韓國數學教育學會誌 시리즈 A <數學敎育>
2011. 11. 제 50권, 제 4호, 489-505.

J. Korean Soc. Math. Ed. Ser. A: The Mathematical Education
November. 2011, Vol. 50, No. 4, 489-505.

중학교 3학년 수학교과서 통계단원에 나타난 요약개념 분석

이 영 하 (이화여자대학교)†
이 은 희 (이화여자대학교 교육대학원)

Ⅰ. 서 론

1. 통계와 통계적 소양(literacy)

통계학(Statistics)은 통계(statistic)를 이용하여 정보를 축적하고 소통하며, 결과의 실용성을 강조하는 학문이다. 통계(statistic)란 수리통계학적으로는 표본 자료(sample data)로부터 구해지는 한 개의 숫자를 뜻하지만, 통계를 쓰는 일반 대중은 숫자 외에도 자료로부터 얻어진 그래프나 표, 그림 등을 모두 통계라고 생각한다. 우리는 서론에서는 후자의 입장을 따르지만, 논의가 심각해지는 다음 장부터는 통계를 숫자 종류로 제한하기로 한다.

현대 사회에서 통계는 그 중요도를 더해가고 있는데, 그만큼 통계에 대한 오해(misinterpretation), 오용(misuse)도 늘어나고 있는 것이 사실이며, 그렇기 때문에 통계적 소양(statistical literacy)이 강조되고 있는 것도 사실이다.(Huff(1954))

특히 한국 사회는 이 점에 있어서 선진국에 비해 매우 낙후되어 있다고 연구자는 판단하고 있는데, 여러 가지 원인들 중에서도 충실하지 못한 초중등 학교 통계 교육 역시 책임의 중요한 당사자라고 생각한다.

사람들이 통계를 제시할 때는 항상 그 배후에 어떤 주장이 있고 그 주장에 대한 근거로 통계가 사용된다고 할 수 있는데, 방송 언론이나 정부 발표 등을 통계전문가의 입장에서 듣고 있노라면, 잘못되거나 허술한 주장[1]이 매우 많다고 느끼는 경우가 많다.

왜 그렇게 느끼는 지를 설명하기 위해, 통계가 어떻게 구해지고 사용되는 지 잠시 생각해 볼 필요가 있다. 통계는 어떤 주장에 대한 목적을 갖고 수집된 자료로부터 계산되고, 그렇게 계산된 통계는, 애초 자신의 자료 수집 목적과 관련된 언어적 주장에 맞도록 해석하여, 그 근거로 제시된다.

즉 자료수집-계산-해석(주장)의 과정에서 준비되고, 계산되고 사용되는 것이 통계이다. 그런데 이 과정에는 도처에 오류의 가능성이 숨어있다. 엉뚱한 자료를 잘못 수집할 수도 있고, 계산을 목적에 맞지 않는 엉뚱한 계산 -가령 평균을 구해야 할 상황에서 표준편차를 구하는 것과 같은- 을 할 수도 있고, 계산 결과를 상식과는 전혀 맞지 않게 해석하여 주장할 수도 있다.

가령 중학교 1학년 학생들에게, 재래시장과 할인점 중에서 어느 쪽의 물건 값이 더 싼지를 조사하여 비교해 보라는 수행평가 문제를 내었다고 하자. 이 경우 당연히 자료를 수집해서 비교해야 하는데, “물건 값”이라는 단어는 이 상황에서 매우 복잡한 정의를 갖는다. 특정 물건의 “물건 값”이라 해도 정가표에 적힌 것인지, 가게 주인이 할인해 준 그 값인지, 정가표를 적지 않고 파는 가게는 어찌 할 것인지, 할인점에서 구매총액에 따라 적립해 주는 포인트는 물건 값에 반영해야 할지 말아야 할지, 조사목적으로 물을 때와 정말 사려고 물을 때, 가게 아저씨의 대답하는 물건 값 차이는 어떻게 처리 할 것인지, 어떤 물건의 값을 조사할 것인지 등, 너무 복잡해지게 된다. 그렇지만 자료를 구성하게 되는 이 “물건 값”

* 접수일(2011년 10월 17일), 수정일(2011년 11월 11일), 게재확정일(2011년 11월 18일)
* ZDM분류 : U2
* MSC2000분류 : 97U20
* 주제어 : 교과서 분석, 대표값, 산포도, 중심 경향값, 통계적 소양, 통계교육

† 교신저자

1) 가령 최근 각 대학의 신입생들의 수학학력 저하로 인해 공대 교수들이 가르치기가 어렵다고 하지만, 정부는 학생들의 수학학력이 저하되지 않았다고 통계 자료를 제시하거나, 아무도 믿지 않는 대학별 취업률 통계를 제시하는 것 등을 예로 들 수 있다.

이 우리가 앞서 처음 질문할 때의 “물건 값”의 정의와 상식적으로 일치되지 않으면, 어떤 통계를 구하여 어떤 주장을 한다 해도 우리는 결과가 믿을 수 없다는 말을 하게 되는 상황이 발생될 수 있게 되는 것이다. 정부는 물가가 3.7%올랐다고 하는데, 시민들은 물가가 40%올랐다고 하게 되는 이유가 바로 양자 사이에 물가의 정의가 서로 다르기 때문인 것이다.

서울 시민 한사람의 1일 수돗물 사용량을 조사할 때, 청구된 월간 수도요금을 가족 수와 날짜 수로 나누면 안 될까? 집집마다 수도관 중간에 새는 물의 양이 다르다면, 전혀 엉뚱한 통계를 얻을 수도 있다.

통계적 소양 교육은 이런 오류의 가능성을 인식하고, 통계를 바르게 사용하는 일에 대한 것이어야 하며, 본 연구는 특히 오류 가능성의 한 지점인 통계 계산에 관련된 합목적성에 대한 문제를 짚어 보려는 것이다.

정보화시대는 컴퓨터의 도움으로 통계 계산 문제에 있어서 공식에 따른 계산의 정확성에 대해서는 의심할 필요가 없고, 따라서 학교에서도 통계 계산은 계산기나 컴퓨터를 사용하도록 권장되고 있으며, 통계 수업의 주제로서 중요하게 생각되지 않는 추세에 있다.

그 대신 통계 계산에서 중요하게 생각되어야 할 것은 언어적으로 진술되는 목적에 비추어 볼 때, 계산된 통계가 목적에 부합된 것인지 여부를 살피는 일이다.

가령 자료의 값들 중에서 중간 정도 크기의 값이 어느 정도인지를 알기를 원했는데, 이를 위해 자료의 표준편차를 구한다면 이는 목적에 맞지 않은 통계 계산을 하고 있는 셈이라는 뜻이다. 이런 터무니없는 오류는 절대 없을 것이라고 생각하는 사람들이 많겠지만, 실제 표준편차가 무엇인지 모른 체 컴퓨터를 이용하여 구할 수도 있을 뿐 아니라, 사실 목적에 맞지 않는 통계는 이 외에도 도처에서 수시로 계산되고 있는 것이 현실이다.

어느 대학입시 지도를 하는 교사가 어느 여학생에게 수능선발 지원 대학을 상담하면서, 여학생 전체의 평균 수능 성적이 높으니, 여자대학에 지원하지 말라고 한다면, 잘못된 통계를 사용하는(엉뚱한 통계 계산을 하는) 것이다.[2] 한미 FTA 협상에서 미국 측이 한국과 미국의 자동차 수출입 문제를 문제 삼으며 양국 간 수출입 차량 수를 근거로 제시한다면 잘못된 통계를 사용한 것이다.[3]

2. 통계교육의 상황변화와 연구문제

A. 통계교육 상황의 변화와 대응

일상생활에서 다루는 통계는 이렇게 다양하지만, 학교 통계 수업에서 다루어야 할 기본 통계로서 전 세계 국가들이 대푯값과 산포도를 꼽고 있다. 우리나라는 오랫동안 평균과 표준편차만을 가르쳐 왔으나 세계적 추세에 맞추어 더 다양한 종류의 대푯값과 산포도를 소개하고 가르치게 되었다.(교육과학기술부(2008))

이러한 변화는 시대 변화에 맞는 매우 중요한 발전의 단초임이 분명하지만, 이런 변화가 대푯값과 산포도의 종류에 대한 지식의 단순한 추가에 그친다거나, 그 계산법에 골몰하는 현재의 대푯값과 산포도 교육의 연장일 뿐이라면, 이것을 발전이라고 할 수 있는 사람은 많지 않을 것이다.

이런 변화는 궁극적으로 통계적 소양의 증대로 이어져야 하며, 이것은 통계의 존립 근거인 ‘계산된 통계 값의 합목적성 평가에 대한 교육’에 획기적 변화의 바람을 몰고 올 것을 기대하는 것이라고 할 수 있다.

지금까지 우리나라 학교수학에서의 통계와 확률교육은 순열과 조합 등의 확률 계산이 강조되어왔다. 즉 스켐프(Skemp)의 관계적 이해의 관점보다는 도구적 이해의 관점이 강조되고 있다고 할 수 있는데, 대학교의 통계학과에서 운영하는 전공과목에서조차 순열과 조합론 내용의 대부분을 그다지 강조하지 않는다는 사실을 중고등학교의 수학교사들이 대부분 모르고 있다는 것이다(김응환, 2007).

7차 개정 교육과정에서 요구하는 교수학습방법에 있어서도 설명식, 소집단 토론식과 활동 중심의 실험수업을 권장하는 등 다양한 방법의 조화를 강조하고 있는데, 이러한 변화의 흐름에 부응하기 위해서는 현재의 학교수학에서는 확률과 통계단원의 궁극적 학습목표를 재검토하고 그에 맞는 효율적인 지도방법을 새롭게 구안할 필요가 있다.(김응환(2007)).

---

2) 그 여학생의 지원가능 모집단위 별 커트라인 통계를 보아야 하며, 개별 상담이 아니라 단순히 여자대학 지원의 유불리만 말하고자 한다 해도 대학에 갈만한(상위 30% 정도) 남녀 학생들의 평균 성적을 비교해야 한다.

3) 수출입 차량 수가 아니라 차량가액 총액을 비교해야 한다.

확률통계 교육에서 확률계산이 더 강조될 수밖에 없는 이유의 한 가지로서, '대푯값과 산포도' 단원의 학습내용의 수준이 중학교 3학년 수학 학습 수준에 비해 매우 낮고, 그 계산은 단순 노동이 아닌가 하는 오해에서 비롯된 것은 아닌지 염려스럽다.

더욱이 대푯값과 산포도의 계산 지도에 있어서 계산기나 컴퓨터를 이용할 것을 강요받고 있는 현재의 교수상황에서, 이 단원의 수업은 교사에게 활로가

없는[4] 난감하고 가르치기 싫은 단원으로 변해가는 것은 아닌지 염려스러운 것이다.

만약 연구자의 이런 추측에 일고의 가치라도 있다면, 이 문제는 결국 '통계 단원에서 무엇을 가르칠 것인가?'의 문제에서 나아가 '통계 단원에서 무엇을, 왜, 가르칠 것인가?'의 문제로 이어지며 결국 초중등학교 확률 통계 단원의 일관된 학습목표 문제와, 이것을 통계적 개념의 학습이라는 다른 표현으로 이어서 생각할 때, '통계 단원에서 가르치고자 하는 통계적 개념이란 무엇인가?'의 문제로까지 비화되는 것을 알 수 있다.

이영하, 남주현(2005)은 선행연구들(Bakker, 2004; Garfield & Ben-Zvi, 2004; Garfield, 2002)을 바탕으로 분포개념, 요약개념, 표본개념의 세 가지를 핵심적인 통계적 개념이라고 주장하였는데, 그 중에서, 주장하고자 하는 사실에 합목적적인 타당한 표본의 정보가 무엇인가 또는 주어진 자료를 어떻게 요약해야 목적에 부합되는 요약이 될 것인가를 고려할 때 이 과정의 이해에 필요한 개념이 요약개념이라는 것이다. (남주현(2007))

어떤 현상에 대한 실증적 근거, 즉 정당화의 한 방법으로서, 주장하려는 바에 대한 근거가 될 수 있는 통계를 구해야 한다. 통계, 즉 자료의 요약이 필요한데, 자신의 주장과 관련 있는 요약이 필요하며, 그것은 요약의 목적과 부합되는 요약, 즉 합목적성이 있고 타당성 있는 요약이어야 한다는 것이며 이것이 요약개념의 핵심이다.

이에 대한 바른 지도는 결국 학교에서 배운 통계를 자신의 생활과 관련 지어 바르게 사용할 수 있고 잘못된 통계에 근거한 통계적 주장을 논리적으로 비판할 수 있는 통계적 소양 교육으로 이어진다고 할 수 있다.

B. 연구문제

요약개념의 교수학습에서 가장 중요한 것은, 언어적으로 제시된 주장과 통계적 요약(통계)이 일관성이 있는지를 바르게 평가할 수 있어야 한다는 것이다.

가령 자동차 수출입 문제에서 "우리는 수입만 많고 수출은 적다."라는 무역 역조 문제의 언어적 표현을 할 때, "이를 뒷받침 하는 통계는 '수(출)입 차량 총 대 수'가 아니라 '수(출)입 차량 가 총액'이어야 하는 이유는 무역 역조 불만의 근본 원인이 돈이기 때문이다."라고 판단할 수 있는, 즉 서로 다른 두 가지 이상의 통계적 요약에 대한 합목적성 평가 능력을 길러야 한다는 것이다.

'자료 중에 중간 정도 크기의 값'[5]을 알기 원하여서, 어떤 통계 공식에 의해 그 값을 구했는데, 이 값이 오히려 중간 정도의 크기를 거의 항상 알려주지 않는다면, 이 통계는 합목적성이 없는 통계이며, 이 경우 그 공식만을 보고도 '목적에 비추어 타당성 없는 통계'임을 미리 평가할 수 있다면, 이 능력이 바로 요약개념 지도를 통해 추구하는 능력인 것이다. 가령 표준편차 공식의 모양만을 보고도, 이것은 중심경향값을 알고자 할 때에는 쓸모없음을 말할 수 있어야 한다는 것이다. 나아가 중심경향값을 알기를 원하여 어떤 공식을 만들었다면, 만약 이 자료의 각각의 값이 b만큼씩 커졌을 때, 알려던 중심경향값도 b만큼 커져야 할 것이라는 논리성이 반영되어야 한다. 즉 자료 $x_1, x_2, x_3, \dots, x_n$로부터 그런 목적으로 통계 $u(x_1, x_2, x_3, \dots, x_n)$을 구한다면 $u(x_1+b, x_2+b, x_3+b, \dots, x_n+b) = u(x_1, x_2, x_3, \dots, x_n)+b$가 성립되어야[6] 비로소, '중간 정도 크기의 값'이 갖추어야 할 (개념)논리적[7] 성질에 그 공식이 부합될 것이라는 뜻이다. 또 '그러므로, 표준편차는 이런 성질을 갖고 있지 않기 때문에 중심경향값으로는 부적절함'을 (연역)논리적으로

4) "평균이나 표준편차 계산을 계산기에 의존한다면 이 단원에서는 무엇을 가르칠 것인가? 계산기 사용법이나 공식암기를 가르치는 것은, 중학교 3학년 학습 내용의 전반적 지적 수준을 고려할 때, 또 수학 교과의 성격을 생각해 볼 때, 아닌 것 같고..."의 고민이 숨어있는 것은 아닐까?

5) 이 값을 앞으로는 중심경향값이라고 부르기로 한다.

6) 이와 같은 성질을 통계 $u(x_1, x_2, x_3, \dots, x_n)$의 위치보존성이라고 한다.

7) 개념 논리적이라는 것은 어휘나 어떤 표현이 자동적으로 내포하는 개념이 스스로 모순된 개념을 갖지 않는다는 뜻이다. 가령 상자라는 표현을 썼을 때, 안팎이 있어야 하는 것은 당연하고, 만약 어떤 상자가 안팎이 없음을 알았다고 주장한다면, 그 경우 상자라는 표현이 개념 논리적으로 타당치 않음을 뜻한다.

설파할 수 있어야 한다는 것이다.

한편 이렇게 언어적으로 설정된 목적에 어떤 통계가 합목적적 타당성이 있는지 여부는, 차량 수출입 문제에서 느낄 수 있는 것처럼, 대개 관련 문제에 대한 다소 전문적인 지식(경제지식, 심리지식, 언어지식, 의학지식, 체육지식 등)이 필요한 경우가 많다. 따라서 이런 문제들은 초중등학교에서는 상식 수준의 지식으로 평가가 가능한 경우로 제한될 수밖에 없지만, 그렇다 해도 전 세계 통계교육에서 빠지지 않고 강조되는 두 가지 종류의 목적을 가진 통계가 바로 대푯값과 산포도라고 할 수 있다.

요약개념은 통계교육의 전 과정에 단계적으로 스며 있어야 하고, 점진적으로 발전되어야 하겠지만, 그 중에서도 중학교 3학년의 '대푯값과 산포도' 단원은 요약개념의 대표적 중심 단원이라고 할 수 있다.

본 연구에서는 중학교 3학년에서 배우는 대푯값과 산포도 단원에 대하여 중심경향값의 목적을 가진 대푯값과 자료의 값들이 얼마나 서로 다르게 흩어져 있는지를 알고자 하는 산포도의 교수학습 내용이 이런 요약 개념 측면의 관점을 잘 반영하고 있는지를 살펴보려는 것이다.

요약개념 측면이란, 중학교 3학년 대푯값과 산포도 단원의 내용들에 대해, 남주현(2007)이 제안한 요약개념 발달 측면을 고려하여, 첫째 '대푯값의 정의와 합목적성', 둘째 '산포도의 정의와 실용성', 셋째 '분포의 대푯값과 산포도', 넷째 '위치 보존성과 단위 보존성' 등의 4가지 측면을 뜻한다.

왜냐하면 이런 검토는, 결국 계산기 사용이 강조되는 상황에서의 수업에서, 기존의 대푯값과 산포도 단원의 교수학습 목표가 수정이 불가피하다고 할 때, 어느 방향으로 수업 목표를 선회할 것인가에 대한 그 방향성을 제시할 수 있을 것이라고 보았기 때문이다. 즉 이제는 "평균 정확히 계산하기"에 매달리는 것이 아니라, 계산기나 컴퓨터가 구한 평균에 대한 합목적성 평가, 제공된 공식의 논리적 타당성을 비판적으로 음미하는, 평가에 초점이 놓인 수업으로의, 목표 전환의 모습을 살피려는 것이다.

Ⅱ. 이론적 배경

1. 대푯값과 산포도 통계에서 요약개념 측면 의미의 구체화

요약(타당성) 개념은 추론하고자 하는 모집단의 정보에 합목적적인, 즉 타당한 표본의 정보가 무엇인가 또는 주어진 자료를 어떻게 요약해야 목적에 부합되는 요약이 될 것인가를 고려할 때 이 과정의 이해에 필요한 개념을 의미한다. 따라서 통계적 요약의 목적이 더 다양하고 복잡해 질 때[8] 이 개념 역시 그에 맞도록 적절하게 발달되어야 한다. 자료를 분류하고, 목적에 맞게 정리하고 모형화하는 요약이 모두 이 개념의 발달에 의존하는 것이다.

이와 관련해, Fisher(1925)는 자료의 축약에서 채택된 통계적 과정의 목표는, 무관한 정보를 제거하고[9] 자료에 포함된 관련된 정보 전체를 분리시키기 위한 것이라고 하였다

대푯값과 산포도 단원에서 앞서 언급한 요약개념발달의 네 가지 관점은 다음과 같이 더 구체화하여 생각하려고 한다.

첫째, 대푯값으로는 다양한 값들이 있고 이들이 대푯값으로 사용되는 것은 경험적인 것이라고 추측되는데, 중요한 것은 자료에 따라, 상황에 따라 더 적절한 대푯값의 종류가 달라진다는 점이다. 그렇기 때문에 얻어진 통계량에 대한 합목적성 검토는 항상 중요하다.

그런데 합목적성을 검토하려면 가장 중요한 것이 대푯값에 대한 언어적 정의이다. 왜냐하면 구한 통계량이 실제로 대푯값의 언어적 정의에 부합되는지를 확인하려면, 대푯값의 정의가 명확해야 하기 때문이다.

따라서 '대푯값의 정의와 합목적성'에서는 대푯값의 정의가 적절한지를 살펴보고, 또 소개되는 여러 종류의 대푯값들이 대푯값으로서의 합목적성을 가지고 있는지를 살펴볼 것이다.

---

8) 목적이 다양해진다는 것은 양적 변화 정도가 아니다. 요약(통계)의 궁극적 목적은 가설검정 같은 통계적 추론까지 이어지는 경우가 많으며, 이것을 염두에 둔 것이다.

9) 이것은 결국 통계의 산포도를 축소시켜, 추론의 신뢰도를 높이는데 기여하게 된다.

둘째, '산포도의 정의와 실용성'에서는 산포도의 정의가 적절한지를 살펴보고, 산포도를 왜 구해야 하며, 그 실용적 가치가 어떠한 지에 대한 충분한 설명과 예가 필요한지를 살펴볼 것이다. 특히 산포도와 예측의 신뢰성과의 관계를 이해시키려 했는지를 살펴보고자 한다. Moore(1990)는 통계학은 산포도에 관한 학문이라고 하였다.

즉, 적당한 실제 자료를 가지고 미래의 일을 예측할 때, 산포도가 작거나 큼이 예측의 신뢰도를 어떻게 변화시키는지 경험하게 해 주는 것이 산포도 교육의 핵심이어야 한다.

셋째, '분포의 대푯값과 산포도'에서는 '자료'의 대푯값과 산포도가 아닌 '분포'의 대푯값과 산포도를 교과서들이 다루고 있는지를 살펴보려고 한다. 가령 대푯값의 하나인 표본의 평균을 구하는 것이 자료 값들의 중간 정도 크기의 값을 알기 위한 것이면 (첫 번째 연구문제에서처럼) 얻어진 표본 평균이 자료 값들 중에서 중간 정도 크기인지 여부만 살펴보아도 충분하다. 그런데 표본의 평균을 구한 것이, 표본의 대푯값을 알기 위해서가 아니라, 모집단의 대푯값을 알기 위해서 구한 경우에는, 문제가 훨씬 더 복잡해지며, 이것은 통계적 추론의 문제로 비화된다. 이 때 이 문제의 구조는 방정식 문제 구조와는 전혀 다르며, 이것을 바르게 이해하기 위해서는, 분포의 대푯값을 먼저 이해해야 할 것이라고 추측된다. 왜냐하면, 가령 모집단의 평균(대푯값 중의 하나)은 그 모집단의 분포의 기대값, 즉 $E(X)=\int_{-\infty}^{\infty} x dF(x)$ (단, $F(x)$는 모집단 분포의 누적분포함수)와 같은 것이기 때문이다. 우리는 교과서의 설명이 처음에는 도수분포나 상대도수 분포에서 이런 것들을 살펴보다가, 고등학교에서는 확률분포에서의 그것들로 확장되기를 기대한다.

분포의 대푯값이나 산포도는 흔히 통계적 추론의 문제에서 모집단의 대푯값과 산포도로 이어지고(같아지고), 결국 표본 자료에 대한 통계적 요약(통계)이 궁극적으로 추구하는 대상(모집단의 그것)의 실재성에 대한 인식으로 이어지기 때문에 중요한 것이다.

우리가 무엇을 구하고자 하는데, 그것이 존재하는지 여부조차 생각해보지 않았다면, 그에 따른 모든 추론의 과정이 모호하고, 바르게 이해되기는 어려울 것이다. 표본(자료)에서 계산이나 비판 등이 이루어지고 있는데, 그 추구하는 실재는 거기에는 없고, 모집단에 있는 것이라면, 구하려는 대상의 실재성은 모호할 것이 충분히 예상되기 때문이다. 이때 표본과 모집단의 대푯값이나 산포도를 이어주는 다리로 연구자가 삼은 것이 바로 분포의 그것들인 것이다.

마지막으로, '위치 보존성과 단위 보존성'에서는 대푯값은 위치 보존성과 단위 보존성이 있고, 산포도는 위치 불변성과 단위 보존성이 있는 것이 각각의 합목적성 측면에서 좋음을 소개하고 있는지에 대해 살펴보고자 한다. 이런 위치 보존성과 단위 보존성은 수학교과의 특징과 고차원적 비판 정신 함양 측면에서 이러한 과정이 필요하다고 볼 수 있기 때문이다. 가령 표준편차가 대푯값으로서 왜 적절치 못한 지 그 이유를 논하라고 한다면, 어떻게 그 이유를 설명할 것인가?

즉 통계를 구하는 공식들이 목적한 언어적 개념에 대해 (개념)논리적으로 타당하고 일관성 있는 수리적 성질을 가지고 있는지에 대해 점검하는 과정이 필요하다. 물론 이미 공식으로 나온 것이니 의심 없이 사용할 수도 있으나, 수학교과의 특징과 고차원적 비판 정신 함양 측면에서, 연구자는 이러한 과정이 필요하다고 생각한다. 나아가 구한 통계량의 합목적성을 재확인 시켜주는 과정이라고 생각할 수도 있다.

이들 네 가지 측면에 대해 좀 더 구체적으로 의미하는 바와 그 가치를 재고해 보기로 한다.

가. 대푯값의 정의와 합목적성

자료나 분포를 요약할 때, 사람들은 우선 자료의 값들의 크기에 주목하는 경향이 있어 보이는데, 이것은 남주현(2007)의 연구에서 대푯값의 역사적 출현 배경에서도 확인할 수 있다. 또 Shaughnessy(1997)의 연구 또한 결국 역설적으로 자료에 대한 교육과정에서의 가장 중요한 인지 특성은 "자료값들의 크기"라는 특성임을 간접적으로 확인시켜주는 것이다. 이것은 결국 요약개념의 발달과정에서 "자료값의 크기"를 요약하는 방법에 대한 교육과정이 필수적임을 간접적으로 보여주는 것이다. 그리고 대푯값 또는 중심경향값이라는 이름으로 세계 여러 국가의 교육과정에서 가르치고 있는 내용이다.

그런데 이 과정에서 중요한 것은 이와 같은 요약이

타당성이 있음을 학생들이 이해해야 한다. 즉 언어적으로 표현된 “자료값들 전체의 크기”를 한 개의 숫자로 요약하려 할 때, “어떤 숫자가 좋겠는가?”라고 한다면 대개 “중간 정도 크기의 숫자”가 좋다는 것이 일반적 상식이고, 그렇다면 우리는 “자료값 전체의 크기”를 요약하는 대푯값은 중심경향값이 좋겠다는 것이 상식이라고 할 수 있다. 그리고 우리는 그런 목적의 요약으로서 대푯값 또는 중심경향값을 가르친다면 얻어진 요약의 결과가 그 목적에 비교적 잘 맞는지 검토할 필요가 있다는 것이다.

가령 지난 5년 동안 해외 명품의 판매량이 5배로 증가했다고 할 때, 한 해 평균 몇 %씩 판매량이 증가한 것인가? 500% 증가 했으니 연평균 100%씩 증가한 것인가? 이 문제의 답은 $(1+x)^5=5$을 풀어서 구할 수 있다고 할 때, 이것을 같은 “평균”이라는 단어를 사용하지만 산술평균이 아니라 기하평균임을 알 수 있다. 이 문제에서 사용한 “평균”이라는 단어의 의미는 중심경향값을 추구하는 단어의 의미가 아니라 “매년 증가율이 일정하다고 할 때의 증가율”을 의미한다. 즉 초등학교 5학년에서 처음 소개되는 “고르게 만들어 주는 숫자”로서의 평균을 뜻한다. 즉 “지난 한달동안 철수가 읽은 책이 10권일 때 한 주일에 읽은 책은 평균 몇 권?”이라는 “매주 읽은 책의 권수가 같다면 주당 읽은 책의 권수”라는 의미의 평균이다. 그리고 이 경우는 산술평균이 그 목적에 맞는 수이다. 새롭게 중앙값, 최빈값이 대푯값으로서 소개될 때, 구하는 방법에 관계없이 그 목적에 맞는 합목적성의 검토가 중요하다.

가령 최빈값을 구했을 때 얻어진 그 숫자가 자료 중에서 가장 많이 관측된 숫자인지 확인하는 것이 아니라, 자료들 중에서 중간 정도의 크기인지를 확인하는 것이 이 경우의 합목적성에 대한 확인이다. 가령 옷가게에서 구하는 옷의 치수의 최빈값은 “어떤 옷이 가장 많이 팔릴까?”를 알려는 목적에 대해서도 적합하다. 이 경우는 중심경향을 알려는 목적이 아니다. 중학교 3학년 통계 단원에서 배우는 최빈값은 중심경향값이며, 그 목적에 대한 합목적성을 확인하는 교육이 필요하고, 옷가게 등의 다른 목적들에 대한 것은 교과서에서는 설명하고 있지 않다는 뜻이다.

나. 산포도의 정의와 실용성

산포도는 대푯값과 함께 소개되는 자료 또는 분포의 대표적 요약인데, 자료 또는 분포를 구성하는 값들의 크기가 얼마나 변화가 심한지를 보여주는 특성을 나타내는 수리적 측도이다. Cobb & Moore(1997)는 통계학의 핵심 개념으로 산포도에 관한 개념이 필수적임을 주장하고 있는데, 이 점에서 많은 통계학자와 통계교육자들이 동의하고 있다. 통계학에서 추구하는 것은 결국 미지의 우연현상 또는 그 원인에 대한 예측과 판단인데 이것이 어려운 이유는 정확히 예측할 수 없는 변화 때문이다.

관측하려는 현상이 숫자로 얻어지는 것일 경우, 관측값들의 변화가 적으면 관측한 값들의 대푯값이 어느 정도 안정되어 있어서 믿을 만하겠지만 그 반대라면 구해놓은 대푯값의 표집 표본 변화에 따른 가능한 변화 폭이 매우 클 것이 예측되므로 신뢰성이 떨어질 것이다. 여기서 대푯값의 신뢰성의 문제가 제기되고, 이 지점에서 핵심적인 역할을 하는 통계가 바로 산포도인 것이다(윤현진 외, 2009).

산포도를 나타내는 수리적 측도로서 대표적인 것이 분산과 표준편차이다. 그런데 이들의 계산은 다소 복잡하다. 학생들의 흥미를 끌기에는 매우 부족하다.

그러나 Cobb & Moore(1997), Wild & Pfannkuch(1999) 등 많은 통계학자, 수학교육학자들이 주장하듯이 이것은 매우 중요한 내용이다. 그렇다면 산포도의 지도는 학생들의 주의를 끌 수 있을 만큼 학생들에게 중요성이 부각되는 내용서술이 필요해 보인다. 즉 산포도는 왜 구해야 하며, 그 실용적 가치가 어떠한 지에 대한 충분한 설명과 예가 필요하다고 생각된다. 적어도 분산이나 표준편차의 계산과정, 방법 등에 치우친 산포도 단원의 설명방식은 그런 목적을 달성하기에는 매우 미흡해 보인다. 구하는 것이 어떤 목적을 가진 숫자이며, 그런 목적을 추구하는 것이 왜 중요한지가 산포도 단원 서술의 핵심이어야 한다.

적당한 실제자료를 가지고, 미래의 일을 예측할 때 산포도가 ‘작다는 것’과 ‘크다는 것’이 예측이 신뢰성을 어떻게 변화시키는지 경험하는 것이 중요하다.

다. 분포의 대푯값과 산포도

분포의 대푯값과 산포도의 실생활 관련 뿌리는 모집단과 같이 방대한 자료에서 각각의 값보다는 도수분포표 형태나 확률분포 형태로 요약된 자료를 생각하는 경우, 즉 분포의 평균, 중앙값, 최빈값, 분산, 표준편차 등을 생

각하는 경우를 말하는 것이다. 우선은 도수분포표에서의 평균, 중앙값, 최빈값, 분산, 표준편차 등을 생각해야 하고, 나아가 확률분포에서의 기댓값, 분산, 표준편차, 중앙값, 최빈값 등을 단계적으로 시기에 맞추어 즉, 도수분포표를 배운 그 무렵 후, 평균, 중앙값, 최빈값, 분산, 표준편차를 가르쳐야 하고, 확률분포를 배운 그 무렵 후부터 평균, 중앙값, 최빈값, 분산, 표준편차를 가르쳐야 한다는 것이다. 이것이 왜 그래야 하는지, 왜 그렇게 중요한지는 이 절의 뒷부분에서 언급하기로 한다.

확률분포에서 최빈값은 확률(밀도)함수 $f(x)$에서 $f(x)$의 값이 최대인 x의 값을 구하는 것이고, 중앙값은 $f(x)$ 그래프의 좌우가 각각 넓이(또는 높이의 합)을 50%씩으로 나뉘는 x의 값을 찾는 것이다. 도수분포의 경우도 확률분포의 경우에 준한다. 가령 히스토그램의 경우, 최빈값은 최빈 계급의 계급값이고, 중앙값은 넓이가 반반쯤으로 나뉘어 지는 계급을 찾아 그 계급에서 넓이를 비례 배분하여 정확히 좌우측의 넓이가 각각 전체 넓이의 50%씩이 되게 한다. 만약 막대그래프라면 좌측 높이들의 합이 우측 높이들의 합과 같아질 만한 막대를 찾아 그 막대의 도수에 따라 비례 배분하되, 만약 그 막대의 x의 값이 한 가지 뿐이라면 그대로 그 막대를 이루는 x의 값이 중앙값이 된다.

중학교에서는 도수분포에서 대푯값과 산포도 등을 가르쳐야 하는데, 계산해서 값을 구하는 것에 초점이 맞추어질 필요는 없다. 다시 말하자면, 컴퓨터나 계산기를 이용하는 것이 좋고 다만 이 경우도 대푯값으로 구한 값이 분포의 중앙을 나타내는지 즉 중심경향값인지 검토하는 것은 여전히 중요한 일이다.

대푯값과 산포도를 구함에 있어서, 분포로 제시된 경우에 대한 것이 자료 값들 자체로 제시된 경우와 구분하여 별도로 언급할 필요가 있음을 주장하는 이유는 분포에도 대푯값과 산포도가 있다는 점을 알게 하려는 것, 그리하여 분포나 모집단의 관심 특성을 요약하는 일에 대한 경험을 갖게 하려는 것이다. 그래서 고등학교에서 모집단의 평균에 대한 신뢰구간 등을 가르칠 때, 모집단 분포의 평균을 표본의 평균으로 추측하는 일에 대하여, 자신이 무엇을 하고 있는지에 대한 것을 명확히 이해할 수 있도록 하는 준비과정이 반드시 필요하다는 것이다.

보통 중학교 1학년에서 히스토그램이 소개된다. 히스토그램은 상대도수의 히스토그램에서 멈추면 안 된다. 상대도수를 계급의 크기로 나눈 것을 밀도도수라 하는데 밀도도수의 히스토그램이 도입되어야 한다. 밀도도수 히스토그램에서 직사각형의 세로(높이)는 밀도도수를, 가로(밑변)는 계급의 크기를 나타냄으로 직사각형 한 개의 넓이는 상대도수를 나타내게 되어 직사각형 전체의 넓이의 합이 1이 된다. 즉 넓이가 상대도수이고 이것이 극한적인 통계적 확률로 연결되면 넓이가 확률이 되고 이것이 확률밀도함수를 이해하는 바탕이 된다. 즉 상대도수의 막대그래프는 높이의 합이 1이고 확률함수로 이어져 높이를 확률로 해석하는 반면, 밀도도수의 히스토그램은 넓이를 확률로 이해하는 확률밀도함수로 이어지는 것이다.

히스토그램을 그리는 이유는 연속변량의 분포상태를 표현하기 위함이다. 연속변량의 분포상태를 표나 그림으로 잘 나타내는 것이 요약 개념 측면에서 바람직한 것이라 할 수 있는데, 대부분의 실용적인 밀도도수의 히스토그램은 확률밀도함수로 수렴하게 된다는 점에서 밀도도수에 대한 도입이 요약 개념에 대한 의미충실한 지도 측면에서 필요하다고 하겠다.(윤현진 외, 2009)

확률밀도함수 $f(x)$에서 분포의 평균은 기댓값

$$E(X)=\int_a^b xf(x)dx=\sum_{i=1}^{k}x_{i:k}\{f(x_{i:k})\Delta x_{i:k}\}$$

(등호 맨 우측의 표현은 확률밀도함수가 밀도도수 히스토그램 $f(x)$인 경우에서, k개의 계급 중에서 i번째 계급의 계급값 $x_{i:k}$에서의 계급의 높이 $f(x_{i:k})$와 계급의 폭 $\Delta x_{i:k}$의 곱은 그 계급의 상대도수임으로, 또 $\sum$(계급값)×(상대도수)=평균임으로, 도수분포표에서의 평균 구하기는, 구분구적법을 고려한 정적분의 정의에서, 확률밀도함수에서의 기댓값과 연결된 주제임을 알 수 있다.)로 기댓값이란 사실상 분포의 평균임을 알 수 있다.

이런 맥락에서 분포에 대한 요약은 도수분포, 상대도수분포, 밀도도수의 분포 등에서 충분히 소개되어야, 비로소 확률밀도함수나 확률함수의 분포의 요약이 자료와 관련성을 갖고 이해될 수 있으며, 비로소 모집단의 분포와 관련되는 $m=E(X)$, $\sigma^2=E[(X-m)^2]$ 등이 명확히 그 실체가 파악되어, 모집단의 평균, 분산 등, 통계적 추론의 명확한 대상으로서의 실재가 확인될 수 있게 되는 것이다. 이것이 분명치 않으면, 자칫 통계적 추론이 무엇을 하는 것인지, 혹시 방정식의 풀이와 유사한 것은

아닌 지 등, 근본적인 문제 자체에 대한 혼란이 발생될 우려가 있다.(이영하, 2009)

한편, 중앙값이나 최빈값에 대한 분포에서의 논의는 이렇게 도수분포로부터 확률분포로 확장 가능하며 산포도에서 분산, 표준편차 등도 같은 형태의 확장이 가능하다. 다만 이 모든 논의는 '자료의 분포'와는 분명하게 구별되어 체계적으로 논의될 필요가 있으며 실용적이고 경험적인 도수분포로부터 점진적으로 추상화되는 과정에 주의를 기울여야 할 것이다(윤현진 외, 2009).

자료에 대한 통계적 요약의 문제와 관련하여 분포의 대응되는 실재에 관한 논의의 결여는 학습자의 개념이미지 속에 그 실체가 규명되지 않음으로써, 평균 따로, 기댓값 따로, 히스토그램 따로, 확률밀도함수 따로, 자료 따로, 확률론 따로 등, 전혀 정합적(coherent) 교육과정의 모습을 기대할 수 없게 되는 것이다. 따라서 자료의 요약 못지않게, 막대그래프, 히스토그램(현재 히스토그램에서의 평균만 도수분포표에서의 평균이라는 이름으로 홀로 소개되고 있음)으로부터 확률밀도함수에 이르기까지 분포의 요약이라는 또 하나의 맥락이 필요하며 이 틀 속에서 여러 가지 모수(모집단의 최빈값, 중앙값 등)와 기댓값들(모평균, 모분산 등)이 소개되어야 할 것이다.

라. 위치 보존성과 단위 보존성

자료 $x_1, x_2, \cdots, x_n$으로부터 계산된 통계량을 $u(x_1, x_2, \cdots, x_n)$으로 적기로 할 때, 통계 u가, 임의의 양수 c에 대하여, $u(cx_1, cx_2, \cdots, cx_n) = c\,u(x_1, x_2, \cdots, x_n)$이 성립하면 단위보존성(scale preserving)이 있고, $u(cx_1, cx_2, \cdots, cx_n) = u(x_1, x_2, \cdots, x_n)$이면 단위불변성(scale invariant)이 있다고 한다. 또 임의의 실수 b에 대하여 $u(x_1+b, x_2+b, \cdots, x_n+b) = u(x_1, x_2, \cdots, x_n) + b$이면 통계 u는 위치보존성(location preserving)이 있고, $u(x_1+b, x_2+b, \cdots, x_n+b) = u(x_1, x_2, \cdots, x_n)$이 면 위치불변성(location invariant)이 있다고 한다.

대푯값과 산포도의 언어적 정의는 위치, 단위 보존성과 불변성을 검토할 것을 요구한다. 대푯값은 자료값들의 크기(위치)를 대표한다고 하였으므로 이런 언어적 정의가 타당성을 지닌 것이려면 자료 $x_1, x_2, \cdots, x_n$의 모든 값들이 모두 b만큼씩 커져서 $x_1+b, x_2+b, \cdots, x_n+b$가 되었을 때 $u(x_1, \cdots, x_n)$과 $u(x_1+b, \cdots, x_n+b)$ 사이에 위치 보존적 관계가 성립해야 한다. 즉 u의 값도 b만큼 커져야 u가 자료의 위치(크기)를 나타내는 성질이 있다고 할 수 있을 것이다. 즉 대푯값은 그 언어적 정의로부터 위치, 단위 보존성이 있어야 타당성 있는 대푯값이라 할 수 있다. 반면 산포도는 주어진 값들의 변화 정도를 나타낸다고 했으므로 값들의 크기가 모두 b만큼씩 커졌을 때 변화 정도에는 변함이 없으므로, 결국 산포도는 위치 불변성이 있어야 하며, 그러나 단위보존성은 있어야 한다고 볼 수 있다.

이렇게 통계값을 구하는 공식들이 목적에 적절한 수리적 성질을 가지고 있는지에 대해 점검하는 과정이 필요하다. 물론 이미 공식으로 나온 것이니 의심 없이 사용할 수도 있으나 수학교과의 특징과 고차원적 비판 정신 함양 측면에서 이러한 과정이 필요하다고 볼 수 있기 때문이다. 나아가 구한 통계량의 합목적성을 재확인 시켜주는 과정이라고 생각할 수도 있다. 결국 실제 사용하는 대푯값(평균 등)의 공식이나 산포도(표준편차 등)의 공식들이 위치, 단위와 관련된 그런 측면을 잘 반영하고 있는지를 검토해 보는 과정이 필요하다는 것이다. 즉 이런 과정이 주어진 계산 공식의 합목적성 또는 타당성에 대해 비판 정신을 갖고 점검해 보는 과정이며 수학교과의 일부로서 이런 과정이 반드시 포함될 필요가 있다고 하겠다(윤현진 외, 2009).

자료의 통계적 요약에 대한 합목적성 검토는 통계학의 전문적 시각에서는 매우 다양하다. 여기서 생각해 본 위치, 단위, 보존성, 불변성의 문제는 그런 전문적 시각의 일부에 포함된다. 그러나 이것은 통계적 요약의 목적과 개념적 일관성을 유지하는 가장 기초적 확인 과정이라고 할 수 있다. 통계적 요약의 합목적성에 관하여 우리는 중등학교에서 전문적일 것을 요구하는 것이 아니라 상식적 언어 개념 수준에서 합목적성을 수리적으로 확인해 보려는 것 뿐이다.

Ⅲ. 연구방법 및 분석

1. 분석대상 및 분석방법

본 연구에서 분석하고자 하는 중학교 3학년 교과서는

모두 14종[10](M1-M14)으로 분석대상 교과서를 다음 내용의 서술방식과 관점에 따라 하나씩 읽어보아 유사한 것끼리 분류하여 각 내용의 서술방식과 관점을 상호・비교하였다. 상호・비교하는 주제별 관심내용은 <표 Ⅲ-1>과 같다.

<표 Ⅲ-1> 분석틀

분석내용	
1. 대푯값의 정의와 합목적성	교과서들의 서술 방식이 대푯값의 정의로서 적절한가? 대푯값의 정의에 비추어 볼 때, 얻은 계산결과의 합목적성을 평가하고 있는가?
2. 산포도의 정의와 실용성	교과서들의 서술 방식에서 산포도의 정의는 적절한가? 산포도의 실용적 중요성에 대해 언급하고 있는가?
3. 분포의 대푯값과 산포도	자료가 아닌 분포의 대푯값과 산포도를 소개하고 있는가? 또 이 과정에서 대푯값과 기댓값의 연결이 자연스럽게 연계될 것이라고 예상되는가?
4. 위치 보존성과 단위 보존성	위치보존성과 불변성, 단위 보존성과 불변성이 소개되고 있는가?

비교하는 주제별 관심내용에 대해 살펴보면, 첫 번째, '대푯값의 정의와 합목적성'에서는 중심경향값으로써 대푯값을 어떻게 정의하였는지 알아보고자 한다. 또한 중심경향값의 합목적성은 구해진 값이 전체 자료와 비교할 때, 실제로도 중간 정도의 크기를 나타내 주어야 하는 것이다. 그런데 어떤 중심경향값은 그렇지 못한 경우가 발생한다는데 문제가 있다.

일반적으로 모든 통계적 요약은 목적을 가지고 자료로부터 계산된 통계량이다. 이것이 목적에 실제로 맞는지 여부를 확인하는 것은 통계량 계산의 지도에서 가장 중요한 것이다. 따라서 대푯값의 정의에 비추어 볼 때, 얻은 계산결과의 합목적성을 평가하고 있는가를 살펴보겠다.

두 번째, '산포도의 정의와 실용성'에서는 산포도를 어떻게 정의하였는지 살펴보고자 한다. 또한 산포도를 나타내는 수리적 측도로서 대표적인 분산과 표준편차를 왜 구해야 하며, 그 실용적인 가치가 어떠한 지에 대한 충분한 설명과 예가 있는지를 살펴보고자 한다. 즉 산포도의 실용성을 자료 자체의 산포도에 관한 문제로써 설명했는지, 자료에서 계산된 통계량의 산포도로써 설명했는지, 구체적인 이야기로 설명했는지를 살펴볼 것이다.

세 번째, '분포의 대푯값과 산포도'에서는 대부분의 교과서에서 대푯값과 산포도에 대해 도입을 할 때 주로 자료에 대한 대푯값과 산포도를 예로 들면서 설명하고 있는데, 이런 자료의 대푯값과 산포도 뿐만 아니라 분포의 대푯값과 산포도도 설명하고 있는지를 살펴볼 것이다. 또한 이 과정에서 대푯값과 기댓값의 연결이 자연스럽게 연계될 것이라고 예상되는지를 살펴볼 것이다. 더 나아가, 분포의 대푯값과 산포도에서 말하는 분포를 중학교 3학년이 되기 전까지 배웠던 분포인 막대그래프, 도수분포표, 도수분포 다각형, 줄기와 잎 그림, 히스토그램, 꺾은선 그래프의 관점에서 다루었는지를 살펴볼 것이다.

마지막으로, '위치 보존성과 단위 보존성'에서는 마지막으로 대푯값은 위치보존성, 단위보존성이 있고, 산포도는 위치불변성, 단위보존성이 있는지를 중학교 3학년이 다룰 수 있는 문제 수준에서 소개하고 있는지를 살펴볼 것이다.

2. 분석내용과 결과

가. 대푯값의 정의와 합목적성

1) 대푯값의 정의

우리나라의 7차 교과서에서는 평균만을 다루고 있는데 7차 개정 교과서로 바뀌면서 중학교 3학년 교과서에서 중앙값과 최빈값이 도입되었고, 이런 평균, 중앙값, 최빈값을 통틀어 대푯값이라는 용어를 사용하고 있다. 중학교에서 가르치고 있는 대푯값이라는 용어를 대학교 교재에서 살펴보자면 잘 찾아볼 수 없다. 대신 '중심위치의 측도(김용대 외, 2009; 구자용 외, 2003)', '자료의 중심에 관한 측도(이외숙 외, 2007)'로 사용하고 있음을 알 수 있다. 대푯값(representative value)이란 용어는 미국에서 60~70년대 주로 사용되던 용어로 현재는 잘 사용되지 않고, 중심경향값, 중심위치의 측도 또는 자료의 중심

10) 본 논문의 교과서 분석 결과의 설명에서는 해당 교과서가 어느 교과서인지 분명히 밝힐 필요가 없으므로, 익명으로 적기로 한다. 구체적인 것을 알고자하면 이은희(2011)를 참고하면 알 수 있다.

에 관한 측도(measure of central tendency)로 사용하고 있다. 그래서 현재 대학교 교재에서 중심위치의 측도, 자료의 중심에 관한 측도란 용어를 사용하고 있는 것이다. 지금 주로 사용하고 있는 용어이기도 하지만 중학교에서 대푯값으로 배웠던 아이들이 대학에 가서 혼동을 일으키지 않도록 하는 배려가 필요해 보인다.

대푯값을 구체적으로 어떤 방식으로 정의하였는지 살펴보고, 그 종류에 대한 설명과 역할에 대해 언급하고 있는지 살펴보았다. 대푯값을 어떻게 구할 지에 대하여 컴퓨터나 계산기에 의존한다면, 대푯값 수업의 초점은 계산방법 대신 구해진 대푯값에 대한 합목적성 평가에 놓일 수밖에 없기 때문에, 대푯값의 바른 정의는 매우 중요한 문제로 대두된다.

14종의 교과서 중 8종의 교과서(M2, M3, M5, M7, M9, M11, M12, M14)가 "자료의 전체적인 특징을 대표하는 것"이라고 되어 있는데, 적절치 못한 정의라고 생각된다. 그 이유는 자료의 특성은 한두 가지가 아니기 때문이다. 그런 정의에 따르면, 자료의 특성을 나타내는 것은 최댓값, 최솟값, 범위, 분산, 표준편차, 기타 자료의 특성을 나타내는 무슨 숫자이든지 대푯값이 될 수 있다. 가령 {32, 31, 37, 30, 36, 39}의 자료의 특성으로 10의 자릿값이 모두 3으로 되어있다는 것 때문에 대푯값을 3 또는 30이라고 한다면 이것이 옳지 않은 이유를 어떻게 설명할 것인가?

또한 교과서 4종(M1, M6, M8, M10)은 "자료의 중심적인 경향을 하나의 수로 나타낸 것"이라고 서술하고 있다. 그런데 자료의 중심적인 경향이라는 말은 다소 생소한 언어이므로 이렇게 정의를 소개할 때에는 중심적인 경향이라는 말을 다시 정의해 주는 것이 옳다고 생각된다. 교과서 M4는 "자료가 분포한 중심의 위치를 대표적으로 나타낸 값"이라고 서술하고 있는데, 드러나 보이는(explicit) 정의로 학생들에게 좀 더 쉽게 다가갈 수 있을 것이라고 생각되나 만약 자료의 값으로 30, 40, 50, 10, 20이 주어져 있을 때, 50을 지칭하는 말로 해석될 수 있다. 따라서 이를 수정·보완하자면 "자료를 크기 순서로 나열할 때 자료가 분포한 중심의 위치를 대표적으로 나타낸 값"이라고 하는 것이 옳은 표현이라 생각된다. 마지막으로 교과서 M13의 정의를 살펴보면 "자료가 어떤 값을 중심으로 분포되어 있는지를 나타낸 값"으로 서술하고 있는데, 가장 타당한 정의로 생각되지만 내포적(implicit)[12] 정의이기 때문에 이해하기 어렵다는 학생이 생기기 쉽다는 단점이 있다.

2) 대푯값의 합목적성

7차 교과서에서는 대푯값이 평균만 소개되어서 대푯값이 곧 평균이라는 오개념을 심어줄 수 있었다. 반면에 7차 개정 교과서에서는 대푯값을 평균 이외에도 중앙값, 최빈값을 소개함으로써 대푯값의 의미를 좀 더 잘 이해할 수 있게 하였다.

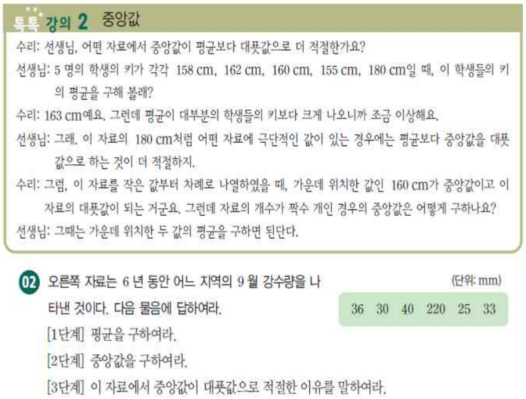
대푯값의 합목적성에 대한 교과서 분석 결과, 대부분의 교과서에서 대푯값의 합목적성에 대한 직접적인 언급은 없었고 초등학교 5학년 때 평균을 배웠다는 전제 하에 중앙값을 도입할 때 자료의 값 중에서 다른 변량에 비해 매우 크거나 매우 작은 값인 극단값이 있을 때 평균이 그 값에 영향을 많이 받는다는 단점을 소개하고 난 후에, 평균이 중심경향값으로써 부적절한 경우, 중앙값을 도입하면서 대푯값의 합목적성을 간접적으로 언급하고 있었다. 이렇게 극단값을 언급하면서 중앙값의 필요성을 언급한 교과서는 12종(M1, M3, M4, M5, M6, M7, M8, M9, M10, M12, M13, M14)이었다.

<그림 III-1>은 교과서 M3의 중앙값 도입에 대한 내용이다. 교과서 M3에서는 중앙값이 평균보다 대푯값으로써 더 적절한 예로 극단값이 존재하는 5명의 학생의 키인 158cm, 162cm, 160cm, 155cm, 180cm를 자료로 주었다. 이 때 평균은 163cm로, 평균이 대부분의 학생들의 키보다 크게 나왔다. 그러므로 선생님은 이때 중앙값인 160cm가 대푯값으로써 적절하다고 설명하고 있다.

중앙값을 도입할 때, 아무런 언급없이 중앙값의 계산과정만을 소개한 교과서는 2종(M2, M11)이었다. 또한 교과서에 나타난 보기나 문제도 극단값이 존재하지 않는 신발의 크기,우리 반 학생들이 등교하는 데 걸리는 시간, 범이네 모둠 학생 9명을 대상으로 지난 한 달 동안 관람한 영화의 편수, 일출시각 등이었다. 이렇게 중앙값을 도입하게 되면, 어떤 경우에 평균을 구하는지 어떤 경우에

12) 내포적(implicit) 표현과 드러나 보이는(explicit) 표현이란, 가령 y를 설명하기 위해 "$f(x)$인 것을 y라 한다." 같은 것을 드러나 보이는(explicit) 표현이라고 하고, "$f(x,y)=0$인 그런 y"라고 표현하는 것을 내포적(implicit) 표현이라고 한다.

중앙값을 구하는지 잘 구별하지 못하는 문제가 발생할 수 있다.


톡톡 강의 2 중앙값

수리: 선생님, 어떤 자료에서 중앙값이 평균보다 대푯값으로 더 적절한가요?

선생님: 5 명의 학생의 키가 각각 158 cm, 162 cm, 160 cm, 155 cm, 180 cm일 때, 이 학생들의 키의 평균을 구해 볼래?

수리: 163 cm예요. 그런데 평균이 대부분의 학생들의 키보다 크게 나오니까 조금 이상해요.

선생님: 그래. 이 자료의 180 cm처럼 어떤 자료에 극단적인 값이 있는 경우에는 평균보다 중앙값을 대푯값으로 하는 것이 더 적절하지.

수리: 그럼, 이 자료를 작은 값부터 차례로 나열하였을 때, 가운데 위치한 값인 160 cm가 중앙값이고 이 자료의 대푯값이 되는 거군요. 그런데 자료의 개수가 짝수 개인 경우의 중앙값은 어떻게 구하나요?

선생님: 그때는 가운데 위치한 두 값의 평균을 구하면 된단다.

02 오른쪽 자료는 6 년 동안 어느 지역의 9 월 강수량을 나타낸 것이다. 다음 물음에 답하여라.

(단위: mm)

36	30	40	220	25	33

[1단계] 평균을 구하여라.

[2단계] 중앙값을 구하여라.

[3단계] 이 자료에서 중앙값이 대푯값으로 적절한 이유를 말하여라.

<그림 III-1> 중앙값의 도입 (M3)

나. 산포도의 정의와 실용성

1) 산포도의 정의

산포도의 정의에 대해 14종의 교과서 중 8종의 교과서(M2, M3, M6, M7, M8, M10, M11, M14)가 서술하고 있는 방식인 "변량들이 흩어져 있는 정도를 하나의 수로 나타낸 값"을 살펴보면, 가장 포괄적이고 가장 적합한 정의라고 볼 수 있다. 그에 비해, 교과서 5종(M4, M5, M9, M12, M13)이 산포도의 정의로써 사용하고 있는 "자료들이 대푯값 주위에 흩어져 있는 정도를 하나의 수로 나타낸 값"이란 표현은 학생들의 이해를 도울 수 있는 정의라고는 생각되나 정확한 정의는 아니라고 생각된다. 이런 표현을 쓰고자 한다면, 대푯값으로써 중앙값과 최빈값을 도입하였듯이 산포도로써 평균을 중심으로 한 표준편차와 분산뿐만 아니라 중앙값을 중심으로 한 평균편차[13](MAD ; mean absolute deviation)의 내용 추가문제나 대푯값이 최빈값인 경우에 난처한 상황이 발생될 우려가 있다[14]. 마지막으로 교과서 M1이 산포도의 정의라고 사용하고 있는 서술방식인 "변량들이 평균을 중심으로 흩어져 있는 정도를 하나의 수로 나타낸 값"은 틀린 표현이라고는 할 수 없으나, 자료의 대푯값으로써 중앙값을 사용한 경우는 어떤 방식으로 산포도를 구할 것인지에 대해 알 수 없어 제한된 의미의 서술방식을 사용하고 있다고 생각된다.

2) 산포도의 실용성

이 절에서는 산포도의 실용성에 대해 자료 자체의 산포도에 관한 문제로써 설명하고 있는지, 자료에서 계산된 통계량의 산포도에 관한 문제로써 설명하고 있는지, 구체적인 이야기로 설명하고 있는지를 살펴보았다. 이 결과를 정리하여 표로 나타내면 다음과 같다. 이 표에 나타난 교과서의 수는 다른 표와 달리 각각의 경우가 배타적이지 않은 경우로 나누어지므로 중복해서 세었다.

먼저 산포도의 실용성을 설명할 때, 자료 자체의 산포도에 관한 문제로써 설명하고 있는지를 살펴보았는데, 모든 교과서인 14종의 교과서에서 이런 방식으로 산포도의 실용성에 대해 설명을 하고 있었다.

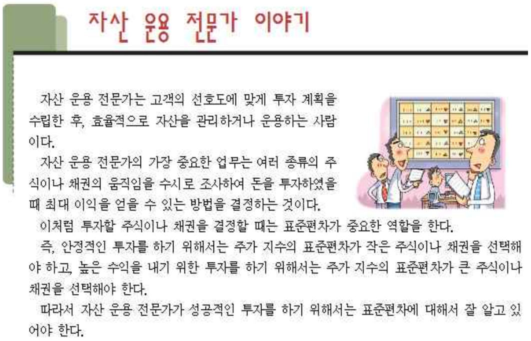
예를 들어 살펴보면, 양궁과 사격에 관한 문제가 빈도수가 가장 높았다. 거의 모든 교과서가 양궁과 사격에 대한 문제를 다루고 있을 뿐만 아니라 산포도를 도입할 때도 양궁과 사격에 관한 문제가 가장 많았다. 그 외에 (그림 III-2>의 만화나(M4 외 6종), <그림 III-3>의 자산운용 전문가(M7)의 실례를 통한 예도 좋은 예라고 생각된다.

<그림 III-2> 산포도의 실용성에 관한 구체적인 이야기 (M6)

---

13) $MAD = \dfrac{\sum_{i=1}^{n} |x_i - 중앙값|}{n}$

14) 산포도는 중심이 되는 값이 존재하지 않아도, 자료 값들 상호 간의 거리의 합으로 정의 될 수 있다.


자산 운용 전문가 이야기

자산 운용 전문가는 고객의 선호도에 맞게 투자 계획을 수립한 후, 효율적으로 자산을 관리하거나 운용하는 사람이다.

자산 운용 전문가의 가장 중요한 업무는 여러 종류의 주식이나 채권의 움직임을 수시로 조사하여 돈을 투자하였을 때 최대 이익을 얻을 수 있는 방법을 결정하는 것이다.

이처럼 투자할 주식이나 채권을 결정할 때는 표준편차가 중요한 역할을 한다.

즉, 안정적인 투자를 하기 위해서는 주가 지수의 표준편차가 작은 주식이나 채권을 선택해야 하고, 높은 수익을 내기 위한 투자를 하기 위해서는 주가 지수의 표준편차가 큰 주식이나 채권을 선택해야 한다.

따라서 자산 운용 전문가가 성공적인 투자를 하기 위해서는 표준편차에 대해서 잘 알고 있어야 한다.

<그림 III-3> 산포도의 실용성에 관한 구체적인 이야기 (M7)

산포도의 실용성 관련 주제 가운데 가장 중요한 것은 산포도와 통계적 추측의 신뢰도 사이의 관련성이다. Moore(1990)는 통계학을 변이성(variability, =산포도)에 관한 학문이라고 보았다. 통계적 추측에 대하여 “내일도 해가 뜰 것이다.”라는 추측을 생각해 보자. 이 경우처럼, 과거 경험 자료의 산포도가 0인 경우(과거에 매일 아침 해는 떴으며 이 사실에 변화는 없었다.)에는 다시 “내일 해가 뜰 것이다.”라고 예측하는 것은 매우 확신할만한 추측이지만, “내일 비가 온다.”의 예측처럼, 과거 경험 자료의 산포도가 큰 경우에는 예측이 어렵다는 점에 유의할 필요가 있다. 이런 맥락에서 산포도에 관한, 통계적 예측 관련, 직관(산포도가 큰 사건일수록 예측의 신뢰도는 약해진다.)을 경험하게 해 주는 교과 내용이 필요한데, 중학교 수준에서 다루기에는 다소 어려운 문제이기 때문에, 모든 교과서에서 이에 관한 문제를 다루지 않는 것으로 보인다. 그러나 고등학교에서 통계적 추론의 문제를 다루기 전에 적어도 언젠가 한 번은 이 문제가 취급될 필요는 있다고 생각한다.

다. 분포의 대푯값과 산포도

김경란(2007)은 "학습자가 효과적으로 교과 내용을 학습하기 위해서는 교과 특성과 체계에 적합하도록 각 교과내용의 연계성을 고려하여 내용을 조직해야 한다."고 하였다. 수학은 위계성이 강한 교과이므로 연계성이 적절히 이루어지지 못하면 기본 개념에 대한 이해 부족으로 다음 단계의 학습이 이루어지지 못한다.

1) 분포의 대푯값

이 절에서는 분포의 대푯값에 대해 설명하고 있는지를 살펴보았다. 여기서 분포의 평균에 대한 문제의 존재 여부는 중학교 1학년 때 도수분포표에서의 평균을 배웠기 때문에 배운 내용 다시보기를 통해 나와 있는 교과서가 종종 있었다. 그러나 이는 중학교 1학년에서 배웠던 내용 또는 배워야 할 내용이라고 생각했기 때문에, 즉 중3교과서에서 다루어야 할 내용이라고 생각되지 않았기 때문에 포함시키지 않았다.

또한 초등학교 3학년에서는 막대그래프를, 초등학교 4학년에서는 꺾은선 그래프를, 초등학교 5학년에서는 줄기와 잎 그림을, 중학교 1학년에서는 도수분포표와 히스토그램, 그리고 도수분포 다각형을 배웠다고 생각했기 때문에 이런 분포들을 중심으로 대푯값에 관한 문제가 존재하는지를 살펴보았다.

분포에 대한 대푯값의 존재 여부를 살펴본 결과, 14종의 교과서 중 분포에 대한 대푯값을 다루고 있는 교과서는 8종(M1, M2, M4, M8, M9, M10, M12, M14)이었다. 분포에 대한 대푯값의 문제를 다루고 있는 교과서는 8종으로 많아 보이는 듯 보이지만 자세히 살펴보면 도수분포표 4종, 줄기와 잎 그림 4종, 도수분포 다각형 3종, 히스토그램 3종, 막대그래프는 2종, 꺾은선그래프 1종으로 각각의 경우를 따져보면 매우 적었다. 즉 자료에 대한 대푯값에 관한 문제는 다양하게 다루고 있었으나 분포에 대한 대푯값은 다양하게 다루고 있지 않았으며, 각 교과서 별로 따져볼 때는 한 두 문제정도로 매우 적은 편이었다.

2) 분포의 산포도

이 절에서는 분포의 산포도, 즉 분산과 표준편차에 대한 내용을 다루고 있는지 살펴보았다. 모든 교과서에서 “도수분포표에서 분산과 표준편차를 구할 수 있게 한다.”고 하고 모든 교과서에서 도수분포표에서의 산포도를 구하는 방법을 제시하고 있었다. 교육과정의 연계성 측면에서 초등학교 때 배웠던 막대그래프, 꺾은선그래프, 줄기와 잎 그림, 그리고 중학교에서 배웠던 도수분포표, 히스토그램, 도수분포다각형에서의 산포도에 관한 문제의 여부를 살펴보았다.

모든 교과서가 분포에 대한 산포도에 관한 문제를 다

루고는 있지만, 도수분포표에서의 산포도 문제를 제외하고 나면 그 문제 수는 적은 편이었으며, 히스토그램은 10종, 막대그래프는 8종, 도수분포 다각형은 2종, 꺾은선 그래프는 2종의 교과서에서 다루고 있었고, 줄기와 잎 그림은 전혀 다루고 있지 않았다. 교과서에서 줄기와 잎 그림을 다루지 않는 이유는 줄기와 잎 그림에 대한 산포도를 계산하는 방법이 자료에 대한 산포도를 계산하는 방법과 유사하기 때문으로 생각되어진다.

라. 위치 보존성과 단위 보존성

1) 위치 보존성

위치 보존성에 대한 문제를 살펴본 결과, 14종의 교과서 중 6종의 교과서인 M1, M2, M3, M8, M10, M12만이 제시하고 있었다. 좀 더 자세히 살펴보면, 교과서 M1의 수학익힘책(p.117)에서는 회당 20점 만점인 4회에 걸친 영어 수행평가 점수를 제시한 후, 총점이 80점인 이 시험을 100점 만점으로 환산하기 위해 각 회별 수행평가 점수에 5점을 더한 새로운 점수에 대한 평균과 분산을 구하고, 환산 전과 환산 후의 점수를 비교하라고 제시하고 있다. <그림 Ⅲ-4>는 교과서 M1의 수학 익힘책에 나온 위치 보존성에 대한 문제이다.

교과서 M2에 대한 문제를 살펴보면, 수학익힘책(p.126)의 문제 16번을 통해 5명의 학생의 수행 평가 성적이 각각 1점씩 감점되면 평균과 표준편차는 어떻게 되는가에 대해 묻고 있다.

활동 2 다음은 두 학생의 4회에 걸친 영어 수행 평가 점수를 나타낸 것이다. 물음에 답하여라.
(단, 수행 평가는 회당 20점 만점이다.)

이름 \ 회	1	2	3	4	총점
박은주	18	14	16	16	64
김창현	17	19	20	16	72

(1) 두 학생의 영어 수행 평가 점수의 평균과 분산을 각각 구하여라.
(2) 100점 만점으로 환산하기 위해 두 학생의 각 회별 수행 평가 점수에 5점을 더한 값을 새로운 점수로 할 때, 새로운 점수에 대한 다음 표를 완성하여라.

이름 \ 회	1	2	3	4	총점
박은주	23				
김창현	22				

(3) 새로 만들어진 영어 수행 평가 점수의 평균과 분산을 구하고, (1)에서 구한 값과 비교하여라.

1. 대푯값과 산포도

117

<그림 Ⅲ-4> 위치 보존성에 대한 문제 (M1)

교과서 M1과 M2에 나온 예는 어느 상황에서 대푯값이 위치 보존성이 있고, 산포도가 위치 불변성이 있는지를 잘 설명해주고 있어 좋은 예라고 볼 수 있다. 나머지 4종의 교과서인 M3, M8, M10, M12에 나온 예는 평균과 표준편차가 주어졌을 때, 각 변량에 2를 더한 경우의 평균과 표준편차를 구하라고 문제를 제시하고 있는데, 이는 대푯값, 산포도의 위치 보존성, 불변성에 관한 이해를 교사의 수업이나 설명에 맡긴 것이라고 추측된다. 학생이 구해진 평균과 표준편차를 음미하여 처음 자료의 그것들과 스스로 비교해 볼 것이라고는 예상되지 않기 때문이다. 교사의 재량권, 수업방법의 다양성에 의존하는 측면에서는 바람직하기도 하지만, 학생 스스로 교사의 도움 없이 이루어지는 학습의 경우, 자기주도적 학습 측면에서는 부족하다고도 할 수 있는 서술이라고 생각된다.

2) 단위 보존성

단위 보존성에 관한 문제를 살펴본 결과, 14종의 교과서 중 6종의 교과서인 M1, M2, M5, M8, M10, M12만이 제시하고 있었다. 좀 더 자세히 살펴보면, 교과서 M1의 수학 익힘책(p.117)에서는 10점 만점인 5회에 걸친 국어 수행 평가 점수를 제시한 후, 이를 100점 만점으로 환산하기 위해 각 회별 수행 평가 점수를 2배한 값을 새로운 점수로 했을 때, 새로 만들어진 국어 수행 평가 점수의 평균과 분산을 구하라고 제시하고 있다. <그림 Ⅲ-5>는 교과서 M1의 수학익힘책에 나온 단위 보존성에 대한 문제이다.

활동 1 다음은 두 학생의 5회에 걸친 국어 수행 평가 점수를 나타낸 것이다. 물음에 답하여라.
(단, 수행 평가는 회당 10점 만점이다.)

이름 \ 회	1	2	3	4	5	총점
임성준	8	9	8	6	9	40
최예린	7	6	10	5	7	35

(1) 두 학생의 국어 수행 평가 점수의 평균과 분산을 각각 구하여라.
(2) 100점 만점으로 환산하기 위해 두 학생의 각 회별 수행 평가 점수를 2배한 값을 새로운 점수로 할 때, 새로운 점수에 대한 다음 표를 완성하여라.

이름 \ 회	1	2	3	4	5	총점
임성준	16					
최예린	14					

(3) 새로 만들어진 국어 수행 평가 점수의 평균과 분산을 구하고, (1)에서 구한 값과 비교하여라.

<그림 Ⅲ-5> 단위 보존성에 대한 문제 (M1)

이는 위치보존성에 대한 문제와 마찬가지로 어떤 상황에서 대푯값과 산포도에 관한 단위보존성이 필요한

지에 대한 이해가 요구되는지 알 수 있는 문제이다. 나머지 5종의 교과서인 M2, M5, M8, M10, M12에서는 평균과 표준편차가 주어진 상황에서 각 변량에 2배를 했을 때, 평균과 표준편차를 비교해 보라는 방식으로 문제를 제시하고 있는데, 이는 교사의 창의적 자율적 수업 측면에서는 바람직하고, 자기 주도적 학습 측면에서는 부족한 표현이라고 생각된다.

위치 보존성과 단위 보존성을 둘 다 다루고 있는 교과서는 5종(M1, M2, M8, M10, M12)이었지만 어떤 상황에서 위치 보존성과 단위 보존성이 필요한지 상황을 제시한 교과서는 M1으로 1종이었다. 따라서 대푯값이 위치 보존성과 단위 보존성이 있고, 산포도가 위치 불변성과 단위 보존성이 있음을 '대푯값과 산포도' 단원에 제시할 뿐만 아니라 언어적으로 어떤 종류의 통계량을 원했는지를 재검토하여, 어떤 상황에서 위치 보존성과 단위 보존성이 필요한지에 대한 설명이 주어질 때, 대푯값과 산포도에 관련된 여러 통계량의 공식을 음미하고 비판적으로 합목적성을 검토해 보는 경험을 제공할 수 있으리라고 생각된다.

Ⅳ. 결론 및 제언

교과서 분석결과를 이 네 가지 관점에서 요약하면 다음과 같다.

첫 번째 관점에서 보면, 대부분의 교과서에서 대푯값의 정의가 적절하지 못하게 서술되어 있었다. 그 뿐만 아니라 대푯값의 합목적성에 대한 직접적인 언급도 없었다. 대푯값의 합목적성에 대해 간접적으로 극단값이 존재할 때 평균은 대푯값으로써 적절하지 않으며 중앙값이 필요하다고 설명하고 있다. 그러나 이런 과정 없이 중앙값을 도입하면서 중앙값의 계산하는 과정만 소개한 교과서도 있었다.

두 번째 관점에서 보면, 산포도의 정의는 적절한 편이었다. 산포도의 실용성 측면에서는 자료 자체의 산포도에 관한 문제는 모든 교과서가 다루고 있었지만 대부분의 교과서가 사격이나 양궁에 대한 문제를 제시하고 있었으며, 자료에서 계산된 통계량의 산포도에 관한 문제는 전혀 다루고 있지 않았다. 또한 구체적인 이야기를 예로 들어 설명하고 있는 교과서가 6종이었는데, 이 6종 중에서도 5종의 교과서가 동일한 이야기를 예를 들어 산포도의 실용성을 설명하고 있었다.

세 번째 관점에서 보면, 분포의 대푯값에 대해서는 14종 중 8종이 다루고 있었다. 또한 분포의 산포도에 대해서는 14종이 다루고 있었으나 다양한 분포에 대해서는 다루고 있지 않았고, 그 문제 수도 매우 적었다. 또한 대푯값과 이후 확률분포의 기댓값의 연계성 문제에서는 대푯값과 기댓값이 어떠한 연관 관계가 있는지에 대해 알 수 있는 내용적 요소나 문제는 전혀 다루고 있지 않았다. 이는 자료의 평균, 즉 각 자료의 합을 도수의 총합으로 나누는 평균의 값과 이후 확률 분포에서의 기댓값을 자연스럽게 연결을 짓지 못하며 심지어는 평균과 기댓값을 다른 것으로 보는 오개념을 낳을 수 있다.

네 번째 관점에서 보면, 대푯값의 위치 보존성과 단위 보존성, 산포도의 위치 불변성과 단위 보존성을 직접적으로 언급하고 있지는 않았고, 문제로써 간접적으로 이를 언급하고 있었다. 그러나 이 둘을 모두 언급한 교과서는 5종에 불과하였다. 또한 5종 중 어떤 상황에서 위치 보존성과 단위 보존성이 필요한지 그 상황을 제시한 교과서는 1종에 불과하였다.

이에 비추어 요약개념과 함께 내용연계성 측면에서 개선되어야 할 점을 제안하면 다음과 같다.

첫째, 현 교육과정에서는 통계자료의 처리, 즉 기술적인 계산측면을 부각하고 있지만, 통계 교육과정에서 요구하고 있는 것은 대푯값과 산포도에 대한 계산보다는 대푯값의 합목적성에 대한 내용이 더 중요하다.

둘째, 대부분의 교과서가 동일하거나 비슷한 예를 들어 산포도의 실용성을 설명하고 있었는데, 조금 더 다양한 예가 필요하다고 생각된다.

셋째, 교육과정의 연계성 측면에서 초등학교와 중학교에서 배웠던 막대그래프, 도수분포표, 도수분포 다각형, 줄기와 잎 그림, 히스토그램, 꺾은선 그래프 등 다양한 분포에 대한 대푯값과 산포도에 대한 문제를 다루어야 하며, 이 과정에서 대푯값과 기댓값의 연결을 자연스럽게 연계될 수 있도록 해야 한다.

넷째, 어떤 상황에서 위치 보존성과 단위 보존성이 필요한지, 그 상황을 제시할 뿐만 아니라 합목적성이 있는 대푯값과 산포도라면, 각각 위치 보존성과 단위 보존성 등과 관련하여 어떤 성질을 갖추어야 하는지, 또 그

런 관점에서 평균이나 중앙값, 최빈값, 범위, 표준편차 등은 그런 성질을 갖춘 통계라고 할 수 있는지에 대한 구체적인 언급이 필요하다고 생각된다.

참 고 문 헌

교육과학기술부 (2008). 교육인적자원부 고시 제2006-75호 및 제2007-79호 초등학교 교육과정 해설 (4). 서울: 대한교과서 주식회사.

교육과학기술부 (2008). 교육인적자원부 고시 제2006-75호 및 제2007-79호 중학교 교육과정 해설 (3). 서울: 대한교과서 주식회사.

교육과학기술부 (2008). 교육인적자원부 고시 제2007-79호에 따른 고등학교 교육과정 해설 (5). 한국보훈복지의료공단 신생인쇄조합.

교육과학기술부 (2009). 고등학교 확률과 통계. 서울: ㈜ 두산.

교육과학기술부 (2011). 중학교 수학 3. 서울: 교학사.

교육과학기술부 (2011). 중학교 수학 3. 서울: 금성출판사.

교육과학기술부 (2011). 중학교 수학 3. 서울: 더텍스트.

교육과학기술부 (2011). 중학교 수학 3. 서울: 도서출판 지학사.

교육과학기술부 (2011). 중학교 수학 3. 서울: 동화사.

교육과학기술부 (2011). 중학교 수학 3. 서울: 두산동아.

교육과학기술부 (2011). 중학교 수학 3. 서울: 미래엔 컬처그룹.

교육과학기술부 (2011). 중학교 수학 3. 서울: 비상교육.

교육과학기술부 (2011). 중학교 수학 3. 서울: 성지출판.

교육과학기술부 (2011). 중학교 수학 3. 서울: 웅진씽크빅.

교육과학기술부 (2011). 중학교 수학 3. 서울: 지학사.

교육과학기술부 (2011). 중학교 수학 3. 서울: 천재교육.

교육과학기술부 (2011). 중학교 수학 3. 서울: 천재문화.

교육과학기술부 (2011). 중학교 수학 3. 서울: 천재문화.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 교학사.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 금성출판사.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 더텍스트.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 도서출판 지학사.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 동화사.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 두산동아.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 미래엔 컬처그룹.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 비상교육.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 성지출판.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 웅진씽크빅.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 지학사.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 천재교육.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 천재문화.

교육과학기술부 (2011). 중학교 수학익힘책 3. 서울: 천재문화.

교육과학기술부 (2011). 중학교 수학 3 교사용 지도서. 서울: 두산동아.

교육과학기술부 (2011). 중학교 수학 3 교사용 지도서. 서울: 비상교육.

교육인적자원부 (2007). 고등학교 확률과 통계 교사용 지도서. 서울: ㈜ 천재교육.

구자용 외 (2003). 통계학(개정판). 서 : 자유 아카데미.

김경란 (2007). 분포개념 발달 관점에서 본 초 중등 확률통계 교육과정 및 교과내용의 분석. p10, 이화여자대학교 교육대학원 석사학위 논문.

김미옥 (1989). 초・중・고 확률 및 통계영역의 연계성에 관한 연구. 이화여자대학교 교육대학원 석사학위 논문.

김민연 (2010). 중학교 수학교과서 개정에 따른 9단계 통계 단원에 대한 7차 교과서 분석. 이화여자대학교 교육대학원 석사학위 논문.

김용대 외 (2009). 통계학개론(제5개정판). 서울: 도서출

판 영지문화사.

김응환 · 이석훈 (2008). 통계교육. 서울: 경문사.

남주현 (2007). 초 · 중등 통계교육을 위한 통계적 방법론에 대한 연구. 이화여자대학교 박사학위논문.

송순희 · 이영하 · 김미옥 (1989). 초 중 고 수학 교과서의 확률 통계 영역의 연계성에 관한 분석(제 1보). 한국수학교육학회지.

윤소진 (2007). R. W. Tyler의 관점에 의한 9단계와 10단계 수학 교과내용의 연계성에 관한 고찰. 성균관대학교 교육대학원 석사학위 논문.

윤현진 외 (2009). 수학과 교육내용 개선 방안 연구 -'이산수학', '확률과 통계' 영역을 중심으로-. 한국교육과정평가원 연구보고 RRC 2009-3-3.

이명근 (1984). 대학 교양과정 연계성에 관한 연구. 연세대학교 대학원 석사학위 논문.

이영하 (2009). 통계적 개념 발달과 통계교육. 서울시 2009 중등 수학과 1, 2급 정교사 자격연수자료, 627-642.

이영하 (2009). 확률과 통계 영역의 지도. 광주광역시 중등학교 정교사(1급) 수학과 자격연수자료.

이영하 (2002). 확률과 통계. 2002년 서울시 1급 정교사 연수자료.

이영하 · 남주현 (2005). 통계적 개념 발달에 관한 인식론적 고찰. 한국수학교육학회지 시리즈 A <수학교육>, 44(3), 457-475.

이외숙 외 (2007). 통계학 입문(제2판). 서울: 경문사.

이정희 (1992). 확률 및 통계 영역의 연계성에 관한 연구. 이화여자대학교 교육대학원 석사학위 논문.

이중권 (2004). 세계 여러 나라의 수학교육과정. 서울: 경문사.

최지안 (2008). 중학교 1학년 통계단원에서 나타난 분포개념에 관한 분석. 이화여자대학교 교육대학원 석사학위 논문.

Bakker, A. (2004). *Design research in statistics education; On symbolizing and computer tools.* Dissertation Utrecht University, the Netherlands.

Cobb G. W., & Moore D. S. (1997). Mathematics Statistics and Teaching, *American Mathematical Monthly* 104, 801-823

Fisher R. A. (1925). *Statistical methods for research workers*. Oliver and Boyd of Edinburgh in Scotland.

Garfield J. B. (2002). The challenge of developing statistical reasoning. *Jour. of Statistics Education*, 10(3).

Garfield J. B., & Ben-Zvi D. (2004). Research on statistical literacy, reasoning and thinking: Issues, challenges, and implications. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking*(pp 397-409). Dordrecht, the Netherlands: Kluwer Academic Publishers

Huff D. (1954). How to lie with statistics, Norton

Moore D. S. (1990). Uncertainty, In the L. S. Steen(Ed), On the shoulders of giants; *New approaches to Numeracy* 95-137, Washington D.C. National Academy Press

Wild C., & Pfannkuch M. (1999). Statistical thinking in Empirical inquiry, *International Statistical Review* 67, 223-266

A summary-concept based analysis on the representative values and the measures of spread with the 9th grade Korean mathematics textbook

Lee, Young Ha†
Ewha Womans University
E-mail : youngha@ewha.ac.kr

Lee, Eun hee
The Graduate School of Education Ewha Womans University
E-mail : flymettm@hanmail.net

This study is an analysis on the focus of textbooks regarding the statistical chapters of "measures of representative(central tendency) and of the spread". Applying the summary-concept criteria of Juhyeon Nam(2007), 4 kinds of aspect of the chapter; (1)definition and its teleological validity of the measures of representative, (2) definition and practical value of the measures of spread (3) distributional form on the measures of representative and of spread (4) location and scale preservation or invariance of the measures of representative and of spread were observed.

On the measures of representative, some definitions were insufficient to check the teleological validity of the measure. Most definitions of the measure of spread were based on the practical view points but no preparation for the future statistical inferences were found even by implication. Some books mention about the measures of representative and of spread for distributions, but we could not find any comments on the correspondence between the sample mean and the expectation of a distribution or population mean. However it is stimulant that some books check the validity of corresponding measures with the location and scale preservation or invariant property, that were not found in the previous curriculum.

* ZDM Classification : U2
* 2000 Mathematics Subject Classification : 97U20
* Key Words : textbook analysis, measure of representative, measure of central tendency, measure of spread, statistical literacy, statistics education
† Corresponding Author