



Original Article

Analysis on Reports of Statistical Testings for Correlation and Regression

Cho, Dong Sook¹⁾ · Chung, ChaeWeon²⁾ · Kim, Jeungim³⁾
Ahn, Sukhee⁴⁾ · Park, SoMi⁵⁾ · Park, Hye Sook⁶⁾

1) Associate Professor, College of Nursing, Eulji University

2) Associate Professor, College of Nursing, Research Institute of Nursing Science, Seoul National University

3) Associate Professor, Department of Nursing, Soonchunhyang University

4) Assistant Professor, Department of Nursing, Chungnam National University

5) Associate Professor, Department of Nursing, Wonju College of Medicine, Yonsei University

6) Associate Professor, Department of Nursing, Catholic Sangji College

상관분석과 회귀분석을 이용한 논문의 통계활용 분석

조동숙¹⁾ · 정재원²⁾ · 김증임³⁾ · 안숙희⁴⁾ · 박소미⁵⁾ · 박혜숙⁶⁾

1) 을지대학교 간호대학 부교수, 2) 서울대학교 간호대학 부교수, 간호과학연구소 연구원

3) 순천향대학교 의과대학 간호학과 부교수, 4) 충남대학교 의과대학 간호학과 조교수

5) 연세대학교 원주의과대학 간호학과 부교수, 6) 가톨릭상지대학 간호과 부교수

Abstract

Purpose: This study aimed to examine the accuracy and adequacy of research papers reporting statistical testings for correlation and regression. **Method:** Original research articles utilized correlation and regression analysis were reviewed from the Korean Journal of Women Health Nursing published from the year 2004 to 2006. Thirty-six papers were evaluated in accordance with formatted criteria in respect to an inclusiveness of research title, accuracy of statistical methods and presentation styles, and errors in reporting statistical outcomes. **Result:** Thirty articles (83.3%) utilized Pearson's correlational analysis, and ten articles did regression analysis. Lack of accurate understanding and interpretation of the statistical method was a main fault. Basic assumptions and diagnostic testings for each statistical method were not

performed or described in most of the studies. Some points like consistency of research questions with statistical methods and criteria for sample size were still left out in part. Details of the presentation in the reporting of outcomes were not complied with the guidelines, which need careful concerns of the writers. Errors in English of result tables were found in more than one third of the tables. **Conclusion:** The outcome would be reflected in the submission guidelines for future writers. To reach the level comparable with internationally recognized nursing journals, concrete knowledge to apply statistical methods should be ensured in the processes of submission, reviews, and editing.

Key words : Correlation, Regression, Statistical analysis

투고일: 2008. 8. 20 1차심사완료일: 2008. 8. 28 2차심사완료일: 2008. 9. 1 최종심사완료일: 2008. 9. 10

• Address reprint requests to : Park, Hye Sook(Corresponding Author)

Department of Nursing, Catholic Sangji College

15 Sangjgil, Andong, Gyeongsangbuk-do, Korea

Tel: 82-54-851-3261 Fax: 82-54-853-3260 E-mail: hsp@csj.ac.kr

서론

연구를 수행하여 과학적이고 타당한 결론을 얻으려면 연구 설계와 적절한 변수 선정이 무엇보다 중요하다. 본 학술지가 학술진흥재단의 등재지가 되기 위해서는 여성건강간호라는 학술지 고유의 전문적 특성을 표방하고 여성, 건강, 간호 관련 주제에 대한 연구결과들을 독자들에게 신속, 정확하게 전달하여야 한다. 연구자들은 전문적인 여성건강간호에 대한 새롭고 과학적인 자료를 제시하기 위해서 변수들 간의 관계를 검증하고자 적절한 통계분석을 활용하게 된다. 또한 통계분석 결과를 제시함에 있어 독자들에게 연구자의 의도를 정확하게 전달하고자 결과표를 제시하고 있다.

독자는 연구자의 연구 의도와 연구 성과를 결과표에서 확인하게 된다. 연구자는 표 제목, 영문표기, 대상자 수, 통계수치 등을 일관성 있고 명료하게 제시하여 연구의 결과를 신뢰할 수 있도록 제곱할 책임이 있다. 즉, 결과표에 의해 연구자와 독자 간의 정보교류와 학문의 발전을 도모할 수 있다는 점에 유념하여야 한다. 본 여성건강간호학회지에 게재된 연구들은 대부분 양적연구가 차지함을 고려할 때 정확한 통계기법의 활용과 제시, 그리고 결과해석이 일관성 있게 적용되었는지에 따라 본 학회지의 질적 수준은 달라진다고 해도 과언이 아니다.

연구자는 연구를 시작하면서 대상자를 선정하고 연구 자료를 수집하게 된다. 수집된 연구 자료를 전산입력하여 오류가 없는 완전한 데이터베이스를 구축하고, 데이터베이스가 완성되면 연구가설을 검증할 수 있는 적절한 통계분석법을 적용하여 연구가설을 검증하게 된다(Korean Association of Medical Journal Editors, 2008). 연구가설을 검증하는 통계 기법의 선택이나 통계분석 결과를 제시하는 것은 연구자의 책임이지만 전문 학술지에 게재된 논문들은 동료로부터 심사과정을 거쳐 연구자가 범할 수 있는 오류와 논리적 비약을 최소화하게 된다. 본 여성건강간호학회지에서도 이러한 중요성을 인식하고 게재 논문 중 차이검정을 이용한 논문의 통계기법의 적절성 혹은 타당성의 문제에 초점을 두고 연구를 수행한 바 있다(Chung et al., 2007). 그리고 이전에도 간호 분야에서는 1986년-1995년까지 간호학회지에 게재된 논문을 대상으로 활용된 통계기법의 타당성을 평가한 연구(Lee & Lee, 1998), 국내 간호학 연구논문에서 사용된 측정도구의 신뢰도와 타당도 연구(Kim & Lee, 2000), 사회적 지지의 효과 모델 및 통계분석방법에 관한 국내 간호논문 분석(Lee & Kim, 2000)과 3년간(1999-2001) 대한간호학회지의 통계기법 활용에 대한 연구(Kang, 2002), 반복 측정 분산분석의 적용논문 분석(Min, 2005)연구 등이 계속되고 있었다.

통계방법을 효율적으로 이용할 수 있는 방안을 적극적으로

탐색하기 위해 통계적 기법을 사용하는 연구자는 올바른 기법의 선택과 적용 여부가 바로 연구의 성패를 가름한다는 사실을 명심해야 한다(Kang, 2002). 이러한 문제는 통계학의 측면에서만 아니라 여성건강 학문분야의 발전에 중요한 저해요인이 될 수 있다. 여성건강간호학회지 게재 논문에 적용된 통계방법들은 t-test, Chi-square test, Correlational analysis, ANOVA, ANCOVA, Regression analysis 등으로 기술통계학과 추론통계학을 사용하였고 분석목적, 분석하고자 하는 변수의 수, 표본 집단의 수, 측정된 자료의 형태 등에 따라 적절한 통계 분석법을 적용하고 있었다(Chung et al., 2007). 그러나 연구자와 심사자마다 연구방법론과 통계분석방법에 대한 지식과 견해가 다르고, 학회지마다 통계값 제시 방식도 다소간의 차이가 있기 때문에 어느 정도의 엄격한 질적 평가기준을 적용하여 심사하고 게재 여부를 결정할 것인지에 대한 다양한 평가가 있어 왔다. 실제로 논문의 심사과정에서 연구목적에 맞지 않는 통계기법을 이용하였거나, 사용한 통계기법에 대한 해석이 타당하지 않은 경우도 발견되고 있다. 또한 본 여성건강간호학회지도 투고 양식이 제시되어 있음에도 불구하고 연구결과를 제시하는 양상에 있어 세세한 부분, 예를 들면 표의 제시 방식, 영문 표현, 미미한 오류에 이르기까지 각 부분을 전체 논문이 통일되게 심사, 편집하는 것은 사실상 불가능하였다.

여성건강간호학회지는 2005년 한국학술진흥재단 등재후보지로 선정되었으며, 2008년 현재 등재지 평가를 신청한 상태이다. 그리고 논문의 영문초록은 여성건강간호학회 홈페이지(<http://women-health-nursing.or.kr>)를 통해 논문의 전문에 포함된 결과표와 함께 영문으로 온라인을 통해 전 세계 연구자들에게 공개되고 있다. 학회지의 질 향상을 위해 여성건강간호학회지에 최근 3년간(2004년-2006년) 게재된 논문의 통계분석방법의 활용을 분석하고자 연구 설계를 확인하였다. 110편의 논문 중 종설을 제외하고, 조사연구가 약 3분의 2이었으며, 22%는 유사 실험연구이었음을 보고한 바 있다(Kim, Park, Park, Chung, & Ahn, 2007). 그리고 2004년-2006년간의 논문에서 차이검정을 위해 t-test, Chi-square test, 그리고 ANOVA를 활용한 논문의 통계활용 분석(Chung et al., 2007)을 시행한 바 있다.

그러나 2004년부터 2006년까지 게재 논문 중 변수간의 상관관계를 확인하고 종속변수를 예측, 설명하는 통계기법에 대한 분석은 이루어지지 않았기에 국내 여성건강간호연구의 발전에 박차를 가할 현 시점에서 상관분석과 회귀분석 통계기법을 활용한 연구에 대한 평가가 필요하다고 생각되어 본 연구를 수행하게 되었다. 여성건강간호학회지가 국제적인 학술지로 발돋움하기 위해서는 여성, 건강, 간호의 국제적 교류 환경변화에 대처하고 본 학회지의 질적 수준을 향상시켜야

한다. 연구자들은 여성의 건강과 간호문제에 대한 이해를 좀 더 높일 수 있는 연구들이 더 많이 도출될 수 있도록 도움을 주고자 상관분석을 실시한 연구와 관련성이 있는 변수들 간의 관계를 확인하기 위해 회귀분석을 활용한 논문을 분석하여 본 학술지가 좀 더 질 높은 학술지로 발전하기 위한 추진력과 가이드라인을 제시하고자 한다.

연구의 목적

본 연구는 최근 3년간(2004년-2006년) 여성건강간호학회지에 발표된 논문을 대상으로 상관분석을 수행한 연구 36편을 선정하였다. 그 중 변수들 간의 관계를 설정하고 결과변수를 예측하고자 회귀분석을 수행한 10편의 연구결과의 통계적 기법 적용 및 결과 제시 양상을 분석하여 향후 여성건강간호학회지에 두고 시 올바른 통계분석방법 적용과 통계분석결과표의 해석방향을 제안하고자 함이다.

연구의 구체적인 목적은 다음과 같다.

- 연구제목, 연구목적의 적합성을 연구 대상자, 연구변수와 연구 설계, 그리고 연구 문제에서 반영하고 있는지를 분석한다.
- 제시한 결과표의 정확성과 일관성(표 제목, 영문표기, 대상자 수, 통계수치)을 분석한다.
- 회귀분석을 실시한 연구에서 독립변수들이 종속변수의 예측요인이라는 근거를 충분히 제시하고 있는지, 표본수가 독립변수의 수를 고려하고 있는지를 분석한다.

연구 방법

1. 연구 설계 및 연구 대상

본 연구는 여성건강간호학회지에 게재된 논문 중 상관분석을 수행한 연구논문 36편을 선정하여 상관분석 통계활용의 정확성을 분석하였고, 36편의 논문 중 회귀분석을 적용한 연구논문 10편의 통계활용의 정확성을 평가한 조사연구이다. 본 연구 대상 논문은 2004년 3월부터 2006년 12월까지 최근 3년간 여성건강간호학회지에 발표된 논문 110편으로 이 중 상관분석을 적용한 36편이었으며, 36편 중 회귀분석을 적용한 논문은 10편이었다.

2. 측정변수

- 1) 상관분석 적용 논문 분석을 위한 변수 선정
- 연구 설계의 유형, 연구목적의 수, 상관분석 표 수, 사용

분석방법 4개 항목

- 연구제목, 연구목적의 적합성 평가는 연구제목이 연구대상, 연구 변수, 연구 설계를 반영하고 있는지, 연구목적에 연구 변수가 기술되어 있는지 4개 항목
- 표 제시양식에서는 표 제목의 정확성, 표본크기(대상자수) 제시, 통계량과 p값 제시의 정확성을 평가한 4개 항목

2) 회귀분석 적용 논문 분석을 위한 변수 선정

- 연구문제가 회귀분석의 필요성과 독립변수들이 종속변수의 설명(예측)요인이라는 근거가 제시되고 있는지를 분석하기 위한 2개 항목
- 분석방법 조사는 연구목적이 연구 변수를 제시하는지, 표본수가 독립변수의 수에 적절한지, 사용한 회귀분석방법은 무엇인지, 회귀분석결과 제시한 표는 몇 개인지, 회귀분석을 위한 연구목적(연구가설)의 수는 몇 개인지 등의 5개 항목
- 통계방법 적용의 정확성은 연구목적에 따른 통계방법 선택을 확인하였는데 영향요인과 예측요인 파악 유무와 변수 선택 방법 유형을 보는 2개 항목
- 결과 해석에서는 회귀분석의 가정에 대한 검토, 회귀분석 전에 변수들 간의 상관관계가 제시되었는지 등 2개 항목
- 통계수치와 결과해석의 정확성은 R²값 제시, 회귀식의 유의성과 설명력 제시, 표준화 계수 b제시, 개별 독립변수의 유의성(p값) 제시와 상대적 중요성 확인, 다중회귀분석의 경우 각 회귀계수의 의미 제시 등의 7개 항목을 선정하였다.

3. 자료 수집 및 분석 절차

논문 분석을 위해 2인의 연구자가 분석틀을 일차 구성하였고 이의 타당도를 6인의 연구자들이 세 차례의 회의를 통해 전원 합의로 분석과 평가 기준을 결정하였다.

여성건강간호학회지 출판위원 6인이 확정된 분석과 평가 기준에 따라 해당 논문의 내용을 검토하여 체크하였으며 자료 분류에 의문이 생긴 경우는 다른 연구자들에게 사례에 대한 의견을 수합하여 모두 일관된 평가기준을 적용, 분석하도록 하였다. 또한 각 항목에서 구체적으로 지적된 사항 및 문체점은 그대로 서술하여 종합하였다.

수집한 자료는 Excel program을 이용하여 분석항목에 따라 빈도와 백분율로 분석하였다.

연구 결과

여성건강간호학회지에 2004년 3월부터 2006년 12월까지 게재된 논문 총 110편 중 상관분석을 수행한 논문은 36편이었다. 선정된 논문을 연구목적에 따라 분석한 결과는 다음과

같다. 한 편을 제외한 대부분의 연구논문(97%, n=35)이 연구 설계를 제시하였는데 상관분석을 수행한 논문의 연구 설계를 보면 조사연구(서술적, 비교)가 16편(44.4%)으로 가장 많았고, 상관성 조사연구 13편(36%), 실험연구 3편(8.3%), 예측연구 2편(5.6%), 이차분석연구 1편, 그리고 연구 설계를 제시하지 않은 논문이 1편 있었다.

1. 상관분석의 연구 제목 및 연구 목적의 적합성

연구자가 기술한 연구 제목이 연구 대상자, 연구 변수, 연구 설계를 반영하고 있는지 여부를 평가한 결과, 36편 모두 연구제목에 연구 대상자를 제시하였고(100%, n=36), 연구 변수를 반영한 연구는 91.6%(n=33)이었다. 또한 72.2%의 논문에서 연구 제목을 통해 연구 설계를 파악할 수 있었다. 연구문제 또는 연구목적의 수는 36편의 연구 중 3편의 논문을 제외한 33편이 상관분석을 수행한다고 기술하고 있었고, 연구문제는 각각 1개씩 설정한 것으로 분석되었다. 연구문제 및 목적에 대한 결과 제시에서는 28편(77.8%)이 1개의 표로 제시하였다. 상관관계를 수행한 연구 5편(13.9%)에서는 연구결과에서 상관관계 표를 제시하지 않았다. 상관분석에 사용된 통계 방법은 Pearson 상관분석이 83.3%, Spearman 상관분석은 11.1%이었는데 통계분석방법을 기술하지 않은 논문도 2편이 발견되어 추후 논문 심사 과정과 편집과정에서 주의가 요구된다고 하겠다(Table 1).

<Table 1> Inclusiveness of the research title and research questions (N=36)

Criteria	f	%
Does the title reflect the research subjects?		
Yes	36	100.0
Does the title reflect the research variables?		
Yes	33	91.6
Does the title reflect the research design?		
Yes	26	72.2
Are the research questions consistent with the correlation analytic method used?		
Yes	33	91.6
No. of tables for correlational analysis		
One	28	77.8
Two	2	5.6
Four	1	2.7
None	5	13.9
Method for correlational analysis		
Pearson	30	83.3
Spearman	4	11.1
Not mentioned	2	5.6

2. 상관분석 연구결과 표 제시 양식의 정확성

상관분석 연구결과 표 제시의 정확성은 표 제목과, 영문 제목표기, 대상자 수와 통계수치 제시 등을 분석하였다.

1) 표 제목의 정확성

결과표 제시 양식의 정확성 및 일관성을 평가하기 위해 36편의 연구논문에서 제시된 상관분석 결과표는 총 33개였다. 표 제목의 정확성에서는 2/3가 바르게 표기하였고(66.7%, n=22), 나머지 1/3(33.3%, n=11)은 표제목이나 변수표기 등의 영문표기에서 오류가 있는 것으로 나타났다(Table 2).

2) 통계수치의 정확성

총 대상자수는 N으로, 소집단 대상자 수는 n으로, 그리고 대상자 분류에 대한 빈도는 n 또는 f 로 표기하는 것이 원칙이다(APA, 2001). 전체 대상자 수(N)를 정확히 제시한 표는 전체의 63.6%에 불과하여 N 표기를 누락한 경우가 예상 외로 많음을 알 수 있었다. 결과표에서 변수들의 range, 평균과 표준편차를 제시한 표는 25개(75.8%)이었다(Table 2).

3) 통계량 및 통계적 유의성 제시

통계량을 제시한 표는 27개(81.8%)였다. p값 제시 형태는 실제 값으로 제시한 표가 19개(57.6%)로 가장 많았지만, range로 제시한 표도 2개(6.0%)가 있었다. 그리고 12개의 표들은 p의 실제 값과 * 표시를 동시에 제시하거나, 유의도의 range 없이 * 표시만 하거나 *p<0.01로 표시하는 등의 다양한 표현 양식을 나타내었다(Table 2).

<Table 2> Accuracy of the table presentation (N=33)

Criteria	f	%
Does the title of table reflect the study results?		
Yes	22	66.7
Any errors in English in title, variables, and classification?		
Yes	11	33.3
Reporting sample size(N)?		
Yes	21	63.6
Reporting range, mean and SD of the variables correctly?		
Yes	25	75.8
Present statistical values?		
Yes	27	81.8
Style of presenting p value		
Actual p value	19	57.6
Range of p	2	6.0
Etc. (*, **, p<0.01)	12	36.4

또한 사용된 연구변수나 변수 값에서 영어표기의 오류가 발견된 경우가 많았다. 한편 통계량을 제시할 때의 정확한 APA(2001) 양식은 통계량을 소수점 두 자리까지 표에 표기하도록 하고, 본문에 인용할 때에는 통계 값과 p값을 괄호를 통해 인용하되, 소문자 p로 표기하도록 되어 있으며, 모두 소수점 두 자리로 표기하도록 하고 있다. 그러나 통계량의 소수점 자릿수는 통일되지 않고 있어 이러한 오류들은 두고 규정에 상세하게 제시하여 향후 통일할 필요가 있다고 본다. 상기 언급한 내용들은 미미한 것으로 간과될 수 있으나 국제화의 일환으로 온라인을 통해 누구라도 우리 학회지 파일을 접할 수 있는 현재의 연구 데이터베이스 시스템을 감안할 때 사소한 영문 표기, 통계량 표기, p값 등의 오류만으로도 여성건강간호학회지 전체 연구물의 가치가 평가절하 될 수 있다는 점에서 연구자들과 심사자들이 유념해야 할 것으로 생각한다.

3. 회귀분석 활용 논문의 정확성 분석

회귀분석을 활용한 논문 10편을 분석한 결과에서는, 9편에서 연구제목이 연구변수를 제대로 반영하고 있었고, 8편의 연구제목이 그 연구의 설계(회귀분석)를 제시하였다. 9편의 논문이 연구문제에서 회귀분석의 필요성을 기술하였지만 연구자가 설정한 독립변수들이 종속변수의 예측요인이라는 이론적, 개념적 근거가 분명하게 제시된 논문은 6편에 불과하고 나머지 4편은 이론적 근거가 불충분하였다(Table 3).

연구목적에 따라 회귀분석에 사용된 분석방법은 다중회귀분석이 7편이었고 로지스틱분석이 2편이었다. 변수 입력을 위해 선택한 방법은 위계적 방법이 1편, 단계적 방법 6편이었고, 구체적 방법을 명시하지 않은 논문이 3편이었다. 회귀분석 결과를 단일 표로 제시한 논문이 7편이었고, 3편은 2개의 표로 제시하였다. 회귀분석을 위한 연구가설은 10편 모두 1개의 가설을 설정한 것으로 분석되었다(Table 3).

각 논문에서 활용된 독립변수의 수를 조사하였는데 독립변수의 수는 4개부터 11개까지로 차이가 많았다. 회귀분석의 특성상 독립변수의 수에 따라 연구대상자 수가 결정되므로 자료수집 전에 반드시 결정하여야 한다. 독립변수가 5개인 논문이 3편이었는데 동일한 수의 독립변수를 연구에서 선정하였다 하더라도 논문에서 선정한 연구 대상자 수의 범위가 컸다. 예를 들면 5개 독립변수에 대해 논문에 따라 연구 대상자 수가 63명부터 166명까지 큰 폭의 차이를 보였다. 향후 연구에서는 통계적 검정력의 이론적 근거를 강화하고, 측정변수에 따른 적정 표본 수 산출 근거를 제시하도록 제안함이 필요하겠다.

<Table 3> Analysis of research purposes and statistical methods (N=10)

Criteria	f	%
Does the table title reflect the variables analyzed?		
Yes	9	90.0
Does the table title reflect the research design?		
Yes	8	80.0
Does the research questions reflect needs for regression analysis?		
Yes	9	90.0
Presenting if the independent variables are the predictors of dependent variable?		
Yes	6	60.0
Analytical methods		
Simple linear regression	1	10.0
Multiple regression	7	70.0
Logistic regression	2	20.0
No. of tables for results		
One	7	70.0
Two	3	30.0
No. of hypothesis described		
One	10	100.0
Method for variable selection		
Hierarchical	1	10.0
Stepwise	6	60.0
Not noted	3	30.0
No. of independent variables		
Four	1	10.0
Five	3	30.0
Six	2	20.0
Ten	1	10.0
Eleven	2	20.0
Not noted	1	10.0

4. 회귀분석 연구결과표 분석

1) 결과 해석

논문에서는 연구가설 검정을 위해 통계분석을 하기 전에 변수들의 정규분포와 선형관계 등을 확인하는 것은 분석결과에 대한 신뢰성과 타당성을 높이기 위해 중요하다. 또한 결과 해석에서 회귀분석의 가정에 대한 검토가 이루어진 연구인지, 회귀분석 전에 변수들 간의 상관관계가 제시되었는지 여부를 확인하는 것도 중요하다. 10편의 논문에서 산점도와 잔차 그림 등을 통한 회귀진단 검토를 시행하여 기술한 논문은 2편이었고, 변수들 간의 상관관계를 제시한 논문은 5편이었다 (Table 4).

(Table 4) Testings for assumptions for regression analysis (N=10)

Criteria	f	%
Check scatter plots and residuals?		
Yes	2	20.0
Check correlations among variables?		
Yes	5	50.0

2) 통계수치 서술 및 해석의 정확성

회귀분석 결과표에서 통계량을 정확하게 제시하고 있는지를 검토한 결과 10편 모두 R²값을 제시하였고, 회귀식의 유의성은 9편에서, 회귀식의 설명력은 7편에서 제시되었고 표준화 계수인 beta값을 제시한 논문은 6편이었다. 개별 독립변수의 유의성(p값)은 9편이 제시하였고, 개별 독립변수의 상대적 중요성을 확인하여 제시한 논문은 7편이었다. 그리고 다중회귀분석의 경우 각 회귀계수의 의미를 명료하게 제시한 논문은 1편에 불과하였다(Table 5).

<Table 5> Accuracy of presenting statistical values(N=10)

Criteria	f	%
Reporting R ²		
Yes	10	100.0
Reporting F value for the significance of regression equation?		
Yes	9	90.0
Describing explanatory power of the regression equation?		
Yes	7	70.0
Presenting standardized b(beta)?		
Yes	6	60.0
Reporting p-value of each independent variable		
Yes	9	90.0
Reporting the relative significance of each independent variable?		
Yes	7	70.0

논 의

1. 상관분석활용 논문의 연구제목, 연구목적의 적합성

상관분석은 변수간의 관계를 알아보기 위한 방법으로, 탐색 연구에서 관련성의 존재 여부를 확인하거나 가설검증연구에서 특정관계에 대한 가설을 검증하기 위하여 사용된다. 따라서 통계기법 중 상관분석을 활용하는 연구 설계는 주로 조사연구와 상관관계연구, 예측설계연구이다. 연구자들이 검토한 연구에서도 상관분석은 대부분 조사연구와 상관관계연구에서 사용되었으며, 일부는 실험연구, 예측연구, 이차분석연구에서

활용되고 있었다. 상관분석을 수행할 필요가 있는지 여부는 연구목적과 연구의 필요성에서 언급하거나 선행 연구나 문헌적 근거를 제시하는 것이 원칙이지만 종종 인구학적 변수들 간의 또는 연구에 포함된 모든 변수들의 상관관계를 분석하여 다량의 표로 제시하는 점은 향후 연구에서 지양해야 할 점으로 본다. 동시에 논문 심사과정에서 논문심사자와 편집인의 보다 세심한 주의가 요구되며, 연구 설계에 맞는 통계분석 방법이 사용되고 있는지에 대한 지속적인 평가도 이루어져야 할 것이다.

바람직한 연구 제목은 연구자의 주요 아이디어가 간략하게 요약되어 제시되어야 한다. 대부분의 논문에서 연구대상자와 연구변수를 반영하고 있었으나 연구 설계 또는 분석의 방향이 반영되지 않는 경우도 28.8%나 되었다. 연구제목에 대상자와 주요 변수뿐만 아니라 연구 설계 또는 분석의 방향이 명확히 제시되어야 논문의 명확성과 함께 문헌의 활용도에도 유용할 것으로 본다. 연구제목에 연구의 필요성과 연구의 목적이 확실하게 나타날 수 있도록 연구 대상자와 연구 변수뿐만 아니라 연구 설계를 구체적으로 기술하여, 연구제목만 보아도 연구자의 의도와 연구의 개요를 알 수 있도록 하는 것이 필요하다고 하겠다. 또한 일부 논문에서는 여러 가지 연구 목적을 제시하였으나 적합한 통계기법을 모두 사용하지 않아 연구목적의 일부만 달성하거나, 전혀 통계분석방법을 기술하지 않고 분석에 따른 결과표도 제시하지 않아 연구의 일관성과 정확성이 문제되는 논문도 있었다. 따라서 논문의 질을 높이기 위해서 연구의 필요성과 목적에 맞는 통계기법의 사용과 결과표 제시, 그리고 포괄적이면서도 간략한 연구제목의 사용이 요구된다.

2. 상관분석 연구 결과표 제시 양식의 정확성

변수들 간의 관련성은 상관계수로 표시되며, 관련성의 강도를 판단하기 위해 p값을 고려해야 한다. 본 연구결과, 83.3%의 논문에서 변수들 간의 관련성은 Pearson 적률상관계수로 확인하였다. 이는 간호학 연구들에서 사용되는 변수가 순위척도나 비율척도보다는 등간척도로 점수화되는 경우가 많기 때문으로 생각된다. 또한 이미 표준화되었거나 널리 활용되고 있는 도구를 사용하는 연구가 많고, 측정된 값을 연구 논문 간에 상호 비교하게 되면서 두 변수 사이의 관계를 상관계수로 수량화하는 일반적 방법을 활용하고 있다. 상관계수 r을 계산하기 위해서는 한 대상자로부터 최소 두 개의 변수가 측정수준이 비율척도 이상이어야 하고, 상관계수 값은 변수간의 상관정도를 비교하는 값으로 간주된다.

상관분석에서 통계적 유의수준은 표본 수에 의해 좌우된다. 표본이 큰 경우에는 “아주 작은 관계”만 있어도 통계적으로

유의하다. 그러나 표본 수에 따라 “매우 유의한” 상관관계이지만 상관계수가 상당히 작을 수 있으므로, 상관계수 r 의 의미에 대해 표본 수를 고려하여 생각해 봐야 한다. 양측검정에서 α 0.05, 보통효과크기 0.30을 기준으로 할 때 0.80의 검정력을 위해 84명의 대상자수가 필요하지만 단측검정에서는 동일한 검정력을 위해 68명의 대상자가 요구된다(Lee, Chung, Kim, Song, & Hwang, 2006). 표본 수가 적을 경우 검정력이 떨어질 수 있으므로, 유의한 결과를 얻기 위해 적절한 표본수를 확보하는 것은 매우 중요하다. 그러나 본 연구에서 분석한 대부분의 상관분석논문에서는 표본 수 선정기준에 대한 언급을 제시하지 않고 있다. 향후 연구에서는 기본 가정 확인과 표본 수 선정기준에 대한 표기를 하도록 하여, 이를 기초로 정확한 결과 해석이 이루어지도록 유도하여 연구결과에 대한 신뢰도를 높여 나가야 할 것이다.

Cohen(1988)은 상관계수 $r=0.10$ 을 작은 효과, $r=0.30$ 을 보통 효과, 그리고 $r=0.50$ 을 큰 효과로 정의하고 있다. 그러나 연구자들이 분석한 논문에서는 상관계수 r 값을 제시하거나, 유의수준을 제시하고 ‘상관관계가 있다’라고만 제시하고 있어, 연구결과를 기술할 때 효과의 정도를 함께 표시하도록 권장할 필요가 있다고 본다. 또한 양측검정으로 100명의 표본일 때 $r=0.20$ 은 0.05수준에서 통계적으로 유의하지만, 10명의 표본인 경우 상관관계는 0.63이상이어야 유의해진다는 점에 유의하여 양측검정과 단측검정에 대한 고려도 검토할 필요가 있다.

상관분석연구에서 표본의 통계로부터 결과를 일반화하기 위해서는 관련되는 변수들 즉 X 와 Y 는 반드시 정규 분포하여야 하며, 변수의 점수는 정규곡선에 근사하여야 하고, X 의 모든 값에 대한 잔차의 분포는 동일한 변이성(등분산성)을 가져야 하며, X 와 Y 의 관계는 선형이어야 한다는 등의 기본적인 변수들의 조건이 갖추어져야 하는데(Lee et al., 2006), 검토한 연구논문들 중 대부분이 각 변수들이 갖추어야 할 조건들을 상세하게 명시하지 않고 있었다.

많은 간호연구가 양적 연구이고 상관분석이 일반적으로 널리 활용되는 상황이므로 이에 대한 통계분석 방법 사용 시 기본 가정에 대한 연구자들의 민감도가 향상된다면 간호학 연구를 통해 수집된 자료를 보다 잘 이해하고 활용하며 나아가 간호 현상에 대한 바른 이해가 이루어질 것으로 생각된다.

상관분석 결과표를 검토하여 표 제시 양식의 정확성과 일관성을 평가한 결과는 전체적으로 만족할만한 수준이 아니었다. 표 제목에 있어 33.3%가 표제목이나 변수표기, 영문표기에서 오류를 나타내었다. 오류의 예로는 서로 다른 표에 같은 표제목이 제시되고 있는 경우, 통계방법을 표 제목으로 표기한 경우, 상관관계 결과표에 변수명이 level로 표기된 경우, 영문표기에서 어휘나 표현 등이 적절하지 못한 경우, 결과표에서 전체 대상자수(N), 변수들의 range, 평균, 표준편차를 정

확하게 제시하지 않은 경우, 그리고 통계량(r)을 표기하지 않은 경우 등이 예외로 많았다.

본 학회지 투고요령에는 표의 수치는 일반적으로 2자리 이상의 평균값은 소수점 한자리, 통계량은 가능한 소수점 2자리로 기록하며, p 값은 *표시 혹은 p 값 중 한 가지로 기록하도록 간략하게 제시되어 있다. 자세한 표 작성 요령이 제시되지 않음에 따라 연구자간에 서로 다른 표기를 야기할 수 있으며, 이는 사소한 표기상의 문제로 인해 독자들로 하여금 여성건강간호학회지의 연구결과물의 질적 수준을 저평가하게 할 수 있다. 그러므로 표 작성을 위한 세부 지침을 개정하여 논문투고규정에 삽입할 필요가 있다.

3. 회귀분석 활용 논문의 정확성 분석

회귀분석은 몇 가지 알려진 증거를 바탕으로 어떤 알려지지 않은 미래의 사건을 예측할 수 있도록 하는 통계방법이다. 즉, 회귀분석은 결과변수를 예측하고 변수들 간의 상호관계를 설명하도록 해주는 유용한 분석기법이다.

본 연구에서 회귀분석을 활용한 논문을 분석한 결과, 연구자가 설정한 독립변수들이 종속변수의 예측요인이라는 이론적, 개념적 근거가 부족한 편이었다. 또한 회귀분석을 수행하기 위하여 요구되는 표본 수는 최소한 독립변수 1개당 10명의 대상자가 확보되어야 함(Lee et al., 2006)에도 불구하고 독립변수 수 제시를 누락한 논문도 있었으며, 대상자수 산출근거를 독립변수 수에 따라 제시한 논문은 찾아보기 힘들었다.

따라서 앞으로 연구결과물의 질 확보를 위해, 연구자는 설정한 독립변수들이 종속변수의 예측요인이라는 분명한 이론적, 개념적 근거, 표본 수 산출근거를 제시하고, 심사자와 편집자는 이를 확인하여야 할 것으로 사료된다.

4. 회귀분석 연구 결과표 분석

회귀분석 연구결과를 바르게 해석하기 위해, 회귀통계분석을 위한 가정의 진단은 중요하며, 이때 중요한 요소는 잔차(residual analysis)이다. 잔차는 실제 점수와 예측점수간의 차이이며 분석이 완벽하다면 잔차는 없게 되어 모두 0이 된다.

잔차에 대한 중요한 가정은 선형성(linearity), 정규성(normality), 등분산성(homoscedasticity), 다중공선성(multicollinearity) 등이다. 두 변수의 관계가 선형이고 종속변수가 독립변수의 개개 값에 정규분포하고 있다면 잔차의 분포는 정규곡선에 근접하여야 한다. 연구자들은 표준화 잔차의 분포를 히스토그램을 통해 쉽게 점검할 수 있으므로 이를 분석과정에서 확인하고 결과에 반드시 언급하도록 노력해야 한다. 등분산성이란 잔차의 분산이 독립변수의 모든 수준에서 일정해야 한다는

것이다. 회귀진단을 위해 예측 값과 독립변수들에 대한 잔차 값을 좌표에 점으로 표시함으로써 등분산성 가정 여부를 진단할 수 있다. 등분산성 평가방법은 회귀표준화 잔차와 회귀표준화예측 값에 대한 산점도를 통해 평가한다. 마지막으로 다중공선성(multicollinearity)을 보이지 않아야 함은 독립변수 간 상관이 낮아야 함을 의미한다. 왜냐하면 독립변수 간 상관성이 지나치게 높으면 자체 공유분산이 높아 종속변수를 설명하는 개별 변수의 변량을 해석하기가 모호해지기 때문이다. 다중공선성 평가방법은 독립변수 간 상관관계 값 확인, 변수의 공차(tolerance), 분산팽창계수(Variance Inflation Factor: VIF)를 통해 평가한다. 본 연구결과 회귀분석 전에 변수들 간의 상관관계가 표로 제시된 경우는 50%, 산점도와 잔차 그림 등을 통한 회귀 진단 검토를 확인한 경우는 20%에 불과하였다. 이는 회귀분석의 기본 가정을 충실히 진단하지 않은 채 단순히 나타난 연구 결과를 해석함으로써 해석상의 오류를 가져올 수 있다. 그러므로 향후 회귀분석 연구결과에서는 잔차 확인에 대한 언급을 반드시 하도록 하는 것이 바람직할 것이다.

정확한 회귀분석 회귀식의 유의성은 F 검정을 통해 확인하고 회귀식의 설명력은 R^2 또는 adjusted R^2 (보정 R^2)로 평가한다. 그러나 R^2 는 대상자의 수와 변수의 수를 고려하지 않은 상태로 계산된 설명력이기 때문에 대상자의 수와 변수의 수를 고려하여 축소공식에 의해 재계산된 값인 Adjusted R^2 를 읽는 것이 바람직하다. 유의성 검정은 단순 선형회귀분석에서는 두 변수 간 상관관계의 유의성을 검정하며 r^2 은 상관관계의 크기에 대한 의미를 나타낸다. 다중상관에서는 전체적인 R의 유의성이나 종속변수의 분산 중 회귀식으로 설명되는 변이(R^2)의 양도 살펴보지만, 동시에 각 독립변수의 유의성도 살펴볼 필요가 있다. R^2 이 유의하다고 해서 모든 독립변수들이 종속변수의 변이를 설명하는데 유의하게 기여하는 것은 아니다. 이를 위해 앞으로 회귀분석을 활용하는 연구자들은 다중 상관이 유의한지 검정함과 동시에 각각의 회귀계수 b의 유의성도 검정하는 원칙적인 단계를 제시할 것을 제언한다. 이는 회귀계수 b의 유의성 검정을 통해 해당 독립변수가 종속변수의 변이를 설명하는데 얼마나 유의하게 기여했는지를 알 수 있기 때문이다.

본 연구 결과 회귀분석결과표에서 R^2 값과 회귀식의 유의성은 비교적 정확하게 제시하고 있는 반면에 회귀식의 설명력, 표준화 계수 b값, 개별 독립변수의 중요성 확인, 다중회귀분석의 경우 각 회귀계수의 의미를 명료하게 제시되어 있지 않았다. 이는 독자에게 연구자의 해석이 바르게 이루어지고 있는지를 확인하기 어렵게 만든다. 따라서 연구자는 통계분석방법에 대한 기본 가정을 충분히 이해하고, 통계수치를 정확히 해석할 수 있는 능력을 갖추는 한편 독자가 바르게 이해할

수 있도록 꼭 필요한 결과표는 제시해야 할 것이다.

결론 및 제언

본 연구는 여성건강간호학회지에 투고된 논문 중 상관분석과 회귀분석을 적용한 논문을 대상으로 통계활용의 정확성과 문제점을 분석하였다. 이러한 시도는 그동안 투고된 논문의 심사와 편집 과정에서 지속적으로 반복된 문제점에 대해 객관적인 검증절차를 거치고자 함이었으며 이를 통해 추후 본 학회지에 투고하고자 하는 연구자들에게 보다 분명하고 정확한 가이드를 제시하고자 함이다. 각 학회지마다 투고규정이 있지만 투고자들의 세심한 검토가 부족하며 세부적이고 구체적인 내용들이 일부 규정에 제시되지 않아 투고자 및 심사자들에게 혼돈을 주고 있음을 알 수 있었다. 이러한 점들은 국내 학술지가 국외 학술지에 비해 취약한 부분들이며 국내 학회지가 출판 전문 인력의 도움 없이 대부분 자체 소속 회원으로 구성된 인력 가동 범위 내에서 모든 과정을 해결하기 때문이 아닌가 여겨진다. 따라서 연구 및 출판에 대한 연구자들의 의욕과 양적인 확대 뿐 아니라 과학적이고 체계적이어야 하는 학술지를 통한 연구 환경의 발전을 위해 보다 구체적인 노력과 지원이 이루어져야 할 시점이라고 본다. 본 연구자들은 연구결과를 투고 및 출판과정에 반영하여 투고규정을 보완하고 연구자들에게 구체적인 정보로 제공함으로써 본 학회지의 질적 수준 향상과 연구자들의 연구역량 증대에 기여할 것으로 기대한다.

References

- APA. (2001). *Publication manual of the American psychological association*. (5th ed.). Washington DC: American Psychological Association.
- Chung, C., Kim, J., Park, H., Ahn, S., Cho, D., & Park, S. (2007). Analysis on reports of statistical testings for mean differences. *Korean J Women Health Nurs*, 13(3), 211-218.
- Cohen, J. (1988). *Statistical power analysis for the behavioral science*. (2nd ed.). New Jersey: Lawrence Erlbaum Association Pub.
- Kang, H. (2002). 최근 3년간(1999-2001) 대한간호학회의 통계기법활용에 대한 연구. *J Korean Acad Nurs*, 32(6), 929-935.
- Kim, E., & Lee, S. (2000). An assessment of reliability and validity of measuring instruments used in nursing researches in Korea from 1996 to 1997. *J Korean Soc Health Inf Health Statistics*, 25(1), 11-26.
- Kim, J., Park, S., Park, H., Chung, C., & Ahn, S. (2007). Analysis of published papers and their keywords in the Korean Journal of Women's Health Nursing. *Korean J Women Health Nurs*, 13(1), 51-59.

- Korean Association of Medical Journal Editors (2008). *Good publication practice guidelines for medical journals*. Seoul: KAMJE, 19.
- Lee, E., Chung, Y., Kim, J., Song, R., & Hwang, K. (2006). *Statistical methods for health care research*. Seoul: Koonja. 253-313.
- Lee, E., & Kim, J. (2000). Major effect models of social support and its statistical methods in Korean nursing research. *J Korean Acad Nurs*, 30(6), 1503-1520.
- Lee, S., & Lee, S. (1998). An assessment of statistical validity of nursing researches published in Korea from 1986 to 1995. *J Korean Soc Health Information Health Statistics*, 23(1), 42-64.
- Min, H. S. (2005). A statistical analysis of nursing research using repeated measures analysis of variance. *Chungnam Journal of Nursing Academy*, 8(1), 60-66.