

초·중·고등학교 확률과 통계 단원에 나타난 표본개념에 대한 분석

이 영 하*, 신 수 영**

본 연구는 고등학교 수학교과에서 배우는 모평균의 신뢰구간 구하기와 같은 통계적 추론 능력을 기르기 위한 방안의 첫 단계연구이다. 통계적 추론과정을 비판적으로 분석하여 신뢰할만한 추론방법으로 이를 인정할 수 있는 표본개념의 형성을 위해, 연구자들은 우연과 필연, 귀납과 연역, 가능성원리, 통계량의 변이성, 통계적 모형 등의 하위 개념들이 형성되어야 한다고 보았다. 그리고 초중등 통계단원의 전 과정에서 이들 개념의 체계적인 발달을 도모해야 한다는 전제 아래, 초·중·고등학교 통계단원을 분석해 본 결과는 아래와 같았다.

첫째, 문제해결 방법 선택의 지도와 관련하여, 통계적 방법을 선택할 문제 상황으로서, 우연적 상황을 필연적 상황과 구분하기 위한 설명이 있는 교과서가 초등학교에는 없고, 중등 수준에서도 매우 드물었다. 둘째 표본의 모집단 관련 의미를 이해시키려는 단계적 준비가 미흡하다고 할 수 있다. 전체와 부분의 모집단과 표본 구분이 고등학교에서 비로소 공식화되고 있으며, 초·중학교에서 사용되는 표본자료는 그것으로부터 얻어지는 계산적 결과에만 초점이 맞추어짐으로서, 학년이 올라감에 따라 모집단을 향한 귀납적 추론의 신뢰성에 대한 비판적 사고의 깊이가 더해지는 모습을 찾아보기 어려웠다. 셋째, 무작위 추출이 갖는 대표성의 의미에 대한 설명보다는 무작위 활동 자체에 대한 설명이 중심이 됨으로서 무작위 추출의 확률적 의미, 즉 무작위 표본을 통해 구해질 통계량의 표집분포에서의 (상속된) 무작위성을 위한 담보로서의 목적에 대한 설명이 없다는 점이다. 넷째 통계적 추론을 수학(연역)적 추론과 구분해 주는 설명이 없을 뿐 아니라, 학습자의 논리성 발달 수준에 맞게 변화하는 가능성원리에 대한 설명, 적용 등을 전혀 찾기 어렵다는 점이다. 다섯째 통계량의 우연변이성과 그에 따른 표집분포의 존재에 대한 이해를 추구하는 설명을 찾기 어렵다는 점이다. 표집분포를 수학적으로 구하는 것은 매우 어려운 과정이지만, 그것의 존재를 인식하느냐 못하느냐는 통계적 추론 자체의 이해 가능성을 달리하는 중요한 문제이기 때문이다.

I. 서론

1. 통계적 추론의 구조적 난해성

A. 인식론적 발문

통계학은 불확실성(uncertainty)을 포함한 어

떤 현상을 예측·추론하기 위해 귀납적 사고를 근간으로 하여 자료를 수집·요약·정리·분석하여 그것을 토대로 합리적인 의사결정을 내리는 학문이다(우정호, 2000-a). 그런데 여기서 '합리적'이라는 말은 어떤 의미로 이해되어야 할까? 이치에 부합된다거나 이성에 비추어 타당하다거나 나아가 상식에 부합된다는 등의 의미로 이해되는 이 표현의 의미는 곰곰이 생각해

* 이화여대 수학교육과, youngha@ewha.ac.kr
** 이화여대 교육대학원, sh7639@daum.net

보면 '귀납적 사고를 근간으로 하여'라는 그 앞의 표현에서 느낄 수 있는 바와 같이 수학교과 의 표준적인 의미와는 사뭇 다른 의미를 내포 하고 있다.

가령 어느 두 복권 한 장의 값도 같고, 상금 도 같을 때, 당첨확률이 하나는 1/(50만)이고, 다른 하나는 1/(100만)이라면 누구나 1/(50만) 의 복권을 사는 것이 상식적으로 '더 합리적'이 라고 할 것이다. 그러나 사실 상 두 복권을 한 번 씩 사서 당첨되는지 살펴보면 모두 다 당첨 되지 않는다는 것은 경험적으로 너무나 잘 아 는 일이다. 사실 이 게임에서 확률의 실질적 영향은 엄청난 반복시행 후에야 비로소 그 의 미를 갖는 것이기 때문이다. 여기서 더욱더 심 각해지는 문제는 '더 합리적'이라는 표현이다. 어떤 수학적 사실을 증명한 서로 다른 두 증명 에서 '더 합리적'인 증명과 '덜 합리적'인 증명 이 있는가?

흔히 통계는 표본을 통해 모집단의 성질을 예 측, 추정, 추론한다. 이것을 통계적 추론(statistical inference)이라고 하는데 따라서 통계적 추론은 전형적 귀납추론이라고 할 수 있다. 이 귀납추 론은, 증명과 같이 표준 수학적 방법에서 유일 하게 인정되는 정당화과정인 연역 논증(deduction) 과는 전혀 다르며, 확률(=논증의 강도)을 함께 이용하여 '더 합리적'인 논증이라는 표현이 인 정되는 귀납논리학이라는 논리학의 비주류 분 야에 속하는 추론이다.

우연현상은 대표성²⁾ 불규칙적 변이성 등의 속성을 지니고 있다. 불확실성의 한 모습은 우 연이며, 우연현상에 대한 추론은 결국 불규칙 한 변이성을 보이는 현상에 대한 추론일 수밖에 없다.

인류는 그에 대해 이랬다저랬다 불규칙적 변

화를 보이는 어떤 설명도 제대로 된 지식으로 인정해오지 않아왔다. 인류에게 지식이란 항상 필연적 인과관계를 전제로 한 것뿐이며, 원인 으로 지시된 것은 필연적으로 그 현상을 불러 일으킬 때, 그 인과관계를 근간으로 한 설명이 비로소 지식으로 대접받을 수 있었다.

반면, 통계적 지식은 그 인과관계가 필연적 은 아니어서 약간의 예외가 있다 해도 대체로 성립되는 것이면 통계적 지식이라는 한계를 갖 고 그 지식에 대한 신뢰도를 붙여 제한된 지식 으로 인정해준다. 통계적 지식 생성의 과정에 핵심적 역할을 하는 것이 통계적 추론인데, 앞 서 설명한바 불규칙적 변이성의 구조적 문제로 인해 우리는 이 추론을 신뢰하기 어렵다고 생 각한다.

B. 표본개념과 통계적 추론

남주현(2007)은 통계적 방법론의 핵심개념으 로 분포개념, 요약개념, 표본개념을 얘기하고 있는데, 그 중에서 통계적 추론의 신뢰성 문제 를 이해하기에 필수적인 개념이 바로 표본개념 이라고 하였다. 여기서 '표본'이라는 단어의 뜻 은 모집단의 일부라는 사전적 의미 외에, 통계 에서 항상 사용되는 '무작위표본'의 무작위성 (randomness), 즉 독립성(independence)과 불규 칩적 변이성(irregular variability), 분포적 항상 성(distributional identity) 등을 상징하는 의미 가 섞여있다.

표본개념은 따라서 (무작위)표본의 이런 속 성을 바탕으로 통계적 추론이 사용하는 특징적 방법들을 이해하기 위한 몇 가지 개념들(우연 과 필연, 통계적 모형, 통계량의 변이성)과 그 속성들을 논리적으로 관련짓기 위해 추론할 때 그것을 이해하기 위해 필요한 몇 가지 개념들

2) 대표성이란 순수한 무작위적 우연현상이 나타내는 속성으로서, 가능한 우연현상 전체의 (무한)분포와 유사한 유한 분포의 출현 확률이 가장 높아지는 특성을 뜻한다. 가령 가능한 경우 전체가 정규분포일 때, 무작 위 표본의 히스토그램은 정규분포 모양으로 된 것을 얻을 확률이 가장 높다는 것을 예로 들 수 있다.

(귀납과 연역, 가능성원리)로 구성된다. 이에 관한 자세한 것은 표본개념의 하위개념들이라는 용어로 다음 장에서 논의하기로 한다.

통계적 추론은 크게 추정(estimation)과 검정(hypotheses testing)으로 구별된다. 또 추정은 다시 점 추정(point estimation)과 구간추정(interval estimation)으로 나뉘어 지는데, 현재 고등학교에서 소개되는 모평균에 관한 신뢰구간 문제는 구간추정의 대표적 문제라고 할 수 있다.

자료를 분석하여 그 특징을 파악하고 그것으로부터 얻은 통계적 결과를 비판적으로 바라볼 수 있는 통계적 소양은 필수적이다. 이러한 지식과 방법은 통계적 추정을 학습함으로써 습득할 수 있다. 통계적 추정은 추출된 자료가 모집단의 특성을 얼마나 유지하는지, 표본으로서 어떤 특성을 가지는지 등을 파악하여 모집단을 합리적으로 추론하고 그 때의 신뢰도를 구하는 것이다(이경화 외, 2005).

추론이란 논리적 사고의 흐름에 대한 일련의 연결된 표현이라 할 수 있으며, 이것이 논리법칙에 맞는 것이어야 그 결과를 신뢰할 수 있는데, 우연현상에 관한 추론은 이 점에 있어서 본질적 속성이 신뢰하기에 충분하지 못함을 앞서 언급하였다.

표본개념의 이해와 적용은 통계적 추론의 근거인 "가능성의 원리(likelihood principle)"라는 논리법칙이 그 핵심으로 작용함을 느낄 수 있을 때 충실해지는데, 이것은 앞의 가능해 보이지 않는 상금을 건 두 가지 게임의 선택에서, 당첨의 실제 여부와는 관계없이, 당첨 확률 1/(50만)의 게임이 더 합리적 선택이 될 수 있는 논리적 근거로 작용한다. 이것은 전통적 논리학의 배반율 같은 논리법칙은 아니어서, 다만 상식처럼 사용되는 법칙이지만, 연구자는 이렇게

'상식'으로 치부되는 것들에 의한 논리성의 정당화방식은 우리가 사용하는 논리적이라거나 합리적이라는 종류의 언어적 표현을 매우 혼란스럽게 만든다고 생각한다. 더욱이 상식적인 것처럼 보이는 가능성원리는 검정(hypotheses testing)에서는 '희귀사건에 관한 통계적 개념(rare event concept of statistics)' 형태로 다시 등장하게 되는데, 이때는 전혀 상식처럼 보이지 않으며, 훨씬 이해되기 어려운 모습으로 등장된다는 점을 지적함으로써, 가능성원리(likelihood principle)는 상식으로 인정되기 보다는 통계적 추론을 위한 중요한 한 개의 논리법칙으로 인정할 필요가 있음을 주장하는 바이다.

C. 통계량의 상속된(inherited) 변이성과 표집 분포

통계적 추정은 표본에서 얻은 정보를 이용하여 모집단의 속성(평균, 표준편차, 등)을 추측하는 방법이다(적분과 통계 지학사, 2010). 통계적 추정에서 가장 유명한 모평균에 대한 구간추정을 이해하기 위해서는 모평균이 μ , 모분산이 σ^2 일 때, 크기(n)가 큰 무작위표본 모뭉 각각의 표본평균 \bar{x} 들의 분포가 정규분포 $N(\mu, \frac{\sigma^2}{n})$ ³⁾를 따른다는 것에 대한 이해가 반드시 선행되어야 한다.

표본평균과 같은 통계량의 분포를 표집분포라고 하는데, 표본평균의 표집분포를 이해하기 위해서는 통계량(표본평균 등)이 확률변수라는 것, 따라서 통계량도 항상 분포를 갖는다는 것(이 분포 $N(\mu, \frac{\sigma^2}{n})$ 를 도출해 내는 것 자체-중심극한의 정리-는 남주현(2007)의 분포개념의 최상위 개념으로 분류된다.)에 대한 인식이 이루어져야 한다. 즉, 한 개의 표본으로부터 계산된 통계량의 값과 함께, 그 본질적 속성을 이해하

3) 평균이 μ 이고 분산이 $\frac{\sigma^2}{n}$ 인 정규분포를 통계학에서 흔히 $N(\mu, \frac{\sigma^2}{n})$ 로 적는다.

여 무작위표본은 추출할 때마다 변한다는 것, 따라서 표본으로부터 계산된 값인 통계량 또한 (상속되어) 변하는 값이라는 것, 그래서 도수분포에 대한 이해와 이것의 확률분포와의 관계에 대한 이해(이들은 모두 분포개념으로 분류됨)를 바탕으로 통계량인 표본평균도 분포를 갖게 된다는 것을 이해할 수 있어야 한다. 이러한 과정과 내용, 논리적 흐름을 이해하는 것이 표본개념의 핵심이다.

2. 연구의 필요성과 연구문제

A. 통계적 추론 지도의 현실

표본개념은 무작위 추출을 통한 표본의 대표성과 함께, 표본의 모집단과는 다른 특징인, 표본 변이성을 이해하고 이로부터 계산된 통계량의 상속된(inherited) 변이성을 인식하면서, 통계량을 확률변수로 인식하고 표집분포의 도입이 자연스럽게 이어지는데 핵심적인 역할을 한다.

통계적 지식을 체계적으로 가르치기 위해서는 학습자의 표본개념 발달 수준을 알고 그에 맞추어 학교 급간의 적절한 연계성을 유지하며 일관성 있게 교과내용을 구성하는 것이 중요하다.

통계적 추정에서의 기본이 되는 표본개념⁴⁾에 대한 지도가 연계성 있게 이뤄지지 않는다면 학생들이 통계적 추정과 관련된 여러 난해한 개념을 익히기 어려우며 그 포괄적인 아이디어를 알지 못 할 것이다(이은섭 외, 2004).

전술한바 통계적 추론의 구조적 난해성은, 우리나라 고등학교 수학교과서 통계적 추론 단원의 지도에 있어서, 그 논리적 과정에 대한 설명을 포기하고 다만 그 실용성에 주목하여 공식만 암기함으로써 문제를 해결하는 형태의 수업이 성행하도록 만들었으며, 그 결과 기술 교과

의 수업처럼 간단한 계산을 통해 답만 구하는 수업으로 이어지고 있다. 여러 선행연구들은 고등학교에서의 모평균의 통계적 추정에서, 상당수의 고등학생들이 신뢰도와 신뢰구간에 관련된 개념을 이해하고 이를 논리적 설명하는 데에 어려움을 느낀다는 것을 확인하였다(류선옥, 2009; 송성수, 2010).

이런 수업이 수학 수업으로서의 정당성을 확보하기는 어렵다고 생각되는데 왜냐하면 수학교과서의 특성상, 학생들이 통계적 추론을 배웠다면, 학생들이 그와 유사한 사고법을 익혀 다양한 실생활 상황에서 스스로 그런 사고를 할 수 있어야 할 것이기 때문이다.

통계적 추론에 관한 본 연구의 분석은 후속 연구를 통해 궁극적으로는 우리나라 고등학교에서 모평균에 대한 구간추정의 논리적 과정을 이해하기 어렵게 하는 원인을 규명하고 나아가 그 해결책을 모색하며 다른 소재, 방법을 통한 통계적 추론의 도입가능성, 또는 현재의 통계적 추론에 대한 개선, 삭제 가능성 등을 포괄적으로 살펴보는 데 그 목적이 있다. 다만, 현 단계에서 본 연구의 목적은 표본개념에 대한 교과서 분석을 통해 초·중등 전 과정에서 통계적 추론을 이해시키기 위한 준비의 실태를 파악하려는 것이 본 연구의 당장의 목적이다.

B. 연구문제

본 연구의 연구문제는 아래와 같이 정하였다.

제7차 개정 교육과정의 개정 방향에 따라 편찬된 초·중·고등학교 수학교과서에 남주현(2007)이 제안한 표본개념의 기본 원리가 적절한 연계성을 가지고 효율적으로 조직되어 있는가?

4) 이은섭의 표본개념은 본 논문에서의 의미와는 차이가 있으나 근본 맥락에서는 일부 공통점이 있어 그대로 사용한다.

II. 이론적 배경

1. 표본개념

A. 표본개념의 의의

“표본개념”은 어떤 목적을 갖고 추출된 표본을 이용하여 통계적 분석을 하였을 때 표본 속성의 변이성을 이해하고, 변이성을 지닌 것임에도 불구하고 이를 통해 어떻게 믿을만한 정보를 획득하는지를 비판적으로 바라보며 정당화하고 그 원리를 이해하는 데 필요한 개념이다.

이 개념을 충실히 갖추기 위해서는 여러 가지 선행되는 하위개념들이 요구된다고 생각된다. 우선 우연현상과 필연현상을 구분할 수 있어야 하고, 모집단과 표본의 구분 및 표본 추출을 위해 전체와 부분을 구별할 수 있어야 한다. 또 추론 형식에 있어서 귀납과 연역의 차이를 이해할 수 있어야 하고, 가능성의 높고 낮음에 따라 합리적 선택을 할 수 있어야 하며, 표본으로부터 계산된 통계량에는 표본 자체의 변이성이 상속되어 있음을 이해할 수 있어야 하고, 우연과 필연이 뒤섞인 현상에 대하여는 이 둘을 분리하여 분석하는 과정이 합리적인 추론을 위해 더 지혜로운 분석임을 느낄 수 있어야 한다. 이들 모두가 하나의 맥락 속에 통합되어 논리적으로 체계적으로 구성된 통계적 추론이 될 때, 우리는 그와 같은 추론이 믿을 만하다는 것에 동의할 수 있게 된다. 왜냐하면 이 통계적 추론의 결과는 우리가 흔히 지식이라고 부르는 것에서 늘 경험하는 것처럼 오류가 전혀 없는 필연적 결과나 결론 또는 정보를 항상 보장해 주지는 못하기 때문이다.

남주현(2007)의 통계 교육적 고찰에 따르면 문제해결방법의 선택과 관련된 우연과 필연, 대

표성과 관련한 무작위추출, 자료 분석에서의 부분과 전체, 귀납과 연역, 추론의 원리와 그 신뢰성에 있어서 가능성 원리와 통계량의 변이성, 통계적 모형 등은 모두 통계적 방법론의 신뢰성 개념에 해당되는 것으로 분류하고 있음을 알 수 있다.

이들 하위개념이 의미하는 바를 더 상세히 살펴보기로 한다.

B. 표본개념의 하위속성

1) 우연과 필연

A) 우연과 필연

“우연과 필연”은 어떤 현상을 분석할 때 이를 통계적 사고로 해결할 것인지 수학적 사고⁵⁾로 해결할 것인지, 문제해결방법의 선택에서 인식론적 분기점에 작용되는 핵심적인 문제이다.

어떤 현상이 필연에 의한 것이라면 논리적 오류가 없는 한 그 결론은 100% 신뢰할만하므로 이 경우 신뢰성을 논할 필요가 없다. 하지만 우연에 의한 것일 경우, 우연에 근거하여 통계적 사고에 의해 얻어진 결론이 신뢰할만한가에 대하여는 의문을 가질 수밖에 없다. 그리하여 “신뢰할만한 자료인가?” 나아가 “분석방법이 타당한가?” 등을 따져보게 되는 것이다. 이처럼 표본개념은 우연과 필연의 구분으로부터 시작된다고 할 수 있다.

현상을 우연과 필연으로 구분하는 것은 주어진 현상을 어떻게 해석할 것인가에 대해 옳은 판단을 내릴 수 있게 하며 이는 현상에 대한 설명 방식에 영향을 준다. 학생들은 사건을 우연과 필연으로 구분함으로써 우연현상의 문제 해결은 통계적 방법으로, 필연현상은 수학적 방법을 사용하는, 상황변화에 따라 적절한 방법을 사용할 수 있다.

5) 본 논문에서의 “수학적 사고”는 수학의 영역 중 대수·기하·해석에서 사용하는 연역적 사고를 말하며 확률·통계에서의 사고는 “통계적 사고”로 구분하여 사용하기로 한다.

B) 무작위성과 대표성

“무작위성과 대표성”은 “모집단을 잘 대표하는 표본을 어떻게 뽑을 것이냐?”의 문제와 관련된다.

자료 분석을 시작하는 단계에서 그 자료가 충분히(관심 모집단 속성에 대한) 대표성을 띄지 않는 자료라면, 오류를 초래하는 경우가 대부분임으로, 자료를 분석하는데 수고를 할 필요가 없을 것이다. 따라서 자료 분석 전에 자료의 신뢰성을 판단하는 것은 필수적이다(이런 이유로 표본개념은 최근 통계교육에서 세계적으로 주목을 받고 있는 탐색적 자료 분석(exploratory data analysis)과도 관련되는 개념이다).

실제 상황에서 모집단의 특성에 대해 신뢰할 만한 판단을 내리고 의사결정을 하기 위해서 모집단을 잘 대표하는 적절한 자료를 수집하는 것은 매우 중요하다(Ben-Zvi et al., 2010). 모집단을 잘 대표하는 표본을 얻기 위해서는 표본 추출 방법에 있어서의 타당성이 필수 사항이다.

대표성을 갖는 표본을 추출하기 위해 가장 자주 사용되는 방법이 무작위추출⁶⁾이다. 무작위추출은 표본을 추출할 때 조사모집단의 모든 추출단위가 표본으로 선택될 확률이 동일하다는 것이고 표본으로 선택되는 추출단위들이 서로 독립적이 되도록 표본을 추출하는 것을 의미한다.

무작위추출에 의한 표본이 기대한대로 높은 확률을 가진 대표성 있는 표본인지의 사실여부는 결과만으로는 알 수 없으므로 무작위 표본에 근거한 추론은 다만 “아마, 그럴 것이다”라고 추측할 뿐이다. 따라서 이런 추론에 대해 신뢰성을 의심하고 꼼꼼히 따져보는 것은 불가피한 필수과정이다.

2) 귀납과 연역

A) 부분과 전체

“부분과 전체”는 무작위추출에 의해 뽑힌 표본에 대해서 “추출된 자료가 모집단의 일부임을 인식하는가?”의 문제이다. 이것은 얻어질 표본의 변이성을 인식한다는 것과 동치이다. 전통적인 통계적 추론은 모수에 관한 추론인 경우가 대부분이다. 이때 표본과 모집단의 속성을 구분하는 것이 중요하다(이은호, 2010). 표본조사의 경우 모든 표본이 모집단과 닮은 것은 아니며(이은호, 2010; 재인용), 표본의 분포는 모집단의 분포 상태를 반영하기를 기대하지만 변이성이 존재하기 때문에 분포의 형태 같은 것들이 항상 비슷할 수는 없다(남주현, 2007). 따라서 표본조사의 경우 표본은 필연적으로 우연변동을 속성으로 갖는다(이석훈 외, 1999).

모집단과 표본의 속성을 안다는 것은 “표본의 정보로부터 모집단의 알고자 하는 정보를 어떻게 얻을 것이냐?”라는 통계적 요약에 관하여, 알고자하는 것을 알려 주는지, 즉 요약의 타당성(요약)개념과 “(불규칙적 변이성을 지닌 표본의 정보인) 그것을 얼마나 믿을 수 있는 것이냐?”의 표본개념과 관련된다.

B) 귀납과 연역

“귀납”과 “연역”은 당면한 문제를 어떤 논리 체계로 분석할 것인가에 대한 생각이다.

수학은 확실한 근거에 의해 판단하는 사고의 논리 체계, 엄밀성과 연역적 사고를 추구하는 반면 통계학은 유용성과 귀납적 사고를 추구하는 경험의 논리 체계이다. 통계적 추론은 대표적인 귀납추론의 예라고 할 수 있다. 학생들은

6) 임의추출법이라고도 하지만 이 경우 “임의”를 “random”으로 해석하지 않고 “조사자 마음대로”로 오해할 우려가 있으므로 본 논문에서는 “무작위”란 용어를 사용한다.

7) 간혹 불규칙성을 무작위성과 동일시하여 결과만으로 무작위 표본 여부를 확인하려는 시도가 있으나 이런 시도는 시도 자체가 무의미하다고 생각된다. 불규칙성은 무작위성의 한 측면이지만 무작위성의 더욱 중요한 속성인 독립성은 결과를 통해 확인될 수 없기 때문이다.

“연역법”과 “귀납법”의 서로 다른 두 가지 해결 방법 중에서, 상황에 따라 적절한 방법을 선택할 수 있어야 한다.

연역법은 전체로 부분을 추론하는 방법으로 의심할 여지가 없는 사실이다. 연역의 추리에서는 우리가 생각하는 의미의 신뢰성은 생각하지 않아도 된다. 하지만 귀납법은 “부분”에서 “전체”로 추론하는 방법으로 “반드시 옳다”고는 말할 수 없는 추론법이다. 따라서 귀납법에 의한 결론은 어떻게 믿을 만한가? 라는 신뢰성의 문제로 이어진다.

귀납과 연역을 학습함으로써 수학적 사고와 통계학적 사고법의 차이를 알 수 있고 이를 구분할 수 있게 된다.

3) 가능성의 원리

학생들은 살아오면서 비형식적인 학습을 통하여 나름대로 사고하는 방법을 익히고 의사결정을 한다. 이러한 의사결정의 신뢰성을 높이기 위해선 사고과정에 대한 논리적인 근거를 제시할 수 있어야 한다. 우리가 상식이라고 말하는 이러한 비형식적 사고흐름인 통계적 추론을 뒷받침할 논리법칙이 필요하다. 이것을 우리는 가능성의 원리(likelihood principle)⁸⁾라고 부르기로 한다.

가능성의 원리는 과거형 언어로 사용할 때 “무작위표본에 의해 관측된 결과는 그와 같은 관측결과를 얻을만한 충분한 확률적 이유가 있다”라고 생각하는 것이다. 이것을 통계적 가설검정으로 이어가면 “회귀사건에 관한 통계적 개념”으로 이어지는데, “회귀한 일이 일어났다면 거기에는 반드시 무슨 이유가 있을 것이라는 생각”을 뜻하며 이 생각은 나아가 과학적

발견의 토대이자 동시에 통계적 가설검정의 핵심 논리로 작용한다.

가능성 원리에 의한 미래형 또는 과거형 추론에 대해 이것을 확률과 통계를 구분하여 적용한 연구자의 생각을 정리하면 아래와 같다. 가능성의 원리의 미래형추론에의 적용은 확률과 관련된다. 확률이 높은 쪽이 일어날 가능성이 더 크므로 확률이 높은 쪽의 사건이 일어날 것이라고 추측하는 것이 확률에서의 가능성의 원리이다. 가능성 원리의 과거형 추론에의 적용은 통계(신뢰도, 유의수준)과 관련된다. 과거형 추론은 미래형 추론처럼 그다지 상식적이지 않다. 그래서 연구자는 가능성원리에 의한 추론을 상식에 의한 추론이라기보다는 논리법칙에 의한 추론으로 이해하려는 것이다.

앞서 “회귀사건에 관한 통계적 개념”은 간접증명법(귀류법)의 추론형태와 관련된다. 동전 던지기의 예를 생각해 보자. 어떤 동전을 5회 던진 결과 모두 앞면만 나왔다고 하자. 만약 이 동전을 한 번 던질 때 앞면이 나올 확률이 $\frac{1}{2}$ 이라고 한다면(귀류법의 가정) 이 결과는 $(\frac{1}{2})^5$ 의 가능성, 즉 약 3%의 가능성이 있는 일이다. 즉, 우연에 의해 이런 일이 일어날 확률은 3%에 불과하다(회귀 사건이다). 이는 가능성의 원리에 맞지 않으므로(충분한 확률을 갖지 못했으므로) $p(\text{앞}) = \frac{1}{2}$ 은, 5회 연속 앞면만 나왔다는 사실(fact, data) 자체를 부정할 수 없는 한, 기각된다. 이는 논리형식상 “귀류법”의 추론형태이다.

이와 같이 간접증명법의 논리를 기반으로 하는 가설검정에서 가능성의 원리에 기초하여 확률적으로 회귀한 사건, 자료 등의 결과를 판단하게 된 것은, 결국 가능성 원리를 통계적 추론의

8) likelihood principle은 통계학에서 사용되는 하나의 중요 원리로서 점 추정론(point estimation theory)에서는 최우추정량(maximum likelihood estimator), 검정론에서는 우도비검정법(likelihood ratio test) 등의 추론적 근거가 되는 원리이다. 연구자들은 이 원리에다가 가능성에 대한 상식적 판단의 의미를 추가하여 통계적 추론에서의 합리성의 논리법칙으로 사용하기로 한다.

<표 II-1> 귀류법과 통계적 가설검정의 증명방법의 비교

	귀류법(간접증명법)	통계적 가설검정
가정 (결론부정)	$\sqrt{2}$ 는 유리수이다	$p(\text{앞}) = \frac{1}{2}$ 이다.
증명과정	$\sqrt{2} = \frac{b}{a}$ (단, a, b 는 서로소인 자연수) $\sqrt{2} a = b$, $2a^2 = b^2$ $2a^2$ 은 짝수이므로 b^2 도 짝수이고 $b = 2k$ 로 나타낼 수 있다. $2a^2 = 4k^2$, $a^2 = 2k^2$ 마찬가지로 $2k^2$ 은 짝수이므로 a 도 짝수이고, $a = 2m$ 으로 나타낼 수 있다.	동전을 5회 던진 결과 모두 앞면이 나오는 상황은 $(\frac{1}{2})^5$ 의 가능성이 있는 일이다(약 3%). 확률이 너무 작으므로 그런 일이 나올 가능성은 거의 없다(가능성의 원리).
모순발견 (오류)	$a = 2m$, $b = 2k$ 이므로 a, b 는 서로소가 아니다. 이는 a, b 가 "서로소인 자연수"라는 것에 모순이다.	"5회 던지기의 결과 모두 앞면만 나왔다"는 관찰 사실은 위에서 "나올 가능성이 거의 없다."라는 사실에 모순이다.
결론	따라서 $\sqrt{2}$ 는 무리수이다.	따라서 $p(\text{앞}) \neq \frac{1}{2}$ 이었을 것이다.

핵심적 논리법칙으로 볼 수 있다는 것이 연구자의 생각이다. 나아가 이것은 통계적 추론에 대한 신뢰성 비판에서 어느 정도를 신뢰해도 되는지(신뢰도 등)에 대한 중요한 근거로서의 역할도 하는 것이므로 통계적 추론의 신뢰성 확보를 위한 핵심적 근거라고 할 수 있다.

4) 통계량의 변이성

통계적 추론은 연역이 아닌 귀납추론이므로 논리적 타당성을 확보하려면(신뢰성이 있으려면) 특별한 검토가 필요하다.

귀납은 "반드시 옳다"고는 할 수 없는 추론이므로 일반화의 오류 가능성이 항상 존재한다. 귀류법에서 "모순이다"라는 명확한 진술은 통계적 추론에서는 "거의 모순이다"와 같은 확률적 진술로 변하는데, 이 확률적 진술의 강도가 곧 추론의 신뢰성을 결정한다.

가령 사건 A가 일어날 확률이 p 라 할 때,

사건 A의 발생여부를 예측해보자. p 의 값이 매우 크다면 우리는 "사건 A가 아마 발생할 것이다"라고 쉽게 예측할 수 있다. 반대로 p 의 값이 매우 작을 때도 "사건 A가 아마 발생하지 않을 것이다"라고 쉽게 예측할 수 있다. 그런데 p 가 $\frac{1}{2}$ 근방의 값일 경우를 생각해보자. 이때는 A의 발생여부를 쉽게 예측하기 어렵다. 사건 A의 발생여부에 대한 변이성을 분산으로 수량화하면 그 값은 $p(1-p)$ 에 비례⁹⁾한다. 그리고 $p(1-p)$ 는 $p = \frac{1}{2}$ 일 때 최대이고, 즉 변이성이 가장 크고, p 가 "0"이나 "1"에 가까워지면 사건 A의 발생여부에 대한 변이성은 작아짐을 알 수 있다. 결국 변이성이 크면 예측이 어렵고, 반대로 변이성이 작아지면 예측이 쉬워진다고 정리해 볼 수 있다.

이렇게 통계적 예측 진술의 강도를 결정하는 것이 통계량의 변이성이고, 가능성의 원리에 논

9) A가 일어나는 사건에 대해 실수 a 를, 일어나지 않은 사건에 대해 실수 b 를 대응시킬 때 이 확률변수 X_A 의 분산 $V(X_A) = E[(X_A - E(X_A))^2]$ 은 $(a-b)^2 p(1-p)$ 로 $p(1-p)$ 에 비례하는 값을 알 수 있다. 단, $E(X_A) = ap + b(1-p)$.

리적으로 근거하여 이를 검토할 수 있다.

모집단에서 하나의 무작위 표본을 추출하여 그로부터 계산한 통계량¹⁰⁾은 하나의 숫자이다. 표본 한 모둠(data set)이 정해질 때에는, 그 숫자는 정해지는 것이지만, 무작위 표본 모둠을 반복 추출¹¹⁾할 때마다 다른 표본 모둠이 나올 것이고 그 때마다 다른 숫자가 계산결과(통계량)으로 얻어질 것임을 안다. 이 때, 상상속의 많은 표본 모둠마다 구해진 통계량(숫자)들로 도수분포표를 만들면 이 표를 그 통계량의 도수분포라고 할 수 있고, 이것으로 상대도수분포표를 만들면, 반복 추출의 횟수가 매우 많은 경우, 이것은 근사적으로 생각할 때, 그 통계량의 (통계적) 확률분포이며, 가상으로 표본 모둠을 무한개 반복 추출했다고 생각할 때의 상대도수분포를 그 통계량의 확률분포, 즉 그 통계량의 표집분포라고 한다. 통계량을 확률변수로 이해한다는 것은 통계량의 외형 - 분포라는 것을 가질 수 없는 한 개의 숫자-가 아니라 그 본질을 파악하여 그 통계량의 표집분포의 존재를 마음속에 상상할 수 있다는 것으로서, 논리전개의 핵심을 정확히 이해하기 위해 필수적인 것이다. 왜냐하면 어떤 사람의 키가 희귀한 정도로 큰 지는 그 사람의 키를 나타내는 숫자를 많은 사람들의 키 값의 분포와 비교할 때, 그래서 그 사람의 키 값이 분포의 한쪽 귀퉁이에 놓일만한 숫자인지 파악할 수 있을 때만 가능하기 때문이다(대개 사람들은 이런 과정을 무의식적으로 하기 때문에 이런 생각을 하고 있다고 느끼지는 않지만, 만약 어느 벽들의 강도가 300psi인데 이 벽들이 다른 벽들에 비해

강한지 약한지 알려면 무엇을 알 필요가 있는지 생각해 보면 곧 이해할 수 있다).

마찬가지로 실제 통계 조사의 한 표본 모둠에서 얻은 통계량인 숫자 한 개가 희귀한 것인지를 알려면, 그 통계량의 모든 가능한 값들의 분포(표집분포)를 알아야 판단이 가능하고, 이 판단이 가능성 원리의 논리로 연결되어야 통계적 추론 전체의 맥락을 이해하고 그 추론을 신뢰할 수 있게 된다. 또, 그 모두를 이해함에 있어서, 최소한의 필수 사항은 한 숫자로 나타난 통계량이 본질적으로는 변이성을 갖고 있는 것이었다는 것을 이해하는 것이며, 그 통계량의 표집분포를 수학적으로 어떻게 구하느냐의 대학수준 어려운 수학 문제와는 별개로 취급되어야 할, 매우 중요한 표본개념 발달의 핵심적 맥락인 것이다.

대개의 통계량의 변이성은 그것의 계산에 사용된 표본의 크기가 커질수록 작아진다.¹²⁾ 어떤 현상에 대한 관측치가 많으면 많을수록 그 관측치에 근거한 현상에 대한 판단이 더 신뢰할 수 있다는 것은 누구에게나 직관적으로 이해 가능하지만, 알고 보면 통계량의 변이성이 이런 속성을 지닌 때문이라고 할 수 있다. 통계적 추론에서 핵심적 역할을 하는 통계량의 변이성(분산 등)이 클수록 그에 근거한 추론은 신뢰성이 약해지고, 반대로 변이성이 작아지면 신뢰성은 높아진다. 이것을 충실히 이해할 수 있을 때 가장 발전된 표본개념의 수준에 도달한 것이라고 추측된다.

5) 통계적 모형

우연현상을 분석하는 사람은 필연적 현상을

10) 표본으로부터 계산되는 표본평균, 표본분산, 표본비율 등과 같이 표본의 특성을 나타내는 계산된 값

11) 실제 통계 조사에서 표본 모둠들을 반복 추출하지는 않는다. 항상 한 모둠의 표본을 얻을 뿐이다. 여기서 실제 얻은 표본 모둠과 동일한 크기의 여러 표본 모둠의 반복 추출을 생각하는 것은 이론적 분석을 위해서 생각하는 것뿐이다. 이 이론의 거울에 실제 얻은 고정된 한 숫자인 통계량을 비취보아야 희귀한 것인지 아닌지 알 수 있기 때문이다.

12) 수렴 통계량(consistent estimator)은 표본의 크기가 한없이 커지면 모집단의 관심통계량에 약하게 수렴(converges weakly in probability)한다.

분석하는 사람과 마찬가지로 어떤 법칙을 얻어 현상을 설명하고 그로부터 어떤 예측을 얻으려 한다. 그런 예측의 주된 장애가 바로 통계량의 변이성이다. 변이성은 예측 불가능하고 해결하기 어려운 원인과 결과를 가져오기 때문에 이를 줄이는 것이 통계적 모형의 신뢰도를 좌우한다(조동기, 2007).

통계적 모형은 언뜻 보기에 우연현상처럼 보이는 어떤 현상을 설명변인인 필연부분과 설명되지 못한 우연부분의 합으로 분리한다.(이것을 가법적 모형(additive model)이라고 한다) 통계적 방법은 이렇게 설정된 모형에서 관찰 현상이 나타내는 변이성 중에 우연 변인의 영향(변이성)을 최소화하여 통계적 추론의 신뢰도를 높이기 위해 노력한다. 따라서 이런 모형의 관점에 의하면 통계적 추론의 오류는 우연변동의 왜방에 의한 것이며 따라서 이것을 잘 통제하는 것이 통계적 방법의 핵심이다. 이러한 과정의 정확성은 채택된 통계적 처리의 신뢰성과 직결된 문제가 된다. 왜냐하면 얻고자 하는 모집단의 정보와 무관한 정보를 깔끔하게 제거할수록 자료 축약¹³⁾의 변이성이 줄어들기 때문이다. 이 문제를 통계학에서는 축약의 충족성(sufficiency) 문제라고 한다(Zacks, 1971).

III. 연구방법 및 결과

1. 분석대상 교과서

초등학교 수학1-1, 수학2-2, 수학3-2, 수학4-2, 수학6-1과 수학기초책(제7차 개정검정) 수학5-2, 수학6-2와 수학기초책(제7차 검정; 개정교과서는 출판되지 않음)

중학교 수학1과 수학1 익힘책 (제7차 개정검정 27종), 수학2와 수학2 익힘책 (제7차 개정검정 17종), 수학3과 수학3 익힘책 (제7차 개정검정 14종)

고등학교 수학의 활용과 수학의 활용 익힘책 (제7차 개정검정 2종) 미적분과 통계기본과 미적분과 통계기본 익힘책 (제7차 개정검정 13종) 적분과 통계와 적분과 통계 익힘책 (제7차 개정검정 10종)

2. 분석틀

앞서 이론적 배경에서 살펴 본 바와 같이 남주현(2007) 연구의 표본개념에 바탕을 두고 표본개념의 원리 및 논리가 되는 하위 속성들을 다음과 같이 분석틀로 정하여 이들의 서술 정도를 관찰 분석하였다.

<표 III-1> 교과서 분석틀

분석내용	세부내용
우연과 필연	우연과 필연
	무작위성과 대표성
귀납과 연역	부분과 전체
	귀납과 연역
가능성 원리	
통계량의 변이성	
통계적 모형	

3. 분석내용 및 결과

A. 우연과 필연

1) 우연과 필연

A) 분석준거

여기에서는 현실에서 필연현상과는 구분되는

13) 한 개의 표본 모듈에는 많은 값들이 있는데, 이것을 목적에 따라 한 두 개의 숫자(통계량/설명변인의 추측값)로 줄이는 것 자체가 자료의 축약이다.

우연현상이 존재함을 얘기하고 있는지, 또 학생들이 우연현상과 필연현상을 구분할 수 있도록 어떤 활동을 제시하고 있는지를 살펴보고자 한다. 주어진 현상에 대한 문제해결방법을 택하는 데 있어서 우연현상은 통계적 사고로 필연현상은 수학적 사고로 해결할 수 있음을 말하고 있는지 살펴보고자 한다. 나아가 필연현상에 대한 분석은 그 결과에 대한 신뢰성을 논할 필요가 없지만 우연현상에서는 그 결과에 대한 신뢰성을 따져봐야 한다는 것을 서술하는지를 살펴본다.

<표 III-9> 우연과 필연의 분석준거

세부내용	준거
우연과 필연의 구분	<ul style="list-style-type: none"> 우연현상의 존재를 서술하는가 우연현상과 필연현상을 구분하여 서술하는가
문제해결방법의 선택	<ul style="list-style-type: none"> 우연현상에는 통계적 사고를 사용해야 함을 서술하는가 필연현상에는 수학적 사고를 사용해야 함을 서술하는가

B) 분석결과

수학1, 수학2, 수학3의 확률·통계 단원의 도입에서 일상생활에 존재하는 우연현상에 대해 소수의 교과서가 언급하고 있으나, 과반수이상의 교과서가 이를 다루고 있지 않으며 우연현

상을 필연현상과 구분하여 설명한 교과서는 없었다.

우연현상의 존재를 말하는 교과서는 그 현상을 확률 혹은 통계의 방법으로 다룰 수 있다고 말한다. 하지만 필연현상은 수학적 사고로 다룬다는 것을 서술하지는 않고 있다.

현행교육과정에서 우연과 필연의 내용을 점진적으로 다루고 있다고 보기는 어렵다.

2) 무작위성과 대표성

A) 분석준거

“무작위성과 대표성”에서는 대표성이 무엇인지 무작위성을 통해 대표성을 얻을 수 있음을 말하고 있는지를 살펴보고자 한다. 모집단의 특성에 대해 타당한 결론을 내리기 위해서는 추출되는 표본이 모집단을 충분히 대표하고 있는지를 확인해야 함을 학생들에게 이해시키

<표 III-12> 무작위성과 대표성의 분석준거

세부내용	준거
타당한 자료의 필요성	<ul style="list-style-type: none"> 타당한 자료의 중요성을 말하는가
자료의 대표성과 무작위추출	<ul style="list-style-type: none"> 무작위추출의 필요성을 자료의 대표성과 관련하여 말하는가
무작위로 추출한 표본의 의미	<ul style="list-style-type: none"> 무작위로 추출한 자료가 어떤 의미를 가지는지 말하는가

<표 IV-1> 우연과 필연의 분석결과

준거	단계(종)								수학1 (27)	수학2 (17)	수학3 (14)	미적분과 통계기본 (13)	적분과 통계 (10)	수학의 활용(2)
	1-1	2-2	3-2	4-2	5-2	6-1	6-2							
우연현상의 존재	2종	5종	2종	5종	8종	2종
필연현상의 존재	1종
우연현상은 통계적 방법	3종	5종	2종	5종	8종	2종
필연현상은 수학적 방법	1종

<표 IV-2> 무작위성과 대표성의 분석결과

단계(종)								수학1	수학2	수학3	미적분과	적분과	수학의 활용(2)
	1-1	2-2	3-2	4-2	5-2	6-1	6-2	(27)	(17)	(14)	통계기본 (13)	통계 (10)	
준거													
타당한 자료의 필요성	7종	4종	2종
자료의 대표성과 무작위추출	13종	10종	2종
무작위로 추출한 표본의 의미

고 있는지를 살펴보고 그 방법이 무작위추출임을 말하는지를 살펴보려는 것이다.

B) 분석결과

고등학교의 선택과목인 미적분과 통계기본, 적분과 통계, 수학의 활용의 몇몇 교과서만이 자료가 대표성을 갖지 않을 때 그 분석 결과에 오류를 보일 수 있음을 보여주며 타당한 자료의 중요성을 말하고 있고 나머지의 교과서는 타당한 표본을 추출하는 것이 왜 중요한지에 대해 서술하고 있지 않다.

표본자료는 초등학교 4학년부터 고등학교 3학년까지의 교과서에서 이용되고 있었다. 하지만 초·중학교에 제시된 표본자료는 자료수집의 목적이나 그것이 표본추출에 의한 자료라는 것을 명확히 하지 않고 있으며, 그 자료가 무작위로 추출된 자료일 때도 무작위추출의 의미나 무작위추출로 추출된 표본의 의미는 전혀 서술하지 않고 있다.

고등학교의 통계적 추정단원에서는 모든 교과서가 자료의 대표성과 관련하여 무작위 추출을 도입하고 그 원리를 설명하고 있으며 몇몇 교과서에서 무작위추출이 아닌 방법으로 추출된 자료를 제시하며 그 대표성을 판단하는 활동을 제시하고 있다.

자료수집의 활동은 주로 중학교 이전까지의 과정에서 다루어지고 고등학교에서는 무작위추

<표 III-14> 부분과 전체의 분석준거

세부내용	준거
부분과 전체의 속성	• 부분과 전체의 속성을 구분하여 말하는가
표본의 변이성	• 부분은 모집단을 반영하지만 부분과 모집단은 다를 수밖에 없음을 말하는가

출의 활동이 다뤄지고 있다. 하지만 그 활동은 자료의 분석에만 초점이 맞춰져 있으며 수집방법보다는 수집된 자료를 분석하는 데 치중하고 있다.

또 고등학교에서의 무작위추출의 활동에서는 추출의 목적이나 추출된 자료의 의미에 관심을 가지기보다는 활동이 이뤄지는 것에만 의미를 두고 있으며 무작위로 추출된 표본이 가지는 의미를 서술하고 있지는 않다.

B. 귀납과 연역

1) 부분과 전체

A) 분석준거

부분과 전체에서는 다음을 준거로 삼아 현행 교과서에서 부분과 전체의 속성을 명확히 구분하고 있는지를 살펴보고자 한다. 단순히 정의에 의한 구분만이 아니라 모집단의 속성과 표본의 속성을 구분하고 있는지를 살펴보려는 것이다.

<표 IV-3> 부분과 전체의 분석결과

단계(종) 준거	수학	수학	수학	수학	수학	수학	수학	수학	수학1	수학2	수학3	미적분과통	적분과	수학의
	1-1	2-2	3-2	4-2	5-2	6-1	6-2	(27)	(17)	(14)	계기본(13)	통계	활용(2)	
부분과 전체의 속성	12종	9종	2종
표본의 변이성	5종	6종	1종

B) 분석결과

미적분과 통계기본, 적분과 통계, 수학의 활용의 대부분의 교과서에서 모집단으로부터 표본을 추출할 때마다 다른 표본이 추출된다는 표본의 변이성을 말하고 있었다. 하지만 이러한 표본의 속성을 모집단의 속성과 명확히 구분하여 설명한 교과서는 없었다.

또 통계적 추정에서 몇몇 교과서에서만 일부는 모집단을 반영하긴 하지만 항상 모집단을 완벽히 대표하지는 않는다는 것을 말하고 있다.

부분과 전체에 대한 내용은 초·중학교의 표본자료에서는 전혀 언급하지 않으며 고등학교 선택과목에 이르러야 이를 서술하고 있어 연계성 있게 교육이 이뤄지고 있다고 보기 어렵다.

2) 귀납과 연역

A) 분석준거

“귀납과 연역”에서는 귀납추론과 연역논증의 의미를 설명하고 있는지를 살펴본다. 통계적 추론은 귀납적 논리임을 학생들이 이해할 수 있

<표 IV-4> 귀납과 연역의 분석결과

단계(종) 준거	수학	수학	수학	수학	수학	수학	수학	수학	수학1	수학2	수학3	미적분과통	적분과	수학의
	1-1	2-2	3-2	4-2	5-2	6-1	6-2	(27)	(17)	(14)	통계기본	통계	활용(2)	
귀납과 연역을 의의	1종
사고체계의 결정	1종

<표 III-17> 귀납과 연역의 분석준거

세부내용	준거
귀납과 연역 을 의의	• 귀납의 의의를 설명하는가
	• 연역의 의의를 설명하는가
사고체계의 결정	• 통계는 귀납적 사고를 사용함을 말하는가
	• 수학은 연역적 사고를 사용함을 말하는가

도록 서술되어 있는지 확인한다. 나아가 주어진 상황에 잘못된 사고체계를 사용했을 때 오류를 보일 수 있음을 서술하는지를 살펴본다.

B) 분석결과

몇몇 교과서에서 귀납을 부분적으로 설명하고 있을 뿐 귀납과 연역의 용어를 사용하며 그 의미를 명확히 설명한 교과서는 전체의 수학교과서 중 1종에 불과했다.

따라서 결정론적 수학적 사고에 익숙한 학생들이 통계의 논리체계인 귀납적 사고를 익힐 기회는 거의 없다고 볼 수 있다.

<표 III-18> 가능성의 원리의 분석준거

세부내용	준거
미래형 추론의 논리적 근거	• 미래형 추론의 논리적 근거를 말하는가
과거형 추론의 논리적 근거	• 과거형 추론의 논리적 근거를 말하는가

C. 가능성의 원리

A) 분석준거

“가능성의 원리”에서는 통계적 추론의 핵심인 가능성의 원리를 말하고 있는지 여부를 확인한다. 또 가능성 원리의 미래형 추론과 과거형 추론에의 적용을 이해할 수 있도록 기술하고 있는지 살펴보고자 한다.

B) 분석결과

수학1, 수학2, 수학3에서는 확률이 어떤 사건이 일어날 가능성이라는 점에 주목하여 그것으로 미래를 예측할 수 있다는 미래형 추론을, 미적분과 통계기본, 적분과 통계, 수학의 활용에서는 추출된 표본을 이용하여 모수를 추론하는 과거형 추론을 다루고 있음을 알 수 있었다. 하지만 그 논리적 근거인 가능성의 원리를 서술한 교과서는 없었다.

D. 통계량의 변이성

A) 분석준거

표본으로부터 계산된 값인 통계량이 변이성을 가지므로 그것의 분포를 생각할 수 있음을

서술하고 있는지를 살펴본다. 또 변이성이 어떤 의미를 갖는지, 나아가 표본의 크기가 변함에 따라 통계량의 변이성, 신뢰도가 어떻게 달라지는지에 대해 서술하고 있는지의 여부를 살펴본다.

B) 분석결과

<표 III-19> 통계량의 변이성의 분석준거

세부내용	준거
통계량의 변이성의 인식	• 표본으로부터 계산된 값인 통계량은 변이성을 가짐을 말하는가
통계량의 변이성의 의미	• 변이성을 계산하는 방법이 표집오차이며 변이성은 예측의 강도(신뢰도)를 결정함을 말하는가
표본의 크기, 통계량의 변이성, 신뢰도와와의 관계	• 표본의 크기, 통계량의 변이성, 신뢰도와와의 관계

고등학교 선택과목의 대부분의 교과서에서 표본을 무작위로 추출하고 통계량을 계산하는 활동을 통해 통계량의 변이성을 인식하도록 하고 있다. 하지만 추론에 있어 변이성이 어떤 의미를 갖는지를 서술하는 교과서는 없었다. 또 표본의 크기, 통계량의 변이성과 신뢰도와와의 관계를 명확히 설명하지 않고, 그 관계를 간접적으로만 인식하도록 하였다. 교과서에 소개된 신뢰구간, 신뢰도, 표본의 크기와 관련된 대부분의 문제는 각각이 지니는 의미와 서로의 관계를 묻기보다는 기초적인 수준에 머물고 있

<표 IV-5> 가능성의 원리의 분석결과

준거	단계(종)												
	수학 1-1	수학 2-2	수학 3-2	수학 4-2	수학 5-2	수학 6-1	수학 6-2	수학1 (27)	수학2 (17)	수학3 (14)	미적분과 통계기본 (13)	적분과 통계 (10)	수학의 활용(2)
미래형 추론의 논리적 근거
과거형 추론의 논리적 근거

<표 IV-6> 통계량의 변이성의 분석결과

단계(종) 준거	수학	수학	수학	수학	수학	수학	수학	수학1	수학2	수학3	미적분과	적분과	수학의
	1-1	2-2	3-2	4-2	5-2	6-1	6-2	(27)	(17)	(14)	통계기본 (13)	통계 (10)	활용(2)
통계량의 변이성 인식	10종	8종	1종
통계량의 변이성의 의미
표본의 크기, 통계 량의 변이성, 신뢰 도와의 관계	6종	3종	1종

<표 III-21> 통계적 모형의 분석준거

세부내용	준거
우연현상 속의 필연변인의 인식	<ul style="list-style-type: none"> 우연변인과 필연변인을 동시에 가지는 현상이 있음을 말하는가
통계적 모형과 신뢰성	<ul style="list-style-type: none"> 한 현상 안에서 필연변인을 찾음으로써 변이성이 줄어들어 추론의 신뢰성을 높일 수 있음을 말하는가

으며 통계량의 변이성은 계산 인자로만 사용되고 있었다.

E. 통계적 모형

A) 분석준거

“통계적 모형”은 곧 추론의 신뢰성을 높이는 방법이다. 먼저 일상생활에서의 우연속성과 필연속성을 동시에 지니는 현상에 대해 언급하고 있는지를 확인한다. 그 현상에서 필연적인 요소

를 찾아서 분리해 갈 때 변이성을 줄일 수 있으며 신뢰도를 높일 수 있다는 것을 서술하고 있는지, 나아가 모집단의 정보와는 무관한 정보를 제거하는 활동을 제시하고 있는지를 살펴본다.

B) 분석결과

수학2, 미적분과 통계기본과 적분, 통계의 소수의 교과서만이 통계적 모형에 대해 서술하고 있었으며, 통계적 모형의 필요성을 변이성과 관련하여 서술한 교과서는 거의 없었다. 또 통계적 모형과 관련된 활동을 제시한 교과서는 전혀 없어 통계적 모형에 대한 학습은 거의 이뤄지지 않는다고 볼 수 있다.

V. 결론 및 제언

첫 번째, 문제해결방법의 선택과 관련된 내

<표 IV-7> 통계적 모형의 분석결과

단계(종) 준거	수학	수학	수학	수학	수학	수학	수학	수학1	수학2	수학3	미적분과	적분과	수학의
	1-1	2-2	3-2	4-2	5-2	6-1	6-2	(27)	(17)	(14)	통계기본 (13)	통계 (10)	활용(2)
우연현상 속의 필연변인의 인식	1종	.	1종	2종	.
통계적 모형과 신뢰성	1종	.

용이다.

주어진 문제에 통계적 방법을 쓸 것인가 말 것인가의 결정은 문제 속의 현상이 우연에 의한 것인지 필연에 의한 것인지를 구분함으로써 결정할 수 있다. 우연과 필연은 중학교 수준에서는 소수의 교과서만이 다루고 있으며 고등학교에 이르러서야 일부 교과서가 우연현상을 설명하고 있다. 이를 필연현상과 비교하여 설명한 교과서는 거의 없었다. 현상에 대한 논의들 통해 우연과 필연을 구분하는 안목을 기를 수 있도록 주어진 현상들을 우연과 필연으로 구분하는 활동을 제시할 것을 제안한다.

필연현상은 수학적으로 해결하기 때문에 신뢰성을 논할 필요가 없지만 우연현상에 대한 분석은 통계적 사고에 의한 것으로 결과의 신뢰성에 의문을 제기해야 함을 언급할 필요가 있다.

두 번째, 표본자료의 의미와 관련된 내용이다.

표본자료는 이미 초등학교 4학년 때부터 사용되고 있으나 초·중학교의 문제에서는 표본자료로부터 계산된 수식 혹은 그래프를 통해 얻은 결과에만 초점을 맞추고 있었으며, 그 자료가 표본으로써 지니는 의미에 대해서는 어떠한 언급도 하고 있지 않고 있다. 비록 고등학교 교육과정 이전에는 표본개념을 다루고 있지 않지만, 연계성의 확보를 위한 대안으로 초·중학교에서는 문제를 이용해서라도 표본개념을 포함시키려는 노력이 필요하다. 자료를 제시할 때, 그것이 표본자료라는 것은 명확히 해야 할 것이며, 자료의 값에만 초점을 맞출 것이 아니라 그 자료가 지니는 의미를 알도록 해야 한다.

세 번째, 자료수집의 활동과 관련된 내용이다.

중학교 1학년의 통계영역에서 주로 이뤄지는 통계조사 활동은 주로 수집된 자료의 분석에만 초점이 맞춰져 있었으며, 제시하고 있는 몇몇 표본추출방법은 무작위추출과는 거리가 있는

방법이었다. 통계조사의 활동을 제시할 때, 활동의 목적에 맞는 대상을 선정하고, 표본을 어떻게 선택해야 하는가에 대해 학생들이 스스로 고민하도록 하고 선택한 방법에 대한 반성이 이뤄지도록 하는 등 그 자료를 수집하는 과정에 더 의미를 둘 필요가 있다.

통계적 추정을 다루는 고등학교의 몇몇 교과서에서는 무작위 추출의 활동을 제시하고 있지만 그 활동은 무작위추출의 방법만을 연습할 뿐, 추출의 목적이 명확하지 않은 활동이었다. 보다 의미 있는 추출활동이 이뤄질 수 있도록 표본추출의 목적이 명확한 자료를 제시할 필요가 있다. 또 무작위 추출이 이뤄진 후에는 표본자료가 지니는 의미를 논의하도록 하여 무작위로 추출된 표본은 대표성을 얻기 위한 최선의 선택임을 인식하도록 할 필요가 있다.

네 번째, 통계적 추론의 정당화에서의 논리와 관련된 내용이다.

연구결과로부터 현행교육과정에서는 논리법칙과 논리적 근거에 대한 학습이 거의 이뤄지지 않음을 알 수 있었다. 통계학에서의 사고체계인 귀납법이 수학적 사고인 연역과는 명확히 구분된다는 것을 서술할 필요가 있다. 또 통계자료에 근거를 두고 판단하는데 있어 추론의 근거가 가능성의 원리임을 서술할 것을 제안한다.

다섯 번째, 표본으로부터 요약된 통계량에 관련된 내용이다.

통계량의 변이성은 표집분포로 연계되는 핵심개념임에도 현행 교육과정에서는 변이성의 존재만 확인하고 변이성이 갖는 의미를 설명하지 않은 채, 이를 계산을 위한 인자로만 사용하고 있다. 통계량의 변이성이 추론의 신뢰도를 결정하는 핵심적인 역할을 한다는 것을 이해하도록 해야 한다. 예측의 강도는 통계량의 변이성에 의해 결정되며, 표본의 크기가 클수록 통계량의 변이성이 작아지며 이에 따라 신뢰

성이 높아진다는 것을 서술할 것을 제안한다.

현재 교육과정에서는 표본개념을 이해할 기회가 충분히 제공되지 않은 상태에서 고등학교에 이르러 통계적 추정을 도입하고 있다. 이런 상황에서 학생들이 통계적 추정을 어려워하고 그 내용을 논리적으로 설명하지 못하는 것은 이상한 일이 아닐 것이다. 자료의 의미와 그 결과의 해석, 통계의 논리와 통계적으로 사고하는 데에 초점을 맞추는 통계교육이 이루어질 것을 제안하며, 학생들이 이러한 표본개념에 바탕을 두고 진정한 통계적 사고를 할 것을 기대해 본다.

본 연구에서 남주현(2007)이 제안한 여러 하위 개념의 정당성 여부에 대해서는 확인하지 않고, 그 개념들이 표본개념 형성과 발달에 핵심적 구성요소라고 보았다. 따라서 본 연구가 현재 우리나라 통계교육의 실상을 정확히 진단할 수 있는지는 분명치 않다고 생각할 수도 있다. 그러나 연구자의 이해 및 남주현의 제안이 전혀 상식과 동떨어진 것이 아니라고 할 수 있다면 본 연구의 결과로서 드러난 표본개념의 교육적 실상은 염려할 만큼 심각하다고 할 수 있다.

본 연구에서는 각 개념이 제시되는 순서의 적절성 여부를 판단할 수는 없고 다만 관련된 주요 하위개념들을 교과서에서 설명하고 있는지와, 그 내용이 연계성을 가지고 제시되고 있는지의 여부만 확인해보았다. 따라서 학생들의 발달단계에 맞춰 표본개념에 대한 연계성을 갖춘 학습이 이뤄지도록 각각의 하위속성을 교육적으로 어떻게 구현하는 것이 좋을 것인가에 대한 연구가 이뤄져야 할 것이다. 표본개념 측면에서 통계적 개념이나 원리 또는 논리들이 아동의 인지발달 시기에 더불어 어떻게 발달해 나가며 어떻게 구조화되는가에 대한 연구를 통해 표본개념의 각 개념이나 원리, 논리들의 발달단계가 입증되고 정립된다면 표본개념에 대한 학습이 체계적으로 이뤄질 수 있을 것이다.

참고문헌

- 교육과학기술부(2010). **적분과 통계**. 지학사.
- 남주현(2007). **초·중등 통계교육을 위한 통계적 방법론에 대한 연구**. 이화여자대학교 대학원 박사학위 청구논문.
- 류선욱(2009). **수 I 통계에 대한 학생들의 오류 및 오개념 분석**. 한양대학교 교육대학원 석사학위논문.
- 송성수(2010). **통계적 개념에 대한 학생들의 오류 유형 및 교과서 분석**. 고려대학교 대학원 석사학위논문.
- 우정호(2000). 통계교육의 개선방향 탐색, **학교수학**, 2(1), pp.1-27. 대한수학교육학회.
- 이경화·지은정(2005). 표본 개념의 교육적 의미와 인식 특성 연구. **수학교육학연구**, 15(2), pp.177-196. 대한수학교육학회
- 이석훈·김응환(1999). **통계와 확률지도론**. 경문사(박문규).
- 이은섭·이승수(2004). 고등학교의 확률·통계 영역에 대한 학습실태와 개선방안, **기초과학연구**, 15, pp.167-176. 강원대학교 기초과학연구소.
- 이은호(2010). **표집분포 개념 이해를 위한 통계교육 소프트웨어 설계 및 구현**. 이화여자대학교 교육대학원 석사학위논문.
- 조동기(2007). **표본평균분포 이해에 관한 연구**. 한국교원대학교 대학원 석사학위논문.
- D. Ben-Zvi and J. Garfield (2010). **통계적 사고의 의미와 교육** [The challenge of developing statistical literacy, reasoning and thinking]. (고은성, 이경화, 지은정 역). 경문사(박문규). (원전은 2004에 출판)
- Zacks, S. (1971). *The theory of statistical Inference*. John wiley & Sons, pp.29-100.

Features of sample concepts in the probability and statistics chapters of Korean mathematics textbooks of grades 1-12

Lee, Youngha (Ewha Womans University)

Shin, Souyeong (Ewha Womans University)

This study is the first step for us toward improving high school students' capability of statistical inferences, such as obtaining and interpreting the confidence interval on the population mean that is currently learned in high school.

We suggest 5 underlying concepts of 'discretion of contingency and inevitability', 'discretion of induction and deduction', 'likelihood principle', 'variability of a statistic' and 'statistical model', those are necessary to appreciate statistical inferences as a reliable arguing tools in spite of its occasional erroneous conclusions. We assume those 5 concepts above are to be gradually developing in their school periods and Korean mathematics textbooks of grades 1-12 were analyzed. Followings were found.

For the right choice of solving methodology of the given problem, no elementary textbook but a few high school textbooks describe its difference between the contingent circumstance and the inevitable one. Formal definitions of population and sample are not introduced until high school grades, so that the developments of critical thoughts on the reliability of inductive reasoning could not be observed. On

the contrary of it, strong emphasis lies on the calculation stuff of the sample data without any inference on the population prospective based upon the sample. Instead of the representative properties of a random sample, more emphasis lies on how to get a random sample. As a result of it, the fact that 'the random variability of the value of a statistic which is calculated from the sample ought to be inherited from the randomness of the sample' could neither be noticed nor be explained as well. No comparative descriptions on the statistical inferences against the mathematical(deductive) reasoning were found. Few explanations on the likelihood principle and its probabilistic applications in accordance with students' cognitive developmental growth were found. It was hard to find the explanation of a random variability of statistics and on the existence of its sampling distribution. It is worthwhile to explain it because, nevertheless obtaining the sampling distribution of a particular statistic, like a sample mean, is a very difficult job, mere noticing its existence may cause a drastic change of understanding in a statistical inference.

* **Key Words** : Sample concept(표본개념), Statistical inference(통계적 추론), 통계교육(statistics education), Textbook analysis(교과서 분석)

논문접수 : 2011. 9. 30

논문수정 : 2011. 11. 3

심사완료 : 2011. 11. 18