

통계자료의 수집 및 분석의 타당성에 관한 연구¹⁾

-광주,전남지역을 중심으로-

이 화 영²⁾

요 약

비학술단체에서 통계자료의 수집과 분석시에 반드시 고려해야할 항목을 나열한 점검표틀 개발하였다. 개발된 점검표틀 활용하여 광주,전남지역의 단체, 언론,금융기관, 산업체에서 일년간(1991년 4월 5일 부터 1992년 4월 4일까지) 통계조사결과 발표된 결과물에 대하여 자료수집과 분석의 타당성을 검토하였다.

1. 서 론

현대사회는 정보화시대이고 모든 분야에서 많은 양의 정보가 보편화된 개인용컴퓨터와 통계패키지의 이용으로 쉽게 통계처리되고 있다. 특히 최근들어 사회단체, 언론,금융기관 및 산업체에서 관심있는 문제에 대한 실태파악을 위하여 통계조사를 실시하고 그 결과를 대중매체인 신문, 라디오 또는 텔레비전을 통하여 많이 발표하고 있다. 그런데 발표결과의 대부분이 dot수분포를 설명하는 기술통계에 그치고 있어 과연 많은 시간과 비용을 들여 수집된 자료들이 얼마나 깊이 있는 통계적 분석이 되어 현대통계학의 목표인 추측통계를 위하여 활용되는지 궁금하다. 아뭏든 이같은 결과는 관련분야 정책수립의 기초자료로 활용되고 있을 것으로 사료된다. 따라서 모든 분야에 있어서 의사결정의 기초로 활용되는 통계자료의 수집과정이 체계적이며 분석시에는 타당한 통계적기법이 적용되었는지 통계전문가들이 반드시 파악해 보아야 할 분야로 생각된다.

우리나라에서도 지방자치제가 도입되어 각종 단체나 기관에서 주민의 생활파악을 할 수 있는 통계, 즉 지역의 특성에 맞는 통계가 절실히 요구되고 있다(국토개발연구원 : 1986, 이화영 : 1991). 그러나 지금까지 우리나라는 모든 정책적 결정이 중앙에서 이루어져 왔기 때문에 지역통계가 상당히 미비한 상태에 있다. 결국 해당지역에서 여러가지의 표본조사를 실시하여 필요한 통계가 작성될 것으로 예상되는 바, 조사는 과학적인 절차를 거쳐서 실시되어야 할 것이며 수집된 자료는 자료의 형태에 따라 타당한 분석기법이 적용되어야 한다. 그렇게 함으로써 결과는 정책 및 의사결정의 기초자료로 활용될 수 있을 것이다.

본 연구는 우리나라의 각종 단체 및 산업체에서 통계자료를 어떻게 수집하고, 수집된 자료의 분석에는 어떠한 통계적기법이 적용되었는지 파악하여 보는데 있으나 연구대상지역은 광주전남지역으로 제한 되었다.

1) 이 논문은 1991년도 교육부지원 한국학술진흥재단의 자유공모(지방대학육성)과제 학술연구 조성비에 의하여 연구되었음.

2) (520-714) 전라남도 나주시 대호동 252 동신대학교 통계학과

2. 통계조사의 활용실태

광주전남지역에 있는 각종 단체, 금융기관과 언론기관의 명단은 1992년 전화번호부를 이용하여 85개 단체, 46개 금융기관, 10개 언론기관이 조사대상에 포함되었다. 또한 광주직할시와 전라남도에서 발행한 기업체현황을 보면 광주시에 1,339개, 전라남도에 1,488개의 산업체가 있음을 알 수 있으나 이 중에서 산업체의 규모가 작은 곳보다는 큰 곳에서 통계조사가 자주 실시될 것으로 사료되어 200인 이상의 종업원을 가진 64개 산업체만 뽑아 조사대상에 포함시켰다. 따라서 통계조사 활용실태 파악을 위한 설문지를 이들 205곳에 우편으로 발송하여 우편조사를 실시하였다. 일정기간이 지난후에 응답이 없는 곳에 대해서는 전화설문과 훈련된 조사원이 직접 방문하여 질문하는 방법을 병행하였다. 그 결과 97곳(43개 단체, 21개 금융기관, 8개 언론기관, 25개 산업체)으로 부터 설문에 대한 응답을 받았으며, 그 응답을 기초로 분석된 결과를 설명하면 다음과 같다.

<표 1> 통계조사 활용실태

문항	내 용	빈도	백분비
조사실시여부	예	39	40.2
	아니오	58	59.8
조사주관	본사	8	20.5
	자체	31	79.5
조사회수	1회	18	46.1
	2회	9	23.1
	3회 이상	12	30.8
조사내용의 통계집계	예	33	84.6
	아니오	6	15.4
통계집계부서	자체	16	48.5
	외부용역	8	24.2
	기타	9	27.3
통계집계결과의 외부발표	예	11	33.3
	아니오	22	66.7
발표형태*	보고서	4	14.8
	신문	10	37.0
	라디오	8	29.6
	T.V.	5	18.5
통계집계를안한이유	통계집계할 사람이 없어서	1	16.7
	집계없이 원하는 정보를 얻어서	5	83.3
조사를 전혀 실시하지 않은 이유	조사를 기획, 수행할 사람이 없음	13	22.4
	본사에서 알아서 함	30	51.7
	기 타	15	25.9
장래 조사실시의 필요성	있 다	22	37.9
	없 다	36	62.1

*: 결과의 외부발표시 중복발표 허용.
 각 항목에 대한 타당성의 평가판정에는 다음과 같은 4개 범주를 적용하였다. 즉,
 PC(Performed Correct) : 정확하게 수행됨
 PI(Performed Incorrect) : 부정확하게 수행됨
 NG(Neglected) : 고려하지 않았음
 NA(Not Applicable) : 해당사항 없음

통계조사를 실시한 경험이 있다고 응답한 곳은 전체의 40.2%인 39곳이었으며, 이들중 조사를 주관한 곳은 본사라고 응답한 경우도 20.5%나 되었다(표1 참조). 이는 주로 응답한 기관 및 단체들이 서울에 본부를 두고 광주전남에는 지사형태로 있기 때문이라 사료된다. 지난 일년간 총 조사회수는 1회 실시한곳이 46.1%로 거의 반을 차지하는 반면에 3회 이상 실시한 곳도 30.8%나 되었다. 이렇게 조사된 내용은 전부가 통계로 집계되지 않았고 약 84.6%만이 통계집계 되었으며 나머지는 통계집계없이 조사내용이 활용되었음을 알 수 있다. 통계로 집계된 곳 중의 약 1/4이 외부에 용역을 맡겨 처리하였다고 응답하였는데 이는 통계처리할 수 있는 인력이 없음을 시사한다고 하겠다. 또한 집계된 결과가 외부에 발표된 것은 1/3뿐이고 나머지는 외부에 발표 없이 내부보고 자료로만 활용되었음이 파악되었다. 외부에 발표된 것 중에서 보고서의 형태를 갖춘것은 4곳뿐이었고 나머지는 신문, 라디오 또는 텔레비전에 간단히 발표된 것이다.

조사를 전혀 실시하지 않은 58곳에서 조사 미실시의 이유중 약 52%는 본사에서 모든것을 처리하기 때문이라고 응답한 반면에, 약 22%에 해당하는 곳에서는 조사를 기획수행 및 통계처리할 직원이 없다고 응답하여 조사대상기관에 통계조사를 담당할 수 있는 인력의 수요는 높을 것으로 예상된다.

3. 점검표

3.1. 점검표의 구성

점검표는 통계조사를 실시하고 자료의 분석을 하는 기관 및 산업체에서 통계학에 대한 이해의 부족으로 범할 수 있는 통계적 오류를 가능한한 줄일 수 있도록 이들이 고려해야할 사항을 참조하여 통계자료수집 및 분석기법을 좀 더 효율적으로 활용할 수 있도록 그 지침을 제시하여 주기 위하여 개발된 것이다.

최와 이(1990)는 Schor와 Karten의 통계적 평가법과 Ford와 Tortora의 점검표를 토대로 하여 통계적기법활용을 위한 3단계(연구논문의 설계과정인 통계적추론 이전의 단계, 통계적 추론의 단계, 결론도출의 단계)점검표를 제시하였다. 그들이 제시한 점검표는 통계기법을 어느 정도 활용하는 학술논문을 평가하기 위한 목적으로 작성된 것이기 때문에 초보적인 통계분석기법을 주로 활용하는 비학술단체의 통계조사분석의 타당성평가를 위하여 적용하기에는 어려움을 안고 있다.

3.2. 점검표

통계자료의 수집 및 분석의 타당성평가를 위하여 개발된 점검표는 <표2>와 같다.

<표 2> 점 검 표

1-단계 (자료수집)				
	PC	PI	NG	NA
1. 연구문제의 형성				
2. 연구가설				
3. 연구의 대상				
1) 목표모집단과 조사모집단				
2) 모집단의 규모, 구조, 특성				
4. 연구설계				
1) 표본추출방법				
2) 대표성				
3) 표본의 크기				
4) 자료수집법				

2-단계(자료분석과 해석)				
	PC	PI	NG	NA
1. 분석기법의 선택				
1) 자료의 성격				
2) 적용절차				
3) 사용된 통계패키지				
2. 결과의 해석				
1) 주요통계량				
2) 유의수준(p값)				
3) 도표법이용				

3.3. 점검항목의 제문제

3.3.1. 자료수집

통계조사의 기획에 있어서 무엇보다도 가장 중요한 것은 그 조사에서 탐구하고자 하는 연구문제(research problem)는 명확하고 간결하게 진술하여야 하며, 필요에 따라 개념규정도 곁들여져야 한다. 그리고 연구문제가 통계조사에서 검증 또는 취급될 수 있는 범위에 한정 된다는 것을 보여 주어야 한다. 또한 연구문제의 중요성은 실질적 유관성과 이론적·방법론적 유관성(김경동; 1988)의 기준에 비추어 명시되어야 한다.

탐구하고자하는 연구문제가 지나치게 일반적이거나 추상적이면 조사연구하는데 애로가 따르게 되므로 이를 보다 구체화할 필요가 있는데, 이러한 구체화는 가설검정을 통해서 획득될 수 있다. 따라서 검증용가설(귀무가설과 대립가설)은 명확하게 제시되어야 하며, 가설이 연구조사와 이론에 대하여 지니는 의미를 밝힌다.

가설이나 이론을 적용할 모집단이 명확하게 규정되어야 한다. 모집단은 조사목적에 의하여 개념상 규정된 목표모집단과 표본을 추출하기 위하여 규정된 조사모집단으로 구분될 수 있는데 이 두 집단은 원칙적으로 일치하여야 한다. 그러나 실제문제에 있어서 일치되지 않는 경우가 많다. 조사모집단은 목표모집단보다 제한되어 있는것이 보통이며, 표본에서 얻은 결과는 조사모집단에 적용되어야지 목표모집단에 적용하면 안된다. 따라서 합리적인 표본을 얻기 위하여 모집단의 규모, 구조와 특성을 살피는 것은 중요하다.

연구설계란 연구가설을 평가하기 위한 구조, 계획 및 전략이라고 할 수 있겠다. 이것은 좋은 자료, 즉 믿을만하고 마땅한 자료 또는 신뢰성(reliability) 및 타당성(validity)있는 자료수집의 문제로 귀결된다. 그러기 위해서 고려되어야 할 사항은 모집단에서 표본의 추출방법, 대표성, 표본의 크기 및 자료수집법이 명시되어야 한다.

3.3.2. 자료분석과 해석

조사가 완료되면 조사표에 기입된 내용을 심사(editing)한 뒤, 부호화(coding), 제표(tabulating)등 기계적인 절차를 거쳐 연구문제 해결을 위한 분석과 해석이 뒤따르게 된다. 통계적 분석에 앞서 자료의 성격에 대하여 충분히 검토해 보는것이 바람직하다. 여기서 자료의 성격이라 함은 변수를 수량화 할 때 그 측정에 사용한 척도의 종류와 그 척도의 연속성의 문제이다(Loether and McTavish; 1980, Blalock; 1972). 왜냐하면 척도의 종류 또는 측정의 수준에 적합하지 못한 통계적기법을 적용하여 아무리 고급 통계적분석을 한다든지, 그럴듯한 결론을 유도한다해도 그 가치는 무의미하게 되기 때문이다. 따라서 자료의 성격에 따라 분석기법이 적절하게 적용되었는지 살펴본다.

개인용컴퓨터와 통계패키지의 보급중대로 연구자들이 비교적 쉽게 자료의 통계적분석을 할 수 있게 되었다. 국내에서 주로 활용되고 있는 통계패키지로는 SPSS (Statistical Package for Social Sciences), SAS (Statistical Analysis System), IMSL (International Mathematical and Statistical Library), SYSTAT (The System for Statistics), TSP (Time Series Processer)와 SHAZAM을 들 수 있겠다. 고급통계분석기법(예를들면 인자분석, 이원분산분석)이 적용될 때 모든 통계패키지가 동일한 결과를 제공하지 못하는 경우가 있다고 지적되어 왔다(Francis ;1973, MacCallum ; 1983, 김병천 ; 1987). 본 연구에서 취급된 결과물들이 사용한 분석기법은 비교적 초보적인 것으로 큰 문제가 제기되지는 않겠지만, 향후 연구자를 위하여 자료분석시 사용된 통계패키지명을 명시하는것도 필요하겠다.

타당한 분석기법을 적용하여 얻어진 결과의 해석은 연구문제에 대한 적절한 추론을 하고 이 추론에 관한 결론을 도출하는 것으로 매우 중요한 작업임은 물론 어려운 일이다. 이같은 목적을 달성하기 위해서는 주요 통계량에 대한 설명과 유의수준에 대한 언급이 필요하다. 또한 분석결과의 해석에 보조역할을 할 수 있는 도표의 사용여부도 검토한다.

4. 자료수집과 분석의 타당성 평가

광주, 전남지역의 단체 및 금융기관, 산업체에서 통계조사를 실시할 경우 자료수집과정과 자료분석시 활용하는 통계적기법의 타당성을 평가하기 위하여 통계조사 실시후 그결과를 통계로 집계하였다는 33곳에 대하여 결과물을 요청하였다. 그러나 대부분이 대외비라는 이유를 내세워 결과표의 공개를 거부하였으며 단지 14곳(9개 단체, 2개 금융기관, 1개 산업체, 2개 신문사)에서만 결과물을 제공하였다. 이들중에는 일년동안 조사를 여러번 실시한 곳도 있어서 수집된 결과물은 30편(단체:9, 금융기관:4, 산업체:2, 언론기관:15)으로 이들 발표물을 대상으로 제시된 점점표를 기초로 하여 자료수집 및 분석의 타당성 평가를 실시한다.

4.1. 자료수집

(1) 연구문제의 형성

신문사에서 자체조사를 실시하여 발표한 결과물 15편중에서 11편은 조사연구의 문제를 충분히 기술하지 않고 기사제목으로 대신하고 있다. 예를 들면 어느 일간지는 “교고 공동지원체제 지지 광주 반대 전남은 찬성”이라는 제목하에 30명의 교육위원을 대상으로 설문조사한 결과를 발표하였다. 반면에 결과발표와 더불어 따로 조사연구의 목적, 방법, 대상등을 상세히 설명한 일간지도 있었다(표3참조).

(2) 연구가설

수집된 결과물의 대부분이 실태파악을 위한 것으로 연구가설은 전혀 고려되지 않았다. 이는 연구대상 기관들이 통계조사는 실시할지라도 그 분석결과를 단순히 문항에 대한 응답분포를 설명하는 기술통계에 그치고, 점정 및 의사결정이나 예측을 위한 추측통계는 전혀 고려하지 못하고 있음을 시사해 준다.

(3) 연구의 대상

가. 목표모집단과 조사모집단: 산업체에서는 전수조사를 실시하였기 때문에 목표모집단과 조사모집단이 일치하였다. 그러나 신문사의 조사 결과물 대부분은 광주·전남지역 주민을 대상으로 표본을 뽑았다고 조사모집단은 명시하였지만 목표모집단에 대하여는 기술하지 않고 있다.

나. 모집단의 규모, 구조, 특성: 표본조사일 경우에는 표본을 뽑기 위한 표본틀(frame)이 준비되어야 하는데 frame의 기본 구성이 되는 모집단의 규모, 구조 및 특성을 설명할 필요가 있

다. 수집된 결과물중 오직 한편만이 지역별, 직업별, 성별 균형을 고려하여 광주전남지역의 직장인의 경제의식 및 실태조사를 실시하였다고 언급하였으나 이들 분포에 대한 구체적인 제시는 없었다. 전수조사인 경우에는 이 문제점이 적용되지 않으며, 나머지 조사결과물에서는 이 점에 대하여 전혀 언급되지 않았다.

(4) 연구설계

- 가. 표본추출방법: 모 일간지의 경우 “...난수표를 이용해 체계적으로 선정한 후, “Quota에 따라 응답자를 선택하는 다단계 무작위추출법(multi-staged random sampling method)을 이용 선정했다”와 같이 표본추출방법이 제대로 설명되었다. 이와 같이 표본추출방법이 명기된 결과물은 8편이었다. 반면에 어느 결과물은 “광주와 전남의 6개 대학의 정치학, 법학, 행정학등 사회과학관련 교수 26명을 대상으로 설문조사 하였다”라고만 기술하여 표본추출방법이 무시된 경우는 16편이나 되었다.
- 나. 대표성: “광주전남지역을 5대 권역별로 구분하여 지역별로 비례하여 표본을 배정하였다”와 같이 대표성을 고려하여 언급한 결과물은 한편뿐이고 나머지 결과물은 이에 대한 언급이 전혀 되지 않았다.

- 다. 표본의 크기: 전수조사를 실시한 6곳을 제외한 표본조사 결과물에 있어서는 조사의 신뢰도 파악을 위하여 표본의 크기 및 이의 결정을 위한 허용 오차한계 및 신뢰수준에 대한 언급이 필요하다. 그러나 단 한편만이 “...조사항목의 중요도와 조사비용을 감안하여 목표 정도 6%하의 표본수를 결정하였다”라고 언급하면서 결정된 표본수에 대하여 95% 신뢰수준하에서의 허용오차표를 제시하였을 뿐 나머지 결과물은 표본의 크기만 제시하였다.
- 라. 자료수집법: 주요 자료수집법에는 개인면접조사법, 배표조사법, 집합조사법, 전화조사법과 우편조사법이 있으며, 각 방법에는 서로 장단점이 있다(김경동). 따라서 결과물에는 자료수집법이 명시되는것이 바람직하다. 수집된 결과물에서 언급된 자료수집법에는 배표조사법(15), 개인면접조사법(4)과 전화조사법(5), 그리고 우편조사법과 혼용된 경우(2)도 있었으며, 이에 언급되지 않은 경우도 4편이나 되었다. 전화조사법이 활용된 5편의 결과물은 모두 신문사에서 활용한 방법인데 이는 응답자가 어떤 시점에서 순간적으로 무엇을 하며 무슨 생각을 하는지를 적은 비용으로 신속하게 알아낼 수 있다는 이점 때문에 많이 활용되었을 것이다. 그러나 이 방법은 Massey(1988)가 지적했듯이 전화가 없는 계층은 표본추출에서 아예 제외되어 표본의 편의가 있으므로 모집단의 전화보급율도 같이 명시하여 주는것이 좋겠다. 한편의 결과물만 전화보급율이 95.3%라고 명시하였는데 이것 또한 조사의 목표모집단은 광주전남지역인 반면에 전국의 전화보급율을 제시하고 있다.

5.2. 자료분석과 해석

(1) 분석기법의 선택

- 가. 자료의 성격: 설문에 응답된 내용을 수량화하는데 사용되는 척도의 종류에는 4가지로 명명형, 순위형, 구간형 그리고 비율형이 있다. 두편의 결과물만이 수집된 자료에 대한 척도의 종류에 대하여 언급하였고 나머지 결과물은 이에 대하여 전혀 언급하지 않았다.
- 나. 적용절차: 30편의 결과물중 6편이 두 변수간의 관계를 보여주는 분할표를 작성 제시하였고 나머지 24편 모두가 명명형인 문항마다의 분포상황을 보여주는 비율 또는 백분비만을 제시하였다.
- 다. 사용된 통계패키지: 자료의 분석을 위하여 사용된 통계패키지를 밝힌 결과물은 두편뿐이고 나머지 결과물들은 이에 대하여 전혀 언급하지 않았다.

<표3> 자료수집의 타당성 평가

점 검 항 목		PC	PI	NG	계	NA
연구 문제의 형성	결과물 편수 (%)	19 (83.3)	11 (36.7)	·	30 (100)	·
연구 가 설	결과물 편수 (%)	·	·	30 (100)	30 (100)	·
목 표 모 집 단 과 조 사 모 집 단	결과물 편수 (%)	10 (33.3)	20 (88.7)	·	30 (100)	·
모 집 단 의 규 모, 구 조, 특 성	결과물 편수 (%)	·	1 (4.2)	23 (100)	24 (100)	6
표 본 추 출 방 법	결과물 편수 (%)	8 (33.3)	·	16 (66.7)	24 (100)	6
대 표 성	결과물 편수 (%)	1 (4.2)	·	23 (95.8)	24 (100)	6
표 본 의 크 기	결과물 편수 (%)	1 (4.2)	23 (95.8)	·	24 (100)	6
자 료 수 집 법	결과물 편수 (%)	11 (36.7)	15 (50.0)	4 (13.3)	30 (100)	·

(2) 결과의 해석

- 가. 주요통계량: 자료의 성격에 따라 적절한 분석기법이 적용되었다 할지라도 통계량은 명확하게 기술되어야 할 것이다. 모든 결과물은 각 문항별로 분포를 설명하는 비율과 백분비를 주로 사용하면서 분포의 집중경향에 대한 언급은 하였으나 분산 및 분포의 형태에 대한 언급은 전혀 없었다. 두 변수간의 관계파악을 위하여 분할표를 제시한 6편의 결과물에는 단지 각 셀의 빈도수와 행백분비 또는 열백분비만 설명할 뿐 독립성 검정을 위한 카이제곱통계량은 전혀 제시되지 않았다.
- 나. 유의수준: 두 변수간의 독립성 관계를 규명하고자 분할표작성을 한 경우에는 카이제곱 통계량과 함께 귀무가설을 기각하는 결정을 내릴때 범할 수 있는 오류의 정도를 나타내는 유의수준(p값)도 함께 제시되어야 하는데 모든 결과물에서 무시되었다.
- 다.도표법이용: 그래프는 수집된 자료의 분석결과를 요약하여 시각적으로 보여주는데, 일반적으로 시각적인 인상은 결과가 숫자의 형태로 있을때보다 훨씬 더 강렬한 인상을 준다. 16편의 결과물이 막대그래프 또는 원주그래프를 이용하여 결과를 설명하고 있다.

5. 요약 및 결론

최근들어 각종 단체 및 언론기관, 금융기관, 산업체에서는 사회문제에 대한 실태파악 또는 사원들의 효율적인 관리를 위하여 통계조사를 통한 자료를 수집하여 분석하는 경우가 부쩍 늘었다. 본 연구에서는 이들에 대하여 통계조사의 활용실태를 파악하고, 자료수집과정과 분석과정에서 고려되어야 할 항목을 나열한 점검표를 개발하고, 이 점검표를 기초로 실제 조사결과 발표물에 대한 타당성평가를 실시하였다.

본 연구에서 제안된 점검표를 기초로 광주·전남지역에서 지난 일년간 통계조사를 실시한 단체, 언론·금융기관 및 산업체에서 발표한 결과물 30편에 대하여 자료수집과정과 분석에 대한 타당성평가 결과를 요약하면 다음과 같다.

자료수집과정에 있어서 고려되어야할 항목중 특히 문제시 되는것은 첫째로 검정할 연구가설이 제시되지 않은 점이다. 이는 대부분의 결과물이 어떤 문제에 대한 실태조사로 단순히 각 문

〈표4〉 자료분석의 타당성 평가

검 검 항 목		PC	PI	NG	계	NA
자료의 성격	결과물 편수 (%)	2 (6.7)	· ·	28 (93.3)	30 (100)	· ·
적 용 절 차	결과물 편수 (%)	30 (100)	· ·	· ·	30 (100)	· ·
사용된 통계패키지	결과물 편수 (%)	2 (6.7)	· ·	28 (93.3)	30 (100)	· ·
주요 통계량	결과물 편수 (%)	· ·	30 (100)	· ·	30 (100)	· ·
유 위 수 준	결과물 편수 (%)	· ·	· ·	6 (100)	6 (100)	24
도 표 법 의 이 용	결과물 편수 (%)	16 (53.3)	· ·	14 (46.7)	30 (100)	· ·

항에 대한 응답분포를 보여주는 기술통계에 그치고 있어 많은 경비와 노력이 투입된 조사내용이 현대통계학의 주를 이루는 추측통계에는 전혀 활용되지 못하고 있다는 점이다. 둘째로, 우리가 무엇을 알려고 하느냐에 따라 정의되는 모집단은 그 규모, 구조와 특성이 명확하게 정의되어야 한다. 왜냐하면 이것은 표본추출에 필요한 정보를 제공하여주기 때문이다. 그런데 대부분의 결과물에 있어서 이 점이 무시되었다. 셋째, 신뢰할 수 있는 표본조사를 하기 위해서는 모집단을 대표할 수 있는 표본추출이 되어야 하며, 표본의 크기 결정을 위한 오차한계 및 신뢰수준이 명시되어야 하나 많은 결과물들이 대표성, 표본크기의 결정기준 및 표본추출방법에 대한 언급이 전혀 없었다.

자료분석과 해석과정에서 문제시되는 것은 첫째, 자료의 성격에 대한 구분은 분석시 적용될 수 있는 분석기법의 결정에 중요한 역할을 하는데 많은 결과물들이 자료의 성격에 대한 언급은 하지 않고 단순히 각 문항에 대한 백분비 또는 비율을 제시하는 것으로 그치고 있다. 둘째, 분석결과에 대한 해석을 도와주는 주요통계량의 제시가 부족했으며 유의수준은 전혀 언급되지 않았다.

각종 단체, 금융언론기관 및 산업체에서 발표한 통계조사 결과물에 대한 자료수집과 분석의 타당성 평가 결과 가장 큰 문제점은 단순한 기술통계의 활용으로 그치고 있다는 것이다. 즉, 의사결정을 위한 추측통계부분은 전혀 다루어지지 못하고 있다. 특히 상관관계 또는 회귀분석과 같은 비교적 초보적인 분석기법이 적용된 결과물도 전혀 없을 뿐만 아니라 다변량해석기법은 아예 적용되지 못하고 있는 점이다. 이는 조사를 기획·실시 및 분석하는 담당자들의 추측통계에 대한 지식 부족 또는 컴퓨터를 이용한 자료분석기법에 대한 미숙에서 초래되는 것으로 사료된다. 따라서 통계조사의 활용이 점증하는 현실을 감안할 때 통계전문가들이 이들 단체·기관·산업체의 통계조사담당자에 대한 자료수집과 분석에 대한 정기적인 교육이 필요하다 하겠다. 이같은 교육은 한국 통계학회의 각 지회에서 해당지역을 담당할 수 있으리라 생각된다. 이렇게 함으로써 통계조사 담당자들에게 통계학의 중요성에 대하여 재인식시킬 수 있고, 통계학을 공부한 전문인이 사회의 어느 기관에서나 꼭 요구된다는 점을 부각시킬 수 있을 것이다.

참고문헌

- [1] 국토개발연구원(1986), 지역통계정비방안연구.
- [2] 김경동, 이은죽(1988), 「사회조사 연구방법」, 박영사.
- [3] 김병천(1987), "개인용 컴퓨터에서의 통계패키지 선택과 활용", 응용통계연구, 제1권 제1호, 75-90.
- [4] 김종호(1992), 표본조사법, 자유아카데미.
- [5] 박홍래(1989), 통계조사론, 영지문화사.
- [6] 윤기중, 안윤기, 김병수(1987), "통계의 오용과 효율적 이용에 관한 연구", 「산업과 경영」 제24권 제2호, 3-37.
- [7] 이화영, 이희춘(1991), "광주전남지역의 DATA BASE 구축방안", 학술연구보고서, 광주은행.
- [8] 최중후, 이재창(1990), 학술논문과 통계적 기법, 자유아카데미.
- [9] Blalock, H. M. Jr. (1972), *Social Statistics*, 2nd Ed. McGraw-Hill, New York.
- [10] Francis, I. (1973), "A Comparison of Several Analysis of Variance Programs," *Journal of American Statistical Association*, 68, 860-865.
- [11] Loether, H. J. and McTavish, D. G. (1980), *Descriptive and Inferential Statistics: An Introduction*, 2nd Ed. Allyn and Bacon, Boston.
- [12] MacCallum, R. (1983), "A Comparison of Factor Analysis Program in SPSS, BMDP & SAS," *Psychometrika*, 48, 223-231.
- [13] Massey, J. T. (1988), "An Overview of Telephone Coverage", Telephone Survey Methodology, John Wiley & Sons, New York.

A Study on the Validity of the Statistical Data Collection and Analysis in Gwangju and Chonnam¹⁾

Hwa-Young Lee²⁾

Abstract

A check list which includes the items that are to be considered in the process of the statistical data collection and analysis by non-scientific organizations is proposed. Based on the suggested check list, the output resulting from the statistical survey conducted by private organizations, banks, organs of expression and enterprises in Gwangju and Chonnam are examined about the validity of data collection and statistical analysis.

1) This paper was supported in part by NON DIRECTED RESEARCH FUND, Korea Research Foundation, 1991.

2) Department of Statistics, Dongshin Univeraity, 252 Daeho-dong, Naju, Chonnam 520-714,