

국내외 문헌정보학분야 연구에서 추론통계 사용에 관한 연구*

A Study on the Use of Inferential Statistics in Library and Information Science Research

노 정 순(Jung-Soon Ro)**

목 차

- | | |
|--------------------|------------------|
| 1. 서론 | 4.2 추론통계사용 분석 |
| 2. 선행연구 | 4.3 추론통계기법 분석 |
| 3. 연구대상 및 방법 | 5. 추론통계 사용상의 문제점 |
| 4. 데이터 분석 | 6. 결론 및 제언 |
| 4.1 통계기법사용 분석 참고문헌 | |

초 록

본 연구는 국내외 문헌정보학분야의 대표적인 10개 학술지에서 2001부터 2004까지 발표된 1,768편의 연구논문을 대상으로 통계사용을 비교하였고, 그 중 추론통계를 사용한 345편의 논문에서 사용된 추론통계기법을 국내외, 학술지 주제영역별, 논문의 연구주제별, 연구방법별, 저자의 전공별로 분석하였다. 2001년 이후 기술통계와 추론통계의 사용은 증가하였으나, 외국에 비하여 국내에서의 사용률은 유의하게 낮은 차이를 보였다. 추론통계를 사용한 345편의 논문은 국내외, 학술지 주제영역별, 연구주제별, 연구방법별, 저자의 전공별 차이가 있었다. 사용된 추론통계기법은 국내외 간에, 학술지 주제영역 간에 차이가 있었다. 전체적으로 다변량분석이 가장 많이 사용되었다. 추론통계를 사용하는데는 가설 설정에서부터 가설을 검증하고, 검증통계량과 유의확률을 제시하고, 모형의 적합성을 해석하는데 보다 주의깊은 논의가 필요한 것으로 분석되었다.

ABSTRACT

This Study analyzed the use of statistics in 1,768 research articles published in 2001-2004 in 4 Korean & 6 English core journals in the field of library and information science. Korean journals made significantly less use of descriptive and inferential statistics. Of the 663 inferential statistics used in 345 of the 1768 articles, the most frequently used inferential technique was multivariate analysis. There was significant difference in inferential methods used in Korean & English journals, also in traditional library science journals and information science journals.

키워드: 추론통계, 문헌정보학, 정보학, 연구논문, 국내외 비교

Inferential Statistics, Statistic Technique, Library and Information Science

* 본 논문은 2005학년도 한남대학교 학술연구비 지원에 의하여 연구되었음.

** 한남대학교 문과대학 문헌정보학과 교수(jsr@mail.hannam.ac.kr)

논문접수일자 2006년 2월 9일

게재확정일자 2006년 3월 17일

1. 서론

학문의 연구목적은 주제분야의 주어진 현상을 과학적 연구를 통해 기술하고 설명하여 이에 대한 포괄타당한 지식체계를 정립함으로써 주어진 현상을 미리 예측하고 통제하기 위한 것이며, 과학적 연구의 목적은 주어진 현상을 객관적으로 검토하여 결론을 내리기 위한 것이다. 객관적인 연구를 위해서는 체계적인 자료 수집과 측정 및 분석이 필요하며, 수집된 자료로부터 얻은 결론을 모집단으로 일반화하기 위하여 통계적 연구방법을 필요로 한다. 통계적 연구방법은 실험 연구와 정량 연구에서 중요한 역할을 하며, 특히 추론통계는 표본집단에서 얻은 데이터를 기반으로 모집단에 대한 가설을 검증한다거나 모형을 추정하고 예측하는데 필수적이다. 정량 연구에서 통계적 기법 특히 추론통계기법은 과학적 연구의 지향도를 나타내는 유용한 지표가 될 수 있다고도 하였다(Wyllys 1978).

그러나 선행연구에 의하면 2000년 이전 문헌정보학에서 추론통계기법의 사용은 매우 미약한 것으로 보고되었다(2.9%(Wyllys 1978) ~ 13.4%(이명희 2003)). 이러한 사용률은 다른 사회과학 분야와 비교할 때 매우 저조한 것이었다(Wallace 1985; 사공철, 최정희 1993). 그 원인은 다수의 미국 문헌정보학과에서 석사 졸업논문을 요구하지 않고(Childers 1984), 다른 사회과학분야(사회복지, 교육, 경영)에 비해 문헌정보학이 교과과정에서 연구방법과 정량분석이 덜 강조하고 있는데서 찾을 수 있다고 하였다(Wallace 1985). 따라서 가설검증을 위한 고급정량분석 연구방법은 문헌정보학에서 중요한 역할을 하지 못하고, 문헌정보학에

서 연구는 예측과 통제 및 연구결과의 일반화를 기할 수 있는 추론통계보다는 변수들 간의 관계현상을 기술하는 기술통계에 더 의지하고 있다고 하였다.

문헌정보학이 우리보다 먼저 소개된 미국에서는 이미 1967-71년 동안 4종의 학술지에 27편의 추론통계를 사용한 논문이 발표되었고(Atherton 1973), 1975년 한 해에는 34편의 추론통계를 사용한 논문이 발표되었다(Wyllys 1978). 그러나 미국에 비해 문헌정보학의 역사와 연구방법론이나 추론통계에 관한 교육의 역사가 짧은 우리나라에서 추론통계를 사용한 논문이 학술지에 처음 발표된 것은 1986년이었다. 1990년까지 총 14편의 논문에서 추론통계가 사용되었다(사공철, 최정희 1993).

미국과 한국에서 문헌정보학과 연구방법론이 연구된 역사와 추론통계가 소개된 역사는 다르지만, 2003년 현재 문헌정보학 대학원 교과목 연구방법론에서 추론통계를 가르치는 학교는 한국은 59%, 미국은 60%로 비슷하고, 통계패키지를 가르치는 것도 23%와 25%로 매우 비슷한 것으로 보고되었다(Park 2004). 이처럼 미국과 한국의 문헌정보학자가 연구방법에서 추론통계를 가르치는 비중은 비슷한데 실제 연구에서 통계 특히 추론통계의 사용도 비슷한가? 2001년 이후 논문에서 추론통계의 사용에는 어떤 변화가 있는가? 본 연구는 이러한 연구 질문을 규명하기 위해 2001년 이후 우리나라와 외국의 주요 문헌정보학 학술지에 발표된 연구논문의 내용을 분석하였다. 보다 구체적인 연구 질문들은 다음과 같다.

1. 연구논문에서 서술통계와 추론통계는 어느 정도 사용되고 있는가?

2. 추론통계를 사용한 논문은 주제분류별, 연구방법별, 저자전공별 차이가 있는가?
3. 연구논문의 하부주제 영역에 따라, 사용된 연구방법에 따라, 저자의 전공분야에 따라 사용된 추론통계기법의 종류에는 차이가 있는가?
4. 위의 현상들은 외국의 연구와 우리나라 연구 간에 차이가 있는가?
5. 위의 현상들은 학술지의 주제 영역(문헌정보학 영역과 정보학 영역)상 차이가 있는가?
6. 추론통계기법을 사용하여 데이터를 통계 분석하고 기술하는 과정에서 문제점은 없는가?

2. 선행연구

문헌정보학 분야의 연구논문을 대상으로 통계기법의 사용에 관한 연구는 1990년 이전 발표된 논문을 대상으로 소수의 연구가 이루어졌으며, 연구결과 문헌정보학에서 추론통계의 사용은 매우 부족하다고 보고되었다.

Atherton(1973)에 의하면 1969-71년 동안 정보학분야 학술지 4종에 발표된 211개 논문들 가운데 27편(12.8%)만 추론통계를 사용하였고, 학술지 1종을 더 추가하여 정보학분야 5종의 학술지 1974년호에 발표된 75편의 연구논문을 대상으로 Atherton의 연구를 되풀이 한 Water(1976)의 연구에서는 10편(13.4%)의 논문에서 추론통계가 사용되었다. Wyllys(1978)에 의하면 Texas대학교도서관에서 구독하는 문헌정보학 학술지 36종의 1975년호에 발표된 1,157

편의 논문을 분석한 결과, 522편(45.1%)이 통계를 사용하였으나, 그 중 34편(통계를 사용한 522편의 논문의 6.5%, 전체 논문 중 2.9%)만이 추론통계를 사용한 것으로 보고되었다.

Wallace(1985)는 교육학, 사회복지학, 경영학, 문헌정보학에서 각각 25종의 학술지를 선정하여 1981년에 발표된 연구논문을 대상으로 통계사용을 분석하였다. 4주제 분야 중 기술통계의 사용은 문헌정보학이 20%(796편 중 158편) 가장 높았으나(경영학은 17%, 사회학은 12%, 교육학 9%), 추론통계의 사용은 문헌정보학이 6%(796편 중 48편)로 가장 낮았다(교육학은 31%, 경영학 26%, 사회복지학 18%).

Enger(1989)는 문헌정보학 문헌의 저자의 소속기관과 통계사용의 관계를 조사하였다. 1985년 Journal Citation Reports에서 문헌정보학분야 학술지 중 영향지수(IF)가 가장 높은 25종의 학술지를 선정하여, 학술지에 발표된 915편의 논문을 분석한 결과, 기술통계는 189편(20.7%)에서, 추론통계는 102편(11.1%)에서 사용되었다. 915편의 논문 중 대학도서관 사서가 가장 많은 239편(26.1%), 2위인 문헌정보학과 교수는 161편을, 3위인 타학과 교수는 125편의 논문을 발표했으나, 추론통계의 사용률은 타학과 교수가 가장 높았고(34.4%, 125편 중 43편), 문헌정보학과 교수는 15.5%(161편 중 25편), 대학도서관 사서는 5.9%(239편 중 14편)였다. 추론통계는 주로 교수들에 의해 사용된다고 보고되었다.

국내에서는 사공철, 최정희(1993)가 문헌정보학에서 통계사용을 사회복지학 및 교육학과 비교하여 분석하였다. 1980~1991년에 3개 분야에서 각각 3종의 학술지에 실린 논문을 분석한 결과,

Wallace의 연구에서와 같이 기술통계는 문헌정보학이 24%로(274편 중 65편) 가장 높은 사용률을 보였으나(사회복지학은 285편 중 18%, 교육학은 308편 중 12%), 추론통계는 문헌정보학이 8%(274편 중 21편)로 가장 낮은 사용률을 보였다(교육학은 29%, 사회복지학 26%).

이명희(2003)는 한국 문헌정보학분야의 대표적인 3개의 학술지 1997-2001년호를 대상으로 연구논문의 내용을 분석한 연구에서 통계사용을 보고하였다. 기술통계는 87편(14.57%)의 논문에서, 추론통계는 80편(13.40%)의 논문에서 사용되었다. 학회지별 추론통계사용은 문헌정보학회지가 18.86%로 가장 높았고, 정보관리학회지는 11.17%, 한국도서관정보학회지가 9.76%로 가장 낮았다.

3. 연구대상 및 방법

본 연구는 문헌정보학 관련 국내 학회지 4종과 외국의 학술지 6종을 대상으로 2001년부터 2004년호에 발표된 연구논문을 대상으로 수행되었다.

〈표 1〉에서 보는 바와 같이 국내 학회지로는 학술진흥재단 등재 학술지 4종 모두를 포함시

켰으며, 외국 학술지는 1997년부터 2000년까지 4년 평균 문헌정보학분야에서 Journal Citation Reports(JCR)의 Journal Impact Factor(영향지수)가 가장 높은 상위 학술지(Schloegl & Stock 2004) 중 정보학 영역과 문헌정보학 영역으로 나누어 각각 IF가 가장 높은 3종을 선정하였다. 상위 7위까지의 학술지 중 리뷰지인 ARIST와 특정 주제에 치우친 Scientometrics을 제외하고 남은 5종은 정보학 영역이 3종, 문헌정보학 전반에 관한 학술지가 2종이었다. 문헌정보학 영역에서 나머지 1종은 4위인 Library and Information Science Research가 리뷰나 경향보다는 과학적 연구방법을 사용한 연구논문을 다루며 문헌정보학 영역에서는 반감기가 가장 높은 학술지이기 때문에 3위인 Library Trends 대신 선정되었다(표 2 참조).

연구대상 논문은 선정된 학술지에 2001년부터 2004년까지 발표된 논문 중 목차에서 저자가 1명 이상 명시되고, 편집인에 의해 '논문'(article)이라고 명시된 것을 대상으로 하였으며, 정규 컬럼이나 학회소식, Editorial, Letters to the editor, Brief communications, An Introduction to the special issue, Book review 등은 제외되었다. 그 결과 총 1,768개의 논문(국내 792, 국외 976)이 분석대상으로 선정되었다(표 3 참조).

〈표 1〉 분석 대상 학술지

	정보학	문헌정보학
외국	Journal of Documentation(JOD) J. of the American Society for Information Science and Technology(JASIST) Information Processing & Management(IPM)	Library Quarterly(LQ) College & Research Libraries(CRL) Library and Information Science Research(LISR)
국내	정보관리학회지	한국문헌정보학회지 한국도서관·정보학회지 한국비블리아학회지

〈표 2〉 1997-2000 평균 학술지영향지수(상위 11학술지)

잡지명	영향지수	반감기
1등 Journal of Documentation	1.52	5.10
2등 JASIST	1.29	7.75
3등 ARIST	1.23	5.15
4등 Library Quarterly	1.05	7.68
5등 College & Research Libraries	0.94	5.93
6등 Scientometrics	0.75	7.53
7등 Information Processing & Management	0.61	6.65
8등 Journal of Information Science	0.55	5.58
9등 Library Trends	0.52	5.68
10등 International J. of Information Management	0.45	5.83
11등 Library & Information Science Research	0.44	6.68

선정된 논문은 연구주제, 저자의 소속기관, 발표된 학술지명, 사용된 연구방법, 가설설정 여부(귀무가설, 연구가설), 검정방법(단측, 양측), 사용된 통계기법, 제공된 통계결과, 사용된 통계 S/W, 추론통계 사용 시 문제점 등의 연구변수로 분석되었다.

논문의 주제분류구분을 위해서는 문헌정보학 분야의 논문을 내용 분석한 기존의 국내외 선행연구(이명희 2003; 손정표 2003 등)와 학술지와 색인지(Information Science Abstract, ALISE(Association for Library & Information Education), E-LIS)에서 사용된 분류표를 면밀히 검토한 결과, 주제를 정보학과 문헌정보학으로 나누어 분류하지 않으면서 어느 한쪽에 치우치지 않은 JITA 분류표가 선택되었다. 문헌정보학에서 컴퓨터와 통신기술의 응용이 정보학이나 정보검색과 같은 별도의 교과목으로 소개되었던 초창기와는 달리, 컴퓨터 통신기술이 문헌정보학 전 교과목에서 응용되는 현시점에서 문헌정보학을 전통적인 도서관학과 정보학으로 나누는 것은, 예를 들면 수동 색인과 분류는 문헌정보학으로 자동 색인과 분류는 정보

학으로 나누는 것은 바람직하지 않다고 판단되었기 때문이다.

JITA 문헌정보학 분류표(JITA Classification Schema of Library & Information Science)는 E-LIS(E-Print in Library and Information Science)에서 사용되는 분류표로, 12개의 대항목과 135개의 하위항목으로 이루어졌다. 본 연구에서는 해당 영역의 문헌이 극히 적은 E(출판, 법률)영역의 문헌을 A(문헌정보학 일반)에, K(시설, 설비)영역의 문헌은 F(경영)에 포함시켜 10구분하였다.

연구논문의 주제를 분석하는 과정에서는 둘 이상의 복수의 주제항목과 관련된 연구논문은 연구자가 임의로 가장 적합하다고 생각되는 단일주제로 분류하였다. 관중보다는 주제가, 대상(Matter)보다는 처리(Energy, Process)가 우선이었다. 그러므로 학교도서관 이용자 조사를 이용자 연구에, 웹문서의 클러스터링은 정보처리(지식조직)에 분류하였다.

논문에서의 통계사용 여부는 수집된 데이터를 비교 분석하는 목적으로 처리된 수량데이터가 언급된 경우 통계데이터가 사용된 논문으로

간주되었다. 선행연구로부터 인용된 수량데이터나 표본집단의 특성을 나타내는 배경정보를 위한 수량데이터만 언급된 논문은 통계데이터가 사용된 논문으로 간주하지 않았다.

한 개의 논문에서 2개 이상의 추론통계기법이 사용된 경우 추론통계를 사용한 논문수를 산출할 때는 한 개로 간주되었으나, 논문에 사용된 각 추론통계기법을 분석한 경우에는 복수로 개개의 기법이 모두 분석되었다. 기술통계만 사용된 경우는 기술통계로, 기술통계와 추론통계 기법이 모두 사용된 경우는 추론통계 기법을 사용한 논문으로 처리되었다. 논문의 연구방법에서는 어떤 통계기법으로 데이터가 분석되었고 기술하고 있으나, 분석된 결과가 기술되지 않은 경우는 사용된 추론통계기법으로 계산되지 않았다. 추론통계는 t검정, ANOVA, 상관계수, x^2 , 회귀분석과 같은 기초 검정기법 외에 다변량검정까지 사용된 모든 통계기법을 조사하였다. ANOVA에서 사후검정에 사용된 Tukey나 Scheffe 등은 별도의 추론통계기법으로 간주하지 않았다.

가설검정을 위해 수집된 데이터는 교차분석 x^2 으로 검정되었다.

4. 데이터 분석

4.1 통계기법 사용 분석

가설1: 논문에서 통계기법의 사용은 국내외 학술지간에, 학술지의 주제영역간에 차이가 있을 것이다.

2001년 이후 2004년까지 발표된 총 1,768편

의 논문을 분석한 결과, 통계를 사용하지 않은 논문은 815편(46.10%)이었으며, 기술통계를 사용한 논문은 608편(34.39%), 추론통계를 사용한 논문은 345편(19.51%)이었다. 2000년 이전의 선행연구에 비해 기술통계와 추론통계의 사용이 증가하였다.

〈표 3〉에서 보는 바와 같이 통계사용률은 국내 학술지가 40.78%임에 비해 외국 학술지는 64.55%로 큰 차이가 있었다. 통계를 사용한 논문의 비율이 가장 높은 학술지는 국내에서는 정보관리학회지(57.89%), 외국에서는 IPM(76.76%)이었다. 국내 학술지에서는 정보관리학회지 논문에서 기술통계(43.06%)와 추론통계(14.83%)가 가장 많이 사용되었고, 외국 학술지에서는 기술통계는 IPM(50.27%)에서 추론통계는 JASIST(30.92%)의 논문에서 가장 많이 사용되었다.

국내외 논문 모두에서 기술통계가 추론통계보다 많이 사용되었다. 기술통계는 국내 논문의 29.92%, 외국 논문의 38.01%에서 사용되었다. 추론통계는 국내(10.86%)보다 외국(26.54%)에서 2.4배 더 많이 사용되었고, 이러한 국내의 통계사용의 차이는 유의한 것이었다($x^2=116.97$, $df=2$, $p=.000$, 표 4 참조).

학술지의 영역별로는 정보학영역 학술지에서 통계사용의 비율은 64.18%(586편)로 문헌정보학영역 학술지의 42.93%(367편)보다 높았다. 기술통계의 사용률은 정보학영역(39.10%)이 문헌정보학영역(29.36%)보다 높았으며, 추론통계의 사용률 또한 정보학영역 학술지(25.08%)가 문헌정보학영역 학술지(13.57%)보다 1.8배 높은 것으로 나타났으며, 이러한 차이는 통계적으로 유의한 것으로 분석되었다($x^2=88.50$, $df=2$, $p=.000$, 표 4 참조).

〈표 3〉 국내외 학술지별 통계 사용 비교(2001~2004)

()안은 비율(%)

	잡 지 명	통계사용안함	기술통계사용	추론통계사용	전체
국내	정보관리학회지	88(42.1)	90(43.1)	31(14.8)	209
	한국문헌정보학회지	130(58.8)	70(31.7)	21(9.5)	221
	한국도서관정보학회지	188(70.4)	53(19.9)	26(9.7)	267
	한국비블리아학회지	63(66.3)	24(25.3)	8(8.4)	95
	전체	469(59.22)	237(29.92)	86(10.86)	792
국외	JASIST	136(33.9)	141(35.2)	124(30.9)	401
	JOD	60(50.8)	33(28.0)	25(21.2)	118
	IPM	43(23.2)	93(50.3)	49(26.5)	185
	CRL	34(26.2)	64(49.2)	32(24.6)	130
	LQ	40(69.0)	13(22.4)	5(8.6)	58
	LISR	33(39.3)	27(32.1)	24(28.6)	84
	전체	346(35.45)	371(38.01)	259(26.54)	976

〈표 4〉 국내외, 주제영역별 통계사용 논문수 비교

	통계사용안함	기술통계	추론통계	전체	통계량, p
국내	469(59.2)	237(29.92)	86(10.86)	792	$\chi^2=116.97,$ df=2, p = .000
국외	346(35.45)	371(38.01)	259(26.54)	976	
문헌정보학	488(57.08)	251(29.36)	116(13.57)	855	$\chi^2=85.50,$ df=2, p = .000
정보학	327(35.82)	357(39.10)	229(25.08)	913	
전체	815(46.10)	608(34.39)	345(19.51)	1768	

4.2 추론통계 사용 논문 분석

1) 주제 분석

가설2: 추론통계를 사용한 논문의 주제분야 별 사용도는 국내외 학술지간에, 학술지의 주제영역간에 차이가 있을 것이다.

통계를 사용한 953편의 논문 중 추론통계를 사용한 논문 345편을 대상으로 논문의 주제를 분석한 결과, 〈표 5〉에서 보는 바와 같이 전체적으로 추론통계는 정보검색/기술 분야(24.93%)에서 가장 많이 사용되었고, 그 다음이 이용자 연구(이용자교육, 독서지도 포함), 정보이용(계량정보학 포함) 순이었다. 정보검색/기술 분야에서 추론통계를 사용한 논문이 가장 많은

이유는 이 분야가 2001년 이후 연구가 가장 활발한 영역이고, 가설검정이 필요한 실험연구 기법을 주로 사용하기 때문인 것으로 보인다 (표 6 참조). 이용자연구 영역의 논문수가 많은 이유는 대학도서관 이용(자)에 관한 연구나 이용자의 온라인(혹은 OPAC)탐색 형태에 관한 조사를 관중이나 OPAC보다는 이곳에 분류했기 때문이며, 가설 검정이 필요한 서베이연구가 주로 사용되기 때문인 것으로 보인다. 수서나 장서보존, 전자문서변환(Digitization), 대출, 상호대차 등에 관한 도서관기술봉사 영역에서 추론통계를 사용한 논문은 매우 적었다.

국내와 외국을 비교할 때, 국내에서는 이용자 연구(22.1%)에서 추론통계를 사용한 논문이 가

장 많으나, 외국에서는 정보검색/기술(30.9%)에서 추론통계를 사용한 논문이 가장 많았다. 추론통계를 사용한 논문의 주제별 분포는 국내와 외국 학술지 사이에 유의한 차이가 있는 것으로 분석되었지만($\chi^2=51.217$, $df=9$, $p=.000$, <표 5> 참조), 5보다 작은 기대빈도를 가지는 셀이 5개(25%)나 되어 계산된 검정 통계량은 χ^2 -분포에 근사하지 않는 것으로 보이나, 셀을 합치지 않고(문헌 수가 작은 분류항목을 다른 분류항목으로 합쳐 χ^2 -분포로 만들 수 있다) 10구분을 유지하였다.

학술지의 영역별로는, 문헌정보학영역 학술지에서는 이용자연구(29.3%)에 추론통계를 사용한 논문이 가장 많았고, 정보학영역 학술지에서는 정보검색/기술(34.9%)에 추론통계를 사용한 논문이 가장 많았다. 문헌정보학영역 학술지에서는 전문직이나 경영과 관중에서, 정보학영역 학술지에서는 정보처리/조직에서 상대적으로 추론통계를 사용한 논문이 많았고, 두 학술지군 사이의 차이는 유의하였다($\chi^2=104.86$,

$df=9$, $p=.000$, 표 5 참조).

2) 연구방법 분석

가설3: 추론통계를 사용한 논문에서 연구방법의 사용은 국내의 학술지간에, 학술지의 주제영역간에 차이가 있을 것이다.

논문에 사용된 연구방법은 7종류로 분류되었다. 이론적 고찰이나 역사적 연구방법을 사용한 논문에서는 추론통계가 사용되지 않았고, 델파이기법이나 사례연구, 질적 연구방법을 사용한 논문에서 추론통계가 사용된 논문의 수는 매우 적어 기타로 합하였다. 특정 도서관이나 시스템을 대상으로 질량분석한 연구는 사례연구로 간주하지 않았고, 질량분석과 함께 정성분석이 사용된 연구를 사례연구로 간주하였다. 시스템설계를 다룬 논문에서 추론통계는 설계된 시스템을 비교평가하기 때문에 유사실험연구로 간주하였다.

전체적으로 추론통계를 사용한 논문에서 사용된 연구방법은 가설검정을 요구하는 실험연

<표 5> 국내외, 주제영역별 추론통계사용 논문의 주제분류 비교

주제분류	국내외		주제영역		전체
	한국	외국	문헌정보학	정보학	
문헌정보학 일반	6(7.0)	12(4.6)	9(7.8)	9(3.9)	18(5.22)
정보이용	10(11.6)	43(16.6)	11(9.5)	42(18.3)	58(16.81)
이용자연구	19(22.1)	63(24.3)	34(29.3)	48(21.0)	82(23.77)
관중	5(5.8)	5(1.9)	9(7.8)	1(0.4)	10(2.90)
경영, 건물	14(16.3)	7(2.7)	16(13.8)	5(2.2)	21(6.09)
전문직(교육)	8(9.3)	8(3.1)	16(13.8)	0(0)	16(4.64)
정보원	10(11.6)	13(5.0)	8(6.9)	15(6.6)	23(6.67)
정보처리/조직	8(9.3)	22(8.5)	4(3.4)	26(11.4)	30(8.70)
도서관기술봉사	0(0)	6(2.3)	3(2.6)	3(1.3)	6(1.74)
정보검색/기술	6(7.0)	80(30.9)	6(5.2)	80(34.9)	86(24.93)
전체	86	259	116	229	345
통계량, p	$\chi^2=51.217$, $df=9$, $p=.000$		$\chi^2=104.86$, $df=9$, $p=.000$		

구와 서베이연구가 대부분(71.6%)을 차지하였다. 국내 연구에서는 서베이연구의 사용률이 가장 높은 반면(44.2%), 외국 연구에서는 실험 연구의 사용률이 가장 높았다(51.0%). 국내와 외국의 추론통계를 사용한 연구에서 사용된 연구방법 사이에는 유의한 차이가 있는 것으로 분석되었다($x^2=46.551$, $df=6$, $p=.000$, 표 6 참조).

문헌정보학영역과 정보학영역의 학술지군에서 추론통계를 사용한 연구에서 사용된 연구방법 사이에서도 유의한 차이를 보였다($x^2=76.038$, $df=6$, $p=.000$, 표 6 참조). 문헌정보학영역의 학술지에서는 서베이연구(54.3%)와 실험연구(19.8%)가, 정보학영역의 학술지에서는 실험연구(56.3%)와 계량정보(16.2%)가 가장 많이 사용되었다.

3) 저자의 전공 분석

가설4: 추론통계를 사용한 논문의 저자의 전공은 국내외의 학술지간에, 학술지의 주제영역간에 차이가 있을 것이다.

논문의 저자는 저자 수와는 상관없이 연구에 참여한 저자(들)이 문헌정보학전공자들인지, 문헌정보학을 제외한 타주제분야 전공자들인지, 아니면 문헌정보학 전공자와 타주제 전공자들이 공동으로 연구를 수행했는지로 구분하였다. 문헌정보학전공자는 문헌정보학과 교수나 현장 사서를 포함하였다. 북미의 경우는 ALA에서 인정하는 학교¹⁾의 교수만을 문헌정보학과 교수로 코딩하였다. 현장사서가 독립 혹은 공동으로 참여하여 수행한 연구는 60편(국내 16편, 외국 44편: 문헌정보학 영역 45편, 정보학 영역 15편)이었다.

전체적으로 추론통계를 사용한 문헌정보학 연구는 문헌정보학전공자(57.7%)에 의해 가장 많이 수행되었고, 타전공자에 의해 혹은 공동으로 연구된 논문이 42.3%를 차지하였다. 국내에서는 추론통계를 사용한 문헌정보학 논문의 대부분(82.6%)이 문헌정보학전공자에 의해 연구되었지만, 외국에서는 49.4%가 문헌정보학전공자에 의해, 나머지 50.6%는 타전공자에 의해 혹은 공동으로 연구되었다. 문헌정보학전공자와 타

〈표 6〉 국내외, 주제영역별 추론통계사용 논문의 연구방법 비교

연구방법	국내외		주제영역		전체
	한국	외국	문헌정보학	정보학	
서베이연구	38(44.2)	57(22.0)	63(54.3)	32(14.0)	95(27.54)
실험	20(23.3)	132(51.0)	23(19.8)	129(56.3)	152(44.06)
평가	11(12.8)	5(12.0)	5(4.3)	11(4.8)	16(4.64)
모형구축	4(4.7)	4(1.5)	4(3.5)	4(1.7)	8(2.32)
계량	7(8.1)	39(15.1)	9(7.8)	37(16.2)	46(13.33)
내용분석	2(2.3)	15(13.5)	6(5.2)	11(4.8)	17(4.93)
기타	4(4.7)	7(2.7)	6(5.2)	5(2.2)	11(3.19)
전체	86	259	116	229	345
통계량, p	$x^2=46.551$, $df=6$, $p=.000$		$x^2=76.038$, $df=6$, $p=.000$		

1) <<http://www.ala.org/ala/accreditation/lisdirb/Alphaaccred.htm>.> 2005.08.10 방문.

전공자와의 공동연구는 외국(13.1%)에 비해 국내(3.4%)에서는 매우 드물었다. 이와 같은 국내와 외국의 차이는 통계적으로 유의한 것이었다 ($x^2=29.167$, $df=2$, $p=.000$, 표 7 참조).

학술지의 주제영역으로 구분하면 전통적인 문헌정보학학술지에서는 추론통계를 사용한 연구가 문헌정보학전공자들에 의해 주도적으로 (84.5%) 이루어졌지만, 정보학영역에서는 문헌정보학 비전공자들(44.5%)이 문헌정보학전공자(44.1%)보다 조금 더 많은 연구를 수행하였다. 공동연구는 문헌정보학영역(9.5%)보다는 정보학영역(11.4%)에서 더 많이 수행되었다. 정보학영역의 학술지와 문헌정보학영역의 학술지 간 저자의 전공별 차이는 통계적으로 유의한 것이었다($x^2=58.151$, $df=2$, $p=.000$, 표 7 참조).

4.3 추론통계기법 분석

국내외의 345편의 논문에는 총 35종의 추론통계기법이 다양하게 사용되었다. 모수통계로는 2개 이하 집단의 평균 검정으로 단일표본 t검정, 대응2표본 t검정, 독립2표본 t검정이 사용되었고, 2개 이상의 집단에 대한 평균검정으로는 일원과 2원 ANOVA가 사용되었다. 상관분석에

는 피어슨과 코사인 상관계수가, 회귀분석으로는 단순선형, 다중선형, 로지스틱, 비선형회귀분석이 사용되었다. 다변량분석으로는 다변량분산분석(MANOVA), 요인분석, 군집분석, MDS, 판별분석, 경로분석, 대응일치분석, 공변량구조분석, 주성분분석, 일반선형모형(GLM)이 사용되었으며 시계열분석도 사용되었다.

비모수통계로는 범주형 데이터의 검증을 위해서는 x^2 (단일표본, 두 독립표본), McNemar(두 종속표본)가 사용되었고, 서열형 데이터의 검증을 위해서는 Kolmogorov-Smirnov(단일표본, 두 독립표본), Mann-Whitney(두 독립표본), Sign, Wilcoxon 대응표본검정, Kruskal-Wallis 일원분산분석(여러 독립표본), Friedman 이원분산분석(여러 종속표본)이 사용되었다. 상관분석에서는 스피어맨과 켄달의 상관계수가 사용되었다.

신뢰도분석에는 Cohen의 카파와 Cronbach의 알파가 사용되었다.

선행연구에서 추론통계 기법은 다섯 개의 기본 검정 통계기법(t검정, ANOVA, 상관계수, 회귀분석, x^2)과 기타로 6구분되었다. 본 연구에서도 선행연구와의 비교를 위해 동일한 6구분을 사용하였으나 기타는 더 세분하여 분석하였다. ANOVA검정에는 1원 ANOVA와 2원

<표 7> 국내외, 주제영역별 추론통계사용 논문의 저자전공 비교

	국내외		주제영역		전체
	국내	국외	문헌정보학	정보학분야	
문정전공	71(82.6)	128(49.4)	98(84.5)	101(44.1)	199(57.68)
문정비전공	12(14.0)	97(37.5)	7(6.0)	102(44.5)	109(31.59)
공동연구	3(3.4)	34(13.1)	11(9.5)	26(11.4)	37(10.72)
전체	86	259	116	229	345
통계량, p	$x^2=29.167$, $df=2$, $p=.000$		$x^2=58.151$, $df=2$, $p=.000$		

ANOVA를 포함시켰으며, MANOVA는 다변량분석에 포함시켰다. 상관분석에는 피어슨, 코사인 외에 비모수 스피어맨과 켄달도 포함시켰다. 회귀분석에는 단순선형회귀, 다중선형회귀, 로지스틱회귀, 비선형회귀까지 포함시켰다. 기타 비모수에는 χ^2 과 켄달과 스피어맨의 상관계수를 제외한 나머지 모든 비모수검정을 포함시켰다.

추론통계는 345편의 논문에서 총 663번 사용되었다. 추론통계기법은 논문당 평균 2종류가 사용되었으며 가장 많이 사용된 논문에서는 6종류가 사용되었다. 추론통계기법은 밝히지 않은 채 유의수준이나 p값만 제시하고 있는 외국논문 7편은 분석에서 제외하고, 총 656회 사용된 추론통계기법이 분석되었다.

전체적으로 추론통계기법은 다변량분석이 가장 많이 사용되었고(126회, 19.2%), 상관분석(119회, 18.1%), t검정(95회, 14.5%), ANOVA(87회, 13.3%), χ^2 (72회, 11.0%), 회귀분석(57회, 8.7%), 신뢰도분석(55회, 8.4%), 기타 비모수검정(45회, 6.9%) 순으로 많이 사용되었다. 다변량분석에서는 요인분석이 44회(35.2%)로 가장 많이 사용되었고, 군집분석(24회)과 MSD(23회)가 뒤를 이었다. 상관분석에서는 피어슨의 상관계수가 가장 많이 82회(68.9%) 사용되었고, 스피어맨과 켄달과 같은 비모수 상관계수는 29회(24.4%) 사용되었다. t검정에서는 독립표본 t검정인지 종속표본 t검정인지를 밝히지 않은 45편을 제외하고는 종속표본 t검정이 27회로 독립표본 t검정보다 많이 사용되었으나, t검정의 종류를 밝히지 않은 45편은 대부분 사용된 표본의 성격으로 보아서 독립표본 2검정을 사용한 것으로 보였다. ANOVA에서

는 2원분산분석이 10회(11.5%) 사용됨에 비해 88.5%가 1원분산분석(77회)을 사용하였다. 회귀분석(57회)에서는 단순선형 회귀분석은 7회(12.2%) 사용됨에 비해 다중선형회귀는 38회(66.7%)로 가장 많이 사용되었다. 기타 비모수검정에서는 Wilcoxon의 대응표본검정이 12회(26.7%) 사용되었고, Kruskal-Wallis(8회), K-S(7회), Mann-Whitney(7회), Friedman(6회) 순으로 많이 사용되었다.

1) 국내외 비교

가설5: 사용된 추론통계기법은 국내외 학술기간에 차이가 있을 것이다.

추론통계기법은 국내 학술지에서는 86편의 논문에서 207회, 외국 학술지에서는 259편의 논문에서 449회 사용되었다. 논문당 사용된 추론통계기법은 국내(2.4회)가 외국(1.7회)보다 높았다.

국내에서는 다변량분석이 가장 많이(19.3%) 사용되었고, 다음으로 t검정이 (16.9%) 사용됨에 반하여, 외국에서는 상관분석이 가장 많이(19.6%) 사용되었고, 다음으로 다변량분석이(19.2%) 사용되었다. 비모수 상관계수와 χ^2 를 포함하여 비모수통계는 주로 외국 논문에서 사용되었다. 단일표본 t검정, 2원 ANOVA, 코사인 상관계수, 비선형회귀분석, 시계열분석, 주성분분석, 카파, K-S, Mann-Whitney, McNemar, Friedman의 검정은 외국에서만 사용되었다. 8개 유형의 추론통계기법의 사용은 국내외 외국 논문에서 통계적으로 유의한 차이를 보였다($\chi^2 = 44.282$, $df = 7$, $p = .000$, 표 8 참조).

〈표 8〉 국내외 추론통계기법 사용

	t	ANOVA	상관	회귀	χ^2	기타 비모수	다변량	신뢰도	전체
국내	35(16.9)	27(13.1)	31(15.0)	22(10.6)	15(7.2)	3(1.4)	40(19.3)	34(16.4)	207(31.55)
국외	60(13.4)	60(13.4)	88(19.6)	35(7.8)	57(12.7)	42(9.4)	86(19.2)	21(4.7)	449(68.45)
전체	95(14.48)	87(13.26)	119(18.14)	57(8.69)	72(10.98)	45(6.86)	126(19.21)	55(8.38)	656

($\chi^2=44.282, df=7, p=.000$)

2) 학술지 주제영역별 비교

가설6: 사용된 추론통계기법은 학술지의 주제영역간에 차이가 있을 것이다.

추론통계기법은 정보학영역의 학술지에서는 229편의 논문에서 416회, 문헌정보학영역의 학술지에서는 116편의 논문에서 240회 사용되었다. 논문당 사용된 추론통계기법은 문헌정보학(2.1회)이 정보학(1.8회)보다 높았다.

문헌정보학영역의 학술지에서는 기타비모수를 제외하고 상관분석(16.7%)과 t검정(15.8%), 다변량분석(15.0%), χ^2 (14.6%) 순으로 비슷하게 사용되었으나, 정보학영역에서는 다변량분석(21.6%)과 상관분석(19.0%)이 가장 많이 사용되었다. t검정과 회귀분석, χ^2 , 신뢰도검정은 정보학영역의 학술지보다는 문헌정보학영역의 학술지에 실린 논문에서 더 많이 사용되었으며, ANOVA와 상관분석, 기타 비모수검정, 다변량검정은 정보학영역 학술지에 실린 논문에서 더 많이 사용되었고 이러한 차이는 통계적으로 유의하였다($\chi^2=33.99, df=7, p=.000$, 표 9 참조). 단일표본 t검정, 판별분석, 비선형회귀분석, 시계열분석, K-S, Mann-Whitney, McNemar, Friedman 검정은 정보학영역의 학술지에서만 사용되었다.

3) 주제분류별 비교

가설7: 사용된 추론통계기법은 주제분류별

차이가 있을 것이다.

주제구분에 따라 사용된 추론통계기법을 분석하면, 논문당 사용된 추론통계기법의 수는 경영분야가 2.7회(57/21)로 가장 높았고 도서 관봉사가 1.3회(8/6)로 가장 낮았다.

정보이용(37.8%)과 정보처리/조직(28.6%), 문헌정보학 일반(27.6%)에서 다변량분석이 가장 많이 사용되었다. 경영에서는 다변량분석과 회귀분석(각각 15.8%), 정보검색/기술에서는 상관계수(26.3%)가 가장 많이 사용되었다. 다변량분석 중 요인분석은 이용자연구(44회 중 17회)에서, 군집분석과 MDS는 계량정보학(군집분석을 사용한 24회중 9회, MDS를 사용한 23회 중 10회)에서 가장 많이 사용되었지만, 군집분석은 문헌의 자동군집 연구 때문에 정보처리/조직 분야(정보처리/조직분야 49회 중 7회)에서 상대적으로 많이 사용되었다. 주제분류별로 사용된 추론통계기법에는 차이가 있는지에 대한 가설검정은 변수가 너무 세분되어(셀이 90개) 기대빈도가 5%이하인 셀이 전체 셀의 20%를 넘기 때문에 χ^2 검정이 적합지 않다고 사료되어 검정하지 않았다(표 10 참조).

4) 연구방법별 비교

가설8: 사용된 추론통계기법은 연구방법별 차이가 있을 것이다.

연구방법별로 논문당 사용된 추론통계기법

〈표 9〉 주제영역별 추론통계기법 사용

	t	ANOVA	상관	회귀	x^2	기타 비모수	다변량	신뢰도	전체
문헌정보학	38(15.8)	30(12.5)	40(16.7)	28(11.7)	35(14.6)	4(1.7)	36(15.0)	29(12.1)	240(36.59)
정보학	57(13.7)	57(13.7)	79(19.0)	29(7.0)	37(8.9)	41(9.9)	90(21.6)	26(6.3)	416(63.41)
전체	95(14.48)	87(13.26)	119(18.14)	57(8.69)	72(10.98)	45(6.86)	126(19.21)	55(8.38)	656

($x^2=33.99$, $df=7$, $p=.000$)

〈표 10〉 주제분류별 추론통계기법 사용

	t	ANOVA	상관	회귀	x^2	기타 비모수	다변량	신뢰도	전체
문정일반	6	4	2	2	6	0	8	1	29
정보이용	4	4	17	9	12	5	34	5	90
이용자연구	26	19	28	16	22	9	36	20	176
관중	3	3	3	3	5	0	2	2	21
경영	7	7	8	9	3	2	9	12	57
전문직	5	2	6	4	6	0	2	4	29
정보원	7	8	7	5	4	4	5	5	45
정보처리/조직	10	5	6	1	5	5	14	3	49
도서관기술봉사	1	3	2	0	1	0	1	0	8
정보검색/기술	26	32	40	8	8	20	15	3	152
전체	95	87	119	57	72	45	126	55	656

의 수는 모형구축(2.9회)과 평가(2.4회), 서베이(2.3회) 순으로 높았다.

연구방법별 사용된 추론통계기법에 차이가 있는지에 대한 검정도 기대빈도 5이하인 셀이 20%를 넘기 때문에 검정하지 않았다. 그러나 서베이연구에서는 상관계수(16.7%)가, 실험연구에서는 t검정(21.3%)과 상관계수(20.9%)가, 평가연구에서는 t검정과 ANOVA, 다변량분

석(각각 18.4%)이, 모형구축과 계량분석에서는 다변량분석(30.4%와 44.3%)이, 내용분석에서는 x^2 (29.0%)이 가장 많이 사용되었다. 통계기법 측면으로 t검정과 ANOVA, 상관분석, 비모수기법은 실험에서, x^2 과 회귀분석은 서베이연구에서, 다변량분석은 계량분석에서 가장 많이 사용되었다(표 11 참조).

〈표 11〉 연구방법별 추론통계기법 사용

	t	ANOVA	상관	회귀	x^2	기타 비모수	다변량	신뢰도	전체
서베이연구	24	18	36	32	33	5	34	33	215
실험	54	48	53	13	15	29	32	10	254
평가	7	7	4	3	4	0	7	6	38
모형구축	1	2	3	3	1	2	7	4	23
계량	3	7	16	6	8	4	35	0	79
내용분석	4	3	6	0	9	2	6	1	31
기타	2	2	1	0	2	3	5	1	16
전체	95	87	119	57	72	45	126	55	656

5) 저자전공별 비교

가설9: 사용된 추론통계기법은 저자의 전공별 차이가 있을 것이다.

논문당 사용된 추론통계기법의 수를 저자의 전공으로 분류하면, 문헌정보학 전공자는 논문당 평균 1.95개(199편에서 389회)의 추론통계기법을, 문헌정보학 비전공자는 1.69개(109편에서 184회)의 추론통계검정을 사용하였다. 그러나 공동연구는 2.24개(37편에서 83회)로 논문당 가장 많은 추론통계를 사용하였다.

세그룹 모두 다변량분석을 가장 많이 사용하였다. 문헌정보 비전공자는 상대적으로 상관분석(21.7%)과 다변량분석(21.7%), 기타 비모수검정(11.4%)을 문헌정보학 전공자보다 많이 사용하였고, 공동연구는 다른 저자군에 비해 ANOVA(18.1%)와 회귀분석(16.9%)을 더 많이 사용하였다. 그러나 그 차이는 통계적으로 유의한 수준은 아니었다($x^2=21.449$, $df=14$, $p=.091$, 표 12 참조).

4.6 비모수통계 사용

가설10: 사용된 추론통계기법에서 비모수통계기법의 사용에는 차이가 있을 것이다.

전체 656회 사용된 추론통계기법을 비모수검정과 기타/모수검정으로 구별하여 분석하였다.

스피어맨과 켄달의 상관계수, x^2 , Cohen의 카파, K-S, Mann-Whitney, McNemar, Sign, Wilcoxon, Kruskal-Wallis, Friedman 검정을 비모수통계로 분류하였다.

전체 656회의 추론통계 중 비모수통계는 149회(22.71%), 기타/모수통계는 507회(77.29%) 사용되었다. <표 13>에서 보는 바와 같이 비모수통계는 국내(10.6%)보다는 외국에서 많이 사용되었으며(28.3%), 이는 유의한 차이를 보였다($p=.000$). 또한 연구방법별로는 내용분석 연구(48.4%)와 기타 연구(37.5%)에서 비모수통계가 전체 평균(22.71%)보다 많이 사용되었다($p=.0024$). 그러나 학술지 주제영역별, 주제분류별, 저자전공별 구분에서는 차이가 없는 것으로 분석되었다.

5. 추론통계 사용상의 문제점

문헌정보학 연구에서 추론통계를 사용하여 가설을 검정하고 결과를 기술하는 과정을 분석한 결과 다음과 같은 문제들이 파악되었다.

첫째는 가설을 세우고 검정하는데의 문제이다. 가설검정이란 표본으로부터 구한 통계량을 이용하여 설정한 가설의 진위여부를 검정하는 것이며, 가설에는 귀무가설과 연구가설이 있다.

<표 12> 저자전공별 추론통계기법 사용

	t	ANOVA	상관	회귀	x^2	기타 비모수	다변량	신뢰도	전체
문정	61(15.7)	53(13.6)	65(16.7)	33(8.5)	51(13.1)	20(5.2)	70(18.0)	36(9.3)	389(59.30)
타전공	22(12.0)	19(10.3)	40(21.7)	15(8.2)	12(6.5)	21(11.4)	40(21.7)	15(8.2)	184(28.05)
공동	12(14.5)	15(18.1)	14(16.9)	9(16.9)	9(10.9)	4(4.8)	16(19.3)	4(4.8)	83(12.65)
전체	95(14.48)	87(13.26)	119(18.14)	57(8.69)	72(10.98)	45(6.86)	126(19.21)	55(8.38)	656

($x^2=21.449$, $df=14$, $p=.091$)

〈표 13〉 비모수통계 사용 비교

		비모수	기타	전체	통계량, p값
국내외	국내	22(10.6)	185(89.4)	207	$\chi^2=25.163$ df=1, p=.000
	국외	127(28.3)	322(71.7)	449	
주제영역	문헌정보학	47(19.5)	193(80.4)	240	$\chi^2=2.112$ df=1, p=.087
	정보학	102(24.5)	317(74.9)	416	
주제분류	문헌정보학 일반	7(24.1)	22(75.9)	29	$\chi^2=9.128$ df=9 p=.426
	정보이용	21(23.3)	69(76.7)	90	
	이용자연구	34(19.3)	142(80.7)	176	
	관중	6(28.6)	15(71.4)	21	
	경영	6(10.5)	51(89.5)	57	
	전문직	8(27.6)	21(72.4)	29	
	정보원	11(24.4)	34(75.6)	45	
	정보처리/조직	12(24.5)	37(75.5)	49	
	도서관기술봉사	2(25.0)	6(75.0)	8	
정보검색/기술	42(27.6)	110(72.4)	152		
연구방법	서베이연구	47(21.9)	168(78.1)	215	$\chi^2=20.325$ df=6 p=.0024
	실험	57(22.4)	197(77.6)	254	
	평가	5(13.2)	33(86.8)	38	
	모형구축	3(13.0)	20(87.0)	23	
	계량	16(20.3)	63(79.7)	79	
	내용분석	15(48.4)	16(51.6)	31	
저자전공	문정전공	84(21.6)	305(78.4)	389	$\chi^2=1.71$ df=2, p=.425
	문정비전공	48(26.1)	136(73.9)	184	
	공동	17(20.5)	66(79.5)	83	
전체		149(22.71)	507(77.29)	656	

일반적으로 방향성이든 비방향성이든 귀무가설에는 차이가 없으나, 연구가설을 비방향성 가설로 세울 때는 양측검정을 수행하고(χ^2 는 제외) 연구가설을 방향성 가설로 세우면 단측검정을 수행할 수 있다. 그러나 대부분의 연구에서는 귀무가설이나 연구가설을 설정하지 않고 양측검정이나 단측검정을 수행하였다. 연구가설을 설정한 경우에도 양측검정인지 단측검정인지 언급이 없고, 심지어 대립가설을 비방향성으로 세우고 단측검정하거나 방향성으로 세우고 양측검정하는 경우도 있었다(Nichols 2003).

두 번째는 사용된 추론통계기법을 분명하게 언급하지 않는 경우이다. 345개의 문헌 중 7개

의 논문에서는 사용된 추론통계기법을 언급하지 않았고, t검정을 사용한 논문 95편 중 45편에서는 단일표본 t검정인지, 독립표본 t검정, 종속(대응)표본 t검정인지 언급이 없었으며, 상관분석을 사용한 논문 118편 중에서도 6편은 어떤 상관계수를 사용했는지 언급하지 않았다.

세 번째는 가설검정의 결과를 제시하는데의 문제이다. 검정결과는 주로 검정통계량과 유의수준이나 유의확률로 제시된다. 유의확률(p값)은 귀무가설을 기각할 최소의 유의수준으로, 유의확률(p값)이 유의수준보다 작으면 귀무가설을 기각하고 크면 귀무가설을 채택한다. 일반적으로 유의수준은 0.05, 0.01, 0.001을 주로

사용하나 사회과학에서는 0.1도 사용한다. 귀무가설의 기각여부를 기술할 때 유의수준보다는 유의확률(p값)이 더 많은 정보를 함축한다. 즉, 귀무가설을 기각하는 확률이 “유의수준 0.05보다 작다” 보다는 “ $p=0.025$ 이다”로 기술하는 것이 보다 많은 정보를 준다.

656회 사용된 추론통계기법 중 t검정(95)과 ANOVA(87), 상관분석(119), χ^2 (72), 기타 비모수(45) 통계를 사용하여 가설을 검정한 418개의 검정 중 43.8%(183)의 검정에서는 통계량과 p값이, 14.4%(60)의 검정에서는 p값만, 22.2%(93)의 검정에서는 통계량과 유의수준이 제공되었다. 그러나 11.2%(47)의 검정에서는 ‘유의수준 0.05에서 유의하였다’와 같이 유의수준만, 7.2%(30)에서는 유의수준이나 p값은 제공하지 않고 통계량만, 1.2%(5)에서는 유의수준도 언급하지 않고 유의하였다 혹은 유의하지 않았더라고만 기술하고 있다.

회귀분석 결과는 모형의 적합도(F통계량과 p값)와, 회귀계수와 회귀계수의 유의성 검정(t), 결정계수(r^2)값으로 제시된다. 회귀분석을 사용한 57편의 논문 중 통계데이터를 제공하지 않는 논문이 1편이었고, 나머지 모든 논문은 회귀계수와 유의성 검정결과를 제시하고 있지만, 모형의 적합도와 결정계수까지 제시한 논문은 11편 뿐이었다.

대부분의 다차원척도법(MDS)에서는 모형의 적합도를 나타내는 S 스트레스값이나 RSQ에 대한 논의없이 집단도표만을 제시하고 있으며, 군집분석에서도 군집수의 적합도에 대한 검정통계량(Hotel ling's T^2 이나 Cubic Clustering

Criterion 방법) 없이 덴드로그램만을 제시한 경우가 주류를 이루었다.

네 번째는 신뢰도검사를 위한 크론바하(Cronbach)의 알파를 사용하는데의 문제이다. 크론바하의 알파는 설문지에서 하나의 개념을 하나의 문항으로 묻는 대신 여러 문항으로 물어 문항들이 하나의 개념(변수)를 얼마나 잘 표현하는지를 알아보는 내적 일치도를 측정하는데 사용되는 신뢰도계수이다. 요인분석으로 묶인 요인의 내적 일치도를 측정하는데도 사용된다. 크론바하의 알파는 0과 1사이의 값이고 1에 가까울수록 내적 일치도가 높다. 일반적으로 0.6 이상이면 문제가 없다고 하나 α 값은 문항수가 많을수록 응답자수가 많을수록 높아지는 경향이 있으므로, 값의 크기가 판단의 근거가 되는 것이 아니라, 그 문항을 제외했을 때 α 값이 적어지느냐 커지느냐를 보고 그 문항을 제외하느냐 그대로 두느냐를 판단해야하는 것이다²⁾.

크론바하의 α 를 이용하여 신뢰도를 검사한 52편의 논문 중 82.7%(43편)는 α 값이 충분히 높거나(0.8이상) 낮을 경우에는 문항을 제거하고 문항제거 시 α 계수를 제시하였다. 그러나 9.6%(5편)는 α 가 낮은데도 무시하고 그대로 사용했으며, 3.85%(2편)는 문항을 제거하지 않고 변수자체를 제거했으며, 3.85%(2편)는 그룹화된 문항들의 일치도 대신 질문지에 사용된 모든 문항의 일치도를 측정하는데 사용하였다.

마지막으로 추론통계를 사용하여 데이터를 분석하는 과정에서 사용하는 여러가지 방법적 선택과 결과해석을 위한 기본적인 데이터를 기술하는데 나타난 문제점이다. 다중회귀분석(38

2) 권세혁, SAS SPSS 활용 설문조사 분석. 자유아카데미, 2004. p.168.

편)에서 변수선택방법(단계, 후진, 전진)을 설명한 논문은 6편 뿐이었고, 다중공선성 문제를 언급한 논문은 3편 뿐이었으며, 이상치와 영향치를 언급한 논문은 한편도 없었다. 대부분의 다차원척도법(MDS)에서는 사용된 척도와 거리계산모형에 대한 설명이 없었고, 군집분석에서도 선택된 군집방법과 유사도측정방법에 대한 설명이 부족하였다.

이 외에도 χ^2 검정에서는 5 이하인 기대빈도셀이 20% 이상일 경우에는 χ^2 분포를 따른다고 볼 수 없음에도 이에 대한 언급없이 χ^2 검정을 수행한 논문이 여럿이었다.

6. 결론 및 제언

본 연구에서는 국내외 문헌정보학 학술지 10종에서 2001년부터 2004까지 4년동안 발표된 연구논문 1,768편을 분석한 결과, 34.39%(608편)의 논문은 기술통계를 사용하였고 19.51%(345편)의 논문은 추론통계를 사용한 것으로 분석되었다.

기술통계의 사용률은 외국 학술지에서는 38.01%로써, 이는 1975년 논문의 42.2%(Wylly 1978)보다는 낮으나 1981년 논문에서의 사용률 20%(Wallace 1985), 1985년 논문에서의 사용률 20.7%(Enger1989)보다는 증가한 것이다. 국내 학술지에서 기술통계의 사용은 29.92%로, 1980-1991년 논문에서의 24%(사공철, 최정희 1993)와, 1997-2001 논문에서의 14.57%(이명희 2003)보다 증가하였다.

추론통계의 사용률은 외국 학술지의 경우는 26.54%로, Atherton(1973)의 12.8%, Water

(1976)의 13.4%, Wylly(1978)의 2.9%, Wallace (1985)의 6%, Enger(1989)의 11.1%보다 최소 2배 이상 증가하였다. 그러나 국내 학술지의 추론통계 사용률은 10.86%로 외국에 비해 현저히 낮았으며, 1980-1991년(사공철 8%)보다는 증가했지만 1997-2001년(이명희 13.4%)보다는 감소한 것으로 나타났다. 한국 논문보다는 외국 논문에서 기술통계와 추론통계가 모두 더 사용되었으며, 문헌정보학영역보다는 정보학영역의 학술지에서 기술통계와 추론통계가 모두 더 사용되었다. 추론통계를 사용한 345편의 논문은 주제분류별로는 한국은 이용자연구, 외국은 정보검색/기술 분야의 논문이, 문헌정보학영역의 학술지에서는 이용자연구, 정보학영역의 학술지에서는 정보검색/기술 분야의 논문이 가장 많았다. 연구방법 측면에서 한국의 논문은 서베이방법, 외국은 실험연구가 가장 많았으며, 문헌정보학영역의 학술지에서는 서베이방법, 정보학영역에서는 실험연구가 가장 많이 사용되었다. 저자의 전공으로 볼 때 한국 논문은 대부분이 문헌정보학을 전공한 저자에 의해 수행되었으나, 외국 논문은 문헌정보학 비전공 혹은 공동연구논문이 더 많았다. 특히 정보학영역의 학술지에서는 문헌정보학 비전공자에 의한 연구가 가장 많았다.

사용된 추론통계기법을 선행연구와 비교한 <표 14>를 보면 2개의 특징을 들 수 있는데, 첫째는 2001년 이후 최근의 연구에서는 다변량분석, 비모수통계 등 기타 통계기법이 가장 많이 사용되었다는 점이다. 국내 논문에서 기타통계의 사용률은 37.2%(77/207)로, 사공철(5%), 이명희(6.25%)와 비교하여 큰 차이를 보인다. 외국 논문에서도 33.2%로 가장 큰 비중을 차지

〈표 14〉 선행연구의 추론통계기법 사용률과의 비교

		t	ANOVA	상관	회귀	χ^2	기타
국내	본연구	16.9	13.1	15.0	10.6	7.2	37.2
	사공철	16	25	20	16	18	5
	이명희	13.75	31.25	17.50	6.25	25.00	6.25
국외	본연구	13.4	13.4	19.6	7.8	12.7	33.2
	Wallace ³⁾	8.1	12.9	21.0	12.9	25.8	19.4

하고 있다.

두 번째 특징은 χ^2 사용의 감소이다. 국내 논문에서 χ^2 의 사용률 7.2%는 이명희(25%), 사공철(18%)에서 비하여 매우 감소한 것이고, 외국 논문에서 사용률 12.7%도 Wallace(25.8%)에 비해 매우 감소한 것이다. Wallace의 분석에서 전체 통계기법 중 가장 많이 사용된 통계기법은 χ^2 이었다. 회귀분석에서도 단순선형회귀보다는 다중, 로지스틱, 비선형 등의 회귀분석이 주로 사용되었다.

사용된 추론통계기법은 국내외간, 학술지의 주제영역간에 유의한 차이가 있었으나, 저자의 문헌정보학 전공과 비전공과의 차이는 없으므로 분석되었다. 신뢰도검사는 주로 국내에서 사용되었으며, 비모수검정은 외국에서 많이 사용되었다. 다변량분석과 비모수검정은 정보학 영역의 학술지에서 더 많이 사용되었다. 연구방법 측면에서 t검정과 ANOVA와 같은 그룹간의 평균비교는 실험연구와 평가연구에서 가장 많이 사용되었고, 다변량분석은 계량정보학에서, χ^2 은 내용분석에서, 회귀분석은 서베이연구에서 가장 많이 사용되었다.

문헌정보학 연구에서 추론통계를 사용하여 가설을 설정하고, 사용된 통계방법과 선택사항들을 언급하며, 검정결과를 기술하는데 여러

가지 문제점들이 파악되었다. 추론통계를 사용한 연구의 신뢰성(Reliability)을 위해서는 연구가설을 세우고 그에 따른 양측 혹은 단측 검정 중 올바른 검정방법을 선택하고, 사용된 추론통계기법을 분명하게 언급하며, 가설검정에 필요한 검정통계량과 유의수준이나 유의확률을 정확히 제공하고, 추론통계를 사용하여 데이터를 분석하는 과정에서 사용된 여러가지 방법적 선택과 결과해석을 위한 기본적인 데이터를 기술하는데 주의깊은 논의가 필요한 것으로 분석되었다.

본 연구결과 나타난 문제점에 대한 제언은 다음과 같다.

첫째 문헌정보학 연구방법론이나 통계학 교과과정에서 t, ANOVA, 상관, 단순회귀, χ^2 분석과 같은 기초통계기법 외에 다변량분석이나 비모수 통계기법에 대해서 보다 많은 관심을 가져야겠다. 문헌정보학 연구에서 가장 많이 사용된 통계분석기법이 다변량분석이었다.

둘째 여러 학문이 합쳐져 이루어진 학제적 응용학문이라는 문헌정보학이 갖고 있는 속성에 부응하여 문헌정보학 전공자와 비전공자간의 공동연구가 활발해져야겠다. 외국의 연구나 정보학영역의 연구에 비하여 국내 연구와 문헌정보학영역의 연구에서는 공동연구의 비율이

3) Wallace의 학술지당 사용 논문수를 비교를 위해 사용률로 변환하였음.

매우 낮았다.

셋째 문헌정보학이 실용학문임을 고려한다면 현장사서들의 현장에서 발생하는 현상을 설명하고 통제하려는 노력이 보다 필요하다 하겠다. 현장사서가 독자적으로 혹은 문헌정보학과 교수나 문헌정보 비전공자와 공동으로 수행한 연구는 345편 중 60편(17.4%)이었다.

끝으로 본 연구에서는 사용된 추론통계기법이 적절한 것인지, 추론통계에서 선택된 방법과 결과를 보다 주의깊게 기술한 논문이 더 좋은 논문인지는 연구하지 않았다. 논문에서 사용된 추론통계에 관한 논의의 정도와 기술방법과 논문의 질(인용빈도)과의 관계를 규명하는 것도 좋은 후속연구가 될 수 있을 것이다.

참 고 문 헌

- 권세혁. 2004. 『SAS, SPSS 활용 설문조사분석』. 자유아카데미.
- 사공철, 최정희. 1993. 문헌정보학 분야의 통계 기법 사용에 관한 계량적 연구. 『도서관학논집』, 20: 31-60.
- 손정표. 2003. 한국의 문헌정보학분야 연구동향 분석 : 1957-2003. 『한국도서관·정보학회지』, 34(3): 9-32.
- 이명희. 2003. 내용분석법에 의한 문헌정보학 학술지 연구논문 분석. 『한국문헌정보학회지』, 36(3): 287-310.
- ALISE (Association for Library and Information Science Education) LIS Research Areas Classification Scheme. [cited 2005.8.10].
<http://www.alise.org/research_class_guide.html>
- Atherton, Pauline. 1973. Developmental paper for area 1 : A science of information. *Research in Information Science: An Assessment*. NATO Advanced study Institute in Information Science, Aberystwyth, Wales, 13-24.
- Childers, Thomas. 1984. "Will the cycle be unbroken? Research and schools of library and information studies." *Library Trends*, 32: 521-535.
- E-LIS JITA Classification Schema of Library and Information Science. [cited 2005. 8.10].
<<http://eprints.rclis.org/jita.html>>
- Enger, Kathy. B, Georgia and Stewart, J. Andrew. 1989. "Statistical methods used by authors of library and information science journal articles." *Library and Information Science Research*, 11: 37-48.
- Hawkins, Donald T., Signe E. Larson, and Bari Q. Caton. 2003. "Information Science Abstracts: Tracking the literature of information science. Part 2: A New taxonomy for information science."

- Journal of the American Society for Information Science and Technology*, 54(8): 771-781.
- Kinnucan, M.T., M.J. Nelson, and B.L. Allen. 1987. "Statistical methods in information science research." *Annual Review of Information Science and Technology*, 22: 147-178.
- Nichols, James, Barbara Shaffer, and Karen Shockey. 2003. "Changing the face of instruction: Is online or in-class more effective?" *Collage and Research Libraries*, 64(5): 378-388.
- Park, Soyeon. 2004. "The study of research methods in LIS education: issues in Korean and U.S. universities." *Library & Information Science Research*, 26: 501-51.
- Schloegl, Christian, and Wolfgang G. Stock. 2004. "Impact and relevance of LIS journals: A scientometric analysis of international and German-language LIS journals - citation analysis versus reader survey." *Journal of the American Society for Information Science and Technology*, 55(13): 1155-1168.
- Wallace, Danny P. 1985. "The use of statistical methods in library & information science." *Journal of the American Society for Information Science*, 38(8): 402-410.
- Water, Nancy Van de, Neil Surprenant, B. K. L. Genova, & Pauline Atherton. 1976. "Research in information science : An Assessment." *Information Processing & Management*, 12: 117-123.
- Wyllys, Ronald, E. 1978. "Teaching descriptive and inferential statistics in library schools." *Journal of Education for Librarianship*, 19(1): 3-20.