한국학교수학회논문집 제 9 권, 제 3 호
Journal of the Korean School Mathematics Society
Volume 9, Number 3, 309-316, September 2006

표집오차(sampling error)와 표집분포(sampling distribution)의 용어 사용에 관한 연구

김응환[1)]

이 논문에서는 현재 중등학교 수학의 통계교육에서 다루고 있는 통계용어의 의미상 혼선과 애매한 내용을 수학교사를 대상으로 알아보고, 표본평균의 확률분포에 대한 지도 영역에 있어서 표집(sampling, 표본추출)의 문맥에서 표집오차(sampling error)와 표본평균의 표집분포(sampling distribution)라는 용어를 도입하여 일관성 있게 사용할 것을 제안하였다. 현행 중고등학교의 수학과의 통계의 용어 정의와 개념설명에 있어서, 교육부가 검정한 12종의 검정 교과서와 국정교과서 간에서도 차이는 물론 의미의 혼선과 함께 정의의 일관성의 부족은 통계를 교육하는 수학교사와 학생들에게 심각한 오개념을 형성하게 만들고, 그 애매함으로 인하여 통계학의 학문 자체에 대한 흥미와 태도의 정의적인 면에서 부정적인 영향을 주고 있음이 발견되었다. 본 연구에서는 표본평균의 확률분포의 효율적인 지도를 위한 표본오차 대신에 표집오차를 사용할 것과 표집분포의 용어를 도입함으로서 통계용어의 정확한 사용을 통하여 교사와 학생들에게 통계용어의 올바른 개념의 형성과 이해는 물론 통계교육의 일관성과 계열성 유지의 필요성을 제기하였다.

주요용어 : 표집오차, 비표집오차, 표집분포, 표준오차, 표본평균의 표집분포, 통계교육, 통계용어.

Ⅰ. 서론

통계학은 데이터의 과학(science of data)이라고 하며 수학은 과학의 언어라고 한다. 중등학교에서 수학교과 속에 포함되어 있는 통계학은 21세기를 맞이하여 모든 국민이 알아야하는 필수 교과가 되었다. 변화가 빠른 요즈음의 학생들에게는 통계학에 대한 새로운 개념을 이해하고 보다 도전적인 문제를 해결하는 경험이 없이 수준 높은 통계학의 학습에 흥미를 느끼기를 기대할 수는 없다(이석훈, 1999). 그리고 시대의 요구에 부응하는 구성주의를 바탕으로 한 토론과 실험 중심의 통계교육의 개선에 대한 주장이 있다(김응환, 2004). 최근 우리나라는 학교교육에서 학생 선진국가일수록 여러 분야에서 수많은 의사결정에 이용되는 통계학의 역할은 그 자리매김을 확실하게 하고 있으며, 온 시민의 통계문맹을 극복시키기 위한 통계교육의 필요성이 새롭게 등장하고 있다(정성석, 2005). 통계학을 배운다는 것은 통계학에서 사용되는 고유 용어를 새롭게 배운다는 의미로 해석할 수도 있다.

1) 공주대학교 수학교육과 (yhkim@kongju.ac.kr)

교육부에서도 7차 수학과 교육과정에서 [수학1]과 [실용수학]와 특히 선택교과인 [확률과 통계]를 독립시켜 그 중요성을 강조하고 있다. 교육목표와 내용면에서도 실생활 응용중심으로 문제해결에 이용할 수 있도록 통계의 실험을 많이 할 것을 주문하고 있다(교육부, 1997). 미국의 NCTM(2000)의 교육과정 규준에서도 자료분석을 강조하며 통계교육의 중요성을 말하고 있다.

통계학에서 사용하는 고유 용어는 그 개념과 정의를 처음 배우게 되는 교사와 학생들에게 정확하게 전달할 수 있게 표현되어야 한다. 그러나 수학교과내용의 선호도 조사에서 보듯이 확률과 통계에 관한 학생들의 인식은 아직도 어려운 학문이며 그 개념을 배우기가 힘든 과목으로 인식되고 있다(김영국, 2000). 그 원인은 여러 가지일 것으로 짐작되지만, 교과구성의 내용면에서 대학의 통계학과에서 조차 가르치는 비중이 낮은 조합론과 확률론 영역의 강조(이상복 외(2005), 김응환(2000))와 통계용어의 난해함에 한 원인이 있음을 말해주고 있다. 왜냐하면 12종의 국내 검정교과서와 국정교과서에서조차도 각종 통계 용어에 대한 일관성과 개념설명의 애매함으로 인한 혼선이 있음이 지적되고 있을 정도이기 때문이다(이상복 외(2005). 특히 Moore(1997)는 "수학은 통계학과 다소 차이가 있다"라고 말하면서 학교수학에서 통계교육을 담당해야하는 수학교사들에게 사고의 전환과 각별한 용어사용의 주의를 환기시키고 있다.

본 연구에서는 교사들이 통계용어에 대한 익숙한 정도와 의미를 어느 정도 이해하고 있는지를 알아보고, 통계의 여러 용어의 애매한 표현과 오개념 형성의 관계를 살펴보며, 이러한 시사점을 바탕으로 교사교육을 위한 표본평균의 분포의 지도에 관한 효율적인 방법의 하나로 표집(sampling)을 행하는 실험활동의 문맥의 상황에서 표집분포와 표집오차 그리고 표준오차의 용어에 대한 사용과 그 개념을 정확하게 교육해야 할 것을 제안하고자한다.

Ⅱ. 본 론

1. 수학교사들의 통계용어에 대한 익숙도 조사

2006년도 A 지역의 중등수학 1급 정교사 자격연수에 입소한 수학교사를 대상으로 통계용어에 대한 익숙도를 조사한 결과는 다음과 같다. 여기서 선정한 용어는 검정교과서 12종과 국정교과서 1종의 통계용어 색인표를 비교해본 결과 서로 일치하지 않기 때문에 대체적으로 일치하는 용어를 참고하고(28종 내외, 국정교과서는 19종), 연구자가 재량으로 선정한 모집단과 표본과 표집(sampling)영역의 용어를 추가한 것이다.

대체로 수학선생님들은 다음 [표 1-1]에서 보여 주고 있듯이 표본을 추출하는 표집의 문맥에서의 표본평균의 확률분포에 대해서는 절반정도(45%)가 익숙하다고 대답하고 있으며, 표본평균의 표집분포에 대해서는(52%) 잘 모르는 것으로 나타났다. 반면에 모비율(52%)이나 표본비율(35%)에 대하여서는 익숙하지 않은 것으로 나타났다. 특히 표준편차에 대해서는 거의 모두 알고 있다고 말하면서 표준오차에서 대해서는 절반가량(45%)의 선생님들이 잘 모르겠다는 답을 하고 있다. 즉 오차(error)의 발생이 표집(sampling:표본추출)의 상황에서 발생하고 그 문맥상에서 연유된다는 사실의 지식이 약한 것으로 나타났다. 익숙하지 않다는 이야기는 잘 모르겠다는 것을 의미한다고 생각할 수 있다. 이러한 사실들은 표집오차, 표준오차 등의 오차에 대한 용어에 대한 표집(표본추출)의 문맥상에서의 이해가 부족한 상태를

나타내며, 표본평균의 표집분포에 대한 개념이해가 부족한 것을 의미한다. 이것은 표집의 문맥상 개념의 이해를 위한 수학교사들의 재교육이 필요함을 시사해주고 있다.

[표 1-1] 통계용어에 관한 익숙도 조사표

순번	용어	익숙도(N=31명)			비고
		1	2	3	
		모른다	보통이다	안다	
1	모집단	0	1(3%)	30(97%)	
2	표본 sample	0	3(10%)	28(90%)	
3	표집 sampling	4(13%)	1(3%)	26(84%)	
4	모평균	1(3%)	0	30(97%)	
5	모분산	2(6%)	1(3%)	28(90%)	
6	표본평균	2(6%)	3(10%)	26(84%)	
7	표본분산	4(13%)	5(16%)	22(71%)	
8	모비율	16(52%)	5(16%)	10(32%)	
9	표본비율	11(35%)	15(48%)	5(16%)	
10	표본평균의 평균	4(13%)	6(19%)	21(67%)	
11	표본평균의 표준편차	6(19%)	15(48%)	10(32%)	
12	표본표준편차	2(6%)	2(6%)	27(87%)	
13	표본평균의 확률분포	5(16%)	12(39%)	14(45%)	
14	표집분포 sampling distribution	15(48%)	8(25%)	7(22%)	
15	표본평균의 표집분포	16(52%)	9(29%)	6(19%)	
16	큰수의 법칙	12(39%)	6(19%)	13(41%)	
17	중심극한정리	8(25%)	12(39%)	11(35%)	
18	확률분포	0	5(16%)	26(84%)	
19	확률변수	4(13%)	0	27(87%)	
20	표준편차 standard deviation	0	1(3%)	30(97%)	
21	표준오차 standard error	14(45%)	10(32%)	7(22%)	
22	표본오차 (영문표기 없음)	8(25%)	11(35%)	12(39%)	
23	표집오차 sampling error	17(55%)	10(32%)	4(13%)	
24	비표집오차 nonsampling error	24(77%)	4(13%)	3(10%)	
25	표본의 크기 size of sample	1(3%)	4(13%)	26(84%)	

2. 여러 교과서의 통계용어 설명의 혼선의 예

앞의 익숙도 조사에서 보았듯이 왜 이렇게 통계학의 용어에 대한 이해가 대체로 익숙하지 않은가는 사범대학의 수학과 교육과정을 조사한 연구를 참고하면(정성석 외, 2005), 사범대학에서의 3-6학점의 확률통계과목 강의로서 그다지 많지 않은 교육에 문제가 있기도 하지만, 다음과 같은 중고등학교의 교과서에서의 용어에 대한 세심한 주의가 이루어지지 않고 있기 때문이라고 생각된다.

여기서는 현행 고등학교의 검인정과 국정교과서에서의 용어에 대한 애매한 혼선의 예를 지적하면서 8차 수학과 교육과정에서의 새로운 교과서의 집필시에는 수학교육학회들과 통계학회의 차원에서 연구와 토의가 이루어져서 이러한 애매함과 오류들이 분명하게 교정되어야 할 것이다. 현행 수학교과서의 통계용어의 설명의 혼선을 알아보자.

1) 표집오차(sampling error)와 표본오차에 관하여

한국교원대학교 국정도서편찬위원회(2003)에서 발행한 국정교과서인 [확률과 통계]교과서의 내용에는 p.128에서 모집단에서 뽑은 일부분을 표본(sample)과 모집단의 용어를 설명하고 있다. 이어서 다음과 같이 **표본오차**에 대하여 설명하고 있다.(p.129)

표본조사의 목적은 모집단에서 뽑은 표본을 바탕으로 모집단의 특성, 즉 모집단의 평균 또는 표준편차 등을 추측하는데 있다. 그러나 이와 같은 경우에는 표본에 의한 추측값과 모집단의 참값 사이에 차이가 발생하기 마련인데 이 차이를 **표본오차**라고 한다.(p. 129) 각주에는 "입력오류, 집계누락, 거짓 응답 등에 의한 오차를 **비표본오차**"라고 한다.(p. 129)

여기서 표본오차라는 용어에 관심을 두고 생각해보기로 한다. 한국통계학회에서 발행한 통계학용어집(1997, 자유아카데미)에서는 표본오차라는 용어는 보이지 않고, **표집오차**(sampling error, p.93과 p. 105)라는 용어를 표시하고 있다. 역시 **비표본오차**라는 용어는 없고, 대신에 비표집오차(non-sampling error, p69와 p.44)라는 용어의 사용을 권장하고 있다.

여기서 논의하고자 하는 것은, 위에서 조사한 수학교사들이 흔하게 접하게 되는 [표준편차, 표준오차, 표본오차, 최대 허용오차, 표집오차, 비표집오차]등에 대한 익숙도 조사에서 잘모른다는 대답을 확인할 수 있다. 이것은 이들 용어가 매우 유사해서 특별한 차이가 없는 듯이 보이고 서로 다른 용어로서 사용되는 이유나 배경에 대하여 관심이 적은 듯 생각된다.

이러한 여러 가지 다양한 용어들의 현실 속에서, 표집오차와 표본오차 중에 어느 용어가 보다 더 표집(sampling, 표본추출)의 문맥상에서 바람직한가를 말하고 싶다. 모집단에서 임의추출한 부분을 표본(sample)이라고 부르고, 표집(sampling)을 실행하는 활동이 동반된 문맥으로 해석하면서 이 때 발생하는 오차는 표본오차라 부르기 보다는 표집오차가 더 그 의미에 가깝다고 할 수 있다. 그러므로 관습적으로 사용하여 왔더라도 이제부터는 표본오차라는 용어를 사용하지 말고 대신에 한국통계학회의 통계용어집에서 사용을 권장하고 있는 표집오차와 비표집오차를 고등학교 수학교과서에 실어서 사용해야 한다고 제안하고자 한다.

이렇게 해야만 대학을 진학하여 통계학을 더 배우거나 사회와 직장에 나가서 한결같이 표집상황 하에서의 표집오차의 용어를 사용함으로써 신문을 보더라도 통계지식과 용어의 일관성과 연결성을 유지해주는 효과와 의사소통의 혼란을 예방하는 기준이 될 것으로 생각되기 때문이다. 보다 근본적으로 말하면 영어 원문에도 표본오차(직역: sample error?)이라는 용어는 없기 때문에 번역에 오류가 있다고 볼 수 있다.

2) 표집분포(sampling distribution)의 용어 사용에 대하여

교육부의 국정교과서인 [확률과 통계](pp. 131-132)에서 "모집단에서 여러 번 표본을 추출하여 평균을 구하였을 때, 그 표본의 평균은 추출되는 표본에 따라 다르다. 여기서는 서로 다른 표본으로부터 구한 **표본의 평균에 대하여 그 분포**를 알아보자."라고 말하고 있다.

몇몇 수학 선생님들과 인터뷰한 결과, 위에서 언급하고 있는 "표본의 평균에 대하여 그 분포를 알아보자"라고 했을 때 그 느낌이 바로 와 닫지 않는다고 한다. 하물며 확실한 표본평균의 성질에 대하여 인지하지 못한 수학교사로부터 고등학교 학생들이 지도를 받는 경우를 생각해보면 통계가 왜 어려운 학문으로 인식되는지를 생각하게 한다.

표본평균의 분포가 왜 이런 현상을 일으키는가는 여러 가지 이유가 있다. 우리 수학교사

들이 표집의 실행단계에서의 확률표본(random sample)에 대한 실험활동의 훈련과 사고가 부족한 사실과 함께 이석훈(1999, p. 208)에서 언급한 것과 같이 sampling 이 표본추출과 표집의 실행적 의미를 가지고 있음에도 불구하고 모집단의 부분집합으로서의 기술적 의미만을 전달하는 표본이라는 용어를 단순하게 쓴 것도 하나의 이유가 된다. 이러한 혼란을 방지하기 위해서는 "표본평균의 그 분포" 또는 "표본평균의 확률분포"라는 문장을 "표본평균의 표집분포"라고 구체적으로 사용한다면 이 표집분포의 용어로 인하여 표본평균의 그 분포가 분명하게 표본추출(표집)에 의한 과정에서 얻어지는 분포로 그 의미가 잘 전달될 것으로 생각된다. 그러므로 새로운 용어의 도입이 어렵다고 하더라도 보다 분명한 표집의 문맥상에서의 표본평균의 분포의 진정한 이해를 위하여 유의미한 표집분포(sampling distribution)의 용어를 도입하여 사용할 것을 제안한다.

3) 표본분산(sample variance)과 표본표준편차(sample standard deviation)의 용어의 정의에 관하여

교육부(2003)의 국정교과서인 [확률과 통계](p. 32)에서는 분산을

$$\sigma^2 = \frac{(x_1 - m)^2 + \cdots + (x_N - m)^2}{N}$$

으로 표기하고 있고, 일부 검정교과서에서는 표본분산으로 대문자 S를 사용한

$$S^2 = \frac{1}{(n-1)}[(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2]$$

을 사용하고, 일부 검정교과서는 소문자 s 로서

$$s^2 = \frac{1}{(n)}[(x_1 - \overline{X})^2 + \cdots + (x_n - \overline{X})^2]$$

으로 표기하고 있다.

일부교과서에서는

$$S^2 = \frac{1}{(n)}[(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2]$$

으로 표기하고 있다(박배훈 외, 2003, p. 325).

이러한 용어의 정의에 대한 혼란 역시 표집의 개념이 전혀 없는 σ^2의 계산 목적과 표집의 개념이 포함될 수밖에 없는 σ^2, S^2의 계산 목적이 서로 구별이 안 되고 있었고, 또한 일부 교과서에서 확률변수인 표본평균과 단일표본(single sample)에서 평균과 분산의 의미를 혼동하는데서 비롯되고 있으며, 표집(sampling)의 문맥상에서 자료의 값에 대한 혼돈이라고 생각할 수 있다. 특히 통계영역에서 문자의 사용시 확률변수로서의 대문자 표기($X, \overline{X}, S^2$)와 각 확률변수에 대한 하나의 구체적인 값인 소문자 표기($x, \overline{x}, s^2$)를 정확하게 구분하여 사용해야 함에도 이에 대한 구분이 불분명한데 기인한다. 그리고 표집활동의 문맥상에서의 표본

분산의 계산식에서 분모에 (n)으로 나누는 것과 (n-1)로 나누는 문제도 모두 (n-1)로 나누는 것으로 일관성 있게 통일해야 한다고 제안한다.

즉, 이러한 혼선을 극복하기 위해서는 확률변수의 지도는 물론 표집이라는 문맥상에서의 표본분산에 대한 통계교육을 할 수 있는 수학과 교육과정의 개정이 필요하다고 하겠다.

Ⅲ. 결 론

지금까지 현행 고등학교와 중학교의 수학과 교육과정에서 혼선을 보이고 있는 통계용어의 혼선에 대하여 그 일부를 살펴보았다. 통계학 지식에 정확한 개념 습득을 위해서는, 그 용어에 대하여 명확하고 간단하면서도 유의미한 설명을 할 수 있도록 학생들과 교사들을 안내해야 한다고 생각된다.

본 연구에서는 표본오차라는 용어 대신에 표본을 추출하는 표집활동의 문맥에서표본오차보다는 표집오차(sampling error)를 사용할 것을 제안하였다. 그리고 표집의 문맥에서 표본평균의 확률분포라는 용어보다 표본평균의 표집분포(sampling distribution)라는 용어를 사용할 것을 제안한다. 실생활과 관련된 통계를 지도함에 있어서 용어의 적절한 사용은 그만큼 쉽게 통계의 지식을 전파할 수 있기 때문이다. 또한 통계용어의 연결성과 일관성을 유지하기위하여 표본분산의 계산식에 대한 통일된 정의와 산점도의 사용을 재차 제안하였다. 앞으로의 교과서에서는 중고등학교의 통계교육에 자료의 수집과 자료의 분석에서 보다 실생활에 사용할 수 있는 내용을 비중있게 다루고, 용어의 의미가 보다 생각하기 쉬운 유의미한 용어를 탐구하며, 특히 학생들을 평가하기에는 다소 어려움이 있을 수 있으나 표집활동을 포함하는 교과서 내용을 구성하는 교육과정에 대한 실제적인 연구가 필요한 시점이라고 생각된다.

참고문헌

교육부 (1997). 7차 수학과교육과정, 대한교과서주식회사.

교육부 (2003). 통계와 확률, 고등학교 교과서, 한국교원대학교 국정도서 편찬위원회.

김영국, 박기양, 박규홍, 박혜숙, 박윤범, 임재훈 (2000). 학교수학의 각 영역에 대한 선호도 연구, 한국수학교육학회 시리즈A, pp. 127-144.

김응환 (2004). 학교수학에서 통계교육의 개선방향, 한국학교수학회 논문집, 제 7권 2호, pp. 51-65.

박배훈, 조민식, 김원석, 이대현, 김원경, 김두성, 정원진 (2003). 고등학교 수학 I 교과서, 법문사.

이상복, 손중권, 정성석 (2005). 응용통계연구, 제 18권 1호, pp. 197-210.

이석훈, 김응환 (1999). 확률과 통계지도론, 경문사.

정성국, 손중권, 이상복 (2005). 통계학과 발전방향에 대한 고찰: 교직과정을 중심으로, 응용통계연구, 제 18권 1호, pp. 211-227, 한국통계학회.

한국통계학회 (1997). 통계학 용어집, 자유아카데미.

Moore, D. S. (1997). Statistics: Concepts and Controversies, 4th ed. New York, W.H. Freeman.

National Council of Teachers of Mathematics (2000). Principles and standards for school mathematics, Reston, Virginia.

김응환

A Study of Using the Terminology of Sampling Error and Sampling Distribution

Kim, Yunghwan[2)]

Abstract

This study examined the ambiguous using the terminology of statistics at mathematics textbook of highschool in Korea and proposed the correct using of sampling error and sampling distribution of sample mean with consistency. And this paper proposed that the concept of error have to teach in context of sampling action in school mathematics.

Key Words : Sampling error, Non-sampling error, Sampling distribution, Standard error, Sampling distribution of sample mean, Statistics education, Statistical terminology.

2) Kongju National University, Department of Mathematics Education (yhkim@kongju.ac.kr)