

기획특집

《조사방법론》에 게재된 논문들의 성격과 경향

A Study on the Characteristics of Articles in *Survey Methodology*

김 규 성 · 이 기 재*
Kyuseong Kim · Kee Jae Lee

이 글에서는 캐나다 통계청에서 발간하고 있는 《조사방법론》(*Survey Methodology*)에 게재된 논문들의 성격과 경향을 개괄적으로 살펴보았다. 논문에서 다루고 있는 내용을 주제별·적용 분야별로 나누어 그 특성을 살펴보고, 연도별로 구분하여 시대적 흐름에 따른 변화를 살펴보았다. 또한 분석에 이용된 데이터의 종류에 따라 논문들을 분류하여 데이터가 논문에 어떻게 이용되는지 알아보았다.

We broadly review the whole articles in *Survey Methodology* published by Statistics Canada. By the classification of subject and application area as well as years, we find some features of the articles. In addition, we classify them according to the sort of data analyzed in the article and investigate how does it use data.

I. 머리말

한국조사연구학회에서 발간하는 학술지인 《조사연구》(*Survey Research*)는 ‘각종 조사와 통계 기법을 활용하는 학문의 발전에 기여하고, 동시에 실제 조사의 질과 신뢰성을 높이고자 조사를 바탕으로 하는 경험적 연구 혹은 이론적 연구논문’의 게재를 일차적인 목표로 하고 있다. 《조사연구》는 이러한 목표로 개별 학문분야에 관계없이 조사와 통계 기법을 활용한

* 김규성(서울시립대학교 컴퓨터·통계학과 부교수, kskim@uos.ac.kr), 이기재(한국방송통신대학교 정보통계학과 부교수, kjlee@mail.knou.ac.kr)

이론적 혹은 경험적 연구논문들 게재한다. 이는 사회학, 경제학, 경영학, 심리학 등 개별 학문 분야에서 이루어진 실제조사에서 자료의 수집 또는 분석과 관련된 문제들을 다룬 논문이거나 혹은 조사 기법에 관련된 논문이라면 《조사연구》에 게재될 수 있음을 시사한다. 이런 《조사연구》의 취지를 염두에 두고 이 글은 《조사연구》에 적합한 통계학 분야의 논문이 가질 성격을 가늠하기 위해 《조사방법론》(*Survey Methodology*)에 실린 논문들을 살펴보았다. 《조사방법론》은 캐나다 통계청에서 발간하는 조사통계 분야의 가장 대표적인 학술지 가운데 하나이다. 이 학술지에 실린 논문들은 조사와 관련하여 통계적 기법 개발의 다양한 측면을 다루는 논문들이다. 《조사방법론》은 조사의 응용에 초점을 맞추고 있는 AAPOR (American Association for Public Opinion)에서 발간하는 《계간 여론조사》(*Public Opinion Quarterly*)에 비해 상당부분 이론적인 내용들을 다루고 있지만, 수리적인 통계이론 보다는 실제 조사 데이터를 이용한 구체적인 방법론에 대한 기법 개발과 평가를 강조하고 있다.

따라서 《조사연구》와 유사한 학술지 발간 목적과 편집 방향을 가진 《조사방법론》에 실린 논문들을 개괄적으로 살펴보는 것은 향후 《조사연구》의 편집 방향 설정에 작으나마 보탬이 되리라고 생각한다.

《조사방법론》은 캐나다 통계청에서 발간하는 학술지로 1975년 창간된 이래 매 6월과 12월에 출판되어 현재 2001년 6월호까지 발간되었다. 권수로는 27권 1호 까지 나왔고 총 478편의 논문이 발표되었으니 한 해 평균 18편의 논문이 6월호와 12월호에 나누어 실린 셈이다. 이 글에서는 1975년 창간호(1권 1호)부터 2001년 6월호(27권 1호)까지 발표된 논문 478편 전체를 분석의 대상으로 하였다. 먼저 다루고 있는 내용에 따라 전체 논문을 주제별로 분류하여 특성을 살펴보았고, 다음에 적용 분야별로 분류하여 적용 분야의 특성을 알아보았다. 또한 연도별로 주제 및 적용 분야의 변화를 살펴보았으며 논문에 사용된 데이터의 이용 방법에 따른 특성을 고찰하였다. 그리고 마지막으로 간단한 요약과 토의를 덧붙였다.

II. 《조사방법론》에 실린 논문의 특성

1. 주제별 특성

《조사방법론》은 편집 방향에서 밝힌 바와 같이 통계적 기법 개발과 관련한 다양한 분야의 논문을 게재하고 있다. 구체적인 분야를 살펴보면, 조사 데이터의 출처 및 수집, 조사 오차, 조사 평가, 조사 방법론 연구, 시계열 분석, 인구통계학 연구, 데이터 통합, 추정 및 분석 기법, 일반적인 조사 시스템 개발 등이다. 그리고 앞서 언급한 바와 같이 데이터 수집 혹은 데이터를 적용한 구체적인 방법론에 대한 개발 및 평가를 중요시한다.

〈표 1〉은 《조사방법론》에 실린 478편의 논문을 내용에 따라 주제별로 분류한 결과이다. 세부 주제별로 보면 조사 데이터를 기반으로 한 추정(estimation)방법 및 이론에 관한 논문이 154편으로 전체의 32.2%를 차지하여 가장 많으며, 분석적인 기법(analytic methods)에 관한 논문이 39편으로 전체의 8.2%를 차지한다. 두 분야를 합하면 193편으로 전체의 40.4%에 달한다. 그리고 표본설계(sample design)에 관련된 논문은 69편으로 14.4%에 해당한다.

전통적인 조사통계 이론은 모집단에서 확률표본을 추출하여 표본 데이터에 근거한 모수 추정방법을 탐구하며, 이 과정에서 표집오차(sampling error)를 줄이는데 이론적인 관심의 초점이 맞추어져 있다. 표집오차를 관리하는 단계로는 표본설계 단계와 추정 단계로 나누어 볼 수 있는데, 《조사방법론》에 실린 54.8%의 논문이 표본설계 및 추정에 관련됨으로써 이 두 가지 주제가 여전히 중요하게 다루어지고 있음을 알 수 있다.

표집오차가 아닌 비표집오차(nonsampling error)는 조사의 전 과정에서 발생할 수 있으며, 조사의 품질을 높이기 위해서는 표집오차는 물론이고 비표집오차를 전반적으로 줄일 수 있어야 한다. 비표집오차를 다룬 논문은 55편으로 전체의 11.5%이다. 표집오차에 비하여 비표집오차는 수치로 계량화하기 힘들고 또한 모형화하기 힘들기 때문에 논문을 쓰기가 어려운 점을 감안하면, 11.5%에 달하는 비표집오차 관련 논문의 수는 이 분

야의 중요성 및 높은 관심을 보여 주고 있다. 다른 주제로는 데이터 수집(data collection) 20편(4.2%), 입력 및 대체(edit and imputation) 18편(3.8%), 데이터 품질(quality assurance) 17편(3.6%), 조사기법 개발(survey development) 31편(6.5%), 시계열 분석 18편(3.8%) 등이다.

〈표 1〉 주제 및 적용 분야별 분류

주제 \ 적용분야	일반 분야	사회경제 통계	센서스	경제 통계	농업 통계	인구 통계	합 계
추정이론	72	43	17	11	5	6	154 (32.2%)
분석적 기법	26	11	1	1	0	0	39 (8.2%)
표본설계	29	15	5	12	7	1	69 (14.4%)
비표집오차	17	19	15	1	2	1	55 (11.5%)
데이터 수집	9	8	0	3	0	0	20 (4.2%)
입력 및 대체	10	1	2	3	2	0	18 (3.8%)
데이터 품질	9	3	2	2	0	1	17 (3.6%)
조사기법 개발	2	19	0	6	4	0	31 (6.5%)
시계열 분석	10	6	0	2	0	0	18 (3.8%)
기타	42	3	3	2	0	7	57 (11.9%)
합 계	226 (47.3%)	128 (26.8%)	45 (9.4%)	43 (9.0%)	20 (4.2%)	16 (3.3%)	478 (100%)

적용분야에 따른 논문의 분류를 살펴보면 일반분야가 226편 (47.3%)으로 전체의 절반을 차지한다. 이 수치는 구체적인 조사에 대한 방법론을 개발하였다 하더라도 이를 일반화한 논문의 수와 처음부터 일반적인 방법론을 서술한 논문의 수를 나타내는 것이다. 따라서 이론을 일반화하려는 연구자들이 많았음을 알 수 있다.

개별분야에서는 사회경제통계(socio-economic statistics)에 관련된 논문이 26.8%로 가장 많고 센서스 관련이 9.4%, 경제통계(economic statistics) 관련이 9.0% 순이다. 이러한 수치는 《조사방법론》의 발행기관이 국가통계작성기관이라는 사실과 밀접한 관계가 있다. 국가의 중요 통계로는 여론조사나 일반 사회조사 보다는 노동력 인구 조사, 가계 조사 등 사회경제적인 지표를 생산하는 대규모 조사가 많기 때문에 《조사방법론》에는 이에 관련된 논문이 많이 게재되고 있다고 해석할 수 있다. 특히 센서스

관련 논문이 45편으로 9.4%를 차지하는 것은 국가통계로서 센서스가 중요한 위치를 차지하고 있음을 보여준다.

2. 연도별/주제별 특성

앞 소절에서 논문의 주제별 특성을 살펴본데 이어 더 세부적으로 연도별 분포를 살펴보았다. <표 2>에는 1990년대에 들어서서 추정이론에 관한 논문이 급격히 많아지는 현상이 나타난다. 1934년 Neyman의 기념비적인 논문이 발표된 이래 추정이론이 꾸준히 개발되어 온 사실을 감안하면 1990년대에 추정관련 논문이 급격히 많아지는 것은 다소 특이한 현상이다(Rao, 1997; Rao and Bellhouse, 1990). 이러한 현상은 《조사방법론》의 편집기획과 관련이 있다. 1990년에 들어 새로운 주제가 생기면 이에 대한 공동의 집중적인 연구를 바탕으로 《조사방법론》을 편집한 것으로 보인다. 그러한 결과 1992년 1호에는 ‘센서스 과소집계 측정방법 및 논점(census undercount measurement methods and issues)’에 관련된 논문이 집중적으로 실렸으며, 2호에는 ‘조사 데이터를 이용한 추론(inferences with survey data)’에 관한 논문이 많이 실렸다. 1994년 20권에는 ‘소지역 추정법(small area estimation)’, 1996년 22권에는 ‘가중치와 추정(weighting and estimation)’, 1998년 24권에는 ‘추적조사와 분석(longitudinal surveys and analysis)’이 주제로 선정되었다.

표본설계 및 비표집오차 관련 논문들은 꾸준히 계속 발표되고 있으며, 다른 분야의 논문들도 연도에 따라 큰 변화가 없다. 전반적으로 연도에 따라 큰 변화가 없는 것은 《조사방법론》의 발행기관이 국가기관이라는 사실과 관련이 있는 것으로 보인다. 국가통계의 속성상 계속되는 조사가 많고 지속적으로 조사의 품질을 관리해야 하기 때문에 이에 대한 지속적인 방법론 개선이 꾸준한 논문의 발표로 나타나는 것이다.

〈표 2〉 연도별/주제별 분류

연도	주제										합계
	A	B	C	D	E	F	G	H	I	J	
1975	5	2	3	3	1	0	1	1	0	3	19
1976	1	2	3	4	0	1	0	2	0	0	13
1977	5	0	0	3	0	0	0	4	0	1	13
1978	2	2	1	3	0	1	0	1	0	1	12
1979	1	0	3	2	1	1	1	0	0	2	11
1980	1	0	1	3	1	0	0	2	0	13	21
1981	1	2	3	1	0	1	1	0	0	1	10
1982	1	0	0	2	1	0	0	0	0	1	5
1983	2	0	2	0	0	0	1	5	0	1	11
1984	1	7	0	1	0	0	0	2	2	0	13
1985	6	0	3	1	1	1	0	0	3	5	20
1986	2	2	0	6	0	6	0	0	1	1	18
1987	8	7	4	1	2	0	0	0	1	3	26
1988	0	0	2	10	0	2	2	1	0	5	22
1989	5	0	2	0	2	0	2	0	1	4	16
1990	10	1	6	2	1	1	0	0	4	7	32
1991	4	1	3	2	5	0	0	2	1	0	18
1992	18	2	2	0	1	0	1	0	0	0	24
1993	2	2	6	2	0	0	4	0	2	0	18
1994	14	1	5	0	1	1	0	0	0	1	23
1995	12	2	3	0	1	0	3	1	0	0	22
1996	11	2	4	1	0	0	0	0	1	1	20
1997	9	0	4	2	0	0	0	2	0	1	18
1998	7	4	2	1	1	0	0	5	0	0	20
1999	13	0	4	1	0	0	1	0	0	3	22
2000	9	0	2	4	1	1	0	2	2	1	22
2001	4	0	1	0	0	2	0	1	0	1	9
합 계	154	39	69	55	20	18	17	31	18	57	478

주: A : 추정 이론, B : 분석적 기법, C : 표본설계,
 D : 비표집오차, E : 데이터 수집, F : 입력 및 대체,
 G : 데이터 품질, H : 조사기법 개발, I : 시계열 분석, J : 기타.

3 연도별/분야별 특성

〈표 3〉에서 연도별, 주제별 발표 논문편수를 살펴보면, 일반분야와 관

련한 논문과 사회경제통계에 관련한 논문은 연도에 관계없이 꾸준히 발표되고 있다. 센서스 관련 논문이 1988년과 1992년에 일시적으로 많아진 것은 1990년 센서스에서 발생한 과소집계(under-count)와 관련이 있다. 센서스 과소집계에 대한 문제 및 해결 방안 모색이 이루어졌으며 관련 논문이 집중적으로 실렸다.

〈표 3〉 연도별/분야별 분류

연도	분야	일반 분야	사회경제 통계	센서스	경제 통계	농업	인구 통계	합계
1975		12	7	0	0	0	0	19
1976		4	5	2	1	1	0	13
1977		5	5	0	2	1	0	13
1978		6	2	1	3	0	0	12
1979		6	2	1	1	1	0	11
1980		13	3	1	2	1	1	21
1981		7	2	0	1	0	0	10
1982		2	3	0	0	0	0	5
1983		2	5	0	1	2	1	11
1984		6	6	0	0	1	0	13
1985		7	4	1	0	1	7	20
1986		14	1	0	2	1	0	18
1987		15	8	1	1	1	0	26
1988		7	1	11	2	0	1	22
1989		6	2	1	4	1	2	16
1990		15	9	4	2	2	0	32
1991		8	3	0	5	1	1	18
1992		6	3	13	1	0	1	24
1993		12	2	1	2	1	0	18
1994		9	8	1	3	2	0	23
1995		10	8	3	0	0	1	22
1996		11	6	0	3	0	0	20
1997		8	8	0	2	0	0	18
1998		5	12	0	2	1	0	20
1999		14	4	1	2	1	0	22
2000		13	3	3	1	1	1	22
2001		3	6	0	0	0	0	9
합 계		226	128	45	43	20	16	478

III. 분석에 이용된 데이터의 특성

조사와 관련된 논문에서 조사 데이터는 논문 구성의 중요한 요소 중 하나이다. 그렇지만 《조사방법론》에 실린 논문들이 모두 조사 데이터를 이용한 것은 아니다. 최근 3년 간(1998년-2000년) 《조사방법론》에 실린 64편의 논문에 이용된 데이터의 종류를 분석한 결과, 실제 데이터를 이용한 논문은 19편(29.7%), 가공 데이터를 이용하여 모의실험을 실시한 논문은 27편(42.2%)로 나타났다. 데이터를 사용하지 않는 논문도 18편(28.1%)에 달한다(〈표 4〉).

〈표 4〉 분석에 이용된 데이터의 종류

연 도	실제 조사 데이터	모의 실험 데이터	기 타	합 계
1998	10	8	2	20
1999	3	10	9	22
2000	6	9	7	22
합계	19 (29.7%)	27 (42.2%)	18 (28.1%)	64 (100%)

데이터를 취급하지 않은 논문이 28.1%에 달하는 현상은 다음과 같이 해석할 수 있다. 조사에는 반드시 데이터가 수반되지만 조사관련 논문에서 반드시 데이터를 취급해야 하는 것은 아니다. 조사 정책, 조사들의 상호 비교, 조사의 특성들을 거시적으로 평가하는 논문에서 반드시 데이터를 이용해 논리를 전개할 필요는 없을 것이다. 이런 경우 개념적인 서술이나 논증이 더 적절할 수 있다. 그렇다 하더라도 조사관련 학술지에서 데이터를 사용하지 않은 논문의 비율 28.1%는 조금 높아 보인다.

데이터를 사용한 경우에도 가공의 데이터를 이용하여 모의실험을 한 논문의 비율(42.2%)이 실제 데이터를 이용한 논문의 비율(29.7%)보다 높게 나타난다. 이것은 모의실험 데이터는 실제 데이터가 아니기 때문에 현실성이 다소 떨어지는 단점이 있기는 하지만, 연구자가 설정한 모형에 따

라 장애 없이 데이터를 만들어 사용할 수 있는 장점이 있기 때문에 나타난 현상으로 보인다. 실제 데이터를 이용하는 경우에는 현실을 그대로 반영하는 데이터라는 점과 실제로 실시된 조사의 방법론 평가에 적절하다는 장점이 있는 반면, 연구목적에 적합한 데이터를 구하기 어려울 수 있고 응답자 비밀보장 등과 관련하여 원시 데이터에 접근이 불가능할 수 있으며 표본수의 제약이 뒤따르는 단점이 있다.

1998-2000년 사이에 《조사방법론》에 실린 논문들이 이용한 실제 조사들을 살펴보면 캐나다 노동력 조사(Labour Force Survey)가 5회로 가장 많으며, 미국 노동력 조사(Current Population Survey)가 2회, 센서스가 2회이다. 그 외에 가격지수조사(Price Index Survey), 국민건강조사(National Population Health Survey) 등이 이용되었다. 이용된 조사들은 대부분 전국을 대상으로 한 대규모 복합조사이며 통계청 등 국가 기관에서 작성하는 국가통계라는 공통점이 있다. 이러한 경향은 역시 《조사방법론》이 캐나다 통계청에서 발간하는 학술지라는 사실과 밀접한 관련이 있으며, 관심의 초점이 국가통계에 맞추어져 있다는 사실을 말해준다.

IV. 맺음말

본 논문에서는 《조사방법론》창간호부터 2001년 6월호까지를 대상으로 하여 《조사방법론》에 실린 논문들을 주제별/적용분야별/연도별 특성을 개괄적으로 살펴보았다. 그리고 분석에 이용된 데이터의 종류에 따라 논문을 분류하여 데이터가 논문에 어떻게 이용되는지를 살펴보았다. 그 결과 《조사방법론》에 실린 연구 논문들은 전국 대상의 대규모 복합조사와 관련한 다양한 통계적 기법을 연구·개발하고 있음을 알 수 있었다. 연구 주제 면에서는 모수 추정, 표본설계 등이 과반수 이상을 차지하며 비표집오차와 관련한 연구도 상당부분 이루어졌다. 또한 적용 분야로는 사회경제통계 관련 조사가 많았고 센서스 및 경제통계, 농업통계, 인구통계 등이 비중 있게 다루어졌다.

20세기 후반 통계학 분야에서 나타나는 한가지 현상은 Kalton(2000)의

지적처럼 조사 연구의 전문화를 들 수 있다. 1975년에 창간된 《조사방법론》은 이러한 전문화 현상을 반영하며 이를 통하여 조사 방법론에 대한 더욱 구체적이고 깊이 있는 연구가 이루어지는 계기가 마련되었다. 21세기 지식정보화사회에서 조사의 필요성은 날로 증가하며 조사의 종류 및 가지 수도 늘어나는 추세에 있다. 이렇듯 양적으로 팽창하는 조사 수요에 발맞추어 조사 방법의 질적 수준을 높여야 하며 좀 더 전문화되고 세련된 조사가 이루어져야 한다. 현재 우리나라의 경우 실제 조사에서 발생하는 다양한 문제들을 학술적으로 다룬 연구논문들을 모아서 게재할 적절한 학술지가 없는 상황이다. 우리 사회에서 이루어지는 많은 조사들이 통계학, 사회학, 경제학, 정치행정학, 심리학 등 다양한 학문분야와 복합적으로 연계되어 있는 점을 고려할 때, 조사 관련 분야의 연구를 활성화하고 동시에 다양한 학문간의 공동연구를 수행하여 우리나라 조사 연구 분야의 질적 수준 향상시킨다는 측면에서 《조사연구》가 학술토론의 무대로써 크게 기여할 수 있기를 기대한다.

참고문헌

- 한국조사연구학회. 2000. 《조사연구》 1(1).
- Kalton, G. 2000. "Developments in Survey Research in the Past 25 Years." *Survey Methodology* 26 : 3-10.
- Rao, J.N.K. 1997. "Developments in Sample Survey Theory." *The Canadian Journal of Statistics* 25 : 1-21.
- Rao, J.N.K. and Bellhouse, D.R. 1990. "History and Development of the Theoretical Foundations of Survey based Estimation and Analysis (with discussion)." *Survey Methodology* 16 : 3-29.