

간호행정학회지 게재논문의 통계학적 방법 사용과 오류

송 기 준¹

¹ 연세대학교 의과대학 의학통계학과

Use and Misuse of Statistical Methods in the Journal of Korean Academy of Nursing Administration

Song, Kijun¹

¹ Department of Biostatistics, College of Medicine, Yonsei University

주요어

통계학적 방법, 통계학적 오류

Key words

Statistical methods,
Statistical errors

Correspondence

Song, Kijun
Department of Biostatistics,
Yonsei University College of
Medicine
134, Shinchon-dong,
Seodaemun-gu, Seoul 120-752,
Korea
Tel: 82-2-2228-2497
Fax: 82-2-364-8037
E-mail: biostat@yuhs.ac

투 고 일: 2012년 12월 17일
수 정 일: 2013년 1월 5일
심사완료일: 2013년 1월 9일

Abstract

Purpose: To do nursing research effectively requires an understanding of fundamental principles of statistical methods. In this article, some key statistical methods which are commonly used in nursing research are identified and summarized. **Methods:** Ninety-two original articles from the Journal of Korean Academy of Nursing Administration were reviewed. Statistical methods were classified and summarized for usage in research and occurrence of common errors. **Results:** Among the original articles reviewed, 58 statistical usages contained errors. Most errors were found in linear regression analysis, Pearson correlation analysis, and chi-square test. From the detection of statistical errors in usage, suggestions for appropriate statistical methods were made. **Conclusion:** In order to improve validity of original articles in the Journal of Korean Academy of Nursing Administration, clearly stated statistical usage and close editorial attention to statistical methods are needed. Understanding statistical methods is part of the process that researchers must use to determine both quality and usefulness of the research. Research findings will be used to guide nursing practice and reduce uncertainty in decision making. However, to understand how to interpret research results, it is important to be able to understand basic statistical concepts. Researchers should also choose statistical methods that match their purposes.

서 론

최근 간호학 연구에서는 수집된 환자 자료의 정보가 방대해지고 분석내용이 복잡해짐에 따라, 이를 분석하고 실제에 적용할 수 있는 적절한 연구설계 방법과 그에 따른 다양한 통계학적 분석 방법이 요구되고 있다. 국내외 대부분의 간호학 학술지에서는 연구자가 주장하는 내용이 통계학적 유의성을 통해 검증되었는지와, 이 과정에서 사용된 통계학적 방법이 타당한지가 연구

논문의 게재 여부를 결정짓는 중요한 요소가 되고 있다. 이에 따라 연구를 수행할 때 적절한 연구설계 방법을 적용하는 것은 필수적인 사항으로 인식되고 있고, 또한 연구를 통해 도출된 결론을 일반화시키기 위해 연구 목적에 부합되는 통계학적 방법을 적용할 필요성이 점차 높아지고 있다. 현재 간호학 연구에서는 통계학적 방법의 사용빈도가 높아지고 적절한 분석 방법의 적용에 대한 필요성이 증대됨에 따라, 통계학자들과의 공동연구를 통해 통계학적 방법의 사용범위를 확대해 나가고 있다. 그러나

아직까지는 인력 및 여건의 부족으로 인하여 통계학적 분석 방법에 대한 이해가 부족하거나, 부적절한 분석 방법을 적용한다거나, 혹은 분석 결과에 대한 정확한 해석이 이루어지지 않는 등, 연구목적에 맞는 연구결과를 도출해내지 못하는 경우가 많다.

본 논문에서는 간호행정학회지에 게재된 논문들의 통계학적 사용방법을 정리하고 아울러 그 때 발생했던 통계학적 오류사항을 정리함과 동시에 올바른 통계학적 방법의 적용 기준을 요약하여 제시하고자 한다. 이를 통하여 연구자들이 연구를 진행함에 있어 통계학적 분석 방법을 적절하게 적용하여 좋은 연구 성과를 얻는데 기여하고자 한다.

방 법

1. 연구 대상

2010년 16권 1호부터 2011년 17권 4호까지 2년간 간호행정학

회지에 게재된 총 92편의 논문을 조사하였다.

2. 자료수집 및 분석방법

게재된 논문에 사용된 통계학적 방법을 정리하여 그 사용도에 대해 조사하였으며 이를 토대로 연구자가 흔히 범하기 쉬운 사용 오류, 해석 오류 등을 정리하였다. 또한 간호학 논문에서 주로 사용되는 통계학적 방법들을 정리하여 그 사용방법과 기준을 제시하였다.

결 과

1. 통계학적 분석 방법의 분류

Table 1은 일반적으로 간호학 연구에 주로 사용되는 통계학적 방법들을 구분한 것이다. 이 방법들을 각 범주로 묶어서 요약

Table 1. Common Statistical Methods Used in Nursing Research

Category	Analysis method	Use
t-test	Independent t-test	독립적인 두 집단의 평균 비교
	Paired t-test	짝을 이룬 자료의 평균 비교
ANOVA	One-way ANOVA	세 개 이상의 독립적인 집단의 평균 비교
	Multiple comparison or post-hoc test	ANOVA 결과가 유의한 경우 특정 집단들의 평균을 개별 비교할 경우
	ANCOVA	통제할 변수를 보정하고 평균을 비교할 경우
	Repeated measures ANOVA	반복측정된 자료의 변화 비교
Nonparametric statistical analysis	Wilcoxon rank sum test (Mann-Whitney U test)	Independent t-test의 비모수적 방법
	Wilcoxon signed-rank test	Paired t-test의 비모수적 방법
	Kruskal-Wallis test	One-way ANOVA의 비모수적 방법
Categorical data analysis	Pearson chi-square test	r×c 형태로 구성된 분할표에서 두 변수간의 독립성 혹은 동질성을 검정할 경우
	Fisher's exact test	분할표에서 적어도 하나의 cell에 대한 기대빈도가 5 미만일 경우 chi-square test 대신 사용
	Mantel-Haenszel chi-square test	순위형 자료로 이루어진 분할표에서 추세 혹은 경향성을 검정할 경우
	McNemar test	이분형자료를 두 번 반복해서 측정했을 때 변화의 유의성을 분석할 경우
	Cochran-Mantel-Haenszel test	두 개의 범주형 변수간의 연관성을 파악할 때 제 3의 변수를 통제해서 분석할 경우
	Kappa index	복수의 관찰자 혹은 관찰방법간의 일치도를 평가할 경우
Correlation analysis	Pearson correlation	두 개의 연속형 변수간의 선형적 상관성을 파악할 경우
	Spearman correlation	Pearson correlation analysis의 비모수적 방법
Regression	Linear regression	종속변수가 연속형인 자료의 회귀분석
	Logistic regression	종속변수가 이분형인 자료의 회귀분석
Multivariate data analysis	Principal component analysis	서로 연관이 있는 변수들을 이용하여 보다 적은 수의 새로운 변수들을 생성하는 경우
	Cluster analysis	유사한 속성을 가진 대상들을 묶어 일정한 수의 집단으로 나누고자 하는 경우
	Factor analysis	관찰된 변수에 근거하여 직접 관찰할 수 없는 요인을 파악하고자 할 경우
	Canonical correlation analysis	두 개 이상의 변수집단간의 상관성을 동시에 분석하고자 할 경우
	Structural equation modelling	변수들 간의 다중의 인과관계 및 상관관계의 추정이 필요할 경우
	Path analysis	변수들 간의 인과관계를 모형을 통해 규명하고자 할 경우

정리하면 다음과 같다.

1) t-검정 (t-test)

일반적으로 t-test라고 하면 독립된 두 집단의 평균을 비교하기 위한 방법을 말하는데, 이 경우에는 구체적으로 independent t-test (혹은 two sample t-test)라는 용어가 적절하다. 이 때 두 집단의 분산이 서로 같은지 다른지에 따라서 검정통계량이 달라 지므로 사전에 분산의 동등성 여부를 검정하여야 한다. 이에 비해 paired t-test는 동일한 대상에 대해 전, 후의 차이를 비교하는 방법으로 전, 후 차이 값의 평균이 통계학적으로 유의한 지 검정하는 방법이다. t-검정을 하기 위해서는 비교하고자 하는 연속형 변수가 정규분포를 따른다는 가정이 필요하다. t-검정 결과를 논문에 제시할 때 각 집단(혹은 전, 후)의 평균과 표준편차만을 제시하는 경우가 대부분인데 이외에 평균차이에 대한 95% 신뢰구간을 제시해주는 것이 정밀도(precision)를 평가하는데 도움이 된다.

2) 분산분석 (ANOVA; Analysis of Variance)

셋 이상의 독립적인 집단 평균에 차이가 있는가를 검정하고자 할 경우에 사용되며, 이 때 비교하는 연속형 변수는 정규분포를 따른다는 가정과 독립적인 집단들의 모분산은 모두 동일하다는 가정이 필요하다. 분산분석에서 측정값에 영향을 미친다고 생각되는 효과를 요인 혹은 인자(factor)라고 하는데, 요인의 수가 하나인 경우를 일요인 분산분석(one-way ANOVA), 둘인 경우를 이요인 분산분석(two-way ANOVA)이라고 한다. 분산분석의 결과가 통계학적으로 유의한 경우 특정 집단 간의 비교를 할 수 있는데, 이 절차를 다중비교(multiple comparison) 혹은 사후검정(post-hoc test)이라고 한다. 다중비교의 방법으로는 주로 Bonferroni 방법, Tukey 방법, 혹은 Scheffé 방법 등이 쓰인다. 공분산분석(ANCOVA; Analysis of Covariance)은 기저변수(예를 들면, 연령과 같은 환자의 기초변수 등)를 통제하는 분산분석법인데, 이 때 분석 결과로 각 집단의 산술평균과 표준편차를 제시하는 것 보다는 최소제곱평균(least square mean)과 표준오차(standard error)를 제시하는 것이 적절하다. 반복측정자료의 분산분석(repeated measures ANOVA)은 어떤 실험을 함에 있어 동일한 실험대상에 대하여 실험 조건을 달리하거나 또는 여러 다른 시점에서 반복적으로 특정한 값을 측정하는 경우에 시점간의 차이 혹은 집단 간의 차이를 비교할 때 사용한다. 반복측정 자료의 분산분석을 적용하기 위해서는 반복적으로 얻어진 관찰치들 간의 상관관계가 복합대칭성(compound symmetry)을 만족한다는 가정이 필요하다.

3) 비모수 통계분석 (Nonparametric Statistical Analysis)

관찰된 자료가 정규분포를 따른다는 가정을 하기 어려운 경우에 적용할 수 있는 방법이다. 비모수 통계분석 결과로 제시하는 대표값은 평균보다는 중위수가 적절하고 퍼짐의 척도 또한 표준편차보다는 범위(range)나 사분위범위(IQR; Inter-quartile Range)가 더 적절하다. 윌콕슨 순위합 검정(Wilcoxon rank sum test)은 두 개의 표본이 동일한 중위수를 갖는지 검정하기 위한 방법으로 모수적 통계분석방법 중 independent t-test에 대응되는 비모수적 방법이다. 일반적으로 윌콕슨 순위합 검정과 맨-휘트니 U 검정(Mann-Whitney U test)은 같은 결과를 제공한다. 윌콕슨 부호순위 검정(Wilcoxon signed-rank test)은 모수적 통계분석 방법인 paired t-test에 대응되는 분석방법이다. 크루스칼-왈리스 검정(Kruskal-Wallis test)은 세 개 이상의 집단의 중위수가 동일한지 검정하기 위한 방법으로 모수적 통계분석 중 일요인 분산분석에 대응되는 비모수적 검정방법이다. 크루스칼-왈리스 검정의 원리에 적합하게 개발된 사후검정 방법으로는 Dunn (1964)의 방법 등이 있다.

4) 범주형 자료분석 (Categorical Data Analysis)

분석하고자 하는 자료가 연속형 자료가 아닌 이산형 자료인 경우에 사용한다. Pearson chi-square test가 대표적인 방법인데, 이는 $r \times c$ 형태의 분할표(contingency table)에서 두 변수간의 독립성(independency)이나 동질성(homogeneity)을 알아보고자 할 때 사용된다. 이 경우 주의할 점은 분할표에서 적어도 하나의 cell 이상의 기대빈도(expected frequency)가 5 미만일 경우에는 Pearson chi-square test보다는 Fisher's exact test를 적용하는 것이 적절하다는 것이다. 또한 $r \times c$ 형태의 자료가 순위형(ordinal) 변수일 경우 그 경향성(trend)을 검정하고자 할 때에는 Pearson chi-square test를 그대로 적용시키기보다는 순위형 변수의 증가 혹은 감소 추세를 검정하는 방법으로 Mantel-Haenszel chi-square test가 적용될 수 있다. McNemar test는 동일 환자에 대해 반복적으로 이분형 변수를 관찰하였을 때 그 변화의 유의성을 분석하는 방법인데, 이때에는 이분형 변수인 경우에서만 사용된다. Cochran-Mantel-Haenszel test는 두 개의 범주형 변수의 연관성을 검정할 때, 이 두 변수에 영향을 줄 수 있는 제 3의 변수를 통제할 필요가 있을 때 적용된다. Kappa 계수(Kappa index)는 둘 이상의 진단 방법 혹은 검사자간의 일치도(agreement)를 판단하기 위하여 사용된다. 그 일치도의 정도를 0부터 1사이의 값으로 판단하는데, 이 값이 0.75보다 크면 매우 일치도가 높은 것으로 평가하고 0.4이하이면 일치도가 낮은 것으로, 0.4에서 0.75사이인 경우는 보통 정도의 일치도라고 평가한다.

5) 상관분석 (Correlation Analysis)

두 개의 연속형 변수들 간에 선형적(linear) 상관성이 존재하는지 알아보기 위한 방법이다. 상관관계의 척도로서는 Pearson이 제시한 상관계수(correlation coefficient)가 대표적이다. Pearson 상관계수를 이용한 상관분석은 분석하고자 하는 변수가 정규분포를 따른다는 가정이 있어야 한다. 상관분석에서 주의할 점은 몇 가지가 있는데, 첫째 상관분석의 결과를 인과관계(causality)로 설명하는 것은 과도한 해석이라는 것이다. 둘째 상관분석 결과 통계학적으로 유의하지 않았다고 해서 두 변수 간에 어떠한 관계도 없다고 결론 내리는 것은 위험하다는 것이다. 이는 두 변수 간에 선형적 상관성 이외에 다른 형태의 상관관계가 존재할 수 있기 때문이다. 마지막으로 상관분석 결과를 두 변수간의 질적인 관계로 확대 해석하면 안된다는 것이다. 어떠한 사건의 이론적 근거없이 상관계수가 크다고 해서 그 관계를 질적인 관계로 해석하는 것은 위험한 발상이기 때문이다. 만약 측정 자료가 정규분포를 따른다고 가정하기 어려운 경우에는 Pearson 상관분석의 비모수적인 방법인 Spearman 상관분석(Spearman correlation analysis)을 사용할 수 있다.

6) 회귀분석(Regression Analysis)

독립변수(independent variable)와 종속변수(dependent variable) 간의 인과관계를 규명하고자 할 때 사용되는 방법이 회귀분석인데, 종속변수가 연속형 변수일 때 적용되는 방법이 선형회귀분석(linear regression)이고 이분형(binary) 변수일 경우 적용되는 방법이 로지스틱 회귀분석(logistic regression)이다. 선형회귀분석을 적용하기 위해서는 종속변수 혹은 회귀모형의 오차(error)가 정규분포를 따르고 서로 독립이며 하나의 분산을 가져야 한다는 세 가지의 기본 가정이 필요하다. 선형회귀분석에서는 독립변수에 대한 회귀계수(regression coefficient)를 최소제곱추정법(least square estimation)을 이용하여 추정하는데, 이 회귀계수의 통계학적 유의성을 검정하여 독립변수가 종속변수에 영향을 미치는지 확인한다. 분석 후에 기본적으로 제시해야할 통계량은 회귀계수와 그에 따른 표준오차이다. 아울러 회귀계수에 대한 신뢰구간 혹은 p-value를 제시해야한다. 연속형 독립변수에 대한 회귀계수의 해석은 기울기(slope)의 의미로 해석이 가능한데, 즉 독립변수가 한 단위 증가할 때 종속변수가 증가 혹은 감소하는 폭을 의미한다. 이에 비해 범주형 독립변수의 경우에는 가변수(dummy variable)의 형태로 분석에 포함시키는데, 이 때 가변수에 대한 회귀계수는 기울기의 의미가 아니라 참조범주(reference category)와 가변수의 범주에 해당되는 종속변수의 평균적 차이를 말한다.

로지스틱 회귀분석에서는 연속형 독립변수와 종속변수의 관심

있는 범주가 발생할 확률 간에 0을 중심으로 S자 형태의 대칭이 된다는 가정이 필요하다. 독립변수가 종속변수에 미치는 영향은 주로 비차비(odds ratio)로써 표현되며, 분석 결과에서는 비차비와 그에 대한 95% 신뢰구간을 같이 표현해주는 것이 적절하다. 가변수화된 독립변수의 비차비는 참조범주에 비해 가변수의 범주일 때 종속변수의 특정 범주가 발생할 확률에 대한 배수로서 직접적인 해석이 가능하지만, 연속형 독립변수의 비차비는 독립변수가 한 단위 증가할 때 종속변수의 특정 범주가 발생할 odds의 증가 혹은 감소분으로 해석해야 한다.

두 개 이상의 독립변수를 동시에 고려하는 다중 회귀분석(multiple regression analysis)을 적용할 때 주의할 점은 독립변수들 사이에 상관관계(다중공선성; multi-collinearity)가 높아서는 안 된다는 것이다. 상관관계가 높은 독립변수들이 동시에 분석에 포함될 경우 해당 독립변수의 회귀계수 추정치에 대한 표준오차가 커져 통계학적 유의성이 떨어지는 결과를 초래할 수 있다. 이는 로지스틱 회귀분석에서도 비슷하게 주의해야 한다. 다중 회귀분석 혹은 다중 로지스틱 회귀분석은 종속변수에 대한 독립변수들의 효과를 보정(adjust)한 후 유의한 인자를 찾고자 할 때 많이 적용되는데, 이 때 예측 모형을 구축하는 것과는 구분해서 이용해야 한다. 예측 모형을 구축하는 것이 목적이라면 분석 결과에 구축된 예측 모형의 성능(performance)과 관련한 지표들을 반드시 제시해야 하며 아울러 모형의 타당성(validation)을 평가 해주어야 한다.

7) 다변량 통계분석(Multivariate Statistical Analysis)

다변량분석은 두 개 이상의 (종속)변수들을 관찰하였을 때 변수간의 관계를 동시에 분석할 수 있는 통계학적 기법을 일컫는 말이다. 다변량분석의 영역에 속하는 분석 방법으로는 주성분분석(principal component analysis), 요인분석(factor analysis), 군집분석(cluster analysis), 판별분석(discriminant analysis), 정준상관분석(canonical correlation analysis), 대응분석(correspondence analysis), 다차원척도법(multidimensional scaling) 등이 있다.

주성분분석은 서로 연관 있는 여러 변수들이 측정되었을 때 이 변수들이 가지고 있는 정보를 최대한 활용하여 보다 적은 수의 새로운 변수들을 생성하는 방법이다. 이를 위하여 주성분분석은 변수들을 선형변환(linear transform)시켜 주성분이라고 부르는 서로 상관되어 있지 않은, 혹은 독립적인 새로운 가공의 변수들을 유도한다. 이 때 각 주성분이 차지하는 변동량(variation)의 크기를 기준으로 원래 자료에 내재해 있는 전체 변동량 중 가능한 많은 부분을 보유하도록 변환시킴으로서 정보의 손실을 최소화하는 변수의 축소를 기할 수 있게 한다. 특히 변수의 축소를 인하여 얻어지는 주성분 점수(principal component

score)들은 다음 단계의 통계분석을 위한 입력자료로 이용되어 일련의 분석과정에서 중간단계의 역할을 하기도 한다.

요인분석은 서로 상관있는 여러 변수들에서 이 변수들 간의 관계를 공통요인(내재적 차원)을 이용하여 밀접한 상관관계를 보이는 변수들끼리 묶어 줌으로써 그 관계를 보다 간결하고 명확하게 해주는 방법이다. 여기에서 공통요인이라는 것은 변수들이 구조적 측면에서 서로 공유하고 있는 확률변수로서, 변수들 간의 상관관계를 생성시키는 가설적이고 관찰할 수 없는 요인을 의미한다. 요인분석에서 요인을 줄이는 기준으로 흔히 아이겐 값(eigen value)이 사용되며 추약되어 생성된 변수의 설명력이 얼마인지는 각 요인의 설명된 분산의 크기로 파악된다. 일반적으로 요인분석은 연구의 목적에 따라 탐색적(exploratory) 접근 방법과 확증적(confirmatory) 방법으로 구분할 수 있다. 전자가 변수들과 공통요인 간에 내재하는 상관관계를 탐구하는 데 그 목적이 있다면, 후자는 공통요인의 개수나 요인들 간의 관계 등 모형에 관한 구체적 가정들을 실제자료를 이용하여 확인하는데 그 초점을 맞추고 있다.

군집분석은 연구 대상들에게서 서로 다른 속성을 가진 변수들을 관찰했을 때, 서로 유사한 속성을 갖는 대상들을 묶어 전체 대상을 일정한 수의 군집(cluster)으로 나누는 분석 방법이다. 이렇게 나누어진 군집은 유사한 성격을 갖는 특성을 보이기 때문에 각 군집에 대한 내적 특징을 파악할 수 있다. 군집의 유형은 상호배반적 군집, 계층적(hierarchical) 군집, 중복 군집, 퍼지(fuzzy) 군집 등으로 나누어 볼 수 있다.

판별분석은 연구대상이 두 개 혹은 그 이상의 집단으로 나누어져 있을 때, 이 집단을 가장 잘 분류할 수 있는 방법을 조사된 변수들을 이용하여 표현하는 분석 방법이다. 즉, 관찰된 변수들을 이용하여 상호배반적인 부분집단으로 분류하기 위해 기준이 되는 판별함수(discriminant function)를 구축하고, 집단이 구분되지 않은 새로운 대상을 판별 기준에 따라 특정한 집단으로 분류하는 과정을 거친다. 실제 최종 판별 결과는 정분류율(classification rate)과 오분류율(misclassification rate)로 표현된다.

앞서 기술한 상관분석은 두 변수간의 선형적 상관성을 파악하는 방법인데 비해, 정준상관분석은 두 개 이상의 변수집단간의 상관성을 동시에 분석하기 위한 방법이다. 이 때 변수집단간의 상관관계는 정준상관계수를 이용하여 측정할 수 있는데, 이는 상관계수의 개념을 일반화 시킨 것으로 볼 수 있다.

8) 구조방정식 모형(Structural Equation Modelling)을 이용한 분석

구조방정식 모형을 이용한 분석은 다차원의 구조로 얻어진 변수들 간에 다중의 인과관계 및 상관관계의 추정이 필요할 때 사

용된다. 구조방정식을 이용하여 직접 변수로 측정되지 못한 단일 개념 혹은 복합 개념에 대하여 추정할 수 있으며, 이 때 추정 과정에서 발생할 수 있는 오차에 대한 설명이 가능하다. 경로분석(path analysis)은 변수들 간의 인과관계를 경로 모형도(path diagram)를 통해 규명하는 방법이다. 구조방정식 모형을 이용한 분석에서 주의할 점은 가능한 최소의 변수만을 이용하여 모형을 구축하는 것이 적절하며, 아울러 분석 결과에 최종 모형의 적합도(goodness of fit)에 관한 각종 통계량들을 반드시 제시해야 한다는 것이다.

2. 간호행정학회지에 사용된 통계학적 분석 방법의 적용 유형

본 논문에서 검토된 원저들에 사용된 통계학적 방법은 총 239건이며 그 빈도는 Table 2와 같다. 사용된 방법을 구체적으로 살펴보면 independent t-test가 53건으로 가장 많이 사용되었고, 그 다음으로 one-way ANOVA 및 Pearson correlation analysis가 많이 사용되었다. 아울러 linear regression analysis가 그 뒤를 이었고 Pearson chi-square test가 다섯 번째로 빈번하게 사용된 방법이었다. 나머지 분석 방법들은 모두 5건 이하로 위의 다섯 가지 분석방법들이 주로 사용됨을 알 수 있었다.

오류 건수를 살펴보면 linear regression analysis에서 가장 많

Table 2. Uses and Errors of Statistical Methods in the Journal of Korean Academy of Nursing Administration

Analysis method	Frequency	Error
Independent t-test	53	2
Paired t-test	1	0
Pearson chi-square test	13	6
Fisher's exact test	4	1
Kappa index	1	0
One-way ANOVA	47	2
Repeated measures ANOVA	2	1
ANCOVA	4	5
Pearson correlation analysis	47	7
Linear regression analysis	37	25
Logistic regression analysis	5	5
Principal component analysis	4	0
Factor analysis	1	0
Cluster analysis	1	0
Canonical correlation analysis	1	0
Structural equation modelling	5	1
Spearman correlation analysis	4	0
Wilcoxon rank sum test (Mann-Whitney U test)	5	1
Kruskal Wallis test	4	2
Total	239	58

은 오류를 보였으며, Pearson correlation analysis가 그 다음으로 많았고, Pearson chi-square test가 그 뒤를 이었다. 나머지의 경우는 모두 5건 이하의 오류를 보였다.

3. 간호행정학회지에 사용된 통계학적 방법의 오류 사례

Table 3은 2년간 조사한 간호행정학회지에 게재된 논문의 통계학적 오류의 사례를 정리한 것이다. 가장 많은 빈도를 차지하고 있는 오류 사례는 linear regression analysis인데, 여기에서 특히 많은 오류를 보였던 예는 다음과 같았다. 먼저, 분석에 포함된 독립변수를 선택한 근거를 제시하지 않은 경우가 많았고, 회귀모형의 기본 가정에 대한 검토를 전혀 언급하지 않은 경우도 다수였다. 아울러 보정이 필요한 상황에서 변수를 보정하지 않고 분석한 경우도 있었으며, 구태여 적용할 필요가 없는 변수

선택 방법을 사용하기도 하였다. 독립변수들 간의 다중공선성의 문제를 고려하지 않은 경우도 있었고, 추정된 회귀계수의 표준오차를 제시하지 않은 경우도 다수 존재했다. 또한 특별한 이유 없이 표준화된 회귀계수만을 제시하는 경우도 있었는데, 일반적으로 회귀분석 결과에 반드시 제시해야 하는 회귀계수 추정치는 scale을 보정하지 않은, 표준화되지 않은 추정치이다. 표준화된 추정치는 필요에 따라 제시하는 것이 적절하다. 아울러 표에 회귀계수 추정치를 표기할 때 표준화되지 않은 회귀계수는 'B'로 표준화된 회귀계수는 ' β '라고 표기를 하는 경우가 대부분이었는데, 이는 특정 통계 software에서 제공하는 표기 방식이지 통계학 분야에서 쓰이는 공식적인 표현법이 아니므로 이를 그대로 적용하는 것은 적절하지 않다. 그리고 linear regression에 포함시킬 독립변수의 적절한 개수를 고려하지 않은 경우도 존재했다. Pearson correlation analysis의 경우에서 벌어진 오류의 대

Table 3. Type of Errors for Statistical Methods in the Journal of Korean Academy of Nursing Administration

Analysis method	Type of error
Independent t-test	Paired data에 independent 자료에 이용하는 방법 적용
Pearson chi-square test	Fisher's exact test를 사용해야할 상황임에도 Pearson chi-square test 사용 Paired data에 independent 자료에 이용하는 방법 적용 순위형 변수에 대한 분석결과, 경향성에 대한 결론 제시
Fisher's exact test	Exact test를 수행했는데, chi-square 통계량 값 제시
One-way ANOVA	동일대상에 대하여 반복적으로 측정된 자료에 one-way ANOVA 사용 사후 검정 결과에 더 큰 의미 부여
Repeated measures ANOVA	분석결과에 대한 부적절한 해석
ANCOVA	불필요한 통계량을 과다하게 제시 부적절한 통계량 제시
Pearson correlation analysis	범주형 변수에 대하여 Pearson 상관분석을 수행 분석결과를 인과관계인 것처럼 확대 해석
Linear regression analysis	회귀모형의 기본 가정에 대한 검토가 없음 분석에 포함된 독립변수 선택의 근거가 없음 변수를 보정하지 않고 분석 특별한 이유 없이 표준화된 회귀계수만을 제시 다중공선성의 문제를 고려하지 않음 회귀계수를 제시하지 않음 표준오차를 제시하지 않음 종속변수가 삼분형 변수인데 선형회귀분석을 적용함 독립변수와 종속변수의 표기를 혼동하여 사용 불필요한 변수선택 방법 사용
Logistic regression analysis	부적절한 독립변수의 개수 연속형 독립변수의 비치비에 대한 잘못된 해석 독립변수들 간의 상관성을 고려하지 않음 독립변수들 간의 상관성을 파악하는데 부적절한 척도 사용
Structural equation modelling	해석의 오류
Wilcoxon rank sum test (Mann-Whitney U test)	부적절한 대표값 제시 Paired data에 independent 자료에 이용하는 방법을 적용
Kruskal Wallis test	부적절한 대표값 제시 부적절한 사후검정 방법 적용

부분은 분석 결과를 마치 인과관계인 것처럼 확대 해석하는 경우였다.

Pearson chi-square test를 적용한 경우 발생한 오류들로는 Fisher's exact test를 사용해야할 상황임에도 Pearson chi-square test를 그대로 사용한 경우가 있고, 자료가 paired data인데 독립적인 자료인 것처럼 분석한 경우가 있었다. 또한 순위형 변수로 이루어진 자료에 Pearson chi-square test를 사용한 후, 경향성에 대한 결론을 내리는 경우도 있었다.

ANCOVA의 경우에는 부적절한 통계량을 제시하는 경우가 많았다. 즉, ANCOVA를 수행하는 경우에는 일반적으로 집단별 최소제곱평균(least square mean)과 표준오차를 제시하는 것이 타당한데, 그냥 산술평균과 표준편차를 제시하는 경우가 대부분이었고 아울러 이러한 통계량을 제시하지 않고 다른 통계량을 과다하게 제시하는 경우도 있었다.

Logistic regression analysis를 사용했을 때 오류의 사례로는 먼저, 분석에 포함시킨 독립변수의 개수가 부적절한 경우로서 종속변수의 분포에 비해 과도하게 많은 독립변수를 포함시킨 상황이었다. 일반적으로 multiple logistic regression을 수행할 때 적절한 독립변수의 개수는 종속변수의 관심 있는 사건의 빈도를 10으로 나눈 수 보다 작게 설정하는 것이 적절하다. 또한 연속형 독립변수의 비차비에 대하여 '몇 배' 등의 개념으로 잘못된 해석을 하기도 하였다. 연속형 독립변수의 비차비는 독립변수가 한 단위 증가할 때 종속변수의 특정 범주가 발생할 odds의 증가 혹은 감소분으로 해석하는 것이 타당하다. 아울러 logistic regression에서도 독립변수들 간의 상관성을 고려하여 분석하는 것이 적절한데 이를 전혀 고려하지 않고 분석이 이루어진 경우도 있었다. 그런데, 독립변수들 간의 상관성을 파악할 때 linear regression에서 쓰이는 척도, 예를 들어 분산확대인자(VIF; variance inflation factor)와 같은 개념을 사용하여 평가한 경우도 있었는데, 이는 logistic regression에서는 정의되지 않는 척도이므로 적절하지 않다.

ANOVA를 사용한 경우의 오류 사례는 동일대상에 대하여 반복적으로 측정된 자료임에도 이를 독립적으로 간주하여 그냥 one-way ANOVA를 적용시킨 경우가 있었다. 또한, ANOVA 결과보다 사후 검정 결과에 더 큰 의미를 부여하여 해석하는 사례도 있었다. ANOVA의 경우, 우선적으로 의미를 부여해야할 결과는 ANOVA를 수행했을 때 F-검정 결과이지 다중 비교와 같은 사후 검정 결과가 아니다. 사후 검정은 부가적으로 수행되는 분석일 뿐이다.

Independent t-test를 적용한 경우 paired data인데 independent t-test를 적용하는 사례가 있었고, structural equation modelling에서는 변수들 간에 역의 관계가 존재할 수 있는 상황을 무시하

고 해석하는 경우가 있었다. Wilcoxon rank sum test를 적용한 논문들에서는 분석방법에 어울리지 않는 대표값을 제시하거나 paired data에 독립적인 자료에 이용하는 방법을 적용하기도 하였다. Kruskal Wallis test에서는 one-way ANOVA에 쓰이는 사후검정 방법을 그대로 적용한 경우가 있었다.

Table 3에 제시되어 있지는 않았지만 논문에 통계분석 방법을 기술하는 과정에서 범하는 오류는 다음과 같은 것들이 있었다. 먼저, 가장 많은 경우의 오류로서 적용된 통계분석방법 일부를 기술하지 않거나, 논문에 실제 적용되지 않은 방법을 기술하는 경우가 있었다. 또한 통계분석방법의 명칭(용어)을 틀리게 기술하거나, 부적절한 용어를 기술한 경우도 있었다. 그 밖의 사항으로는 분석에 이용된 통계 프로그램 (혹은 software)을 기술하지 않은 경우가 있었다. 아울러 총 92건의 논문 중 통계학적 의사결정의 원칙을 기술한 경우는 단 세 건의 논문 밖에 없었다. 통계학적 의사결정의 원칙을 기술하는 것은 연구자가 어느 정도의 유의수준(significance level)에서 통계학적 유의성을 평가할 것인지를 정의하는 것인데, 이는 통계분석 결과에 대한 평가를 가능하게 하는 잣대를 정하는 것이므로 통계분석을 하는 경우라면 반드시 기술해 주어야할 중요한 요소이다. 통상 암묵적으로 5%의 유의수준을 많이 적용한다고 하여 이를 생략하는 것은 적절하지 못하고 공식적으로 기술해줄 필요가 있다. 일반적으로 통계학적 방법이 적용되는 학술논문에서 반드시 기술되어야할 세 가지 요소가 첫째, 분석에 사용된 통계분석 방법의 명칭, 둘째, 통계분석에 실제 이용된 통계 프로그램, 마지막으로 통계학적 의사결정의 원칙이다.

기술통계량 표현에서 범하는 오류 중에는 분석할 자료에 부적절한 기술통계량(평균, 표준편차 등)을 제시하는 경우가 있었다. 다시 말하면 자료를 요약하여 나타내는 기술통계량은 자료의 중앙(center)을 나타내주는 지표와 자료의 퍼짐의 정도를 나타내는 지표로 구분할 수 있다. 중앙을 표현하는데 쓰이는 통계량으로는 평균(mean), 중위수(median) 등이 있는데, 중위수는 자료에 극단적으로 크거나 작은 값이 존재하는 경우 혹은 자료의 분포가 왜곡되어(skewed) 있는 경우처럼, 평균을 그 대표값으로 쓰기에 부적절한 경우 사용된다. 퍼짐의 정도를 나타내는 통계량으로는 분산(variance), 범위(range) 등의 통계량이 상황에 따라 적절히 활용되는데, 분산은 자료의 변동이 평균을 중심으로 얼마나 퍼져 있는지를 나타내는 지표이고 여기에 제곱근을 취한 것이 표준편차(standard deviation)이다. 범위는 평균을 쓰기 부적절한 상황에서 중위수를 쓰는 것처럼 표준편차에 대체적으로 쓰이는 지표이다. 이러한 원칙에 근거하지 않고 기술통계량을 제시하는 경우가 있었다.

기타의 오류로서는 분석 결과 table에 p-value를 제시하지 않

은 경우가 있었고, p-value를 통계프로그램에 나온 표기법 그대로 '0.000', '0.00' 등으로 표기한 경우도 있었다. 이 경우에는 '(0.001)', '0.01' 등으로 표기하는 것이 적절하다.

분석 결과와 결론을 기술할 때 나타난 오류의 사례로서는 먼저, 통계학적으로 유의하지 않은(차이를 보이지 않은) 결과에 대하여 '비슷하다' 혹은 '동일하다' 등으로 해석하는 경우가 있었는데, 이러한 해석을 내리기 위해서는 자료에서 통계학적 검정력(power)을 계산했을 때 일정 값(통상 80%) 이상이 얻어졌을 때 타당하다. 만약 그렇지 않은 경우 그냥 '통계학적으로 유의한 차이가 없었다.' 라는 식의 해석만이 가능하다. 또한 연구에 적용된 설계 방법이나 통계분석 결과는 빈약한데 비해 얻어진 결과에 대해 지나치게 확대 해석하는 경우도 존재했다.

논 의

지금까지 간호행정학 연구에서 기본적으로 적용되는 통계분석 방법 및 올바른 사용 방법에 대하여 개괄적으로 소개하였다. 아울러 2010년 16권 1호부터 2011년 17권 4호까지 2년간 간호행정학회지에 게재된 총 92편의 논문을 조사하여 논문에 사용된 통계학적 방법의 사용 현황과 그 오류 등을 정리하였다. 간호행정학회지에 게재된 대부분의 논문이 통계학적 분석 방법을 사용하고 있었으며 그 결과를 통해 연구자가 주장하고자 하는 결론에 도달하고 있었다. 하지만 일부 논문에서 통계학적 방법의 부적절한 적용이나 미비한 해석으로 연구의 성과를 완벽하게 얻어 내지 못한 점은 아쉬운 부분이라 하지 않을 수 없다. 앞서 언급한 오류들은 통계분석 방법의 사용 용도와 해당 분석 방법의 적용에 대한 정확한 인식이 부족하거나 혹은 의심 없이 관습적으로 사용하던 행위에서 기인한 것으로 보여진다. 그러나 통계학을 전공하지 않은 연구자의 입장에서 모든 통계학적 방법 및 그에 따른 특성들을 정확히 파악하는 것이 결코 쉬운 일이 아님을 주지의 사실이다.

본 연구를 통해 제시한 통계학적 분석 방법들의 적용 기준을 통해 연구자가 자료의 특성을 정확히 파악하여 연구목적에 부합되는 통계분석 방법을 사용한다면 연구결과에 대해서 더 정확한 해석을 할 수 있으며 그에 따른 올바른 결론을 내릴 수 있을 것이다.

결 론

대량으로 얻어지는 전산화된 정보와 가속적으로 발전하는 통계분석 패키지 등을 통해 편리하게 자료를 얻고 분석할 수는 있지만, 부적절한 통계분석 방법의 적용은 연구의 결과를 왜곡하

고 아울러 연구의 가치를 떨어뜨리게 할 수 있다. 또한 통계분석 방법은 적절하게 적용하였으나 그 결과를 올바르게 해석하지 못한다면 좋은 연구 결과를 도출해낼 수 없을 것이다. 본 연구를 통해 통계학적 분석 방법들을 소개하고 그 사용 오류를 정리하는 것은 연구자가 자료의 분석과 해석을 하는데 있어 조금이나마 도움이 되게 하고자 하는 것이다. 또한 이러한 일련의 검토와 지적은 향후 간호행정학회지의 발전과 탁월한 연구 성과의 도출을 위해서도 그 일조를 할 것이라고 판단한다.

REFERENCES

- Agresti, A. (2002). *Categorical data analysis* (2nd. ed.). New York: John Wiley & Sons.
- Altman, D. G. (1998). Statistical reviewing for medical journals. *Statistics in Medicine*, 17, 2661-2674. [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19981215\)17:23<2661::AID-SIM33>3.0.CO;2-B](http://dx.doi.org/10.1002/(SICI)1097-0258(19981215)17:23<2661::AID-SIM33>3.0.CO;2-B)
- Altman, D. G. (2000). Statistics in medical journals: Some recent trends. *Statistics in Medicine*, 19, 3275-3289. [http://dx.doi.org/10.1002/1097-0258\(20001215\)19:23<3275::AID-SIM626>3.0.CO;2-M](http://dx.doi.org/10.1002/1097-0258(20001215)19:23<3275::AID-SIM626>3.0.CO;2-M)
- Altman, D. G., Goodman, S. N., & Schroter, S. (2002). How statistical expertise is used in medical research. *Journal of the American Medical Association*, 287, 2817-2820. <http://dx.doi.org/10.1001/jama.287.21.2817>
- Altman, D. G. (1994). The scandal of poor medical research. *British Medical Journal*, 308, 283-284. <http://dx.doi.org/10.1136/bmj.308.6924.283>
- Altman, D. G. (2002). Poor-quality medical research: What can journals do? *Journal of the American Medical Association*, 287, 2765-2767. <http://dx.doi.org/10.1001/jama.287.21.2765>
- Armitage, P., & Colton, T. (1998). *Encyclopedia of biostatistics*. New York: John Wiley & Sons.
- Bailar, J. C., & Mosteller, F. (1988). Guidelines for statistical reporting in articles for medical journals: Amplifications and explanations. *Annals of Internal Medicine*, 108, 266-273.
- Dillon, W. R., & Goldsrein, M. (1984). *Multivariate analysis*. New York: John Wiley & Sons.
- Emerson, J. D., & Colditz, G. A. (1983). Statistics in practice. *New England Journal of Medicine*, 309, 709-713. <http://dx.doi.org/10.1056/NEJM198309223091206>
- Fisher, V. B., & Lumley, H. (2004). *Biostatistics: A methodology for the health sciences* (2nd. ed.). New York: John Wiley & Sons.
- Forthofer, R. N., Lee, E. S., & Hernandez, M. (2007). *Biostatistics: A guide to design, analysis, and discovery*. New York: ELSEVIER.
- Hair, J. F., & Anderson, R. E. (1995). *Multivariate data analysis* (4th. ed.). New Jersey: Prentice Hall.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applies logistic regression* (2nd. ed.). New York: John Wiley & Sons.

- Lachin, J. M. (2000). *Biostatistical methods: The assessment of relative risks*. New York: John Wiley & Sons.
- Landis, J. R., & Koch G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- MacArthur, R. D., & Jackson, G. G. (1984). An evaluation of the use of statistical methodology in the Journal of Infectious Disease. *Journal of Infectious Diseases*, 149, 349-354. <http://dx.doi.org/10.1093/infdis/149.3.349>
- Redmond, C. K., & Colton, T. (2001). *Biostatistics in clinical research*. New York: John Wiley & Sons.
- Rennie, D. (1998). Freedom and responsibility in medical publication: Setting the balancing right. *Journal of the American Medical Association*, 280, 300-302. <http://dx.doi.org/10.1001/jama.280.3.300>
- Rosner, B. (2011). *Fundamentals of biostatistics* (7th. ed.). Pacific Grove: Duxbury.
- Shoukri, M. M., & Pause, M. M. (1998). *Statistical methods for health sciences* (2nd. ed.). New York: CRC Press.
- Sterne, J. A., & Davey, S. G. (2001). Sifting the evidence: What's wrong with significance tests? *British Medical Journal*, 322, 226-231.