

CAS 계산기를 활용한 고등학교 정규분포곡선의 교수-학습을 위한 시사점 탐구

조정수 (영남대학교)

본 연구는 고등학교 통계 영역의 확률분포에 제시되어 있는 정규분포를 이항분포에서 정규분포로의 근사, 정규분포곡선의 탐구, Monte Carlo 방법에 의한 정규분포곡선의 넓이 탐구, 정규분포곡선의 선형변환, 그리고 여러 형태의 정규분포곡선 탐구 등의 내용을 중심으로 CAS 계산기를 활용하여 탐구해보고자 한다. CAS 계산기의 도구적 기능인 사소화, 실험, 시각화, 집중의 측면에서 볼 때 지필로서는 교육과정에 제시된 확률분포의 목표를 달성하기 불가능하다고 판단된다. 따라서 본 연구에서는 CAS 계산기를 활용하여 정규분포곡선의 다양한 성질을 탐구하고 이러한 과정과 결과로부터 정규분포곡선에 대한 교수학적 시사점을 도출하고자 한다.

I. 서 론

고등학교 통계 영역의 정규분포에 제시되어 있는 정규분포의 확률밀도함수

$$f(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-m}{s})^2}$$
(상수 $m, s (s \neq 0)$)는 어떤 함수로부터 유도된 것일까? 이 함수는 어떤 역사발생적 배경을 가지고 있는가? 만약 이 함수가 $f(x) = e^{-x^2}$ 로부터 유도된 것이라면 $\sqrt{2\pi}$ 와 $\frac{1}{2}$, 그리고 상수 $m, s (s \neq 0)$ 은 어떤 역할을 하는 것일까? 정규분포곡선의 넓이는 어떤 과정과 절차를 거쳐 구해지는가? 본 연구는 정규분포와 관련된 이러한 의문을 해결하고자 하는 목적으로 수행되었다.

개정된 2007년 수학과 교육과정에 따르면 통계 영역의 확률분포는 ‘미적분과 통계 기본’과 ‘적분과 통계’에서 다루도록 하고 있다(교육과학기술부, 2008). ‘미적분과 통계 기본’과 ‘적분과 통계’에는 “정규분포의 뜻과 그 성질을 이해하게 한다, 표준정규분포와 표준정규분포를 따르는 확률변수를 알고 이를 활용할 수 있게 한다, [그리고] 이항분포와 정규분포 사이의 관계를 이해하게 한다”(p. 202, 284)라고 제시되어 있다. 하지만 이들 수학과 선택 과목의 교육과정을 보면 여전히 과거와 마찬가지로

* 접수일(2010년 1월 15일), 심사(수정)일(1차: 2010년 1월 25일), 게재확정일자(2010년 2월 9일)

* ZDM 분류 : K64

* MSC2000 분류 : 97D40

* 주제어 : CAS 계산기, 정규분포, 정규분포곡선

지필환경에 크게 의존함으로써 확률분포에 대한 성질과 관계를 탐구를 통해 이해하게 하는 것이 아니라 확립된 사실로서 전달하는데 초점을 두고 있다. 지필환경에서 확률분포의 다양한 성질과 관계를 파악한다는 것은 사실상 불가능하므로 이를 손쉽고 깊이 있게 탐구하여 그 성질과 관계를 이해하기 위해서는 공학의 효율적 활용이 적극적으로 필요하다. 본 연구에서는 CAS 계산기 공학의 측면에서 이러한 문제점을 해소해보고자 한다.

컴퓨터 대수시스템인 CAS(Computer Algebra Systems) 기능을 가진 CAS 계산기는 학교수학의 교수-학습에 미치는 긍정적인 영향과 수학과 교육과정의 강조점과 구조의 변화, 그리고 교수-학습 방법상의 크고 다양한 변화가 있음이 이미 밝혀졌다(Heid, 2003; Lagrange, 2005; Tall, David, Cynthia, 2008). 특히, Kutzler는 수학의 교수-학습을 위한 교수학적 도구로서의 CAS를 말하면서 사소화(trivialization), 실험(experimentation), 시각화(visualization), 집중(concentration)의 네 가지 도구적 역할을 있다고 주장했다(Kutzler, 2003).

정규분포의 다양한 성질 탐구를 위하여 본 연구에서 CAS 계산기를 활용한 이유는 여기에서 다루는 내용인 이항분포에서 정규분포로의 근사, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 의 곡선 탐구, Monte Carlo 방법에 의한 정규분포곡선의 넓이 탐구, 정규분포 함수의 선형변환, 그리고 여러 형태의 정규분포곡선 탐구 등의 내용이 CAS 계산기의 도구적 기능인 사소화, 실험, 시각화, 집중의 측면에서 볼 때 지필로서는 교육과정에 제시된 목표를 달성하기 불가능하다고 판단되었기 때문이다. 따라서 본 연구에서는 CAS 계산기를 활용하여 정규분포곡선의 다양한 성질을 탐구하고 이러한 과정과 결과로부터 정규분포에 대한 교수학적 시사점을 도출하고자 한다.

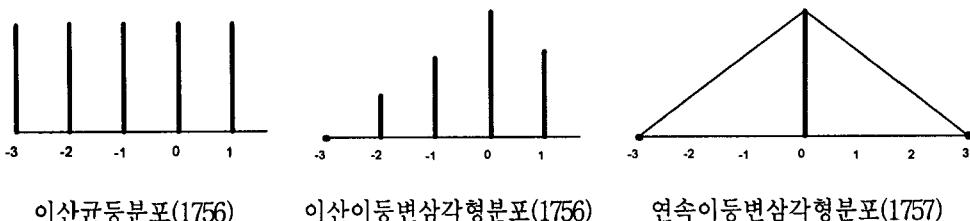
II. 정규분포곡선의 배경 지식

1. 정규분포곡선의 역사발생적 배경

분포 개념은 18세기에 에러이론(theory of errors)에서 출현하였는데, 이 시기는 에러를 줄이는 방법으로 산술평균의 사용이 타당한지에 대한 논쟁이 한 장이었다. 이때는 개별 에러를 판단하는 것이 불가능했기 때문에 에러들 사이의 관계를 파악하고자 했다. 이러한 전환을 실행에 옮기 사람이 Simpson이었다. 1756년 Simpson은 여러 번의 관찰에 의해 구한 관찰값들의 평균이 한 번의 관찰값 보다 더 타당하다는 주장을 하기 위해 단순확률함수를 사용했는데, 이 때 이들 관찰값의 에러들 사이의 관계를 파악했다(Bakker, 2004).

Simpson이 제안한 최초의 에러분포(distribution of errors)는 모든 값에 대해 동일한 확률을 가지는 이산 균등 분포(discrete uniform distribution)였다. 그 뒤 Simpson은 비율적으로 확률을 가지는 이산 이등변삼각형 분포(discrete isosceles triangle distribution)를 추정했다. 1년 뒤에는 연속 이등변

삼각형 분포를 제안했다. 다음 그림은 Simpson이 제안한 이들분포의 모양이다.



<그림 II-1> Simpson이 제안한 에러분포

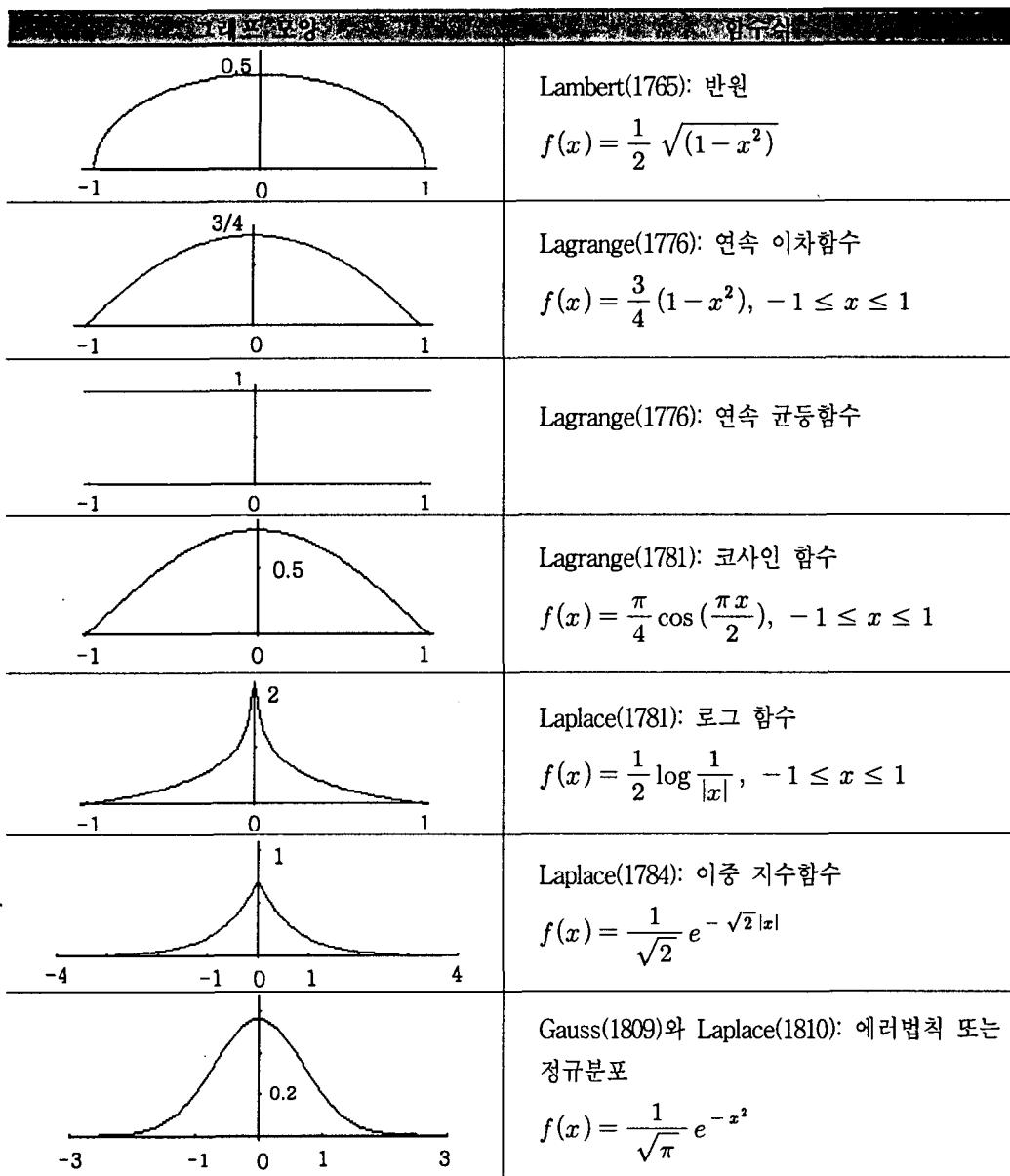
Simpson이 확률분포의 아이디어를 발표한 후 여러 수학자가 대안적 에러법칙(laws of errors)을 제시하였는데, 이들 중에는 Lagrange, Lambert, Daniel Bernoulli, Laplace, Gauss 등이 있었다. 이들은 Simpson과는 달리 현대적인 함수 형태로 확률분포를 제시하였다. 예를 들어, Lambert는 에러분포를 함수로 나타내지 않은 채 모수에 대해 관측 자료에서 발생 확률(가능성)이 가장 높은 추정을 말하는 최우법¹⁾(最尤法, method of maximum likelihood)을 발표했다.

수학적으로 볼 때 정규분포는 여러 방법으로 표현할 수 있다. 역사적으로 최초의 방법이며 여전히 아주 보편적인 방법은 이항분포 $bin(n, p)$, $n \rightarrow \infty$ 의 극한으로부터 정규분포를 구하는 것이다. 이것은 De Moivre-Laplace의 극한정리(limit theorem)의 결과인데, 이 정리는 많은 수의 독립 확률변수의 합은 근사적으로 정규분포에 가깝다는 중심극한정리(central limit theorem)의 특수한 경우이다. 예를 들어, 사람들의 키 데이터의 경우 부모의 키, 식생활 습관, 운동량 등의 여러 가지 요인이 키에 영향을 준다. 이들 요인은 정규분포를 따르지 않더라도 이들의 합은 정규분포에 근사한다. 많은 현상들이 정규분포로 설명되는 이유가 바로 이 때문이다(Wilensky, 1997).

정규분포와 정규분포곡선은 에러법칙, 빈도법칙, Gaussian 곡선, Laplace-Gauss 곡선 등의 여러 이름으로 불린다(Stigler, 1999). 여기에 언급되지 않은 명칭으로 Freudenthal은 'De Moivre 분포'라고 불렀는데, 그 이유는 De Moivre가 처음으로 이 함수를 정의했기 때문이다. 정규분포곡선에 대한 Gauss의 연구 결과는 천문학을 비롯하여 사회과학, 인류학 등의 연구에 사용되기 시작했다. 다음 그림은 정규분포곡선의 역사적 발달 과정을 그래프와 함수식 관점에서 살펴본 것이다.

정규분포곡선의 특징 중의 하나는 대칭성이다. 하지만 Bessel, Fechner, Edgeworth 등은 이 대칭성 가정을 거부했다. 오래 전부터 에러분포는 대칭성을 가지는 것으로 추정되었다. 1838년 Bessel은 최초로 이 대칭성 가정에 의문을 가졌다(ESS, 1981). 그리고 그 당시 대부분의 학자들과는 달리 1874년 Fechner는 심지어 대부분의 데이터 분포는 비대칭적이라고 추정했다.

1) 최우법은 최대 가능성(최대 우도)를 나타내는 것으로 최우 추정법(maximum likelihood estimation)이라고도 한다.



<그림 II-2> 정규분포곡선의 역사적 발달 과정(Bakker, 2004)

2. 정규분포곡선에 대한 교과서 분석

정규분포곡선의 도입 방식을 살펴보기 위하여 대표적으로 두 출판사의 고등학교 수학 교과서를 분석하였다. <그림 II-3>의 교과서(임석훈, 2002) 발췌 내용의 경우 정규분포와 정규분포곡선의 용

어, 그리고 함수식, 정규분포 기호 사용법을 제시하고 있다. 그리고 정규분포곡선의 모양이 m 과 σ 에 따라 어떻게 달라지는지를 그림으로 제시하고 있지만, σ 가 변할 때를 보면 $x=0$ 에서 곡선이 대칭을 이루고 있어 이 곡선이 표준정규분포도 아니면서 이 점이 평균으로 되어 있는 오류가 있다. 또 m 이 변할 때를 보면 평균인 m 이 0, 1, 4 등으로 되어 있어 표준정규분포도 아니면서 평균이 0이 되는 오류가 있다. 그리고 이런 제시와는 아무 연관 없이 박스 안에 정규분포곡선의 성질을 요약해놓고 있다.

▶ 정규분포에 대하여 알아보자.

연속확률변수 X 의 확률밀도함수 $f(x)$ 가 다음 식과 같을 때, X 의 확률분포를 정규분포라고 하고, 함수 $f(x)$ 의 그래프를 정규분포곡선이라고 한다.

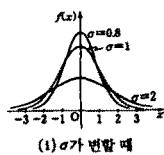
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} (-\infty < x < \infty)$$

여기서 m 과 $\sigma(\sigma>0)$ 은 각각 평균과 표준편차를 나타내는 상수이며, e 는 그 값이 2.71828...인 무리수이다.

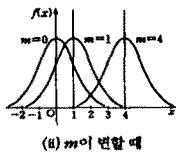
또, 평균이 m , 표준편차가 σ 인 정규분포를 기호로 다음과 같이 나타낸다.

$$N(m, \sigma^2)$$

한편, 정규분포곡선은 m 과 σ 의 값에 따라 다음과 같이 그 모양이 정해진다.



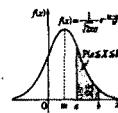
(a) σ 가 변할 때



(b) m 이 변할 때

<그림 II-3> 정규분포곡선의 도입 방법

확률변수 X 가 정규분포를 따를 때, 정규분포곡선과 x 축 사이의 넓이는 1이다. 또, X 의 값이 구간 $[a, b]$ 에 있을 확률 $P(a \leq X \leq b)$ 는 오른쪽 그림에서 색칠한 부분의 넓이이다.



일반적으로 정규분포곡선은 다음과 같은 성질을 가진다는 것이 알려져 있다.

정규분포곡선의 성질

- (1) 직선 $x=m$ 에 대하여 대칭인 평균의 곡선이고 평균선은 x 축이다.
- (2) 곡선과 x 축 사이의 넓이는 1이다.
- (3) $x=m$ 일 때, 최댓값 $\frac{1}{\sqrt{2\pi}\sigma}$ 을 가지다.
- (4) m 이 일정할 때, 표준편차 σ 가 커지면 곡선은 양쪽으로 퍼지고 σ 가 작아지면 곡선은 좁아진다.
- (5) σ 가 일정할 때, m 이 변하면 대칭축의 위치는 바뀌지만 곡선의 모양은 같다.

또 다른 교과서(임재훈 외, 2002)에서의 정규분포곡선 도입 방식을 살펴보면 위의 교과서와 거의 동일함을 <그림 II-4>를 통해 알 수 있다. 한 가지 차이점은 정규분포곡선의 모양이 m 과 σ 에 따라 어떻게 달라지는지를 그림으로 제시하면서 앞에서 지적했던 오류는 보이지 않지만 정규분포곡선의 y 축이 확률밀도함수 $f(x)$ 임을 빠뜨리고 있다.

위의 <그림 2>와 같은 곡선 가운데 연속확률변수 X 의 확률밀도함수 $f(x)$ 가

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

으로 주어지는 경우가 있다. 이러한 연속확률변수 X 의 확률분포를 정규분포라고 한다.

여기서 m 과 $\sigma(\sigma>0)$ 은 상수이고, e 는 2.71828...인 무리수이다.

이 때, 연속확률변수 X 의 평균과 표준편차는 각각 m , σ 임이 알려져 있다.

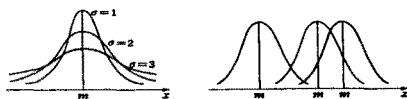
이와 같이 평균이 m , 표준편차가 σ 인 정규분포를 기호로

$$N(m, \sigma^2)$$

과 같이 나타낸다.

또, 확률밀도함수 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$ 의 그래프를 정규분포곡선이라고 한다.

정규분포곡선은 m 과 σ 의 값에 따라 다음과 같이 그 모양이 결정된다.



정규분포곡선 $y=f(x)$ 에는 다음과 같은 성질이 있다.

정규분포곡선의 성질

- ① 곡선과 x 축 사이의 넓이는 1이다.
- ② 직선 $x=m$ 에 대하여 대칭이다.
- ③ $x=m$ 일 때 최댓값 $\frac{1}{\sqrt{2\pi}\sigma}$ 을 가지며, x 축이 평균선이다.
- ④ 평균 m 이 일정할 때, 표준편차 σ 의 값이 커지면 곡선이 낮아지면서 옆으로 퍼지고, 표준편차 σ 의 값이 작아지면 곡선의 중앙 부분이 높아진다.
- ⑤ 표준편차 σ 가 일정할 때, 평균 m 이 변하면 대칭축의 위치는 변화지만 곡선의 모양은 일정하다.

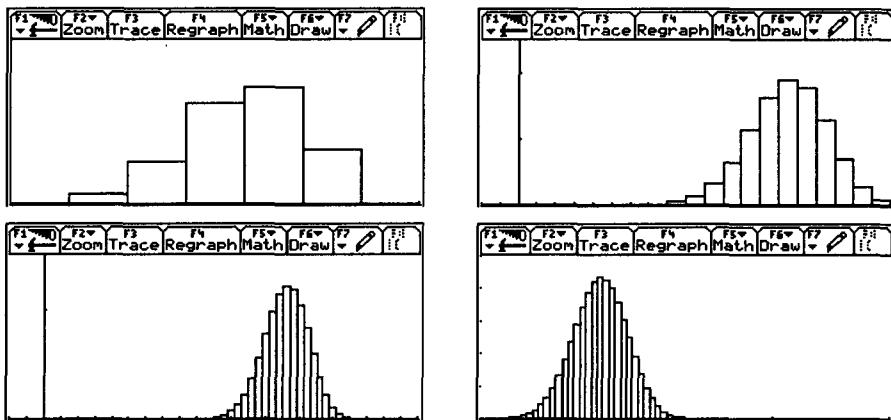
<그림 II-4> 정규분포곡선의 도입 방법

이 두 교과서를 보면 확률밀도함수의 모양과 정규분포곡선의 성질을 어떤 과정을 통해 얻게 되었는지 전혀 알 수 없으며, 확률밀도함수의 m 과 σ 의 다양한 값에 의한 그래프의 모양 변화와 그로부터의 성질 도출은 불가능함을 알 수 있다.

III. 정규분포곡선의 성질 탐구

1. 이항분포의 정규분포로의 근사

n 회의 독립시행에서 어떤 사건이 일어나는 횟수를 X 라 하면 X 는 $0, 1, 2, \dots, n$ 의 값을 갖는 확률변수이고 X 가 k 일 확률은 $P(X=k) = {}_n C_k p^k q^{n-k}$ 인 이항분포를 따른다. 여기서 성공확률 p 를 고정시켰을 때, 시행횟수 n 을 증가시키면 이항(확률)분포의 그래프는 종모양 곡선에 가깝게 된다. 예를 들어, 아래의 그림은 $p=0.7$ 일 때 이항분포의 시행횟수 $n=5, 20, 50, 100$ 일 때의 분포 모양의 변화를 CAS 계산기의 통계 프로그램인 List Editor로 실험한 것이다.



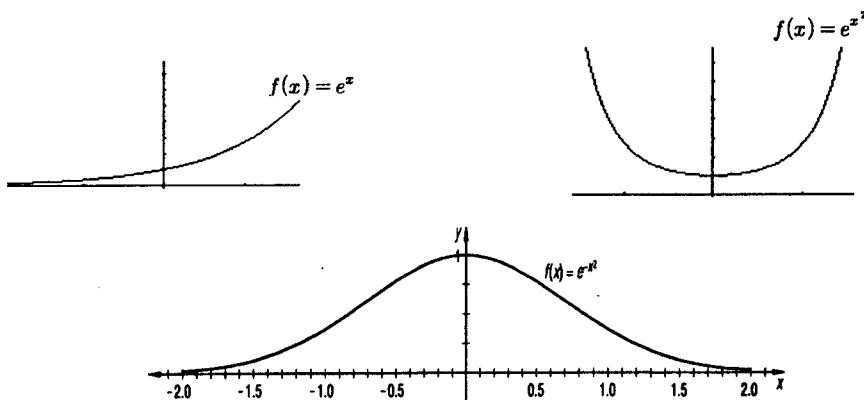
<그림 III-1> 이항분포의 정규분포로의 근사

<그림 III-1>을 보면 이항분포는 n 이 커짐에 따라 정규분포로 근사한다는 사실을 실험을 통한 결과의 시작화로 알 수 있다. 이 예로부터 CAS 계산기를 사용함으로써 수학교실에서 특히 지필로 다루기 어려운 이러한 분포나 함수의 근사 성질을 실험으로 확인할 수 있는 것은 탐구 활동을 증대시키고 수학적 개념에 대한 깊은 이해와 개념들 사이의 연관성을 인식시킬 수 있다.

2. 정규분포곡선의 원형인 $f(x) = e^{-x^2}$ 의 모양

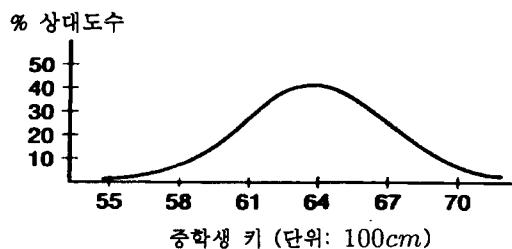
정규분포의 그래프를 정규곡선(normal curve)이라 하며 앞에서도 언급했듯이 19세기 초 Gauss와

Laplace가 처음으로 연구했다. 정규곡선의 원형은 $f(x) = e^{-x^2}$ 이며, $f(x) = e^x$ 와 $f(x) = e^{x^2}$ 으로 부터 유도될 수 있다. 본 연구에서는 고등학교 수학과 교육과정과의 일치를 위하여 정규곡선을 ‘정규분포곡선’이라고 하고, 함수 $f(x) = e^{-x^2}$ 를 ‘정규분포곡선의 원형’이라고 부르기로 한다.



<그림 III-2> 정규분포곡선의 원형 모양

많은 자연 현상의 데이터가 정규분포를 따른다. 예를 들어, 어느 교육청 산하의 중학교에서 무작위로 1000명의 남학생을 뽑아 키를 조사하였는데, 이 학생들의 상대도수분포의 그래프는 다음과 같다.



<그림 III-3> 중학생 남학생의 키의 분포

이 중학교 남학생의 키의 분포 그래프는 정규분포곡선의 원형을 적절하게 선형변환함으로써 구할 수 있다. 정규분포곡선의 원형과 이 그래프로부터 선형변환된 그래프를 종종 종모양 곡선 (bell-shaped curve)이라고 한다.

3. 정규분포곡선의 원형의 성질 탐구

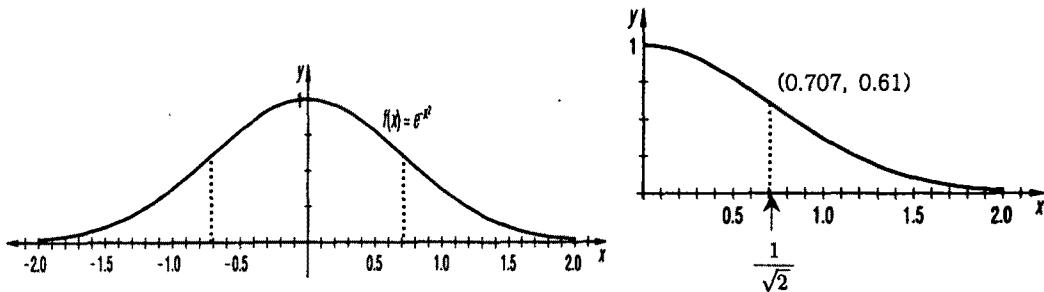
정규분포곡선의 원형 $f(x) = e^{-x^2}$ 의 그래프로부터 다음의 성질들을 알 수 있다.

첫째, $(0,1)$ 이 최댓값이며 y 절편이 1이다.

둘째, $f(-x) = e^{-(-x)^2} = e^{-x^2} = f(x)$ 이므로 y 축 대칭이다.

셋째, x 축이 점근선이다. $|x|$ 가 커지면 x^2 도 커진다. 그런데 e^{-x^2} 을 하면 $\frac{1}{e^{x^2}}$ 이므로 이 큰 값의 역수가 되어 x 의 값이 커질수록 $f(x) = e^{-x^2}$ 의 값은 점점 더 작아지게 된다.

넷째, 이 곡선의 곡률변화가 또 다른 성질이다. 이 그래프는 y 절편 근처에서는 곡선이 볼록하고, y 절편에서 멀어질수록 오목하다.



<그림 III-4> $f(x) = e^{-x^2}$ 의 변곡점의 x 좌표

이 그래프의 볼록과 오목이 변하는 변곡점은 위의 그림에서 보듯이 두 점이다. 이 변곡점을 구하기 위하여 $f''(x) = 0$ 인 점을 조사해 보면 $x = \pm \frac{1}{\sqrt{2}} \approx \pm 0.707$ 이며 이 두 점의 y 값을 구해보면 $y \approx 0.61$ 이다. 아래의 CAS 계산기의 결과를 보면 이 사실을 알 수 있다.

$\frac{d}{dx}(y_2(x))$	$-2.000 \cdot x \cdot e^{-x^2}$	$\text{solve}\left(4 \cdot x^2 - 2 \cdot e^{-x^2} = 0, x\right)$ $x = -.707 \text{ or } x = .707$
$\frac{d^2}{dx^2}(y_2(x))$	$(4.000 \cdot x^2 - 2.000) \cdot e^{-x^2}$	$\text{solve}\left(4 \cdot x^2 - 2 \cdot e^{-x^2} = 0, x\right)$ $x = \frac{-\sqrt{2}}{2} \text{ or } x = \frac{\sqrt{2}}{2}$

<그림 III-5> CAS 계산기에 의한 $f(x) = e^{-x^2}$ 의 변곡점의 x 좌표 계산 결과

(1) 적분법에 의한 $f(x) = e^{-x^2}$ 의 그래프와 x 축 사이의 넓이 계산

$f(x) = e^{-x^2}$ 의 그래프와 x 축 사이의 넓이는 $\int_{-3}^3 (e^{-x^2}) dx \approx 1.772414$ 이며 이 값은

$\sqrt{\pi} \approx 1.7725$ 와 같다. 그리고 $|x| > 3$ 인 경우 $f(x) = e^{-x^2}$ 아래에 추가되는 넓이는 거의 없다.

그 이유는 $\int_{-\infty}^{\infty} (e^{-x^2}) dx \approx 1.772453$ 이므로 $-3 \leq x \leq 3$ 의 범위에서 $f(x) = e^{-x^2}$ 넓이의 약

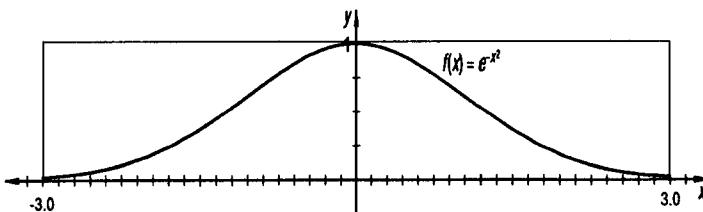
99.9% 를 차지하기 때문이다. 따라서 정규분포 곡선 아래의 넓이는 $\sqrt{\pi}$ 와 같다.

$\int_{-3}^3 (e^{-x^2}) dx$	1.7724147
$\int_{-\infty}^{\infty} (e^{-x^2}) dx$	1.7724539
$\int_{-\infty}^{\infty} (e^{-x^2}) dx$	1.7724146965167
$\int_{-\infty}^{\infty} (e^{-x^2}) dx$.99997791

<그림 III-6> CAS 계산기에 의한 $f(x) = e^{-x^2}$ 그래프와 x 축 사이의 넓이 계산 결과

(2) Monte Carlo 방법에 의한 $f(x) = e^{-x^2}$ 의 그래프와 x 축 사이의 넓이 계산

아래 그림과 같이 $x = -3$ 과 $x = 3$, $y = 0$ 와 $y = 1$ 사이에서 $f(x) = e^{-x^2}$ 아래의 넓이를 포함하도록 직사각형을 그린다. n 개의 점을 직사각형 안에 무작위로 뿌렸을 때, 정규분포 곡선 아래에 위치하는 점의 개수를 m 이라 하자. 그러면 이 곡선 아래의 넓이는 근사적으로 $\frac{m}{n} \times (\text{직사각형의 넓이})$ 가 된다.



<그림 III-7> Monte Carlo 방법을 적용하기 위한 직사각형 만들기

실제로 CAS 계산기를 이용하여 다음과 같이 시뮬레이션 프로그램을 작성하여 $n = 3000$ 개의 점을 이 직사각형 안에 무작위로 뿌리는 실험을 10번 실시했다(Kelly, 1997). 이 때 곡선 아래에 뿌려진 점의 개수는 908, 928, 914, 843, 891, 879, 885, 897, 900, 842개였다. 이 10개의 평균은 888.7이므로 $\frac{m}{n} = \frac{888.7}{3000} \approx 0.296$ 이다. 따라서 정규분포곡선 아래의 넓이는 $0.296 \times 6 \approx 1.7774$ 이며 이 값은 적분을 이용한 $\sqrt{\pi} \approx 1.7725$ 와 상당히 일치함을 알 수 있다.

```

F1 F2 F3 F4 F5 F6
Control I/O Var Find... Mode
:Monte()
:Prgm
:ClrIO
:Input "how many points",n
:0→m
:For i,1,n
:rand(→x
:rand(→y
:If y<e^(-x^2)
:m+1→m
:EndFor
:approx(m/n)→t:Disp m,t:EndPrgm

```

<그림 III-8> CAS 계산기의 프로그래밍 도구를 이용한 Monte Carlo 방법의 시뮬레이션 프로그램

IV. 정규분포곡선의 선형변환

정규분포곡선의 원형인 $f(x) = e^{-x^2}$ 를 선형변환시킴으로써 이 함수를 표준정규분포의 확률밀도 함수 $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 로 바꿀 수 있고 이로부터 표준정규분포곡선에 대한 중요한 성질을 추출 할 수 있다.

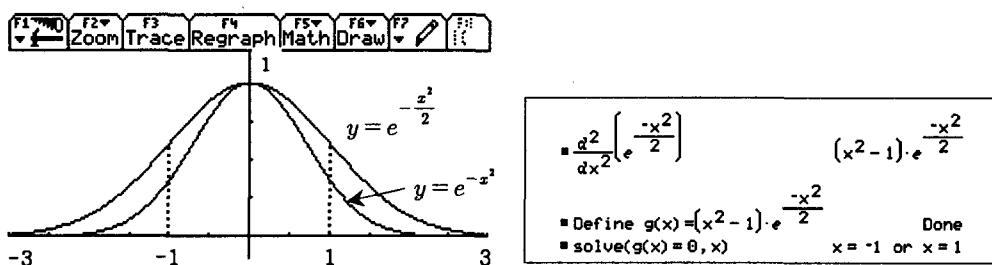
1. 정규분포곡선의 변곡점 $x = \pm \frac{1}{\sqrt{2}}$ 를 $x = \pm 1$ 로 변환

$f(x) = e^{-x^2}$ 의 변곡점 $\pm \frac{1}{\sqrt{2}}$ 을 1로 바꾸려면 x 좌표값을 $\sqrt{2} \times x$ 로 하면 된다. 이 변환을

식으로 나타내면 $S_1 : (x, y) \rightarrow (\sqrt{2}x, y)$ 이다. 따라서 이 변환을 적용하여 다음과 같다.

$$\begin{aligned} x' &= \sqrt{2}x \text{이므로 } x = \frac{1}{\sqrt{2}}x' \\ y' &= y \text{이므로 } y = y' \\ y' &= e^{-(\frac{1}{\sqrt{2}}x')^2} = e^{-\frac{(x')^2}{2}} \\ \therefore y &= e^{-\frac{x^2}{2}} \end{aligned}$$

변환을 시킨 함수 $y = e^{-\frac{x^2}{2}}$ 와 원래 함수 $y = e^{-x^2}$ 의 그래프를 함께 그려보면 아래 그림과 같이 최댓값 1은 불변이다. 그렇지만 이계도함수를 구하여 변곡점을 구해보면 변곡점의 위치가 $x = \pm 1$ 로 바뀌어 있고 이 변곡점에서 곡선이 y 축 양의 방향으로 올라가서 곡선 아래의 넓이가 증가되었음을 알 수 있다.



<그림 IV-1> 변곡점 변환에 의한 $y = e^{-\frac{x^2}{2}}$ 의 그래프와 CAS 계산기에 의한 결과

2. 정규분포곡선의 넓이를 1로 변환

앞에서 살펴본 바와 같이 $f(x) = e^{-x^2}$ 곡선 아래의 넓이는 $\int_{-\infty}^{\infty} (e^{-x^2}) dx \approx \sqrt{\pi}$ 이고, S_1 의

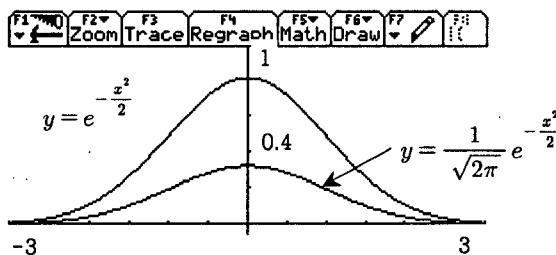
변환에 의해 $f(x) = e^{-\frac{x^2}{2}}$ 의 곡선 아래의 넓이는 $\sqrt{\pi}$ 보다 커졌다. 그렇다면 이 넓이는 결국 얼마인가? 넓이를 변환하기 위해서는 “ $S: (x, y) \rightarrow (ax, ay)$ 로 변환하면 변환된 넓이는 원래 넓이의 $|ab|$ 배가 되어 $|ab| \times (\text{원래 넓이})$ 이다”는 스케일 변환정리(Area Scale Change Theorem)를 사용한다. 그런데 앞의 $S_1: (x, y) \rightarrow (\sqrt{2}x, y)$ 의 변환과 스케일 변환정리에 의하면 $f(x) = e^{-\frac{x^2}{2}}$ 의 넓이는 $|\sqrt{2} \cdot 1| \times \sqrt{\pi} = \sqrt{2\pi}$ 이다. 변곡점 $x = \pm 1$ 에서 y 축으로 올라감으로써 넓이가 $\sqrt{\pi}$ 에서 $\sqrt{2\pi}$ 로 증가되었으므로 이 넓이를 1로 만들려면 스케일을 $S_2: (x, y) \rightarrow (x, \frac{y}{\sqrt{2\pi}})$ 로 하면 된다. 그러면

$$x' = x \text{이므로 } x = x'$$

$$y' = \frac{1}{\sqrt{2\pi}} y \text{이므로 } y = \sqrt{2\pi} y'$$

$$\sqrt{2\pi} y' = e^{-\frac{x^2}{2}}$$

$$\therefore y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



$\blacksquare \text{ Define } f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$\blacksquare \int_{-2}^2 f(x) dx$.954
$\blacksquare \int_{-1}^1 f(x) dx$	$\blacksquare \int_{-3}^3 f(x) dx$.997
.683	$\blacksquare \int_{-\infty}^{\infty} f(x) dx$	1.000

<그림 IV-2> 넓이 변환에 의한 $y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 의 그래프와 CAS 계산기에 의한 결과

위의 그래프와 CAS 계산기에 의한 적분 결과를 보면 $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 인 표준정규분포곡선은 다음과 같은 성질을 가진다.

첫째, x 축과 곡선 아래의 넓이는 1이다.

둘째, 변곡점은 $x = \pm 1$ 일 때이며 $-1 < x < 1$ 에서 볼록하고 $|x| > 1$ 에서 오목하다.

셋째, 이 함수는 짹수(우)함수이므로 $x = 0$ 인 y 축 대칭이다.

넷째, 이 함수의 최댓값은 $f(0) = \frac{1}{\sqrt{2\pi}} \approx 0.3989$ 이다.

다섯째, 모든 실수에 대해 $f(x) > 0$ 이다.

여섯째, $x \rightarrow \pm\infty$ 일 때, $f(x) \rightarrow 0$ 이다.

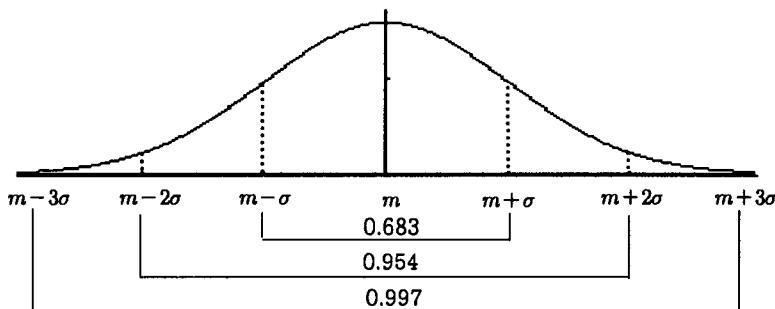
그리고 위의 적분 결과를 보면 $-1 < x < 1$ 사이의 넓이는 0.683이며 이는 전체 넓이의 68.3%에 해당한다. $-2 < x < 2$ 사이의 넓이는 0.954이며 이는 전체 넓이의 95.4%에 해당한다. 그리고 $-3 < x < 3$ 사이의 넓이는 0.997이며 이는 전체 넓이의 99.7%에 해당함을 알 수 있다.

정규분포를 $N(m, \sigma^2)$ 로 나타낼 때, 이 정규분포를 따르는 확률변수 X 에 대하여 확률 $P(a \leq X \leq b)$ 의 값은 정규분포곡선의 그래프와 x 축 사이의 넓이와 같다. 또, 확률 $P(m - k\sigma \leq X \leq m + k\sigma)$ ($k = 1, 2, 3$)의 값은 지금까지의 결과에 따르면 다음과 같음을 알 수 있다.

$$P(m - \sigma \leq X \leq m + \sigma) = P(m - \sigma < X < m + \sigma) \approx 0.683$$

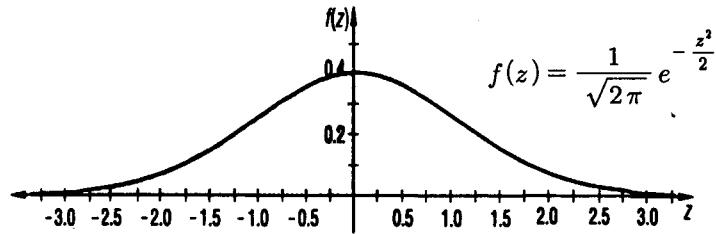
$$P(m - 2\sigma \leq X \leq m + 2\sigma) = P(m - 2\sigma < X < m + 2\sigma) \approx 0.954$$

$$P(m - 3\sigma \leq X \leq m + 3\sigma) = P(m - 3\sigma < X < m + 3\sigma) \approx 0.997$$



<그림 IV-3> 정규분포의 평균 m 과 표준편차 σ 와 정규분포곡선의 넓이 사이의 관계

일반적으로, 표준정규분포의 확률밀도함수의 x 축을 z 축으로 바꾸어 표현하므로 표준정규분포를 나타내는 함수는 다음과 같이 표현되며 그래프의 축도 z 축($-\infty < z < \infty$)과 $f(z)$ 축으로 바뀐다.

<그림 IV-4> 확률변수 z 에 의한 표준정규분포곡선

3. 여러 형태의 정규분포곡선과 성질

표준정규분포곡선에 대해 임의의 선형변환과 스케일 변환을 주더라도 그 결과는 여전히 정규분포곡선이다. 특히, m 과 s 가 상수이고 $s \neq 0$ 일 때, 다음 식의 그래프도 곡선 아래의 넓이가 1인 정규분포곡선이다.

$$f(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-m}{s})^2}$$

따라서 이러한 곡선도 확률분포를 나타낼 수 있다. 이 확률밀도함수에 대해 m 과 s 를 변화시켜본 그래프는 <그림 IV-5>와 같고 이로부터 중요한 성질을 발견할 수 있다.

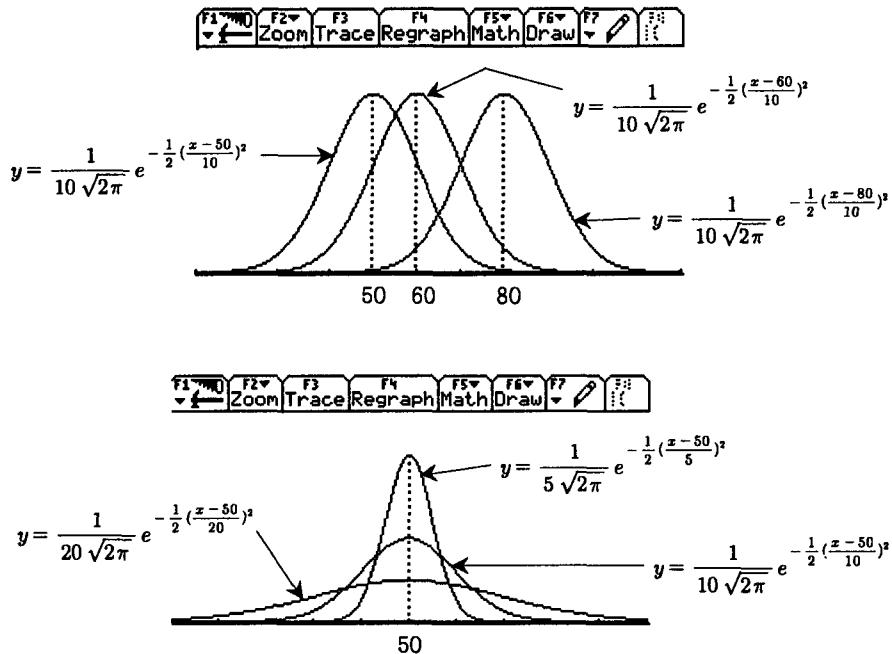
첫째, 확률밀도함수에서 $m > 0$ 이면 그래프가 오른쪽으로 평행이동하고 $m < 0$ 이면 그래프가 왼쪽으로 평행이동한다. 그리고 그래프의 모양은 변하지 않는다.

둘째, s 는 곡선의 퍼짐 정도를 나타낸다. s 가 커질수록 곡선의 퍼짐 정도는 커지고 작을수록 중심으로 모인다.

셋째, 이 곡선은 직선 $x = m$ 에 대하여 대칭이다.

그리고 이 경우 m 은 이 확률분포의 평균이고 s 는 표준편차를 나타낸다. 1730년대 Abraham DeMoivre가 평균이 μ 이고 표준편차가 σ 인 정규분포에 해당하는 다음과 같은 곡선의 방정식을 발견함으로써 정규분포에 대한 완벽한 수학적 식이 완성되었다. 이 식에서 상수 $\frac{1}{\sigma\sqrt{2\pi}}$ 는 앞에서 언급했듯이 곡선 아래의 넓이를 1로 만들기 위한 것이다.

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



<그림 IV-5> 상수 m 과 $s(s \neq 0)$ 에 따른 여러 형태의 정규분포곡선의 모양 변화

V. 결 론

본 연구는 고등학교 통계 영역의 확률분포 단원에 제시되어 있는 정규분포곡선과 정규분포의 확률밀도함수의 다양한 성질을 CAS 계산기를 활용하여 탐구하고, 이로부터 정규분포곡선의 교수-학습을 위한 시사점을 도출하고자 하였다. 이 목적을 달성하기 위하여 이항분포에서 정규분포로의 근사, 정규분포곡선의 탐구, Monte Carlo 방법에 의한 정규분포곡선의 넓이 탐구, 정규분포곡선의 선형변환, 그리고 여러 형태의 정규분포곡선 탐구 등의 내용을 중심으로 CAS 계산기를 활용하여 탐구해보았다. 본 연구의 결과로부터 정규분포곡선의 교수-학습을 위해 다음과 같은 시사점을 발견하였다.

첫째, 이항분포가 정규분포로 근사한다는 사실을 교과서처럼 그림으로 제시해서는 학생들에게 그 의미를 충분히 전달할 수 없으므로 CAS 계산기와 같은 공학을 사용하여 실험할 수 있는 기회를 제공해 주어야 할 것이다.

둘째, 현행 고등학교 수학과 교육과정에 의하면 적분을 배우기 전에는 x 축과 정규분포곡선 아래의 넓이를 구할 수 없으므로 이를 경우 간단한 프로그램을 활용하여 Monte Carlo 방법으로 구하도록 한다. 이렇게 함으로써 확률에 의한 넓이 계산이 가능함을 알 수 있게 하고 수학내용 사이의 연결성에 대한 인식을 증대시킬 수 있다.

셋째, 현행 고등학교 수학과 교육과정에 의하면 상수 $m, s(s \neq 0)$ 에 의해 선형변환된 함수

$f(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-m}{s})^2}$ 를 정규분포로 먼저 도입함으로써 복잡한 함수에 의한 두려움으로 인해 정규분포곡선에 대한 이해를 방해하고 있다고 본다. 이런 도입 방법보다는 먼저 정규분포곡선의 원형인 $f(x) = e^{-x^2}$ 에서 $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 로의 선형변환을 통해 도입하는 것이 복잡한 함수식에 대한 부담을 덜고 함수 형태에 대한 이해를 높일 수 있다고 본다.

넷째, 현행 고등학교 수학과 교육과정에서는 정규분포 $f(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-m}{s})^2}$ 를 먼저 도입하고 $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 을 표준정규분포로 도입하고 있다. 이것은 위에서도 지적했듯이 이미 선형변환된 함수식을 먼저 도입함으로써 학생들의 함수식에 대한 개념과 이해를 어렵게 하고 있다고 본다. 따라서 본 논문에서 제시했듯이 표준정규분포곡선을 먼저 도입하는 것이 학생들의 인지적 부담을 줄이는 방법이라고 본다.

다섯째, 현행 고등학교 수학과 교육과정에 제시되어 있는 정규분포곡선의 함수식은 $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ 의 꼴로 주어져있다. 정규분포곡선의 이 함수식을 보면 $x - m$ 은 평균으로부터의 거리와 방향을 나타내는 편차이므로 이를 제곱한 것은 편차제곱이다. 이것을 표준편차의 제곱 σ^2 인 분산으로 나눈다는 것은 무의미하다. 따라서 평균으로부터의 거리와 방향을 표준편차로 나누어 단위 길이로 만드는 과정이 표현되도록 $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-m}{\sigma})^2}$ 으로 나타내는 것이 더 타당하다고 제안한다.

참 고 문 헌

- 교육과학기술부 (2008). 고등학교 교육과정 해설5: 수학. 서울: 교육과학기술부.
- 임석훈 · 기호삼 · 김동수 · 이방수 (2002). 수학I. 서울: 천재교육.
- 임재훈 · 이경화 · 김진호 · 윤오영 · 반용호 · 조동석 외 (2002). 고등학교 수학I. 서울: 두산.
- Bakker, A. (2004). *Design research in statistics education on symbolizing and computer tools*. Doctoral dissertation, Utrecht University, 2004.
- ESS (1981). *Encyclopedia of statistical sciences* (Eds. Kotz, S. & Johnson, N. L.). New York: Wiley & Sons.
- Heid, K. (2003). Theories for thinking about the use of CAS in teaching and learning

- mathematics. In James, T. F., Al, C., Carolyn, K., Lin, M., & Rose, M. Z.(Eds.), *Computer algebra systems in secondary school mathematics education* pp.33-52. Reston, VA: NCTM.
- Kelly, B. (1997). *Investigating statistics with IT-92*. Burlington, Ontario: Brendan Kelly Publishing Inc.
- Kutzler, B. (2003). CAS as pedagogical tools for teaching and learning mathematics. In J. T. Fey, A. Cuoco, C. Kieran, L. McMullin, & R. M. Zbiek(Eds.), *Computer algebra systems in secondary school mathematics education* pp.53-72. Reston, VA: NCTM.
- Lagrange, J. B. (2005). Transposing computer tools from the mathematical sciences into teaching. In Dominique, G., Kenneth, R., & Luc, T.(Eds.), *The didactical challenge of symbolic calculators: Turning a computational device into a mathematical instrument* pp.67-82. New York: Springer.
- NCTM (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
- Tall, D., Smith, D., & Piez, C. (2008). Technology and calculus. In M. K. Heid & G. W. Blume(Eds.), *Research syntheses* pp.207-258. Reston, VA: NCTM.
- Wilensky, U. (1997). What is normal anyway? Therapy for epistemological anxiety. *Educational Studies in Mathematics*, 33, pp.171-202.

Pedagogical Implications for Teaching and Learning Normal Distribution Curves with CAS Calculator in High School Mathematics

Cho, Cheong-Soo

Department of Mathematics Education, Yeungnam University, Gyeongsan, Korea, 712-749

E-mail : chocs@ynu.ac.kr

The purpose of this study is to explore normal distribution in probability distributions of the area of statistics in high school mathematics. To do this these contents such as approximation of normal distribution from binomial distribution, investigation of normal distribution curve and the area under its curve through the method of Monte Carlo, linear transformations of normal distribution curve, and various types of normal distribution curves are explored with CAS calculator. It will not be able to be attained for the objectives suggested the area of probability distribution in a paper-and-pencil classroom environment from the perspectives of tools of CAS calculator such as trivialization, experimentation, visualization, and concentration. Thus, this study is to explore various properties of normal distribution curve with CAS calculator and derive from pedagogical implications of teaching and learning normal distribution curve.

* ZDM Classification : K64

* 2000 Mathematics Subject Classification : 97D40

* Key Words : CAS(Computer Algebra Systems) Calculator, normal distribution, normal curve