



국내 커뮤니케이션 연구의 통계분석방법의 현황과 문제점

지난 10년간 <한국언론학보> 게재 논문의 내용분석

이병관 한양대학교 광고홍보학과 교수
김재민 한양대학교 광고홍보학과 석사과정
김주환 한양대학교 광고홍보학과 석사과정
장다연 한양대학교 광고홍보학과 석사과정
권나현 한양대학교 광고홍보학과 석사과정

Current Status and Problems of Statistical Analysis in Communication Research

Content Analysis of Studies in the Korean Journal of Journalism & Communication Studies over the Past 10 Years*

Byoungkwan Lee**

Professor, Dept. of AD & PR, Hanyang University

Jaemin Kim***

Graduate student, Dept. of AD & PR, Hanyang University

Juhwan Kim****

Graduate student, Dept. of AD & PR, Hanyang University

Dayeon Jang*****

Graduate student, Dept. of AD & PR, Hanyang University

Nahyun Gwon*****

Graduate student, Dept. of AD & PR, Hanyang University

Although statistical analysis methods are essential tools for quantitative researchers, controversy has persisted over the objectivity of inference through statistical procedures. The purpose of this study is

* This study was supported by the Hanyang University(HY-2019-G)(본 연구는 2019년 한양대학교 교내연구비 지원으로 연구되었음(HY-2019-G)).

** gogreen@hanyang.ac.kr

*** jmini369@naver.com

**** feelth1147@naver.com

***** jdy6237@hanyang.ac.kr, corresponding author

***** nahyun0224@naver.com

to review the use and trends of statistical analysis methods in quantitative research papers published between 2010 to 2019 in the Korean Journal of Journalism and Communication Studies and to discuss its practices and problems of statistical use in communication research in Korea. For this purpose, 612 quantitative research papers using statistical analysis methods were content-analyzed out of a total of 845 papers published across 10 years. In this 612 quantitative research, the basic characteristics of the study, as well as the reliability between coders, the number of coders, and power for experimental design, were reviewed according to the methodology used. This study also checked for the use of null hypotheses and multi-item measures, and more specific details including the statistical packages, the basic assumptions of statistical analysis methods used in individual studies, such as t-test and regression, and the description of statistical results. Research papers were collected, reviewed, and analyzed by four graduate students majoring in advertising and public relations. The inter-coder reliability was measured by Krippendorff alpha, and the reliability for each item was between 0.71 and 1, thus ensuring a stable level of reliability. As a result of the analysis, research using surveys was the most common among the research methods, and regression was the most frequently used statistical method except for descriptive statistics. The most commonly used statistical package was SPSS. The current study found problems of over-reliance on a specific statistical package, erroneous interpretation of statistical analysis results, and misuse of statistical methods for yielding contrived results. In particular, over-reliance on a specific statistical package was related to over-reliance of specific statistical analysis methods such as partial eta squared and cronbach alpha. At the same time, in most studies, wrong practices were also found. For example, important information regarding the processes and results of statistical analysis was not provided, such as basic statistical assumptions and correlation between major variables, and neither was information on confidence intervals as supplementary indicators of null hypothesis significance testing. Given that this study empirically identified the criticisms and discussions that have been continuously raised about statistical analysis methods and reporting, this study holds implications. Through this study, it is hoped that future Korean communication researchers will actively carry out research on various problems and alternatives raised in individual statistical analysis methods, thereby enriching academic discussions.

Keywords: Statistical analysis, Quantitative research, Communication research, Content analysis

“어떤 연구자는 통계를 슬주정뱅이가 가로등을 붙을 밝히는 도구가 아닌 몸을 지탱하는 도구로 사용하는 것처럼 이용한다.”

- 앤드류 랭(Andrew Lang, 1990: Pelsma, 1937, p. 72 재인용)

1. 들어가는 말

통계분석방법은 자신의 연구결과를 입증하려는 양적연구자들에게 있어 필수적 도구임에도 불구하고, 통계적 절차를 통한 추론의 객관성에 대한 논란은 계속되어 왔다(이병관·손영곤·오현정·박선화, 2015). 이러한 논란에는 ‘통계의 오용(misusing statistics)’이 그 중심에 서 있다. 통계의 오용은 ‘왜곡 혹은 인위적인 결과를 생산하기 위해 통계적 방법, 테크닉, 모델을 사용하는 것’과 ‘통계적 방법에 관한 중요한 정보를 연구자에게 공개하지 않는 것’ 등 크게 두 가지 유형으로 구분될 수 있다(Gardenier & Resnik, 2002). 이는 통계의 사용에 있어 연구자의 의도적 혹은 비의도적인 잘못된 관행에서 비롯된다.

이러한 관행은 통계학을 전공으로 하지 않고 통계학자들에 의해 개발된 통계분석 알고리즘을 가져다 사용해야만 하는 대부분의 응용과학 분야, 특히 사회과학 연구자들에게는 현실적 문제이다. 예를 들면, 피셔(Fisher, 1925)와 노이만과 피어슨(Neyman & Pearson, 1928) 등에 의해 고안된 이후 거의 100년 가까이 사회과학 연구자들에게 신화처럼 사용되어 온 영가설 유의도 검증(null hypothesis significance testing: 이하 NHST)은 그동안 많은 비판과 대안이 제기되었음에도 불구하고(예: Bakan, 1966; Boster, 2002; Cohen, 1994; Hunter, 1997; Levine, Weber, Hullett, Park, & Lindsey, 2008; Nickerson, 2000; Schmidt & Hunter, 1997) 여전히 견재하다. 크론바흐(Cronbach)의 신뢰도 계수인 α 역시 1951년 그의 최초 논문(Cronbach, 1951)이 *Psychometrika*에 게재된 이후 개념의 혼돈과 방법론적·통계적 결합에 대한 비판에도 불구하고(예: Borsboom, 2006; Shevlin, Miles, Davies, & Walker, 2000; Sijtsma, 2009; Yang & Green, 2011) 사회과학 분야에서 독보적인 사용을 보여 왔다. 실제로 크론바흐가 1951년 최초로 게재한 논문 “Coefficient alpha and the internal structure of tests”의 게재이후 2004년까지 적어도 5590회의 인용이 이루어졌고(Cronbach & Shavelson, 2004), 최근까지도 연간 약 325회의 사회과학 분야의 인용 지수(social science citations index)를 보이는 것으로 나타났다(이병관·손영곤·강경희, 2018). 국내의 경우, 1998년 창간호부터 2014년 여름호까지 <홍보학연구>에 게재된 총 317편의 논문 중 통계분석방법이 사용된 전체 논문을 대상으로 통계분석방법을 메타 분석한 연구(이

병관 외, 2015)에 따르면, 신뢰도 검증이 필요한 다중문항 측정 도구를 사용한 143편의 논문 중 신뢰도를 보고하지 않은 논문 18편을 제외한 125편의 98.4%(123편)가 α 를 신뢰도 지수로 사용한 것으로 나타났다.

최근 이러한 무비판적인 통계분석방법의 사용에 대한 자성의 움직임이 목격되고 있다. 예를 들면, 많은 문헌들이 위에서 언급한 '통계적 방법에 관한 중요한 정보를 연구자에게 공개하지 않는 것'을 비판하고 이에 대한 가이드라인을 제시해왔다(예: Bailar & Mosteller, 1988; Boster, 2002; Wilkinson & APA Task Force on Statistical Inference, 1999). 이 가이드라인에는 분석에 사용된 방법론적 절차(예: 연구 설계, 샘플링, 측정, 검정력(Power) 등)나 기술 통계 결과(예: 표본오차, 데이터의 분포나 변동량)를 상세히 기술하지 않는다거나, 필요한 통계적 가정(예: 정규성, 동분산성 혹은 구형성 등)을 검증하지 않으며, NHST의 대안 지표(예: 효과크기, 신뢰구간 등)를 보고하지 않고, 메타 연구를 가능하게 하는 통계치(예: 이변량 상관관계 행렬)를 보고하지 않는 등의 문제가 비판의 대상이 되어 왔다.

본 연구의 목적은 국내 커뮤니케이션 연구에 있어 지금까지의 통계분석방법의 사용 현황과 경향을 살펴보고, 그 결과를 바탕으로 국내 양적 연구의 통계 사용에 있어서의 문제점을 논의하는 것이다. 이를 위해 본 연구는 국내 커뮤니케이션 연구의 대표 학술지라고 할 수 있는 <한국언론학보>에 게재된 양적 연구들에 대한 내용분석을 수행하고, 이를 통해 국내 커뮤니케이션 연구의 통계 사용의 관행과 문제점을 논의하고자 한다.

2. 분석의 틀

전술한 바와 같이 본 연구는 <한국언론학보> 논문들의 통계분석방법 사용 현황과 경향을 알아보는 것에 목적을 두고 있다. 연구 대상 논문 분석 틀을 구성함에 있어, 사회과학 분야에서 수행된 연구들을 대상으로 진행되었던 이병관 외(2015)의 연구에서 사용된 분석 틀을 기초로 통계 방법 분석에 대한 고찰을 진행한 국내외 연구들을 참고하여 수정 및 보완하였다(남기성, 2004; Bailar & Mosteller, 1988; Boster, 2002).

우선, 본 연구는 통계분석 기법을 활용한 논문을 분석 대상으로 하기 때문에 질적 연구 논문을 제외한 양적 연구(quantitative research) 논문을 분석 대상으로 선정하였다. 이어 각 연구에서 사용된 방법론에 따라 표본 크기, 표집 대상 등 연구의 기본적 특성을 점검하였다. 특히, 내용분석 연구방법을 사용한 경우 결과의 신뢰도(inter-coder reliability), 코더 수(number

of coder), 코더 간 신뢰도 분석방법(analysis method of inter-coder reliability)을 추가적으로 확인하였고 실험 연구의 경우 피실험자 할당방법과 실험 효과 크기(power)를 검토하였다. 기본 유목들을 점검한 후, 영가설사용, 다중문항 측정 도구 사용 등에 대하여 확인하고 나아가 t 검증, 회귀분석 등 개별 연구에서 활용된 통계분석방법에 대한 기본 가정, 통계 결과 제시 등의 보다 구체적인 내용들에 대한 검토를 아래와 같이 진행하였다.

첫째, 개별 연구의 측정도구에 관하여 측정변인의 측정수준, 구성변인의 타당도(validity) 및 신뢰도(reliability), 결측치(missing data) 처리 등에 대해 검토가 이루어졌다. APA 가이드라인에 따르면 심리 측정을 위해 둘 이상의 문항을 사용하여 모집단의 자료를 수집할 경우, 타당도, 신뢰도, 표준오차 등의 결과에 영향을 미칠 수 있는 측정도구의 속성을 요약 및 보고하도록 권고하고 있다(Wilkinson & APA Task Force on Statistical Inference, 1999). 연구를 위해 수집한 데이터 분석에 있어서 특정 정보가 누락되는 결측치가 존재할 수 있는데 이는 참여자, 연구 설계, 참여자와 연구 설계의 상호작용 다양한 원인으로 발생할 수 있다(McKnight, 2007). 일반적으로 결측치를 처리하는 방법으로는 하나의 응답이라도 누락될 경우 해당 데이터를 삭제하는 목록삭제(listwise deletion), 누락된 응답항목만을 삭제하고 나머지는 유지한 채 분석하는 한 쌍 제거(pairwise deletion), 일정한 과정을 통해 생성된 하나의 대푯값으로 대체하는 단일대체(single imputation), 다양한 기법을 활용하여 추정된 값을 대체하는 다중대체(multiple imputation) 등이 있다(Graham, 2012). 개념 타당도를 확보하는 방법에는 탐색적요인분석(exploratory factor analysis: EFA), 확인적요인분석(confirmatory factor analysis: CFA), 다속성-다방법(multi-trait multi method: MTMM) 기술, 집단 간 비교 등이 있고, 내용타당도를 확보하는 방법으로는 표면타당도, 액면타당도, 논리적 타당도 등이 있다. 측정의 타당도 못지않게 측정 대상을 얼마나 일관성 있게 측정하는가에 대한 측정의 신뢰도 역시 중요하다. 측정 도구의 적절성을 판단할 수 있는 핵심 요소의 하나인 신뢰도는 측정하고자 하는 현상이나 특성을 각기 상이한 조건하에서 얼마나 안정적으로 비슷한 결과를 산출하는가를 살펴보는 것이다(이병관 외, 2018). 신뢰도는 관찰된 효과 크기에 영향을 미칠 수 있기 때문에, 신뢰도가 확보되지 않은 항목으로 측정할 경우 효과가 실제보다 축소되는 제2종 오류(Type II Error)가 나타날 수 있다(Ritter, 2010). APA 가이드라인에서도 측정문항에 대한 타당도와 신뢰도를 반드시 보고하도록 규정하고 있다.

둘째, 개별 연구들을 수행하는데 있어 어떤 통계 패키지(statistical package, 예: SPSS, R, SAS 등)를 사용했는지, 기술 통계치(descriptive statistics, 예: 평균, 표준편차 등)를 어떻게 처리했는가와 함께 어떤 방법을 통해 연구가설에 대한 추론을 하였는지(예: 영가설 검증과

신뢰구간 추정) 등에 대해 분석을 실시하였다. 엣지 컴퓨팅(edge computing), 클라우드(data cloud), 인공지능(artificial intelligence: AI) 등의 4차 산업혁명 기술의 발전과 함께 데이터 양이 폭발적으로 증가하고 있으며, IDC에서 발간한 'Data Age 2025(Reinsel, Gantz, & Rydning, 2018)'에 따르면 2025년의 총 데이터 양은 175제타바이트(zeta byte: ZB, 약 1조 1000억 GB)까지 증가할 것으로 전망하였으며 지금도 하루에 약 3엑사바이트(exa byte: EB, 약 10억 GB)의 데이터가 생성되고 있다. 이러한 변화는 사회과학 분야에서의 새로운 연구방법(예: 네트워크 분석, 시계열 빅데이터 분석 등)을 등장 시켰다. 또한 방대한 데이터 처리와 함께 쉽게 데이터를 파악하게 해주는 표, 그래프 나아가 3D형태의 시각화 데이터 자료를 생성해주는 새로운 통계 프로그램(예: R, Python, NodeXL, Java)이 사용되기 시작했다. 이에 본 연구에서는 해당 통계 프로그램의 사용 경향에 대해 점검하였고, 이와 함께 NHST의 사용 행태에 대해 분석하였다. NHST는 p 값(유의수준)이 특정가설, 즉 영가설의 정확성을 정량화하지 않는다는 점, 통계적으로 유의미한 p 값(.05 또는 .01 이하)을 얻기 위해 의도적이고 가차 없는 시도(예: 유의미한 p 값으로 떨어질 때까지 표본 크기를 늘리는 것)를 하는 이른바 p -해킹(p -hacking), 유의미한 p 값만으로 영가설을 기각하고 연구가설을 채택하는 등(박준석, 2015) 잘못된 지식과 신념, 그리고 남용으로 인해 비판받아 왔으며(Bakan, 1966; Boster, 2002; Hunter, 1997; Levine et al., 2008; Nickerson, 2000), 효용성에 대해 문제(Cohen, 1994; Schmidt & Hunter, 1997)가 제기되기도 하였다. 이에 따라 여러 선행 문헌들(예: Bailar & Mosteller, 1988; Boster, 2002; Wilkinson & APA Task Force on Statistical Inference, 1999)은 NHST를 사용할 때에는 대안적인 추정치(예: 효과 크기(effect size), 신뢰구간(confidence interval) 등)를 함께 보고하는 것을 권고하고 있다.

셋째, 본 연구에서는 개별 통계분석방법 및 검증방법의 활용 현황과 통계 결과 제시에 대한 사항을 살펴보았다. 일반적으로 개별 통계분석방법은 각 방법을 실행하기 전에 확인해야 하는 기본적인 통계 가정(basic statistical assumption)들이 존재하며, 통계분석을 실시한 이후 추가 검증을 해야 하는 사후검증이 있기도 하다. 평균 비교를 목적으로 하는 t 검증(T -test)과 분산분석(ANOVA)의 경우 기본적으로 각 집단의 정규 분포(normal distribution) 혹은 정규성(normality)을 가정하고 있으며 단일 표본 t 검증과 대응표본 t 검증을 제외하고, 두 집단 이상을 비교하는 경우 집단의 변량의 동질성(homogeneity of variance)을 전제로 하고 있다. 이외에도 집단 내 분산분석(within ANOVA 혹은 repeated measure ANOVA)의 경우 구형성(sphericity)에 대한 검증이 등분산 검증의 대체로 실행되어야 한다. 또한 분산분석은 세 집단 이상의 평균을 비교하는데 사용하는 통계방법으로써 개별 집단 간의 차이를 파악하기 위해서는

투키(Tukey)의 HSD, 피셔(Fisher)의 LSD, 본페로니(Bonferroni) 보정 등의 사후 분석(post-hoc comparison)을 실시해야만 한다. 이에 본 연구는 t 검정과 분산분석을 활용한 분석 대상 논문에 대하여 전제 가정을 지키는지에 대한 여부, 분산분석의 경우 사후 분석을 실시하였는지, 추가적으로 η^2 , 부분(partial) η^2 , ω^2 등의 효과 크기 제시 현황을 중점적으로 파악하였다.

넷째, 연구에서 사용된 각 변인들 간의 인과성과 설명력의 확인을 목적으로 한 회귀분석 또한 여러 선행 가정이 충족되어야 한다. 일반적으로 회귀분석(regression)은 오차의 정규성에 대한 가정(normality of the error distribution), 독립변인과 종속변인의 선형관계에 대한 가정(assumption of linearity), 독립변인의 등분산성에 대한 가정(assumption of homoscedasticity), 주어진 X 에 대해 오차항은 독립이어야하며 그 평균은 0이어야 한다는 가정(conditional mean of zero) 등이 점검되어야 한다. 이러한 주요 가정들 이외에도 두 가지 이상의 독립변인이 투입되는 다중회귀분석의 경우 독립변인 간의 상관관계가 없어야 하는 다중공선성(multicollinearity)을 추가로 검증할 필요가 있다. 또한 회귀분석의 유형에 따라 이러한 가정들은 달라지는데, 예를 들면 로지스틱 회귀분석의 경우 독립변인과 종속변인의 선형관계에 대한 가정이나 오차의 정규성에 대한 가정은 필요치 않으나 오차항의 독립성(independence of errors)이나, 연속형 독립변인에 대한 로짓의 선형성(linearity of independent variables and log odds), 독립변인 간의 다중공선성(little or no multicollinearity among the independent variables) 등에 대한 검증이 필요하다. 특히 로지스틱 회귀분석은 최소자승법(ordinary least square: OLS)과는 달리 최대우도(maximum likelihood: 이하 ML로 표기)추정법을 사용하기 때문에 보다 큰 표본 크기가 필요하다.¹⁾ 본 연구에서는 이러한 가정에 대한 검증 여부와 회귀분석의 종류, 방법과 더불어 회귀분석의 효과 크기로 사용되는 R^2 , 수정(adjusted) R^2 , 오즈비(odds ratio: OR) 등의 사용 양상을 살펴보았다. 한편, 단독으로 사용되는 빈도는 낮으나 회귀분석, 요인분석 등을 실행하기에 앞서 사용되는 상관분석의 경우 상관분석의 종류(이변량 상관분석, 편상관분석)와 변인의 척도 수준에 따라 결정되는 상관계수의 형태(Pearson의 r , Spearman의 ρ 등)를 살펴보았다.

끝으로, 본 연구는 구조방정식모형(structural equation modeling: 이하 SEM으로 표기)의 사용 양상에 대해 살펴보았다. 비교적 짧은 역사에도 불구하고, SEM은 인과관계에 대한

1) 로지스틱 회귀분석을 위한 적절한 표본 크기에 대해 합의된 것은 없지만 일반적으로 몇몇 가이드가 제시되어 왔다. 이를 위해서는 페두치, 콘카토, 켐퍼, 홀포드, 그리고 페인슈타인(Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996)과 쉐이(Hsieh, 1989)의 연구를 참고하길 바란다.

가설 검증을 위한 강력한 도구로서 최근 사회과학 분야에서 대중적으로 사용되어 왔다. 이러한 대중적 사용과 관련하여 SEM의 오용에 대한 우려도 여러 학자들(예: 강남준, 1999; 이기중, 2016; Baumgartner & Homburg, 1996; Holbert & Stephenson, 2002; MacCallum & Austin, 2000)에 의해 제기되고 있다. 본 연구에서 SEM 사용 시 필수적으로 고려해야만 하는 다양한 요인들을 중심으로 SEM 이용 행태에 대해 살펴보았다.

3. 연구방법

본 연구에서는 지난 10년 간 한국언론학회의 <한국언론학보>에 게재된 논문을 분석 대상으로 선정하였다. 2010년(제54권 1호)부터 2019년(제63권 6호)까지 발간된 논문들에 대해서 내용분석을 실시하였으며, 구체적인 연구논문 수집과 검토, 자료 입력은 광고홍보학 전공 대학원생 4명이 담당하여 진행하였다. 코더 간 신뢰도 검증은 전체 표본 중 10%에 해당하는 62편의 논문을 무작위로 추출하여 개념적인 틀에 따라 각자 분석한 뒤, 이를 상호 교차적으로 비교 검증하였다. 코더 간 신뢰도 수준이 낮게 나오는 항목에 대해서는 토론을 거쳐 최종 합의에 도달한 뒤 30개의 논문을 무작위로 재추출하여 2차 코딩을 실시하였다. 이때, 코더 간 신뢰도가 만족할 만한 수준으로 도출되지 않아 총 3차에 걸쳐 코더 간 신뢰도 수준을 높이기 위한 토론 및 교육이 진행되었다. 결과적으로 각 항목에 대한 크리펜돌프 알파(Krippendorff, 2004) 신뢰도 수준은 0.71에서 1 사이로 나타나 안정적인 수준의 신뢰도를 확보하였다. 연구를 분석하기 위한 기본 분석 단위는 개별연구이며, 하나의 연구에서 중복 응답이 필요한 내용에 대해서는 중복응답을 기준으로 결과를 제시하였다. 또한 하나의 논문에서 두 가지 이상의 연구를 진행한 경우에는 각각의 개별 연구로 코딩을 진행하였으며 본 연구를 위한 자료 처리는 R programming language & statistical software(R Core Team, 2020)를 이용하였다. <한국언론학보>에서는 2010년부터 2019년까지 총 845편의 논문이 발표되었고, 이 중 612편이 양적 연구방법을 사용한 것으로 나타나 전체 논문의 약 72% 정도가 분석에 사용되었다.

4. 연구 결과

1) 연구방법 및 통계분석방법 이용 현황

분석대상논문에서 어떠한 연구방법론(research method)을 사용했는지를 파악하기 위해 기본적으로 각 논문의 연구방법을 살펴보았다. 1차적으로 개별 연구의 방법론 부분에 정확히 명시되어있는 경우 해당 방법론으로 분류하였으며(예: 설문조사, 서베이, 설문조사 전문기관, 실험, 내용분석, 메타분석 등), 크롤링(crawling)과 같이 기계 수집을 사용하거나 Kirkwic와 같이 텍스트 마이닝을 활용한 연구는 빅데이터 연구로 분류하였고, 이외의 2차 자료 분석 등은 기타 유목으로 분석하였다.

논문 발표 시기에 따른 연구방법 사용 현황을 살펴보았을 때, 설문조사(survey)를 사용한 연구가 모든 시기를 통틀어 가장 많았다(Table 1) 참조. 각 연도별 발표된 전체 논문 대비 실험연구(experiment)를 사용한 논문의 비율은 약 20%정도로 동일했다. 빅데이터 연구는 2011년 1편에서 2015년 6편까지 증가하였고, 2019년을 제외하면 2014년 이후 매년 한 편 이상 발표되고 있었다. 2차 자료를 이용한 연구도 48편이 있었는데, 기존 서베이 데이터를 사용한 연구가 22편, 시정률 데이터 등 내부 자료를 활용한 분석이 24편 존재했다.

Table 1. Research Method across Publication Years

	Method						Total
	Survey	Experiment	Contents Analysis	Meta	Big data	Other	
2010	35(44)	20(25)	20(25)	-	-	5(6)	80(100)
2011	38(54)	14(20)	13(18)	-	1(1)	5(7)	71(100)
2012	44(59)	14(19)	9(12)	1(1)	2(3)	4(5)	74(100)
2013	41(55)	15(20)	13(17)	-	-	6(8)	75(100)
2014	46(53)	16(19)	14(16)	-	6(7)	4(5)	86(100)
2015	28(49)	14(20)	13(19)	2(3)	6(9)	7(10)	70(100)
2016	28(49)	11(19)	10(18)	-	5(9)	3(5)	57(100)
2017	17(40)	9(21)	7(17)	-	4(10)	5(12)	42(100)
2018	12(32)	8(22)	11(30)	-	2(5)	4(11)	37(100)
2019	16(53)	6(20)	2(7)	1(3)	-	5(17)	30(100)
Total	305	127	112	4	26	48	622

Note. Number of publications with percentages in the parentheses.

연구방법별로 사용된 통계분석방법을 살펴보면, 대부분이 기술통계(98.4%)이었고, 그 다음이 회귀분석(44.1%), 분산분석(33.4%), 상관분석(28.1%), *t* 검증(26.5%) 등의 순이었다. 이에 비해 시계열분석, AHP 분석, 판별분석 등은 활용 빈도가 매우 낮았다. 기타 통계분석방법을 사용한 경우는 33건이 존재했는데, 그래인저 인과관계(granger causality)는 4건, 교차지연 상관분석 2건, 맨-휘트니 U검정 2건 등이었다.

통계처리를 위해 사용한 프로그램에 대해서는 전체의 65.8%에 해당하는 409편에서 특별한 언급이 존재하지 않았다. 통계 패키지를 밝힌 연구는 거의 대부분 SPSS를 활용하여 통계처리를 진행하고 있었고, 적지만 R이나 Python을 활용하여 통계처리를 진행했다고 밝힌 연구도 존재했다. 기타 통계 패키지에는 Netdraw(5건), HLM(3건), Expert Choice(3건) 등이 있다(〈Table 2〉 참조).

Table 2. Statistical Method and Packages

	Type	Frequency	%
Statistical Method*	Descriptive Statistics	612	98.4
	Regression	274	44.1
	ANOVA	208	33.4
	Correlation	175	28.1
	T-test	165	26.5
	Structural Equation Modeling	133	21.4
	Factor Analysis	129	20.7
	χ^2	100	16.1
	Network Analysis	30	4.8
	Cluster Analysis	21	3.4
	Time Series	8	1.3
	AHP	5	0.8
	Discriminant Analysis	1	0.2
Other	33	5.3	
Statistical Package*	SPSS	104	16.7
	AMOS	79	12.7
	UCInet	12	1.9
	NodeXL	10	1.6
	R	8	1.3
	LISREL	8	1.3
	Python	6	1
	Other	49	7.9
	Not Specified	409	65.8

Note. *Multiple Responses($n = 622$).

메타분석과 2차 자료 분석을 제외하고 연구방법별로 사용된 통계분석방법을 살펴본 결과, 모든 연구방법에서 기술통계를 가장 많이 사용했다. 기술통계를 제외하면 설문조사의 경우 회귀 분석, 상관분석, SEM이 많이 사용되었으며, 실험연구의 경우 분산분석이 가장 많이 사용되었다. 내용분석의 경우 대부분 교차분석을 분석에 사용하고 있었으며, 빅데이터 연구는 네트워크 분석을 가장 활발하게 사용하는 것으로 나타났다(〈Table 3〉 참조).

Table 3. Statistical Method by Research Method

	Survey	Experiment	Contents Analysis	Big data	Total
Descriptive Statistics	301(52.8)	127(22.3)	110(19.3)	22(3.9)	560(98.3)
Regression	184(32.3)	45(7.9)	13(2.3)	4(0.7)	246(43.2)
ANOVA	67(11.8)	108(19)	24(4.2)	2(0.4)	201(35.3)
Correlation	117(20.5)	21(3.7)	14(2.5)	5(0.9)	157(27.5)
T-test	67(11.8)	76(13.3)	12(2.11)	1(0.2)	156(27.4)
SEM	111(19.5)	15(2.6)	-	-	126(22.1)
EFA	109(19.1)	13(2.3)	2(0.4)	1(0.2)	125(21.9)
χ^2 -test	23(4)	9(1.6)	62(10.9)	3(0.5)	97(17)
Network Analysis	2(0.4)	1(0.2)	3(0.5)	19(3.3)	25(4.4)
Cluster Analysis	16(2.8)	1(0.2)	2(0.3)	2(0.3)	21(3.7)
Time Series Analysis	-	-	5(0.9)	2(0.3)	7(1.2)
AHP	5(0.9)	-	-	-	5(0.9)
Discriminant Analysis	1(0.2)	-	-	-	1(0.2)
Other	7(1.2)	4(0.7)	7(1.2)	5(0.9)	23(4)
Total	305(53.5)	127(22.3)	112(19.7)	26(4.6)	570(100)

Note. Multiple responses($n = 570$), Number of publications with percentages in the parentheses.

2) 연구방법에 따른 표본 크기 및 표본 특성

연구방법에 따른 표본 크기를 살펴본 결과는 <Table 4>와 같다. 추가적으로, 연구방법에 따른 표본 특성을 살펴본 결과, 총 622편 중 267편이 일반인을 표집대상으로 삼고 있었다. 설문조사를 사용한 연구는 일반인(65.2%)을 표집대상으로 하는 경우가 가장 많았으며 그 다음으로 대학(원)생(16.7%), 청소년(16.1%)의 순이었다. 설문조사의 표집대상의 기타 항목에는 국내 거주 외국인, 북한 이탈 주민 등이 포함되어 있었다. 실험연구의 경우는 대학(원)생(55.1%)이 가장 많았으며, 그 다음으로 일반인(37.8%)이었다. 내용분석과 빅데이터 연구의 경우 표집 대상의 대부분이 기사(37편)였으며, 트위터나 유튜브 같은 SNS를 표집 대상으로 선정한 경우도 다수 존재했다(<Table 5> 참조).

Table 4. Sample Size by Research Method

Method	Total	M	SD	Min	Med	Max
Survey	305	560.58	472.88	21	395.0	3032
Experiment	127	252.03	166.75	46	200.0	1002
Contents Analysis	112	2290.14	7764.76	4	659.0	76444
Big Data	26	126750.7	310247.2	87	3246.0	1080288
Meta	4	676.00	917.31	153	251.5	2048
Other	48	3836.05	9308.66	34	921.0	56207

Table 5. Sample Characteristics by Research Method

	Sample Characteristics					Total
	General Public	Professionals	College Students	Under-aged	Other	
Survey	199(65.2)	23(7.5)	51(16.7)	49(16.1)	24(7.9)	305(100)
Experiment	48(37.8)	-	70(55.1)	8(6.3)	1(0.8)	127(100)
Contents Analysis	-	-	-	-	112(100)	112(100)
Big Data	-	-	-	-	4(100)	4(100)
Meta	-	-	-	-	26(100)	26(100)
Other	20(41.7)	2(4.2)	-	2(4.2)	24(50)	48(100)

Note. Multiple responses($n = 622$), Number of publications with percentages in the parentheses.

본 연구는 실험연구를 수행한 연구 중 실험 집단의 설계(예: 무선 배정 혹은 준 실험설계 등) 및 검정력(power) 제시 여부 또한 살펴보았다. 먼저 실험 집단의 설계의 경우, 실험연구방법을 사용한 127편의 논문 중 89%인 113편의 논문이 실험 집단의 설계방법을 제시하고 있었고, 실험 집단의 설계방법을 언급하지 않은 논문은 14편(11%)으로 나타났다. 또한 실험연구 전 검정력 분석에 대한 결과를 제시한 논문은 단 1편 존재하였다(〈Table 6〉 참조).

Table 6. Experiment Group Assignment Method and Power Analysis

	Type	Frequency	%
Experiment Assignment Method	Random	85	66.9
	Quasi	28	22.1
	Not Specified	14	11
Power Analysis	Provided	1	0.8
	Not Specified	126	99.2

Note. $n = 127$

3) 측정도구의 신뢰도 및 타당도 검증

연구에서 측정도구의 신뢰도 및 타당도 검증방법을 분석한 결과는 다음과 같다(〈Table 7〉 참조). 신뢰도와 타당도 검증이 필요한 다문항 측정도구를 사용한 논문은 총 423편(68%)이었으며, 다문항 측정도구를 사용하였는지 정확히 알 수 없는 경우도 19편(3%) 있었다. 다문항 측정도구를 사용한 423편의 논문 중 94%인 396편이 신뢰도를 밝히고 있었으며, 타당도는 44.2%인 187편만 밝히고 있었다. 신뢰도를 밝힌 논문 모두가 크론바흐 α 를 사용하고 있었으며 10편의 논문은 크론바흐 α 값에 추가적으로 상관계수를 보고하고 있었다. 타당도 검증을 위해서는 탐색적 요인분석을 사용한 경우가 많았으며, 13.2%인 56편의 논문은 확인적요인분석을 사용했다. 탐색

적요인분석과 확인적요인분석 모두를 사용한 논문도 27편(6.4%) 존재했다. 추가적으로 안면 타당도, 판별 타당도 등을 이용하여 타당도 검증을 진행한 논문도 5편이 있었다.

Table 7. Publications Reporting Validity and Reliability of Research Instruments

	Type	Frequency	%
Use of Multi-item Scale	Use	423	68
	Didn't Use/Not Specified	19	3
Reliability*	Cronbach's α	396	93.6
	Correlation Coefficient	10	2.4
	Not Specified	27	6.4
Validity*	Exploratory Factor Analysis	99	23.4
	Confirmatory Factor Analysis	56	13.2
	Both	27	6.4
	Other	5	1.2
	Not Specified	241	57

Note. $n = 622$ *Multiple responses($n = 423$).

내용분석, 메타, 빅데이터 연구를 제외하고, 총 438편의 논문 중 결측치 제거에 대해 명시한 연구는 12편이 있었으며 12편의 논문 모두 완전 제거(listwise) 방식을 이용해 결측치 처리에 대한 내용을 보고하였다(〈Table 8〉 참조).

Table 8. Replacement of Missing Values

	Type	Frequency	%
Replacement of Missing Values	Pairwise	-	-
	Listwise	12	2.7
	Not Specified	426	97.3

Note. $n = 438$

4) 영가설 유의도 검증과 신뢰구간

〈Table 9〉에 따르면, 연구 대상이 된 총 622개의 연구 중 49개의 연구를 제외한 573개의 연구(92.1%)가 NHST를 이용하여 가설검증을 시도한 것으로 나타났다. 이들 573개의 연구를 대상으로 신뢰구간을 제시하였는지 여부를 확인하였을 때, 신뢰구간을 밝히고 있는 연구는 42편(7.3%)에 불과하여 거의 대부분의 연구에서 신뢰구간을 제시하지 않는 것으로 나타났다.

Table 9. Publications Using NHST and Reporting Confidence Interval

	Type	Frequency	%
Using NHST	Use	573	92.1
	Didn't use	49	7.9
Confidence Interval*	Provided	42	7.3
	Not Provided	531	92.7

Note. $n = 622$ *Only for research using NHST($n = 573$).

5) 통계분석방법 사용 현황

(1) t 검증

통계적 추론을 위해 t 검증을 사용한 연구는 총 165건이었다. 이 중 독립표본 t 검증(independent T-test)을 사용한 연구가 136건(82.4%)으로 가장 많았으며, 대응 표본 t 검증(paired T-test)이 22.4%, 단일표본 t 검증(one sample T-test)이 2.4%의 순으로 나타났다. 한편, 독립표본 t 검증을 위한 선행조건인 동질성 검증에 대한 결과가 제시되었는가를 분석한 결과, 동질성 검증이 필요한 전체 136건의 연구 중 3건의 연구만이 동질성 검증을 결과를 제시한 것으로 집계되었다(〈Table 10〉 참조).

Table 10. T-Test: Type and Test of Assumption

	Type	Frequency	%
Test of Homogeneity of Variance*	Tested	3	2.2
	Not Tested or Provided	133	97.8
Type of T-test**	One Sample T-test	4	2.4
	Independent Sample T-test	136	82.4
	Paired Sample T-test	37	22.4

Note. *Only for independent t-test($n = 136$). **Multiple responses($n = 165$).

(2) 분산분석

분석 대상 연구에서 분산분석을 활용한 연구는 총 208편으로 나타났으며, 그 중 일원분산분석(one-way between ANOVA)을 실시한 연구가 116건으로 전체의 55.8%를 차지하여 가장 많았다. 그 다음이 이원분산분석(two-way between ANOVA) 19.7%, 공분산분석(ANCOVA) 18.8%, 일원반복측정분산분석(one-way within ANOVA) 9.1%의 순으로 나타났다.

분산분석에 앞서 분석 자료가 통계적 가정을 충족하였는지에 대한 검증, 즉 동질성, 정규성, 구형성 검증결과를 제시하였는지를 각 분산분석의 종류별로 살펴보았다. 먼저 정규성 검증은

분산분석의 종류와 상관 없이 그 결과를 제시하여야 함에도 불구하고 208건의 연구 중 4건 (1.9%)만이 검증 결과를 제시하고 있었다. 동질성 검증의 경우 집단 간 분산분석 및 혼합 요인 분산분석을 수행한 경우 결과를 제시하여야 하며, 본 연구에서는 총 191건의 연구 중 6건 (3.1%)의 연구에서만 검증 결과가 제시되었다. 구형성 검증은 집단 내 분산분석과 혼합 요인 분산분석의 경우 제시하여야 하며, 본 연구에서는 총 36건의 연구 중 7건(19.4%)에서만 검증 결과를 제시하였다.

사후검증 결과를 제시한 연구는 208건 중 87건(41.8%)이었으며, 사후검증방법으로는 Scheffe 검정(Scheffe test, 21.6%), 투키의 HSD(6.3%), 본페로니 검정(3.4%) 등을 사용했다. 기타 사후검증방법에는 일반 t 검증(10건), 대비검정(2건)등이 있었다. 2개 이상의 방법을 사용하여 사후검증 결과를 제시한 연구도 2편 확인하였다. 효과크기와 관련하여, 효과크기를 밝히고 있는 연구는 60건(28.8%)에 불과하였으며, 대부분이 부분 η^2 (44건, 21.2%)를 사용해 효과크기를 보고하고 있었다. 효과크기로 η^2 가 제시된 경우도 14건(6.7%) 있었으며, 둘 다 제시한 경우도 2건 존재했다(〈Table 11〉 참조).

Table 11. ANOVA: Type, Test of Assumptions, Post-hoc Analysis, and Effect Size

	Type	Frequency	%	
Test of Homogeneity*	Tested	6	3.1	
	Not Tested or Provided	185	96.9	
Test of Normality	Tested	4	1.9	
	Not Tested or Provided	204	98.1	
Test of Sphericity**	Tested	7	19.4	
	Not Tested or Provided	29	80.6	
ANOVA Method***	One-way Between Groups	116	55.8	
	One-way Within Group	19	9.1	
	Two-way Between Groups	41	19.7	
	Two-way Within Group	4	1.9	
	Two-way Mixed	7	3.4	
	Three-way Between Groups	8	3.9	
	Three-way Mixed	7	3.4	
	Other	ANCOVA	39	18.8
		MANOVA	12	5.8
		MANCOVA	8	3.9
ATI		1	0.5	

	Type	Frequency	%
Post-hoc Analysis***	Scheffe	45	21.6
	Tukey	13	6.3
	Bonferroni	7	3.4
	LSD	5	2.4
	Duncan	2	1
	Dunnett	1	0.5
	Other	17	8.2
	Not Tested or Provided	121	58.2
Effect Size	η^2	14	6.7
	Partial η^2	44	21.2
	Both	2	1
	Not Tested or Provided	148	71.2

Note. $n = 208$ *Only for between&mixed ANOVA($n = 191$). **Only for within&mixed ANOVA($n = 36$). ***Multiple responses($n = 208$).

(3) 회귀분석

회귀분석방법을 활용한 총 274건의 연구 가운데 98.2%에 해당하는 270건의 연구가 다중회귀분석을 사용하였으며, 단순회귀분석은 9건(3.3%)의 연구에서 사용하였다. 또한 대부분의 연구가 선형회귀모델(89.6%)을 사용하고 있었으며, 로지스틱 모델을 사용한 경우도 27건(9.7%) 있었다. 결정계수와 관련해서는 87.6%에 해당하는 240건의 연구가 이를 제시하고 있었으며, R^2 를 제시한 연구가 57.3%로 가장 많이 차지하고 있었고, 그다음 수정 R^2 를 제시한 연구(16.8%), 그리고 둘 다 제시한 연구(13.1%)의 순서였다.

한편 분석에 사용된 변인들에 대해 단순 및 다중회귀분석의 선행 조건인 정규성, 선형성, 영조건부평균(conditional mean of zero) 검증 결과를 제시한 연구는 정규성 5건(2%), 선형성 2건(0.8%)에 불과하였으며, 영조건부평균의 경우 검증 결과를 제시한 연구가 존재하지 않았다. 한편, 다중회귀분석의 경우 다중공선성 검증 결과가 제시된 연구는 59건(23%)에 불과하여 대부분의 연구에서 회귀분석의 기본 가정을 기술하고 있지 않았음을 알 수 있다.

로지스틱 회귀분석의 경우에는 독립변인 간의 다중공선성에 대한 검증 결과를 제시한 연구 4편(14.8%)을 제외하고 기본 가정인 오차항의 독립성, 연속형 독립변인에 대한 로짓의 선형성의 검증 결과를 제시한 연구는 없었다(〈Table 12〉 참조).

Table 12. Regression: Type, Test of Assumptions, and Provision of Determination Coefficients

	Type	Frequency	%	
Test of Normality*	Tested	5	2	
	Not Tested or Provided	245	98	
Test of Linearity*	Tested	2	0.8	
	Not Tested or Provided	248	99.2	
Test of zero conditional mean*	Tested	0	0	
	Not Tested or Provided	250	100	
Test of Multicollinearity	Multiple Regression**	Tested	59	23
		Not Tested or Provided	186	77
	Logistic Regression***	Tested	4	14.8
		Not Tested or Provided	23	85.2
Test of independence of errors***	Tested	0	0	
	Not Tested or Provided	27	100	
Test of linearity of independent variables and log odds***	Tested	0	0	
	Not Tested or Provided	27	100	
Type of Regression Model	Simple Regression	4	1.5	
	Multiple Regression	265	96.7	
	Both	5	1.8	
Type of Regression Analysis****	Linear Model	249	89.6	
	Non-Linear Model	2	0.6	
	Logistic Model	24	8.6	
	Multinomial Logistic Model	3	1.1	
Coefficient of Determination	R^2	157	57.3	
	Adjusted R^2	46	16.8	
	R2 and Adjusted R^2	36	13.1	
	Odds Ratio	1	0.4	
	Not Provided	34	12.4	
Correlation	Provided	97	35.4	
	Not Provided	177	64.6	

Note. *Only for simple&multiple linear regression($n = 250$). **Only for multiple linear regression($n = 245$). ***Only for logistic regression($n = 27$). ****Multiple responses($n = 274$).

(4) 상관관계분석

상관관계분석을 활용한 175건의 연구 중 대부분이 이변량 상관분석(bivariate correlation 혹은 zero-order correlation)을 실시하였으며, 편상관분석(partial correlation)을 사용한 연구는 4편 존재했다. 상관계수의 경우 171건(97.7%)이 피어슨의 적률상관계수를 사용하였고 4건이 스피어만의 순위상관계수를 사용했다(〈Table 13〉 참조).

Table 13. Type of Correlation Analysis and Correlation Coefficient

	Type	Frequency	%
Test of Correlation Analysis	Zero-order Correlation	171	97.7
	Partial Correlation	4	2.3
Type of Regression Model	Pearson's r	171	97.7
	Spearman's ρ	4	2.3

Note. $n = 175$

(5) 탐색적요인분석

탐색적요인분석을 사용한 총 129건의 연구 중, 요인추출방법을 밝힌 연구는 118건(91.5%)이었으며, 이 중 114건(88.4%)이 주성분분석(principle component analysis)을 활용하였다. 주성분분석을 사용하지 않은 4건의 연구는 ML추정법, 일반화최소제곱법(generalized leased square), 주축분해법 등을 요인추출방법으로 사용하였다. 또한 62%에 해당하는 80건의 연구가 요인회전방법을 제시했는데, 69건(53.5%)이 배리맥스(varimax) 방식을, 5건(3.9%)이 오블리민(oblimin) 방식을 사용하였다. 기타회전방법으로는 프로맥스(promax), 해리스-카이저(Harris-Kaiser) 등이 사용되었다. 탐색적요인분석 수행에 따른 변인들의 구형성 검증을 제시하였는지 여부를 살펴본 결과, 20.2%에 해당하는 26건의 연구에서만 구형성 검증 결과를 제시한 것을 알 수 있었다(〈Table 14〉 참조).

Table 14. Exploratory Factor Analysis: Test of Assumption, Method of Extraction and Rotation

	Type	Frequency	%
Test of Sphericity	Tested	26	20.2
	Not Tested or Provided	103	79.8
Method of Extraction	Principle Component	114	88.4
	Other	4	3.1
	Not Specified	11	8.5
Method of Rotation	Varimax	69	53.5
	Oblimin	5	3.9
	Other	6	4.7
	Not Specified	49	38

Note. $n = 129$

(6) 구조방정식

전체 분석 대상 연구 중 구조방정식을 활용한 연구는 총 134건으로, 이 중 24건(17.9%)이 확인적요인분석의 목적으로, 46건(34.3%)이 경로분석의 목적으로, 그리고 64건(47.8%)이 확인적

요인분석과 경로분석의 목적으로 구조방정식을 활용하였다. 한 건을 제외한 연구가 순환모델(recursive model)을 이용하였으며, 전체의 약 90%에 해당하는 120건의 연구가 자유도(degree of freedom)를 제시하고 있었지만, 자유도를 제시하지 않은 연구도 14건(10.4%) 존재했다. 사용된 통계 패키지의 경우 AMOS가 79건(59%), LISREL 8건(6%), Mplus 7건(5.2%), 기타 3건(2.2%)으로 나타난 한편, 전체의 27.6%인 37건은 구조방정식에 사용된 통계 패키지를 언급하지 않았다. 구조방정식에 활용한 분석 자료의 경우, 단 한 편의 연구에서 공분산 행렬(covariance matrix)을 사용하였다고 언급하고 있었고 나머지 133건의 연구(99.3%) 모두 자료의 성격을 밝히지 않았다. 또한 모수추정방법에 대해서 전체의 80.6%에 해당하는 108건의 연구가 이를 밝히지 않고 있었다. 모수추정방법에 대해 밝힌 연구 26건 중 25건(96.2%)이 ML추정법을 사용하였고 1건만이 완전정보최대우도법(full information maximum likelihood)을 사용하였다. 구조방정식의 선행 가정인 정규성, 선형성, 다중공선성 충족 결과를 제시하였는지를 살펴보았을 때, 정규성 검증은 12건(9%), 다중공선성 검증 8건(6%)의 연구에서 제시하고 있었으며, 선형성 검증 결과를 제시한 연구는 단 한 편도 없었다. 모델의 적합도 검증을 위한 지표로는 χ^2 가 가장 많이 활용되었으며(127건, 94.8%), 그 다음으로는 RMSEA(126건, 94%), CFI(120건, 89.6%), NNFI(82건, 61.2%) 등의 순이었다(〈Table 15〉 참조).

Table 15. Structural Equation Modeling: Test of Assumptions, Estimation of Population, Purpose, Type of Model, Degrees of Freedom, Packages, Data of Analysis, and Use of Fit indices

	Type	Frequency	%
Test of Normality	Tested	12	9
	Not Tested or Provided	122	91
Test of Linearity	Tested	-	-
	Not Tested or Provided	134	100
Test of Multicollinearity	Tested	8	6
	Not Tested or Provided	126	94
Estimation of Population	Maximum Likelihood	25	18.7
	Full Information Maximum Likelihood	1	0.7
	Not Specified	108	80.6
Purpose of Usage	Confirmatory Factor Analysis	24	17.9
	Path Analysis	46	34.3
	Both	64	47.8
SEM Model	Recursive	133	99.3
	Nonrecursive	1	0.7

	Type	Frequency	%
Degrees of Freedom	Provided	120	89.6
	Not Provided	14	10.4
Statistical Package	AMOS	79	59
	LISREL	8	6
	Mplus	7	5.2
	R	1	0.7
	SAS	1	0.7
	SmartPLS	1	0.7
	Not Specified	37	27.6
Data of Analysis	Raw Data	-	-
	Covariance Matrix	1	0.7
	Not Specified	133	99.3
Fit Indices*	χ^2	127	94.8
	RMSEA	126	94
	CFI	120	89.6
	NNFI(TLI)	82	61.2
	NFI	62	46.3
	GFI	59	44
	IFI	26	19.4
	AGFI	24	17.9
	SRMR	22	16.4
	RMR	21	15.7
Correlation	Provided	63	47
	Not Provided	71	53

Note. *Multiple responses($n = 134$).

(7) 내용분석

연구방법론으로 내용분석을 사용한 연구 112건 외에도 내용분석 기법을 활용한 연구를 포함하여 총 125건의 연구에 대해 살펴본 결과, 코더 수는 2명이 74건(59.2%)으로 가장 많았으며, 그 다음으로는 3명(17건, 13.6%)의 순이었다. 코더수를 제시하지 않은 연구도 17건(13.6%)이 존재했다. 코더 간 신뢰도 분석 결과를 제시한 연구는 111건(88.8%)으로, 이 중 코더 간 신뢰도 분석방법을 제시한 연구는 106건이었고, 신뢰도 분석에 사용한 자료의 수를 밝힌 논문은 96건이었다. 분석방법으로는 코헨의 카파(Cohen's kappa)를 이용한 연구가 39건(31.2%)으로 가장 많았으며, 그 다음으로 홀스티의 파이(Holsti's phi) 31건(24.8%), 크리펜돌프의 알파(Krippendorff's alpha) 25건(20%) 등의 순이었다. 15.2%에 해당하는 19건의 연구는 코더 간 신뢰도 분석방법을 밝히지 않았다(〈Table 16〉 참조).

Table 16. Content Analysis: Number of Coders and Test of Reliability

	Type	Frequency	%
Number of Coders	1	1	0.8
	2	74	59.2
	3	17	13.6
	4	7	5.6
	5	4	3.2
	6	4	3.2
	8	1	0.8
	Not Specified	17	13.6
Number of Contets	Provided	96	76.8
	Not Specified	29	23.2
Test of Intercoder Reliability	Tested	111	88.8
	Not Tested or Provided	14	11.2
Reliability Coefficient*	Cohen's Kappa	39	31.2
	Holsti's Phi	31	24.8
	Krippendorff's Alpha	25	20
	Scott's Pi	11	8.8
	Cronbach's Alpha	4	3.2
	Other	5	4
	Not Specified	19	15.2

Note. *Multiple responses($n = 125$).

5. 결론 및 논의

본 연구의 목적은 국내 커뮤니케이션 연구에 사용된 통계분석방법의 양상을 고찰하고, 그에 대한 문제점을 파악하고자 하는 것이었다. 이를 위해 국내 커뮤니케이션 연구의 대표 학술지라고 할 수 있는 <한국언론학보>에 2010년부터 2019년까지 게재된 총 845편의 논문 중 양적 연구방법을 사용한 612편을 대상으로 분석이 수행되었다. 본 연구에서 도출된 연구의 결과는 다음과 같이 논의될 수 있다.

첫 번째 경향은 특정 통계 패키지 사용의 편중이다. 연구 결과에 따르면, 총 622건의 연구 중(중복응답) 통계 프로그램을 밝힌 213건 중 104건(48.8%)의 연구가 SPSS를 사용하고 있는 것으로 나타났으며, 409건(65.8%)의 프로그램을 밝히지 않은 연구 역시 SPSS를 사용했을 개연성이 크다. 최근 데이터 과학의 관심과 함께 R, python, 기타 텍스트 마이닝이나 기계 학습을 위한 패키지 등이 점차 증가하고 있는 추세이기는 하지만, SPSS에 대한 과도한 집중은 여전히

다고 할 수 있다. 물론 특정 통계 프로그램에 대한 집중 자체가 문제가 되는 것은 아니다. 그러나 이러한 SPSS에 대한 과도한 집중은 통계분석방법의 여러 문제를 야기할 수 있다는 점에 경계할 필요가 있다.

예를 들면, 많은 연구들이 효과크기의 지표로 부분 η^2 의 사용에 문제를 제기해 왔다(예: Bakeman, 2005; Fritz, Morris, & Richler, 2012; Levine & Hullett, 2002; Olejnik & Algina, 2003; Richardson, 2011). 특히 집단 내 분산분석이나 요인설계(factorial design) 분산분석의 경우 부분 η^2 가 효과크기로 사용되는 것은 설명 변량(explained variance)이 1을 넘어 해석에 어려움이 있거나 모집단 추정치를 과대추정하게 되는 이유로 문제가 있다. 그러나 대부분의 사회과학 연구자들이 부분 η^2 를 무차별적으로 사용하고 있는데, 주로 사용하는 통계 패키지인 SPSS가 효과크기로 부분 η^2 만을 제시해주기 때문이다(Levine & Hullett, 2002; Olejnik & Algina, 2003; Richardson, 2011). 실제로 본 연구의 결과에서도 이러한 문제는 발견되었는데, 효과크기를 밝히고 있는 60건의 연구 중 72% 이상인 44건이 부분 η^2 를 사용한 것으로 보고하였고, 집단 간 일원분산분석을 제외한 집단 내 분산분석이나 요인설계 분산분석에서 총 42건의 연구(약 69%)가 부분 η^2 를 보고하였다. 또한 η^2 를 사용했다고 밝힌 17건의 연구 중 5개의 연구는 SPSS를 사용한 것으로 밝히고 있어 부분 η^2 를 η^2 로 잘못 보고한 것으로 추측된다. 실제로 여러 연구들(예: Levine & Hullett, 2002; Richardson, 2011)이 이러한 SPSS의 부분 η^2 와 η^2 의 잘못된 표기 혹은 혼동 문제를 강력히 비판하기도 하였다. 이에 관해 부분 η^2 와 η^2 의 대안으로 표준화(generalized) η^2 나 표준화(generalized) ω^2 등을 제안하는 연구들이 많이 있지만(예: Bakeman, 2005; Fritz et al., 2012; Levine & Hullett, 2002; Olejnik & Algina, 2003), 오직 부분 η^2 만을 제시해주는 SPSS만을 사용하는 경우 이러한 문제는 해결될 수 없다.

이러한 특정 통계 프로그램에 대한 과도한 이용은 SEM에서도 나타난다. 사실 SEM은 이론적 개념과 행렬을 이해하지 못하면 쉽게 구현할 수 없는 통계방법임에도 불구하고, AMOS 등과 같은 손쉬운 그래픽 인터페이스(graphical user interface: GUI)를 기반한 SEM 분석 패키지를 통해 많은 연구자들이 SEM을 거의 '관행적'으로 사용하고 있는 것이 현실이다(이병관 외, 2015). 바움가트너와 홈버그(Baumgartner & Homburg, 1996)가 지적한 바와 같이 SEM은 "비숙련된 연구자에게는 위험한 도구"(p. 140)일 수 있다. 본 연구의 결과에서 알 수 있듯이 SEM을 사용한 연구에서 절반 이상의 연구가 통계 패키지로 AMOS를 사용한 것으로 나타났다(79건, 59%), 37건(27.6%)의 연구는 아예 사용된 통계 패키지를 밝히지 않았다. 이러한 상황은 설정된 연구모델의 이론적 배경의 부족, 모델의 부정확한 기술 및 부적절한 추정방법

(estimation methods), 심리 측정 척도의 속성에 대한 부정확한 보고, 표본자료나 표본크기 보고 누락, 부적절한 모델수정(model modification) 과정, 연구 모집단에 대한 기술 누락 등과 같은 SEM의 오용(이병관 외, 2015; Boomsma, 2000; Holbert & Stephenson, 2002)으로 이어진다. 본 연구 결과에서도 단 한편의 연구만이 SEM에 활용된 입력 자료로 공분산 행렬을 사용하였다고 언급하고 있었고, 나머지 133건의 연구는 이에 대한 정보를 정확히 밝히지 않고 있었다. 공분산 행렬을 입력 자료로 사용한 경우와 상관관계 행렬을 입력 자료로 사용한 경우는 서로 다른 결과를 산출할 수도 있다는 점(강남준, 1999; 이기중, 2016; Holbert & Stephenson, 2002; Kline, 1998)에서 SEM에 활용된 입력 자료에 대한 정보를 보고하지 않는 것은 문제가 있다. 또한 SEM에서는 분석된 자료의 성격에 따라 추정절차가 달라져야하기 때문에 분석에서 사용된 모수추정 방식을 반드시 보고해야 한다. 예를 들어, 본 연구에서 모수추정 방식을 제시한 총 26건의 연구에서 25건(약 96%)의 연구는 ML추정법을 가장 많이 사용한 것으로 나타났는데, 이는 다변량 정규성이 위배되는 적은 표본 규모에서는 바람직하지 않은 추정방법이다. 표본 크기가 충분히 큰 경우, ML추정법은 정규성이 위반되지 않을 정도로 견고하지만(Hu, Bentler, & Kano, 1992), 표본 크기가 작은 경우에는 다변량 정규분포가 위배되어 옳은 모델이 기각될 가능성이 보다 높기 때문이다. ML추정법을 위해서는 최소 100개에서 150개까지의 표본규모가 필요하다는 것이 일반적이지만(Ding, Velicer, & Harlow, 1995), 일부 보수적인 학자들은 최소한의 표본규모를 200이상으로 할 것을 주장하고 있다(Anderson & Gerbing, 1998; Bearden, Sharma, & Teel, 1982; Boomsma, 2000; Chou & Bentler, 1995). 이렇게 자료의 정규성 가정이 위배되었을 때에는 가중최소제곱법(weighted least squares: WLS)을 사용해야 하는데, WLS추정을 사용하려면 표본의 크기가 200~500 혹은 그 이상으로 매우 커야 하는 조건이 뒤따른다(Kline, 1998). SEM은 자료의 분포 특성, 특히 다변량 정규성에 매우 민감하기 때문에 분석 전 기본적인 통계 가정들을 점검하는 것은 필수적이다(이병관 외, 2015). 이외에도 적은 표본으로 인해 정규성 가정이 위배되었을 때 적합한 추정방법에 대해 다양한 대안적 방법들에 대한 논의가 있어 왔다(예: 점근분포자유법(asymptotically distribution free: ADF): Browne, 1984; Huang & Bentler, 2015, 요인점수회귀법²⁾(Factor Score Regression: FSR): Croon, 2002; Lu, Kwan, Thomas,

2) 요인점수회귀법(FSR)으로는 일반적으로 regression FSR, Bartlett FSR, bias avoiding method, bias correcting method 등의 4가지 방법이 사용된다. 이 방법들의 설명과 비교에 대한 논의는 데브리저, 메이어, 그리고 라실(Devlieger, Mayer, & Rosseel, 2016)을 참고하기 바란다. 최근 발표된 편향교정 요인점수 경로분석방법(bias-corrected factor score path analysis: BCFSPA)은 FSR의 대안으로 개발된 추정방법이다. 최근의 연구에

& Cedzynski, 2011; Skrondal & Laake, 2001). 본 연구의 분석에 따르면 정규성 미제시 122건(91%), 선형성 미제시 134건(100%), 다중공선성 미제시 126건(94%)으로 나타났으며, 모수추정방식을 보고하지 않은 연구 또한 108건(80.6%)으로 나타났다. 결과적으로 대부분의 연구들이 기본적인 가정과 더불어 모수추정방식에 대해 주의를 기울이지 않는 것으로 나타났는데, 이는 수행된 연구결과의 검증 및 확장을 시도하고자 하는 후속 연구자의 반복연구(replication study)를 불가능하게 할뿐 아니라 도출된 연구결과의 질이나 타당성, 신뢰성에 부정적인 영향을 미칠 수 있다.

본 연구는 특정 통계분석방법에 대해 절대적으로 의존하고 있는 경향 또한 발견하였다. 앞서 언급한 바와 같이 크론바흐의 α 는 1951년 그의 최초 논문(Cronbach, 1951) 이후, 수없이 많은 비판(예: Borsboom, 2006; Shevlin et al., 2000; Sijtsma, 2009; Yang & Green, 2011)에도 불구하고 사회과학 연구자들에게 있어 측정의 신뢰도를 재는 척도로 독보적인 자리를 지켜왔다. 크론바흐 α 는 다중 항목 구성개념의 내적 일관성 신뢰도(internal consistency reliability)를 추정하기 위해 사용되는 계수로, 크론바흐(1951)는 당시 전통적인 신뢰도 측정 방법인 스피어만-브라운(Spearman-Brown)의 반분법(split-half approach)에 대한 문제를 제기하며, 이에 대한 대안으로 관찰 점수(X) 변량에 대한 진점수(T) 변량의 비율로 정의한 α 계수를 제안하였다. 그러나 널리 알려진 것과는 달리, α 계수 산출 공식은 크론바흐에 의해 최초로 제시된 것은 아니다. α 계수는 1951년 크론바흐의 최초 논문(Cronbach, 1951)이 나오기 이전에 이미 여러 학자들에 의해 개발된 것일 뿐 아니라, '알파 계수(α coefficient)'라는 명칭도 크론바흐가 명명한 것은 아니다. 이에 대해 크론바흐는 최근 α 계수에 대한 회고 논문(Cronbach & Shavelson, 2004)에서 "당혹스럽다(embarrassment to me)"(pp. 396-397)고 소회를 밝힌 바 있다(이병관 외, 2015).

α 계수 탄생의 역사적 배경이 어떻든 간에 크론바흐 α 는 개념적·방법론적으로 많은 문제를 안고 있다. 예를 들면, 시즈즈마(Sijtsma, 2009)는 크론바흐 α 가 내적 일관성은 물론 구성개념의 단일 차원성을 재는 지수가 아니라고 비판한다. 시즈즈마에 따르면, 내적 일관성이나 단일 차원성이 낮아도 크론바흐 α 는 얼마든지 높을 수 있으며, 그 역도 성립한다. 슈미트(Schmitt, 1996)는 동질성 개념과 구분되는 개념으로서 내적 일관성을 '항목들 내에서의 상호

따르면, BCFSPA 방법은 R에서만 분석할 수 있다는 한계에도 불구하고 작은 표본 크기에서 중간 크기의 표본 크기, 다수준(multilevel) 모델, 비정규 변인이 포함된 데이터 등 다양한 설정에서 수행결과가 우수한 것으로 보고되었다. 이에 대한 논의와 분석을 위한 R syntax는 켈시(Kelcey, 2019)의 연구를 참고하길 바란다.

관계성(interrelatedness)'으로 규정하고, 내적 일관성이 구성 개념의 동질성을 평가하는 단일 차원성 지수로는 적절하지 않음에도 불구하고 크론바흐(1951)가 내적 일관성과 동질성을 동일 개념으로 사용하고 있다고 지적하였다. 또한 크론바흐 α 는 신뢰도를 과소 추정하거나 측정 가능한 신뢰도의 범위, 즉 α 계수가 1을 넘는 등과 같은 통계적 약점도 비판의 대상이 되고 있으며 (예: Sijtsma, 2009; Schmitt, 1996; Ten Berge & Sočan, 2004), 크론바흐 α 가 심리측정 척도에서는 거의 충족될 수 없는 가정인 타우-등가 모형(tau-equivalent model) 가정에 의존한다는 것, 척도 개발 과정에서 '항목이 삭제되었을 때 알파(alpha when items deleted)'를 산출한 후 그 값을 최종 지표로 보고한다는 것, 그리고 크론바흐 α 가 점 추정치(point estimation)로 보고된다는 것 등으로도 많은 비판을 받고 있는 현실이다(Dunn, Baguley, & Brunsten, 2014).

그러나 이러한 개념의 혼돈과 방법론적·통계적 결함에 대한 비판에도 불구하고, 크론바흐 α 에 대한 국내 커뮤니케이션 연구자들의 신뢰는 절대적이다. 본 연구의 결과에 따르면, 신뢰도 지표로서 크론바흐 α 계수를 사용한 연구는 전체의 약 93.6%로 타 방법에 비해 압도적으로 많이 사용되고 있다. 이러한 경향은 국내 커뮤니케이션 연구뿐만 아니라 사회과학 분야 전반에 걸쳐 나타나는데, 이에 대해서는 여러 이유가 있을 수 있겠지만, 다중항목 신뢰도 척도로 크론바흐 α 만을 제시³⁾ 하는 SPSS의 과도한 사용이 주된 이유일 수 있다(Borsboom, 2006; McNeish, 2018). 사실, 오래전부터 '최대 하계(greatest lower bound: Bentler & Woodward, 1980; Jackson & Agunwamba, 1977)'나 '맥도날드의 오메가(McDonald's ω : McDonald, 1999)' 등과 같은 크론바흐 α 의 대안 지수에 대한 많은 논의가 있었고, 대안 지수를 추정하기 위한 새로운 통계 프로그램이 발전해 왔음에도 불구하고, 크론바흐 α 에 대한 국내외 연구자들의 절대적 신뢰는 변하지 않은 듯 보인다. 크론바흐 α 가 가진 대중적인 인기의 이유는 보르스붐(Borsboom, 2006)의 말처럼 "크론바흐 α 가 특정 인기 통계 프로그램의 마우스 클릭 시퀀스 상의 기본 옵션"(p. 433)이기 때문일지도 모른다.

이와 같이 특정 통계 소프트웨어 사용에 의해 개선되지 않는 여러 통계적 문제는 대안적인 통계 소프트웨어의 사용을 통해 해결할 수 있다. 과거에는 일반 연구자가 대중적으로 접근 가능한 통계 소프트웨어인 SPSS 혹은 SAS에 의존할 수밖에 없었다면, 최근에는 다양한 통계 소프트웨어(예: R, python, Jamovi, JASP 등)가 오픈소스로 제공되고 있어 연구자의 선택의 폭

3) 물론 크론바흐 α 이외에도 spss가 신뢰도 지표로서 split-half, guttman, parallel 등을 산출해주는 하지만, 이들은 다중 항목의 내적 일관성 척도로서 일반적으로 사용되는 지표는 아니다.

이 넓어졌다. R과 Python의 경우 처음 익숙해지는 데에는 시간이 소요되지만 t 검정부터 머신러닝까지 다양한 통계방법을 상황에 맞춰 사용할 수 있으며, 온라인 지원 커뮤니티 형성이 잘 되어 있어 통계분석에 관한 도움을 구하기도 쉽다는 장점이 있다. 또한 SPSS와 같이 그래픽 인터페이스에 익숙한 연구자의 경우 R을 기반으로 개발된 Jamovi나 JASP과 같은 그래픽 기반의 무료 통계 소프트웨어가 좋은 대안이 될 수 있을 것이다. Jamovi의 경우 앞서 다뤘던 부분 η^2 의 대안인 표준화 η^2 와 크론바흐 α 의 대안인 맥도날드의 ω 를 산출할 수 있으며, JASP의 경우는 베이저안(bayesian) 통계방법을 지원한다. 또한 이 두 통계 소프트웨어는 구조방정식모형을 보다 쉽고 직관적으로 지원해주고 있다.

앞서 언급한 문제 이외에도 본 연구는 ‘통계적 방법에 관한 중요한 정보를 연구자에게 공개하지 않는’ 잘못된 관행에 대해 이야기하고자 한다. 이러한 관행은 각각의 통계분석방법 전반에 걸쳐 나타나고 있는데 본 연구 결과를 통해 t 검증, 분산분석, 회귀분석 등을 수행하기 전 필수적 통계 가정을 검증한 내용을 보고한 연구는 많지 않았다는 것을 알 수 있었다. 또한 회귀분석이나 SEM 분석 결과를 보고할 경우 후속 연구 및 메타분석을 위해 중요 변인들 간의 상관관계 행렬을 제시하는 것이 필수적임에도 불구하고(Boster, 2002; Wilkinson & APA Task Force on Statistical Inference, 1999), 전체 274건의 회귀분석 연구 중 35.4%, 전체 134건의 SEM 연구 중 47%만이 상관관계 행렬을 제시하고 있었다. 상관관계 행렬의 보고에 대해서는 <한국언론학보> 논문 작성 규정에도 권고사항으로 제시되어 있음을 고려할 때, 이러한 비율은 매우 낮은 것이라고 할 수 있다.

끝으로, NHST 사용의 문제에 관해 논의할 필요가 있다. 앞서 언급한 바와 같이 NHST는 여러 학자들로부터 지속적으로 비판을 받아왔다. 이에 대해 현재 여러 사회과학분야의 연구자들은 NHST 사용에 대한 자성과 대안의 필요성에 대해 점점 목소리를 높이고 있다. NHST와 관련하여 가장 큰 오해는 NHST의 검증결과로 얻어진 p 값을 영가설이 참이라는 확률로 해석하는 것이다. 즉, $p < .05$ 라는 의미는 영가설이 참인 확률이 5% 미만인 것이 아니라, 영가설이 참일 때 해당 가설의 결과가 일어날 확률이 5% 미만이라는 것을 말한다. 예를 들어, $p < .001$ 라는 것은 영가설이 참일 때 얻어진 결과의 기대 확률이 천분의 1보다 작다는 것을 의미하는 것이다(Nickerson, 2000). 다시 말하면, p 값은 ‘유의성의 측정치(measures of significance)’가 아니라, 영가설이 참일 때, 검증 통계치의 조건적 확률을 말하는 것이다(Bakan, 1966). 적지 않은 수의 연구자들이 p 값이 작으면 작을수록 효과의 크기가 큰 것이라는 잘못된 믿음을 갖고 있지만(Nickerson, 2000), p 값은 효과의 크기를 보여주는 지표가 아니다(Cohen, 1994; Rosenthal, 1991). 우리가 잘 알고 있는 바와 같이 p 값은 표본의 크기에 매우 민감한데, 만일

표본 수가 작으면 p 값이 높아져 영가설을 기각할 수 없더라도 효과크기는 클 수 있고, 이러한 경우 제2종 오류(Type II error)는 증가하게 된다. NHST가 가진 여러 문제들과 관련하여 1999년 미국심리학회(American Psychological Association)의 과학학술위원회(APA Board of Scientific Affairs)는 태스크포스 팀(Task Force on Statistical Inference)을 구성하여 NHST의 폐기 여부에 대해 논의하였다. 물론 이 논의의 결과로 NHST를 폐기하라는 권고가 나오지는 않았지만, NHST를 이용한 연구 결과의 분석에서는 반드시 NHST를 보완할 수 있는 여러 방법들을 함께 제시하도록 권고하였으며(Wilkinson & Task Force on Statistical Inference, 1999), NHST를 보완할 수 있는 방법으로는 ‘효과크기 제시’, ‘신뢰구간 제시’, ‘그래픽의 형태 제시’, ‘통계적으로 유의미하지 않은 결과도 효과크기 제시’ 등이 여러 학자들에 의해 제시되었다(예: 이병관 외, 2015; Cohen, 1994; Levine et al., 2008; Nickerson, 2000; Thompson, 2002; Wilkinson & Task Force on Statistical Inference, 1999). 2015년 2월에는 사회심리학 저널인 *Basic and Applied Social Psychology*에서 NHST가 분석방법 및 해석으로 타당하지 않으며, 향후 NHST 사용을 금지하는 내용을 담은 새로운 편집 규정을 공지하였는데, 이 편집 규정은 향후 이 저널에 투고하는 모든 논문은 p 값, t 값, F 값, ‘통계적으로 유의미한 이라는 표현’ 등을 사용할 수 없도록 규정하고 있다(Trafimow & Marks, 2015). <한국언론학보> 또한 2015년부터 논문 작성 규정 개정을 통해 모든 연구자들은 NHST와 함께 신뢰구간도 반드시 제시할 것을 규정하고 있다. 그러나 이러한 규정에도 불구하고 큰 변화는 나타나지 않았다. 2015년을 기점으로 개정 전, 후의 신뢰구간 제시의 변화를 살펴본 결과, 2015년 이전에 수록된 386건의 연구 중 8건(2.1%)이, 2015년 이후에 수록된 236건의 연구 중 34건(14.4%)만이 신뢰구간을 제시하고 있었다. 이러한 결과는 우리 연구자들과 학계 모두가 최근 진행되고 있는 NHST와 관련한 변화에 대해 여전히 소극적이라는 것을 방증하는 것이다.

본 연구에서 도출된 연구 결과와 논의는 향후 커뮤니케이션 연구를 위해 통계분석방법을 수행하는 연구자에게 반드시 고민해야 할 중요한 교훈을 준다는 점에서 의의를 갖는다. 그럼에도 불구하고 본 연구는 개별 통계분석방법에 대한 구체적인 문제점과 해결책을 제시하는 것이 미흡했다는 한계를 지닌다. 다만, 본 연구는 개별 분석방법에 있어서의 여러 구체적인 문제점과 대안을 제시하는 것 보다는 큰 관점에서의 현상과 문제점을 파악하는 것이 주된 목적이었으며, 개별 분석방법에 있어서의 구체적인 문제점과 대안은 또 하나의 연구 주제라고 할 수 있다. 구조방정식의 예를 들면, 본 연구는 구조방정식 사용에 있어서 ‘입력 자료로 사용한 행렬의 보고’, ‘모수추정 방식의 보고’, ‘통계 가정의 점검 여부’, ‘사용된 통계 패키지’ 등을 살펴보았다. 그러나 구조방정식이라는 단일 분석방법과 관련한 연구만 해도 방대한 주제를 가지고 진행되고 있으며(예: 구

조방정식 일반: 강남준, 1999; 이기중, 2016; Baumgartner & Homburg, 1996; Holbert & Stephenson, 2002, 구조방정식 연구의 보고: Boomsma, 2000; Raykov, Tomer, & Nesselroade, 1991 등), 구조방정식의 모델 적합도 사용과 관련해서도 많은 연구(예: 변인의 수와 모델 적합도와의 관계: Kenny & McCoach, 2003, fit index의 골든 룰: Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004, GFI와 표본 수: Marsh, Balla, & McDonald, 모델 오추정과 모델 적합도와의 관계: Beauducel & Wittmann, 2005, 모델의 구조: Ding, Velicer, & Harlow, 1995 등)가 수행되고 있다. 이는 비단 SEM에만 국한되지 않고, 분산분석이나 회귀분석 등과 같은 다른 분석방법에서도 동일할 것이다. 따라서 본 연구와 같이 커뮤니케이션 연구에서 사용되는 모든 분석방법에 관한 논의에서 특정 방법에 대해 자세히 논의를 하는 것은 연구의 범위로 보나 지면의 한계로 보나 현실적으로 한계가 있어 본 연구에서는 특정 분석방법의 문제를 제시하고 그 대안에 대한 독자의 이해를 위해 몇 가지 사례만을 언급할 수밖에 없었다. 우리 학계에서 개별 통계분석방법에서 제기되는 문제점과 대안에 대한 연구의 부족에는 다양한 이유가 있겠지만, 향후 국내 커뮤니케이션 연구자들도 개별 통계분석방법에서 제기되는 여러 문제점과 대안에 대한 연구들을 적극적으로 수행함으로써 이에 대한 학문적 논의가 풍성해지길 바란다. 이를 위해서 엄격한 과학적 절차를 통해 자료를 수집 및 분석하고, 이를 해석한 결과를 통해 연구자의 가설을 추론하고 논의하는 연구자의 노력과 함께 이러한 연구를 지속적으로 장려하고 학술지에 적극적으로 심사하고 게재하려는 학계의 관심과 노력이 필수적으로 수반되어야 할 것이다.

국내 커뮤니케이션 분야의 통계분석의 사용에 있어 발견된 여러 문제는 오래전 호텔링, 바트키, 데밍, 프리드만, 그리고 호텔(Hotelling, Bartky, Deming, Friedman, & Hoel, 1948)이 잘못된 통계 사용의 관행에 대해 비판한 인용구를 떠오르게 한다. “불행히도 너무도 많은 사람들이 마치 기도하듯이 통계를 수행하고 있다. 그 기도는 아주 오래전 쓰여진 존귀한 책에 나와 있는 공식만을 단지 읊을 뿐이다.”(Hotelling et al., 1948, p. 103). 이 인용구는 엄격한 기준과 비판적인 검토 없이 기존의 관행만을 의지한 채 통계분석을 수행하는 연구자의 무능과 부주의함을 비판하고 있다. 통계의 오용이 의도적 기만이라는 연구자의 윤리적 해이를 반드시 수반하는 것은 아니다. 그럼에도 우리는 통계의 오용을 실수, 무능, 편견, 수용 가능한 관행으로부터의 심각한 이탈로서 이해해야 한다(Gardenier & Resnik, 2002). 엄격한 과학적 절차를 따르지 않은 채 편의로 자료를 수집 및 분석하고, 이를 해석한 결과를 통해 연구자의 가설을 추론하고 논의하는 것은 무의미하며, 이러한 상황에서 연구자가 밝히고자 하는 ‘진실’은 이미 연구자가 설명할 수 없는 전혀 다른 곳에 존재하게 될지도 모른다(이병관 외, 2015).

References

- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *10*, 411-423.
- Bailar, J. C. & Mosteller, F. (1988). Guidelines for statistical reporting in articles for medical journals: amplifications and explanations. *Annals of Internal Medicine*, *108*(2), 266-273.
- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, *66*, 1-29.
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*(3), 379-384.
- Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, *13*, 139-161.
- Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*, *12*(1), 41-75.
- Bentler, P. M., & Woodward, J. A. (1980). Inequalities among lower bounds to reliability: With applications to test construction and factor analysis. *Psychometrika*, *45*(2), 249-267.
- Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling*, *7*, 461-483.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425-440.
- Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample size effects on chi square and other statistics used in evaluating causal models. *Journal of Marketing Research*, *19*, 425-430.
- Boster, F. J. (2002). On making progress in communication science. *Human Communication Research*, *28*(4), 473-490.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of co-variance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62-83.
- Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural Equation Modeling* (pp. 37-59). Thousand Oaks, CA: Sage.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*(12), 997-1103.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.
- Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, *64*(3), 391-418.

- Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure modeling* (pp. 195-223). Mahwah, NJ: Erlbaum.
- Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement, 76*(5), 741-770.
- Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling, 2*, 119-144.
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399-412.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2-18.
- Gardenier, J., & Resnik, D. (2002). The misuse of statistics: concepts, tools, and a research agenda. *Accountability in Research: Policies and Quality Assurance, 9*(2), 65-74.
- Graham, J. W. (2012). *Missing data: Analysis and design*. New York: Springer Science & Business Media.
- Holbert, R. L., & Stephenson, M. T. (2002). Structural equation modeling in the communication sciences, 1995-2000. *Human Communication Research, 28*(4), 531-551.
- Hotelling, H., Bartky, W., Deming, W. E., Friedman, M., & Hoel, P. (1948). The teaching of statistics [a Report of the Institute of Mathematical Statistics Committee on the Teaching of Statistics]. *Annals of Mathematical Statistics, 19*, 95-115.
- Hsieh, F. Y. (1989). Sample size tables for logistic regression. *Statistics in Medicine, 8*(7), 795-802.
- Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*(2), 351-362.
- Huang, Y., & Bentler, P. M. (2015). Behavior of asymptotically distribution free test statistics in covariance versus correlation structure analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(4), 489-503.
- Hunter, J. E. (1997). Needed: A ban on the significance test. *Psychological Science, 8*(1), 3-7.
- Jackson, P. H., & Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: Algebraic lower bounds. *Psychometrika, 42*(4), 567-578.
- Kang, N. (1999). The problems of using structural equation model(SEM) in communication research. *Korean*

Journal of Journalism & Communication Studies, 44(1), 5-51.

- Kelcey, B. (2019). A robust alternative estimator for small to moderate sample SEM: Bias-corrected factor score path analysis. *Addictive Behaviors*, 94, 83-98.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York, NY: The Guilford Press.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.
- Lee, B., Sohn, Y., Oh, H., & Park, S. (2015). Current status and future direction on the use of statistical methods in quantitative research papers in public relations: A systematic analysis of journal of public relations, 1998~2004. *Journal of Public Relations*, 19(1), 270-298.
- Lee, B., Sohn, Y., & Kang, K. (2018). How Reliable is the Reliability of the Celebrity Endorsement Scales?: Applying the Reliability Generalization Meta-Analysis. *The Korean Journal of Advertising and Public Relations*, 20(3), 35-77.
- Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research*, 28(4), 612-625.
- Levine, T. R., Weber, R., Hullett, C., Park, H. S., & Lindsey, L. L. M. (2008). A critical assessment of null hypothesis significance testing in quantitative communication research. *Human Communication Research*, 34(2), 171-187.
- Lu, I. R., Kwan, E., Thomas, D. R., & Cedzynski, M. (2011). Two new methods for estimating structural equation models: An illustration and a comparison with two established methods. *International Journal of Research in Marketing*, 28, 258-268.
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual review of psychology*, 51(1), 201-226.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103(3), 391.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York, NY: The Guilford Press.
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412-433.
- Nam, K. (2004). A study of choice for statistical method on social science. *Journal of the Korean Data*

Analysis Society, 6(5), 1255-1265.

Neuman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. *Biometrika*, 20A(1/2), 175-240.

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological Methods*, 5(2), 241-301.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8, 434-447.

Park, J. (2015). Bayesian statistics as a solution to the problems of NHSTP: The case of psychology. *The Korean Journal for the Philosophy of Science*, 18(2), 135-147.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12), 1373-1379.

Pelsma, J. R. (1937). *Essentials of debate*. New York, NY: Thomas Y. Crowell Company.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>

Reinsel, D., Gantz, J., & Rydning, J. (2018). *Data age 2025: the digitization of the world from edge to core*. Retrieved from <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74(1), 145.

Rhee, K. (2016). Wrong applications with overall model evaluations and its corrections in structural equation modeling. *Survey Research*, 17(1), 71-83.

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6(2), 135-147.

Ritter, N. L. (2010). *Understanding a widely misunderstood statistic: Cronbach's*. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans. Retrieved from <https://files.eric.ed.gov/fulltext/ED526237.pdf>.

Rosenthal, R. (1991). Effect sizes: Pearson's correlation, its display via the BESD, and alternative indices. *American Psychologist*, 46, 1086-1087.

Schmidt, F. L., & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of

- significance testing in the analysis of research data. In L. L. Harlow, S. A. Mulaic, & J. H. Steiger (Eds.), *What if there were no significance tests* (pp 37-64). London, UK: Psychology Press.
- Shevlin, M., Miles, J. N. V., Davies, M. N. O., & Walker, S. (2000). Coefficient alpha: a useful indicator of reliability? *Personality and Individual Differences*, 28(2), 229-237.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350-353.
- Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, 66, 563-576.
- Thompson, B. (2002). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling & Development*, 80(1), 64-71.
- Ten Berge, J. M., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69(4), 613-625.
- Trafimow, D. & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, 37(1), 1-2.
- Wilkinson, L., & Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604.
- Yang, Y., and Green, S. B. (2011). Coefficient Alpha: a reliability coefficient for the 21st Century? *Journal of Psychoeducational Assessment*, 29, 377-392.

최초 투고일 2020년 8월 3일
 게재 확정일 2020년 9월 28일
 논문 수정일 2020년 10월 5일

부록

- 강남준 (1999). 커뮤니케이션 연구에서 구조방정식 모형(SEM)의 활용가능성: SEM 사용의 문제점을 중심으로. <한국언론학보>, 44권 1호, 5-51.
- 남기성 (2004). 사회과학에서 통계분석방법의 선택에 관한 연구. <한국자료분석학회>, 6권 5호, 1255-1265.
- 박준석 (2015). 영가설 유의성검증 절차의 문제점들에 대한 해결책으로서의 베이지언 통계학: 심리학의 경우. <과학철학>, 18권 2호, 135-147.
- 이기중 (2016). 구조방정식모형의 모형평가 오, 남용과 교정. <조사연구>, 17권 1호, 71-83.
- 이병관·손영곤·강경희 (2018). 유명인 모델의 공신력 척도는 얼마나 신뢰할 수 있는가?: 신뢰도 일반화 메타 분석의 적용. <한국광고홍보학보>, 20권 3호, 35-77.
- 이병관·손영곤·오현정·박선화 (2015). 통계분석방법의 경향과 과제: 1998년 창간호부터 2014년 여름호까지 게재된 논문의 내용분석. <홍보학연구>, 19권 1호, 270-298.

국내 커뮤니케이션 연구의 통계분석방법의 현황과 문제점 지난 10년간 <한국언론학보> 게재 논문의 내용분석

이병관

(한양대학교 광고홍보학과 교수)

김재민

(한양대학교 광고홍보학과 석사과정)

김주환

(한양대학교 광고홍보학과 석사과정)

장다연

(한양대학교 광고홍보학과 석사과정)

권나현

(한양대학교 광고홍보학과 석사과정)

본 연구는 2010년부터 2019년까지 <한국언론학보>에 게재된 양적 연구 논문을 대상으로 통계 분석 방법 사용의 현황과 경향을 살펴보고, 이를 통해 국내 커뮤니케이션 연구의 통계 사용의 관행과 문제점을 논의하는 것에 그 목적이 있다. 이를 위해 10년간 <한국언론학보>에 게재된 총 845편의 논문 중 통계분석방법을 사용한 612편을 대상으로 내용분석을 진행하였다. 분석 결과, 지나치게 특정 통계분석방법에 의존하거나, 해석의 문제, 인위적인 결과를 생산하기 위한 통계적 방법의 오용 현상이 발견되었다. 이와 함께, 대부분의 연구들에게서 통계 방법에 관한 중요한 정보를 제시하지 않는 잘못된 관행도 발견되었다. 본 연구는 통계 분석 방법 및 보고에 대해 지금까지 많은 통계 학자들로부터 지속적으로 제기 되어온 비판과 논의를 실증적으로 확인했다는 점에서 그 의미를 가진다. 도출된 결과를 바탕으로 통계 분석의 오용, 남용, 관행적 사용 등에 대한 문제점이 논의되었다.

핵심어: 통계 분석 방법, 양적 연구, 커뮤니케이션 연구, 내용 분석