경우도 있는데, 예를 들면 컴퓨터의 체크로 에러가 지적되어도 조사표를 확인하면 잘못이 아니라고 하는 것이 있다. 이것은 이미 기술한 것이지만 의심스러운 데이터를 될수 있는한 체크로 찾으려 하고 있기 때문이며 이러한 데이터에 대해서 자동수정을 하는 것은 좋지 않으므로 자동수정을 피하기 위한 연구를 해둘 필요가 있다.

세번째는 부호변환이다. 조사항목의 부호는 컴퓨터 처리의 일보다도 조사표의 기입의 용이함, 내용검사나 부호기입의 용이성이라고 하는 인간

의 노력을 줄이는 것을 주안점으로 설계되어 있기 때문에 본포이후의 집계 프로그램에 적당하다라고는 말할 수 없는 경우가 있다. 예를 들면 국세 조사에 있어서 「연령 각세」, 「5세 계급」, 「10세 계급」 등의 연령별 통계표가 만들어져 있지만 조사표에는 신고자의 연령(생년월)이 기입되어 있을 뿐이며, 「5세 계급」, 「10세 계급」의 코드는 일부러 기입하지 않는다. 그 이유는 「연령각세」가 있으면 컴퓨터에 의해 「5세 계급」, 「10세 계급」의 코드를 만들어 낼 수 있기 때문이다. 프로그램이 몇개로 나누어져 분업하고 있는 경우, 각각의 프로그램으로 코드 변환하면 중복이 되기도하고 오류를 범하기도 하므로, 체크단계에서 행하는 것이다. 코드 변환의 가운데에는 계산으로 구해지지 않는 것도 있다.

예를 들면 산업소분류나 직업소분류 코드를 연속번호로 변환하는 경우이다. 이 번호의 필요성은 본포 프로그램으로 산업 또는 직업의 소분류별의 통계표를 만들때에 사용하기 때문이다.

네번째는 체크 프로그램의 출력 데이터에는 입력정보(조사표의 내용)를 반드시 부가해 두는 것이다. 물론 후속의 프로그램에서 필요가 되는 조사항목은 출력 되지만 조사표 번호, 묶음 번호라고 하는 직접 집계에는 관계없는 것도 기록해 둔다. 본포 프로그램에서 문제가 발생할 경우 조사표를 거슬러 올라갈 수 있기 때문이다. 또 입력 데이터의 항목이 있는 것은 코드 변환에 의해 별도 정보로 형을 바꾸지만 코드 변환시의 오류를 생각하면 그 기본이 되는 코드는 기록해 두어야 한다. 체크 프로그램과 같이 수작업부문과의 「뒤엎힘」이 있으면 프로그램만의 재수정은 어려우므로 본래의 코드가 있으면 그것에서 코드변환만을 고쳐할 수 있기 때문이다.

(3) 본포 프로그램의 구성

분포 프로그램에서는 체크가 끝난 개별 데이터(체크 프로그램의 출력 데이터)로 분포표를 작성하는 것이므로 개별 데이터의 처리횟수를 극히 적게 하는 것이 기본적이라고 생각된다.

일반적으로 통계조사의 통계표는 종류가 많고, 게다가 하나의 통계표의 셀의 수도 많기 때문에 하나의 분포 프로그램으로 전부의 통계표 분포를 구할 수 없는 일이 많다. 그러므로 통계표를 몇개의 그룹으로 나누어 각 그룹마다 분포 프로그램을 만들면 개별 데이터를 여러번 처리하는 것이 된다. 그래서 분포 프로그램에서는 프로그램의 내용을 극히 단순화하고 프로그램수를 가능한한 적게 하도록 구성을 취한다. 내용 단순화 방법의 하나로 는 나중에 산출할 수 있는 합계란을 전부 빼면 메모리가 절약되고, 또한 합계산출을 위한 명령을 사용하지 않으면 그 부분의 메모리를 절약해서, 조금이라도 많은 통계표의 메모리를 할당할 수 있게 된다. 예를 들면 [표 11]의 통계표의 모든 셀의 수는 120셀로 그 가운데 합계란은 64셀 분을

표11 합계란(사선부분)의 삭제

	총 수					남					여				
	총 수	미혼	유배우	사별	이별	총 수	미혼	유배우	사별	이별	총 수	미혼	유배우	사별	이별
총 수															
15~19세															
20~24세															
25~29세															
30~39세															
40~49세															
50~59세															
60세이상															

차지하고 있다. 대개 받은 나중에는 계산할 수 있는 합계란인 것이다. 게다가 합계란을 삭제하면 분포의 방법도 간단하게 된다. 분포의 방법은 통계표를 [그림 30]과 같이 2차원의 배열로 생각하면 이해하기 쉽다. 분포해야 할 위치(셀)를 $P(X, Y)$ 로 나타내고, X 는 표측의 「행」을 Y 는 표두의 「열」로 하면 표측, 표두의 항목코드가 일련번호로 나타내져 있는 경우 그대로 「첨자」로서 쓸 수 있다. 이 X, Y 를 「매트릭스 코드」라고 부르고 있다.

그림30 분포의 방법

		표 두								
		1	2	3	4	5	6	7	8	예
표 측	1									
	2		원점							
	3									
	4					$P(x, y)$				
	5									

↓ 행

예를 들면 표측이 「연령 5세 계급」으로 이것이 일련번호이며, 표두가 「남녀별」로 「1, 2」와 연결번호라면 분포해야 할 셀 P 를 찾는 것은 입력 데이터의 「연령 5세 계급 코드」와 「남녀별 코드」를 그대로 명령 가운데에 사용할 수 있다. (구체적으로 표현은 생략). 그러나 통계표의 표측, 표두에 합계란이 있으면 일련번호의 코드라 할지라도 합계란이 방해를 해서 그대로 명령 가운데에 사용할 수 없다. 분포 프로그램에서의 통계표의 메모리 전개와 분포를 쉽게 하기 위한 매트릭스 코드와는 밀접한 관계가 있고, 메모리 전개의 방법(예를 들면 합계란을 포함해서 전개할지 여부)에 의해 매트릭스 코드의 번호기입은 변한다.

분포 프로그램을 단순화하기 위하여 앞에 설명한 체크 프로그램으로 매트

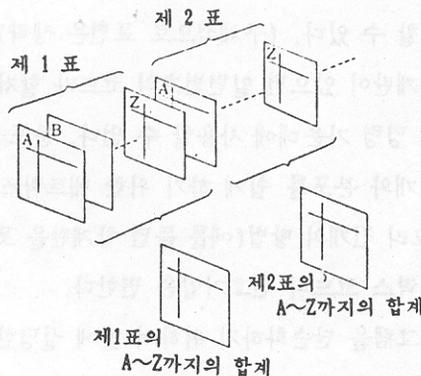
릭스 코드를 부여하는 방법도 있지만 통계표의 변경에 의한 매트릭스 코드를 다시 부여하기 때문에 좋은 방법은 아니다. 오히려 체크 프로그램에서는 각 항목의 코드를 일련번호로 출력하는 방법이 좋다.

(4) 합산 프로그램의 구성

본포 프로그램에서는 될 수 있는 한 합계란을 설정하지 않고 내역의 셀에 분포하였지만 이 합계에는 2종류가 있다. 하나는 표내의 셀의 합계이며, 또 하나는 표끼리의 합계이다. 예를 들면 표내의 셀의 합계는 통계표의 표측, 표두에 있는 합계이며, 표끼리의 합계는 통계표가 「市區町村」 등의 지역구분으로 명시되어 있는 경우의 「縣計」나 「시군계」, 「군부계」가 이것에 해당한다. 이 가운데 합산 프로그램이 취급하는 합계는 표끼리의 합계이다. 표끼리의 합계는 바꾸어 말하면 통계표의 난의 항목에 관한 합계이다.

따라서 지역계라고는 할 수 없는 난의 항목이 「남녀별」이라면 「남녀계」를 구하는 것이 된다.

그림31 표끼리의 합계



간단한 예로서 [그림 31]로 설명하자.

이 그림에서의 통계표는 표측, 표두의 항목은 생략하고 있지만, 난의 항목은 「市區町村」으로 「A시」에서 「Z촌」까지가 분포 프로그램으로 산출되어 있다. 이제부터 「縣計」를 구하는 것이지만 일반적으로 분포 프로그램에서는 하나의 프로그램으로 가능한 많은 통계표를 구하도록 하고 있기 때문에 [그림 31]과 같이 같은 종류(같은 표번호)의 통계표가 출력 화일에 정연히 늘어서 있는 의미는 아니다. 정돈하기 위해서는 소트프로그램이라고 해서 순번을 정돈하는 범용 프로그램을 필요로 하는데 이것은 컴퓨터 메이커가 작성해서 제공 하므로 우리들이 작성할 필요없이 이용만 하면 된다. 이 소트 프로그램을 사용해서 [그림 31]과 같이 소트 하고나서 합계를 구한다.

합계를 구하는 방법은 먼저 구하고자 하는 통계표와 같은 크기의 메모리를 준비하고 입력 데이터(A市, B市 라는 지역단위로 써내는 것과 마찬가지로)를 읽어서 표번호가 같은 데이터를 준비한 메모리에 가산해 간다. 그리고 표번호가 다른 데이터를 읽어들이는 때, A시에서 Z촌까지의 합계가 메모리에 구해지게 된다. 여기서 Z촌의 데이터까지를 가산하기 위하여 난외 항목이 Z촌인 데이터가 메모리에 읽어 들여졌는지 어떤지를 비교명령으로 점검하는 방법을 사용하면, 만약 제 1표에 Z촌의 데이터가 존재하지 않는 경우에는 제 2표의 Z촌의 데이터를 찾을때까지 가산을 계속해 가는 것이 되기 때문에 틀린 통계표까지 가산하는 것이 된다. 일부만 컴퓨터 처리를 다시하는 경우에 데이터 전부를 사용하지 않고 일부만 처리할 때라든가, 혹은 특정한 대상, 예를 들면 「실업자」에 관해서 지역구분별의 통계표를 작성한때에 Z촌에 「실업자」가 없으면 Z촌의 데이터는 존재하지 않기 때

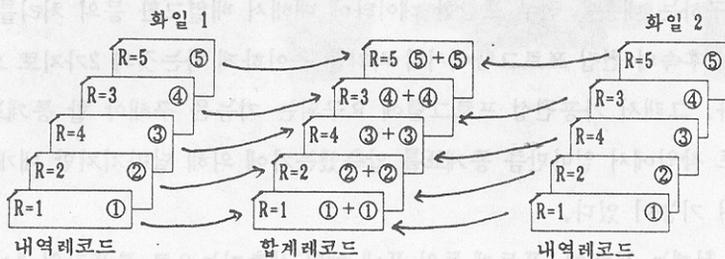
문에(이러한 경우는 출력화일에 써내지 않는다) Z층을 눈금으로 합계를 구할 수는 없다.

합계의 종류로는 대계, 중계, 소계와 같이 레벨이 있다. 통계표의 구조가 어느정도의 계층구조로 되어 있는가에 의해 구해야 할 합계의 레벨이 다르게 되지만 1회의 합산 프로그램의 처리로 모든 레벨의 합계를 구하는 것은 피하는 편이 좋다. 그 이유는 통계표마다 그 구조가 다르기 때문이기도 하지만 합계의 레벨이 달라지고 합산 프로그램의 내용이 복잡하게 되기 때문이다. 프로그램은 단순한 만큼 오류는 적기 때문에 합산 프로그램의 1회의 처리로는 각 통계표 하나의 레벨 합계를 구하는 정도의 프로그램 구조로 하는 편이 좋다.

이와 같이 하면 합산 프로그램은 5가지의 기능 프로그램 가운데 가장 단순하게 되어 범용화가 하기 쉽게 된다. 「합산처리」는 범용화한 합산 프로그램을 사용하지 않으면 각 레벨의 합계를 구하기 위하여 「소트」, 「합산」을 반복 사용하므로 그때마다 합산 프로그램을 작성한 것은 매우 비효율적인 일이 된다. 합계의 레벨은 통계표의 구조에 의존하는데 합계를 구하는 방법으로 직렬형과 병렬형 2가지가 있다. 직렬형의 합산은 [그림 31]에서 보는 것처럼 내역의 데이터가 순서있게 늘어서 있어 그것을 가산해서 합계를 구하는 방법이며, 이것은 전표를 넘기면서 합계를 구하는 것과 유사하다. 병렬형의 합산은 [그림 32]와 같이 2개 이상의 입력화일에서 데이터를 읽어서 각각을 가산해서 합계를 구하는 방식이다.

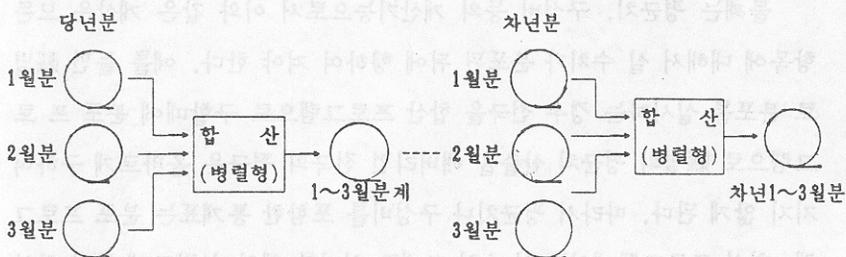
이 방식은 매월의 집계결과에서 [그림 33]과 같이 4분기의 합계를 구하는 경우에 많이 사용된다.

그림32 병렬형의 합산



주 R은 란의항목을 나타낸다.

그림33 사반기마다 합산



(5) 가공편성 프로그램의 구성

분포 프로그램에서는 집계시간의 단축, 프로그램 수의 삭감 등을 목적으로 하여 통계표를 최소한의 내용으로 한정시키는 것을 설계의 기본으로 했기 때문에 이것에 의해 구해야 할 통계표를 후속의 프로그램으로 복원하지 않으면 안된다. 그 하나가 합산 프로그램에서의 각종 합계의 산출이며, 또 하나가 가공편성 프로그램에서의 각 기능의 산출이다. 따라서 이 프로그램의 목적은 앞단계까지의 프로그램 처리된 데이터에 대해서 그 내용을

보충하여 통계표를 완전히 만족하는 요약 데이터를 작성하는 것과 통계표에서 구하는 내용을 전부 포함한 데이터에 대해서 배열교환 등의 처리를 행하여 후속의 편집 프로그램에서의 처리를 용이하게 하는 것의 2가지로 요약된다. 그래서 가공편성 프로그램에 요구되는 기능은 구해야 할 통계표와 분포 작업에서 얼마만큼 통계표를 압축했는가에 의해 달라지지만 대개 다음의 기능이 있다.

첫째는 표측계, 표두계 등의 표내 계의 산출기능으로 분포작업에서 통계표의 내역부분의 셀을 압축한 경우에 산출한다. 내역만으로 처리하는 이유는 메모리를 절약하고 요약 레코드의 길이를 짧게하여 처리시간을 단축하며 분포 프로그램의 단순화를 도모하기 위함이다.

둘째는 평균치, 구성비 등의 계산기능으로서 이와 같은 계산은 모든 항목에 대해서 실 수치가 분포된 뒤에 행하여 져야 한다. 예를 들면 縣별로 분포를 실시하는 경우 전국을 합산 프로그램으로 구할때에 분포 프로그램으로 縣별의 평균치 산출을 해버리면 전국의 평균은 올바르게 구하여지지 않게 된다. 따라서 평균치나 구성비를 포함한 통계표는 분포 프로그램, 합산 프로그램에서는 실 수치 그대로 처리를 행하며 전국계 등의 합산치도 실수치로 구하고 나서 평균치, 구성비를 계산한다.

또 매월의 평균으로 평균치(구성비)를 포함한 통계표의 연간 수치나 연평균 수치의 산출도 전국계와 같은 모양의 취급을 하기 때문에 기본이 되는 월보의 요약 데이터는 실 수치로 계산되어야 한다.

셋째는 분포항목의 삭제기능이다. 이미 산출한 분포항목을 삭제한다고 하는 것에 의문을 갖겠지만, 분포 프로그램으로 작성한 분포항목은 통계표의 출력에 직접 필요한 항목만이라고 할 수만은 없기 때문이다. 통계표의 출력에 불필요한 항목으로 분포하는 것은, 예를 들면 [표 12]의 통계표

와 같이 「주로 일한자」의 내역은 「종업원 규모」인 것에 대해서 「틈틈이 일한자」의 내역은 없다. 그러나 「주로 일한자」와 「틈틈이 일한자」의 합계인 총 수는 「종업원 규모」의 내역을 가지고 있다. 그 때문에 「틈틈이 일한자」에는 「종업원 규모별」의 명시는 필요하지 않는데도 불구하고, 분포 프로그램으로는 [그림 34]와 같이 「종업원 규모」를 구해두어야 한다. 그리고 가공편성 프로그램으로 「주로 일한자」와 「틈틈이 일한자」의 「종업원 규모」를 각각 가산해서 「총수」를 구함과 동시에 「틈틈이 일한자」의 합계를 구하고 나서 「종업원 규모」를 삭제하는 것이다. 이 예와 같이 분포 항목을 삭제하는 방법은 통계표를 구성하는 각 항목의 분류가 일정하지 않고 구성이 불규칙하면 분포 프로그램이 복잡하게 되기 때문에 세세한 분류에 합해서 프로그램을 단순화하기 위하여 행하고 있다.

그림34 분포프로그램을 위한 통계표

취업 상태 종업원 규모		표 두 항목
주로 일한 자	1 ~ 9 인	이곳을 확대해 둔다
	10 ~ 19 인	
	⋮	
	관 공 서 미 상	
틈틈이 한 자	1 ~ 9 인	
	10 ~ 19 인	
	⋮	
	관 공 서 미 상	

표12 「총수」가 세밀한 통계표

남 녀 별	취업 상태	중 업 원 규 모	4 5 ~ 5 4 세						재 학 자	교 육 미 상
			총 수	졸 업 자						
				총 수	소 학 · 중 학	고 교 · 구 중	단 대 · 고 전	대 학 · 대 학 원		
A	총 수	12646	12633	5971	5068	472	1123	0	13	
	1~9인	5401	5397	2966	2092	127	212	0	3	
	10~19인	914	914	470	359	27	53	0	0	
	관공서 중업원규모 미상	1268 12	1267 8	278 3	429 4	160 0	299 1	0 0	1 5	
B	주로 일한자	10651	10639	4905	4224	412	1098	0	11	
	1~9인	3986	3983	2177	1519	88	199	0	2	
	10~19인	791	791	409	303	21	57	0	0	
	관공서 중업원규모 미상	1214 11	1214 6	261 3	501 3	156 0	297 1	0 0	1 4	
C	틈틈이 일한자	1995	1994	1066	844	60	24	0	2	

네번째는 분포항목의 배열교환 기능이다. 분포 프로그램의 사정으로 표측, 표두의 항목을 옮겨놓고 분포하는 것이 있는데, 통계표의 명시대로 하는 경우가 이것에 해당한다. 또 앞의 예 [표 12] 통계표를 [그림 34]와 같이 하지 않고 「틈틈이 일한자」를 「총수」의 "A"의부분으로 분포하여 "A"의 합계치를 먼저 구해서 "C"의 부분으로 옮기고 나서 "B"를 "A"에 더한다고 하는 방법을 취하는 것이 있다. 이때에는 배열교환(A에서 C)과 합계의 산출(A+B)이 조합되어 실행된다.

다섯번째는 제로 데이터의 복원기능이다. 「제로 데이터 (혹은 제로 레코드)」라는 것은 분포한 값이 「제로」의 셀이 써내는 단위 (후술)로 연속해 있는 데이터(레코드)이며, 통계표의 분류가 세세하게 되는 만큼 제로의 셀은 많아진다. 따라서 「제로 데이터」가 많아 지지만 합산의 산출에

는 제로 데이터와 관계없고, 제로 데이터를 삭제해서 처리시간을 줄일 수 있기 때문에 진단 프로그램에서는 제로 데이터를 뺀 상태로 처리한다. 그리고 가공편성 프로그램으로 통계표에서 요구하는 형식으로 복원하는 것이다. 제로 데이터의 발생은 분포 프로그램에서 자기매체(예를 들면 자기 테이프)로 써내는 형식에 따라 좌우된다. 써내는 형식은 메모리에 확보한 통계표를 표단위로 써내는 「표단위 써냄」과 통계표의 표측의 행단위로 써내는 「행단위 써냄」이 있고, 제로 데이터의 판단은 써내는 단위로 모든 내용이 「제로」인지 어떤지에 의해 행한다. 따라서 표단위로 써내는 방식으로는 하나의 통계표의 모든 셀이 제로때가 제로 데이터이며, 이와 같은 일은 적다. 행단위 써내는 방식으로는 표측의 각 행에 대해서 1행 전체가 모두 제로인 때가 제로 데이터이며, 표측 항목의 분류가 세세하면 제로 데이터의 발생은 많게 된다. 어느쪽의 써내는 방식이라도 써내는 데이터에 대해서 모두 제로인지를 점검하여 제로의 경우는 자기매체에 쓰지 않도록 하면 제로 데이터의 삭제가 행하여 진다. 제로 데이터의 복원은 간단하다. 구하여지고 있는 통계표를 메모리에 받아 그 내용을 모두 제로로 해두고, 입력 데이터(제로 아닌 데이터 만으로 되어 있다)를 읽어서 메모리에 분포하도록 하면된다. 제로 데이터로서 삭제한 데이터는 입력에는 존재하지 않으므로 메모리는 제로가 된 그대로 되어 있다.

여섯째는 파생표의 작성이다. 다른 통계표(부모)에서 수치를 꺼내서 새로운 통계표(자식)를 만드는 것으로 언뜻 보기에 유용하고 합리적인것 같이 생각되지만, 통계표의 변경에 따라 부모편 통계표의 변경이 자식의 통계표를 만들어지지 않게 하거나 또 자식의 변경이 부모에서 수치를 뺄 수 없게 되는 일이 있는것 같이 상호 영향을 주는것, 또 재작성하기에 처리가 표단위로 할 수 없는 염려가 있는것, 그리고 결정적인 것으로 프로그

램의 단순화에 역행하고 있다고 하는 이유 등으로 파생을 작성하는 경우는 적다. 그러나 「실수치」, 「구성비」의 2개로부터 만들어지는 통계표에서는 「실수치」의 통계표에서 「구성비」의 통계표를 작성하는 경우가 많다.

일곱째는 통계표의 인쇄를 위한 준비이다. 이것은 후속의 편집 프로그램으로 처리하는 방법이 좋은지 어떤지에 의해 정하지만 내용은 2가지가 있다. 하나는 데이터의 분할이고, 두번째는 프린트 개행(원고나 인쇄물에서 행을 바꿈) 문자의 부가이다. 이것에 대해서는 편집 프로그램에서 설명한다.

이상으로 중요한 기능에 대해서 설명하였는데, 가공편성 프로그램에서는 통계표마다 사용되는 기능이 다르고 오버레이라고하는 복잡한 구조의 프로그램으로 해서 하나의 프로그램으로 그 조사의 모든 통계표의 가공 편성을 행하게 하므로 범용의 프로그램으로 하는것은 어렵다.

(6) 편집 프로그램의 구성

편집 프로그램은 인쇄된 결과표를 보기쉬운 형태로 조정하는 것으로, 통계표의 작성은 몇번의 연산과정을 거쳐 편집 프로그램에 이르기 때문에, 그 도중의 과정에서 데이터는 계산하기 알맞는 형식을 하고 있고, 통계표로서의 보기 쉬움, 인쇄형식 등은 무시된다. 그래서 편집 프로그램에서는 계산중심의 데이터를 보기 쉽게 만들기 위하여 다음과 같은 기능을 가진다.

첫번째 기능은 통계표의 체제(형식)를 조정하는 것이다. 이것에는 「색인의 부여」와 「셀의 편집」이 있다. 색인의 부여는 인쇄된 통계표를 누가 보아도 이해할 수 있도록 하기 위해서이며, 통계표의 표번호나 표제의 「표제 색인」, 표두항목의 「표두 색인」, 표측 항목의 「표측 색인」의 3종류가 있고, 가다가나, 영문자 등을 사용해서 표시한다. 셀의 편집은

유효숫자의 좌측에 있는 제로를 소거하거나, 소숫점이나 쉼표를 넣어 숫자의 3자리마다 공백을 넣는 등이 대표적인 것이며 이들은 컴퓨터의 명령으로 미리 준비되어 있으므로 그것을 사용한다.

두번째 기능은 데이터의 분할이다. 컴퓨터의 인쇄장치는 1행에 인쇄할 수 있는 문자수(자리수)가 정해져 있으며, 대체로 130자리수 전후이다. 따라서, 인쇄해야할 데이터가 제한 자리수를 넘는 경우에는 몇개로 분할해야 한다. 예를 들면 [그림 35]는 2개로 분할한 예이다. 좌측의 제 1분할을 먼저 인쇄하며, 다음에 우측의 제 2분할을 인쇄한다.

그림35 데이터 분할의 예

남여·입주시기 전 주 지	(1) 총 수	(2) 0 세	(3) 1~4	(4) 5~9	(5) 세계급	(19) 80~84	(20) 85세이상
총 수					.		
출 생 시 부 터	제 1 분 류				.	제 2 분 류	
소 화 34년 이 전	(건반10컬럼분)				.	(후반10컬럼분)	
35년 ~ 39년	↓				.	↓	
40년1월~44년9월	제 1 항목				.	제 2 항목	
자市區町村내					.		
자市내타區					.		
縣내타市町村					.		
타 縣					.		
국 의					.		
44년 10월이후					.		
(상 동)					.		
남					.		
(상 동)					.		
여					.		
(상 동)					.		

새번제 기능은 통계표를 인쇄할때의 행간격이나 페이지를 바꾸는 부분을 지정하는 정보를 부여하는 것이다. 이것을 CCC(캐리지 콘트론 캐릭터)라고 부르며, 0에서 9까지의 숫자로 준다. 9를 제외한 숫자는 1행을 인쇄하고 나서의 행을 바꾸는 수를 나타내며 1은 1행을 인쇄하면 즉시 아래의 행으로 행을 바꾸는 것으로 「싱글 스페이스」라고 부른다. 2는 1행을 인쇄하고 1행 거른 다음의 행으로 행을 바꾸는 것으로 「더블 스페이스」라고 부르고 있다. 2~8까지는 인쇄행 간에 적당한 공백행을 새로 만들어 내는 기능을 가지고 있지만, 0은 인쇄해도 행을 바꾸지 않는 것으로 겹쳐서 인쇄된다. 9는 인쇄하고 나서 페이지를 바꾸는 기능이 있다.

편집 프로그램에서는 편집완료된 데이터를 자기매체(주로 자기 테이프)에 출력하고 인쇄 전용의 별도 프로그램에 의해 인쇄한다. 그렇지 않으면 대량 인쇄의 통계표를 심사해서 「인쇄가 불선명」, 「종이가 파손되어 있다」 등의 인쇄장치 측의 문제가 발생했을때 대처할 수 없기 때문이다. 자기 테이프에 출력되어 있으면 특정한 페이지, 표를 재차 인쇄할때 이미 인쇄하면 좋은 만큼의 데이터가 기록되어 있으므로 그 데이터를 컴퓨터로 찾아내는 수고 정도 이지만, 직접 인쇄를 채용하고 있으면 편집 프로그램 각 기능을 다시고쳐 하게 되며 특정한 페이지, 표만을 인쇄 할 수는 없다.

10. 기타 프로그램의 구성

집계수치의 보기쉬움의향상, 통계보고서 작성의 신속화, 결과표 심사의 효율 향상 등을 위하여 보조적(집계 그것이 아니라고 하는 의미로)인 프로그램이 다음과 같이 몇개 있다.

- (1) 한자 프린터(일본어 페이지 프린터)용 프로그램

컴퓨터에서 계산된 통계숫자는 이미 기술한 「편집 프로그램」으로 처리하는 것에 따라 컴퓨터의 출력장치인 라인 프린터(인쇄장치)를 위한 데이터가 만들어져 인쇄되지만, 라인 프린터는 컴퓨터가 원래 미국에서 만들어진 것이기 때문에 영자, 숫자 및 약간의 기호를 인쇄할 수 있는 기능밖에 가지고 있지 않으므로 예를 들면 표제나 표측, 표두의 항목 부분을 일본어로 인쇄할 수는 없다. 그 대신에 라인 프린터의 활자는 일정한 크기이며 문자와 문자사이의 간격과 행간이 일정하기 때문에 「편집프로그램」은 어려운 종류의 프로그램이 아니므로 극히 용이하게 결과표의 인쇄를 할 수 있는 것이다.

그러나 일본에 컴퓨터가 보급하게 됨에 따라 일본어로 인쇄를 할 수 없는 것은 불편했기 때문에 일본어 인쇄장치의 연구 개발이 행하여져 현재는 인쇄장치로서 충분할 만큼 완성되어 있다. 통계표의 인쇄에 「한자 프린터」를 이용한 것은 통계국·통계센터에서는 1970년 국세조사부터 이다. 한자 프린터를 이용하는 것의 유리한 점으로서 2가지를 들 수 있다. 첫번째는 통계표의 「보기쉬움」이다. 표두나 표측, 표두의 항목이 영어나 로마자로 인쇄되는 것과 일본어로 인쇄되는 것중에 어느쪽이 보기 쉬운가는 새삼스럽게 말할 필요가 없다고 생각한다. 「보기쉬움」이 중요한 것은 제표의 최종단계에서 수작업에 의한 「결과표 심사」를 할때 인쇄된 결과표가 수회에 걸쳐 여러가지 판단기준에 의해 심사되기 때문이다. 두번째는 통계보고서 작성의 신속화를 도모하기 위하여 「版下」(판목을 뜨기 위한 밑글씨)로서의 이용이다.

통계보고서를 위한 「판하」의 작성에는 2가지의 방법이 있다. 하나는 표제, 표측항목, 표두항목을 인쇄한 대지(사진이나 그림을 붙이는 두꺼운

종이)를 미리 준비해 놓고 컴퓨터의 라인 프린터에 의해 인쇄된 집계숫자 부분을 잘라내어 그 대지에 붙이는 방법과 또 하나는 표제, 표측항목, 표두항목의 일본어의 부분과 집계숫자 부분을 컴퓨터에 의해 조합시켜 「한자 프린터」로 인쇄하는 방법이다.

전자에 대해서는 통계표마다 대지를 준비하여 컴퓨터로 부터의 숫자부분을 도려내고 붙이는 정도이므로 각별히 프로그램에서 주의할것은 없다.

후자에 대해서는 한자 프린터의 기능을 전제로 해서 「일본어 편집 시스템」을 개발해야만 한다. 게다가 어떠한 조사에도 대응할 수 있도록 범용화가 요구된다. 참고로 통계국·통계센터에서 개발한 시스템에 대해서 기술하면, [그림 36]과 같이 5가지의 부분으로 구성되어 있다.

첫째는 레이아웃 시트의 작성이다.

레이아웃 시트라고 하는것은 한자 프린터의 최소 활자 폭과 최소 행을 바꾸는 폭으로 모눈눈금을 만들어 그것에 통계표의 체재를 고려해서 이기한 것이다. 이 레이아웃 시트는 모든 통계표에 대하여 작성한다.

둘째는 표제나 표측, 표두의 각 항목에 사용되어 있는 문자(일본어, 영어 등)의 작성이다. 이 문자는 레이아웃 시트에서 골라내어 워드프로세서와 같은 일본어 입력기를 사용해서 한자변환을 하면서 소요 문자를 플로피 디스크에 기록한다. 문자는 통계표마다 전부입력하는 것이 원칙이지만, 통계표의 성격에서 예를 들면 「직업산업 분류」나 「都道府縣」과 같이 공통해서 사용되는 문자는 한번만 입력하며 나중에 컴퓨터에 의해 카피를 한다.

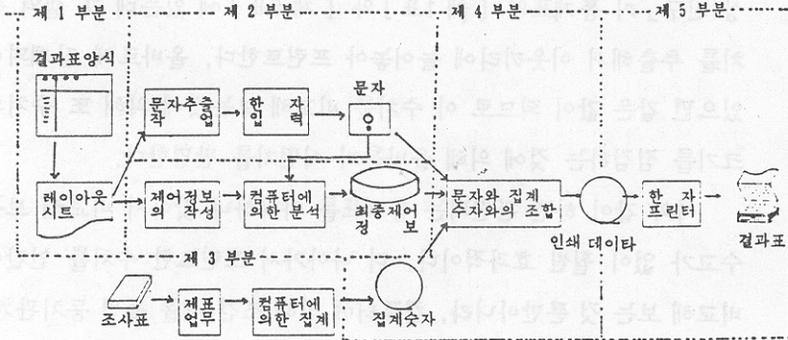
셋째는 한자프린터로 인쇄하는 종이의 「어느위치」에 「어느정도의 크기」로 인쇄하는것 인가를 지정하는 정보(제어정보)의 작성이다. 수작업으로 활자를 뽑아서 조판을 하는 처리과정을 컴퓨터가 처리하게 하는 것이

다. 제 2단계에서 입력한 문자와 제 3단계의 제어정보를 컴퓨터로 분석해서 최종의 제어정보를 작성한다.

넷째는 최종 제어정보에 따라서 입력한 문자와 집계숫자 부분을 조합해서 인쇄행을 만드는 프로세스이다. 이것에 따라 한자 프린터로 인쇄하는 데이터가 만들어 진다.

다섯째는 한자 프린터에서의 인쇄이다. 이 「일본어 편집 시스템」에서는 제 3의 제어정보를 작성하기 위하여 통계표의 양식을 레이아웃시트에 이기하고 있다. 모눈눈금의 레이아웃시트에 기입하고 있는 최대의 이유는 인쇄해야 할 문자의 위치(1 페이지의 위에서부터의 행수, 왼쪽 끝에서의 자리수로 나타낸다)를 알기 위해서이며, 눈금이 없는 백지에서는 그 위치를 알수 없기 때문이다.

그림36 일본어편집 시스템



통계표의 양식에서 보고서로 만들때의 체제(1페이지 중에서 표제의 위치나 글자의 크기 등 겉보기의 아름다움)를 생각하면서 레이아웃시트에 기입하

는 것은 의외로 수고와 시간이 걸리고, 또 기입사항의 정정이나 위치의 변경이 엄청난 작업이 되는 것이므로 앞으로는 대형의 화면 표시장치를 보면서 워드프로세서와 같은 감각으로 통계표의 체재를 생각하는 방향으로 나아가야 할 것이다.

(2) 결과표 심사용 프로그램

결과표가 올바른지 어떤지의 판단을 최종적으로는 인간이 행하여야 하는 것이며, 그를 위해서 필요한 정보를 컴퓨터가 제공한다고 하는 것이 결과표 심사용 프로그램의 기본적인 사고방식이다. 심사용 프로그램에서의 필요정보의 제공에는 다음의 3가지 방법이 있다.

(가) 프린트 법

결과표 심사에서 필요하게 되는 수치를 각각 관련하는 그룹으로 나누어서 프린트하는 방법이다. 예를 들면 노동력 조사에서 「15세 이상 인구」가 통계표의 [제 1표]와 [제 6표]에 있는데 그 셀의 수치를 추출해서 이웃끼리에 늘어놓아 프린트한다. 올바르게 집계되어 있으면 같은 값이 되므로 이 수치를 비교해 보는 것에 의해 또 수치의 크기를 점검하는 것에 의해 올바른지 어떤지를 판명한다.

이와 같이 하면 관련되는 통계표를 하나하나 찾아서 비교해 보는 수고가 없이 훨씬 효과적이다. 더 나아가서 프린트한 수치를 인간이 비교해 보는 것 뿐만 아니라, 컴퓨터에 의해 조건(예를 들면 등치관계, 대소관계 등)의 충족 여부를 체크하고 만약 충족시키고 있지 않으면 그것을 나타내는 「기호」를 부여하여 프린트하면 더욱 효율적이 된다. 이 프로그램은 심사의 대상으로 되는 셀을 추출하여 프린트하는 정도

는 것은 의외로 수고와 시간이 걸리고, 또 기입사항의 정정이나 위치의 변경이 엄청난 작업이 되는 것이므로 앞으로는 대형의 화면 표시장치를 보면서 워드프로세서와 같은 감각으로 통계표의 체재를 생각하는 방향으로 나아가야 할 것이다.

(2) 결과표 심사용 프로그램

결과표가 올바른지 어떤지의 판단을 최종적으로는 인간이 행하여야 하는 것이며, 그를 위해서 필요한 정보를 컴퓨터가 제공한다고 하는 것이 결과표 심사용 프로그램의 기본적인 사고방식이다.

심사용 프로그램에서의 필요정보의 제공에는 다음의 3가지 방법이 있다.

(가) 프린트 법

결과표 심사에서 필요하게 되는 수치를 각각 관련하는 그룹으로 나누어서 프린트하는 방법이다. 예를 들면 노동력 조사에서 「15세 이상 인구」가 통계표의 [제 1표]와 [제 6표]에 있는데 그 셀의 수치를 추출해서 이웃끼리에 늘어놓아 프린트한다. 올바르게 집계되어 있으면 같은 값이 되므로 이 수치를 비교해 보는 것에 의해 또 수치의 크기를 점검하는 것에 의해 올바른지 어떤지를 판명한다.

이와 같이 하면 관련되는 통계표를 하나하나 찾아서 비교해 보는 수고가 없이 훨씬 효과적이다. 더 나아가서 프린트한 수치를 인간이 비교해 보는 것 뿐만 아니라, 컴퓨터에 의해 조건(예를 들면 등식관계, 대소관계 등)의 충족 여부를 체크하고 만약 충족시키고 있지 않으면 그것을 나타내는 「기호」를 부여하여 프린트하면 더욱 효율적이 된다. 이 프로그램은 심사의 대상으로 되는 셀을 추출하여 프린트하는 정도

이므로 구조는 간단하지만, 프린트 방법으로는 합계와 내역이 일치해 있는지 어떤지의 체크를 하지 않으므로 수작업으로 확인해야 한다.

(L) 자동심사법

「자동심사」라고 하는 호칭법에는 약간의 설명이 필요하다. 앞에도 언급하였지만 올바른지 어떤지의 최종의 판단을 인간이 내리는데, 「자동심사법」에 따랐으므로 인간이 판단하지 않아도 좋다고 하는 것은 아니다. 프린트법에서 합계와 내역과의 관계가 계산되지 않는 부분을 보충해서 인간의 판단을 필요로 하지 않는 기계적으로 할 수 있는 것을 거의 전부 처리한다는 의미에서 이 호칭법을 사용한다. 기계적으로 할 수 있는 기능을 포함하면, 프로그램이 상당히 대규모가 되므로 이것도 「결과표 자동심사 시스템」이라고 하는 처리체계를 개발해야 한다.

프로그램의 구성으로서는 결과표의 난외항목, 표측항목, 표두항목의 구조를 각각 정의하는 것에 따라 심사지시(합쳐야 할 셀이라든가, 합계와 내역의 계산식 등)를 4종류의 외부 정보에 의해 자동적으로 생성하는 방법을 채용하고 있다. 외부 정보는

- ① 심사대상 데이터에 대해서 주로 물리적 성질을 기술한 것이며
- ② 대상데이터의 분류항목
- ③ 결과표의 구조
- ④ 예외적 심사지시이며 이것들을 통계표마다 기술하는 것에 따라, 「검산」, 「정치 체크」, 「표간대조」, 「항목간 체크」를 행하는 것이다.

(ㄷ) 데이터베이스법

심사에 필요한 정보를 데이터베이스를 이용해서 얻는 방법이다. 심사정보에는 정형적 정보, 예를 들면 김산, 항목간 대조, 시계열 비교 등과 같이 집계시에 그 대상이나 방법이 정해져 있는 것과 결과 수치를 보면서 여러가지의 각도에서 분석적으로 점검하는 비정형적 정보가 있다.

전자에 대해서는 결과표의 집계시에 병행해서 산출하면 좋고, 프린트법이나 자동심사법이 여기에 속한다. 후자에 대해서는, 예를 들면 집계수치가 사회적 배경과 모순하는 것 같은 경우로 조사표 데이터를 점검 할때 데이터베이스를 이용해서 해당 데이터를 검색하는 것이 데이터베이스 이용법이다. 집계수치의 이상함에 의해 원인탐구의 방법이 확정되는 경우에 데이터베이스법의 위력이 발휘되지만, 정형적 정보의 제공에도 산출의 기초로 되는 수치의 수집에 있어서 데이터베이스는 유효하게 활용되어야 하며, 금후 통계 데이터베이스가 정비되는 것과 더불어 더욱더 이용이 증가될 것이다.

11. 프로그램 개발

통계집계에 필요한 프로그램의 기능에 대해서 그 개략을 설명해 왔지만, 마지막으로 프로그램의 작성과정에 대해서 언급하고자 한다. 이것은 대개 5개의 부분으로 구성되어 있다.

(1) 분석

통계조사의 구조를 이해하는 것이 맨처음이다. 그렇게 하기 위해서는

「조사의 안내서」, 「조사표」, 「통계표 서식」 등의 조사관계 자료를 수집하며, 첫번째로 조사의 실사가 어떠한 형식으로 행하여지는가를 파악하는 것이다. 특히 都道府縣이나 市區町村에서의 조사표 정리방법이나 통계국·통계센터에의 조사표 송부방법을 충분히 이해해야 한다. 두번째로 조사표에서 어떠한 종류의 통계표를 작성하고 있는가를 검토한다.

여기에서는 통계표를 그냥 막연히 바라보고 있는것이 아니라 조사항목이 통계표의 어느 장소에 사용되고 있는지, 또 수작업으로 그 통계표를 만든다고 하는 경우의 순서를 명확히 이해할 수 있는가 하는 관점에서 수집자료를 분석한다.

다음으로 제표 관계의 자료(예를 들면 엔트리에서의 「입력요령」, 「체크요령」, 내용검사 및 부호기입 등의 「제표업무 수속」 등)를 검토하며 조사표가 어떤 배열(순번)로, 어떠한 부호가 붙여지며 컴퓨터에 주어지는 것인가를 이해함과 동시에 「체크요령」에서 대강의 체크 처리 순서를 검토한다. 체크처리 순서는 수작업부분에 큰 영향을 주기 때문이다.

(2) 시스템 설계

분석단계는 각각의 자료를 기본으로 조사의 구조, 조사대상, 조사표, 제표업무, 통계표라고 하는 말하자면 점으로서 존재하는 개별사항의 이해가 중심이었지만 그들의 점을 서로 연결하는 선을 생각하지 않으면 안된다. 이것이 시스템 설계이다.

시스템 설계에서는 우선 첫째로 집계 기본방침을 정할 필요가 있다. 예를 들면 집계 시스템으로서 기능별의 프로그램으로 할것인지, 프로그램은 반복 사용되는지 아니면, 한번만 그 과정을 거치는 것인지, 데이터 체크에서 정정의 방법은 온라인 방식인지 배치 방식인지, 그밖에도 여러가지

종류가 있지만 시스템의 골자를 확정하는 것이다. 둘째는 통계의 질을 어느 수준으로 유지하기 위하여 필요한 컴퓨터로 부터 얻는 자료의 종류를 정한다. 논리적으로 집계 순서를 생각해 프로그램을 작성하면 통계표는 얻어지지만 대규모 조사와 같이 제표업무가 오랜기간 계속되는 경우에는 집계 질을 일정하게 유지하기 위하여 각종의 자료를 기본으로 한 적절한 조치를 필요로 한다. 그를 위해서는 예를 들면 「에러 데이터의 건수」, 「컴퓨터로 자동수정한 건수」, 「주요 조사항목별의 데이터 갯수」 등을 산출하고 판단자료로 사용한다. 셋째는 조사표(입력화일)가 여러가지의 컴퓨터 처리를 거쳐 결과표(출력 화일)에 이르기까지의 처리의 순서와 화일을 설계한다. 국세조사와 같이 통계표의 구분, 즉 「추출속보 집계」, 「추출 상세집계」, 「제 1차 ~ 제 3차 기본집계」, 「근무지·통학지 집계」 외에 여러가지 있지만 순서나 화일의 중복을 피해서 설계를 한다.

(3) 프로그램의 작성

처리순서 및 화일의 형식을 확정하면 개별 프로그램의 설계에 들어간다. 우선 프로그램에 요구되는 기능이나 이용가능한 장치, 기억용량 등 하드웨어의 제약조건을 고려하면서 「프로그램 사양서」를 작성한다. 이 사양서는 프로그램 로직(논리, 처리수속)을 간결하게 나타낸 것이고, 지금부터 플로우차트를 그려 프로그램의 코딩을 행한다. 플로우 차트나 코딩에는 컴퓨터 부문에서의 표준형식이 있고, 통일해서 부주의에 따른 실수를 될 수 있는한 막도록 하고 있지만 여기까지의 작업은 전부 프로그래머의 머리 속에서 조립된 것으로, 올바른지 어떤지 객관적으로 알기 위하여 프로그램의 테스트를 할 필요가 있다.

(4) 프로그램 테스트

이 부분은 3단계로 나누어 행한다. 제 1단계는 프로그램을 컴파일 (기계어로 번역하는 것 - 컴퓨터는 기계어로 밖에 움직이지 않으므로) 하는 것이다. 프로그램 언어에는 영문법과 같이 작성하는 방법의 규칙이 있고, 그것에 따라서 프로그램을 작성하지 않으면 컴파일 할때에 컴퓨터의 쪽에서 「에러」가 지적된다. 따라서 문법위반과 같은 작성방법에서 기인하는 에러는 이 단계에서 전부 올바르게 할 수 있다. 제 2단계는 「시험 테스트」이다. 프로그래머의 잘못된 생각에 따른 에러는 컴파일에서는 발견할 수 없다. 예를 들면 본래 가산해야 할 부분을 감산해 버렸어도 에러의 지적은 되지 않는다.

따라서 실제로 컴퓨터를 움직여서 답이 올바른지 어떤지 확인할 수 있는 방법이 필요하다. 프로그램을 단일체로서 테스트하는 것을 「시험 테스트」라고 부르며, 프로그래머의 가상대로의 계산 결과가 얻어지도록 데이터를 만들어 테스트를 한다.

여기에서 중요한 것은 계산 결과를 예상할 수 있는 데이터를 사용하는 것이다. 만약 적당하게 고른 데이터를 사용하면 결과가 올바른지 어떤지를 판정할 수 있도록 그 데이터를 수집해 두어야 하며, 그 수고는 무시할 수 없다. 제 3단계는 「최종점검 테스트」이다. 시험테스트에서 개개의 프로그램에 에러가 없어도 각 프로그램을 처리의 흐름에 따라서 연속해서 움직이면 무엇인가의 에러가 발생할 수 있다. 그래서 실제의 데이터를 1부분, 혹은 1개월분을 사용해서 종합 테스트를 하는 것이다. 이 목적은 대량의 데이터를 흐르게 하는 것에 따른 예상외의 데이터가 현실로 존재하는지도, 프로그래머의 잘못된 생각에 따른 집계결과 오류의 발견에 있다. 특히 후자

의 에러는 시험테스트에서 발견할 수 없다.

테스트 데이터를 만드는 프로그래머가 잘못된 논리로 프로그램을 만들고 있어도 본인은 깨닫지 못하기 때문에 테스트 데이터를 만들어도 올바른 답(실제로는 틀림)을 얻은 것이라고 생각하고 있기 때문이다. 화이날 테스트에서는 집계결과를 본 연산과 같도록 「결과표 심사」를 행하며 제 3자의 판정을 기다리는 것이다.

(5) 도큐먼트의 정리

프로그램이 완성되면 프로그램 도큐먼트의 정리를 한다. 프로그램 작성에서는 집계의 전체에 관계되는 도큐먼트, 개별 프로그램의 도큐먼트 등과 수많은 자료가 만들어 지는데 이것들이 흩어지지 않도록 하나의 화일로 통합하는 일이 중요하다. 동시에 컴퓨터의 조작 부문에 필요한 도큐먼트를 작성한다. 그것들은 화일을 중심으로 한 처리의 흐름도, 화일보존의 정보, 연산시간의 추계, 화일레코드 수의 추계 등으로 구성되어 있지만, 이 가운데서 연산시간의 파악은 대단히 어렵다. 그 이유는 현재의 컴퓨터가 프로그램을 다중처리 하고 있기 때문에 프로그램들이 서로 영향을 주어 처리 속도의 정확한 추계를 할 수 없기 때문이다. 그 때문에 일반적으로는 화이날 테스트에서의 처리시간으로 추계하지만 과거 처리의 실적 시간이나 유사조사의 실적시간을 참고로 해서 정한다.

(6) 프로그래밍 계획

지금까지 설명한 (1)~(5)의 업무를 행하기 위하여 프로그래밍 계획을 작성한다. 각각의 업무에 대해서 「공량(총 업무량)」을 추계하여 「능력」로 부터 프로그래머의 수나 프로그래밍 기간을 산출하는 것은 제표계획

과 아무런 변화가 없지만 공량이나 능률이 수작업으로 할때의 경우와 달라 추계를 하기 어렵다. 예를 들면 분석 시스템 설계와 같은 주로 생각하는 업무에서는 허비한 시간과 업무의 공량과는 비례하지 않는다. 따라서 지식인 작업에 있어서의 공량이나 능률을 어떻게 측정하는 지가 문제이다. 프로그래밍 업무는 몇 사람으로 이루는 팀을 편성해서 행하는 것이 일반적이다. 팀 구성은 소규모 조사에서 3~4인, 중규모 조사에서 6~7인, 대규모 조사에서 10인 전후이며 가장 경험이 풍부하고 기술적으로도 뛰어나 있는 자를 팀리더로 한다. 리더는 분석, 시스템 설계를 주로 행함과 동시에 팀의 각자에게 능력에 따라서 어떠한 프로그램을 분담시킬지 정한다. 프로그램을 완성하기 까지 필요로하는 시간은 소규모 조사에서 3개월, 중규모 조사에서 6개월, 대규모 조사에서는 거의 1년정도 이지만, 팀의 구성원이라든가 소요기간이라든가와 같이 대충의 짐작이며 조사의 질적인 어려움, 구성원의 능력, 관련자료의 정비상황 등으로 상황은 변동한다. 따라서 지식 집약형 업무의 공량이나 능률을 노동집약형의 그것과 같은 확실도로 추정하는 것이 금후의 과제가 된다.

지금까지 통계조사의 제표에 대해서 수작업면과 컴퓨터의 면에서 해설하였지만, 통계 그것이 중요하기는 하지만 화려하지는 않고 그 중에서도 제표에 대해서는 더욱 알려지지 않고 있다.

이번의 간단한 글에서 다소나마 이용자에게 참고가 되었으면 하는 바램입니다.