

모의실험을 통한 표본 강의평가제의 실현 가능성 탐구

The Practicability of the Sample Course Evaluation System through Simulation

김용태*, 김성윤**, 이상준***

세명대학교 교양대학*, 배화여자대학교 경영과**, 세명대학교 정보통신학부***

Yong-Tae Kim(dataminer@semyung.ac.kr)*, Seong-Yoon Kim(donakim@baewha.ac.kr)**,
Sang-Jun Lee(leesangjun@semyung.ac.kr)***

요약

현재 국내 대학들이 강의평가의 저조한 참여율을 보완하기 위해 부가한 강제성은 학생들의 불성실한 응답을 야기하여 평가결과의 타당성을 크게 저해하는 요인으로 지적되고 있다. 본 연구는 전수조사로 진행되는 강의평가제의 문제점을 지적하고 표본 강의평가제를 그 대안으로 제시하면서 표본평균을 이용해 모평균을 추론하는 것이 통계적으로 의미 있다는 것을 모의실험을 통해 확인하고자 한다. 이를 위해 2016년 2학기 S 대학에서 개설된 강좌의 소형, 중형, 대형, 사이버 강좌를 모집단으로 가정하였고, 각 강좌의 강의평가 평균점수를 모평균으로 가정하였다. 그리고 표본수의 결정을 위하여 직전년도 동일 교과목에 대한 평균과 표준편차를 사용하였고, 신뢰수준은 95%, 오차한계는 ± 0.25 로 하였다. 데이터 분석도구인 R을 이용한 모의 실험 시행결과 소형, 중형, 대형, 사이버 강좌 모두 모평균을 포함하는 신뢰수준이 95% 이상 근사하는 것으로 나타나 대표성 있는 표본 강의평가제의 실현 가능성을 확인하였다.

■ 중심어 : | 모의실험 | 강의평가제 | 표본조사 | 신뢰수준 | 표본크기 |

Abstract

The current system of making course evaluations mandatory in universities in Korea in order to supplement low participation levels is a major hindering factor for the validity of the evaluation results due to students' insincere responses. This study points out the problems in course evaluations that are in the form of surveys and instead proposes a sample course evaluation system that utilizes a sample mean to deduce a population mean, attempting to prove through simulation that this is statistically significant. Thus, small, medium, and large scale courses as well as cyber courses that were available in S University in the fall 2016 semester were set as the population, and each course's evaluation average grade was set as the population mean. Furthermore, we used the average grade and standard deviation of the same courses from the previous year in order to decide the sample number, the reliability level was set as 95%, and the margin of error was set as ± 0.25 . As a result of carrying out a simulation using the data analysis tool R, all the courses (small scale, medium scale, large scale, and cyber) showed to have a reliability level close to 95% including the population mean, and consequently practicality of the representative sample course evaluation system was proven.

■ keyword : | Simulation | Course Evaluation | Sample Survey | Confidence Level | Sample Size |

* 이 논문은 2017학년도 세명대학교 교내학술연구비 지원에 의해 수행된 연구임

접수일자 : 2017년 11월 02일

심사완료일 : 2017년 12월 13일

수정일자 : 2017년 12월 13일

교신저자 : 이상준, e-mail : leesangjun@semyung.ac.kr

I. 서론

1990년 중반부터 도입된 강의평가제는 초기 찬반의 견이 활발히 논의되었지만 1997년 대학종합평가의 평가항목으로 강의평가제 적용여부가 공식적으로 포함되면서 현재 대부분의 대학에서 실시하고 있다. 특히 수업에 대한 가치판단의 수단뿐만 아니라 대학교육의 질 향상이 교육개혁의 주요 요인으로 부각되면서 강의평가의 중요성과 이에 대한 관심은 점점 높아지고 있다 [1]. 강의평가의 목적은 크게 세 가지로 구분할 수 있는데 첫째, 강의평가 결과를 교수들에게 제공하여 강의를 개선하고, 둘째, 교수 효과성을 정확히 측정하여 보상과 처벌의 기초자료로 활용하며, 셋째, 학생들에게 교과목을 선택할 때 도움이 되고자 교수와 교과목에 대한 정보를 제공하기 위함이다[2].

정보통신의 발달과 더불어 강의평가제는 기존의 설문지 방식에서 벗어나 온라인 방식의 체제로 변화되었다. 현재 국내 대부분의 대학에서 온라인 방식으로 진행하고 있는 강의평가는 성적열람 또는 수강신청의 선행조건으로 학생들이 강의평가에 의무적으로 참여해야 하는 강제성을 부여하고 있다. 이에 따라 학생들은 수강했던 모든 교과목을 평가하는 과정에서 성적열람에 대한 조바심, 응답에 대한 피로도 등으로 인해 무성의한 응답과 왜곡된 응답을 하게 되고, 강의평가 결과를 활용해야 하는 학교의 입장에서 학생들의 성실한 참여를 절실히 요구하고 있는 실정이다.

강의평가의 저조한 참여율을 보완하기 위해 부가한 강제성은 모든 문항에 대해 동일한 선택지에 응답하는 등 불성실한 응답비율을 높게 하는 요인으로 보고되고 있다[3][4]. 한 대학교 신문사의 강의평가에 관한 설문조사 결과 34%의 학생들이 강의평가의 질문을 읽지도 않고 버튼만 대충 누른다고 하였고, 강의평가에 제시된 모든 문항에 똑같은 점수로 응답한 학생들의 일관적 응답비율이 50% 이상 되는 경우도 다수 발견되고 있다 [5]. 또한 이러한 문제점에 대한 심층면담 결과 반복되는 질문으로 인해 무성의한 응답, 성적확인을 위한 형식적 응답(동일한 번호로 찍기, 생각 없이 만점주기) 등의 부정적인 의견들이 제시되었대[6].

본 연구는 현재의 강제적인 강의평가 방식으로 얻은 결과를 강의평가분석 및 개선점 도출을 위한 유용한 데이터로 사용되는 것은 문제가 있다고 보고 전수조사로 진행되는 강의평가제 대신 대표성 있는 표본 강의평가제를 제안하고자 한다. 보다 구체적으로 1) 대학은 강좌 크기에 따른 최적의 표본수를 산출하여 각 강좌에 참여한 학생들을 추출하고 2) 학생은 해당 학기에 수강한 모든 교과목에 대해 강의평가를 하는 것이 아니라 본인이 표본으로 선정된 최소 강좌에 대해서만 강의평가를 하는 것이다. 3) 이렇게 되면 모든 교과목에 대한 평가를 하는 것이 아니기 때문에 학생 스스로 강의평가에 대한 부담감을 줄일 수 있고, 강의에 대한 대표성을 가진 책임감 있는 응답을 할 것으로 유추할 수 있다.

이를 위해 강좌를 규모별로 소형, 중형, 대형, 사이버 강좌로 구분하고, 각 모집단 N 명 중에서 완전 랜덤방식으로 표본 n 명을 선택하여 이 n 명의 표본평균으로 모평균을 추론하는 것이 통계적으로 의미 있다는 것을 모의실험의 신뢰구간 추론을 통해 확인하고자 한다. 즉 신뢰구간 추론은 신뢰수준 $(1-\alpha) \cdot 100\%$ 에서 수행되므로 이러한 동일한 실험(모집단 N 에서 표본 n 을 추출하여 모평균의 신뢰구간을 추정)을 m 회 반복하면 이 반복한 m 번의 실험 중에서 모평균 μ 를 포함할 비율이 약 $(1-\alpha)$ 가 될 것이다. 따라서 신뢰구간이 모평균 μ 를 포함하는 비율이 $(1-\alpha)$ 로 근사한다면 표본평균을 이용하여 모평균을 추론하는 것이 통계적으로 의미가 있다고 할 수 있을 것이다.

II. 이론적 배경

강의평가는 대학교 강의의 질적인 효율성을 목적으로 타당한 평가도구를 사용하여 강의자의 교수학습 활동의 정보를 수집하고 교수 프로그램에 대한 가치를 판단하는 교육적인 의사결정이다[7]. 또한 강의평가는 강의의 질을 진단·평가하여 이를 개선하기 위한 정보를 수집하는 형성적 목적과 교육프로그램의 지속성 여부 및 교수의 승진이나 재임용, 정년보장 등에 활용하기 위한 총괄적 목적으로 활용되고 있는데[1], 누가 평가를

받으며 누가 평가하는지, 무엇을 평가해야 하는지, 어떤 평가도구를 어떤 방법으로 시행해야 하는지, 어떤 기준으로 평가해야 하는지, 평가결과는 어떻게 처리되며 어떻게 활용되는지, 평가결과는 타당하고 신뢰할만한지 다양한 관점에서 연구들이 이루어지고 있다[8].

2000년대에 들어와 국내 대학의 강의평가가 온라인 방식으로 일반화되고 학생들의 응답률을 높이기 위해 강제하는 방식으로 전환되면서 학생들의 무성의한 응답의 문제는 평가결과의 타당성과 신뢰성에 대한 근본적인 의문을 제기하고 있다[9]. 이러한 학생의 무성의한 응답은 학생의 성별, 학년, 전공, 성취도, 공부시간, 수업 난이도, 수강환경, 교수 연령 등과 함께 강의평가 점수에 통계적으로 유의미한 영향을 미치는 요인으로 제시되고 있다[10]. 미국의 주요 대학들이 온라인의 편리함과 학생들의 참여도를 높이는 장점에도 불구하고 아직까지 인쇄물과 OCR 카드로 평가하고 있는 것을 고려해보면 성실한 응답을 유도하기 위한 방법을 재고해야 할 필요성이 있다[11].

모든 학생을 대상으로 하는 온라인 강의평가는 전수 조사의 성격을 가진다. 전수조사보다 표본조사에서 일관적 응답패턴이 줄어들고 신뢰도가 높아졌다는 연구 결과에서 확인할 수 있듯이 전원 강의평가 참여라는 양적인 지표 대신에 표본조사를 통한 성실한 응답을 유도하려는 노력이 필요하다[12]. 강의개선의 효과를 체험하지 못한 학생들은 강제적인 강의평가에 무성의하게 한 번으로 응답하여 비표본오차를 크게 만든다. 무성의한 일관적 응답은 내적일관성을 높여 신뢰도계수를 매우 크게 만들기 때문에 이런 강의평가 결과를 가지고 수업을 개선하여 학생들에게 피드백하고 교수업적을 평가하는 것은 사실상 의미가 없다. 따라서 강의평가 자료를 분석하기 전에 학생들이 얼마나 성실하게 응답한 자료인지 먼저 강의평가 데이터의 품질을 확인하는 것이 필요하다[13].

III. 실험설계

1. 모수추정

유한 모집단에서 표본수의 결정은 모평균과 모비율의 추정에 따라 달라지는데, 모평균을 추정할 때는 수식 (1)을 사용하고 모비율을 추정할 때는 수식 (2)를 사용한다[14][15].

$$n = \frac{N(z_{\alpha/2} s)^2}{Nd^2 + (z_{\alpha/2} s)^2} = \frac{(z_{\alpha/2} s)^2/d^2}{1 + \frac{(z_{\alpha/2} s)^2/d^2}{N}} \quad (1)$$

$$n = \frac{[z_{\alpha/2}^2 p(1-p)]/d^2}{1 + \frac{[z_{\alpha/2}^2 p(1-p)]/d^2}{N}} \quad (2)$$

여기에서 d 는 신뢰수준 $(1-\alpha) \cdot 100\%$ 에서 오차한계이며, s 는 표본 표준편차, p 는 모비율, $z_{\alpha/2}$ 는 신뢰계수, N 은 모집단의 크기이다. 일반적으로 신뢰수준은 95%를 많이 사용하므로 본 연구에서도 신뢰수준은 95%로 정의하였다. 그리고 표본 표준편차는 사전에 알 수 없으므로 사전조사의 결과나 이전조사의 결과를 이용할 수 있으며, 모비율 p 도 사전조사의 결과나 분산을 최대화할 수 있는 0.5를 이용할 수 있다.

그리고 모평균의 신뢰구간 추정을 통하여 표본평균을 이용한 모평균의 추론이 타당한 것인지 판단하기 위한 모평균의 $(1-\alpha) \cdot 100\%$ 신뢰구간은 다음 수식 (3)과 같다.

$$P\left(-t_{\frac{\alpha}{2}}(n-1) \leq \frac{\bar{x} - \mu}{\sqrt{\left(\frac{N-n}{N}\right) \frac{s^2}{n}}} \leq t_{\frac{\alpha}{2}}(n-1)\right) = 1 - \alpha \quad (3)$$

이러 수식 (3)에서 $\frac{\bar{x} - \mu}{\sqrt{\left(\frac{N-n}{N}\right) \frac{s^2}{n}}}$ 을 모평균(μ)만 남

기고 나머지를 부등호 왼쪽과 오른쪽 항으로 이항하여 정리하게 되면 신뢰하한 수식 (4), 신뢰상한 수식 (5)와 같은 모평균의 신뢰구간을 얻을 수 있다.

$$L = \bar{x} - t_{\alpha/2}(n-1) \sqrt{\left(\frac{N-n}{N}\right) \frac{s^2}{n}} \quad (4)$$

$$U = \bar{x} + t_{\alpha/2}(n-1) \sqrt{\left(\frac{N-n}{N}\right) \frac{s^2}{n}} \quad (5)$$

즉, $P(L \leq \mu \leq U) = 1 - \alpha$ 을 이용하여 신뢰구간을 산출할 수 있고, 표준오차는 유한모집단 수정계수 $\left(\frac{N-n}{N}\right)$ 을 곱한 $\sqrt{\left(\frac{N-n}{N}\right) \frac{s^2}{n}}$ 이다.

2. 데이터

본 연구는 전수조사로 진행되는 강의평가제의 강의평가점수를 표본 강의평가제를 통해 추정하는 것이 목적이기 때문에 먼저 모평균의 추정에 대한 표본수를 결정해야 한다. 연구자들은 모집단의 크기를 일반적인 강의개설 기준인 1) 10명 이상 30명 미만인 소형 강의, 2) 30명 이상 50명 미만인 중형 강의, 3) 50명 이상인 대형 강의, 4) 대형 강의보다 많은 인원수를 보이며 온라인상에서 수강하는 사이버강의로 구분하였다.

2016년 2학기에 개설된 S 대학의 강좌에 대한 강의평가 결과를 바탕으로 소형, 중형, 대형, 사이버 강좌를 모집단으로 설정하였고, 선택된 강좌의 강의평가 평균점수를 모평균(μ)으로 가정하였다. 이를 위해 2016년 2학기 강의평가 자료에서 소형 강좌($N=15$), 중형 강좌($N=40$), 대형 강좌($N=60$), 사이버 강좌($N=150$)에 대한 모집단의 크기를 결정하였고, 이를 대표할만한 교과목을 인원수를 고려하여 각각 하나씩 선정하였다. 그리고 표본수의 결정을 위하여 직전년도 동일 교과목에 대한 표준편차를 사용하였고, 신뢰수준은 95%, 오차한계는 ± 0.25 로 설정하였다.

3. 실험절차

모의실험은 강의평가제를 표본조사로 진행하더라도 전수조사로 진행되는 강의평가 결과를 대표할 수 있다는 것을 증명하고자 표본평균을 이용해 모평균을 추론하였으며, 다음과 같은 절차로 진행하였다.

- ① N 명에 대한 모평균(μ)을 계산한다.
- ② 모집단 N 명에서 완전 랜덤방식으로 표본 n 명을 선택한다.

- ③ 표본 n 명을 이용하여 모평균(μ)에 대한 신뢰수준 $(1 - \alpha) \cdot 100\% = 95\%$ 에서 구간추정을 한다.
- ④ 이 신뢰구간 내에 모평균(μ)을 포함하고 있으면 1, 아니면 0으로 하여 현재까지 반복한 것에 대해 모평균(μ)을 포함하고 있는 비율(1의 비율은 신뢰수준을 의미)을 계산한다.
- ⑤ 2단계와 4단계의 과정을 m (5,000)회 반복하여 모의실험을 시행한다.

IV. 실험결과

본 연구는 표본에 의한 모평균의 추론이 타당하지 않 아보고자 모집단 N 에서 표본 n 을 선택하여 신뢰수준 $(1 - \alpha) \cdot 100\%$ 에서 모평균(μ)에 대한 신뢰구간을 추정하고, 모평균(μ)가 신뢰구간 내에 포함되는지 살펴 보았는데 구체적인 실험결과는 다음과 같다.

1. 표본수 산출

먼저 소형, 중형, 대형, 사이버 강좌별 표본수의 결정을 위해 직전년도 동일 교과목 자료에서 소형, 중형, 대형, 사이버 강좌의 모평균(μ)과 모표준편차(σ)를 산출하였고, 이후 결정된 모수들을 이용하여 강좌별 표본수(n)를 산출하였다.

[표 1]의 결과에서 보는 바와 같이 소형 강좌의 경우 모평균 4.558, 모표준편차 0.590으로 나타났고, 중형 강좌의 경우 모평균 4.316, 모표준편차 0.763으로 나타났 다. 그리고 대형 강좌의 경우 모평균 4.405, 모표준편차 0.673으로 나타났고, 사이버 강좌의 경우 모평균 4.242, 모표준편차 0.699로 나타났다. 이를 바탕으로 표본 강의평가제의 추론을 위한 강좌별 표본수를 소형 강좌 9명, 중형 강좌 19명, 대형 강좌 20명, 사이버 강좌 26명으로 산출하였다.

```

Small Course.R x
Source on Save
Run
Source
1 S.Data<-read.table("small_data.txt",sep="\t",header=TRUE)
2 head(S.Data)
3 N<-15 # 모집단 크기
4 n<-9 # 표본 크기
5 rp<-5000 # 모의실험 반복수
6 alpha<-0.05 # 유의수준
7 Total<-6-apply(S.Data[,16:27],MARGIN=1,FUN=mean) # 학생별 강의평가 점수
8 pop.mean<-mean(Total) # 모평균
9 s.mean<-rep(NA,rp) # 표본 평균 초기화
10 s.std<-rep(NA,rp) # 표본 표준편차 초기화
11 adj.s.std<-rep(NA,rp) # 유한모집단 수정 표본 표준편차 초기화
12 L.Bound<-rep(NA,rp) # 신뢰하한 초기화
13 U.Bound<-rep(NA,rp) # 신뢰상한 초기화
14 signif.n<-rep(0,rp) # 모평균 포함여부
15 attach(S.Data)
16 for (i in 1:rp) {
17   sample.num<-sample(1:N, n) # 표본번호 선택
18   sample.data<-Total[sample.num] # 선정 표본번호의 강의평가 점수
19   s.mean[i]<-mean(sample.data) # 표본 평균
20   s.std[i]<-sd(sample.data) # 표본 표준편차
21   adj.s.std[i]<-sqrt((N-n)/N)*s.std[i] # 유한모집단 수정 표본 표준편차
22   t.inv<-qt(1-alpha/2,n-1) # t-분포 값
23   L.Bound[i]<-s.mean[i]-t.inv*adj.s.std[i]/sqrt(n) # 신뢰하한
24   U.Bound[i]<-s.mean[i]+t.inv*adj.s.std[i]/sqrt(n) # 신뢰상한
25   if (all(pop.mean>=L.Bound[i], pop.mean<=U.Bound[i])) { signif.n[i]<-1 }
26 }
27 detach(S.Data)
28 signif.p<-cumsum(signif.n)/(1:rp)
29 signif.n
30 signif.p
31 plot(signif.p, type="l")
32 result<-data.frame(pop.mean,s.mean,s.std,adj.s.std,L.Bound,U.Bound,signif.n,signif.p)
33 write.csv(result, file="small_result.txt", row.names=FALSE)
34 mean(signif.n)
34:16 (Top Level) R Script
    
```

그림 1. 모의실험 R 코드

표 1. 표본수 산출

	소형 강좌	중형 강좌	대형 강좌	사이버 강좌
μ	4,558	4,316	4,405	4,242
σ	0,590	0,763	0,673	0,699
N	15	40	60	150
n	8,816 \approx 9	18,893 \approx 19	19,029 \approx 20	25,033 \approx 26

2. 모의실험 R 코드

본 연구는 실제 실험환경을 설정하기 어렵기 때문에 모의실험을 5,000회 진행하였는데, [표 1]에서 산출한 각 강좌별 표본수를 바탕으로 데이터분석 도구인 R을 이용하여 모의실험을 구현하였다[16][17][18]. [그림 1]은 소형 강좌의 모의실험을 위한 R 코드의 예시이고,

구체적인 수행내용은 다음과 같다.

- ① 1행 : 강의평가 자료 불러오기
- ② 3-6행 : 모의실험에 필요한 모집단 및 표본크기, 모의실험 반복횟수, 유의수준 설정하기
- ③ 7-8행 : 강의평가 모평균 설정하기
- ④ 9-15행 : 표본 통계량 및 모의실험 결과 저장공간 초기화하기
- ⑤ 16-26행 : 모집단에서 설정한 표본의 수만큼 표본을 추출하여 평균을 계산하고, 실제 모평균을 포함하고 있는지 판단하는 모의실험 과정을 반복횟수 만큼 실행하기
- ⑥ 28행 : 반복횟수 만큼 모의실험 한 결과 실제 모평균을 포함하고 있는 비율을 계산하기
- ⑦ 31행-34행 : 반복횟수별로 실제 모평균을 포함하

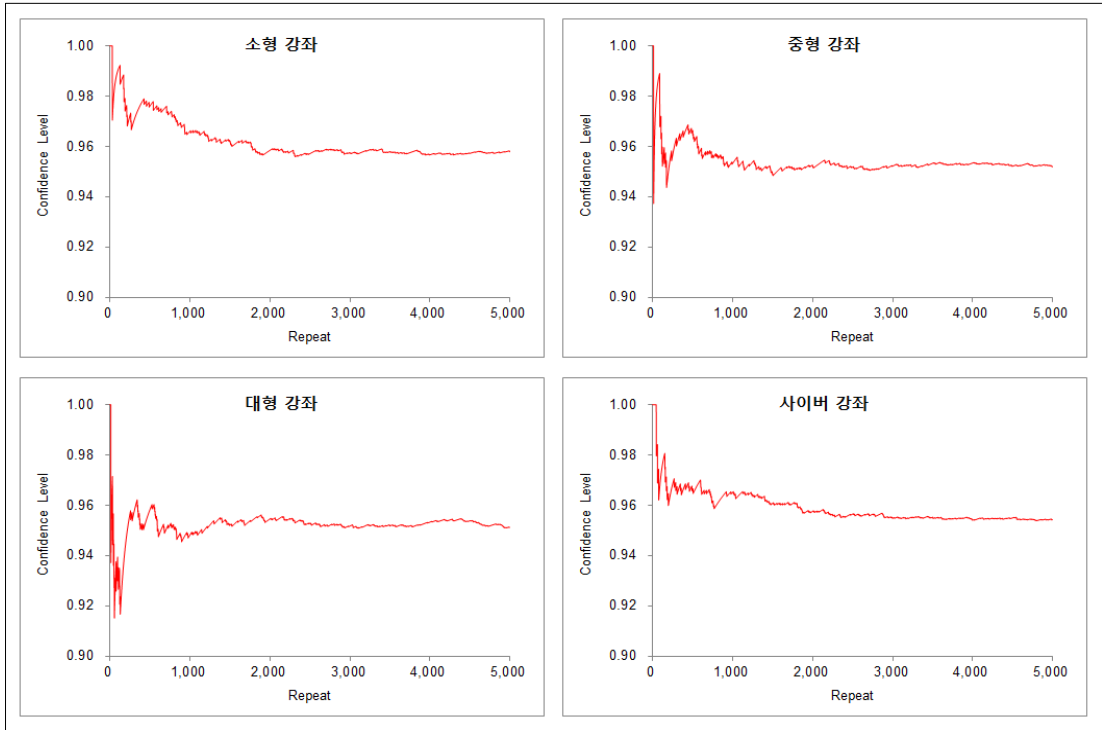


그림 2. 소형, 중형, 대형, 사이버 강좌의 모의실험 결과

는 비율을 그래프로 작성하고 저장하기

3. 모의실험 결과

전수조사로 진행되는 강의평가의 모평균을 추론하기 위해 5,000회 반복 시행한 모의실험 결과는 다음 [그림 2]와 같다. 먼저 소형 강좌의 경우 모평균을 포함한 신뢰수준은 96.2%로 수렴하는 것으로 나타났고, 중형 강좌의 경우 모평균을 포함한 신뢰수준은 95.4%로 수렴하는 것으로 나타났다. 또한 대형 강좌의 경우 모평균을 포함한 신뢰수준은 95.2%로 수렴하는 것으로 나타났으며, 사이버 강좌의 경우 모평균을 포함한 신뢰수준은 95.9%로 수렴하는 것으로 나타났다. 이를 종합해볼 때 소형, 중형, 대형, 사이버 강좌 모두 신뢰수준은 95% 이상으로 근사하는 것으로 볼 수 있어 모집단 N 명 중에서 단순 무작위 표본추출로 표본 n 명을 선택하여 이 n 명의 표본평균으로 모평균을 추론하는 것은 통계적으로 의미 있다고 할 수 있다.

V. 결론 및 시사점

현재 국내 대학들이 온라인 강의평가의 단점인 저조한 참여율을 보완하기 위해 부가한 강제성은 강의평가 결과의 타당성을 크게 저해하는 요인이다. 이에 본 연구는 전수조사 강의평가 제도의 단점을 보완하기 위한 대안으로 모의실험을 통해 대표성 있는 표본 강의평가제의 실현 가능성을 제시하였다. 이를 위해 모집단 N 에서 표본 n 을 선택하여 신뢰수준 $(1-\alpha) \cdot 100\%$ 에서 모평균(μ)에 대한 신뢰구간을 추정하고 이 신뢰구간을 이용하여 모평균(μ)이 신뢰구간 내에 포함되는지 살펴보았다. 보다 구체적으로 신뢰구간이 모평균(μ)을 포함하는 비율이 $(1-\alpha)$ 로 근사한다면 표본평균을 이용하여 모평균을 추론하는 것이 타당한 것인지를 판단하였는데, 모의실험 결과 소형, 중형, 대형, 사이버 강좌 모두 신뢰수준은 95%로 근사하는 것으로 나타났다. 따라서 소형, 중형 강좌의 경우 약 50% 정도의 표본을 통해 유의한 강의평가가 가능하다고 볼 수 있고,

특히 소형, 중형 강좌보다는 대형, 사이버 강좌에서 본 연구가 제안한 표본 강의평가제의 운영 가능성이 훨씬 높다고 할 수 있다.

강제성 있는 강의평가는 저조한 참여율 문제를 해결할 수 있지만, 반면에 불성실한 응답률을 증가시키는 문제를 불러일으킨다. 강의평가제의 타당성과 신뢰성을 높이기 위해서는 강의평가 도구뿐만 아니라 강의평가제의 운영 전반에 걸쳐 강의평가 결과를 왜곡하는 요인들에 대한 심층 분석이 필요하다. 진수조사를 통해 응답률을 높이는 방법에만 치중한 나머지 형식적인 평가로 그치는 것이 아니라 학생들이 성실하게 응답한 자료를 바탕으로 결과의 객관성, 신뢰성, 공정성을 높일 처방과 대책을 고민해야 한다.

학생들이 강의평가의 모든 설문 문항에 대해 성실하게 응답한 경우에도 동일 응답이 50% 이상 나온다면 이는 극단적으로 말해 하나의 문항으로 강의 만족도를 구하는 것이 보다 효율적일 것이라는 견해를 보더라도 [9] 각 대학은 학생들의 응답 패턴을 분석하여 강의평가의 신뢰성과 타당성을 높일 필요가 있다. 강의평가의 문항개선, 조사방법의 체계성, 운영방식의 합리성 등이 뒷받침 되더라도 학생들의 강의평가에 대한 인식과 적극적인 참여태도가 강의평가 결과에 대한 신뢰도와 타당도를 높일 수 있는 주요 요인임을 고려할 때, 대학에서는 강의평가의 중요성을 더욱 인지하고 강의평가의 전반적인 측면에서 깊이 있는 질적 연구 등 지속적인 개선노력을 해야 한다.

참 고 문 헌

- [1] 한신일, 김혜정, 이정연, “한국대학의 강의평가실태 분석,” 교육행정학연구, 제23권, 제3호, pp.379-403, 2005.
- [2] 김명화, “강의평가의 타당도와 신뢰도,” 아시아교육연구, 제6권, 제3호, pp.1-24, 2005.
- [3] 하오선, 정민호, “강의평가 응답분석을 통한 강의평가도구 개선방안,” 열린교육연구, 제22권, 제3호, pp.273-294, 2014.
- [4] 양길석, “대학 강의평가 일관적 응답의 경향성과 영향력 분석,” 제27권, 제2호, pp.255-278, 2014.
- [5] 홍경선, “대학교 강의평가에 나타난 일관적 응답 분석,” 교육정보미디어연구, 제12권, 제2호, pp.97-127, 2006.
- [6] 신소영, 권진희, “강의평가결과의 신뢰성 제고 방안 연구: 강의평가제도의 운영을 중심으로,” 순천향 인문과학논총, 제35권, 제4호, pp.115-145, 2016.
- [7] 오숙영, “강의평가점수 영향력 요인에 대한 사후조정 연구,” 교육평가연구, 제28권, 제4호, pp.1225-1254.
- [8] W. Hoy and C. Miskel, *Educational Administration: Theory, Research, and Practice*, New York: McGraw Hill, 2005.
- [9] 이원석, 이현우, 정영근, “대학 강의평가에서 실시 방식이 학생들의 불성실한 응답에 미치는 영향에 관한 연구,” 교육방법연구, 제24권, 제3호, pp.547-561, 2012.
- [10] 최보금, 김재용, “위계적 선행모형을 활용한 대학생 강의평가 관련 요인 탐색: 무성의 응답의 영향을 중심으로,” 열린교육연구, 제21권, 제1호, pp.77-100, 2013.
- [11] 김성숙, 김학일, “학기중 적용을 위한 강의진단평가도구 개발과 타당성 분석,” 교육평가연구, 제21권, 제1호, pp.55-78, 2008.
- [12] 안수현, 이상준, “교육수요자 만족도조사의 신뢰성 제고방안 탐색: S대학교 사례를 중심으로,” 디지털융복합연구, 제15권, 제12호, pp.71-77, 2017.
- [13] 한경수, 최숙희, 박재철, “강제적인 대학 강의평가의 문제점,” 한국통계학회논문집, 제18권, 제1호, pp.35-45, 2011.
- [14] M. Milton, *Head First Data Analysis: A Learner's Guide to Big Numbers, Statistics, and Good Decisions*, O'Reilly Media, 2009.
- [15] D. Griffiths, *Head First Statistics: A Brain-Friendly Guide*, O'Reilly Media, 2008.
- [16] 서민구, *R을 이용한 데이터 처리&분석 실무*, 길벗, 2014.
- [17] M. J. Crawley, *The R book*, John Wiley &

