

## 탐색적 자료분석과 학교수학에서의 통계지도

김 응 환<sup>1)</sup>

### I. 서 론

21세기 사회는 정보화와 세계화로 그 성격을 예견할 수 있는데, 이에 대비하기 위한 학교수학의 중점은 당연히 단순기능인의 양성보다는 자기 주도적으로 자신의 관리와 지적 가치를 창조할 수 있는 적극적이고 능동적인 인간 양성에 두어져야 한다. 이러한 인간을 키우기 위해 교육부는 현행 6차 수학과 교육과정을 대폭 개선하여 2001년부터 적용할 제7차 수학과교육과정(1997년 12월 31일 고시)을 새롭게 발표하였다.

제7차 교육과정은 6차교육과정의 기본철학을 유지하면서 그 기본방향을 “교육주체를 위한 학습자 중심의 교육과정”으로 정하고 이를 구현하고자 하였다. 특히 학교수학에서의 수학교육은 수학적 힘(mathematical power)을 신장시키는데 두고 있다. 이 수학적 힘이란 문제해결력의 신장의 관점보다는 보다 광의의 개념으로 정보화 사회에서의 수학을 이용한 정보이해능력, 획득한 정보가 타당한지 여부를 가리는 판단력과 의사결정력, 수학을 사용한 다른 사람과의 의사소통능력, 실생활이나 다른 교과영역에 수학적

지식을 적용하여 문제를 구성하고 이를 해결하는 문제해결능력 등을 포함하고 있는 것을 말한다.

이러한 주변 교육상황의 변화는 일선에서 학생들을 직접 가르치는 중등학교의 수학교사에게 새로운 변신을 요구하며 교재연구와 교육방법에 대한 탐구를 기대하게 한다. 특히 확률통계 영역의 교육내용은 그야말로 불확실성이 점점 확대되고 있는 미래의 정보화 사회에서의 합리적인 판단력과 정보처리능력의 신장에 초점을 두어 지도해야하는 중요성은 아무리 강조해도 지나치지 않을 것이다.

전통적인 통계지도에서는 과도하게 수치적으로 편향되어 있는 반면에 현대적인 통계지도 면에서 자료분석은 시각적 표현인 그래픽적 해석에 많이 의존하고 있다. 이러한 이유는 오늘날 컴퓨터의 하드웨어와 소프트웨어적 과학 발전에 힘입은 바 크다고 하겠다. 즉 자료분석의 방법론은 컴퓨터 환경에 맞추어 매우 다양하고 풍성하게 발전하고 있다.

그래서 본 연구에서는 중등학교의 확률통계지도 영역에서 자료분석 단원의 지도에 초점을 맞추고 학생중심의 입장에서 통계지도의 학습보조자료로서의 사용과 수행평가를 할 수 있는 내용을 다루고, 학생의 호기심을 불러일으키며 상상력을 자극하고 스스로 창조적 발상을 할 수 있도록 도와줄 수 있는 몇 가지 방법론을 제안하고자 한다. 특히 중등수학교육수준의 학교수학에서 탐색적 자료분석과 수학교육과의 관계를 살펴보고 학

1) 공주대학교 수학교육과

e\_mail : yhkim@knu.kongju.ac.kr

생들로 하여금 확률통계단원을 학습하는 경우 보다 창의적이고 발견적인 자료분석을 통한 통계학습에 임할 수 있도록 하는 여러 가지 탐색적 자료분석방법을 소개하여 학교수학의 수업방법과 내용에 풍성함을 제공하는데 그 목적이 있다.

본 연구의 내용으로는 2장에서 현재의 수학교육과 자료분석을 검토한다. 3장에서는 줄기와 잎 그림, 상자그림, 별 그림, 얼굴 그림 등을 다룬다. 결론에서는 탐색적 자료분석과 수학교육에 대한 끝맺음을 한다.

## II. 학교수학과 자료분석

현행 대부분의 중등학교에서는 여전히 과밀한 다인수 학급에서 특히 능력차이가 다양한 학생들이 함께 교육을 받고 있는 현실이며, 이를 극복하고자 수준별 이동수업이 다소 무리하게 추진되고 있는 실정이다. 이러한 현실이 갑자기 개선되기는 어려울 전망이다. 하지만 점진적으로 개선되는 속도가 빨라질 것이다.

제7차 수학과 교육과정은 초등학교 1학년부터 고교 1학년까지 10년간을 국민공통 기본교육기간으로 정하고 이 기간 동안 10단계의 단계형 수준별 교육과정을 운영한다. 이 개정의 기본방향을 수학적 힘의 신장으로 정하고 이를 구현하기 위하여 다음과 같은 실천항목들을 제시하고 있다.

1) 개인의 능력수준과 진로를 고려한 수학교육

고교 2, 3학년에서 선택형 과목으로 미분적분학, 확률통계, 이산수학, 실용수학을 제시하고 있으며 계열에 구분없이 자기의 적성과 능력, 진로에 맞게 수학과목을 선택적으로 학습할 수 있다.

2) 수학의 기본지식을 갖게 하는 수학교육

수학 지식의 기본적 개념과 원리를 충실하게 전달하여야 한다.

3) 학습자 활동을 중시하는 수학교육

학생 스스로 실제 상황을 수학적 모델로 바꾸고 수학적 원리나 법칙을 탐구 발견 창안하며, 자기 주도적으로 수학을 실생활의 문제해결에 활용할 수 있도록 가르친다. 학생들은 소집단 별로 다른 학생과 토론을 충분히 할 수 있도록 하고, 개별적 학습활동도 할 수 있도록 해야한다.

4) 수학교육에 흥미와 자신감을 갖게 하는 수학교육

수학교육에 흥미와 자신감을 갖게 하기 위하여 수학의 내적, 외적 가치와 필요성을 인식시킬 수 있도록 수업내용전개와 학습활동이 주어져야한다.

5) 계산기 컴퓨터 및 구체적 조작물을 학습 도구로 활용하는 수학교육

계산기와 컴퓨터의 학습도구로의 사용이 적극적으로 이루어져야 한다. 특히 현장교사들에 의해 만들어진 구체적 조작물의 학습보조자료의 사용이 권장되어야 한다.

6) 다양한 교수학습방법과 평가방법을 활용하는 수학교육

교사중심의 설명식 학습과 소집단활동, 개별화 학습 등 다양한 교수학습방법이 조화를 이루어 효율적인 수업이 되도록 하고, 평가도 학습지도의 일부가 되어야 한다. 특히 진단평가, 형성평가, 총괄평가 외에 집단학습평가, 수행평가, 포트폴리오 평가 등이 적절히 사용되어야한다.

위와 같은 관점에서뿐만 아니라 현재의 학교수학 수업에서 새롭게 확률통계단원의 효율적인 지도방법을 구안해야할 필요를 느끼게 된다.

통계의 영역에 대한 것은 일반적으로 두 가지로 구분된다. 하나는 고전적 수리통계와 최근에 개발되고 있는 현대적 자료분석이다. 예를 들면 수리통계(mathematical statistics)는 주로 특정한 모델에서의 모수의 추정과 숨겨져 있는 모집단의 비밀을 찾아서 먼저 결정된 유의수준에서의 통계적 가설검정을 포함하는 공식적인 통계적 추론을 주로 의미

한다. 이와는 대조적으로 자료분석(data analysis)은 데이터 속에서 통찰을 얻어내고자 하는 목적이 있다. 이것은 자료몽치로부터 정보내용을 요약하거나 예상할 수 없었던 현상을 발굴해 내는 것을 의미한다. 광의의 새로운 용어로 자료채굴(data mining)을 의미한다. 이러한 발견적 과정은 자료의 탐색적 분석과 밀접한 관련이 있다. 그리고 모델은 그들의 평가의 관점과 마찬가지로 본질적으로 자료로부터 출발되며, 자료에서 찾고자 하는 것이 밝혀지고 새롭게 얻어지는 결과물인 것이다.

현재의 중등학교 통계자료분석 단원에서의 지도항목들을 살펴보면 그 범주가 오랫동안 변하지 않고 있으며 상당히 제한되어 있음을 알 수가 있다. 그 예를 살펴보면 자료의 요약으로서 막대그래프, 히스토그램, 도수분포표, 원그림 등이 그 지도 항목에 들어있다. 이들이 중요하지만 융통성이 있고 창의적인 자료의 분석요약방법의 표현이 개발되고 있음을 알려줄 필요가 있다.

그래서 다음 장에서는 재미있고 아무라도 조금만 관심을 기울인다면 배울 수 있는 탐색적 자료분석 방법을 몇 가지 소개하고 학생과 교사들이 수학교육에 흥미와 관심의 폭을 넓힐 수 있고 실생활에 용이하게 사용할 수 있는 아이디어를 제공하고자 한다.

### Ⅲ. 탐색적 자료분석

존 투키(John Tukey, 1977)는 자료분석의 숲을 다음 두 가지로 구별하고 있다. 그 하나는 탐색적 자료분석(exploratory data analysis)으로, 이것은 주로 통찰을 위한 자료의 탐색을 의미하고, 다른 하나는 확증적 자료분석(confirmatory data analysis)으로, 이것은 어떤 의미를 찾아서 얻어진 통찰에 대한 증거의 강력함을 강조하는 것이다. 이러한 자료분석은 수학이라고 하기보다는 실

제 현상에 대한 실험과학(empirical science)에 가깝다고 말할 수 있다.

그래픽은 왜 유용할까? 인간은 숫자들 모임으로서의 패턴보다는 시각적 패턴(visual patterns)을 보다 손쉽게 인식한다. 그림은 우리가 미처 자료에서 발견하기를 기대하지도 않았는데 드라마틱하게 어떤 것을 밝혀주기 때문에 탐색적 자료분석은 특별히 중요한 것이다. 이와는 대조적으로 수치적 요약은 취급하기 좋게 설계되지 않으면 자료의 숨겨져 있는 어떤 구조를 그대로 덮어 버릴 수 있는 경향이 있다.

통계그래픽은 과학과 예술의 극적인 만남이라고 말할 수 있다. 최근에는 소비자 중심의 다양한 정보를 처리하기 위하여 여러 가지의 그래픽 방법이 새롭게 고안되고 있다. 다음은 그래픽 방법을 이용한 탐색적 자료분석의 몇 가지 예를 들어 학교수학의 교수학습 방법의 개선에 아이디어를 제공하기로 한다.

#### A. 줄기와 잎 그림

학교수학에서 자료분석 방법으로 도수분포표와 히스토그램을 이용하여 어느 집단의 속성을 파악하고 자료를 요약하는 방법을 배우고 있다.

줄기와 잎 그림(stem and leaf plot)은 수치자료의 분포특징을 파악하기 위하여 사용되는 아주 유용한 그래픽 자료분석 방법으로 히스토그램(histogram)의 내용과 아주 유사하다. 이는 미국의 통계학자 Tukey(1977)에 의해 개발된 것이다. 이 절에서는 줄기와 잎 그림의 작성방법에 대하여 논하기로 한다.

다음 [표 1]에서와 같이 공주시의 한 고등학교 2-4반 학급의 5월 모의고사 수학 성적이 얻어졌다. 이 자료에 대한 줄기와 잎 그림의 작성방법은 다음과 같은 절차를 밟는다.

[표 1] 모의고사로 측정된 수학점수

68 77 80 67 53 36 61 64 52 44 56 43  
 74 57 37 61 70 59 61 67 68 62 68 52  
 80 47 64 53 78 67 52 43 38 51 77 57  
 61 67 68 75 48 42 57 57 56 63

- 1) 측정된 자료의 값을 우선 전체적으로 읽어본다. 자연스럽게 최대 최소값을 발견할 수 있고 자료의 수치에 대한 전반적인 특징을 파악한다. 이 경우 수학점수들은 36에서 80점 사이의 자연수로 구성되어 있음을 알 수 있다. 그리고 첫째 줄의 4개의 점수 값을 읽어보면 68 77 80 67 이다.
- 2) 이 값들을 적당히 다음과 같이 줄기와 옆으로 분리한다.

$$\begin{aligned} \text{줄기} & \quad \text{옆} \\ 68 &= 6 \times 10 + 8 \\ 77 &= 7 \times 10 + 7 \end{aligned}$$

- 3) 종이의 왼쪽에 열을 맞추어 줄기값의 최소값과 최대값 사이의 숫자를 빠뜨리지 않고 써 넣는다. 그리고 이 열 옆에 세로줄을 긋는다. 세로로 그은 수직선의 오른쪽에 옆의 값을 하나씩 순서대로 기입한다.

9		678
8		437382
7		3267923217776
6		7814117828471783
		740875
		00
N		46

이 경우에 줄기 6에 옆이 8 과 7이 달린 순서는 관찰된 순서일 뿐이다. 일차로 분석이 끝난 후에 서열을 정렬(sorting)할 수도 있다.

- 4) 이와 같은 과정을 모든 관측한 자료값에 대하여 반복 실시한다.
- 5) 줄기그림이 완성되면 각 줄기의 도수를 파악하여 도수의 합계가 총도수와 일치하는지를 확인한 다음 일치하면 줄기의 옆 왼쪽

에 각 줄기의 도수를 기록한다.

보조적으로 누적도수를 각 줄기에 기록할 수도 있다. 자료의 중위수(중앙값)를 포함하는 줄기에는 깊이 대신 도수를 괄호로 표시한다. 이와 같은 작업순서로 다음의 [그림 1]과 같은 줄기와 옆 그림의 그래프가 얻어진다.

3	3	3		678
9	6	4		437382
22	13	5		3267923217776
16	16	6		7814117828471783
8	6	7		740875
2	2	8		00
N	46			

[그림 1] 수학성적의 줄기와 옆 그림

일단 작성된 줄기와 옆 그림에서 줄기의 수를 늘리거나 줄일 수가 있다. 즉 각 줄기를 합할 수도 있고 한 줄기를 분할 할 수도 있다. 이를 조정함으로써 가장 보기 좋은 모양의 그림을 얻을 수 있기 때문이다. 줄기와 옆 그림은 줄기의 수에 따라 상당히 달라진다는 사실을 알 수 있다.

한 자료에 대하여 여러 가지 형태의 줄기와 옆 그림이 얻어질 수 있는데 어느 그림이 좋다고 할 수 있을까. 이것의 대답은 사용용도와 자료의 성격을 고려하여 실험자가 자유롭게 결정하면 된다. 시행착오가 따르더라도 여러 개의 줄기와 옆 그림을 그려보아 자료의 특성이 어떻게 다르게 나타나는가를 검토하여보고 이를 토대로 비교하여 적절한 것을 택하는 것이 자료분석의 태도이다.

줄기와 그림의 역할을 하는 전통적인 방법은 히스토그램이다. 이 히스토그램과의 비교를 통한 줄기와 옆 그림의 특징을 살펴보면 다음과 같다.

- 1) 공통점은 전체적인 길모양은 동일하다는

것이다. 어느 것이나 그 구간에 도수에 비례하는 막대기등을 가지고 있다.

2) 차이점은 줄기와 잎 그림이 자료분포의 대략적인 형태를 보기 위하여 작성되는 그래프로서 히스토그램에 비하여 정보의 손실이 없다는 것이다. 그 이유는 줄기와 잎 그림에서는 구간내의 자료들이 숫자 그 자체로 생생하게 구별이 되어 있는데 반해 히스토그램에서는 그 구간에서의 도수의 크기만을 보여준다. 이것은 히스토그램은 자료가 갖고 있는 특성을 축약하는 과정에서 정보의 손실이 발생한다는 것을 알 수 있다. 예를들어 히스토그램에서는 중위수를 구할 경우 그 값이 속하는 구간의 보간법을 통해 근사값을 계산하지만 줄기와 잎 그림에서는 정확한 중위수를 구할 수 있다.

3) 줄기와 잎 그림인 경우 줄기의 수를 줄이거나 늘이는 데에 원래 작성된 표를 이용할 수 있으나 히스토그램에서는 다시 원자료(raw data)에 의존하여 도수분포표를 새로 만들 수밖에 없다. 탐색적 자료분석에서 한 자료에 대해 여러 번 히스토그램이나 줄기와 잎 그림을 만들어보는 시행착오를 하는데 있어서의 효율성에 차이가 있다.

4) 히스토그램에서는 계급 구간의 폭을 임의로 할 수 있으나 줄기와 잎 그림에서는 구간 폭에 있어 그다지 자유롭지 못하다. 구간폭은 반드시 1, 2, 5의  $10^d$ (d는 정수)이어야 하기 때문이다.

5) 히스토그램에서는 자료의 크기가 관계가 없으나 줄기와 잎 그림에서는 대량의 자료에 대해서는 분석하기가 쉽지 않은 것이 단점이다.

다음은 실제로 한 학급의 학생들로부터 얻을 수 있는 여러가지의 자료 중 하나인 각 학생들의 생년월일을 기초자료로 하여 직접 줄기와 잎 그림의 연습을 위한 수업자료로서 활용하여 보았다.

[표 2] 2-4반의 학생들의 생년월일

1/11	2/22	1/20	1/5	2/16	3/20	4/30
4/27	12/04	11/29	11/10	11/16	10/26	7/15
7/7	7/17	6/27	6/20	6/12	8/3	8/23
11/26	11/15	4/10	5/28	5/27	6/25	6/11
6/21	6/9	1/27	2/12	3/1	2/28	4/23
9/16	9/25	4/7	5/6	7/24	8/2	9/23
9/29	12/23	4/1	2/16			

이미 학생들은 자기 학급의 수학적 자료를 이용하여 줄기와 잎 그림을 그리는 방법을 숙지하고 있다. 여기서 위 [표 2]와 같은 구체적인 자기학급의 생년월일 자료를 연습자료로서 열린 문제로 제공하면 학생들은 새로운 문제상황에 접하게 되며 도전의욕을 갖게된다.

이때 교사는 문제 자료의 제시를 해주고 안내심을 가지고 기다리면서 새로운 발상을 학생 스스로 할 수 있도록 충분한 자유시간과 기회를 주고 독창적인 자료분석을 발견을 유도해야 할 것이다. 각 학생들의 여러 가지 궁리들이 오고간 뒤에 자기의 분석 결과물을 만들고 서로 간의 그림을 비교 토의할 수 있도록 하면 좋다.

이 자료의 줄기와 잎 그림은 다음 [그림 2]와 같다. 여기서 줄기는 달(월)이고 잎은 날짜가 된다. 특히 잎은 두 자리수가 하나의 개체이므로 콤마로서 도수를 구분하였다.

4	1	11, 20, 05, 27
5	2	22, 16, 12, 16, 28
4	3	20, 01, 23, 07
4	4	30, 27, 01, 10
3	5	06, 27, 28
7	6	11, 21, 25, 27, 20, 12, 09
4	7	15, 07, 17, 24
3	8	03, 23, 02
4	9	16, 25, 23, 29
1	10	26
5	11	29, 10, 16, 26, 15
2	12	23, 04
N=46		

[그림 2] 생년월일의 줄기와 잎그림

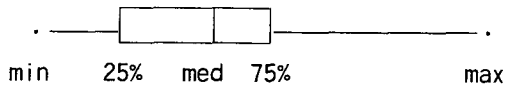
[표 2]의 관찰값을 토대로 이 학급의 생년월일에 대한 분포를 줄기와 잎 그림인 [그림 2]로 살펴볼 수가 있으며 이는 각 학생의 생일 날짜가 살아있어 정보손실이 없이 매우 생생하게 자료를 요약한 장점을 충분히 알 수가 있다. 위 [그림 2]에서도 군집(cluster)의 존재여부와 자료의 집중도가 높은 구간, 대칭성의 여부, 자료의 범위와 산포, 특이점의 존재여부 등을 쉽게 파악할 수 있다. 즉 6월에 태어난 학생이 제일 많으며 10월을 제외하고는 거의 같은 비율로 분포되어 있음을 알 수 있다. 특히 각 자료가 자료의 정보를 보존하고 있기 때문에, 히스토그램에서는 알 수 없는 것으로 2월에 생일이 같은 학생이 있음을 쉽게 발견할 수 있다. 이 예는 각 학생의 신상과약과 인성교육에도 이 줄기와 잎 그림의 자료분석방법을 이용하면 매우 유익함과 함께 학생들의 발견적 욕구를 채워줄 수 있어서 통계의 자료분석에 대한 동기유발과 함께 재미있음과 실용성을 느끼게 할 수 있다.

B. 상자그림(box plot)

자료분석시에 줄기와 잎 그림은 각 자료의 모든 것을 펼쳐놓아야 하지만, 때때로 우리는 각각의 구조와는 달리 전체적인 자료의 분포에 대한 인상을 얻고자 할 때가 있다. 이러한 목적을 위하여 개발된 것이 상자그림(box plot)이다. 이것을 알아보기로 한다.

상자그림은 다음 [그림 3]과 같이 x축을 따라 상자를 길게 그린 그림이다. 이 상자그림의 아이디어는 자료의 50%를 중점적으로 표현하기 위한 상자를 그림으로써 자료의 전체적 분포와 자료의 위치를 파악하려는 것이다. 이 상자그림에는 5가지 척도(최소값, 25%지점수, 중위수, 75%지점수, 최대값)를 포함하고 있다. 이때 여기서 필요한 5가지 수치는 앞에서 배운 줄기와 잎 그림에서 쉽

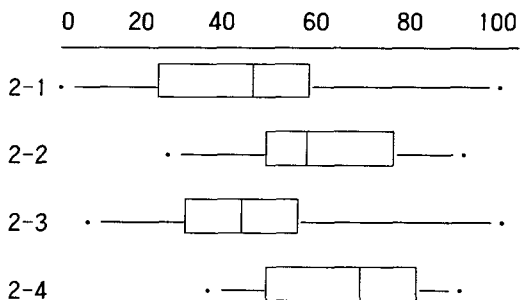
게 얻어질 수 있는 값들이다. 전체 자료의 대부분이 어떻게 분포되어 있는가를 확인 할 수가 있으며 여러 집단의 상호비교에 유익한 그래픽방법이다.



[그림 3] 상자그림

우리들은 자료의 모임으로부터 자세한 나무가 아니라 전체적인 숲을 보고 싶은 경우도 종종 있다. 특히 자료의 여러 개 집합을 비교할 필요가 있을 경우가 바로 이 경우이다.

예를 들어 여기 어느 중학교의 2학년 4개 학급의 중간고사 수학성적이 얻어졌다고 할 경우에 이들 각 학급간의 성적을 비교하고 싶다고 하자. 보통은 평균과 표준편차만을 가지고 비교하게 되는데 이것은 자료분석의 목적 상 다소 건조하지 않을까 생각된다. 이때 다음과 같은 [그림 4]처럼 각 학급의 성적을 비교해본다면 순간적으로 이 학교 2학년 전체의 수학성적의 개요를 쉽게 비교 파악할 수 있을 것이다.



[그림 4] 4개 학급의 수학성적

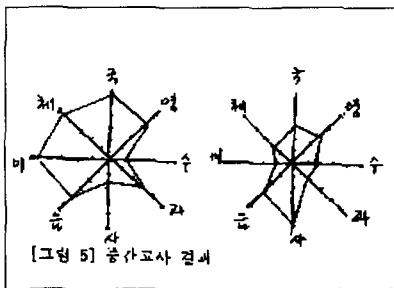
C. 별그림(star plot)

보통 우리는 변수를 1, 2개를 취급하였으나, 별그림은 여러 개의 변수를 함께 취급하여 각 개체의 특징을 비교하는 방법이다. 그리는 방법은 우선 변수의 개수만큼 직교좌표축을 바탕으로 축을 그리고, 전체자료에서 최대 최소값을 비교하여 축의 눈금을 적절하게 결정한다. 그리고 각 변수의 값을 점을 찍은 후에 서로 선으로 연결하여 그림을 그린다. 예를 들어 다음 [표 3]

[표3] 개인별 중간고사성적

학생	국	영	수	과	사	음	미	체
1	수	우	가	미	양	우	수	수
2	미	미	양	양	수	우	가	양

과 같은 2명의 중간고사 성적이 개인별로 얻어진 결과를 별그림으로 그려보자.



위 [그림 5]에서 1번 학생의 전체적 성적을 알 수 있으며 수학이 매우 저조함을 볼 수 있다.

D. 얼굴 그림(face plot)

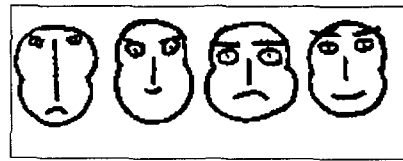
중등학교의 학교수학에서 통계의 그래픽 지도내용으로는 막대그래프, 히스토그램, 산포도, 원그래프, 꺾은선그래프 등이 있다. 이들은 대체로 1차원이나 2차원의 자료에 대한 요약물 다루고 있다. 그러나 오늘날에는

정보를 처리하고자 하는 요구가 크게 일어나고 있으며, 다차원의 자료를 2차원으로 축약하여 묘사하는 방법은 없을까 하는 궁리를 하게 된다.

얼굴 그림은 다차원의 자료를 2차원의 얼굴 모습에 표현하는 방법이다. 문득 이런 자료분석을 하려면 컴퓨터를 사용해야겠다는 생각이 들지도 모르나, 얼마든지 손으로 손쉽게 나타낼 수가 있다.

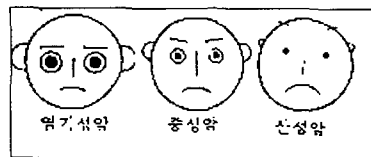
이 얼굴 그림을 고안한 사람은 러시아의 체르노프(chernoff. H)이고, 그가 이 그림을 고안하게 된 동기는 “인간의 마음에 대한 적응성은 사람의 얼굴을 접했을 때에 최대한 발휘된다.”라는 신념에 근거한 것이었다.

다음 얼굴들은 체르노프의 포로야구의 피쳐의 성적 기록을 나타낸 것이다. 이 그래프는 컴퓨터로 그려진다.



[그림 6] 프로야구 선수의 성적

다음 얼굴들은 일본 게이또 대학의 키로우 선생에 의해 고안된 얼굴로 암석의 자료를 바탕으로 컴퓨터로 그린 것이지만 체르노프의 얼굴보다 부분적으로 단순한 곡선이 이용되고 있어서 친하기 쉬운 얼굴을 하고 있다.



[그림 7] 암석의 특성

이상과 같이 얼굴 그림을 그리는 데는 특별한 약속은 없고 각 용도에 따라 자유로운 발

상으로 그릴 수 있다는 것이다. 또한 자연스럽게 다차원의 자료가 2차원의 얼굴 그림으로 자연스럽게 표현되어 있으며 그 얼굴들로부터 전체적인 인상이 받아들여지는 느낌을 가질 수 있다. 나름대로 이 얼굴그림의 장점을 서술해 본다면

- 1) 얼굴의 연속적인 변화를 통하여 자료의 특성을 나타낼 수 있다.
- 2) 다차원의 자료를 1, 2차원으로 차원축소를 할 수 있다.
- 3) 시각적으로 강렬한 인상을 줄 수 있다.
- 4) 유사한 얼굴끼리의 군집을 만들 수도 있다.

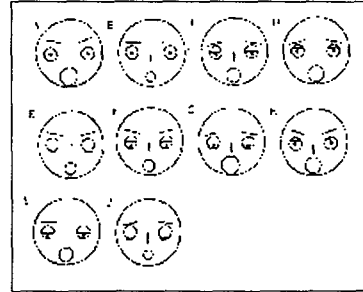
앞의 얼굴 그림을 보면서 현재 우리의 학교 수학의 통계그래픽에 대한 지금까지의 지도는 조금은 고정적이 아니었던가 생각이 든다. 자료의 요약과 정리의 지도에서 부여된 그래프의 역할을 고려해본다면, 그래프의 본질은 자신의 의도를 소비자에게 효과적으로 전달하고 상대방이 알고 싶은 정보를 얻도록 도와주는 도구로 생각할 수 있다. 이러한 측면을 주목한다면 실제 통계수업 시간에 정말 여러 가지의 그래프가 지도되는 것도 좋은 일이라 생각된다.

다음의 예는 어느 중학교의 10명의 학생에 대한 학기말 성적(수, 우, 미, 양, 가)을 바탕으로 얼굴그림을 표현한 결과물이 [그림 8]이다.

[표 4] 학기말 성적표

학생	A	B	C	D	E	F	G	H	I	J
수학	수	수	미	우	가	미	양	우	양	가
음악	수	가	미	우	양	양	수	우	미	가
미술	수	미	가	양	우	우	양	수	미	가

(수학=눈, 음악=입, 미술=눈썹)



[그림 8] 학기말 성적

이상과 같이 얼굴그림은 각 학생의 전체성에 대한 얼굴 인상을 대응시키고 있어서 너무 재미있다. 즉 노래를 잘하는 사람은 입이 커다랗고, 수학을 잘하면 두눈이 또랑또랑하고, 수학을 못하면 졸린 눈을 하고 있는 것을 볼 수가 있으며, 성적이 우수한 학생들의 얼굴형을 따라 군집으로 분류할 수도 있다.

실제로 수업시간에 얼굴그림의 지도에서 학생들의 발상을 제한시키면 효과가 줄어들 수 밖에 없다. 여기서의 강조점은 자기의 표현을 자기 마음대로 마음껏 할 수 있는 기회를 학생들에게 제공하고 충분히 집단적으로 토의하고 이 그래프를 통하여 서로의 친구들과 분석하고자 하는 자료에 대한 의사소통과 다른 학생의 표현을 비교하여 자기와의 차이를 인정하면서 여러 장점을 마음에 새기도록 하면 유익할 것이다.

## IV. 결 론

결론적으로 이러한 자료분석을 수학교육의 학습지도로 어떻게 연결할 수 있을까?

첫째, 탐색적 자료분석 즉 실제 그래픽 분석방법은 어떤 상태를 발견하는 세계 즉, “아하!”를 통계 속으로, 수학교육 속으로 안내할 수 있다. 평균과 표준편차를 계산하고 유의수준을 정하고 추정을 행하고 공식을 통



하여 노력하는 것보다 학생들에게 흥미를 자극할 수 있고 창조적 상상력과 합리적 분석력을 기를 수 있게 한다. 지금까지는 확률과 통계를 학교수학에서 지도하면서 입시 때문에 과도하게 확률이론과 순열조합에 치중하여 지도하는 경향이 있어서 통계하면 매우 난해하고 골치아픈 분야라는 왜곡된 인식과 기억을 가지고 있었던 것이 사실이다. 앞서 제시한 여러 탐색적 자료분석을 학교수학의 수업시간에 적극적으로 활용함으로써 통계는 보다 유익하고 쉬우며 실생활에 아주 가까이 있음을 인식할 수 있는 계기가 될 수 있을 것이다.

둘째, 탐색적 자료분석은 자연스럽게 응용분야인 확증적 통계영역으로 인도할 수 있기 때문에 그래픽을 이용한 탐색적 자료분석 방법은 사전에 가르쳐지거나, 적어도 고전적 통계와 병행해야 한다. 학생들이 좀더 통계의 이론적인 면을 배우기 전에, 추상대수 전에 계산과 연습을 공부하는 것처럼, 데이터를 다루고 자료분석 작업을 직접 참여해봄으로써 자료에 대한 느낌과 감각을 얻을 수 있게 될 것이다.

셋째, 탐색적 자료분석에서 사용되는 수학은 그리 어려운 수학이 아니기 때문에 고전적인 통계가 요구하는 수준보다 낮은 수학적 배경을 가진 사람들에게도 가르쳐 질 수가 있다. 그래픽방법에 대해서도 이것은 사실이다. 현재의 학교수학의 교육영역에서 학생들이 그래프를 사용하거나 만드는데 교육환경의 어려운 문제는 어느 정도 해소되었다. 즉 조금만 노력하면 가까이에서 그래픽을 이용하는 소프트웨어를 접할 수 있기 때문이다. 초등학교나 중등학교에서 이러한 주제는 좀더 일찍 조기에 교육되어도 무방하다. 중등학교 고학년 수준의 수학교육에서도 수학을 좋아하지 않는 학생들을 위해서도 가르쳐질 수 있다고 생각한다. 그리고 대학을 뛰어넘어서 비즈니스나 경영분야의 일반인들에게도 가르쳐질 수 있다.

끝으로 중등학교에서 자료분석의 지도에

서 실제 수업을 통한 여러 가지 효율적인 자료정리와 요약에서의 통계 그래픽의 교수학습지도방법의 개발을 다음의 연구과제로 남긴다.

## 참 고 문 헌

- [1] 교육부(1995). 제 6차 수학과교육과정의 해설.
- [2] 교육부(1997). 제 7차 수학과교육과정, 교육부 고시 제 1997-15호[별책 8], 대한교과서주식회사.
- [3] 구광조,전평국,류희찬(1993), 수학교육과정과 새로운 평가방향, 서울, 경문사
- [4] 김승동, 박달원(1993). 제6차 교육과정에 관한 고찰. 공주대학 논문집.
- [5] 김응환, 이석훈(1995). 통계교육의 발전을 위한 제안. 충남과학연구지 제22권 제2호.
- [6] 박기석, 김응환, 유승욱(1998). 고등학교 통계수업에서 줄기그림의 활용에 관한 연구, 과학교육연구, 제 29집, 공주대학교 과학교육연구소.
- [7] 박세희, 정광식, 강병개(1995). 고등학교 수학1 교과서, 동아서적(주)
- [8] 백운봉, 허명희 (1987), EDA 탐색적 데이터분석, 서울, 박영사
- [9] 송혜향, 김동재(1995). 통계학. 청문각.
- [10] 허명희, 이태림(1993), 탐색적 자료분석, 방승통신대학교 출판부
- [11] 竹内善和(1983), 顔形グラフを用いた統計指導, 數學教育, 明治圖書, No. 287
- [12] Hoaglin,D.C., Mosteller, F., and Turkey,J.W.(1982), Understanding Robustness and Exploratory Data Analysis, Wiley, New York.
- [13] Tukey, J.W.(1977) , Exploratory Data Analysis, Addison-Wesley, MA
- [14] Velleman,P.F., and Hoaglin,D.C.(1981), Applications, Basics and Computing

of Exploratory Data Analysis,  
Duxbury Press, Boston.

- [15] Watts, D. G.(1991), Why is introductory statistics difficult to learn? and what can we do to make it easier ?, The American Statistician, November. Vol.45. No.4.

## Exploratory Data Analysis and Teaching of Statistics in School Mathematics

Yung-Hwan Kim<sup>1)</sup>

### ABSTRACT

This paper will present some basic and simple graphical methods of exploratory data analysis for the instrument of data analysis at school mathematics. Human beings perceive visual patterns more readily than patterns in collections of numbers. This is especially important in exploratory data analysis because pictures dramatically reveal things that we did not expect to find in the data set.

Here are graphical methods as the stem and leaf plot, the box plot, the star plot and the face plot. These methods impulse the motivation of students in real life. And the subject can be taught in secondary school with several applications. Also It is important for students to get a feel for working with and manipulating data before studying the more theoretical aspects of statistics.

---

1) Department of mathematics education, Kongju national university, chungnam, 314-701, Korea