韓國數學敎育學會誌 시리즈 E
<數學敎育 論文集>
제 11집, 2001. 1. 367-387.

J. Korea Soc. Math. Ed. Ser. E
Communications of Mathematical Education
Vol. 11, Jan. 2001. 367-387.

Excel과 Mathview를 활용한 고등학교 통계지도

김 지 곤 (상명대학교)

통계 수업에서는 실제 자료를 직접 다루는 활동이 있어야 하며 이것은 계산기 또는 컴퓨터가 있어야 한다. 계산기 기능만으로는 많은 양의 자료를 반복해서 다루는데 문제가 있다. 그렇다고 통계 응용프로그램을 사용하는 것은 통계의 원리와 과정을 배우는 것이 목적인 고등학교 통계 교육과정에 맞지 않다. 이러한 면에서 볼 때 Excel은 계산기적 기능과 통계적 기능을 모두 갖추고 있으며 또한 시중에 많이 보급된 장점도 있다. 통계문제를 처리할 때 Excel의 계산기 기능을 사용하여 교과서의 원리대로 계산한 후 Excel의 통계기능을 이용하여 검토해 봄으로써 향후 통계 응용프로그램을 다루는 기초를 쌓을 수 있다. 그러나 Excel은 적분기능이 없어 연속분포에서 적분이 필요한 경우는 MathView를 사용하였다. 제 7차 고등학교 통계 교육과정의 내용을 Excel을 활용하여 지도하는 모델을 개발해 봄으로써 새 수학과 교과서 개발을 위한 기초연구를 제공하고자 한다.

1. 서 론

요즈음 교사들은 열린 수업, 수준별 수업, 학습자의 자기 주도적 학습능력 향상, 컴퓨터를 활용한 수업 등 제 7차 교육과정에 맞는 모형을 개발하기 위해 시범학교, 시범수업에 참여하고 적응하기 위해 많은 노력을 하고 있다. 시대의 변화에 따라 학생들은 특히 수학분야에 있어서는 생각하고 계산하기는 싫어하면서 컴퓨터를 사용하는 것은 매우 흥미로워 한다. 지금 현재로서는 1 대 1 컴퓨터를 활용한 수업은 어려워서 시범적으로 보여주는 수업을 하고 있지만, 7차 교육과정 중에는 경제발전과 기술발전으로 인한 컴퓨터가격의 하락으로 학생 개별로 컴퓨터를 사용한 수업이 이루어질 수 있을 것이다. 컴퓨터를 활용하여 지도함으로서 싫어하는 수학에 대한 흥미를 유발시킬 수 있고 교육의 수월성을 추구할 뿐만 아니라 학생들로 하여금 컴퓨터에 친숙하고 스스로 활용할 수 있는 능력과 태도를 길러 주게 될 것이다. 또한 수학적 사고력과 창의력 문제해결력 향상을 위해 학습자의 직접적인 경험이 필요한데 학생들이 직접 컴퓨터에 자료를 입력하여 결과를 봄으로서 이해도 빠르고 장기 기억에 효과적일 수 있다.

그러나 결과만 빠르게 나타내는 컴퓨터의 활용이 수학적 원리와 과정을 중요시 여기는 수학교육에서 효과적이라고 할 수 있을까? 는 의문이 남는다. 여기에 대해서는 수학의 각 분야에 관한 각 소프트웨어의 분해적 접근 가능성 등 여러 가지 수학교육에 미칠 영향에 대한 연구가 필요하다. 그러나 많은 자료를 처리해야할 통계 수업에서는 빠른 시일 내에 교과서의 본문에서부터 컴퓨터의 활용

* 본 연구는 2000학년도 상명대학교 자연과학연구소 학술연구비에 의하여 수행되었음.

이 도입되어야 하며 각 계산과정 별로 분해가 가능하고 늘리 보급된 Excel이 교과서에 도입될 수 있는 소프트웨어 중의 하나라 생각되어진다.

Excel의 기능 중 통계 수업에서 가장 강한 장점으로 생각되어 지는 기능은 어떤 한 셀에서 변수를 사용하여 얻어진 식의 값을 인접 셀에서 변수의 값을 변화하면서 다음의 값을 구하기 위해서는 변화된 변수의 값과 식을 반복하여 입력할 필요 없이 채우기 핸들을 이용하여 복사하면 되고, 복사된 셀을 클릭하여 보면 그 셀에 맞는 변수의 값으로 식이 변화된 것을 확인할 수 있는 것이 그래픽 계산기에서 할 수 없는, 어떤 통계 프로그램에서도 할 수 없는 Excel만의 장점으로 결과보다도 과정을 중요시 여기는 통계교육에서는 가장 채택할 만한 프로그램이다. 통계 개념이나 문제 해결을 위해서는 Excel의 사용 방법을 설명해 가면서 학생 스스로도 그것을 사용할 수 있게 해야한다. 그렇게 하면 종이와 연필 없이 통계수업을 진행할 수 있고 복잡한 반복계산에서 해방되므로 통계의 원리를 이해할 수 있는 여유를 갖게 된다. 이러한 Excel의 강력한 데이터 처리 능력, 즉 계산기적 능력과 통계처리능력을 접하면서 상명대학교 수학교육과 3학년 통계수업에 적용해 본 결과 매우 좋은 반응을 얻었다. 처음의 우려와는 달리 계산은 Excel로 하면 되니까, 분해된 과정으로 문제를 해결해야 하므로 이론시간에 더욱 흥미를 보였고 이론의 중요성을 인식하였다. 자료의 정리부터, 분산분석, 회귀분석, 비모수통계 등 통계의 모든 분야를 Excel로 교과서의 원리대로 처리할 수 있었고 진도도 빨랐고 학생들도 통계의 처리에 대해 자신감을 갖고 흥미로워했다. Excel은 대학생뿐만 아니라 고등학교 통계수업에 활력을 불어넣을 수 있으므로 교과서의 본문에서부터 활용할 가치가 있는 것으로 생각되어 제 7차 교육과정의 통계내용을 이산적인 것은 Excel로 연속적인 변수는 MathView를 활용하여 교과서의 식의 전개에 따르면서 통계문제를 처리할 수 있음을 확인하였다. 따라서 7차 교육과정의 통계내용은 본문에서부터 Excel을 도입해도 무리가 없다는 것을 알아보았다.

2. 제 7차 수학교육과정에서 확률 통계 과정의 단계별 내용 체계

제 7차 교육과정은 국민공통기본 교육과정과 고등학교 선택중심 교육과정으로 구성되어 제6차 교육과정과는 다른 점이 많으므로 확률 통계에 관한 교육과정의 편성을 알아보면 아래와 같다. 특히 지난번 교육과정까지는 모든 고등학생들이 똑 같은 과정의 확률 통계를 배웠어나, 제 7차 교육과정에서는 고등학교 2학년부터 선택과목제로 운영되므로 각 선택과목에 관한 내용체계를 알아야할 필요가 있다. 수학에 관한 선택과목은 실용수학, 수학 I, 수학II, 미분과 적분, 확률과 통계, 이산수학 6개 교과목이다. 확률, 통계의 내용이 나오는 과목은 실용수학, 수학 I, 확률과 통계, 이산수학 4개 과목이다. 특히 실용수학, 확률과 통계, 이산수학 3개 과목은 10단계 수학에 도달 여부에 관계없이 학생들이 선택할 수 있는 과목으로서, 연속확률변수에서 적분 기호를 사용하지 않는 특색이 있다. 수학 I 에서의 확률, 통계는 현재 수학 I 의 내용과 큰 차이가 없으나 적분을 이용하여 연속확률변수의 평균과 표준편차를 구하는 것은 지도하지 않고, 통계적 추정은 모집단이 정규분포인 경우만 다룬다. 각 단계별 및 각 선택과목별 확률 통계 내용은 아래표와 같다.

< 각 단계별 및 각 선택과목별 확률 통계 내용 >

국민 공통 기본 교육 과정

<1-가 단계>
(1) 분류하기
① 사물이나 사람을 미리 정한 한 가지 기준 에 따라 분류하여 각각의 개수를 셀 수 있다.

<2-나 단계>
(1) 표의 작성
① 실생활에서 찾을 수 있는 구체적인 자료의 크기를 조사하여 표로 나타낼 수 있다.
② 조사된 자료를 ○와 같은 표시나 간단한 그림을 이용하여 그래프로 나타내고, 자료의 크기를 비교할 수 있다.
③ 표나 그래프가 자료의 크기를 나타내고 비교하는데 편리하다는 것을 알 수 있다.

<3-나 단계>
(1) 자료의 정리
① 생활에서 발생하는 실제적인 자료들을 수집, 분류, 정리하여 표를 만들고, 이를 막대그래프로 나타내고 읽을 수 있다.
② 적절한 소재를 선택하여 자료를 수집하고, 이를 분류, 정리하여 알맞은 그래프로 나타내고 여러 가지 사실을 찾을 수 있다.

<4-나 단계>
(1) 꺾은선그래프
① 연속적인 변량에 대한 자료를 표로 만들고, 이를 바탕으로 꺾은선그래프를 그릴 수 있으며, 여러 가지 사실을 찾아 낼 수 있다.
② 막대그래프와 꺾은선그래프를 비교하여 그 차이점을 이해하고, 각각의 특성과 용도를 안다.
(2) 여러 가지 그래프로 나타내기
① 실생활에서 찾을 수 있는 여러 가지 자료를 목적에 맞는 그래프로 나타내고, 여러 가지 사실을 알 수 있다.

<5-나 단계>
(1) 자료의 표현
① 자료를 정리하여 이를 줄기와 잎 그림으로 나타내고 자료의 특성을 파악할 수 있다.
② 평균의 의미를 알고, 주어진 자료의 평균을 구할 수 있다.

<6-가 단계>
(1) 비율그래프
① 띠그래프와 원그래프의 의미를 알고, 이를 활용할 수 있다.

<6-나 단계>
(1) 경우의 수와 확률
① 경우의 수의 뜻을 알고, 이를 구할 수 있다.
② 경우의 수를 바탕으로 확률의 의미를 안다.

<7-나 단계>
(1) 도수분포와 그래프
① 도수분포표, 히스토그램, 도수분포다각형을 이해한다.
② 주어진 자료를 표나 그래프로 나타내고, 이를 읽을 수 있다.
③ 도수분포표에서 평균의 뜻을 알고, 이를 구할 수 있다.
(2) 상대도수의 분포와 누적도수의 분포
① 상대도수의 분포와 누적도수의 분포를 이해하고, 이를 그래프로 나타낼 수 있다.

<8-나 단계>
(1) 확률과 그 기본 성질
① 간단한 경우의 수 또는 상대도수를 이용하여 확률의 뜻을 안다.
② 확률의 기본 성질을 이해하고 간단한 확률의 계산을 할 수 있다.

<9-나 단계>
(1) 상관도와 상관표
① 상관도와 상관표를 알고, 주어진 자료를 상관도와 상관표로 나타낼 수 있다.
② 상관도와 상관표를 보고, 두 변량 사이의 상관관계를 알 수 있다.

<10-가 단계>
(1) 산포도와 표준편차
① 산포도와 표준편차를 이해하고, 이를 구할 수 있다.

실용 수학

실용수학은 10 단계 수학 도달 여부에 관계없이 학생 들이 실생활에 필요한 수학을 학습하기 위하여 선택할 수 있는 과목이다

(가) 생활 통계
(1) 자료의 정리와 요약
① 관찰된 자료를 여러 가지 그래프나 표로 나타내고, 그 자료의 전체적인 경향과 분포를 파악할 수 있다.
② 관찰된 자료의 평균과 분산을 구하고, 그 의미를 이해한다.
(2) 확률과 통계의 활용
① 확률의 뜻을 알고, 이를 활용할 수 있다.
② 복권 등의 기대값을 구할 수 있다.
③ 이항분포를 실생활의 문제 해결에 활용할 수 있다.
④ 정규분포를 실생활의 문제 해결에 활용할 수 있다.
⑤ 여론 조사의 결과를 해석할 수 있다.

수학 I

(가) 순열과 조합

(1) 경우의 수

① 합의 법칙, 곱의 법칙을 이해하고, 이를 이용하여 경우의 수를 구할 수 있다.

(2) 순열

① 순열의 뜻을 알고, 순열의 수를 구할 수 있다.

② 원순열, 중복순열, 같은 것이 있는 순열을 이해하고, 그 순열의 수를 구할 수 있다.

(3) 조합

① 조합의 뜻을 알고, 조합의 수를 구할 수 있다.

(4) 이항정리

① 이항정리를 이해한다.

② 이항정리를 이용하여 여러 가지 문제를 해결할 수 있다.

(나) 확률

(1) 확률의 뜻

① 통계적 확률과 수학적 확률의 뜻을 알고 그 관계를 이해한다.

② 확률의 기본 성질을 이해한다.

(2) 확률의 계산

① 확률의 덧셈정리를 이해하고, 이를 활용할 수 있다.

② 여사건의 확률의 뜻을 알고, 이를 활용할 수 있다.

③ 조건부확률의 뜻을 알고, 이를 구할 수 있다.

④ 사건의 독립과 종속의 뜻을 이해한다.

⑤ 확률의 곱셈정리를 이해하고, 이를 활용할 수 있다.

⑥ 독립시행의 확률을 이해하고, 이를 활용할 수 있다.

(다) 통계

(1) 확률분포

① 확률변수와 확률분포의 뜻을 안다.

② 이산확률변수의 기대값(평균)과 표준편차의 뜻을 이해하고, 이를 구할 수 있다.

③ 이항분포의 뜻을 이해하고, 이항분포에서 평균과 표준편차를 구할 수 있다.

④ 정규분포의 뜻과 그 성질을 이해한다.

(2) 통계적 추정

① 모집단과 표본의 뜻을 안다.

② 표본평균과 그 분포를 이해한다.

③ 표본평균과 모평균의 관계를 이해한다.

④ 모평균을 추정할 수 있다.

<학습 지도상의 유의점>

① 적분을 이용하여 연속확률변수의 평균과 표준편차를 구하는 것은 지도하지 않는다.

② 통계적 추정은 모집단이 정규분포인 경우만 다룬다.

확률과 통계

(가) 자료의 정리와 요약

(1) 자료의 정리

① 관찰된 자료를 도수분포표와 히스토그램으로 나타 내고, 그 자료의 분포와 특성을 파악할 수 있다.

② 관찰된 자료를 줄기와 잎 그림으로 나타내고, 그 자료의 분포와 특성을 파악할 수 있다.

(2) 자료의 요약

① 대표값으로서 평균, 중앙값, 최빈값을 구할 수 있다.

② 산포도로서 범위, 분산, 표준편차를 구할 수 있다.

(나) 확률

(1) 확률

① 확률의 뜻과 성질을 알고, 실생활에서 일어나는 여러 가지 우연 현상을 이해한다.

② 순열과 조합을 이용하여 확률을 구하고, 실생활과 관련된 확률 문제를 해결할 수 있다.

(2) 조건부확률

① 조건부확률을 이해하고, 이를 구할 수 있다.

(다) 확률변수와 확률분포

(1) 확률변수

① 이산확률변수와 연속확률변수의 뜻을 안다.

② 이산확률변수의 기대값과 분산을 구할 수 있다.

(2) 확률분포

① 이항분포를 이해하고, 이를 활용할 수 있다.

② 정규분포를 이해하고, 이를 활용할 수 있다.

<학습 지도상의 유의점>

① 확률변수는 직관적으로 이해할 수 있도록 한다.

② 연속확률변수에서 적분 기호를 사용하지 않는다.

③ 연속확률변수의 기대값은 다루지 않는다.

(라) 통계적 추정

(1) 표본의 뜻

① 모집단과 표본의 관계를 이해한다.

② 표본평균의 분포를 이해한다.

(2) 구간추정

① 모평균을 구간추정하고, 그 결과를 해석할 수 있다.

② 모비율을 구간추정하고, 그 결과를 해석할 수 있다.

<학습지도상의 유의점> ① 모평균에 대한 추정은 모집단의 분포가 정규분포일 때만 다루고, 모비율에 대한 추정은 표본의 크기가 클 때만 다룬다. **이산 수학** (가) 선택과 배열 (1) 순열과 조합 ① 합의 법칙, 곱의 법칙을 이해하고, 이를 이용하여 여러 가지 경우의 수를 구할 수 있다. ② 순열과 조합의 뜻을 알고, 어떤 주어진 조건을 만족 하는 순열이나 조합을 모두 나열할 수 있다.	③ 일일이 나열하지 않고도 어떤 주어진 조건을 만족 하는 순열이나 조합의 수를 구할 수 있다. ③ 유한집합을 서로소인 몇 개의 집합의 합집합으로 나타낼 수 있는 방법의 수를 구할 수 있다. ④ 어떤 자연수를 몇 개의 자연수의 합으로 나타낼 수 있는 방법의 수를 구할 수 있다. ⑤ 주어진 조건에 맞는 여러 가지 분배의 수를 구할 수 있다.

3. 고등학교 통계내용의 EXCEL과 MATHVIEW 활용

3.1 도수분포

주어진 자료에서 도수분포표를 작성하고 평균과 표준편차를 구하는 것에 대해서는 중학교에서 이미 학습하였다. 그러나 중학교에서는 단순히 평균과 표준편차를 공식에 의해서만 계산하여 구하였으나, 가평균을 이용하거나 계급의 크기를 이용하여 평균과 표준편차를 계산하는 간편 계산 방법은 학습하지 않았다. 따라서 중학교에서 배운 평균과 표준편차를 복습한 후 가평균을 이용한 간편 계산 방법을 이용하여 평균과 표준편차를 구하는 방법을 이해시킨다. 여기서 기능의 숙달보다는 원리를 이해시키고, 처음에는 Excel의 계산기 기능을 이용하여 단계적으로 계산해보며 그 결과를 Excel의 통계기능을 통하여 검토해 본다.

3.1.1 도수분포표

Excel을 이용한 도수분포표를 작성하기 위해 주어진 자료를 <표 1>에서와 같이 A2에서 H6까지 입력한 후 D10부터 D17 까지 20, 30, 40, 50, 60, 70, 80, 90을 입력한다. 20은 $x \leq 20$, 30은 $20 < x \leq 30$, … , 90은 ($80 < x \leq 90$)을 의미한다. F10부터 F17 까지 계급값 15, 25, 35, 45, 55, 65, 75, 85를 입력한다. G10에서 G17까지 선택한 후 함수마법사 f_x 를 선택하고, 함수 종류는 통계, 함수 이름은 FREQUENCY를 선택한 후 Data_arry 에는 A2:H6를 Bins_array 에는 D10:D17을 입력한다. 이 때 주의 할 것은 확인을 클릭하거나, [Enter]를 쳐서는 안되고 반드시 [Shift],[Ctrl] 두키를 누른 상태에서 [Enter]를 쳐야한다.

상대도수는 H10에서 =G10/G18을 입력한 후 채우기 핸들로 H11부터 H17까지 복사한다. 최대값, 최소값, 중앙값, 최빈값, 범위는 <표1>에 나타나 있는 데로 입력하면 된다.

<표 1> 도수분포표

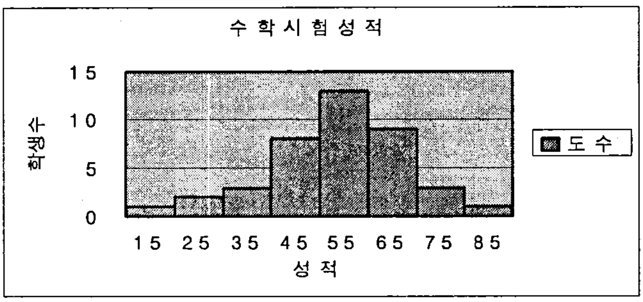
	A	B	C	D	E	F	G	H
1	수학 시험 성적							
2	35	61	52	76	45	71	56	63
3	46	52	43	14	57	82	36	51
4	27	74	54	42	69	34	55	64
5	66	49	53	58	66	56	53	57
6	48	58	61	43	62	21	45	62
7								
8	최대값	82		도수분포표				
9	최소값	14		Excel기준	계급	계급값	도수	상대도수
10				20	10<x≤20	15	1	0.025
11				30	20<x≤30	25	2	0.05
12				40	30<x≤40	35	3	0.075
13				50	40<x≤50	45	8	0.2
14	중앙값	54.5		60	50<x≤60	55	13	0.325
15	최빈값	61		70	60<x≤70	65	9	0.225
16	범위	68		80	70<x≤80	75	3	0.075
17				90	80<x≤90	85	1	0.025
18						계	40	1

=MAX(A2:H6)
=MIN(A2:H6)
=MEDIAN(A2:H6)
=MODE(A2:H6)
=B8-B9
조심 : 본문참조

3.1.2 히스토그램과 줄기-잎그림

도수가 나타나 있는 G10에서 G17까지 선택한 후 도구상자의 챠트 마법사를 클릭한다. 세로막대형을 선택한 후 다음을 클릭하고 계열을 클릭한 후 X(항목)축레이블(Ⅰ)의 입력란에 커서를 두고서 챠트 마법사를 계급값이 보이도록 옆으로 옮긴 후 계급값이 있는 F10부터 F17까지 선택한다. 그러면 X축 항목이 계급값으로 표시되어진다. 마침을 클릭하고 나온다. 이 상태는 막대그래프 상태이기 때문에 히스토그램으로 바꾸기 위해서 그림의 막대부분을 클릭한 후 오른쪽 마우스를 클릭하고 데이트 계열서식을 선택한다. 오른쪽 끝에 있는 옵션을 클릭한 후 간격너비를 0으로 입력한 후 확인을 클릭한다.

Excel에서는 기본적으로 줄기-잎그림 기능이 없다. 매크로를 사용하여 프로그램을 짜야한다. 서울대학교 통계학과에서 개발한 Excel을 사용한 통계프로그램 KESS를 이용하여 아래의 줄기-잎그림을 그렸다. 자료를 한 열로 만들고 통계분석-그래프-줄기잎그림을 선택하면 된다.



<그림 1> 히스토그램

```
줄기-잎 그림(Stem-and-Leaf Plot)
변수명: 성적자료

Stem Unit: 10  Leaf Unit: 1

   1      1   4
   3      2   17
   6      3   456
  14      4   23355689
 ( 13)    5   1223345667788
  13      6   112234669
   4      7   146
   1      8   2
```

<그림 2> 줄기-잎 그림

3.1.3 평균과 표준편차

(1) 변량 $x_1, x_2, \cdots, x_n$의 평균을 m이라고 하면

$$m=\frac{x_1+x_2+\cdots+x_n}{n}=\frac{\sum_{i=1}^{n} x_i}{n} \tag{3.1}$$

(2) 변량 $x_1, x_2, \cdots, x_n$에 대응하는 도수가 각각 $f_1, f_2, \cdots, f_n$일 때 평균을 m이라고 하 면

$$m=\frac{x_1 f_1+x_2 f_2+\cdots+x_n f_n}{f_1+f_2+\cdots+f_n}=\frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} \tag{3.2}$$

(3) 가평균 a를 이용한 평균 계산

$$m=a+\frac{(x_1-a)f_1+(x_2-a)f_2+\cdots(x_n-a)f_n}{f_1+f_2+\cdots+f_n}=a+\frac{\sum_{i=1}^{n}(x_i-a)f_i}{\sum_{i=1}^{n} f_i} \tag{3.3}$$

(4) 가평균 a와 계급값 c를 이용한 평균 계산

$$m=a+\frac{\frac{(x_1-a)}{c}f_1+\frac{(x_2-a)}{c}f_2+\cdots\frac{(x_n-a)}{c}f_n}{f_1+f_2+\cdots+f_n}\times c=a+\frac{\sum_{i=1}^{n}\frac{(x_i-a)}{c}f_i}{\sum_{i=1}^{n} f_i}\times c$$

$$=a+\frac{\sum_{i=1}^{n} u_i f_i}{\sum_{i=1}^{n} f_i}\times c=a+c\times\overline{u_i} \qquad (\text{단, } u_i=\frac{x_i-a}{c},\ \overline{u_i}\text{는 } u_i\text{의 평균}) \tag{3.4}$$

(5) 분산

$$\sigma^2=\frac{\sum_{i=1}^{n}(x_i-m)^2 f_i}{\sum_{i=1}^{n} f_i} \tag{3.5}$$

$$=\frac{\sum_{i=1}^{n}(x_i^2-2mx_i+m^2)f_i}{\sum_{i=1}^{n} f_i}=\frac{\sum_{i=1}^{n} x_i^2 f_i-2m\sum_{i=1}^{n} x_i f_i+m^2\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} f_i}$$

$$= \frac{\sum_{i=1}^{n} x_i^2 f_i}{\sum_{i=1}^{n} f_i} - m^2 = \overline{x_i^2} - m^2 \tag{3.6}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - A)^2 f_i}{\sum_{i=1}^{n} f_i} - \left(\frac{\sum_{i=1}^{n}(x_i - A) f_i}{\sum_{i=1}^{n} f_i}\right)^2$$

$$= c^2\left(\frac{\sum_{i=1}^{n} u_i^2 f_i}{\sum_{i=1}^{n} f_i} - \left(\frac{\sum_{i=1}^{n} u_i f_i}{\sum_{i=1}^{n} f_i}\right)^2\right) \tag{3.7}$$

(6)표준편차

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - m)^2 f_i}{\sum_{i=1}^{n} f_i}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 f_i}{\sum_{i=1}^{n} f_i} - m^2} = \sqrt{\overline{x_i^2} - m^2} \tag{3.8}$$

$$= c\sqrt{\frac{\sum_{i=1}^{n} u_i^2 f_i}{\sum_{i=1}^{n} f_i} - \left(\frac{\sum_{i=1}^{n} u_i f_i}{\sum_{i=1}^{n} f_i}\right)^2} \tag{3.9}$$

3.1.4 Excel을 이용한 자료의 평균과 표준편차 계산

(a). (3.1)식을 이용한 평균을 구하기 위하여 C16 셀에서 =SUM(A2:H6)/COUNT (A2:H6)을 입력하면 평균 52.925가 나타난다. 이 평균계산이 맞는가를 확인하기 위하여 C17셀에서 =AVERAGE(A2:H6)을 입력하면 같은 평균값 52.925가 나타난다.

(b). (3.5)식을 이용하여 분산을 구하기 위해 A9 셀에서 =A2-C16을 입력하여 A2셀의 평균과의 편차를 구한 후 A9 셀부터 H13 셀 까지 채우기 핸들을 사용하여 복사하면 아래의 편차표가 만들어진다. 편차표의 정확도를 확인하기 위하여 H14 셀에서 =SUM(A9:H13)을 입력하면 1.14×10^{-13}이 나타나 편차 표가 정확하다는 것을 확인함으로서 편차의 의미를 음미해본다. 분산을 구하기 위해 G16 셀에서 =SUMSQ(A9:H13)/COUNT(A9:H13)을 입력하면 분산 202.1194 가 나타난다.

(c). (3.6)식을 이용하여 분산을 구하기 위해 G17 셀에서 =SUMSQ(A2:H6)/COUNT (A2:H6)-(C16)^2을 입력하면 분산 202.1194가 구해진다. 이 분산을 확인하기 위해 G18 셀에서 =VARP(A2:H6)을 입력하면 202.1194가 나타나 옳음을 알 수 있다.

(d). (3.7)식을 이용하여 표준편차를 구하기 위해 E14 셀에서 =SQRT(G16)를 입력하면 표준편차 14.21687 이 구해진다. 표준편차를 확인하기 위해 E15 셀에서 =STDEVP(A2:H6)을 입력하면 같은 값

14.21687 이 나온다.

<표 2> 자료의 평균과 표준편차

	A	B	C	D	E	F	G	H	I
1					수학 시험 성적				
2	35	61	52	76	45	71	56	63	
3	46	52	43	14	57	82	36	51	
4	27	74	54	42	69	34	55	64	
5	66	49	53	58	66	56	53	57	
6	48	58	61	43	62	21	45	62	
7									
8	<편차표>	=A2-C16							
9	-17.925	8.075	-0.925	23.075	-7.925	18.075	3.075	10.075	
10	-6.925	-0.925	-9.925	-38.925	4.075	29.075	-16.925	-1.925	
11	-25.925	21.075	1.075	-10.925	16.075	-18.925	2.075	11.075	
12	13.075	-3.925	0.075	5.075	13.075	3.075	0.075	4.075	
13	-4.925	5.075	8.075	-9.925	9.075	-31.925	-7.925	9.075	
14	=SUM(A2:H6)/COUNT(A2:H6)			표준편차1 :	14.21687		편차의 합	1.14E-13	
15				표준편차2 :	14.21687	=SUMSQ(A9:H13)/COUNT(A9:H13)			
16	평균1	: $\Sigma x_i/n$	52.925		분산1	: $\Sigma(x_i-m)^2/n$	202.1194		
17	평균2	확인	52.925		분산2	: $\Sigma x_i^2/n-m^2$	202.1194	=VARP(A2:H6	
18	=AVERAGE(A2:H6)				분산3	확인	202.1194		
19				=SUMSQ(A2:H6)/COUNT(A2:H6)-(C16)^2					
20									

3.1.5 Excel을 이용한 도수분포표에서의 평균과 표준편차 계산

도수분포표에서 적용되는 (3.2)~(3.9)식을 이용하여 아래 표를 사용하면서 평균 분산 표준편차를 구해보자.

(a). (3.2)식을 이용하여 평균을 구하려면 B11 셀에서 =SUM(B4:B10)을 입력하여 도수의 합 53을 구한 후 C4 셀에서 =A4*B4 입력 후 C10까지 채우기 핸들로 복사한 후 C11 셀에서 =SUM (C4:C10)을 입력하면 총합 9102.5 가 나타난다. C13 셀에서 =C11/B11 입력하면 평균 171.74528 이 구해진다.

(b). (3.4)식을 이용하여 평균을 구하려면 D4 셀에서 =(A4-167.5)/5를 입력하여 D10까지 채우기 핸들로 복사한 후 E4 셀에서 =D4*B4 입력한 후 E10 까지 채우기 핸들로 복사한다. E11 셀에서 =SUM(E4:E10)을 입력하면 가평균과 계급값을 이용한 편차의 합 45가 나타난다. C13 셀에서 =E11/B11*5+167.5를 입력하면 가평균 167.5와 계급값 5를 이용한 평균 171.7452801이 구해진다.

<표 3> 도수분포표의 평균

	A	B	C	D	E	F	G
1			=(A4-167.5)/5	$u_i = (x_i - 167.5) / 5$			
2		키					
3	키 : x_i	학생수 : f_i	$x_i f_i$	u_i	$u_i f_i$	=D4*B4	
4	152.5	1	152.5	-3	-3		
5	157.5	3	472.5	-2	-6		
6	162.5	4	650	-1	-4		
7	167.5	10	1675	0	0		
8	172.5	15	2587.5	1	15		
9	177.5	17	3017.5	2	34		
10	182.5	3	547.5	3	9	=E11/B11*5+167.5	
11	계	53	9102.5		45		
12							
13		평균 1	171.74528	평균 2	171.74528		
14							
15		=C11/B1					

<표 4> 도수분포표의 표준편차

	A	B	C	D	E	F	G	H
1			=(A4-C13)^2*B4		=A4^2*B4			
2		키						
3	키 : x_i	학생수 : f	$x_i f_i$	$(x_i-m)^2 f_i$	$x_i^2 f_i$	u_i	$u_i f_i$	$u_i^2 f_i$
4	152.5	1	152.5	370.3809	23256.25	-3	-3	9
5	157.5	3	472.5	608.7843	74418.75	-2	-6	12
6	162.5	4	650	341.901	105625	-1	-4	4
7	167.5	10	1675	180.2243	280562.5	0	0	0
8	172.5	15	2587.5	8.543966	446343.8	1	15	15
9	177.5	17	3017.5	562.985	535606.3	2	34	68
10	182.5	3	547.5	346.9918	99918.75	3	9	27
11	계	53	9102.5	2419.811	1565731		45	135
12								
13		평균 : m	171.745283		=D11/B11			
14								
15		분산 1	$E[(X-m)^2]$	45.65682			표준편차 1	6.756983
16					=E11/B11-C13^2			
17		분산2	$E[X^2]-(E[X])^2$	45.65682			표준편차2	6.756983
18					=5^2*(H11/B11-(G11/B11)^2)			
19		분산3	$c^2(E[U^2]-(E[U])^2$	45.65682			표준편차3	6.756983
20								
21				=5*SQRT(H11/B11-(G11/B11)^2)				

(c). (3.5)를 이용한 분산을 구하기 위하여 <표4>의 D4 셀에서 =(A4-C10)^2*B4을 입력한 후

D10까지 채우기 핸들로 복사한다. D11 셀에서 =SUM(D4:D10)을 입력하면 평균과의 편차의 제곱합 2419.811이 나타난다. D15 셀에서 =D11/B11을 입력하면 분산 이 구해진다.

(d). (3.6)식을 이용한 분산을 구하기 위하여 E4 셀에서 =A4^2*B4를 입력한 후 E10까지 채우기 핸들로 복사한다. E11 셀에서 =SUM(E4:E10)을 입력하면 변량의 제곱합 1565731이 나타난다. D11 셀에서 =E11/B11-C13^2를 입력하면 분산 45.65682가 구해진다.

(e). (3.7)식을 이용한 분산을 구하려면 H4 셀에서 =F4^2*B4를 입력한 후 H10까지 채우기 핸들로 복사한다. H11 셀에서 =SUM(H4:H10)을 입력하면 가평균과 계급 값을 이용한 편차의 제곱합 135가 나타난다. D19 셀에서 =5^2*(H11/B11-(G11/B11)^2)를 입력하면 분산 45.65682가 구해진다. 표준편차를 구하려면 H15 셀에서 =SQRT(D15)를 입력하면 표준편차 6.756983이 나타난다.

3.2 확률분포

변수 X 가 취할 수 있는 모든 값이 $x_1, x_2, x_3, \cdots, x_n$ 이고, X 가 이들 값을 취할 확률 $p_1, p_2, p_3, \cdots, p_n$ 이 정해져 있을 때 이 변수 X를 확률변수라 하고, 확률변수 X 가 취하는 값 x_i 와 X 가 x_i 를 취할 확률 p_i 와의 대응관계를 확률변수 X 의 확률분포라 하고 이때, 이 대응관계는 $P(X=x_i)=p_i$ (단, $i=1,2,3,\cdots,n$) 과 같이 식으로 나타낼 수도 있고, 이것을 표로 나타냈을 때 이 표를 확률분포표라 한다. 확률분포는 $0 \le P(X=x_i) \le 1$, $\sum_{i=1}^{n} P(X=x_i)=1$ $P(X=x_i \text{ or } X=x_j)=P(X=x_i)+P(X=x_j)$ (단, $i \ne j$) 인 성질이 있다.

3.2.1 확률변수의 평균과 분산과 표준편차

(1) 확률변수 X 의 평균 또는 기대값을 m 또는 $E(X)$ 라 하면

$$m=E(X)=\sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n \tag{3.10}$$

(2) 확률변수 X 의 평균이 m 일 때, $(X-m)^2$ 의 평균 $E((X-m)^2)$ 을 X의 분산이라 하고, $V(X)$ 또는 $\sigma^2(X)$ 로 나타내기로 하고, 분산의 음이 아닌 제곱근을 표준편차라 하고, $\sigma(X)$ 로 나타내기로 하면,

$$V(X)=\sum_{i=1}^{n}(x_i-m)^2 p_i=\sum_{i=1}^{n} x_i^2 p_i - m^2 \tag{3.11}$$

$$\sigma(X)=\sqrt{\sum_{i=1}^{n}(x_i-m)^2 p_i}=\sqrt{\sum_{i=1}^{n} x_i^2 - m^2} \tag{3.12}$$

(3) 평균, 분산, 표준편차의 성질

$$E(aX+b)=aE(X)+b\ , \qquad V(aX+b)=a^2 V(x) \tag{3.13}$$

$$\sigma(aX+b) = |a|\sigma(x), \qquad V(X) = E(X^2) - (E(x))^2 \tag{3.14}$$

(4) 이항분포

어떤 시행에서 사건 E 가 일어날 확률이 p 라 하자. 이 시행을 n 회 독립으로 반복할 때, 사건 E 가 일어날 횟수를 확률변수 X 라 하면 X 의 확률분포가

$$P(X=x) = {}_nC_x p^x (1-p)^{n-x}, x = 0, 1, 2, \cdots, n \tag{3.15}$$

일 때 이항분포라 하고, $B(n, p)$로 나타낸다. 이 분포의 성질로 평균, 분산, 표준편차는

$$E(X) = np, \qquad V(X) = np(1-p), \qquad \sigma(X) = \sqrt{np(1-p)} \tag{3.16}$$

이다.

3.2.2 Excel을 이용한 확률분포의 계산

흰 구슬 네 개, 푸른 구슬 세 개가 들어 있는 주머니에서 동시에 세 개의 구슬을 꺼낼 때, 나오는 푸른 구슬의 개수를 확률변수 X 라고 하자. 이 때 X의 확률분포를 구하여 보자.

(a). 푸른 구슬이 나오는 개수 x를 B2부터 E2까지 0,1,2,3을 입력한다. 확률을 계산하기 위해 B3 셀에서 =COMBIN(3,B3)*COMBIN(4,3-B3)/COMBIN(7,3)을 입력한 후 채우기 핸들로 E3 셀까지 복사한다. F3 셀에서 =SUM(B3:E3)를 입력하여 합이 1임을 확인한다.

(b). (3.10)식을 이용하여 평균을 구하기 위해 B4 셀에서 =B2*B3를 입력한 후 E4셀 까지 채우기 핸들로 복사한다. F4셀에서 =SUM(B4:E4)을 입력하면 X의 평균 1.285714가 구해진다.

(c). (3.11)의 첫째 식을 이용하여 분산을 구하기 위하여 B5셀에서 =(B2-B8)^2*B3를 입력한 후 E5 셀까지 채우기 핸들로 복사한다. F5 셀에서 =SUM(B5:E5)를 입력하면 X의 분산 0.489796 이 구해진다.

(d). (3.11)의 둘째 식을 이용하여 분산을 구하기 위하여 B6 셀에서 =B2^2*B3를 입력한 후 E6 셀까지 채우기 핸들로 복사한다. F6 셀에서 =SUM(B6:E6)을 입력하면 제곱평균 2.142857이 나타난다. B10 셀에서 =F6-B8^2를 입력하면 또한 X의 분산0.489796이 구해진다.

(e). (3.12)의 첫째 식을 이용하여 표준편차를 구하기 위해 F9 셀에서 =SQRT(SUM(B5:E5))를 입력하면 표준편차 0.699854가 구해진다.

(f). (3.13), (3.14)의 식을 확인해 보기 위하여 $Y=2X+3$인 확률변수 Y의 확률분포표를 만들고 위와 같은 방법으로 계산하면 식(3.13), (3.14)가 성립함을 확인할 수 있다.

<표 5> 확률분포표 및 평균과 표준편차

	A	B	C	D	E	F	G
1	확률분포표						
2	푸른구슬의수: X	0	1	2	3	계	
3	P(X=x)	0.11428571	0.514286	0.342857	0.028571	1	
4	x*P(X=x)	0	0.514286	0.685714	0.085714	1.285714	: 평균
5	$(X-m)^2P(X=x)$	0.18892128	0.041983	0.174927	0.083965	0.489796	: 분산1
6	X^2*P(X=x)	0	0.514286	1.371429	0.257143	2.142857	
7							
8	평균	1.28571429					
9	분산1	$E[(X-m)^2]$	0.489796		표준편차1	0.699854	
10	분산2	$E[X^2]-(E[X])^2$	0.489796		표준편차2	0.699854	
11							
12							

=B2*B3

=SUM(B4:E4)

=F6-F4^2

=SQRT(F6-F4^2)

3.2.3 Excel을 이용한 이항분포의 계산

한 개의 주사위를 7번 던질 때 1의 눈이 나타나는 횟수를 X라 하면 X는 $n=7,\ p=\frac{1}{6}$인 이항분포를 따른다.

(a). (3.15)식을 이용하여 이항분포표를 만들기 위해서 B3셀에서 0 입력 후 Ctrl 키를 누른 상태에서 I3까지 채우기 핸들로 복사하면 확률변수 0,1,2,3,4,5,6,7 이 나타난다. B4셀에서 (3.15)식을 이용해서 =COMBIN(7,B3)*(1/6)^B3*(5/6)^(7-B3)을 입력한 후 I4까지 채우기 핸들로 복사하면 이항분포표가 만들어진다. J4에서 =SUM(B4:I4)을 입력하여 합이 1 임을 확인한다.

(b). 다음으로 Excel에서 주어진 이항분포 함수마법사를 이용하여 확률을 계산하기 위해 B5셀에서 =BINOMDIST(B3,7,1/6,FALSE)을 입력 후 I5까지 채우기 핸들로 복사하면 같은 이항분포표가 만들어진다.

(c). 확률 $P(2\le X\le 5)$를 구하기 위해 B9셀에서 =SUM(D4:G4)를 입력하면 확률 0.33007545가 나타난다. $P(6\le X)$를 구하기 위해 B10셀에서 =SUM(H4:I4)를 입력하면 구할 수 있다.

(d). 이항분포의 평균, 분산, 표준편차는 앞의 확률분포에서 구한 방식으로 구할 수 있고, 또한 식 (3,16)을 이용하여 구할 수 있다.

<표 6> 이항분포표, 평균, 표준편차

	A	B	C	D	E	F	G	H	I	J
1		=COMBIN(7,B3)*(1/6)^B3*(5/6)^(7-B3)								
2	이항분포									
3	X	0	1	2	3	4	5	6	7	계
4	P(X=x)	0.27908165	0.391	0.234	0.078	0.016	0.0019	0.0001	4E-06	1
5	확인P(X=x)	0.27908165	0.391	0.234	0.078	0.016	0.0019	0.0001	4E-06	1
6	X*P(X=x)	0	0.391	0.469	0.234	0.063	0.0094	0.0008	3E-05	1.167
7		=BINOMDIST(B3,7,1/6,FALSE)								
8										
9	P(2≤x≤5)	0.33007545								
10	P(6≤x)	0.0001286								
11	평균1	$\sum X_iP(X=x_i)$:	1.167	=SUM(B6:I6)						
12	평균2	n*p :	1.167	=7*(1/6)						
13		=(B3-C9)^2*B4								
14	$(X-m)^2*P(X=x)$	0.37986113	0.011	0.163	0.263	0.125	0.0276	0.0029	0.0001	0.972
15	$X^2*P(X=x)$	0	0.391	0.938	0.703	0.25	0.0469	0.0045	0.0002	2.333
16		=B3^2*B4		=SUM(B12:I12)						
17	분산1	$E[(X-m)^2]$:	0.972							
18	분산2	$E[X^2]-(E[X])^2$:	0.972	=J13-C9^2						
19	분산3	np(1-p) :	0.972		=7*(1/6)*(5/6)					

3.2.4 연속확률변수 와 정규분포

(1) 연속확률변수

확률변수 X가 어떤 구간의 모든 실수값을 연속적으로 취할 때, X를 연속확률변수라고 한다. 확률변수 X가 어떤 구간 $[a, b]$의 모든 값을 취하고, 이 구간에서 정의된 연속함수 $f(x)$에 대하여

$$[1]\ f(x)\geq 0,\quad [2]\ \int_a^b f(x)dx=1,\quad [3]\ P(\alpha\leq X\leq\beta)=\int_\alpha^\beta f(x)dx \quad (\text{단}, a\leq\alpha\leq\beta\leq b)$$

이 성립할 때 함수 $f(x)$를 확률밀도함라고 한다.

(2) 정규분포

연속확률변수 X 의 확률밀도함수 $f(x)$ 가

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad -\infty < x < \infty \tag{3.17}$$

일 때 정규분포라 하며 X 의 평균은 m, 분산은 σ^2 인 성질이 있고 $N(m,\sigma^2)$으로 나타낸다. 특히, 평균이 0, 표준편차가 1인 정규분포 $N(0,1^2)$ 을 표준정규분포라고 한다.

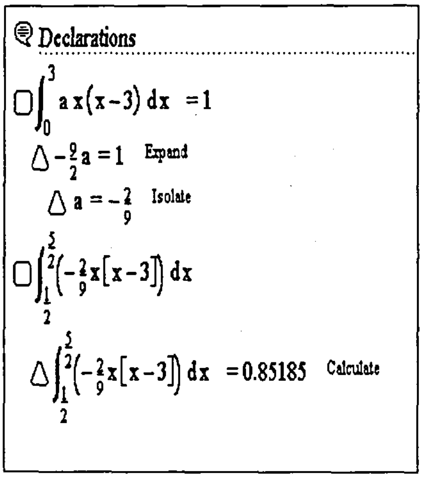
(3) 이항분포와 정규분포 사이의 관계

확률변수 X 가 이항분포 $B(n,p)$를 따르고, n 이 클 때는 X 는 근사적으로 정규분포 $N(np, np(1-p))$를 따른다. (3.18)

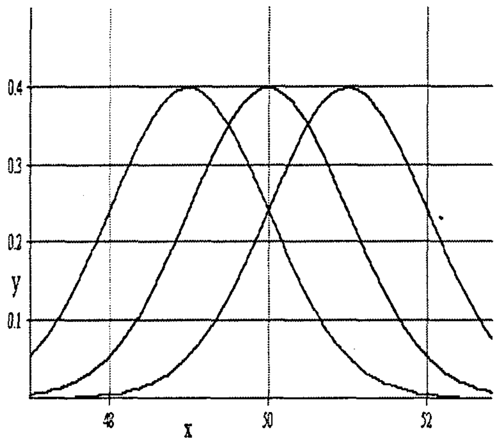
3.2.5 Excel과 MathView를 이용한 연속확률변수와 정규분포

(a) <그림 3>에서 확률변수 X의 확률밀도함수 $f(x)$가 $f(x)=ax(x-3)$ $(0\leq x\leq 3)$으로 주어질 때, a의 값을 구하고 확률 $P(\frac{1}{2}\leq X\leq\frac{5}{2})$의 값을 MathView를 이용하여 구하여 보자. 먼저 도구상자에서 정적분 아이콘( )을 클릭하여. 식과 범위를 입력하고 Expand, Isolate 아이콘을 클릭하면 a가 구해진다. 앞의 확률을 구하기 위해서도 식과 범위를 입력해서 Calculate 아이콘을 클릭하면 된다.

(b) MathView를 이용하여 평균(그림4)과 표준편차(그림5)의 변화에 따른 정규분포곡선의 변화를 MathView의에니메이션 기능을(그림6)이용하여 시각적 연속적으로 보여준다.



<그림 3> MathView로 확률계산



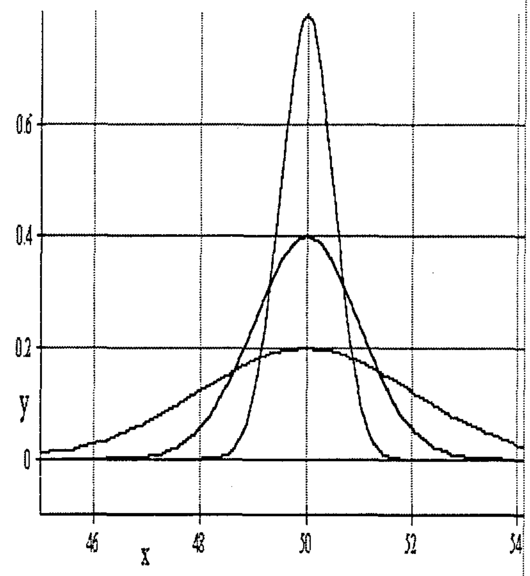
<그림 4> 평균변화에 정규분포 곡선변화

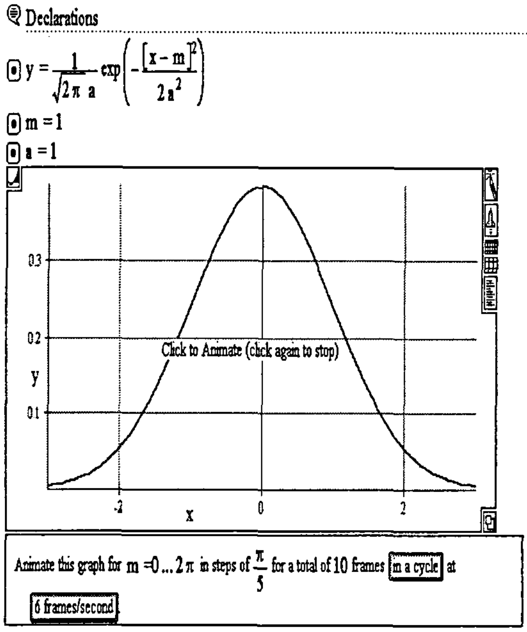
(c) <표 7>에서 어느 고등학교 학생 500명의 키의 분포가 평균이 168, 표준편차가 5인 정규분포 $N(168, 5^2)$을 따른다고 할 때

(1) 몇 명 정도가 160~170 사이에 있을 것인가를 알아보자. 먼저 Excel의 누적 정규분포함수를 이용하여 계산하기 위해 B4 셀에서 =NORMDIST(170,168,5,TRUE) -NORMDIST (160 ,168 ,5,TRUE)을 입력하면 확률 0.601이 나타나고 C4 셀에서

=B4*500 하면 300명 정도가 160~170 사이에 있음을 알 수 있다.

(2) 165 이하인 학생은 몇 명인가를 계산하기 위해 B5 셀에서 =NORMDIST(165,168,5, TRUE)을 입력하면 확률 0.274이 나타나고 C5 셀에서 =B5*500을 입력하면 137명 임을 알 수 있다.





<그림 5> 평균=50, 표준편차 = 0.5, 1, 2 일 때의 정규분포

<그림 6> 평균과 표준편차의 변화에 따른 정규분포 곡선의 에니매이션

(3) 173 이상인 학생은 몇 명인가를 계산하기 위해 B6 셀에서 =1-NORMDIST(173,168 5, TRUE)을 입력하면 확률 0.159가 나타나고 C6 셀에서 =B5*500을 입력하면 79명 정도임을 알 수 있다.

(4) 500명중에서 위에서부터 5%이내 즉, 25등 이내에 들어가려면 키가 몇 이상이어야 하는가를 계산하기 위해 C7 셀에서 =NORMINV(1-0.05,168,5)을 입력하면 176.2 이상임을 알 수 있다.

위의 방법은 Excel의 정규분포함수를 이용했어나, 고등학교 과정에서는 각 변량을 표준화시켜 누적 표준정규분포 표를 이용하여 계산하므로 교과서 방식대로 다시 계산해 보기로 하자.

(5) $P(160 \le x \le 170)$을 구하기 위해 E$ 셀에서 160 와 170을 표준화시킨 후 표준정규분포함수를 이용하는 식 =NORMSDIST((170-168)/5)-NORMSDIST((160-168)/5)을 입력하면 확률 0.601이 나타난다. 나머지는 같은 방법으로 하면 <표7>같이 계산되어진다.

(d) 이항분포에서 시행의 횟수 n 이 크면 정규분포에 근사해 짐을 Exel과 MathView를 이용하여 시각적으로 보여 줄 수 있다. Exel에서 n=10, 20, 30, 50 일 때 이항분포함수를 이용하여 확률분포표를 만든 후 MathView로 가져와서 산점도를 만든다. 각 각의 평균과 분산을 계산하여 그에 해당하는 산점도에 정규분포식을 적고 주메뉴에서 Graph-Additional-Add Line Plot를 클릭하면 정규분포곡선이 아래 <그림 7, 8, 9, 10 >과 같이 나타나 시각적으로 n 이 커지면 이항분포는 정규분포곡선에 근사함을 보여 줄 수 있다.

<표 7> 정규분포 확률계산

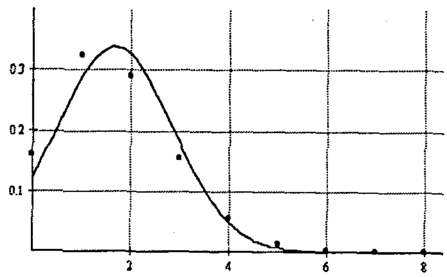
=NORMDIST(170,168,5,TRUE)-NORMDIST(160,168,5,TRUE)

	A	B	C	D	E	F
3	범위	확률	인원	정규범위	확률	인원
4	P(160≤X≤170)	0.601	300.3	P(-1.6≤Z≤0.4)	0.601	300.3
5	P(X≤165)	0.274	137.1	P(Z≤-0.6)	0.274	137.1
6	P(173≤X)	0.159	79.3	P(1≤Z)	0.159	79.3
7	P(x≤X)=0.05		176.2	P(z≤Z)=0.05	z=1.644853	176.2

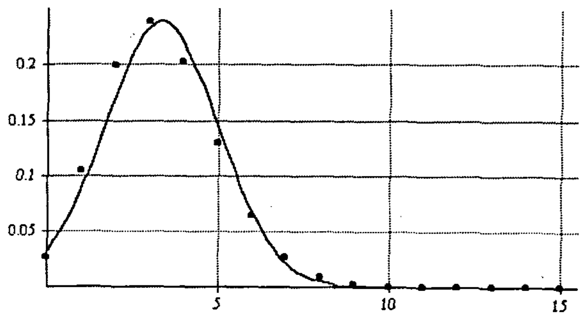
=NORMINV(1-0.05,168,5)

=NORMSINV(1-0.05)*5+168

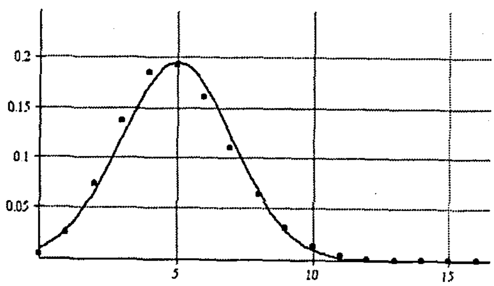
<그림 7> p=1/6, n=10 일 때 이항분포와 정규분포



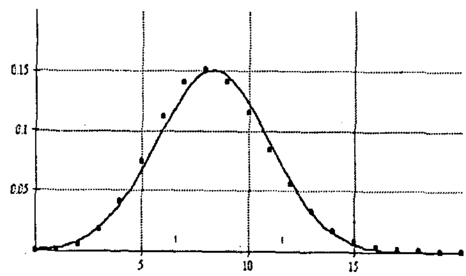
<그림 8> p=1/6 n=20 이항분포와 정규분포



<그림 9> p=1/6, n=30 이항분포와 정규분포



<그림10> p=1/6,n=50 이항분포와 정규분포



3.3 통계적 추측

3.3.1 모집단과 표본, 표본평균의 확률분포, 모평균의 추정

(1) 표본조사에서 조사의 대상이 되는 자료 전체를 모집단이라 하고, 조사하기 위하여 모집단에서 뽑은 자료의 일부를 표본이라고 한다. 어떤 모집단에서 임의추출된 크기 n인 표본을

$X_1, X_2, \cdots, X_n$이라 하고, 그 평균을 $\overline{X}$라고 하면, $\overline{X}$를 표본평균이라 하고, 이 때의 표준편차를 표본표준편차라고 한다.

(2) 모평균 m, 모표준편차 σ인 모집단에서 크기 n인 표본을 임의로 복원추출할 때, 표본평균을 $\overline{X}$라고 하면 $E(\overline{X}) = m$, $\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}$ (3,19) 이다.

(3) 모집단의 분포가 정규분포일 때는 $\overline{X}$의 분포는 정규분포 $N(m, \frac{\sigma^2}{n})$을 따르고, 모집단의 분포가 정규분포가 아닐 때에도 n이 충분히 크면 $\overline{X}$의 분포는 근사적으로 정규분포 $N(m, \frac{\sigma^2}{n})$을 따른다.

(4) 표본에서 얻은 통계적 수치를 이용하여 모집단의 성질을 추측하는 것을 추정이라고 한다. 표본평균 $\overline{X}$를 이용하여 모평균 m을 추정하면

$$95\%\text{의 신뢰구간} \quad [\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}, \quad \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}] \tag{3.20}$$

$$99\%\text{의 신뢰구간} \quad [\overline{X} - 2.58\frac{\sigma}{\sqrt{n}}, \quad \overline{X} + 2.58\frac{\sigma}{\sqrt{n}}] \tag{3.21}$$

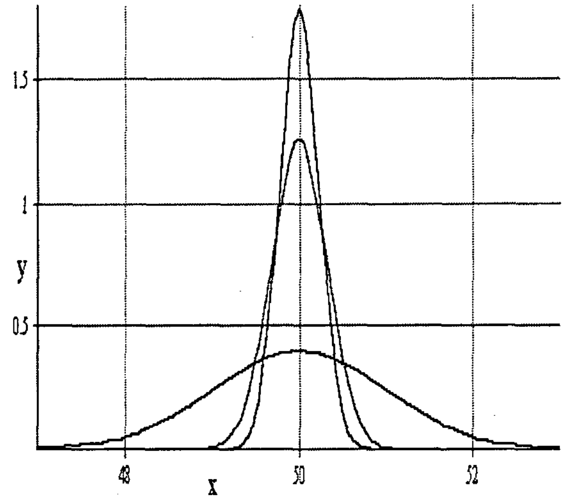
3.3.1 Excel과 MathView를 이용한 통계적 추측

(a) (3.19)식을 이용해 n=1 일 때 평균과 표준편차를 구하여 MathView에서 정규분포식을 입력하고 주메뉴에서 Graph-y=f(x)-Linear을 클릭하여 그래프를 그린 후 n=10, 20 일 때에는 Graph-Additional-Add Line Plot를 클릭하면 정규분포곡선이 아래 <그림 11 >과 같이 나타나 시각적으로 n 이 커지면 표본의 평균은 모평균에 가까이 나타날 확률이 높다는 것을 보여준다. 또한 표본의 개수 n 의 변화에 따른 정규분포곡선의 변화를 MathView의 에니메이션 기능을(그림12) 이용하여 시각적 연속적으로 보여준다.

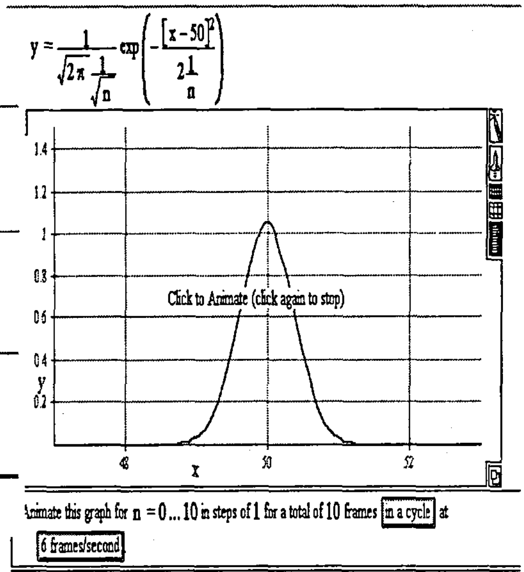
(b) (1). X의 분포가 정규분포 $N(230, 30^2)$을 따른다고 할 때 표본평균을 $\overline{X}$라고 할 때 $P(227 \le \overline{X} \le 236)$과 $P(\overline{X} \le 221)$을 구할 때 <표 8> 과 같이 하면 된다.

(2). 어느 고등학교 학생 중 400명을 임의추출하여 키를 조사하였더니 평균 168cm, 표준편차 10cm 이었다. 이 학교 학생 전체의 키의 평균 m 을 신뢰도 90 %로 추정하는 것은 <표 8 >과 같이 하면 된다.

(3). X의 분포는 표준편차가 10.5인 정규분포를 따르고 있음을 알고 있다. 신뢰도 95 %로 신뢰구간의 폭을 2 이하로 되게 하려면 몇 명 이상을 조사하여야 하는가를 계산하기 위해서는 <표 8>과 같이 하면 된다.



<그림 11> 모평균=50, 모표준편차=1,표본의크기=1, 10, 20에 따른 표본평균 분포함수



<그림 12> 표본의 개수 n 의 변화에 따른 표본평균 분포의 변화 애니메이션

<표 8> 표본평균의 확률, 모평균의 구간추정, 표본의 수

	A	B	C	D	E	F	G
1		=NORMDIST(236,230,3,TRUE)-NORMDIST(227,230,3,TRUE)					
2	P(227≤표본평균≤236) :	0.818595					
3	P(표본평균≤221) :	0.00135					
4			=168+NORMSINV(0.95)*10/SQRT(400)				
5							
6	신뢰도 95% 모평균 추정 :	167.1776	≤ m ≤	168.8224			
7	신뢰도 95% 일때 표본의 수:	423.5364	≤ n				
8							
9	=(2*1/2*1.96*10.5)^2						
10							

4 결 언

본 연구는 Excel을 7차 교육과정의 통계 내용에 무리 없이 도입될 수 있는지 알아보았고, 연속적인 경우는 MathView를 이용하여 자료의 변화에 따른 그래프의 변화를 시각적으로 다양하게 또는 애니메이션으로 보여줄 수 있음을 보였다. 그러나 무엇보다 중요한 것은 수학교육의 목적에 맞게 다양한 교과서와 교재를 개발했더라도 수강 신청하는 학생이 없다면 무슨 소용이 있겠는가.

제 7차 교육과정에서는 고등학교 2학년부터 수학의 6개 과목 모두가 선택과목으로 지정되어 있어

수학이 싫은 학생은 수학과목 모두를 포기하고 졸업할 수 있으며 대학에 진학할 수 있도록 되어있다. 따라서 수능시험도 미국의 수능시험인 AHSME(American High School Mathmatics Examination)와 같이 미분 적분 등 여러 가지 선택과목이 빠진 10단계 교육과정 내에서만 출제될 것이고, 미분 적분 등이 선택과목으로 포함된다 하더라도 일류 몇 개 대학을 제외하고는 수학 선택과목을 요구하지 않을 것이다. 이것은 대학의 필수과목 변화와 편입시험 변화 과정에서 찾아볼 수 있다. 현재 일류대학을 포함한 거의 대부분의 대학에서 자연과학, 및 공과대학에서 조차 수학을 필수에서 선택으로 돌렸고, 편입시험에서는 거의 없어졌다. 대학입시에서 조차 문과 이과가 교차 지원되므로 편입시험에서 수학시험을 요구하면 우수한 학생이 새로 수학준비를 하기가 부담스러워 수학을 보지 않는 학교로 지원하기 때문이다. 이러한 현상으로 볼 때 과연 몇 개 대학에서 고등학교의 선택과목인 수학을 끝까지 요구할 수 있을지 의문스럽다.

따라서 지금과 같은 방식의 교과서와 이론 중심의 수업이 진행된다면 학생들의 반 이상은 수학과목을 회피할 것이고 고등학교 수학교사의 수도 반으로 줄어들 것으로 예상되어진다. 자연과학, 공학, 사회과학의 이론개발 분야로 나갈 학생을 위해 수학교육 목적에 충실한 교과서와 수업도 중요하고 필요하지만, 그 이외의 분야로 나갈 학생들, 또는 10단계 수학에 도달하지 못한 학생들에게는 수학교육의 목적에 다소 미흡하더라도 그래픽 계산기, 컴퓨터 등을 과감히 활용하여 흥미 있게 수업을 진행할 수 있는 교과서와 수업방법이 개발되어 지고 교육부 및 수학교육자들에게 용인되어져야한다. 완전히 포기되어지는 것 보다는 결과만이라도 활용할 능력이 있는 것이 바람직하기 때문이다.

이와 같은 이유에서 제 7차 교육과정의 수학교과서는 지금까지 교육부의 까다로운 심사규정에 의한 똑 같은 내용의 똑 같은 수준의 틀을 벗어나 교과서 저자에 따라 수준 면에서 다양하고, 컴퓨터를 본문에서 도입하는 등 방법 면에서 다양한 교과서가 나와서 각 학교의 실정에 맞게 다양하게 교과서를 선택함으로서 많은 학생들이 수학과목을 선택할 수 있도록 되어야 할 것이다. 특히 통계분야는 많은 자료를 처리해야 하는 실용분야이므로 교과서의 본문에서부터 도입되어야하고 가장 늘리 보급되고 단계적 계산이 가능한 Excel이 바람직하다고 생각한다. Excel을 사용한 열린 수업 과 수준별 수업을 위한 통계수업 모형은 앞으로 연구되어야 할 과제이다.

참 고 문 헌

강옥기 (1998). 그래픽 계산기를 활용하는 수학과 교수-학습 자료 모형 개발 연구, 대한수학교육학회 논문집 8(2), pp.453-474.

곽민정 외 (1999). EXCEL 97 데이터 처리, 서울: 자유아카데미.

김명렬 외 (1998). 고등학교 수학(Ⅰ), 서울: 중앙교육진흥연구소.

문교부 (1997). 제 7차 교육과정; 수학과 교육과정. 서울: 대한교과서주식회사.

송문섭 · 조진섭 (1999). EXCEL에 기초한 통계학 입문, 서울: 자유아카데미.

신동선 외 (1998). 고등학교 수학(I), 서울: 교학사.

안병곤 (2000). 초등학교 수학과 교수-학습에서의 엑셀 프로그램의 활용, 한국수학교육학회 시리즈 E <수학교육 논문집> 10, pp.81-96, 서울: 한국수학교육학회.

허혜자 (1998). Mathematica를 활용한 수학 지도, 대한수학교육학회 논문집, 8(2), pp, 541-550.

N. Scott Hoffner (1997). *MathView User's Guide*, Canada: Waterloo Maple Inc.