대한수학교육학회지 수학교육학연구 제 22 권 제 2 호
Journal of Educational Research in Mathematics
Vol. 22, No. 2, 117-136. May 2012

정규분포에 대한 교수학적 변환 방식과 학생들의 이해 분석[1)]

신 보 미*

이 연구는 정규분포에 대한 교수학적 변환 방식과 학생들의 이해의 특징을 분석함으로써 고등학교 통계 단원에서 정규분포를 지도하는 것과 관련된 시사점을 얻는데 목적을 두었다. 이를 위해 정규분포의 역사 발생 과정과 학문적 지식으로서 정규분포의 의미를 확인하여 정규분포의 교수학적 변환 방식을 살펴보기 위한 4개의 분석 관점을 추출하고 이를 토대로 미국, 영국, 우리나라 교과서의 정규분포에 대한 교수학적 변환 방식을 분석하였다. 또한 정규분포에 대한 학생들의 이해 특징과 관련된 선행 연구 결과를 살펴봄으로써 고등학생들의 정규분포에 대한 이해의 특징을 기술하는데 필요한 분석 관점을 4가지로 구체화하였으며, 이를 토대로 분석한 학생들의 이해 특징을 4가지로 요약하였다.

Ⅰ. 서론

정규분포는 연속확률분포 중에서 가장 널리 이용되는 대표적인 분포로(박정식 · 윤영선, 2004), 자료 분석에서 통계적 추정으로 나아가기 위해 확률 분포를 지도하는 과정에서 중요한 통계적 모델이 된다(Batanero, Tauber, & Sanchez, 2004). 제 7차 개정 교육과정의 '미적분과 통계 기본'에서 통계적 추정은 모집단의 분포가 정규분포인 경우만 다루도록 하여, 모평균의 신뢰구간을 구하는 방법은 모집단의 분포가 정규분포라는 사실에 기초하여 유도된다(교육인적자원부, 2007). 이는 '미적분과 통계 기본'의 통계 영역이 확률분포와 통계적 추정이라는 두 단원으로 구성되어 있음을 감안할 때, 확률 분포 단원에서 다루는 정규분포 개념이 다음 단원인 통계적 추정의 주요 내용 요소를 이해하는데도 필수적인 요소가 됨을 시사한다. 그럼에도 고등학교 교육과정에서 정규분포를 다루는 방식에 대한 교수학적 연구는 그리 많지 않다(허명회, 2007; 조정수, 2010; Batanero et al., 2004).

Cobb & Moore(1997)는 정규분포와 같은 통계적 모델을 이론적으로 도입하여, 분포의 성질을 연역적으로 설명하는 것은 귀납 과학으로서 통계적 실제를 지도하는데 한계가 있다고 지적한 바 있다. 이들에 따르면 학생들은 자료 분석에 관한 실질적인 경험을 바탕으로 확률 분포 모델을 직관적으로 구성해 볼 필요가 있다. 우정호(2008)는 우리나라 통계 교육의 주요 문제 중 하나는 고등학교 통계 단원에서 정규분포가 이론적인 확률분포로 도입된 다음 이를 토대로 전형적인 확률 문제 해결을 위한 알고리즘에의 숙달이 주요 교육 내용으로 지도됨으로써 통계적 사고 교육을 소홀히 하고 있는 점이라고 비판하였다.

이에 이 연구는 고등학교 통계 단원에서 정규분포를 지도하는 것과 관련된 시사점을 얻기 위하여 미국, 영국, 우리나라 교과서의 정규분포에

* 전남대학교, bomi0210@jnu.ac.kr
1) 이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 연구되었음(NRF-2011-332-B00475).

대한 교수학적 변환 방식을 분석하고 고등학생들의 정규분포에 대한 이해의 특징을 살펴본다. 이를 위해 우선 정규분포의 역사 발생 과정과 학문적 지식으로서 정규분포의 의미를 확인하여 정규분포의 교수학적 변환 방식을 살펴보기 위한 분석 관점을 추출하고 이를 바탕으로 미국 Key Curriculum Press의 Statistics in Action 교과서, 영국 SMP(School Mathematics Project)의 Statistics 1 교과서, 제 7차 개정 교육과정의 '미적분과 통계 기본'교과서 13종을 분석하고 그 특징을 기술한다. 또한 정규분포에 대한 학생들의 이해 특징과 관련된 선행 연구 결과를 살펴보고 고등학생들의 정규분포에 대한 이해의 특징을 기술하는데 필요한 분석 관점을 구체화한다. 이러한 분석 관점에 기초하여 검사 문항을 개발하고, 일반계 고등학교 학생을 대상으로 지필검사를 실시한 다음 그 결과를 바탕으로 학생들의 정규분포에 대한 이해의 특징을 기술한다.

II. 이론적 배경

1. 정규분포의 역사 발생

정규분포는 확률 문제를 탐구하는 과정에서 1733년 De Moivre에 의하여 이항분포의 극한 분포로 유도되었다(Hald, 1990). 이 시기에 진행된 정규분포에 대한 연구는 이항분포에 바탕을 둔 여러 사건의 확률과 기댓값에 관련된 것으로 자료 분석이나 통계적 추론과는 다소 거리가 있었다. 한편, 18세기 중반부터 천문학을 중심으로 관측 값에 대한 연구가 시작되어, 오차를 설명하는데 적절한 분포를 찾는 문제, 측정 결과로부터 참값을 추정하는 문제 등이 주요한 연구 주제가 되었다. 이러한 연구의 동향 하에서 1809년 Gauss는 정규분포를 오차의 분포로 이용할 수 있는 근거를 제시함으로써 De Moivre가 만든 정규분포와 오차분포의 관계를 설명하였다(Hald, 1998: 351-357). 또한, Bessell는 이러한 정규분포를 오차분포로 쓰는 것이 타당한지를 경험적으로 검증하려고 시도하였다. 그는 별의 위치에 대한 관측 값의 오차분포를 조사하여 그 결과가 가정된 오차 분포와 비교적 일치한다는 결과를 얻었다(Stigler, 2005).

19세기에 접어들면서 천문학에서의 오차이론은 사회 현상의 자료를 설명하는데 활용되기 시작하였다. Quetelet는 사람에 대한 측정 결과에서 그 자료 값이 평균으로부터 떨어진 편차를 설명하는데 정규분포를 이용하였다. 그는 '평균적인 사람(average man)'이라는 개념을 정의하고 이를 사회에서 모범의 역할을 하는 일종의 참값처럼 다루었다(조재근, 2008). 이를테면 그는 '병사 5,738명의 가슴둘레 분포는 병사 한 명의 가슴둘레를 5,738번 측정한 오차 분포와 다르지 않다(Stigler, 2005: 422)'고 생각하였다. 그는 모집단 자료의 변이성(variability)을 그 자체로 인정하기 보다는 고정된 참값의 오차로 간주하였다는 점에서 참값에 대한 오차분포를 정규분포로 설명하려고한 천문학의 관점을 따르고 있다고 볼 수 있다.

Quetelet이후 통계학이 천문학을 넘어서 보다 넓게 활용될 수 있는 구조를 만들어 낸 역할을 한 사람은 Galton, Edgeworth, Yole, Pearson이다(Stigler, 2005). 특히, Galton은 Quetelet와 달리 정규분포를 고정된 참값의 오차를 나타내는 분포로서 보다는 모집단 자료의 변이성 패턴을 해석하는 개념적 도구로 활용하였다(남주현, 2006). 이를테면, Galton은 부모의 키에 따른 자식의 키의 변화에 있어 왜 변이성이 증가하지 않는지에 의문을 품고 이를 합리적으로 설명하기 위해 정규분포 곡선을 활용하였다. 그는 강낭콩 씨앗의 무게에 대한 실제 자료[2]를 바탕으로, 퀸컹크스(Quincunx) 장

치를 이용한 실험 결과[3])를 이항분포의 아이디어에 기초하여 해석함으로써 '평균으로의 회귀이론(regression to the mean)'을 탄생시켰다. Pfannkuch & Wild(2004)는 Galton의 이러한 분석이 통계적 사고가 참값이라는 중심보다는 모집단 자료 자체의 변이성에 집중하도록 만든 최초의 역사적 사건이라고 평가하였다.

19세기 말부터는 점차 비대칭인 분포를 고려하는 경향이 생겨나기 시작하였다(임지현, 2008). Fechner는 비대칭인 분포를 따르는 자료가 존재할 수 있음을 가정하였으며, Edgeworth는 정규분포만을 믿는 연구 동향을 비판[4]하였다. 19세기 말의 이러한 추세에 힘입어 Pearson은 치우친 곡선이라는 개념을 발전시켰으며, 보다 많은 현상을 해석할 수 있는 변형된 분포에 대한 연구를 진행하였다.

이상 정규분포의 역사 발생 과정으로부터 정규분포 개념에는 이항분포의 근사 분포로서의 의미와 고정된 참값에 대한 측정값의 오차 분포로서의 의미가 함께 담겨 있음을 확인하였다. 이와 더불어 정규분포는 모집단에 있는 자료의 변이성 패턴을 해석하는 분포로 다루어질 수 있음을 알아보았다. 또한, 19세기 말에 접어들면서는 자연 현상이나 사회 현상의 통계 자료를 보다 적절히 해석하기 위해 정규분포뿐만 아니라 치우친 분포(skewed distribution)와 같이 다양한 분포 모델이 개발되었음을 확인하였다. 이 점을 고려하여 미국, 영국, 우리나라 교과서의 정규분포에 대한 교수학적 변환 방식을 살펴보기 위한 분석 관점을 다음과 같이 추출할 수 있다.

관점 1. 정규분포는 이항분포의 근사 분포, 고정된 참값에 대한 측정값의 오차 분포, 모집단의 변이성 패턴을 나타내는 분포 중 어떤 의미로 다루어지는가?

관점 2. 정규분포를 따르지 않는 자료의 예와 그 분포가 다루어지는가?

2. 학문적 지식으로서 정규분포

대부분의 대학 교재(고승곤·양완연·오현숙, 2003; 김상익 외, 2005; 김우철 외, 2006; 박정식 · 윤영선, 2004; Hogg & Craig, 2002; Hogg & Tanis, 2005)에서 정규분포는 다음과 같이 특정한 확률밀도함수를 갖는 분포로 정의된다.

정규분포는 평균 μ 와 표준편차 σ 에 의해 결정되는 분포로서 다음과 같은 확률밀도함수를 갖는다.

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

(김우철 외, 2006: 111)

정규분포에 대한 이러한 정의이후, 정규분포 곡선과 그 곡선의 성질, 정규분포를 따르는 확률변수 X 의 대표적인 확률 값[5]) 등이 소개된다.

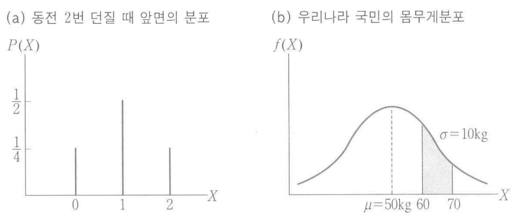
2) Galton은 무게에 따라 7개의 집단으로 나눈 강낭콩 씨앗을 친구들에게 키워달라고 부탁하였다. 그 결과 각 집단에 있는 강낭콩의 무게는 각각 정규분포곡선을 따를 뿐만 아니라 그 중심이 되는 무게는 각각 다르지만 흩어진 정도는 비슷하다는 사실을 알게 되었다. 7개의 자식 집단에서 강낭콩의 무게는 정규분포를 따르지만 그 분포의 중심은 부모의 무게보다는 전체 집단의 평균에 더 가까웠다. 즉, 변이성의 증가는 집단 전체의 변이성 정도를 유지하는 선에서 그치게 되며, 이를 '복귀유전'이라는 개념으로 정리할 수 있다(조동기, 2007: 9).

3) 이 장치의 꼭대기에 있는 깔때기에 작은 구슬들을 넣으면 구슬은 안에 박혀 있는 못들 사이로 떨어지게 된다. 이 때, 구슬은 하나의 못에 부딪히게 되는데 그 후 왼쪽이나 오른쪽으로 떨어질 가능성은 반반이 된다. 이는 앞면과 뒷면이 나올 가능성이 같은 동전 여러 개를 던졌을 때와 같은 무작위 효과를 발생시켜 최종적으로 바닥에 있는 칸막이 방에 정규분포와 같은 모양으로 쌓이게 된다(임지현, 2008: 16).

4) "이론은 잘못된 원칙과는 구별되어야 한다. 그 원칙은 어떤 집단의 통계 값들이 하나의 봉우리를 가지는 곡선으로 나타난다면 그 곡선은 반드시 정규곡선의 일종이어야만 한다는 원칙이다(Stigler, 2005)."

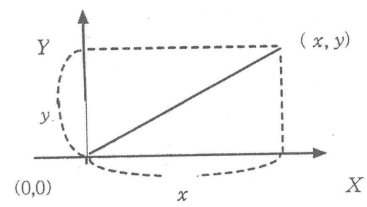
특히, 박정식 · 윤영선(2004)은 연속확률 분포인 정규분포에서의 확률은 이산확률분포에서와 달리 주어진 정규분포의 확률밀도함수에 대한 넓이로 나타남을 다음과 같이 설명한다.

> 그림에서 보면 이산확률분포는 높이로 확률을 나타내지만 연속확률분포는 높이가 아니라 넓이로 확률을 나타내게 된다. 우리나라 국민의 몸무게의 분포가 평균 $\mu=50kg$, 표준편차 $\sigma=10kg$인 정규분포를 따른다고 하면 국민 중에서 몸무게가 60~70kg에 있을 확률은 전체 넓이 중에서 파란 부분이 차지하는 비율로써 나타낼 수 있다(박정식 · 윤영선, 2004: 158-159).



이는 정규분포에서의 확률은 확률밀도함수에 대한 넓이가 됨을 설명할 때, 위와 같이 이산확률분포 그래프와 연속확률분포 그래프 사이에 존재하는 차이를 인식하도록 하는 것이 중요함을 보여준다. 이산확률분포 그래프의 y축은 확률변수 X에 대한 확률 $P(X)$를, 연속확률분포 그래프의 y축은 확률변수 X에 대한 확률 $P(X)$가 아니라 확률밀도함수의 함숫값 $f(X)$가 된다.

한편, 구자홍(1999)은 정규분포의 확률밀도함수가 $f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$와 같이 $y=e^{-x^2}$의 형태가 되는 이유를 오차분포인 정규분포의 의미로부터 다음과 같이 유도하였다.

> 좌표평면 상에서 참값을 $(0,0)$, 측정값을 (x,y)이라 하자. 오차의 분포가 방향과는 무관하고 오직 원점으로부터의 거리에만 관련되며 직교하는 두 방향의 오차는 서로 독립이라고 가정하자. 예를 들면, 좌표평면의 원점에 서서 원점을 겨냥하고 작은 공을 떨어뜨려 공이 떨어진 위치를 측정하는 경우가 이상의 가정에 따른다고 볼 수 있다. 이 때, 오차분포의 확률밀도함수를 φ라고 하면 주어진 가정에 의하여 다음 관계식이 성립한다.
>
> $$\varphi(x^2+y^2)=\varphi(x^2)\cdot\varphi(y^2)\ \cdots ①$$
>
> ①에서 x와 y는 각각 원점에서 X축 방향으로 벗어난 변의 길이와 Y축 방향으로 벗어난 변의 길이이다. 즉, 떨어뜨린 공이 X축 방향으로 x만큼 벗어나고, Y축 방향으로 y만큼 벗어났다면 이 공은 참값인 원점으로부터 $\sqrt{x^2+y^2}$ 만큼의 오차를 발생시켰다고 볼 수 있다. 이 때, X축 방향으로의 오차 x와 Y축 방향으로 오차 y는 서로 독립이다.
>
> 
>
>
> 한편, 함수 방정식 $f(x+y)=f(x)\cdot f(y)$의 해 $f(z)$는 $f(z)=e^{hz}$이므로 구하고자 하는 φ가 확률밀도함수임을 감안하면 임의의 실수 h에 대하여 식 ①의 해는 다음과 같이 나타낼 수 있다.
>
> $$\varphi(x^2)=k\cdot e^{-h^2x^2}$$
>
> φ가 확률밀도함수가 되도록 상수 k을 결정하면 오차분포에 관한 확률밀도함수를 얻을 수 있다(구자홍, 1999: 4).

이상의 설명은 실제 관측 값들의 오차를 경험

5) $P(\mu-\sigma \leqq X \leqq \mu+\sigma)=0.6827$, $P(\mu-2\sigma \leqq X \leqq \mu+2\sigma)=0.9545$, $P(\mu-3\sigma \leqq X \leqq \mu+3\sigma)=0.9973$(고승곤 외, 2003: 193)

적으로 관찰한 결과로부터 유도하기에는 거의 불가능해 보이는 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 와 같은 복잡한 함수가 어떻게 정규분포의 확률밀도함수가 되었는지를 논리적으로 이해하는데 도움을 준다. 그러나 이러한 유도 방식은 대학 학부 수준의 통계학 교재에도 거의 실려 있지 않을 정도로 전문적인 수학 내용을 담고 있다. 이는 정규분포의 확률밀도함수가 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 와 같이 되는 이유를 합리적으로 이해하기 위해서는 상당한 수학적 지식이 필요함을 시사한다.

이상으로 대학 통계학 교재와 통계 분야의 선행연구를 통해 학문적 지식으로서 정규분포를 분석한 바에 의하면, 연속확률분포인 정규분포에서의 확률은 이산확률분포의 확률에 비추어 그 차이점이 설명됨을 확인하였다. 또한 정규분포의 확률밀도함수는 오차분포인 정규분포의 의미에 기초하여 논리적으로 유도될 수 있으나, 그 유도 방식을 적절히 이해하는 데는 상당히 전문적인 수학적 소양이 필요함을 예측할 수 있었다. 이러한 사실을 고려하여 미국, 영국, 우리나라 교과서의 정규분포에 대한 교수학적 변환 방식을 살펴보기 위한 분석 관점을 다음과 같이 추출할 수 있다.

> 관점 3. 정규분포에서 확률을 구하는 방법이 어떻게 설명되는가?
> 관점 4. 정규분포의 확률밀도함수가 어떻게 다루어지는가?

3. 정규분포에 대한 이해를 다룬 선행 연구

Batanero et al.(2004)에 따르면 정규분포에 대한 학생들의 이해를 다룬 연구가 거의 없으며, 현재까지 이루어진 연구의 대부분은 대학생 이상의 성인을 연구 대상으로 삼고 있다. 정규분포에 대한 학생의 이해를 다룬 선행 연구가 이처럼 제한적이기는 하나, 이 연구가 학교 교육과정을 통해 정규분포를 학습한 고등학생들의 이해 특징을 살펴보는데 그 목적을 두기 때문에 이하에서는 연구대상이 성인이라고 하더라도 정규분포와 관련된 이해의 특징을 기술한 선행 연구의 구체적인 내용을 간략하게 살펴본다.

Wilensky(1997)는 상당히 전문적인 수학적 소양을 지닌 대학원생 2명을 연구대상으로 정규분포에 대한 이해를 살펴보는 면담을 실시하였다. 그는 연구대상에게 ‘사람의 키와 같은 자료를 해석하는데 왜 정규분포를 사용할 수 있는가?’라는 질문을 제시함으로써 면담을 시작하였다. 연구대상은 정규분포와 관련된 전형적인 확률 문제는 능숙하게 해결할 수 있었지만 정규분포의 개념적 의미와 관련된 이와 같은 질문[6]에 대해서는 적절한 응답을 하지 못하였으며, 특정 자료가 정규분포를 따른다는 것이 어떤 의미인지를 설명하는데도 상당한 인식론적 불안(epistemological anxiety)을 보였다.

Huck, Cross, & Clark(1986)은 연구에 참여한 대학생들 대부분이 정규분포를 따르는 어떤 과목의 시험 점수는 최댓값과 최솟값이 존재하지 않는다고 생각하는 오류를 보였다고 하였다. 연구대상은 이론적 분포로서 정규분포와 시험 점수라는 실제 자료 사이의 차이를 인식하지 못하고, 정규분포는 확률변수 X가 $-\infty < X < \infty$에서 정의되기 때문에 정규분포를 따르는 분포는 유한인 분포라고 하더라도 최댓값과 최솟값이 존재할 수 없다고 설명하였다. Huck(2009: 23)은 100,000명의 키는 근사적으로 정규분포를 따를 수 있지만 측정 대상이 되는 그룹의 수가 아무

6) 이는 19세기에 Galton이 부모의 키에 따른 자식의 키의 변화에 있어 왜 변이성이 증가하지 않는가에 의문을 두었던 사실과 같은 맥락에 있는 질문으로 Galton은 이 질문에 답하는 과정을 통해 모집단의 변이성 패턴을 나타내는 분포로서의 정규분포가 지닌 의미를 설명한 바 있다.

리 많다고 하더라도 키의 최댓값과 최솟값은 항상 존재할 것이므로 사람의 키의 분포가 정확하게 정규분포가 되는 것은 아니라고 하였다.

Batanero, Tauber, & Meyer(1999)는 대학생을 대상으로 구체적인 자료 집합들을 제시하고 이러한 자료 집합이 정규분포를 따른다고 보는 것이 적합한지 그렇지 않은지에 대해 설명해 보도록 하였다. 그러나, 이 연구에서 거의 모든 학생들은 이러한 질문에 적절한 설명을 하지 못하였다. Keith & Bower(2003)는 전문가를 대상으로 한 실험에서 연구대상 대부분이 정규분포를 따르지 않는 자료는 통계적으로 뭔가 문제가 있는 것으로 판단하는 경향이 있다고 지적하였다.

이상 정규분포에 대한 성인들의 이해를 다룬 선행 연구를 분석한 바에 의하면, 연구대상들은 정규분포와 관련된 전형적인 확률 문제는 잘 해결하면서도 정규분포에 대한 개념적 의미와 관련된 질문에는 적절하게 답하지 못하는 경우가 있었다. 또한 이론적 모델인 정규분포와 실제 자료간의 차이를 구분하지 못하기도 하고, 어떤 자료가 정규분포를 따르는지를 판단하는 데도 어려움을 겪는 사례가 있었다. 이상을 고려하여 고등학교 학생들의 정규분포에 대한 이해 특징을 분석하기 위한 분석 관점을 다음과 같이 구체화할 수 있다.

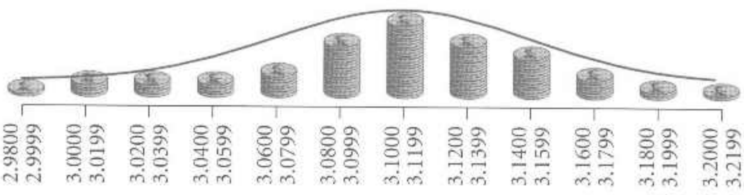
관점 5. 학생들이 정규분포와 관련된 전형적인 확률 문제를 어떻게 해결하는가?
관점 6. 학생들이 정규분포에 대한 개념적 이해와 관련된 질문에 어떻게 답하는가?
관점 7. 학생들이 이론적 모델인 정규분포와 경험적 자료 사이의 관계를 어떻게 인식하는가?
관점 8. 학생들이 정규분포를 따르지 않는 자료의 예를 들 때 어떤 특징이 있는가?

III. 정규분포의 교수학적 변환 방식

1. 미국 교과서

우리나라와 달리 미국은 국가 수준의 강한 구속력을 갖는 단일한 교육과정이 존재하지 않기 때문에 교과서에 출판사와 저자의 고유한 관점이 반영될 여지가 많다. 이 연구에서 살펴본 출판사 Key Curriculum Press의 교과서에 대하여 박경미 외(2002: 320, 323)는 수학적 개념을 풍부한 맥락과 더불어 실생활과의 관련 속에서 제시하는데 주안점을 두면서 주제 중심 전개방식을 따르고 있는 교재로 설명하고 있다. 그는 이러한 경향의 교과서가 미국 NCTM 규준집의 아이디어를 구현하고자 한 Core-Plus, UCSMP, IMP, STEM 등의 대형 프로젝트 결과 출판된 교과서들과 그 맥을 함께 하며 이와 같은 유형의 교과서들이 점차 미국 교과서들의 주류를 이루는 추세라고 하였다. Key Curriculum Press의 Statistics in Action은 고등학생을 대상으로 한 통계 입문과정의 교재로, 실생활 문제를 통계적으로 해결하는데 필요한 다양한 탐구 활동을 다루고 있는 것이 특징이다(Watkins, Scheaffer, & Cobb, 2004: vii).

Key Curriculum Press의 Statistics in Action에서 정규분포는 측정값의 오차 분포로서 처음 도입되고 이후 모집단의 변이성 패턴을 나타내는 분포로 설명된다. 이 교과서에는 우선 테니스 공 한 개의 지름을 학급 전체 학생들이 측정해 보고 그 결과를 서로 비교해 보도록 한다. 이를 통해 개별 측정값이 종 모양 형태로 분포됨을 직관적으로 확인하도록 하고 이로부터 정규분포를 종 모양 형태의 곡선을 갖는 분포로 정의한다. 참값이 존재하는 테니스공 한 개의 지름에 대한 측정값의 오차가 어떻게 분포하는지를 탐구하도록 하는 이러한 활동은 정규분포를 측정값의 오차분포로 소개하려는 의도를 담



[그림 Ⅲ-1] 최근에 생산된 1센트짜리 동전 100개의 무게
(Watkins, Scheaffer, & Cobb, 2004: 27)

고 있다고 볼 수 있다.

정규분포를 이상과 같이 정의한 다음에는 최근 생산된 1센트짜리 동전 100개의 무게에 대한 그래프가 [그림 Ⅲ-1]과 같은 경우를 제시함으로써 종모양의 곡선을 갖는 분포는 모집단의 변이성 패턴을 설명하는데도 활용할 수 있음을 언급하고 있다.7) 이 교과서는 정규분포를 통해 테니스 공 한 개의 지름에 대한 측정값의 오차분포를 표현할 수 있으며, 최근에 생산된 1센트짜리 동전들의 무게와 관련된 모집단 자료의 변화 패턴도 설명할 수 있다고 언급한 다음, 이 점이 정규분포를 통계적으로 매우 중요하고 가치 있는 도구가 되도록 한다고 기술하고 있다.

그러나 [그림 Ⅲ-1]에서는 정규분포 곡선이 1센트짜리 동전 100개의 무게에 대한 점그래프와 겹쳐 그려짐으로써 연속확률밀도함수의 그래프인 정규분포 곡선이 히스토그램의 다른 표현으로 오해될 가능성을 가지고 있다. 이를테면 학생들은 최근에 생산된 1센트짜리 동전의 무게가 '3,1000g에서 3,1199g사이에 있을 확률'을 주어진 구간에서 정규분포 곡선으로 둘러싸인 영역의 넓이가 아니라 주어진 구간에서 히스토그램의 높이로 설명할 가능성이 있다. 박영희(2002: 681-682; 임지현, 2008: 73)는 이와 같이 히스토그램을 활용하여 연속확률밀도함수의 곡선에 대해 설명할 때는 히스토그램의 기둥 높이가 연속확률밀도함수의 함숫값이 되는 것은 아니므로 우선 상대도수 히스토그램을 구한 다음, 그 세로 눈금을 계급의 폭으로 나눈 값으로 변환하여 설명할 필요가 있다고 지적한 바 있다.

한편, 이 교과서는 정규분포를 측정값의 오차분포와 모집단의 변이성 패턴을 나타내는, 종모양의 그래프를 갖는 분포로 설명한 다음 정규분포의 성질 및 이를 활용한 확률 문제 해결 방법, 표준정규분포와 표준화, 이항분포와 정규분포의 관계 등 정규분포와 관련된 전반적인 내용을 다룬다. 그러나 정규분포에 대한 전반적인 특징을 설명하는 이와 같은 과정에서 정규분포의 확률밀도함수 $f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 는 전혀 도입되지 않는다.

또한 이 교과서는 '실생활에서 얻은 어떤 자료도 정확하게 정규분포를 따르지는 않으며, 근사적으로 정규분포 곡선에 가까울 뿐이다(Watkins, Scheaffer, & Cobb, 2004: 27)'는 기술을 제시함으로써 이론적인 모델인 정규분포와 실제 통계 자료 사이의 차이를 간접적으로 설명하고 있다. 정규분포를 설명한 다음 장에서는 균등분포(uniform distribution)와 치우친 분포(skewed distribution) 등을 다룸으로써 정규분포를 따르지 않는 통계 자료가 실제로 존재함을 구체적으로 보이고 있다.

7) 이에 앞서 '미국에서 생산되는 1센트짜리 동전의 무게는 3.110g정도가 되어야 하며, 허용 오차는 0.130g정도이다'라고 설명한다.

2. 영국 교과서

이 연구에서 살펴본 Statistics 1: Foundations of sampling and estimation은 SMP(School Mathematics Project) 교과서로 표본추출(sampling)과 추정(estimation)의 기초를 지도하는데 주로 활용된다. 이는 16-19세의 학생을 대상으로 한 고급 수준의 교과서이며 수학적 능력과 함께 응용력을 기르는데 목표를 두고 있다. 이 교과서의 각 장은 실제적인 활동을 통해 새로운 아이디어를 개발할 수 있는 과제로 구성되어 있다(SMP, 2005: ⅳ).

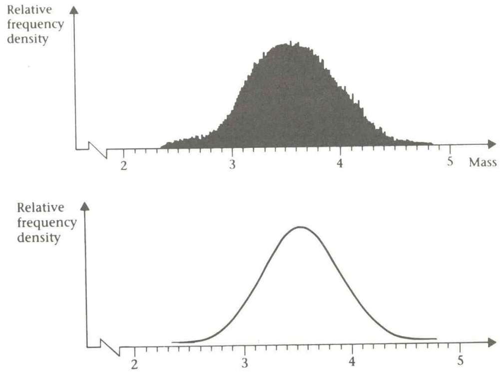
SMP 교과서의 Statistics 1에서 정규분포는 Key Curriculum Press의 Statistics in Action와 달리, 측정값의 오차 분포의 관점에서 소개되지 않고 모집단의 변이성 패턴을 나타내는 분포로서 도입된다. 이 교과서는 남자의 키에 대한 150개의 자료를 비롯한 여러 통계 자료를 제시하고 이를 히스토그램으로 정리해보는 활동을 통해 정규분포 곡선을 설명하는데, 여기에는 정규분포를 모집단의 변이성 패턴을 나타내는 분포로서 다루려는 의도가 있다고 볼 수 있다.

다만 이 교과서는 히스토그램을 활용하여 연속확률분포인 정규분포를 도입할 때, 히스토그램의 기둥 높이를 연속확률밀도함수의 함숫값으로 간주하지 않도록 하기 위하여 상대도수 밀도(relative frequency density)라는 개념을 [그림 Ⅲ-2]과 같이 먼저 정의한다.

> The height of each block on the histogram now becomes
>
> $$\text{height} = \frac{\text{relative frequency}}{\text{width of interval}} = \textbf{relative frequency density}$$

[그림 Ⅲ-2] 상대도수 밀도의 정의(SMP, 2002: 33).

그런 다음, 계급의 폭이 아주 작고 자료 값이 아주 많은 경우 [그림 Ⅲ-3]과 같이 신생아 몸무게의 상대도수 밀도 히스토그램을 부드러운 곡선으로 표현하면 종모양이면서 대칭인 함수의 그래프를 얻게 됨을 설명하고, 이와 같은 그래프를 정규분포 곡선으로 정의한다.

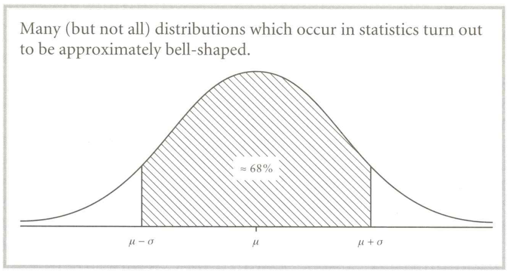


[그림 Ⅲ-3] 신생아 몸무게에 대한 상대도수밀도 히스토그램(SMP, 2002: 35).

이 교과서는 이상과 같이 정규분포 곡선을 정의한 다음 표준정규분포 곡선을 설명하고 표준정규분포 곡선의 식을 유도하는 활동을 진행한다. 우선, k의 값에 관계없이 $f(x)=ke^{-\frac{1}{2}x^2}$의 그래프와 표준정규분포 곡선의 모양이 비슷함을 확인하도록 하고, 컴퓨터 프로그램을 통해 $y=e^{-\frac{1}{2}x^2}$의 곡선 아래 넓이를 구하는 활동을 진행한다. 이를 통해 $f(x)=ke^{-\frac{1}{2}x^2}$의 곡선 아래 넓이가 1이 되기 위해서는 k의 값이 얼마가 되면 좋을지 추측해 보도록 함으로써 $k=\frac{1}{\sqrt{2\pi}}$임을 설명한다. Statistics in Action과 마찬가지로 Statistics 1도 정규분포의 확률밀도함수를 도입하지 않지만, 대신 일반적인 정규분포의 확률밀도함수에 비해 상대적으로 간단한 표준정규분포의 확률밀도함수가 $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$와 같이 되는 이유를 간접적으로 설명하고 있다.

한편, 이 교과서는 정규분포 및 표준정규분포 곡선과 그 식을 도입하기 앞서 우선 구체적인 자료 집합 5개[8]에 대한 히스토그램을 각각 그린 다음 평균 m과 표준편차 s에 대하여 각각 $m\pm s$, $m\pm 2s$, $m\pm 3s$의 비율을 구하도록 하는 탐구활동을 제시한다. 그런 다

음 이러한 활동에서 얻은 히스토그램 각각의 형태를 서로 비교하고 그 모양이 종 모양인 것과 그렇지 않은 것이 있음을 확인하도록 한다. 이로부터 '전부는 아니지만 많은 통계 자료의 분포가 근사적으로 종모양의 형태를 띤다'고 [그림 Ⅲ-4] 과 같이 설명한다.



[그림 Ⅲ-4] 종 모양 분포(SMP, 2002: 30)

이상과 같은 정규분포 곡선의 도입 과정은 자료 집합에 대한 히스토그램이 종모양의 형태를 띠는 것과 그렇지 않는 것이 존재함을 구체적으로 보임으로써 모든 통계 자료가 정규분포를 따르는 것은 아님을 직접적으로 경험할 수 있도록 한다. 그러나 이 교과서는 정규분포 이외의 균등분포나 치우친 분포에 대해서는 언급하고 있지 않으며, 이는 Statistics in Action의 전개방식과 대조를 이룬다.

3. 미적분과 통계 기본

미적분과 통계 기본 교과서 13종 중 11종은 Statistics 1과 같이 정규분포를 모집단의 변이성 패턴을 나타내는 분포의 관점에서 도입한다. 이들 교과서는 본문 내용을 시작하기 앞서 '학생 100명의 몸무게', '50명의 수리영역 점수' 등과 같은 소재의 탐구활동을 제시하는데 여기에는 정규분포를 모집단에서 추출된 여러 자료 값의 변화 패턴을 나타내는 분포로 소개하려는 의도가 있다고 볼 수 있다. 한편, 13종 중 2종의 교과서는 '주어진 손목시계 1개의 길이' 또는 '주어진 곡선 1개의 길이'를 학급 학생들이 측정해 보는 탐구활동을 제시한다. 이들 교과서는 참값을 갖는 하나의 대상에 대한 측정값의 오차분포로 정규분포를 소개하려는 것으로 볼 수 있다.

이와 같은 탐구활동 이후 모든 교과서가 단원의 맨 처음 시작부분에서 정규분포를 다음과 같이 특정한 확률밀도함수를 갖는 분포로 정의한다.

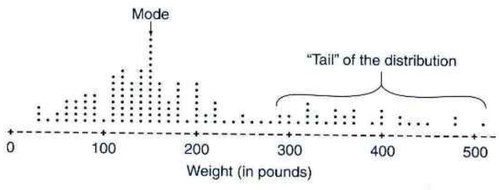
> 연속확률변수 X가 모든 실수 값을 갖고, 확률밀도함수 $f(x)$ 가 $f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$ $(-\infty<x<\infty)$으로 주어질 때, X의 확률분포를 정규분포라 하고, 이것을 기호로 $N(m,\ \sigma^2)$ 과 같이 나타낸다(김해경 외, 2009: 183)

앞 장에서 살펴보았듯이 정규분포의 확률밀도함수가 $f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 이 되는 이유를 고등학교 교육과정에서는 합리적으로 설명하는 것은 거의 불가능하다. 그럼에도 현재와 같이 정규분포를 처음 도입하는 맥락에서부터 '확률밀도함수가 $f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 와 같은 분포'라고 제시하는 것은 학생들이 이러한 정규분포의 정의만을 단순히 수용하도록 만들 우려를 갖고 있다. 학생들은 이와 같이 복잡한 함수가 어떻게 발견되었으며, 탐구활동에서 다루어본 구체적인 통계 자료의 분포가 어떻게 이와 같은 복잡한 함수를 따르게 되는지, 그러한 분포의 확률밀도함수가 왜 $y=e^{-x^2}$의 형태를 띠게 되는지를 거의 이해할 수 없기 때문이다. 나아가 '확률밀도

8) ① 줄의 길이가 50cm, 처음 시작 각도가 10°인 진자가 10회 왕복하는데 걸린 시간에 대한 100개의 자료, ② 줄의 길이가 50cm, 처음 시작 각도가 100°인 진자가 10회 왕복하는데 걸린 시간에 대한 자료, ③ 60세 이전 성인 남자의 키에 대한 150개의 자료, ④ 60세 이전 성인 여자의 키에 대한 150개의 자료, ⑤ 자동차가 어느 교차로를 통과하는데 걸린 시간에 대한 243개의 자료

함수가 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 와 같은 분포'라는 정의에는 통계적 모델로서 정규분포가 갖는 이항분포의 근사 분포로서의 의미도, 참값에 대한 오차분포로서의 의미도, 모집단 자료의 변이성을 나타내기 위한 분포로서의 의미도 드러나지 않는다. 이러한 관점에서, Statistics in Action이나 Statistics 1의 전개 방식이 미적분과 통계 기본에서 정규분포를 정의하는 방식에 시사하는 바가 있다. 이들 두 교과서는 정규분포의 확률밀도함수 자체를 소개하지 않고 구체적인 통계 자료의 히스토그램에 대한 탐구활동을 토대로 종 모양 곡선을 갖는 분포로 정규분포를 정의하거나, 마찬가지 방법으로 정규분포 곡선을 정의하고 그와 관련된 상당한 논의를 진행한 다음 표준정규분포의 확률밀도함수만을 간단하게 다룬다.

한편, 정규분포에 대한 정의와 더불어 미적분과 통계 기본 교과서 6종은 '자연 현상이나 사회 현상에서 관측하는 자료의 수를 한없이 크게 하면 좌우 대칭인 종모양의 곡선에 가까워진다' 고 설명한다. 또, 10종의 교과서는 '자연현상이나 사회현상에서 관측되는 자료의 대부분이 정규분포를 따른다'고 기술한다. 여기에는 정규분포의 통계적 가치를 '자료 수를 늘리면 가까워지는 분포' 또는 '현상으로부터 얻은 자료의 대부분이 따르는 분포'라는 두 가지 의미로 설명하려는 교수학적 의도가 담겨 있는 것으로 보인다. 그러나, 통계 자료 중에는 관측하는 자료의 수를 크게 하여도 좌우가 대칭인 종모양의 곡선에 가까워지지 않는 경우가 존재한다[9]. 이를 테면, 사람들이 태어난 달은 자료의 수를 늘릴수록 균등분포에 가까워진다. 또한, 야생 곰의 몸무게 자료에서 극단적인 값은 자료의 중심에 대하여 대칭으로 나타날 수 없기 때문에 그 분포는 자료의 수를 늘릴수록 [그림 Ⅲ-4]와 같이 왼쪽으로 치우친 분포에 가까워진다.



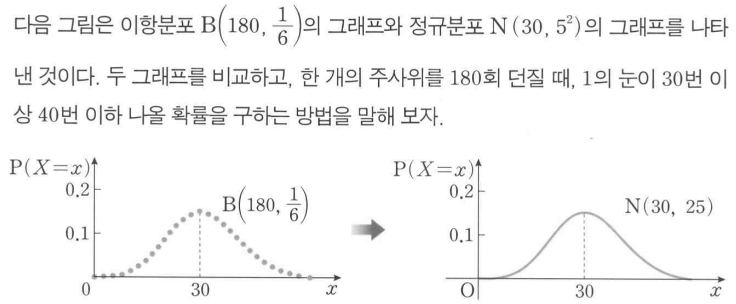
[그림 Ⅲ-4] 야생 곰의 몸무게
(Watkins, Scheaffer, & Cobb, 2004: 30)

한편, 허명회(2007)는 몸무게나 키 이외에 실제 관측 자료가 정규분포를 따르는 경우가 우리 주변에 얼마나 많은가에 의문을 제기한 바 있다. 그는 정규분포가 갖는 통계적 의의는 실생활에서 얻어지는 자료의 대부분이 따르는 분포라는 점에 있는 것이 아니라, 이항분포의 근사 분포라는 점에 있다고 설명하였다. 허명회에 따르면 이항분포의 극한으로서 정규분포는 그 개념의 역사 발생 맥락과도 일치하며, 시행횟수가 매우 큰 이항분포에서 어떤 사건이 일어날 확률 값을 지필환경에서 구하고자 할 때에도 유용하다. 이상은 정규분포를 다루는 과정에서 '자료 수를 늘리면 가까워지는 분포' 또는 '현상으로부터 얻은 자료의 대부분이 따르는 분포'와 같은 미적분과 통계 기본 교과서의 설명이 필수적인지에 대한 반성적 검토가 필요함을 시사한다.

미적분과 통계 기본 교과서는 단원의 말미에서 이항분포와 정규분포의 관계를 다룬다. 모든 교과서가 주사위나 동전 등의 구체적인 예를 통해 시행 횟수가 커질수록 이항분포의 그래프가 정규분포 곡선에 가까워짐을 보임으로써 '확률변수 X 가 이항분포 $B(n,\ p)$ 를 따를 때, n 이 충분히 크

9) 통계 자료 중에는 국민들의 연봉 액수처럼 자료의 수가 충분히 크더라도 오른쪽으로 치우친 분포를 따르는 것도 얼마든지 존재한다(Moore, 2001).

면 X는 근사적으로 정규분포 $N(np,\ npq)$ 을 따른



[그림 Ⅲ-5] 이항분포 그래프와 정규분포 그래프의 관계
(유희찬 외, 2009: 199)

다'고 설명한다. 그러나 이 과정에서 6종의 교과서는 이항분포 그래프의 y축과 정규분포 그래프의 y축에 다음과 같이 $P(X=x)$ 을 동일하게 제시하고 있다.

이는 이산확률분포인 이항분포 그래프의 y축은 확률변수 $X=x$ 에 대한 확률 $P(X=x)$ 을 나타내는 반면, 연속확률분포인 정규분포 그래프의 y축은 확률 $P(X=x)$ 가 아니라 확률밀도함수 f 에 대한 함숫값 $f(x)$ 를 나타냄을 간과한 결과로 볼 수 있다. 장대흥(1999)은 컴퓨터의 비약적인 발달로 시행횟수가 매우 큰 이항분포에서의 확률 값을 구하는 것이 그리 어렵지 않게 되었으므로, 이항분포의 정규근사가 주는 통계적 의의는 그 근삿값의 계산에 있기보다는 아주 다른 성질을 갖는 이산확률분포와 연속확률분포가 적절한 조건하에서는 서로를 넘나들 수 있다는 점이라고 설명하였다. 이러한 맥락에서 볼 때, 학교수학에서 이항분포와 정규분포 사이의 관계를 다룰 때는 두 분포를 적절한 조건하에서 '같게' 볼 수 있다는 점과 함께, 이항분포 그래프에서 $P(X=x)$ 와 정규분포 그래프에서 $f(x)$ 의 관계와 같이 두 분포가 원래는 어떻게 서로 '다른' 성질을 가지고 있었는지를 분명히 할 필요가 있다. 이는 연속확률분포 단원에 대한 지도상의 유의점으로 '확률밀도함수 $f(x)$ 가 확률 그 자체를 의미하는 것은 아님을 강조하도록 한 (교육과학기술부, 2008)' 교육과정의 취지와도 일치한다.

4. 논의

이상 정규분포의 교수학적 변환 방식에 대하여 Key Curriculum Press의 Statistics in Action, SMP의 Statistics 1, 제 7차 개정 교육과정의 미적분과 통계 기본을 분석한 결과를 앞서 추출한 4가지 분석 관점에 비추어 정리하면 다음과 같다.

관점 1. 정규분포는 이항분포의 근사 분포, 고정된 참값에 대한 측정값의 오차 분포, 모집단의 변이성 패턴을 나타내는 분포 중 어떤 의미로 다루어지는가?

Statistics in Action은 정규분포를 통해 고정된 참값에 대한 측정값의 오차 분포와 모집단 자료의 변이성 패턴을 나타낼 수 있다고 설명하면서 이런 점에서 정규분포가 통계적으로 중요한 위치를 차지한다고 기술하고 있다. 반면, Statistics 1 및 미적분과 통계 기본은 탐구활동 과제 등의 특징을 고려해 볼 때, 정규분포를 모집단 자료의 변이성 패턴을 나타내는 분포로 설명하려는 경향을 보인다.

관점 2. 정규분포를 따르지 않는 자료의 예와 그

분포가 다루어지는가?

Statistics in Action은 정규분포를 다룬 다음 장에서 균등분포나 치우친 분포를 자세하게 다룬다. Statistics 1 및 미적분과 통계 기본은 정규분포 이외의 분포를 구체적으로 다루고 있지는 않다. 그러나 Statistics 1은 정규분포 곡선의 도입 과정에서 자료의 분포가 종모양의 형태를 띠는 것과 그렇지 않는 것을 동시에 제시함으로써 모든 통계 자료가 정규분포를 따르는 것은 아님을 학생들이 경험적으로 알 수 있도록 하고 있다.

관점 3. 정규분포에서 확률을 구하는 방법이 어떻게 설명되는가?

Statistics 1은 상대도수 히스토그램을 활용하여 연속확률분포인 정규분포를 도입할 때, 연속확률분포에서의 확률은 상대도수 히스토그램에서와 달리 구간에서 곡선으로 둘러싸인 부분의 넓이임을 설명하기 위하여 상대도수 밀도히스토그램(relative frequency density histogram)이라는 개념을 도입한다. 이를 통해 상대도수 히스토그램에서의 기둥 높이와 연속확률밀도함수의 함숫값이 서로 같지 않음을 설명한다. 반면, Statistics in Action 및 미적분과 통계 기본은 정규분포의 확률밀도함수 f 에 대하여 $f(x)$ 가 확률을 의미하는 것처럼 오해할 여지가 있는 그림을 담고 있다.

관점 4. 정규분포의 확률밀도함수가 어떻게 다루어지는가?

Statistics in Action과 Statistics 1은 정규분포의 확률밀도함수를 전혀 도입하지 않는다. 다만, Statistics 1은 정규분포 곡선을 종 모양 곡선으로 정의하고 그와 관련된 상당한 논의를 진행한 다음 표준정규분포의 확률밀도함수만을 간단하게 다룬다. 그러나 미적분과 통계 기본은 정규분포를 처음 도입하는 맥락에서부터 '확률밀도함수가 $f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 와 같은 분포'로 정규분포를 정의한다.

Ⅳ. 정규분포에 대한 학생들의 이해

이 연구의 두 번째 목적은 학교 교육과정을 통해 정규분포를 학습한 학생들의 이해 특징을 살펴보는데 있다. 이에 일반계 고등학교 학생을 대상으로 지필검사를 실시하기 위하여 앞서 추출한 분석 관점 5, 6, 7, 8을 토대로 검사 문항을 개발하였다. 문항 1-1은 정규분포와 관련된 전형적인 확률 문제[10]로 분석 관점 5에 대한 답을 얻기 위한 것이다. 문항 1-2는 학생들이 이론적 모델로서 정규분포와 경험적 자료간의 차이를 구별할 수 있는지 알아보기 위한 문항으로 분석 관점 7에 답하기 위함이다. 문항 2는 학생의 정규분포에 대한 개념적 이해의 특징을 살펴보기 위한 것으로 분석 관점 6과 관련된다. 문항 3에서는 정규분포를 따르지 않는 자료의 예를 들도록 하여 학생들이 어떤 자료가 정규분포를 따른다는 말의 의미를 어떻게 이해하고 있는지 간접적으로 살펴보기 위한 것으로 분석 관점 8과 관련된다.

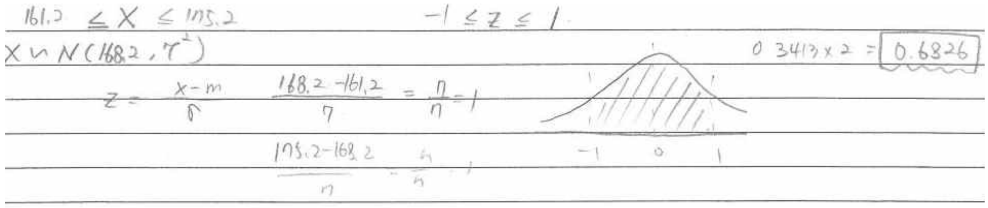
이상과 같이 개발된 검사 문항을 활용하여 2012년 1월에 **광역시 소재 4개의 일반계 고등학교 상위권 학생 92명을 대상으로 지필검사를 진행하였다. 각 문항에 대한 학생들의 응답 결과로부터 드러난 특징을 요약하면 <표 Ⅳ-1>과 같다.[11]

<표 Ⅳ-1> 정규분포에 대한 학생들의 이해 특징

유형	내용	문항
유형 1	표준화 공식을 통해 표준정규분포로 변환하여야 확률을 계산할 수 있다.	1-1
유형 2	정규분포를 따르는 자료는 정규분포의 모든 성질을 공유한다.	1-2
유형 3	연속확률변수는 정규분포를 따른다.	2, 3
유형 4	자료의 수가 많으면 정규분포를 따른다.	2

10) 문항 1-1과 같은 유형의 문항은 미적분과 통계 기본 교과서 13종 모두에서 다루고 있는 전형적인 확률 문제이다.

11) 각 문항별 응답 결과의 세부 내용과 응답 인원수의 구체적인 내용은 <부록 2>을 참조하기 바란다.

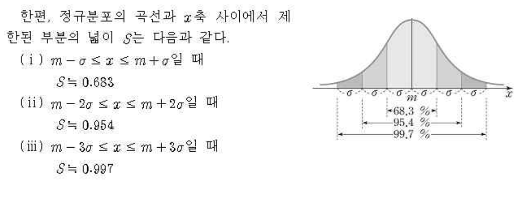


[그림 Ⅳ-1] 문항 1-1에 대한 응답의 예

유형 1. 표준화 공식을 통해 표준정규분포로 변환하여야 확률을 계산할 수 있다.[12)]

문항 1-1에서 81명의 학생이 $X \sim N(168.2,\ 7^2)$ 인 확률변수 X를 공식 $Z = \dfrac{X-m}{\sigma}$ 을 사용하여 표준정규분포를 따르는 확률변수 Z로 변환시켜 [그림 Ⅳ-1]과 같이 해결하였다.

문항 1-1에서 평균은 $m = 168.2$ 이고 표준편차가 $\sigma = 7$ 이므로 $P(161.2 \leq X \leq 175.2)$ 은 $P(m-\sigma \leq X \leq m+\sigma)$ 로 고칠 수 있기 때문에 정규분포 곡선의 성질에 의해 구하는 확률이 약 0.68이 됨을 곧바로 알 수 있음에도 불구하고 단지 2명의 학생만이 이 방법을 사용하여 문제를 해결하였다. 학생들은 정규분포를 표준정규분포로 변환하는 표준화 방법을 숙지하고 있고 이를 이용하여 문제를 해결할 수는 있지만, [그림 Ⅳ-2]와 같은 정규분포 곡선의 주요한 특징을 충분히 이해하고 있다고 보기 어렵다.


한편, 정규분포의 곡선과 x축 사이에서 제한된 부분의 넓이 S는 다음과 같다.

(i) $m-\sigma \leq x \leq m+\sigma$일 때
$S \fallingdotseq 0.683$

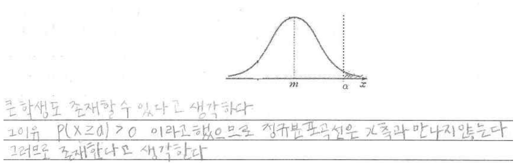
(ii) $m-2\sigma \leq x \leq m+2\sigma$일 때
$S \fallingdotseq 0.954$

(iii) $m-3\sigma \leq x \leq m+3\sigma$일 때
$S \fallingdotseq 0.997$

[그림 Ⅳ-2] 정규분포 곡선의 특징(우정호 외, 2009 : 201)

유형 2. 정규분포를 따르는 자료는 정규분포의 모든 성질을 공유한다.[13)]

문항 1-2에서 학생 45명은 키가 무한히 큰 사람이 존재할 수 있다고 주장하였다. 이들은 [그림 Ⅳ-3]과 같이 확률이 0보다 크거나 정규분포 그래프가 x축에 무한히 접근한다는 이론적인 사실에만 집중함으로써 잘못된 응답을 하였다. 이는 Huck, Cross, & Clark(1986)의 연구대상이 정규분포를 따르는 확률변수는 최대값과 최소값을 가질 수 없다고 응답한 사례와 같은 오개념에 해당한다고 볼 수 있다.



[그림 Ⅳ-3] 문항 1-2에 대한 응답의 예(1)

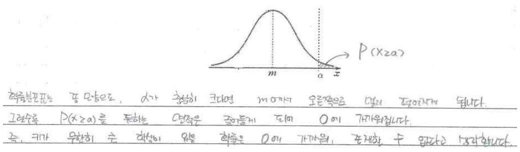
주어진 문제에 적절한 답을 제시한 학생 47명 중에서도 42명은, 이론적 모델인 정규분포와 실제 자료 사이의 관계를 직접적으로 고려하여 사람의 키가 정규분포를 따르기는 하지만 이는 근사적인 의미임을 명시적으로 설명하기 보다는 [그림 Ⅳ-4]와 같이 단지 구하는 확률이 거의 0에 가깝기 때문에 키가 무한히 큰 사람은 존재할 수 없다고 응답하였다.

---

12) 관점 5(정규분포와 관련된 전형적인 확률 문제를 어떻게 해결하는가?)에 따른 분석 결과

13) 관점 7(이론적 모델인 정규분포와 경험적 자료 사이의 관계를 어떻게 인식하는가?)에 따른 분석 결과

Garfield & Ben-Zvi(2010: 49)는 통계 교육에서 다루어져야 할 중요한 통계적 아이디어 9개에 ‘통계적 모델에 대한 이해’를 포함시켰다.

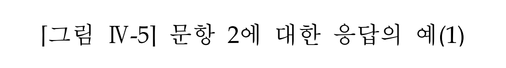


[그림 Ⅳ-4] 문항 1-2에 대한 응답의 예(2)

이들에 따르면 통계적 모델을 바람직하게 이해한 학생은 정규분포와 같은 모델을 문제 해결에 이용할 수 있어야 하고 이로부터 자료 값을 예측할 수 있을 뿐만 아니라, 모델이 실제 자료를 모델링하는데 적합한지를 평가하고 구체적 자료와 모델 사이의 경험적 차이를 인식할 수 있어야 한다. 이러한 맥락에서 볼 때, 정규분포에 대한 검사대상의 이해 정도가 충분히 깊다고 보기에는 어려운 점이 있다. 검사대상은 통계적 모델인 정규분포와 구체적인 자료 사이의 차이를 인식하지 못하고 정규분포를 따르는 자료는 정규분포가 갖는 모든 성질을 공유해야 한다고 생각하는 경향이 있었다.

유형 3. 연속확률변수는 정규분포를 따른다.[14)]

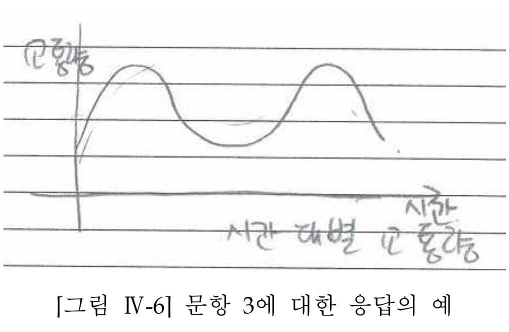
검사대상 중 20명은 문항 2에서 사람의 키와 같은 자료를 해석하는데 정규분포를 사용하는 이유를 ‘사람의 키가 연속이기 때문에’라고 응답하였다.



[그림 Ⅳ-5] 문항 2에 대한 응답의 예(1)

사람의 키가 연속이기 때문에 이를 해석할 때 정규분포를 사용한다는 학생들의 의견에는 연속확률변수는 항상 정규분포를 따른다는 암묵적인 이해가 담겨 있다. 연속확률변수와 정규분포사이의 관계에 대한 학생들이 이러한 이해는 연속확률분포로서 정규분포만을 다루고 정규분포이외의 연속확률분포에 대한 안내가 거의 없는 현행 교육과정의 지도 방식과 무관하지 않는 것으로 보인다. Gu, Huang, & Marton(2004)는 수학적 개념의 본질이 학생들에게 분명히 이해되기 위해서는 개념적 변이(concept variation)로서 개념을 담고 있는 예를 다양화시킴과 동시에 비개념 변이(non-concept variation)인 반례를 통해 개념의 의미가 혼동되는 예도 폭넓게 다룰 필요가 있다고 지적한 바 있다. 정규분포를 따르는 자료의 예만을 제시하는 것은 정규분포라는 개념을 학생들이 분명하게 이해하게 하는데 적절한 접근 방식이 아닐 수도 있다.

그러나 정규분포의 개념을 분명히 하기 위하여 정규분포를 따르지 않는 자료의 예를 다루는 것이 학생들에게 새로운 인지적 부담을 주게 될 것이라는 우려도 있을 수 있다. 실제로 정규분포를 따르지 않는 자료의 예를 묻는 문항 3에서 대부분의 학생들이 부적절한 예를 들거나 응답 자체를 하지 못하였다. 그러나 검사대상 중 22명의 학생은 정규분포를 따르지 않는 자료의 예를 [그림 Ⅳ-6]과 같이 바르게 설명하기도 하였다.

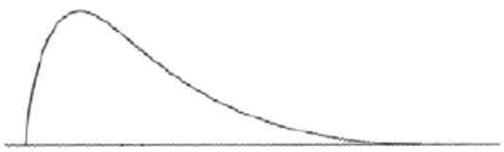


[그림 Ⅳ-6] 문항 3에 대한 응답의 예

14) 관점 6(정규분포에 대한 개념적 이해와 관련된 질문에 어떻게 답하는가?)와 관점 8(정규분포를 따르지 않는 자료의 예를 들 때 어떤 특징이 있는가?)에 따른 분석 결과

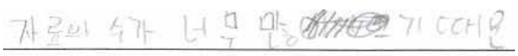
이는 정규분포를 따르지 않는 자료의 예를 전혀 다루지 않는 현행 교육과정의 지도 방식에 비추어 볼 때, 상당히 긍정적인 결과로 볼 수 있다. 정규분포를 따르지 않는 자료의 예로 균등분포나 치우친 분포의 개념을 본격적으로 도입하지 않으면서도 다음과 같이 간단한 설명을 제시하는 것이 가능할 수 있다.

> …연속확률변수 중에는 이상과 같이 정규분포를 따르는 것도 있지만 그렇지 않은 예도 있다. 이를테면 어느 학교의 수학성적의 분포는 다음과 같이 한쪽으로 치우친 분포가 될 수 있다.



유형 4. 자료의 수가 많으면 정규분포를 따른다.[15)]

문항 2에서 25명의 학생은 사람의 키가 정규분포를 따르는 이유는 사람의 수가 매우 많기 때문이라고 설명하였다.



[그림 Ⅳ-7] 문항 2에 대한 응답의 예(2)

이와 같이 응답한 학생들은 자료의 수가 충분히 크면 항상 정규분포를 따르는 것으로 이해하고 있다고 볼 수 있다. 선행 연구(고은성 · 이경화, 2011; Keith & Bower, 2003)는 정규분포에 대한 이러한 오개념이 예비교사나 전문가들에게도 나타나며 이는 자료의 수가 충분히 크면 그 자료의 '평균의 분포'가 근사적으로 정규분포에 가까워진다는 중심극한정리를 부적절하게 해석한 결과에 기인한 것이라고 설명하면서 정규분포 개념 지도와 관련하여 이러한 오개념이 발생하지 않도록 교수학적 경각심을 가질 필요가 있다고 지적하였다. 이러한 맥락에 비추어 볼 때, '자연현상이나 사회현상에서 관측하는 자료의 수를 한없이 크게 하면 좌우 대칭인 종모양의 곡선에 가까워진다'와 같은 미적분과 통계기본 교과서의 설명에 대한 재검토가 필요하다.

Ⅳ. 결론

이 연구는 정규분포에 대한 교수학적 변환 방식과 학생들의 이해의 특징을 분석함으로써 고등학교 통계 단원에서 정규분포를 지도하는 것과 관련된 시사점을 얻는데 목적이 있다. 이를 위해 정규분포의 역사 발생 과정과 학문적 지식으로서 정규분포의 의미를 확인하여 정규분포의 교수학적 변환 방식을 분석하기 위한 4개의 관점을 추출하였으며, 정규분포에 대한 학생들의 이해 특징과 관련된 선행 연구 결과를 살펴봄으로써 고등학생들의 정규분포에 대한 이해의 특징을 기술하는데 필요한 분석 관점을 4가지로 구체화한다.

이상과 같은 교수학적 변환 방식의 분석 관점을 토대로 미국, 영국, 우리나라 교과서에서 정규분포에 대한 교수학적 변환 방식을 살펴보았다. 우리나라 교과서의 경우 정규분포를 모집단의 자료 패턴을 나타내는 분포로 소개하려는 의도 하에 관련된 실생활 소재로 탐구 활동을 구성하고 있으나, 실제 정규분포를 정의하는 상황에서는 '확률밀도함수가 $f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 와 같은 분포'라고 이론적인 맥락으로 설명한다. 또

15) 관점 6(정규분포에 대한 개념적 이해와 관련된 질문에 어떻게 답하는가?)에 따른 분석 결과

한 정규분포의 특징을 기술하는 과정에서는 정규분포의 확률밀도함수에 대한 함숫값이 확률을 의미하는 것처럼 오해할 여지가 있는 그림을 제시하기도 한다.

한편, 정규분포에 대한 학생들의 이해의 특징을 그 분석 관점에 기초하여 살펴본 바에 따르면, 학생들은 정규분포와 관련된 확률 문제를 해결할 때, 표준화 공식을 활용한 표준정규분포로의 변환을 주로 이용한다는 사실을 알 수 있었다. 또한 이론적인 모델로서 정규분포와 경험적 자료의 차이를 인식하지 못하고 정규분포를 따르는 모든 자료는 정규분포가 지니는 이론적 특징을 모두 공유하여야 한다고 생각하는 경향이 있었다. 더불어, 연속확률변수는 항상 정규분포를 따른다거나 자료의 수가 많으면 정규분포를 따른다는 오개념을 지니고 있었다.

참고문헌

고승곤 · 양완연 · 오현숙(2003). **일반통계학**. 서울: 교우사.

고은성 · 이경화(2011). 예비교사들의 통계적 표집에 대한 이해. **수학교육학연구, 21**(1), 17-32.

교육인적자원부(2007). **수학과 교육과정**. 서울: (주) 대한교과서.

교육과학기술부(2008). **고등학교 교육과정 해설 5 수학**. 서울: (주) 미래엔 컬처그룹.

구자흥(1999). 가우스의 오차론에 근거한 정규분포 배경의 역사적 고찰. **한국수학사학회지, 12**(1), 1-12.

김상익 외(2005). **통계학의 이해와 적용**. 서울: 민영사.

김우철 외(2006). **현대통계학**. 서울: 영지문화사.

김해경 외(2009). **미적분과 통계 기본**. 서울: 더텍스트.

남주현(2006). **초 · 중등 통계교육을 위한 통계적 방법론에 대한 연구**. 이화여자대학교대학원 박사학위 논문.

박경미 외(2002). 한국, 일본과 미국, 영국의 수학 교과서 비교. **학교수학, 4**(2), 317-331.

박영희(2002). 연속확률변수 개념의 직관적 이해에 관한 고찰. **학교수학, 4**(4), 677-688.

박정식 · 윤영선(2004). **현대통계학**. 서울: 다산출판사.

우정호(2008). **학교수학의 교육적 기초**. 서울 : 서울대학교출판문화원.

우정호 외(2009). **미적분과 통계 기본**. 서울: 두산동아.

유희찬 외(2009). **미적분과 통계 기본**. 서울: 미래엔 컬처그룹.

임지현(2008). **분포개념의 지도에 관한 고찰**. 서울대학교 교육학석사학위 논문.

장대흥(1999). 이항분포의 정규근사에 대한 고찰. **응용통계연구, 12**(2), 671-681.

조동기(2007). **표본평균분포 이해에 관한 연구**. 한국교원대학교대학원 석사학위 논문.

조재근(2008). 19세기 중반 오차와 정규분포의 역사. **한국통계학회 논문집, 15**(5), 737-752.

조정수(2010). CAS 계산기를 활용한 고등학교 정규분포곡선의 교수-학습을 위한 시사점 탐구. **수학교육논문집, 24**(1), 177-193.

허명회(2007). 고등학교 수학Ⅰ 통계에 대한 고찰. **응용통계연구, 20**(1), 159-165.

Batanero, C., Tauber, L., & Meyer, R. (1999). From data analysis to inference: A research project on students' understanding of the normal distribution. *International Statistical Institute, 52nd Session*, Finland.

Batanero, C., Tauber, M., & Sánchez, V. (2004). Students' reasoning about the normal distribution, In D. Ben-Zvi and J. Garfield

(Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 257-276). Dordrecht: Kluwer Academic Press.

Cobb, G. W. & Moore, D. S. (1997). Mathematics, statistics and teaching, *The American Mathematical Monthly, Nov.,* 801-823.

Garfield, B. J. & Ben-Zvi, D. (2010). *Developing Students' Statistical Reasoning*. USA: Springer.

Gu, L., Huang, R., & Marton, F. (2004). Teaching with variation: A chinese way of promoting effective mathematics learning. In L. Fan, N. Y. Wong, J. Cai, & S. Li (Eds.), *How Chinese Learn Mathematics: Perspective from Insiders* (pp. 309-347). Singapore: World Scientific.

Hald, A. (1990). *A history of probability and statistics and their applications before 1750*, NY: John Wiley & Sons, Inc. John Wiley & Sons, Inc.

Hogg, V. R., & Craig, T. A. (2002). **수리통계학개론**(이재창·이용구 역). 서울: 경문사 (영어 원작은 1996년 출판)

Hogg, V. R., & Tanis, A. E. (2005). *Probability and statistical inference*. USA: Prentice Hall.

Huck, S., Cross, T. L., & Clark, S. B. (1986). Overcoming misconceptions about z-scores. *Teaching Statistics, 8*(2), 38-40.

Huck, W. S. (2009). *Statistical Misconceptions*. NY: Routledge.

Keith, M. & Bower, M. S. (2003). *Some misconceptions about the normal distribution*. American Society for Quality: Six Sigma Forum.

Moore, S. D. (2001). *Statistics : Concepts and controversies*. USA: W. H. Freeman and company.

Pfannkuch, M. & Wild, C. (2004). Towards an understanding of statistical thinking, In D. Ben-Zvi and J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 257-276). Dordrecht: Kluwer Academic Press.

SMP(2002). *Statistics 1: Foundations of sampling and estimation*. UK: Cambridge University Press.

Stigler, S. (2005). **통계학의 역사**(조재근 역). 서울: 한길사 (영어 원작은 1986년 출판).

Watkins, E. A., Scheaffer, L. R., & Cobb, W. G. (2004). *Statistics in Action*. USA: Key Curriculum Press.

Wilensky, U. (1997). What is normal anyway? Therapy for epistemological anxiety, *Educational Studies in Mathematics*, 33, 171-202.

A Study on a Didactic Transposition Method and Students' Understanding for the Normal Distribution

Shin, Bo Mi (Chonnam National University)

The goal of this study is to investigate a didactic transposition method of text books and high school students' understanding for the Normal Distribution. To accomplish this goal, framework descriptors were developed to analyse the didactic transposition method and interpret the students' understanding based on the Historico-Genetic process of the Normal Distribution, the meaning of the Normal Distribution as a scholarly knowledge and the results of previous studies on students' understanding for the Normal Distribution. This study presented several recommendations for instruction of the Normal Distribution according to the results of analysing the didactic transposition method and interpreting the students' understanding in terms of the developed framework.

* Key Words : Normal distribution(정규분포), didactic transposition(교수학적 변환), statistical education(통계 교육), students' understanding for the Normal Distribution (정규분포에 대한 학생들의 이해).

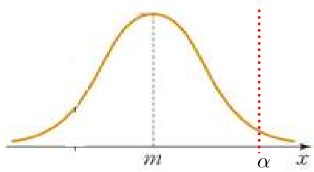
논문접수 : 2012. 3. 27
논문수정 : 2012. 4. 23
심사완료 : 2012. 5. 10

<부록 1> 지필검사 문항

1. 윤호네 고등학교 전체 학생의 키는 평균이 168.2 cm이고 표준편차가 7 cm인 정규분포를 따른다고 한다.

(1-1) 윤호네 고등학교에서 한 명을 택할 때, 이 학생의 키가 161.2 cm 이상 175.2 cm 이하일 확률을 구하여라.

(1-2) 윤호네 고등학교 학생의 키를 확률변수 X라 하고, 다음 그림을 윤호네 고등학교 학생의 키 X의 분포라고 할 때, 충분히 큰 α에 대하여 $P(X \geqq \alpha) > 0$ 이다. 이는 윤호네 고등학교 학생 중에 키가 무한히 큰 학생도 존재할 수 있음을 의미하는가? 자신의 생각과 그렇게 생각하는 이유를 쓰시오.



2. 사람의 키는 정규분포를 따른다고 알려져 있다. 사람의 키가 무작위로 분포되지 않고 일정한 패턴인 정규분포를 따르는 이유가 무엇이라고 생각하는가?

3. 정규분포를 따르지 않는 자료의 예를 2가지 이상 들어보시오.

<부록 2> 지필검사 결과

<table>
<tr><th>문항
번호</th><th colspan="2">응답 결과</th><th>반응수</th></tr>
<tr><td rowspan="3">1-1</td><td colspan="2">표준화 공식인 $\dfrac{X-m}{\sigma}$ 을 사용</td><td>81</td></tr>
<tr><td colspan="2">$X\sim N(m,\ \sigma^2)$ 일 때, $P(m-\sigma \leqq X \leqq m+\sigma)=0.6826$ 을 이용</td><td>2</td></tr>
<tr><td colspan="2">무응답</td><td>9</td></tr>
<tr><td rowspan="8">1-2</td><td rowspan="3">Ⅰ.
존재할 수 있다</td><td>Ⅰ-1. 확률이 0이 아니므로</td><td>29</td></tr>
<tr><td>Ⅰ-2. 정규분포 그래프는 x축에 붙어 있지 않기 때문에</td><td>8</td></tr>
<tr><td>Ⅰ-3. 확률이 0보다 크므로 학생 수가 무한히 많다면 가능</td><td>8</td></tr>
<tr><td rowspan="5">Ⅱ.
존재할 수 없다</td><td>Ⅱ-1. 정규분포 그래프도 언젠가는 x 축에 닿을 것이므로</td><td>2</td></tr>
<tr><td>Ⅱ-2. 확률이 0에 가깝기 때문</td><td>29</td></tr>
<tr><td>Ⅱ-3. 확률이 0에 가깝고 학생 수가 무한이 아니기 때문에</td><td>11</td></tr>
<tr><td>Ⅱ-4. 확률분포는 확률끼리의 관계로 확률을 예측하기 위한 것이지 실제로 그런 값이 존재하는 것을 의미하지는 않는다.</td><td>4</td></tr>
<tr><td>Ⅱ-5. 사람은 무한히 클 수 없다.</td><td>1</td></tr>
<tr><td rowspan="6">2</td><td colspan="2">실제 사람들의 키는 평균 근처에 몰려 있고 평균과의 차이가 많이 나는 자람의 수는 점점 작아지므로</td><td>7</td></tr>
<tr><td colspan="2">사람의 키는 연속이니까</td><td>20</td></tr>
<tr><td colspan="2">사람의 수는 매우 많기 때문에</td><td>25</td></tr>
<tr><td colspan="2">사람의 키에는 평균이 있으므로</td><td>5</td></tr>
<tr><td colspan="2">사람의 키는 무작위 확률로 결정되므로</td><td>6</td></tr>
<tr><td colspan="2">무응답</td><td>29</td></tr>
<tr><td rowspan="3">3</td><td colspan="2">적절한 예</td><td>22</td></tr>
<tr><td colspan="2">부적절한 예</td><td>39</td></tr>
<tr><td colspan="2">무응답</td><td>31</td></tr>
</table>