

210-6
19148

제2회 「통계의 날」 기념 논문집

1996



210-6
19148

통 계 청

National Statistical Office

머 리 말

국내외의 급속한 환경변화는 우리나라 통계작성체계의 변화와 새로운 통계개발을 요구하고, 또한 올해의 OECD가입은 우리 통계인에게 우리나라 통계를 선진국수준으로 도약시킬 수 있는 계기를 마련해 주었습니다.

이제까지의 경제통계 위주의 통계작성방식에서 환경, 복지, 문화·여가 등 국민의 삶의 질을 측정할 수 있는 통계생산과 통계의 질적 개선을 한층 더 요구하고 있습니다. 이러한 과제를 수행하는 데는 어느 한 개인이나 기관의 독자적인 노력만으로는 이루어질 수 없는 우리 모두의 과제입니다.

우리 통계인은 올바른 통계를 만들기 위해 많은 노력을 기울여야 하고 국민들은 통계조사에 진실된 응답으로 협조함으로써 보다 정확하고 신뢰성 있는 통계를 작성할 수 있도록 다함께 노력하는 자세가 어느때보다도 필요한 시기라고 생각합니다.

이 논문집은 통계인들의 통계작성 노력을 한데 모으고 국민들의 통계에 대한 인식제고와 통계부문의 투자를 확대할 수 있는 계기를 마련하고자 개최된 제2회 「통계의 날」 기념행사의 일환으로 실시한 통계심포지엄 발표논문과 대학생을 대상으로 한 통계경진대회 입상 작품을 수록하였습니다.

본 책자가 여러분의 업무수행에 조금이나마 보탬이 될 수 있기를 바라며, 내년도에도 훌륭한 논문이 다수 발표되어 우리나라 통계발전을 위한 초석이 되길 바랍니다.

아울러 그 동안 통계발전에 공헌하신 통계작성기관, 학계 등 관계자 여러분에게 깊은 감사를 드립니다.

1996. 12

통계기획국장 김일현

목 차

I. 심포지움 발표논문 3

1. 제1주제 : 인구구조의 변화와 정책방향
 ◇ 발 표 : 김일현 / 통계청 통계기획국장 5
2. 제2주제 : 기초통계의 부족 및 활용방안
 ◇ 발 표 : 팽동준 / 한국은행 조사2부장 26
3. 제3주제 : 통계의 시의성과 정확성
 ◇ 발 표 : 이재창 / 고려대학교 교수 38

II. 통계경진대회 입상논문 49

《1등-1편》

1. 농구경기의 그래픽컬 표현
 ◇ 김정수, 김혜숙 / 중앙대학교 응용통계학과 51

《2등-2편》

2. 나무구조에 의한 B형간염 조기진단 모형연구
 ◇ 이춘식 / 한국방송통신대학교 응용통계학과 60
3. 원그리기 자료에 대한 통계분석
 ◇ 김범준 / 목원대학교 응용통계학과 73

《3등-3편》

4. 서울특별시 구별 특성에 따른 쓰레기 배출량 분석
 ◇ 서두성, 이종현, 이규학 / 고려대학교 응용통계학과 90
5. 고려대학교 학생들의 사회현상 및 시사에 대한 인지도 조사
 ◇ 최원혁, 장우성, 김영덕 / 고려대학교 통계학과 106
6. 밖에서 본 5.18
 ◇ 김재춘, 박장호 / 동신대학교 통계학과 128

I. 심포지움 발표논문

인구구조의 변화와 정책방향

김 일 현

통계청 통계기획국장

1. 서 언

현재 우리나라에서 인구통계를 작성할 수 있는 기본적인 자료원은 매 5년마다 전국적으로 조사되는 인구센서스와 국민의 신고로 이루어지는 호적제도 및 주민등록제도 세가지로 볼 수 있다. 이 제도로 부터 작성되고 있는 통계내용을 살펴보면 인구센서스로 부터는 인구정태통계, 호적제도로 부터는 인구동태통계 그리고 주민등록제도로 부터는 인구이동통계가 만들어지고 있다.

이와같은 제도의 연혁을 살펴보면 우리나라에서는 통일신라시대부터 조선시대에 이르기까지 호구성적이라는 제도로부터 파생된 것으로 볼 수 있다. 당시 호구성적제도의 목적은 징병, 賦稅, 徭役, 신분 혼란의 방지등으로서 오늘날의 통계작성목적과는 다소 거리가 있었다. 그러다가 1896년 9월 1일자로 호구조사규칙이 제정되면서 인구조사가 비로서 오늘날과 같은 통계목적을 가지게 되었고, 이후 인구센서스, 호적제도 및 주민등록제도로 분화 운용된 것이다.

이러한 인구통계를 작성할 수 있는 각종 제도로부터 만들어진 인구정태 및 동태통계는 각종 정책집행에 대한 평가자료로서 뿐만 아니라 각종 경제사회개발계획 수립을 위한 기초자료로서 매우 유용하게 사용된다.

본고에서는 인구통계작성제도를 검토하기보다는 통계청에서 작성하는 각종 인구정태 및 동태통계를 이용 인구구조의 변화와 그에 따른 정책방향을 살펴보고자 한다.

2. 우리나라의 인구성장추이

2.1. 인구규모 성장추이

우리나라 인구구조의 변화추이에 관한 고찰에 앞서 각종 역사기록에 의한 우리나라의 인구성장추이를 시대흐름별로 개략적으로 살펴보고자 한다.

한반도에 인간이 살기 시작하면서부터 지금에 이르기까지 인구의 성장과 변화는 역사변화와 때를 같이하고 있다. 이러한 점을 고려하여 우리사회의 인구변동의 시대구분을 하면 다음과 같다.

- ① 근대화 이전의 인구(1910년 이전)
- ② 식민지시대의 인구(1910~1945년)
- ③ 혼란기의 인구(1945~1960년)
- ④ 경제성장기의 인구(1960년 이후)

(1) 근대화 이전의 인구

먼저 삼국시대 이전의 상황을 살펴보면 지금부터 약 2000여년전 당시 한반도지역에는 낙랑, 현도, 임둔, 진번 이름을 가진 한사군과 남쪽지방에는 마한, 변한, 진한 이름의 삼한과 이외 옥저, 동예 같은 부족연맹국들이 혼재하고 있었던 것으로 분석되고 있다. 史書에 기재된 당시의 호구수를 살펴보면, 낙랑은 62,812호에 406,748명, 현도는 45,006호 221,845명, 마한은 10만호, 변·진한은 4~5만호, 옥저는 5천호, 동예는 2만호 정도로만 파악되고 있다. 이를 합해보면 약 283천호가 되는데, 이는 당시 한반도에 거주하고 있는 최소한의 호수가 아닌가 싶다. 그리고 호당 평균가구원수를 6명으로 간주한다면 당시 한반도에 거주하고 있었던 최소한의 인구수를 170만명으로 계산하여 볼 수 있다.

한편 삼국시대에 대한 기록은 삼국유사에서 단편적으로 찾아볼 수 있는데, 삼국시대 전성시 호수의 합계(신라 178,936호, 고구려 210,508호, 백제 152,300호)가 541,744호로 파악되고 있다. 만약 호당 평

균가구원수를 5명이라 가정하면 삼국시대에 총인구는 2,708,720명으로 추정해 볼 수 있다.¹⁾

고려시대의 경우는 인구수가 오직 宋史 고려조에서만 발견되는데, 여기서 총인구를 2백 10만명으로 기록하고 있다. 이 기록은 당시 고려측으로부터 제공받은 수치를 기록한 것으로서 신뢰성이 높지만 이 기록만 갖고 고려시대의 정확한 인구수를 추정하기에는 부족하다고 볼 수 있다.²⁾

조선시대로 넘어와서는 17세기 중엽까지 호구수에 대한 기록을 거의 찾아볼 수 없다. 단지 1400년 전후에 몇가지 기록이 있으나 호구수가 터무니없이 적게 나타나, 당시의 제도가 상당히 취약했거나 아니면 호구를 헤아리는 방법이 나중과 달랐기 때문으로 볼 수 있다. 조선시대에서 인구수가 7백만명 수준으로 파악된 것은 임진왜란이 끝나고 100여년이 지난 17세기 후반부터이다. 이 때부터 700만명 수준의 인구규모는 1900년경까지 계속되는데, 아마도 인구동태형태가 多産多死이었기 때문에 인구의 변동이 그다지 없었기 때문으로 풀이될 수 있다.

그러나 액면 그대로 조선시대의 인구를 700만명 수준이라고 단정 짓기에는 뒤에서 살펴보는 바와 같이 漏戶·漏口를 감안하여 본다면 조선시대에서의 실제 호수는 약 200만호에 인구수는 약 천만명 수준이 아니었나 싶다.³⁾ 이렇게 조선시대의 인구를 천만명 수준으로 보더라도 오늘날의 우리민족의 인구수(1995년 현재 남한 44,851천명, 북한 23,261천명, 해외동포 500만명⁴⁾으로 합계 7천 3백만명)는 조선시대보다 7배 이상이 늘어난 것으로 분석하여 볼 수 있다.

1) 윤종주, 우리나라 古代人口에 관한 小考, 인구학회지, 1985

2) 진단학회, 한국사, 을유문화사, 1969

3) 法制處, 度支志(제1권), 법제자료 제91집, 1977

4) 外務部 集計에 의하면 1995年末 현재 우리나라 해외동포수는 4,938,345명임. 이 중 中國에는 1,926천명, 미국은 1,661천명, 일본은 659천명, 옛 소련에 459천명 등이 거주하는 것으로 분석되고 있음.

〈표 1〉

시대별 호구기록 자료

시 대	년 도	戶	口	자 료 출 처
삼국시대 ¹⁾	고구려 전성시(491)	210,508	—	삼국유사 권1 고구려조
	백 제 전성시(554)	152,300	—	삼국유사 권1 변한 백제조
	신 라 전성시(576)	178,936	—	삼국유사 권1 진한조
	고구려 멸망시(668)	697,000	—	舊唐書 卷199 東夷傳
	백 제 멸망시(660)	760,000	—	삼국유사 권1 太宗春秋公條
고 려 ²⁾	고 려 전성시	—	2,100,000	宋史 高麗傳
조 선 ³⁾	인조 17년(1639)	441,827	1,521,156	戶口總數
	현종 7년(1666)	1,108,351	4,107,156	戶口總數
	숙종 16년(1690)	1,514,000	6,952,907	戶口總數
	고종 원년(1864)	1,703,450	6,828,521	
	순종 2년(1907)	2,333,087	9,781,671	경찰관서조사(5. 20)
일 제 하 ³⁾	1910	2,742,103	12,935,282	民籍法 實施調査(5. 10)
	1925	3,720,773	19,522,945	제1회 簡易國勢調査結果

자료 : 1) 윤종주, 우리나라 고대인구에 관한 小考, 한국인구학회지, 1985

2) 진단학회, 한국사, 을유문화사, 1969

3) 통우회, 한국통계발전사, 1989

주 : 1) 고구려 및 백제 멸망시의 수치는 기록에 戶로 나타나 있으나 일반적으로 인구 수로 보고 있음.

2) 1907년 및 1910년의 수치는 경찰계통을 통한 조사결과로서 조선총독부의 統計年表에 수록된 연말기준 수치보다는 적게 나타나나, 1906년 내부의 조사결과와 비교하게 되면 월등히 많아 이 당시 조사를 엄격하게 하였음을 반증함.

(2) 식민지시대의 인구변동(1910~1945)

일본의 식민지하 인구는 우리나라 역사상 엄청난 변화를 보인 시기이다. 비록 식민지 통치를 그 목적으로 하고 있었지만 인구조사가 정확하게 이루어졌기 때문에 당시의 인구에 관한 연구를 쉽게 할 수 있다.

국세조사가 시작된 1925년 이후의 인구성장을 보면 1925년에 약 1,902만명 정도에서 1944년에는 약 2,512만명으로 이 기간에 약 610만명의 인구증가를 보여 연평균증가율은 1.5% 정도를 나타내고 있다.

(3) 혼란기(1945~1959)

1945년 해방이후부터 1955년 사이에 우리나라의 인구추세는 커다란 혼란을 겪게 된다. 우선 1945년 해방과 함께 우리나라는 남·북한의 둘로 나뉘어졌다. 불행히도 해방이후 북한의 인구에 대해서는 극히 단편적인 자료밖에 나오지 않고 있다. 따라서 1945년 이후 우리나라의 인구추세는 주로 남한지역에 국한하여 살피는 도리밖에 없다. 또한 1950~1953년 사이의 한국동란은 여러가지로 우리나라의 인구현상에 영향을 주고 있다.

1945년 남한지역의 인구는 약 1,614만명으로 추산된다. 이 인구는 1949년에 2,017만명 정도로 늘어났다. 이 기간의 연평균 증가율은 5.7%로 이는 현재 기록에서 찾아 볼 수 있는 한 우리 역사상 가장 높은 증가율이다. 이러한 엄청난 인구의 증가는 무엇보다도 당시 일본과 만주로부터의 대량 귀환이동과 이북으로부터의 귀환 및 피난민의 유입에 의해 설명된다. 1945~1949년 사이 만주 및 이북으로부터의 이입자 총수는 약 250만명⁵⁾에 달하고 있으며, 이는 1945년 당시 인구의 12%에 해당하는 수치이다. 이러한 대량적인 인구이입은 인구압력을 급격히 증가시켰음은 물론, 이미 정치적·경제적으로 혼란 하였던 당시의 사회를 더욱 혼란스럽게 만드는 결과를 초래하였다. 귀환민들은 거의가 과거 농촌에 거주하였던 농촌출신이지만 돌아올 때에는 농촌에 생활기반을 갖지 못한 사람이 대부분이었다. 따라서 이들은 귀국후 농촌으로 돌아가기보다는 자기출신지역의 도시에 정착하는 경향이 강했다. 해방 직후 도시에 집중되었던 약품 및 식량 구호활동도 또한 이들의 도시집중을 부채질한 것으로 풀이된다. 한편 피난민은 거의가 서울 등 몇몇 대도시로 정착을 하고 있다. 결과적으로 1945~1949년 사이에 우리나라에서는 모든 도시가 엄청난 인구의 증가를 경험하게 된다.

5) 해방이후 1949년까지 만주, 일본 및 북한으로부터 들어온 한국인수는 한국은행의 1949 한국경제연감에서는 2,198,642명, 한국산업은행 조사부의 한국산업경제10년사에서는 2,482,365명, 미공표된 1949 인구센서스에서는 1,880,000명으로 기록하고 있다.

1949년과 1955년 사이의 인구추세는 6·25사변에 의해 결정된다. 즉, 높은 사망률과 이북으로부터의 인구유입, 그리고 경계선(휴전선)의 변화에 의해 인구변동은 설명된다. 이 기간에는 상당한 인구유입이 있었음에도 불구하고 전쟁동안의 높은 사망률 때문에 총인구의 증가율은 극히 낮은 수준에 머물렀다. 전쟁기간(1950~1953)만을 따지면 전쟁으로 인해 발생한 추가사망이 약 165만명, 이북으로 납치된 인구가 30만명, 그리고 이북으로부터의 피난민이 65만명으로 결국 약 130만명의 인구가 순전히 전쟁때문에 감소한 것이 되었다. 또한 1949~1955년 사이에는 전쟁으로 인해 남한내에서도 심한 피난민이동이 발생하였다. 서울 및 기타 북쪽 도시인구의 남쪽, 특히 경상도지역 도시로의 이동, 북쪽지방의 도시에서 농촌으로의 이동이 대표적인 경우이다.

1955년 이후 우리나라의 인구는 해외이동에 의해 별 영향을 받지 않은 폐쇄적 인구의 특징을 나타낸다. 그러나 1955~1960년 사이의 인구성장과 그 유형은 한국동란 동안의 인구추세에 의해 주로 결정된 것으로 해석할 수 있다. 전쟁이 끝나자 소위 '아기붐'현상이 나타났다. 아기붐은 다른 말로 출생률의 급격한 상승이다. 이는 전쟁기간 별거하던 젊은 부부들의 재결합과 그 사이 연기되었던 결혼이 한꺼번에 몰리는 바람에 나타나는 현상으로 풀이된다. 전쟁기간에 15% 정도 감소하던 출생률이 1955~1960년 사이에는 정상적인 경우보다 10% 더 높아져 전쟁을 전후해 출생률은 25%의 차이를 보인다. 한편 사망률도 전쟁기간에 보급된 항생물질이 전쟁이후 대량으로 퍼짐에 따라 급속히 감소해 인구의 자연증가는 유례없이 높은 수준을 보이게 된다.

1955~1960년 사이 우리나라의 연평균 인구증가율 2.9%는 해외인구이동의 영향을 받지 않은 것으로서는 우리의 역사상 최고의 수준임에 틀림없다. 또한 전쟁기간의 피난민들의 귀환이동과 이북으로부터 온 피난민의 재정착과정은 1955~1960년의 국내 인구재분포의 가장 중요한 특징이었다. 이러한 이동은 북쪽지역의 도시, 즉 서울

및 경기·강원지역 도시에서 급격한 인구성장을 초래하였다.

전반적으로 1945~1960년 사이의 혼란기는 1960년 이후 인구 및 사회변동에 중요한 의미를 갖는다. 이 사이의 혼란은 현대적인 도시화의 기틀과 함께 출생률 변천의 기틀을 마련하는데 결정적인 역할을 하였다. 우선 피난민 이동이기는 하지만 단기적인 대규모의 이동은 이동의 억제요인인 지역적인 숙박을 없애고 다른 지역에 대한 지식과 연관을 넓히는데 큰 공헌을 하였다. 전쟁기간에 어쩔 수 없는 많은 인공유산의 실시는 사람들로 하여금 전통적인 인공유산에 대한 규범적 숙박을 덜 느끼게 하였다. 또 전쟁과 이동 때문에 초혼연령은 급격히 올라갔다. 1940년에 17.8세이던 여자의 평균 초혼연령이 1960년에는 21.5세로 높아지고 있다. 남자의 경우는 21.8세에서 25.4세로 올라가고 있다. 한편 이러한 혼란 때문에 야기된 경제적인 어려움은 사람들에게 자녀수를 억제하도록 압력을 가하는 가장 중요한 원동력이 되었다.

(4) 1960년 이후의 인구추이

1960년대에 들어오면서 우리 사회는 새로운 인구변천기에 접어들게 된다. 전통적으로 높았던 출생률이 계속 떨어지는 현상이 나타났으며 1960년 이후 사망률도 계속해 다시 떨어지기 시작했다. 그러나 사망률의 저하속도보다 출생률의 저하속도가 더 빨라 인구증가율도 계속 떨어지게 되었다. 2.9%를 보이던 1955~1960년의 인구증가율은 80년대 중반부터 1% 수준으로 감소했다.

2.2. 人口變動率의 推移

앞에서 주로 戶口調査나 인구센서스 결과를 기초로 우리나라의 人口規模 및 人口成長率 推移를 살펴 보았다. 그러나 이 수치에는 人口増加의 자연적 요인(出生·死亡) 뿐만 아니라 사회적 요인인 이동까지 포함되어 있어 엄밀한 의미의 出生率 및 死亡率만을 기초로 한 人口變遷理論을 살펴보기가 어렵다. 또한 센서스 인구를 기초로 한

센서스간 人口成長率은 인구센서스마다 각기 다르게 나타나는 인구수의 누락(Under-Count)과 중복(Over-Count) 때문에 정확한 인구증가율이 되기가 어렵다.

〈표 2〉 센서스 인구추이

	센서스 인구 ¹⁾	연평균 증가율(%)
1925. 10. 1	19,020,030	
1930. 10. 1	20,438,108	1.4
1935. 10. 1	22,208,102	1.7
1940. 10. 1	23,547,465	1.2
1944. 5. 1	25,120,174	1.8
1945	16,140,000 ²⁾	
1949. 5. 1	20,166,756	5.7
1955. 9. 1	21,502,386	1.0
1960. 12. 1	24,954,290	2.9
1966. 10. 1	29,159,640	2.7
1970. 10. 1	31,435,252	1.9
1975. 10. 1	34,678,972	2.0
1980. 11. 1	37,406,815	1.5
1985. 11. 1	40,419,652	1.6
1990. 11. 1	43,390,374	1.4
1995. 11. 1	44,606,199	—

주 : 1) 센서스 인구는 외국인을 제외한 한국인 수치이며, 1944년 이전 자료는 남·북한 합계임. 1995년 자료는 외국인을 포함한 잠정 집계 결과임.

2) 권태환·권두섭, 인구의 이해, 서울대 출판부, 1990에서 인용

따라서 出生 및 死亡申告를 기초로 未申告 또는 遲延申告分을 감안하여 推定한 자료로서 우리나라 人口動態率 推移를 살펴보면 〈표 3〉과 같다. 동표에서 1960년 이전년도를 대상으로 한 金哲의 推定値는 出生·死亡에 있어 未申告나 遲延申告分 등을 감안하지 않고 신고된 결과를 있는 그대로 사용한 것인 반면, 崔仁鉉의 推定値는 未申告나 遲延申告分을 감안하여 추정한 것으로서 金哲의 推定値보다 높게 나타나는 특징을 보이고 있다.

동표를 기초로 우리나라의 人口變動率 推移를 살펴보면 다음과 같다.

먼저 出生率을 살펴보면, 1960년 이전에는 연간 出生率이 人口千名當 40명이상으로 거의 일정하게 높은 수준을 나타내고 있다. 만약 이 수준을 자녀수로 나타낸다면 1960년 이전에는 한명의 여자가 평균 6명이상의 자녀를 가지고 있었던 것으로 추측해 볼 수 있다.

그러나 이와같은 높은 出生率은 1960년대 초반부터 저하를 나타내기 시작하여 1990년에는 불과 人口千名當 15.6명이 출생한 것으로서 매우 낮은 수준을 보이고 있다. 이를 자녀수로 나타내 본다면 지난 1960년에는 한 여자가 평균적으로 6명의 자녀를 가졌으나 1990년에는 2명에도 미치지 못하는 1.6명에 불과한 수준인 것이다.

반면에 死亡率은 1910년대부터 꾸준히 감소하는 양상을 보이고 있다. 年間 死亡率이 1916~20년간의 3%의 높은 수준에서 해방 직후인 1946~50년간에는 1.6%로, 1970년에는 1%로, 그리고 1990년에는 0.6%로 나타나 우리나라의 死亡率이 꾸준히 감소하였음을 알 수 있다.

따라서 人口增加要因인 出生率에서 人口減少要因인 死亡率을 빼준 自然人口增加率은 1930년 이전에는 높은 出生率과 높은 死亡率로 인하여 2% 이하를 기록하였으나, 1956~60년간에 出生率은 계속 높은 반면에 死亡率의 감소로 인하여 人口增加率이 오히려 높아져 3%나 되고 있다. 그러다가 1960년 이후에는 지속적인 出生率 및 死亡率의 감소로 인하여 人口增加率도 떨어져 1970년에는 2% 수준, 1985년 이후에는 1% 수준으로 되었다.

〈표 3〉

出生率, 死亡率 및 自然人口增加率 推移

(단위 : 인구천명당)

	崔 仁 鉉 推 定 ¹⁾			金 哲 推 定 ²⁾		
	出 生 率	死 亡 率	自然人口 增 加 率	出 生 率	死 亡 率	自然人口 增 加 率
1916~20	47.5	31.6	15.9	31.5	24.7	6.8
1921~25	48.0	29.5	18.5	36.3	20.8	15.5
1926~30	45.9	26.4	19.5	37.5	21.6	15.9
1931~35	45.5	23.3	22.2	30.9	20.3	10.6
1936~40	43.3	21.4	21.9	32.6	18.4	14.2
1941~45	42.1	19.5	22.6	35.4	18.5	16.9
1946~50	39.9	15.8	24.1			
1951~55	41.0	14.3	26.7			
1956~60	43.0	12.8	30.2			
1960 ³⁾	42.1	12.1	30.0			
1970 ³⁾	29.9	9.5	20.4			
1980 ³⁾	23.4	6.7	16.7			
1985 ³⁾	16.4	6.2	10.2			
1990 ³⁾	15.6	5.8	9.8			
1994 ⁴⁾	16.5	5.5	11.0			

자료 : 1) 경제기획원, 1966 인구센서스 종합분석 보고서, 1970, p.30

2) 金哲, 한국의 인구와 경제(일어판), 岩波書店, 1965, p.24.

3) 통계청, 장래인구추계 : 1990~2021, 1991.4

4) 통계청, 1994 인구동태통계연보, 1995

주 : 1) 이 표에서의 동태율은 신고를 기초로 추정한 것으로, 인구센서스를 기초로 한 연평균 인구증가율(이민율이 포함되어 있음)하고는 수치가 다르게 나타남.

2) 金哲의 추정치는 출생·사망 신고에 있어 미신고나 지연신고분을 감안하지 않고 신고결과를 그대로 사용한 것이며, 崔仁鉉의 추정치는 미신고나 지연신고분을 감안한 것임.

3. 장래인구구조의 전망

3.1. 총인구수 및 연평균증가율

통계청에서는 1995년 인구센서스결과를 기초로 인구추계작업을 진

행중에 있다. 본 논문에서는 새로운 인구추계결과를 사용하기가 시기적으로 적절치 않아 1990년 인구센서스 결과를 기초로 한 인구추계 작업결과를 이용하여 분석하였다.

1990년 인구센서스를 기초로 한 추계(구추계)결과와 1995년 인구센서스를 기초로 한 추계(신추계)결과에서 나타날 차이점은 두가지로 요약 유추해 볼 수 있다.

첫째는 1995년 연앙추계인구는 신추계에서 다소 많아지리라고 보는데 이는 지난 5년간의 출산율이 당초 예측하였던 것보다 다소 높았기 때문으로 당연한 결과이다.

둘째는 구추계에서 2021년에 최대인구 5,059만명으로 인구정지상태에 도달될 것으로 예측하였는데, 신추계에서는 인구정지상태 도달시기가 몇년정도 늦추어지고 최대인구규모도 다소 많아지리라고 생각된다. 그러나 기본적인 인구구조의 변화추이는 구추계나 신추계 모두 같은 방향으로서 큰 차이는 없다고 본다.

구추계에 이용된 기초인구(Base population)로서의 1990년 年央人口(Midyear population)는 42,869천명인데 32년후인 2021년에는 최대인구인 5,059만명에 이르고, 그 이후부터는 총인구수 자체가 감소되는 부(-)의 인구성장을 나타내게 될 것으로 추계되었으나 실제에 있어서는 인구학적 여건으로 보아 인구가 감소하지는 않을 것으로 본다.

1990년의 인구증가율은 1%를 밑도는 0.98% 수준이고, 점차 감소하기 시작하여 1995년에는 0.93%가 되고 2000년에는 0.83%로 낮아지며, 2021년에는 0%를 기록해 우리나라 인구의 증가가 극대화 될 것으로 전망된다.

3.2. 인구구조의 변화

(1) 1990년까지의 인구구조

연령별 인구구조의 類型은 일반적으로 사회의 발전단계를 나타내는 지표가 될 수 있으며, 低開發國家에서 開發途上國家 그리고 先進

國家로 발전됨에 따라 인구구조도 피라미드형에서鐘型 그리고 항아리형으로 변화된다. 우리나라 人口構造의 變化推移를 살펴보기에 앞서 인구구조의 특징을 살펴보면 다음과 같다. 즉 연령별로 인구수가平滑(Smooth)하지 않고 몇가지 역사적인 사건에 따라 기복이 있음을 보인다. 1945년 전후의 해방시기와 1950~53년간의 한국전쟁 시기에는 출산율의 감소 등으로 인구가 인접 연령층 보다도 적게 나타난다. 전쟁후에는 '베이비 붐(Baby boom)' 현상으로 인하여 인구수가 갑작스럽게 늘어났으며 이러한 현상은 그 당시만 영향을 미치는 것이 아니라, 25년이 지난후 즉 한世代가 지난후에도 인구구조에 영향을 미치는 요인이 되고 있는 것이다.

우리나라의 연령별 인구구성의 變化過程을 살펴보면, 1960년대까지는 다분히 후진국 특징을 지니고 있었으나 1970년대부터는 선진형과 후진형이 혼합된 양상을 나타내기 시작하였다. 이것이 80년대 초반부터는 비교적 선진형에 접근하고 있음을 볼 수 있다. 인구구조에 있어서 후진형의 특징은 높은 幼年人口 구성비와 생산가능연령의 낮은 인구구성비를 들 수 있다.

그런데 우리나라 유년인구 구성비의 추이를 살펴보면 1970년 이전까지는 모두 40% 이상을 차지하고 있어 전형적인 피라미드형 인구구조의 저변을 형성하고 있음을 알 수 있다. 그러다가 1970년 이후 유년인구 비율은 1975년에 35.7%, 1980년에 33.7%, 1985년에 30.2% 그리고 1990년에는 25.8%로 계속 떨어지는 모양을 보이고 있다. 이와 같은 유년인구 비율의 감소는 생산가능연령 인구비율의 증가를 의미하는 것으로서, 1960년의 54.8%에서 1980년에는 62.2%로 그리고 1990년에는 69.2%로 높아졌다.

이러한 인구변화는 扶養比의 감소를 초래하게 된다. 부양비는 유년인구와 노년인구를 합하여 生産可能年齡人口로 나누어준 비율인데, 이 비율은 1960년의 83%에서 1980년 60.7% 그리고 1990년에는 44.5%로 줄어든 것으로 나타났다. 이것은 扶養負擔이 줄어들고 있다는 점에서 바람직한 현상이다.

또한 인구구조의 변화추이를 한마디로 요약한 指標로는 平均年齡 資料가 있다. 우리나라의 인구구조가 피라미드형에서 항아리 형으로 점차 이행됨에 따라, 평균연령도 1960년의 23.1세에서 1990년에는 29.5세로 6.4세가 증가하여 인구의 노령화 현상이 진행되고 있음을 알 수 있게 하고 있다.

(2) 장래인구 구조

우리나라의 인구구조는 예측한 대로 代替出產力(Replacement level of fertility)이하의 낮은 출산력 수준이 계속됨에 따라 향후 0~14세 幼年인구의 비율이 지속적으로 감소하여 '90년의 25.8%에서 2021년에는 15.8%까지 낮아질 전망이다.

상대적으로 65세이상의 老齡인구가 전체인구에서 차지하는 비율은 유년인구 비율의 감소, 평균연령의 상승 등으로 계속 늘어나게 되는데, 1990년의 5.0%에서 2000년에는 6.8%, 2021년에는 13.1%까지 될 것으로 전망된다. 이같이 인구의 노령화현상을 맞이하게 됨에 따라 현재 선진국들이 경험하고 있는 노인문제를 우리나라도 겪게 될 것이며, 최근 노인의 數 및 노인가구의 빠른 증가는 사회·경제 문제와 관련하여 심각성을 재촉하게 된다. 따라서 이에 대한 研究 및 對策이 필요한 실정이다.

한편으로 生産可能 年齡人口라고 볼 수 있는 15~64세 인구비율은 유년인구비율의 감소로, 1990년 69.2%에서 점차 늘어 1995년에는 71%까지 된 후 그 이후에는 전체인구의 72%를 계속 차지할 것으로 전망되어 노동력인구의 극대화가 이루어진다. 즉, 우리나라는 향후 계속하여 상당기간 勞動力供給의 總量的 측면에서 어떤 문제의 발생 보다는 오히려 이 생산가능연령인구의 量的 증가가 여자인구의 經濟活動 參加와 더불어 고용기회를 더 많이 創出하도록 하는 압력을 가할 수 있는 가능성이 크다는 것이다.

그러나 이는 15~64세 인구 全體的인 측면에서 그렇다는 것이지 연령층으로 세분하여 살펴보면 勞動供給에 있어 많은 문제가 발견된

다. 즉, 총인구 중 15~64세 인구가 차지하는 비율은 72%정도가 계속 되겠지만, 인구의 노령화 현상으로 인하여 壯年層 勞動力 인구수는 계속 빠른 속도로 증가하지만, 상대적으로 값싼 勞動力을 제공할 수 있는 소년층의 絶對人口數는 오히려 줄어드는 현상을 경험하게 될 것이다. 따라서 기술의 선진화가 수반되지 않는 한 노동의 생산성 문제가 야기될 것이고 이에 대한 어떤 政策的 對應이 절실하게 된다.

또한 이들 生産可能年齡人口가 생산활동을 할 수 없는(0~14세와 65세이상) 인구에 대해서 갖는 扶養負擔의 상대적 크기를 나타내는 扶養比(Dependency ratio)가 1990년의 44.5%에서 1999년에는 38.6%로 떨어지다가 이후에는 다시 높아져서 2021년에는 40.6%가 될 것으로 전망된다. 말하자면 인구구조의 노령화에 따라서 생산인구보다는 소비인구가 漸增하게 된다. 결국 우리나라 인구의 동태적 변화 양상이 저출생, 저사망의 형태로 進行되어 감에 따라 인구구조는 0~14세의 유년인구가 전체인구의 40%이상을 차지하는 후진국의 피라미드형(Pyramid)에서 모든 연령층에 인구가 고르게 분포되는 鐘型으로 바뀌어 가고 있음을 보이고 있다. 이렇게 변화되는 인구구조는 인구학적으로 장래인구를 결정하는데 뿐 만 아니라 경제적으로는 노동력공급이라는 측면과 財貨 및 서어비스의 수요 측면에서도 매우 중요한 역할을 하는 것이다. 앞에서도 언급한 바와 같이 이러한 인구구조를 先進型이라고 하는데 우리나라는 그 변화가 너무나 빠른 속도로 이루어졌으며 이러한 생산활동이 극대화 될 수 있는 인구구조가 형성되었을 때 노동력을 충분히 활용할 수 있는 정책이 수반된다면 생산의 극대화(生産性 增大)가 가능할 것이고 富의 蓄積이 이루어져 우리의 先進化도 容易하게 된다는 것이 특징이라 하겠다. 모든 先進國의 발전단계를 보면 대체로 인구구조가 鐘型인 단계에서 경제 발전이 最上이었음을 간과해서는 안 될 것이다.

〈표 4〉

人口構造의 推移

(단위 : 천명)

년도	총인구	0~14세		15~64세		65세이상		부 양 비		
		인 구	(%)	인 구	(%)	인 구	(%)	계	유년	노령
1960	25,012	10,588	42.3	13,698	54.8	726	2.9	82.6	77.3	5.3
1970	32,241	13,709	42.5	17,541	54.4	991	3.1	83.9	78.2	5.7
1980	38,124	12,951	34.0	23,717	62.2	1,456	3.8	60.7	54.6	6.1
1985	40,806	12,305	30.2	26,759	65.6	1,742	4.3	52.5	46.0	6.5
1990	42,869	11,077	25.8	29,648	69.2	2,144	5.0	44.5	37.3	7.2
1995	44,851	10,400	23.2	31,908	71.1	2,543	5.7	40.6	32.6	8.0
2000	46,789	9,917	21.2	33,705	72.0	3,168	6.8	38.8	29.4	9.4
2010	49,683	9,510	19.1	35,505	71.5	4,668	9.4	39.9	26.8	13.1
2021	50,586	7,989	15.8	35,972	71.1	6,625	13.1	40.6	22.2	18.4

주) 扶養比는 0~14세 人口와 65세 以上 人口를 15~64세 人口로 나눈 수치임.

4. 인구정책방향

앞으로 더욱 광범위해지는 우리나라의 인구문제에 대하여 구체적인 人口政策 方向을 人口調節(Population influencing)政策과 人口對應(Population responsing)政策의 측면으로 제시하고자 한다.

4.1. 인구조절정책

출산력에 대한 정책을 보면, 현재의 우리나라 출산력 수준이 低出生率으로 정착된 상태이나 그것이 轉換期的인 일시 현상이냐는 논외로 하더라도 21세기를 향한 장기적 출산력 정책을 고려하면 가능한 한 저출산 수준이 유지되도록 하여야 할 것이다. 그 이유는 인구규모를 최대한 억제하여야 하고, 母子保健 增進을 통한 가정복지의 극대화가 중요하기 때문이다. 저출산력 유지를 위한 구체적 방안으로는, 첫째, 小子女가정을 형성함으로써 가정복지가 향상되도록 하는 것과 母子保健의 측면에 대한 출산조절정책을 상호 연계시켜 추진하고,

둘째, 여성고용에 있어서 차별철폐의식 및 제도확보를 통해 여성의 사회 경제적 지위향상과 역할을 개발하는 정책을 촉진하고, 셋째, 소득이 낮은 계층일수록 자녀에 대한 경제적 가치를 높이 평가하므로 지역간 사회계층간 소득분배의 적정화를 통한 근본적 출산조절을 도모하고, 넷째, 출산조절정책의 내실화 및 피임보급의 질적향상을 들 수 있다.

家族計劃 事業의 政策을 보면 가족계획이 가지고 있는 원래의 의미가 피임이나 출산 억제만을 뜻하는 것이 아님에도 불구하고, 高出産 시대의 가족계획은 주로 출산억제 중심의 경직된 목적으로만 해석되어 왔다. 그동안 가족계획 사업이 추진되어 오면서 국민들의 자녀가치관이 크게 변화하였고, 피임에 대한 수용도 또한 점진적으로 향상되었다. 이와 같이 높은 피임 실천율은 단순히 정부의 강력한 가족계획사업 추진력의 결과로만 볼 수 없고, 피임실천의 동기가 사회변화에 따라 개인이나 집단수준에서 小子女觀이 확고하게 이루어진 결과로 해석할 수도 있다.

장기발전 구상으로서 향후 가족계획사업은 피임보급의 질적 향상, 피임보급대상자 선정, 새로운 피임보급망의 구축, 피임비용부담 등 전반적인 사업관리 및 운영의 개선이 요구되며, 이는 기본적으로 기존의 政府主導型 사업에서 民間主導型 사업으로 전환하는 것을 전제하고 있다. 이렇게 될 때 피임보급의 제1차적 목적도 출산력 저하를 위한 물량공급에 있지 아니하고 가정복지 차원에서의 가족보건 향상에 있게 될 것이다.

勞動力 및 人口資質 政策을 보면, 21세기를 앞둔 우리나라의 경우 인구학적측면에서의 인구의 고령화에 의한 사망력의 변동을 예상할 수 있다. 사망력이 가지는 의미는 인구정책 목적으로만 중요한 요인으로 간주되는 것이 아니고, 보건학적으로도 국민건강수준을 나타내는 지표로서 대단히 중요시 된다. 따라서, 앞으로의 사망력 정책은 우선 예방 가능한 질병의 사인 제거 및 질병의 조기발견을 촉진하도록 건강검진을 의료보험 급여에 포함시키는 방안, 지역사회 보건사업 활

동강화 등이 주안점이 되어야 한다.

국민의 인구자질은 건강을 토대로 한 생물학적 자질과 교육을 중심으로 한 사회적 자질로 논의된다. 사회적 자질의 확충을 교육정책의 역할로 미루고 생물학적 자질확충에 초점을 맞추면 크게 세가지가 주목된다.

첫째, 정신보건 향상 정책이다. 갈등과 경쟁의 현대사회에서 파생되는 각종 사회병리 현상을 극복하기 위해 인간관계를 개선하고 집단내의 정신장애 발생을 최소화시켜 나가도록 하는 전략은 집단 정신보건향상 정책의 중요 과제가 아닐 수 없다.

둘째, 선천성 요인에 의한 심신의 손상을 예방할 수 있는 대책의 개발이 필요하다.

셋째, 사망률의 저하는 국민보건의 향상을 의미하며 국민건강의 향상은 곧 인구자질의 향상을 말한다고 할 수 있는데, 특히 영유아 사망수준의 저하를 유도하기 위한 모자보건 사업은 인구자질 정책의 중요과제이다.

人口移動 政策을 보면 21세기의 고도산업화된 우리사회에 대응하는 바람직한 인구이동 정책은 적극적인移出政策을 통한 과잉인구상태의 해소, 수도권을 비롯한 대도시 지역 과밀인구의 지방분산, 지방 중소도시의 육성·활성화, 농업구조 개선을 통한 농촌 과잉인구의 해소와 농촌경제의 활성화 등이 이루어지는 방향에서 추진되어야 할 것이다.

특히, 수도권 및 대도시 지역의 인구과잉집중을 억제·분산하기 위해 지방자치제 실시와 병행하여 중앙행정 기능의 대폭적인 지방이양, 대폭적인 사회간접자본의 지방투자를 통한 지방의 기반시설 확장과 교육·문화·보건의료·교통·금융 등 편익시설의 확충, 중소기업의 지방 중소도시 유치를 통한 대도시 지역의 인구분산과 중소도시의 활성화 등 지방 중소도시의 개발·육성을 통해서 대도시 인구의 지방유입 기반을 조성해야 할 것이다.

4.2. 인구대응정책

인구성장율의 안정화 및 경제발전으로 인해 21세기 인구대응정책의 초점은 사회복지 특히 노인복지 정책 등에 맞추어 질 것이 요구된다. 이에 아래에서는 이를 중심으로 정책방향을 제시하기로 한다.

老人福祉 政策을 살펴보면 노인인구는 그간 보건·의료분야의 발달로 인한 사망력 저하와 평균수명의 연장, 출산력 저하로 인한 小家族化, 연령구조의 변화로 인하여 급속한 양적 증가와 더불어 질적, 구조적 변화를 보여 왔으며, 이 추세는 21세기 이후에도 지속될 것으로 예상된다. 노인인구의 증가는 첫째로 인구학적 측면에서 산업화, 도시사회화의 지속적 진보와 小家族化·核家族化 경향에 따라 노인동거 가구수는 감소하는 반면 노인단독가구의 급증과 더불어, 시설수용 노인수의 증가를 가져올 것으로 예상되며, 둘째로 경제적 측면에서 노인의 상대적 빈곤문제의 심화 및 이에 대처하는 가족의 노부모 부양문제와 정부의 대책이 요구된다. 셋째로 사회문화적 측면에서 노인존경의 윤리적 과제 및 바람직한 노인문화 형성 등의 과제를 제기하며, 기타 신체적·정서적 측면에서도 많은 문제를 제기하고 있다.

노인문제에 대한 노인복지 정책의 방향은 크게 두가지 측면으로 접근해야 한다. 우선, 전통적인 미풍양속에 따른 노인의 가족내 부양 상황에서는 경로사상의 고양을 위한 구체적이고 효율적인 대책으로서 경로행사 내실화, 경로를 위한 학교교육 강화등의 마련과 주택·생계지원등의 노인부양가족에 대한 제반 우대책 확보 및 3세대 주택 건립등의 기본대책이 있어야 한다.

다음으로 사회보장적 기능을 강조한다면 노인단위의 생계보장책으로 연금제도 확립, 노령수당 지급 확대, 생활보호법 범위의 확대등이 요구되며, 노인주택 건설과 노인 요양원 확충등의 대책에 역점을 두어야 할 것이다.

아울러 노인으로 하여금 긴 노후생활 기간을 활력을 갖고 지내도

록 하는 사회참여 기회의 확대, 여가 활용을 위한 시설확충 및 프로그램 개발, 노인 교육의 내실화 등의 대책 강구를 통해 노후 생활을 전향적으로 보람있게 지내도록 지원해야 한다.

人力需給 政策을 살펴보면, 우리나라의 인력수급 구조 전망에 있어서의 문제점은 양적 측면보다는 구조적 측면에 있다는 점이다. 구조적 문제는 특히 고학력 과잉, 저학력 과소라는 학력구조의 불균형으로 나타날 것으로 보이며, 이러한 불균형은 갈수록 심화될 전망이다. 또한, 인력공급을 산업별로 살펴보면, 농림·광공업등 유형재 산업은 점차 초과공급 상태가 심화될 전망이다. 이에 비하여 건설·판매·운수·금융업 등은 초과수요가 심화될 전망이어서 대조를 이룬다. 기타서비스, 에너지 산업은 초과공급상태가 停滯 내지 완만한 감소를 보일 것이다.

결국, 앞으로 다가올 高學歷 시대에 대비하여 인력수요 측면에서는 기술진보에 의한 산업인력의 수급구조가 개편되어야 한다. 아울러 인력공급 측면에서는 기술 및 직업 교육을 강화하여 高學歷 인구의 고용흡수력을 높이는 방향으로 전개되어야 할 것이다.

또한, 산업구조 조정에 소요되는 기간동안 人力再配置 및 再訓練에 힘써야 하며, 무역 및 해외투자를 활성화하여 인력을 활용하는 방안도 강구해야 할 것이다.

5. 결 언

우리나라의 인구구조는 향후 저출산에 따른 인구감소와 노령화현상을 경험하게 될 것이다. 이와같은 현상은 사회구조의 근본적인 변화를 가져와 지금까지 성장·확대 일변도였던 경제 사회정책의 재검토를 요구하게 된다.

정부에서는 이와같은 인구구조의 변화를 감안 1996년 6월 4일자로 출산억제 위주의 기존 인구정책을 전환하기로 하였다. 이 정책의 초점은 인구의 자질을 높이는 방향으로 되어 있다. 구체적으로는 의료

보험분만급여의 2자녀 제한, 부양가족 소득공제혜택 2자녀제한, 2자녀 불임가정의 공공주택 입주우선권 부여, 공무원의 학비보조수당 2자녀제한 등이 1996년중에 모두 없어질 것이다.

그러나 인구정책은 20~30년 또는 그 이후의 효과를 측정해서 결정되는 특징이 있어 정책전환의 신중과 정교함이 요구된다. 인구밀도가 세계 3위라는 사실만 갖고 인구억제정책을 쓸 수 없거니와 노동력 부족만을 놓고 인구정책을 논하기는 어렵다. 정치·경제·사회·문화등 모든 면에서 고려된 인구정책이어야 한다.

향후 인구정책은 앞으로의 경제성장속도, 산업간의 합리적 균형과 조화, 국토의 가용면적과 자원 등이 집약적으로 분석하는 것과 아울러 인구의 질을 높이는 문제가 최우선적으로 다루어져야 한다. 즉 건강한 인구, 삶의 질의 향상이 이루어질 수 있는 인구의 유지에 초점을 주어야 한다.

끝으로 인구문제와 관련된 통계자료가 정밀분석되어야 한다. 막대한 인력과 예산을 투입하여 실시한 인구센서스결과나 국민이 제출한 각종 인구동태에 관한 정보들이 충분히 심층분석되어야 하겠다. 따라서 이를 바탕으로 해서 사회·경제적 필요와 효과가 종합화된 인구정책이 수립되어야 할 것이 요구된다.

參 考 文 獻

- 권태환·권두섭, 인구의 이해, 서울대 출판부, 1990
조남훈외, 인구규모 및 구조변동에 따른 정책과제와 대응방안, 한국보건사회연구원, 1995. 12
양해영, 인구의 질을 높이자, 서울신문 1996. 6. 7자 논단
김일현, 인구증가율 1% 달성의 의미와 향후 전망, '88 언론인 심포지움자료, 대한가족계획협회, 1988
김일현외, 장래인구추계와 인구학적특징 : 1990년 인구센서스 및 신인구추계를 중심으로, 한국인구학회지, 제4권제1호, 1991

통계청, 1990~2021 장래인구추계, 1991

통계청, 1994 인구동태통계연보, 1995

통계청, 1995 인구주택총조사 잠정보고서, 1996. 4

통계청, 1995 인구주택총조사 2% 표본 속보집계결과 보도자료,
1996. 7

기초통계의 부족 및 활용방안

팽 동 준

한국은행 조사2부장

1. 머리말

통계는 사회경제의 실상이나 그 변화의 상황을 정확히 파악하고 그에 대한 자료를 제공함으로써 국민의 의사결정에 기여한다는 점에서 soft infrastructure라고 말할 수 있다. 이와 같은 통계의 의의 및 역할을 근거로 하여 통계수준을 발전·유지시켜 나가기 위해서는 부단한 개선의 노력이 필요하다. 즉 사회경제정세는 급격히 변화하므로 이에 부응하는 통계조사의 체계적인 정비·보완이 필요해져 가고 있는 것이다.

또한 통계에 대한 수요가 다양화됨에 따라 이에 대응한 자료의 요청이 많아져 가고 있는 한편, 통계를 둘러싼 환경도 크게 변화되어 어려움을 더해가고 있다.

매년 우리나라의 중앙기관나 지방자치단체 또는 민간기관에서 작성하는 통계는 370개 이상이나 되며 이는 통계조사의 대상이 된 개인, 국가, 법인 등의 협력에 의하여 작성되고 있다. 이러한 통계는 정부시책은 물론 기업경영 등에 중요한 기초자료로 활용되며 더욱 중요한 사실은 이들 통계를 분석 검토하여 이론을 세우거나 장래에 대한 예측을 하여 이 사회가 나아갈 방향을 제시한다는 것이다.

그러므로 세계화, 개방화 진전에 따른 사회 각 부문의 무한경쟁을 지원하기 위해서는 부족한 기초통계의 개발과 함께 통계 선진화도 시급히 요청되고 있으며 각국의 통계정보를 신속히 입수하여 제공하여야 한다. 또한 지방화시대의 조기정착을 위해 절대적으로 부족한

지역통계의 생산과 지방자치단체의 통계작성능력을 빠른 시일내에 함양시키는 것도 우리의 당면과제라 할 수 있다.

2. 기초통계의 개념

기초통계란 1차통계라고도 하며 집단에 속하는 개체의 수 또는 개체의 특성을 총체적으로 나타내는 통계로서 일반적으로 통계조사를 실시하여 그 결과에서 직접 얻어진 통계를 말한다. 그리고 기초통계는 통계의 작성을 주목적으로 실시조사를 하여 얻어진 조사통계와 법령에 의한 개인, 단체의 신고, 보고, 신청, 인·허가 등과 같이 다른 행정업무에 수반하여 수집된 자료로부터 작성한 보고통계로 나눌 수 있다. 이에 반해 가공통계(2차통계)는 1차통계에 어떠한 연산을 하여 얻어진 통계로서 1차통계에 비하여 해석적 특성이 있는 통계를 말한다. 가공통계에는 집단 특성치인 평균, 산포도, 지수, 상관계수 등 뿐만 아니라 국민소득통계와 같은 추계에 의한 통계도 있다.

이러한 기초통계는 인구, 경제, 사회 등에 관해 우리나라의 실질상태를 파악하고 국민의 생활향상에 도움이 되는 것이라야 한다. 이 때문에 국가의 기본적인 중요한 통계를 작성하는데 있어서는 행정정책의 기획·입안을 위한 기초적 정보의 제공뿐만 아니라 폭넓은 국민일반의 이용을 위한 정보제공이라는 면에 대해서도 심분 배려할 필요가 있다. 또한 국민의 부담과 협력에 의해 얻어진 통계는 국민의 공유재산이므로 신속히 제공되고 폭넓게 이용되도록 하는 것도 중요하다.

3. 기초통계의 현황 및 문제점

3.1. 기초통계의 작성현황

1996년 1월 1일 현재 우리나라에서 작성하고 있는 승인통계는 총

372종으로서 지정통계 36종, 일반통계 336종이다. 작성방법별로 보면 전체 372종의 통계중 기초통계는 343종으로 그중 조사통계는 176종, 보고통계는 167종이며, 가공통계는 29종이다. 또한 작성기관을 정부기관과 민간지정기관으로 구분하여 볼 때 정부기관에 의해 작성되고 있는 통계는 269종이며 민간지정기관에서 작성하고 있는 통계는 103종이다.

참고로 일본의 경우에는 439종으로 우리나라보다 67종이나 많다. 그리고 기본통계인 지정통계는 우리나라의 경우 36종인데 반해 일본은 62종으로 상당한 차이를 보이고 있어 우리나라의 기초통계가 일본에 비해 상대적으로 크게 부족하다는 것을 알 수 있다.

우리나라의 승인통계 작성현황(1996. 1. 1현재)

(단위 : 종)

	계	종 류 별		작 성 방 법 별		
		지정통계	일반통계	조사통계	보고통계	가공통계
－ 정 부 기 관	269	30	239	104	142	23
○ 중앙행정기관	219	30	189	84	127	8
○ 지방자치단체	50	－	50	20	15	15
－ 민 간 지 정 기 관	103	6	97	72	25	6
계	372	36	336	176	167	29

자료 : 통계목록(통계청, 1996. 1)

일본의 통계작성 현황 (1995년 현재)

(단위 : 종)

계	지정통계	신고통계	승인통계
439	62	167	210

자료 : 일본 총무청 통계국 통계기준부

3.2. 기초통계의 문제점

(1) 통계의 조사 및 공표상의 문제점

기초통계를 조사할 때 예산제약 등으로 양질의 조사요원 확보가 어려울 뿐만 아니라 이들에 대한 교육훈련이 부족하며, 또한 정확한 통계를 작성하기 위해서는 개인 또는 사업체에서 정확한 정보제공을 전제로 하나 사생활 침해, 사업체 기밀보호를 이유로 응답기피 또는 거부현상이 날로 증가함에 따라 수집된 통계의 질이 저하될 우려가 높은 것이 현실이다.

그리고 경제통계의 경우 현재의 경기국면에 대한 정확한 진단과 처방에 필수적인 각종 통계들의 속보치가 발표된다면 시의적절한 경기정책의 시행이 가능하게 된다. 정부의 경기정책과 기업의 경영활동은 때를 놓치지 않는 것이 매우 중요하며 생산·투자·성장 등 주요 경제통계들이 선진국들보다 늦게 집계된다면 효율적인 경기정책 마련의 기회를 놓치는 요인이 될 수도 있다. 물론 방대한 양의 통계를 신속·정확하게 집계하기 위해서는 많은 예산과 인력이 필요할 것이다. 그렇지만 주요 통계들이 다소 늦은 시점에서 잠정치가 발표되고 나중에 확정치로 발표되고 있는 현재의 통계집계방식보다는 먼저 대표성 있는 자료들을 취합하여 우선 속보치를 발표한 후 추후에 계속 수정을 하는 방법으로 발표빈도수를 늘림으로써 보다 신속하게 집계·공표할 수가 있을 것이다.

(2) 기초통계의 부족

현재 OECD 가입을 앞두고 제출하여야 할 약 80여종의 통계의 개선 또는 개발이 시급한 과제로 떠오르고 있다. 즉 기존에 생산되고 있는 많은 통계도 선진국 수준으로 기준과 질을 제고 시키는 것이 필요하며 많은 통계, 특히 사회부문, 환경부문의 통계의 개발이 요청되고 있다.

그리고 이제까지 경제개발중심의 정책으로 경제통계에 치우쳐 사회부문의 통계가 부족한 점 등 국가기본통계가 균형있게 발전하지

못하였다. 이는 우리나라의 전체 승인통계중 경제·산업부문의 통계가 상당부분을 차지하고 있으며 특히 보건, 사회, 복지부문통계는 주로 행정보고통계에 의존(전체 55종중 보고통계가 35종으로 64%)하고 있는 것을 보아도 알 수 있다.

환경통계의 정비 및 개발 또한 시급한 과제이다. 환경은 삶의 질을 측정하고 평가하는데 필수적인 요소이지만 현행의 거시경제지표(GDP 등)로는 환경오염이나 자연자원의 감모 등에 의한 국민생활에 대한 질의 변화를 반영하기 어렵다. 현재 환경통계는 승인통계 9종, 행정보고통계 17종 등이 있지만 국민경제활동과 환경과의 상호작용을 파악할 수 있는 체계도 없어 환경정책수립에 애로가 많은 것도 사실이다.

한편 산업구조의 변화에 따라 비중이 계속 높아지고 있는 서비스업 부문의 통계가 매우 취약하며 금융, 보험, 증권 등의 통계를 제외한 정보산업, 부동산업, 사회서비스업 부문의 통계가 크게 부족하고 국제간에 서비스 교역의 증가로 인력, 기술, 지적재산권에 대한 교역정보가 주요 관심사로 부각되고 있으나 이에 대한 통계도 거의 전무한 상태이다.

(3) 통계작성 주기상의 문제점

최근에 급격한 경제성장으로 국민경제 각 부문의 스톡구성이 크게 변화하고 또 사회간접자본의 확충과 주택보급의 확대 등 복지국가화에 관심이 높아지면서 국부통계의 필요성이 크게 강조되고 있다. 그럼에도 불구하고 우리나라의 국부통계는 현재 10년 주기로 작성되고 있을 뿐 연간 시계열로 작성되지 않고 있다. 국부통계의 미비는 국민대차대조표의 작성을 어렵게 하고 국민경제의 스톡이 플로우에 미치는 자산효과를 올바르게 파악할 수 없게 함은 물론, 국제기준에 맞는 국민계정체계의 작성 또한 어렵게 하고 있다. 따라서 국부통계의 편제는 인적·물적자원이 대량으로 소요되는 방대하고 어려운 작업이라고 하나 국민경제에서 이 통계가 갖는 중요도를 감안할 때 미

국·일본 등과 같이 연장지표를 이용하여 매년 작성하는 것이 바람직하다.

(4) 통계의 계절변동조정계열 미작성

미국과 일본 등 주요 선진국에서는 경제통계의 대부분을 계절조정을 하여 원계열과 함께 발표하고 있다. 이러한 계절조정 시계열은 원계열로부터 계절변동요인을 제거함으로써 경제변수의 움직임을 제대로 파악하는데 그 목적이 있다. 그러나 현재 우리나라에서는 실업을 등 일부 경기지표를 제외하고는 계절조정자료를 작성하지 않고 있으며 이에 따라 경제분석을 실시할 경우 경제학자들은 분석목적에 따라 다양한 방법으로 계절조정을 실시하고 있다. 계절조정방법에는 여러가지가 있지만, 현재 가장 많이 쓰이고 있는 방식은 캐나다 통계국이 개발한 X-11 ARIMA이다. 그러나 이 패키지는 계절조정방법이 매우 복잡하고 옵션에 따른 자의성도 많이 개입된다는 문제점을 안고 있다. 또한 이 방법은 선진국가의 실정에 맞게 개발되어 있어 특히 태음력과 양력을 동시에 사용하는 우리나라에서는 명절과 공휴일을 조정하는 방법이 매우 어렵다. 일본의 경우 X-11 ARIMA가 갖는 문제점을 개선하기 위하여 자국의 실정에 맞는 MITI법 및 EPA법을 개발하여 경기지표 등에 이용하고 있다. 따라서 우리나라도 우리실정에 맞는 계절조정법을 개발하는 것이 시급하며, 통계의 이론적·기술적 분야도 부단히 개발해 나가는 것이 필요하다.

4. 기초통계의 확충 및 활용방안

4.1. 통계인력 및 예산의 확충

우리나라의 통계관련 인력현황은 1994년 2월 1일 현재 92개의 작성기관에 5,494명(정부기관+민간지정기관)이고 이들에 의하여 작성되는 승인통계는 372종(1996. 1. 1 현재)에 이른다. 그러나 극히 일부를 제외한 다른 통계작성기관이 자체적으로 통계를 기획하고 조사

하는 기능을 수행한다는 것은 매우 어려운 실정이다.

우리나라의 통계인력 및 통계관련 예산현황 (1994. 2. 1현재)

	통계작성기관수	인력(명)	예산(백만원)
- 정 부 기 관	41	4,697	13,976
○ 중앙행정기관	26	3,977	10,642
○ 지방자치단체	15	720	3,304
- 민 간 지 정 기 관	51	797	5,523
계	92	5,494	19,469

자료 : 「'94년 통계활동현황조사」 결과(통계청, 1994. 9)

또한 통계인력을 주요 선진국들과 비교해 보면, 인구 100만명당 통계인력은 우리나라는 106명으로 독일 222.3명, 일본 121.2, 캐나다 241.3명, 네덜란드 197.6명 등에 비해 크게 부족한 편이다.

주요국과의 통계인력 비교

(단위 : 명)

		전체통계인력 ¹⁾	인구백만명당 통계인력
한	국	4,672	106.0
독	일	17,230	222.3
일	본	14,964	121.2
캐	나	6,400	241.3
네	덜	3,000	197.6

주 : 1) 중앙통계기관과 타부처 및 중앙은행, 지방조직을 합한 것이며, 민간지정기관은 제외된 것임.

자료 : 제1회 「통계의 날」 기념논문집(통계청, 1995. 12)

한편 우리나라의 통계관련 예산도 통계인력과 마찬가지로 상당히 부족한 편이다. 아래의 표를 통해 주요 EU국가들과 비교해 볼 때 우리나라의 통계예산이 상당히 열악하다는 것을 알 수 있다.

주요 EU국과의 예산 비교

(단위 : %)

		GDP대비 통계예산	
		중앙통계청	통계부처합계
한	국 (1993)	0.01	—
독	일 (1989)	0.03	0.04
스	페 인 (1991)	0.03	0.05
네	덜 란 드 (1990)	0.05	0.05
영	국 (1991)	0.01	0.03
프	랑 스 (1990)	0.02	0.03

주 : ()내는 비교연도임
 자료 : EU 12개국의 통계제도(통계청, 1996. 1)

우리나라가 OECD에 가입하게 되면, 작성통계중 인구통계를 제외한 대부분 통계를 개선해야 할 것으로 파악되고 있다. 특히 환경통계를 비롯, 외국인 투자기업의 산업활동통계, 해외원조통계 등 17종의 통계는 새로이 개발해야 한다. 그러나 현재의 인력과 적은 예산으로는 기초통계의 한계를 극복하고 양질의 통계를 공급하기에는 매우 어려운 상황이다.

우리나라가 명실상부한 선진국 대열에 합류하고 세계화를 앞당기기 위해서는 통계에 대한 정책적인 배려와 과감한 투자가 요구된다. 선진국의 통계담당기관은 많은 고급인력의 보유를 통해 질 높은 통계의 생산은 물론 연구·분석에도 상당한 투자가 이루어지는 등 행정의 과학화에 앞장서 나가고 있는 실정이다.

따라서 우리나라를 선진사회로 이끌어 가는데 반드시 이루어야 할 통계분야의 선진화를 위하여는 자료생산의 측면에서 통계조직을 확대하고 소요되는 예산을 합리적으로 확보·활용함은 물론 통계연구소의 설립 등을 통해 통계의 이론적·기술적 분야도 부단히 개발해 나가야 할 것이다.

4.2. 정보공개 확대

정보화의 진전은 정보의 대량축적과 신속한 유통을 가능하게 하였다. 정보는 일상생활에 직·간접적으로 큰 영향을 미치며, 이에 따라 이를 소유한 자와 그렇지 못한 자간에는 큰 차이를 가져올 수 있다. 과거 우리나라에서는 각종 통계를 발표하지 않거나 제한된 수치만 공개함으로써 이로 인해 사회적·경제적 낭비와 손실이 생길 수 있다고 하겠다.

최근에는 과거와 달리 많은 기초통계가 일반인들에게 공개되고 있으며 이는 계속 확대되고 있는 추세이다. 그러나 아직도 많은 통계가 여러 가지 이유로 일반인들에게 공개되지 못하고 있는 실정이다.

따라서 국내 각 기관에서 생산되는 각종 통계정보를 축적·제공하는 통계 데이터베이스를 이용기관의 다양한 목적에 맞도록 개편 및 확충하고 세계화를 지원하기 위하여 국제기구(OECD, IMF 등)와 주요 선진국으로부터 각종 통계정보를 신속히 입수하여 각종 매체를 통하여 제공할 필요가 있다. 그리고 통계작성기준, 조사기법, 조사항목의 국제화로 자료의 국제간 비교성을 제고하고 국제기구에서 권고하고 있는 신기준의 조속한 적용을 추진하여야 할 것이다.

4.3. 기초통계의 신규 개발

국민생활의 질을 측정할 수 있는 사회복지통계의 개발과 쾌적한 생활과 환경보호에 필요한 기초 환경통계 및 지표를 개발하고 서비스업부문 통계 등의 기초통계를 생산할 때 가공통계·분석통계의 생산지원에 역점을 두는 등 다양한 기초통계의 생산으로 국가통계의 균형발전을 도모하여야 하겠다.

그리고 대부분의 선진국에서 이미 사용하고 있는 계절조정방법을 우리나라 실정에 맞게 개발하여 정확한 경제분석 및 경기 판단에 필요한 여건을 마련해야 하겠다. 또한 지방화시대에 필요한 지역통계를 체계적으로 작성할 필요성이 증대되었는데 중앙기관에서는 대규

모 조사를 통하여 세분된 지역통계를 생산하여 제공하고 지방정부에서는 표본조사를 개발하여 자체적으로 활용토록 하여야겠다.

한편 최첨단 전산장비를 활용하여 행정업무를 전산화하고 이를 이용하여 다양한 행정통계를 생산할 뿐만 아니라 행정통계의 DB를 각 부처에 구축한 후 상호연결하여 각 기관이 다목적으로 이용할 수 있도록 해야 할 필요성이 있다.

4.4. 통계 공표시기의 조기화

통계조사 결과에 대한 통계이용자의 다양하고 신속한 제공 요청에 부응하기 위해서는 공표항목의 선택에서부터 통계 집계기간의 단축이나 통계조사의 기획·설계로부터 공표까지 통계조사의 전반에 걸친 개선을 통해서 공표시기의 조기화를 적극 모색해야 할 것이다.

또한 경기지표로 이용되는 통계 등 특히 조기공표에 대한 수요가 큰 것에 대해서는 부분적인 집계치나 속보치 등의 공표도 적극적으로 검토해야 할 것이다. 또한 공표일자를 외국처럼 지정하는 방안도 검토되어야 하겠다.

4.5. 통계 제공방법의 개선

현재는 컴퓨터와 각종 통신기술이 급속도로 발전하고 있으며 거의 모든 기관과 국민들이 이를 이용하고 있다. DB를 구축하는 목적은 데이터를 축적하는 목적도 있지만, 이를 이용하는 사람들의 편의를 도모한다는 목적도 있다. 실제 방대한 자료를 수집하여 경제분석을 하고자 할 경우 분석자들은 사용하고 있는 패키지에 적합하도록 각종 통계자료책자와 일부 통계생산기관의 데이터베이스를 이용하여 자료를 정비하는데 많은 시간을 투입하고 있는 실정이다. 만약, 미국과 일본 등 선진국들처럼 자료가 잘 정비되어 있고 사용자에게 맞게 서비스가 이루어진다면 경제진단이나 예측 등 경제분석을 하는데 많은 시간이 절약되고 그 만큼 연구의 질도 제고될 수 있는 것이다.

21세기에는 정보의 원천이 되는 통계자료의 수집, 축적, 가공, 분

석 등의 활동이 매우 중요한 자산가치를 가질 것이며 단순한 상품이 아닌 지식과 정보가 국제경쟁력의 핵심이 될 것이다. 이와 함께 통계정보의 역할이 한층 중요해지고 있기때문에 국민 각계각층의 통계에 대한 다양한 수요에 적절히 대응해 가면서 신속하고 이용하기 쉬운 형태로 자료를 제공해 갈 필요가 있다.

5. 맺음말

최근에 우리나라 통계의 수준은 과거의 단순한 통계자료의 생산에서 본격적인 통계분석을 통해 고급정보를 제공하는 등 급격히 향상되고 있다. 또한 통계자료의 이용을 편리하게 해주는 서비스적 측면도 많이 개선되고 있으며 이와 같은 추세는 앞으로도 계속되어지리라 믿는다.

그러나 다른 한편으로 우리나라의 기초통계는 세계화의 진전 및 OECD 가입을 앞두고 앞에서 지적한 바와 같은 여러 가지 문제점들에 직면해 있다. 즉 다양한 통계에 대한 수요가 증증할 것으로 예상되고 특히 환경통계 및 사회복지통계에 대한 필요성은 더욱 절실해지고 있으며 이에 따라 통계기관의 통계업무도 급증하리라 생각된다. 이러한 문제점들을 해결하기 위해서는 통계인력 및 예산의 확충, 정보공개의 확대, 주요한 기초통계의 신규개발 등의 노력이 꾸준히 필요하리라 생각된다.

그러나 통계의 질적 개선은 작성자나 작성기관의 노력뿐 아니라 국민의 이해와 협력이 있어야 비로소 그 목적이 실현된다. 기본적으로 양질의 통계는 통계조사에 대한 국민의 성실하고 정확한 응답에 달려있기 때문이다. 그러므로 통계에 대한 국민의 인식이 올바른 방향으로 바뀌도록 유도하여야 하며 이를 위해서는 통계의 보급·개발, 통계교육의 충실은 물론 기업자료 등의 기밀보호와 국민의 프라이버시의식 등을 배려하면서 개별정보의 보호를 확보하는 일이 무엇보다도 중요하다 하겠다.

통계의 시의성과 정확성

이 재 창

고려대학교 통계학과 교수

1. 서 론

실험계획을 통해 얻게 되는 결과나 자연현상 또는 사회 현상을 관찰하는 것은 모든 과학에서 무엇보다 중요하다. 현상을 관찰한다는 것은 엄밀히 말한다면 측정행위가 된다. 개수를 헤아리는 것이나, 전압이나 온도와 같이 도구나 기기를 이용해 현상의 어떤 특성을 약속된 단위의 숫자로 표현하거나 또는 행정조직을 통해 보고된 내용의 집계, 그리고 대화와 같은 방법으로 상대방의 생각을 정성적 또는 정량적인 척도로 반응을 기록하는 것들은 모두 광의의 측정이다.

통계란 이러한 관찰로 얻어진 데이터를 토대로 가공한 정보라고 말할 수 있다. 따라서 통계와 측정은 불과분의 관계를 가지며 측정이 제대로 됐을 때에만 이로부터 얻어진 통계가 의미를 가질 수 있다. 측정된 결과를 우리는 데이터라고 하는데 이렇게 측정이란 통계 생산 과정에서 가장 중요한 원료가 된다.

측정이란 원래 사물의 어떤 특성을 약속된 단위(單位)로 숫자화하는 과정(process)을 말한다. 어떻게 보면 이런 정의와 개념이 자연과학에만 적용되는 것으로 생각하기 쉬우나 실은 같은 개념이 사회과학에도 엄밀히 적용되어야 할 것은 쉽게 이해할 수 있다. 자연과학에서의 측정은 과학의 발달과 꼭 같은 속도로 발전해왔다. 과학의 발달정도는 측정 능력으로 직접 평가되기도 한다. 여기에는 자연과학의 측정은 측정과 관련된 많은 요인(要因)을 가능한 한 불변(不變)인 상황에서 측정기술을 고도화시켰고 측정의 정확도는 엄청나게

발달하고 있는 것이다. 반면에 사회과학의 측정이란 많은 경우에 측정 대상에서부터 측정환경은 물론 측정방법 등이 지속적으로 변하고 있어 “측정기술”의 발달보다도 이들의 변화에서 오는 어려움이 측정 결과에 압도적으로 더 큰 영향을 주고 있다.

물론 자연과학에서 측정은 측정과 관련된 환경을 상당수준으로 불변인 안정 상태를 유지하며 가능한 한 환경의 변화에 크게 영향을 받지 않도록 측정기술을 발전시켜왔다. 가령 길이라는 특성을 살펴보면 역사적으로 이와 관련된 단위는 여러 차례 바뀌어왔다. 현재 쓰고 있는 길이의 단위는 SI에 정의된 바와 같이 크립톤(Kr)86램프로부터의 등직선을 택해 이 파장의 1,650,763.73배가 1미터라고 한다. 이렇게 정의된 길이의 단위를 재현하는 능력은 그 나라의 과학기술의 수준을 나타내고 있다. 즉 측정을 하나의 과정으로 보면 이 과정을 동질적인 것으로 유지하면서 실험을 통해 측정결과를 얻어내게 된다.

측정이란 이와 같이 숨겨진 고정된 참값을 찾는 일회적 행위가 아니고 주어진 조건에서 반복되는 관측으로 이의 특성을 나타내는 대표값을 추정하는 과정(process)이다. 참값의 의미는 합의되기 어려운 개념이다. 그러나 측정행위가 통제상태(under control)에 있는가는 측정결과를 받아들이거나 이를 이용하여 또 다른 의사결정을 하는데는 매우 중요한 판단기준이 된다. 측정이란 하나의 과정이 통제상태에 놓이고 경제적으로 수궁할만한 수준에 있는가는 많은 조건과 기준으로 검토함으로써 가능하다. 여기서 “경제적으로 수궁 가는 수준”이란 측정 결과를 이용하여 의사결정을 했을 때 측정의 정확도가 미흡하여 의사결정에 오류를 범하게 되고 이때 기대되는 손실과 연결된다. 따라서 여기에는 「사용목적에 적합한 정확도」라는 개념으로 각각의 경우에 판단기준은 같지 않아 확일적으로 다루기는 어렵다. 그러나 일반적으로 최종추정치가지는 정확도를 분산과 편향(bias)으로 나타내고 있으나, 의사결정 문제와 비교했을 때에는 이용되는 측정 결과의 시의성(timeliness)에 의해 판단될 수 있고, 또 다른

경우에는 총계된 결과치의 형태(일반적으로 Aggregate된 것)가 사용자의 의사결정 문제와 관련하여 얼마나 적절한지에 달렸다고 볼 수 있다. 따라서 통계의 유용성을 판단하는 기준으로는 통계의 정확성, 시의성 그리고 관련성을 들 수 있고 정확성은 다시 측정과정으로 설명된다. 통계의 사용목적에는 통제(control), 평가, 상황판단 및 예측, 그리고 정책입안과 같은 의사결정을 들 수 있다. 이들 각각의 경우에 정확성, 시의성 및 관련성은 기본적인 판단기준으로 된다. 물론 사안에 따라 이들의 수준이 달라질 것이며 많은 경우 시의성과 정확성은 주어진 일정 수준의 비용으로는 서로 상반적 관계(Trade-off)가 있다. 정확도의 제고를 위해서는 시의성이 어느 정도 희생되는 경우는 경제통계에서는 흔히 있는 일이다.

사회과학의 통계와 자연과학의 통계는 다르다고 생각할 수 있으나 통계 생산 과정의 「측정」면에서 보면 공통적인 문제도 많으며 또한 자연과학 측정의 경험으로부터 여러 가지를 배울 수도 있을 것이다.

2. 측정과 측정과정

자연과학에서는 측정결과 그 자체에 대한 평가보다는 이들이 얻어진 측정과정에 대한 평가로 측정의 신뢰성이나 정확성이 논의된다. 따라서 정확도 또는 불확도(Uncertainty)에 관한 서술에 앞서 측정과정의 특성을 살펴볼 필요가 있다.

정확도를 논하기에 앞서 측정과정(measurement process)의 두 가지 특성을 이해하는 것이 편리하다. 첫째로 동일한 측정방법, 절차, 기기, 측정자 및 측정환경을 가진 측정과정을 통해 동일한 대상을 반복하여 측정하더라도 이들의 값은 항상 같지 않고, 둘째로 두 개의 측정과정으로 구한 측정치들은 극한평균치(limiting mean), 다시 말해 많은 값의 평균치는 같지 않다.

측정과정을 구체적으로 명확히 정하기 위해서는 아래와 같은 몇 가지 질문을 자문하여 보면 현재의 측정과정이 잘 설정되어 있는지

를 알 수 있다.

(1) 안정상태에 있는 사물을 같은 기기로 재측정했을 때 어떤 주어진 구간 내에서 첫번값과 일치할 것인가?(반복성 : Repeatability)

(2) 만일 재측정이 일정시간이 경과한 후에 행해졌다면 일치성에는 변화가 없겠는가?

(3) 만일 동일 제조자가 만든 다른 기기를 사용한다면 무슨 결과를 기대할 것인가?

(4) 타제조자가 만든 기기를 사용한다면 결과에 어떤 차이가 기대되는가?(재현성 : Reproducibility)

(5) 기하적 조건(측정의 방향 등)이 측정에 어떠한 영향을 주는 지?

(6) 환경조건(온도, 습도, 기압 등)은 어떤 영향을 주는지?

(7) 측정결과가 측정절차에 어떻게 의존하는지?

(8) 측정자가 바뀌면 측정치에 어떤 차이를 초래하는지?

(9) 기기의 편향이나 유지용 표준(reference standards)은 교정되었으며 그 불확도는 최종 측정치에 어떻게 나타나는지?

등의 질문을 해보면 측정과정은 어떻게 정의되어야 할 것인지에 대한 대략의 개념을 얻을 수가 있다.

2.1. 측정 방법

측정방법에 따라 측정과정은 결정된다. 측정방법과 관련하여 고려되어야 하는 사항은 아래와 같은 것들을 들 수 있다.

(1) 단위의 선정

(2) 측정의 물리적 체계

(3) 교정곡선의 문제(Calibration Curve)

(4) 편향(Bias)

(5) 구체성(Specificity)

(6) 교란(Interference)

(7) 양립성(Compatibility)

(8) 소급성(Traceability)

2.2 측정과정의 횡적비교

위에서는 측정방법의 개발과 관련된 문제를 논의하였으나 실제 측정방법이 채택되고 측정과정이 현장에서 운용되기 위해서는 측정실험실간의 수평적인 일치가 필요하게 된다. 각종의 측정 데이터가 발표되었을 때 측정치의 질(quality)이나 이에 대한 보증방법이 적절히 유지되어야만 하며 이들 방법 등이 표준화가 되고 각 측정실험실의 결과가 의미 있게 하기 위해서는 이들간의 비교가 신빙성 있게 행하여져야만 한다.

측정과정이란 일련의 숫자만을 산출하는 것이 아니고 측정대상이 되는 특성과 관련하여 유용한 정보를 가져다주어야 하는데 데이터의 유용성은 이의 신빙성에 크게 의존하는 것이다.

흔히 생기는 일로 두개 이상의 연구기관이나 정부기관에서 발표한 결과가 서로 상이하여 어느 것을 믿기도 어렵다면 이는 이를 응용하는 사람들에게는 매우 당혹스러운 일이 아닐 수 없다. 따라서 측정과정의 소상한 정보는 최종이용자에게는 매우 중요한 정보가 된다. 한 측정기관의 신뢰성은 타기관과의 비교에서 가능한 것으로 이러한 비교는 측정기관의 능력향상과도 직결된다.

3. 통계의 오차와 시의성

앞에서는 자연과학에서의 측정과정을 설명했다. 그러나 이러한 개념은 많은 경우에 사회과학의 측정에서도 따져 볼 수 있는 것이다. 사회과학에서는 오차에 관해 많은 연구가 진행되고 있으며 오차를 줄이려는 노력과 오차를 표현하는 방법도 많이 발전해 왔다. 경제통계에 관한 연구로 유명한 Osker Morgenstern은 1963년 출판한 그의 저서 「On The Accuracy of Economic Observations」(경제적

관측의 정확도에 관하여)에서 “통계학적인 분명한 의미는 정의되어 있으나 일반적으로 오차(Error)란 서술에 있어서 불완전(imperfect)하거나 불충분(incomplete)한 표현”이라고 쓰고 있다. 자연과학에서 말하는 오차보다는 협의로 생각할 때 그는 Th nen이 한계효용이론과 관련하여 토지와 관련된 실험을 실시함에 이 모든 과정을 자세한 기록을 보존한 예외는 인정하나 일반적으로 경제통계란 실험 계획에 의한 결과가 아니고 단지 기업이나 정부의 활동에서 파생되는 부수물로서 측정, 묘사, 기록된 것으로 경제학자가 알고 싶어하는 정확한 상황과는 다르다고 이야기하고 있다. 이들은 보편적으로 경제적인 정의나 과정에 의존하기보다는 법적인 의미에 의존하는 경향이 있다.

자연과학과 사회과학에서의 데이터 사용에서 보여주는 특징중의 하나는 자연과학에서는 데이터 생산자와 사용자가 보통 일치하지만, 사회과학의 경우는 그러하지 않다는 점이다. 따라서 데이터를 생산한 과학자는 이렇게 얻은 데이터로부터 최대한의 정보를 꺼내야 하고 이 데이터가 남들에게 공개되더라도 이미 더이상 새로운 분석이 필요 없다고 생각되어야 한다. 뿐만 아니라 그는 데이터 생산과정의 모든 것을 소상히 알고 있으며 실험에 참여한 모든 사람의 능력을 알고 있어 데이터의 질에 관해서도 신뢰도나 정확성을 판정하는 데는 어려움이 없다.

반면에 사회과학에서는 이와는 전혀 다르다. 데이터를 사용하는 사람은 데이터에 관한 자세한 특성이나 과정을 알지 못하는 것이 보통이다. 뿐만 아니라 데이터를 생산(수집)하는 사람과 이를 가공(분석)하는 사람도 전혀 다르다. 전통적으로 사회과학 데이터 사용자들은 데이터 생성과정이나 이에 따른 통계계산과정에 대해서 알아보고자 하지도 않는다.

경제통계의 경우를 보면 데이터가 수집된 시점이나 상황에 관해서 자세하고 정교하며 객관적으로 서술한 것은 별로 없고, 비슷한 상황에서 데이터를 반복적으로 얻고자 노력하지도 않는다. 재현성이란

일반적으로 가능하지 않다. 그러나 경제통계의 특징중의 하나는 이들이 장기간 동안 축적되어 시계열을 이루는 것으로 이는 주어진 시점에서의 정확성보다 시계열상의 과정에서 과거, 현재, 미래가 서로 비교될 수 있고 지역간의 비교도 가능케 하는 비교성을 유지하는 것이 더욱 의미가 있다고 볼 수 있다.

경제통계와 관련된 측정에서는 표본으로 추출된 기업이나 개인이 주어진 설문에 응답하는 형식이거나, 관련된 행정조직을 통해 보고한 것을 집계하는 경우가 많다. 어느 경우이건 측정은 이루어져야 하며 측정과정은 통제상태(under control)에 놓여야 한다. 그러기 위해서는 자연과학의 측정과정에서와 같이 몇 가지 중요한 질문으로 측정이 제대로 이루어지는 지를 파악할 필요가 있다. 여기서 중요한 것은 질문자와 응답자와의 상호작용에 의해 데이터가 수집되며 어떤 경우에는 질문자에 의해 측정되며 또 다른 경우에는 응답자가 스스로 측정하기도 한다는 점이다. 결국 사람과 사람의 상호작용이 측정 과정에 개입되어 자연과학의 경우보다는 매우 복잡하다. 뿐만 아니라 많은 경우에는 기록된 결과로부터 응답이 도출되기도 하지만, 경우에 따라서는 응답자의 기억에 의존하거나 주관적 판단에 의해 대답되기도 한다. 따라서 측정과정을 자연과학에서와 같이 정형화된 모형으로 만들기는 어렵다. 그러나 몇가지 요인은 측정과정의 안정성을 보여주고 이로부터 의미있는 데이터를 공급해 줄 수 있는 지를 판단하기 위해 꼭 밝혀져야 한다.

최근의 추세는 개인의 사생활(privacy) 보호를 저해하지 않는 수준에서 원시자료(raw data)를 공개하여 통계이용자가 경우에 따라서는 자기가 필요한 정보를 직접 가공해 쓸 수 있도록 하는 것이다. 또한 데이터 수집과정을 소상히 밝혀야 하는 이유는 이로부터 가공된 통계의 정확성을 판단하는 중요한 기준으로 될 수 있기 때문이다.

Morgenstern은 경제통계는 몇몇 예외적인 경우를 제외하고는 「실험계획」의 결과가 아니라고 지적하고 있다. 그리고 경제통계를 포함

한 사회과학적 통계에서의 어려운 점 중에서 또 하나는 「의도적인 허위응답」으로 들고 있다. 그는 이와 관련하여 경제통계 생산에서의 조사자의 전문성결여를 들고 있으며, 설문지의 문제점, 경제통계의 속성상 대량관측(massive observations)의 불가피성과 이에 따른 통계생산과정에서 부수되는 문제들, 측정의 대상이 되는 현상의 정확하고도 객관적인 정의가 결여된 경우, 분류상의 어려움, 측정시점의 결정이나 반복측정간의 시간적 간격 등이 오차의 중요한 근원이 된다고 하였다. 이러한 것들은 통계학에서는 비표본오차(nonsampling errors)로 불리우는 것들로 수식이나 숫자로 표현하려는 노력은 있지만 아직 큰 성과를 보지는 못하는 것들이다. 이에 관한 논란은 이들이 계통오차(systematic error)인가 또는 우연오차(random error)인가로 부터 시작하여, 이들 오차가 통계가공단계별로 전달(propagation of error)되는 것을 추정하는 문제에까지 이르고 있다.

그러나 경제통계의 경우에는 경제적 의사결정이, 다른 사회과학의 경우와도 같이, 정량적 정보(quantitative information)에만 의존하지 않고 정성적 정보(qualitative information)를 동시에 이용하기 때문에 오차의 수량적 서술보다는 데이터수집의 과정과 이에 관련된 상황을 설명해 주는 것이 바람직하다.

경제적 의사결정에 또한 중요한 요소로는 통계의 시의성을 들 수 있다. 많은 경우에 정확정보다 시의성이 더 큰 문제로 대두될 수 있다. 의사결정이나 정책집행이 적절한 시기를 놓치면 아무리 정확한 통계에 의한 결정도 이에 따른 손실을 회복하기 어렵다. 가령 기업체들을 대상으로 하는 통계조사에서 회수율(응답율)을 높이기 위해 상당기간을 기다리며 집계를 늦출 수 있다. 응답율이 높으면 일반적으로 정확성도 높아지게 된다. 이렇게 필요한 시점에 통계를 생산, 보급하는 시의성과 높은 응답율로 얻어지는 추가적 정확성은 서로 상반되는 기준(Trade-offs)이 된다. 따라서 모든 조건(비용 등)이 일정한 상황에서는 사용목적에 따라 어디에다 중점을 두어야 하는지

를 조심스럽게 판단해야 한다. 통계를 생산하는 주기의 결정도 이에 버금가는 중요한 문제이다. 조사 및 생산주기의 선정은 비용의 측면에서 뿐만 아니라 앞서 이야기한 바와 같이 어떤 조사에 소요되는 최소시간(time lags)도 고려해야 한다. 뿐만 아니라 기업의 경우와 같이 특정한 시점(예를 들면 세무보고가 필요한 시점 등)을 고려해야 얻고자 하는 데이터가 수집될 수 있다.

Steven Landefeld(1996)는 미국상무성 경제분석국 (Bureau of Economic Analysis)의 경우를 들어 비용을 줄이고 응답자의 부담을 줄여주기 위해서는 가장 시의적절한 데이터는 상대적으로 가장 작은 표본으로 구해진 월별통계라고 했다. 따라서 매분기, 또는 매년 생산되는 통계는 월별통계에 사용된 표본보다 크고 정확도 또한 높아 월별 통계의 사후 수정에 이용된다. 5년 주기로 Benchmark로 활용되는 경제활동센서스에 의해 앞의 통계들도 사후에 수정이 이루어진다. 따라서 시의성에 의해서 정확성이 결정되는 셈이다. 그렇게 볼 때 통계의 정확도에 관한 수준은 몇 가지 차원에서 보아야 한다고 했으며, 이들은 주어진 시점의 정확도, 경제적 성장의 한 주기로 볼때의 정확도 그리고 장기적으로 경제성장의 추세를 감안했을 때의 정확도로 구분하여 평가하여야 한다고 한다.

통계의 시의성과 정확성의 연구는 경제통계에서는 가장 중요한 척도로 볼 수 있으며 앞으로 심도있게 다루어져야 할 것이다. Jacob Ryten(1966)은 그의 연구결과에서 시의성과 정확성의 최종적인 결정은 주관적인 것이라고 결론을 짓고 있다. 그러나 그는 경제통계의 사용자들이 시의성과 정확성을 동시에 상향적으로 요구하고 있다는 것을 문제점으로 들고 있다. 이런 요구를 적절히 수용할 수 있도록 하는 방안은 예산을 늘이는 것 이외에는 현재로서는 알려져 있지 않다.

4. 결 론

통계의 정확성 문제는 어제 오늘 제기된 것은 아니다. 자연과학의

측정과 같은 모형을 사회과학적 응용에는 도입하기가 쉽지는 않지만 유사한 연구는 필요하다고 본다. 사회과학의 통계는 그동안 조사방법이나 표본추출이론 등이 크게 발전해 오고 있지만 사회나 인간의 복잡성은 이러한 방법의 발전보다 더욱 빠른 속도로 심화되고 있어 결과적으로는 통계의 질이 떨어지고 있다는 이야기까지도 나오게 된 것이다. 사회과학의 통계는 모든 사람의 관심을 불러일으킬 만큼 일상생활과 밀접한 관계를 갖고 있다. 뿐만 아니라 정치적 불신이나 사회적 무관심, 개개의 이기적 행동 등은 이런 모든 문제를 악화시키고 있다. 한 나라의 자연과학의 발달정도를 알기 위해 그 나라의 과학측정 능력을 보듯이 한나라의 사회적 수준은 그 나라가 생산하는 경제통계의 질에 의해 판단될 수도 있다. 요구에 비해 협조는 줄어들고 있다는 불평은 세계 각국의 통계생산자들이 이구동성으로 하고 있다. 정확성과 시의성의 제고 또한 학문적 연구뿐만 아니라 한 나라의 국민적 지원과 협조가 있을 때 가능하다. 시의성의 문제는 통계가 통계사용자의 사용목적에 적합하게 하며, 사용자가 통계를 이용함으로써 부가적 이익을 최대화하도록 하는 노력을 의미한다. 통계는 이미 이의 존재 자체가 부가가치로 평가되고, 이에 따른 모든 노력도 비용과 효과로 설명될 수밖에 없다.

참고문헌

- (1) Oskar morgenstern (1963), On the Accuiracy of Economic Observations, Princeton University Press
- (2) Steven Landefeld (1996), Accuracy, Timeliness and Rel-
evance, unpublished report
- (3) Jacob Ryten (1996), "Timeliness and Reliability : A neces-
sary Trade off", unpublished Report
- (4) 이재창 (1979), "과학적 측정방법과 데이터의 공신", 「환경과
공해」, vol. 3, No. 1.

Ⅱ. 통계경진대회 입상논문

농구 경기의 그래픽컬 표현

김정수, 김혜숙

(지도교수 : 박상규)

중앙대학교 응용통계학과

〈요 약〉

농구 경기를 전 후반 점수와 더불어 주요한 경기 흐름을 함께 보여주는 그래픽컬한 표현 기법을 제안한다. 이 그래프는 기존의 경기 결과를 위주로 발표되는 것과는 달리 경기의 흐름에 따른 점수의 변동과, 주요한 경기 내용이 함께 표현되기 때문에 경기의 전체적인 내용을 보다 이해하기 쉽고 경기 승패의 원인까지 분석할 수 있는 장점이 있다. 또한 이 그래프는 기존의 스프레드시트 프로그램에서 쉽게 만들어 낼 수 있기 때문에 많은 사용이 기대된다.

1. 서 론

농구 경기는 다른 스포츠에 비해서 경기의 흐름이 매우 빠르고 경기 내용에 변화가 많다. 따라서 많은 경우 경기 결과만을 보고 그 경기의 전반적인 흐름을 이해하기는 어렵다. 즉, 단순히 경기 결과를 보고 A팀이 B팀을 이겼기 때문에 A팀이 B팀보다 우수하다든지 점수의 차이가 많이 나는 경기 결과가 있다고 해서 경기가 일방적이었는지 하는 결론을 내리기가 어렵다는 것이다. 또한 경기 결과와 함께 보여주는 선수들의 개인 기록에도 중요한 정보가 누락될 수 있다는 것이다. 예를 들어 C라는 선수는 득점을 많이 하기는 하나 경기에 별로 영향을 주지 못할 수 있고 D라는 선수는 득점이 적기는 하나 경기에 결정적인 영향을 줄 수 있고, 또한 E라는 선수는 한 개의 실책을 했지만 열 개의 실책을 한 F라는 선수보다 경기에 더 영향을 줄 수 있다는 것이다. 이는 이러한 정보가 경기와 연관되어 고

려되지 않기 때문이다. 이러한 단점을 극복하는데는 경기 결과에 따른 부연 설명이 필연적으로 따라야 하는데 지난 95~96 농구 대잔치 결승전에 관련된, 한 신문에서 발췌한 농구경기 기사의 일부를 참조해 보자.

“기아는 김유택(13점, 5턴공잡기)과 한기범(4점, 5턴공잡기)이 제 공권을 장악하고 ‘농구 달인’ 허재(21점, 5턴공잡기, 5도움주기)와 김영만이 코트를 휘저으며 짜릿한 승리를 거두었다. 기아의 노련미는 경기 막판 한층 돋보였다. 상무 문경은(21점)의 자유투와 이상민(16점, 5턴공잡기)의 골 밑 슛으로 종료 1분 48초를 남기고, 74-70, 4점차까지 따라 잡히는 위기를 맞았으나 강동희의 자유투 2발이 그대로 바구니에 빨려 들어가 76-70으로 불을 껐다. 이어 77-73으로 다시 추격당하던 기아는 종료 58초전에 교체 선수 박규훈이 자유투 2발을 고스란히 성공시켜 79-73으로 달아나면서 승부에 종지부를 찍었다.”

이 신문은 선수들의 개인 별 기록과 함께 위 기사를 싣고 있다. 경기 후반부의 경기 상황이 비교적 상세하게 설명되어 있어 단순히 경기결과를 보는 것보다는 경기 전체를 이해하기 쉽다. 하지만 공간적, 시간적인 한계 때문에 경기 전체의 내용을 위와 같이 부연 설명할 수는 없다.

본 연구에서는 잘 알려진 단순한 그래픽컬한 표현을 통해서 농구 경기를 결과와 더불어 경기 내용까지 보다 잘 이해할 수 있는 방법을 제시하려고 한다.

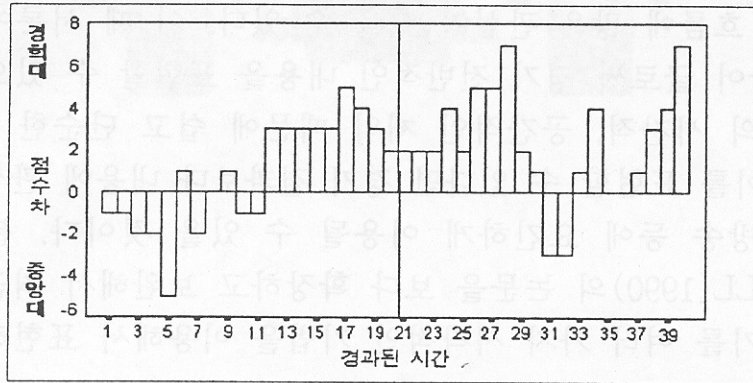
2. 그래픽컬한 표현기법

최근 단순하고 이해가 빠른 시각적인 그래픽 기법들이 복잡하고 장황한 통계적인 수치보다 훨씬 선호 받고 있으며 많은 분야에서 개발되어지고 있다. 농구같이 경기의 흐름이 매우 빠른 스포츠는 단순히 경기 결과나 개인 기록이 크게 의미가 없을 수도 있고 경기 결과

만큼 경기 흐름에 많은 관심이 있을 수 있다. 이 때 서론에서 언급한 바와 같이 글로써 경기 전반적인 내용을 표현할 수 있으나 신문이나 방송의 시간적, 공간적인 제약 때문에 쉽고 단순한 그래프를 이용해서 이를 표현할 수 있다면 경기 결과보다 내용에 관심이 많은 신문이나 방송 등에 요긴하게 이용될 수 있을 것이다. 본 연구는 WESTFALL(1990)의 논문을 보다 확장하고 보완해서 최근 있었던 농구경기를 여러 가지 시각적인 기법을 이용해서 표현해 보고자 한다.

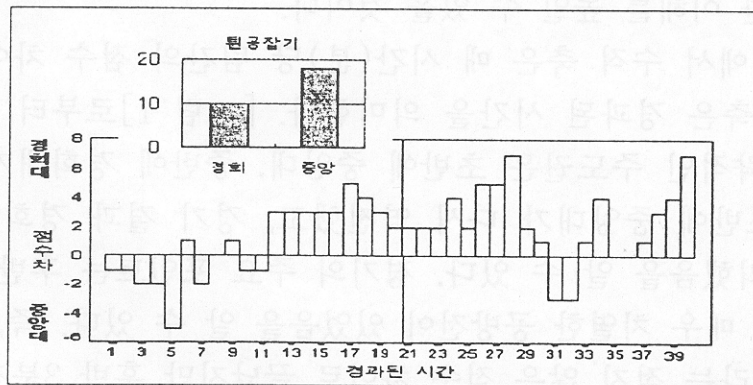
아래 경기는 1996년 대학 농구 연맹전에서 경희대와 중앙대의 경기이다. 전체적인 경기 내용은 경희대가 전반 중반부터 후반 중반까지 경기를 주도했으나 경희대의 실책과 중앙대의 우세한 튄공잡기로 후반 10분경 중앙대가 역전했다. 하지만 11분경 중앙대 주공격수인 6번 선수가 5반칙 퇴장당하고 공방전이 계속되다가 17분경 경희대가 다시 주도권을 잡게 되었다. 이러한 상황을 그래프로 표현함으로써 경기에 대한 이해를 높일 수 있을 것이다.

[그림 1]에서 수직 축은 매 시간(분)당 팀간의 점수 차이를 의미하며, 수평축은 경과된 시간을 의미한다. [그림 1]로부터 40분간의 경기의 대략적인 주도권은 초반에 중앙대, 중반에 경희대가 잡았으며 후반 초반에 중앙대가 다시 역전했고, 경기 결과 경희대가 8점 차이로 승리했음을 알 수 있다. 경기의 주요 포인트는 후반 9분에서 16분까지로 매우 치열한 공방전이 있었음을 알 수 있다. 즉, 경기 결과는 8점이라는 적지 않은 점수 차이로 끝났지만 후반 3분전까지 매우 치열한 경기였음을 알 수 있다.



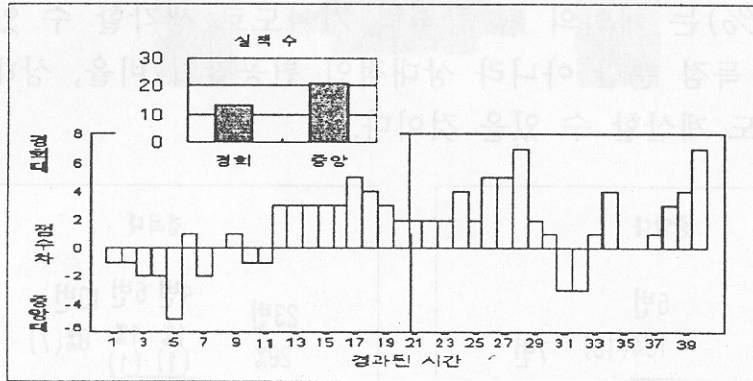
[그림 1]

[그림 2]는 [그림 1]을 수정 보완한 것이다. 즉, 경기의 중요한 승패 요인 중의 하나인 투공잡기의 수를 [그림 1]에 포함시킨 것인데 패한 중앙대가 경희대보다 오히려 많은 투공잡기를 했음을 알 수 있고, 이것이 패인은 아니라는 추론을 할 수 있다.



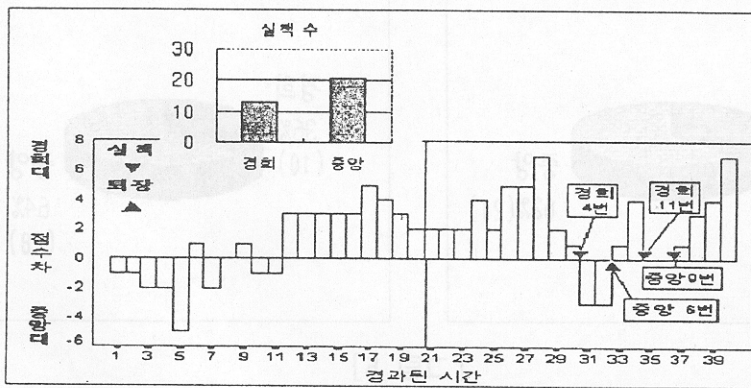
[그림 2]

[그림 3]은 역시 중요한 승패 요인 중의 하나인 실책 수를 [그림 1]에 포함시킨 것이다. 중앙대의 실책이 경희대보다 훨씬 많은 것으로 보아 중앙대의 패인은 실책임을 알 수 있다.



[그림 3]

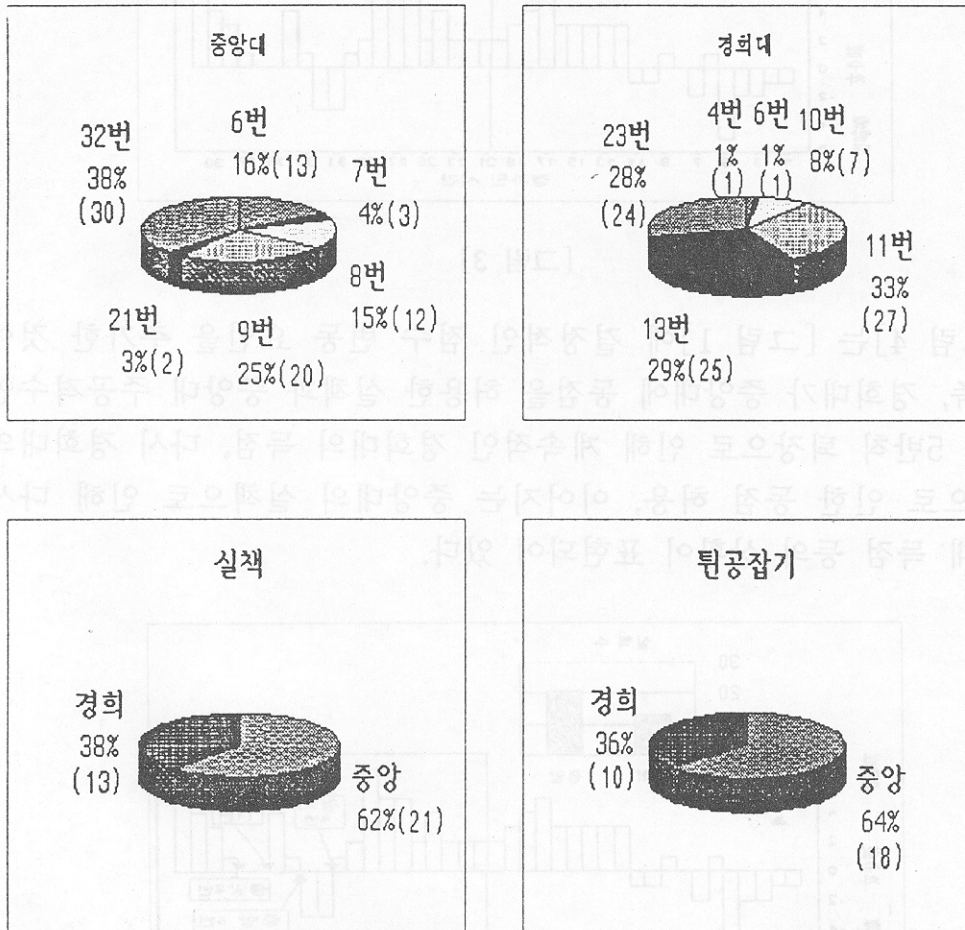
[그림 4]는 [그림 1]에 결정적인 점수 변동 요인을 추가한 것이다. 즉, 경희대가 중앙대에 동점을 허용한 실책과 중앙대 주공격수인 6번의 5반칙 퇴장으로 인해 계속적인 경희대의 득점, 다시 경희대의 실책으로 인한 동점 허용, 이어지는 중앙대의 실책으로 인해 다시 경희대 득점 등의 상황이 표현되어 있다.



[그림 4]

[그림 5]는 각 팀들의 개인 별 득점 및 팀별 실책과 튄공잡기 기록이다. 이 개인별 득점기록은 개개인의 공격 능력을 판단하는데 중요한 의미를 가진다. 절대적인 수치가 전체 팀 중에서의 개개인의 공격 능력을 판단할 수 있다면, 팀 득점에서 나누어 얻어지는 상대

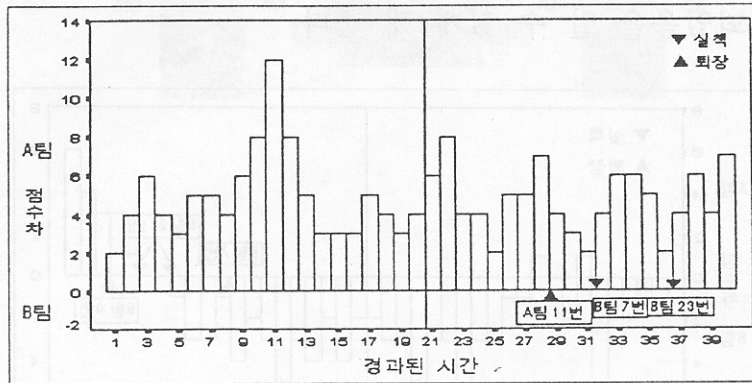
적인 수치(%)는 개인의 팀 공격력 기여도로 생각할 수 있을 것이다. 이것은 득점 뿐만 아니라 상대적인 튄공잡기 비율, 상대적인 실책 비율 등도 계산할 수 있을 것이다.



[그림 5]

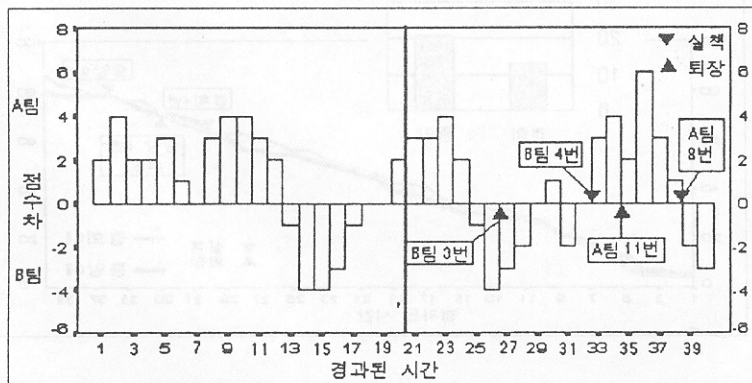
[그림 6], [그림 7], [그림 8]은 위에서 언급한 그래픽컬한 표현 기법을 이용해 몇 가지 농구 경기를 나타내 본 것이다.

[그림 6]은 A팀이 일방적으로 B팀을 이긴 경기이다. 두 팀간의 최고 점수 차이는 12점이고, A팀이 B팀보다 전력상으로 우세하다는 것을 보여 주고 있는 경기이다. 하지만 A팀의 선수가 퇴장당했을 때 B팀은 선수들을 실책으로 좋은 기회를 살리지 못했음을 알 수 있다.



[그림 6]

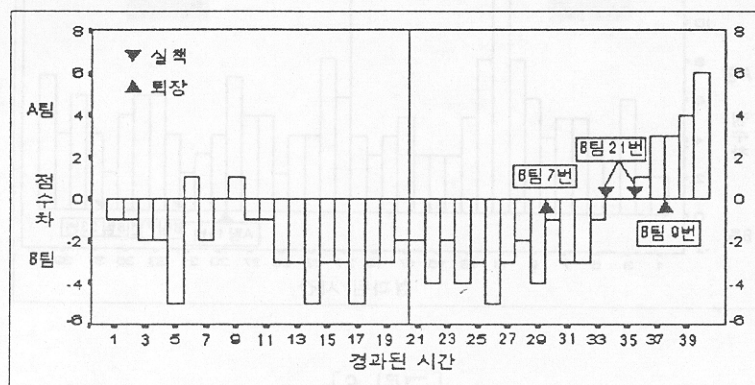
[그림 7]은 전반 12분까지 A팀이 앞서다가 B팀에 의해 역전을 당하고, 다시 후반전 들어 B팀이 역전을 반복하는 치열한 경기였음을 알 수 있다. 후반 6분경 B팀 주전 선수인 3번 선수가 퇴장을 당하고, 결정적인 실책으로 인해 A팀에게 리드를 빼앗겼다. 그러나 후반 15분경 A팀 11번 선수가 퇴장을 당함으로써 B팀은 추격의 실마리를 잡을 수 있었다. 또한 A팀의 실책으로 인해 B팀이 경기를 뒤집을 수 있었다.



[그림 7]

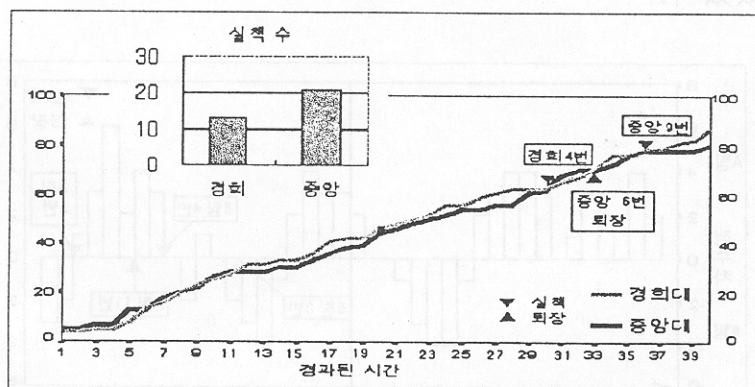
[그림 8]은 B팀이 전체적인 경기의 주도권을 쥐고 있었으나 주전 선수들의 연속적인 퇴장과 연이은 실책으로 말미암아 다 이긴 경기

를 놓치게 되었음을 알 수 있게 해 준다.



[그림 8]

[그림 9]는 [그림 4]의 표시를 시간에 따른 누적 점수의 관점에서 그려본 것이다. 두 누적 점수 그래프가 서로 만나는 점의 수로 전체적인 경기의 흐름을 파악할 수 있고 두 그래프가 매우 근접하면 할수록 경기는 치열했다는 것을 알 수 있을 것이다. 경우에 따라서 시각적인 효과가 더 클 수 있으리라 생각된다.



[그림 9]

3. 결 론

본 논문은 농구 경기를 시각적인 기법을 통해서 많은 정보를 보다 쉽게 전달하고 경기 전반을 보다 쉽게 이해하는 방법을 제시했다. 기존의 방법이 결과 위주의 방법이라면 본 연구 논문에서 제시하고 있는 방법은 경기의 과정에 보다 많은 비중을 두는 표현방법이라 하겠다. 제시된 방법은 경기 결과보다는 훨씬 많은 정보를 주고 있으며 그래프를 만드는 것도 로터스, 엑셀 등 기존의 많은 소프트웨어에서 쉽게 만들 수 있어 많은 활용이 기대된다. 이러한 방법을 잘 활용하면 경기 전체를 쉽게 이해할 수 있을 뿐 아니라 선수 개개인의 활약을 질적으로 평가할 수 있는 기준으로 사용할 수 있을 것이다.

참 고 문 헌

- [1] PETER H. WESTFALL(1990). Graphical Presentation of a Basketball Game. The American Statistician, Vol.44, No.4, 305~307.

나무구조에 의한 B형간염

조기진단 모형연구

이 춘 식

(지도교수 : 이태림)

한국방송대 응용통계학과

1. 서 론

현대인의 질병중 B형간염은 발생순위가 높고, 건강과 생명에 큰 영향을 주는 중요한 질병으로 급성간염의 경우 건강한 다른 사람들에게 직, 간접으로 전파시킬수도 있는 위험이 있다. B형 간염은 1963년 부름버그(Blumberg)에 의해 Australia antigen이라 불리기 시작하여 1970년 초 까지 간염관련항원(HAA)으로 불리어 오다가 1972년 미국 과학원 간염위원회가 종전의 전염성 간염을 A형 간염으로, 혈청간염을 B형간염으로 용어를 통일하였고 이들 병원체를 간염 A바이러스와 간염 B형 바이러스로 명명하였다. 이외에도 비(非)A, 비B, 혹은 수혈에 의해서만 전파되는 비A, 비B간염의 병원체를 간염 C바이러스라 부른다. 그러나 간염 B바이러스는 HB표면항원및 항체, HB크아 항원및 항체등이 개발되어 현재는 백신제조와 활용이 가능해졌다.

근래에 와서는 간염 B바이러스 감염은 여러 형태의 간염뿐 아니라 간세포성 암과 강한 연관성이 있다는 사실이 계속 보고되어 나라에서 감염율이 높아 아프리카의 어느 나라에서는 15%를 상회하는가 하면 미국과 같은 선진국의 HB표면항원 양성률은 0.3%인데 우리나라는 약 8%로 추정되고 있다. 세계에는 약 3억의 인구가 감염되어 있다는 추산이다. 이러한 중요성을 인식하여 우리나라도 1986년 B형

간염을 제3종 전염병으로 지정하였다. B형간염의 역학적 특성을 보면 성별 양성률은 남성이 훨씬 높고, 연령별로는 10대가 높은 경향을 보이다가 40대가 되면 감소한다. 지역별로는 농촌주민의 양성률이 높고 또한 평야나 산간지역에 비해 해안지역 주민이 월등히 높다. 사회경제적 상태별로 볼 때도 상류층에 비해 하류층에 양성률이 높고, 직업별로는 혈액 및 혈액재제에 폭로기회가 많은 의료인의 양성률이 높다. 구체적으로 1975년 조사결과에 의하면 일반인이 7.2%, 공혈자 8.7%, 의사 9.9%, 간호사 10.2%, 검사실기사 10.8%, 의과 및 간호학과 학생이 7.3%였다. 전파양식은 혈액, 타액, 정액, 질분비물 등이 감염성이 있고 이러한 감염성체액이 경제적으로 혹은 경점막 폭로로 되는데 오염된 주사기와 바늘 의료기기등은 주요한 매개 역할을 한다. B형간염의 감염은 상처나 열, 오염 혹은 감염된 혈액에 점막이 폭로될때 확산률이 높다.

이에 본 논문에서는 B형 간염의 조기진단을 통한 예방이 중요하다고 사료되어 B형간염의 진단방법인 간기능 효소(GPT, GOT)검사 및 총단백, 암부민, AG비, 콜레스테롤, 요소질소 등의 혈액성분조사 자료로 B형간염진단을 위한 나무구조 진단모형을 개발하여 기존의 간기능효소(GOT, GPT)만에 의한 방법과 진단의 효율성을 비교 검토하고자 한다.

2. 나무구조에 의한 진단모형

2.1. 나무구조모형

나무구조 진단모형은 나무형태로 나타내진 의사결정 즉 분류규칙을 모형화 한것으로 실질적인 진단이나 분류를 쉽게 적용할수 있는 분류기법이다. 이분적인 단순 형태외에도 나무구조 분류방법은 실용적이고 다른 모수적인방법 즉 로지스틱이나 판별분석등에 대해서도 비슷한 진단력을 나타내주고 있다. 즉 특이도와 민감등으로 산출된

모형의 정확도가 같은수의 변수를 모형에 참여 시켰을때 모수적 방법보다 더 향상되었다.

나무진단 모형의 적합은 이미 알고 있는 진단군과 관련된 변수자료를 정보로 하여 나무분류모형을 작성하고 이에 의하여 새로운 자료 즉 환자가 왔을때 이 나무진단 모형에 의해 짧은 시간내에 최소의 변수만을 조사하여 진단을 즉 간염 감염여부를 진단하고자 하는 것이다. 나무구조에 의한 분류모형도 다음과 같은 응용분야를 갖는다.

- 1) 농수산분야에서 고해상도 사진촬영에 의한 작황량 예측
- 2) 의학분야 환자의 병력과 검사자료에 의한 계량적 진단
- 3) 범죄 연구에서 목소리 음파에 의한 범인 판별
- 4) 레이더 자료에 의한 잠수함에서의 적군, 아군 판별작업
- 5) 초음파에 의한 화학물질 성분분류
- 6) 최근 10년간의 기상자료에 의한 내일의 기상 예측

2.2. CART기법에 의한 분류

CART는 Classification and Regression Tree방법으로 이분류 나무구조로 작성된 진단 분류 모형을 단계적으로 비모수적 방법으로 적합시킨다. CART에 의한 진단모형은 모수적 방법과 비교하여 정확하고 이해및 설명이 쉬우며 통계 이론적으로도 정확한 방법으로 발표된 바있다. 간염진단과 같은 의학분야의 적용에 CART는 매우 유용하여 조사된 변수들간의 관련성을 밝혀주고 중요도를 지수로 나타내주는 강력한 방법이다.

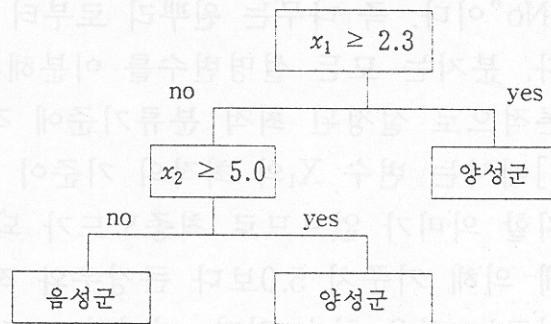
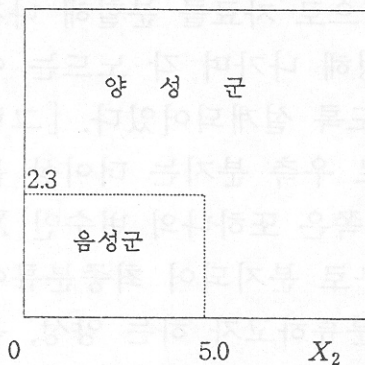
CART분류기법은 조사관찰된 각종 임상자료의 변수간 관련성의 유형을 예측하고 밝히는데 효율적인 비모수적 방법으로 나무형태 분류모형은 단순히 [그림 1]과 같이 각 변수 마다 일정기준을 만족하는가 아닌가에 의해 2분류를 해나감으로서 작성된다. 즉 반응변수 (Y =양성 또는 음성)의 결과를 예를들어 두개의 설명변수 X_1 과 X_2

에 의해 나타내게 된다. 비교를 위해 로지스틱 회귀모형의 분류개념과 그림으로 나타내서 변수 X_1 과 X_2 에 의해 반응변수가 어떻게 전체 자료를 분류해 나가는가 나타내 보았다.

나무구조에 의한 분류규칙

변수별 분류기준 : $X_1 \geq 2.3$ or $X_2 \geq 5.0$

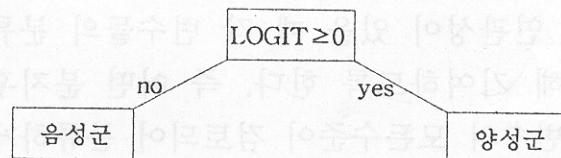
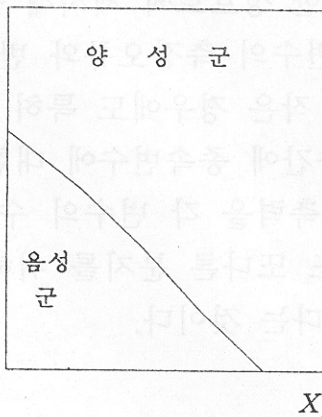
X_1 분류 분할공간



Logistic Regression Rule

classify as high risk if $\text{LOGIT} \doteq -1.2 + 0.52 \cdot X_1 + 1.34 \cdot X_2 \geq C$

X_1 분류 분할공간



$P_c = \exp(c)/(1+\exp(c))$ = "critical probability"

Example: $c=0$ implies $P_c = 0.5$

[그림 1] 나무분류모형과 로지스틱모형의 기하학적 비교

두방법모두 두변수 X_1 과 X_2 에 의해 전체공간을 반응변수의 양성과 음성의 두 부분공간으로 나누어 나감을 알수 있는데 차이점은 로지스틱모형이 나무구조에 비해 X_1 과 X_2 의 교차항을 포함 회귀식의 형태임을 알수있고 분류식을 함수식으로 나타낼수 있는데 비해 나무구조 분류모형은 단순한 질문 즉 " $X_1 > 2.3$ 인가", " $X_2 > 5.0$ "을 만족하는가등의 조건에 의해 분할되어 나감을 알 수 있다. 이 예에서 분할규칙은 이항 나무구조이고 모든 질문에서의 결과는 "Yes" 아니면 "No"이다. 즉 나무는 왼쪽부터 계속해서 자료를 분할해 나간다. 분지는 모든 설명변수를 이분해서 결정해 나가며 각 노드는 이론적으로 설정된 최적 분류기준에 적합하도록 설계되어있다. [그림 1]에서는 변수 X_1 의 최적의 기준이 2.3으로 우측 분지는 더이상 분지할 의미가 없으므로 최종노드가 되고, 왼쪽은 또하나의 변수인 X_2 에 의해 기준치 5.0보다 큰경우와 작은경우로 분지되어 최종분류에 이르는 것을 알수 있다. 마지막 노드들은 분류하고자 하는 양성, 음성의 두결과중 하나가 되는 분류결과임을 알수 있다. 이러한 결과는 또한 예측치에 대한 정보와 변수간의 분류에 기여하는 중요도등을 포함한다.

CART 알고리즘내에 자동적인 단계적 적합과 사후점검에 의한 모형의 정확 산출 분류모형의 오분류율을 또한 정보로써 제시해 준다. CART에 의한 분류모형은 이상치와 관측변수의 측정오류와 변수간에 상호 관련성이 있거나 예측에 기여도가 작은 경우에도 특히 로버스트함을 이론적으로 밝힌바 있다. 즉 변수간에 종속변수에 대한 상호 연관성이 있을 때 각 변수들의 분류 예측력을 각 변수의 수준에 의해 기여하도록 한다. 즉 어떤 분지후에도 또다른 분지를 위해 모든변수의 모든수준이 검토되어 분류하게 된다는 것이다.

3. 분석자료 및 결과

3.1. 분석자료

본 논문에서 분석된 자료는 대한 적십자사 대구 경북지사 혈액원의 1995년도 헌혈자의 혈액검사중 GPT이상치(45H)를 추출하여 얻은 17863명중 1985명을 계통추출하여 분석했다. 조사항목은 <표 1>에서와 같이 B형간염, C형간염의 감염여부와 간염진단에 주요 조사항목으로 대한적십자사에서 실시하고 있는 혈청조사로 S-GPT, S-GOT의 수치와, 총단백, 알부민, A/G비, 콜레스테롤양, 요소질소, AIDS항체검사, 매독항체 검사등을 조사한 결과이다. 조사대상자의 성별 B형간염과 C형간염의 감염분포는 <표 2>, <표 3>과 같다.

<표 1> 헌혈자의 검사항목

검 사 항 목	설	명
B형 간염항원 (HBsAg)	음성(-) 양성(+)	B형간염 병원체(바이러스)가 인체내에서 발견되지 않음. B형간염을 일으킬 수 있는 병원체가 몸속에 있어 발병가능성이 있음.
C형 간염항체 (Anti HCV)	음성(-) 양성(+)	C형간염 병원체(바이러스)에 감염된 흔적이 없다는 의미 C형간염 병원체에 감염된 흔적이 있어 만성간염으로의 발전 가능성 있음.
혈 청 S-GPT/S-GOI	정상치	GOT는 간과 심근에, GPT는 간에 많이 있는 효소로서 간 질환이 있으면 혈청내에서 증가됨.
총단백 알부민 A G 비		건강한 사람의 혈청단백은 거의 일정한 분량이나 짐칸에 따라서는 감소함. 또 혈청단백은 크게 알부민(A)과 글로블린(G)으로 나누어짐. 건강인은 일정수준이 유지되나 질병이 있을 경우 변화가 생긴다.
콜레스테롤 (Cholesterol)		혈액속의 지방질 성분중의 하나로 지방이 많은 식사를 오래 계속하거나 간, 신장, 췌장, 담도등의 질환으로 콜레스테롤이 증가되고 혈관질환등이 생길 수 있음.
요 소 질 소 (BUN)		체내 단백질은 영양소로 이용된뒤 요소로 분해되어 신장을 통해 방뇨된다. 따라서 신장기능이 나빠지면 요소배설이 잘안되어 혈액속의 그 양이 증가된다. 또 체내에서는 단백질이 많이 파괴될 때에도 증가합니다.

〈표 2〉 B형간염의 성별 감염분포

B형 간염 성 별	양 성	음 성	계
남	5,520 (30.78)	10,994 (61.30)	16,514 (100.00)
여	943 (66.41)	477 (33.59)	1,420 (100.00)
계	6,463 (36.04)	11,471 (63.96)	17,934 (100.00)

〈표 3〉 C형간염의 성별 감염분포

C형 간염 성 별	양 성	음 성	계
남	311 (3.38)	16,203 (96.62)	16,514 (100.00)
여	48 (3.38)	1,372 (96.62)	1,420 (100.00)
계	359 (2.00)	17,575 (98.00)	17,934 (100.00)

각종 진단을 위한 검사치중 진단에 유의하고 주로 이용되는 SGOT의 결과를 간염유형과 성별로 기본통계량을 구한 결과는 다음 〈표 4〉와 같다.

〈표 4〉 혈청효소(SGOT)의 질병 유형별 성별 기본 통계량

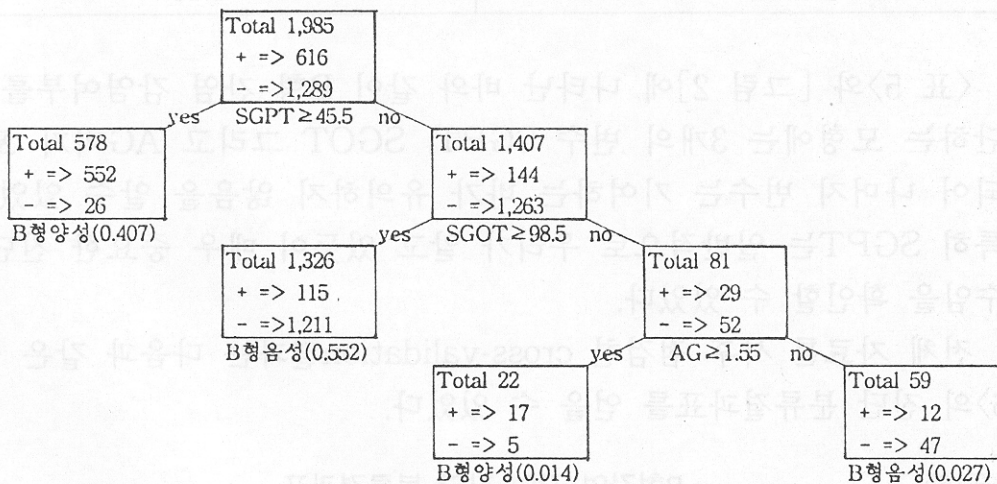
질병유형	성 별	감염여부	N	평 균	표준편차	P-value
B형간염	남	양 성	5,520	33.65	37.97	0.0000
		음 성	10,994	52.61	82.49	
	여	양 성	943	26.94	48.26	0.0000
		음 성	477	45.84	46.51	
C형간염	남	양 성	311	44.49	75.52	0.6566
		음 성	16,203	46.31	71.28	
	여	양 성	48	22.63	9.02	0.0001
		음 성	1,372	33.66	49.26	

3.2. CART에 의한 나무구조 진단모형 결과

조사된 변수인 연령, 성별, 직업, 혈액효소 SGOT, SGPT, 총단백, 알부민, AG비, 콜레스테롤, 요소질소 등의 변수를 넣어 총 조사자 17,863명중 1,985명을 계통추출하여 CART에 의해 수행시킨 결과 총진단력이 91.7%이고 3개의 변수만이 참여된 나무구조 진단모형을 적합시킬수 있었다.

(1) B형 간염 진단모형

B형 간염의 감염진단을 위한 적합된 나무구조 분류모형은 다음의 [그림 2]와 같다.



[그림 2] B형 간염의 감염진단을 위한 나무구조 분류모형

관여된 변수의 각기준치는 실제 쓰이고 있는 SGPT기준치 45가 유의함을 확인할 수 있었고, 그외 SGOT의 98.5, AG비의 1.55가 유의한 기준치로 제시됨은 매우 유용한 결과였다. 이 모형의 민감도는 0.818, 특이도는 0.976이고 가양성률은 0.052, 가음성률은 0.092로

매우 진단력이 높은 바람직한 진단모형임을 알수있었다. B형간염 진단에 참여한 모든 변수의 기여도를 100을 기준으로 한 상대적인 지수로 나타낸 결과는 다음 <표 5>와 같다.

<표 5> B형간염 감염 진단모형의 변수 중요도

변 수	중요도 지수
1. S G P T	100.00
2. S G O T	46.61
3. 알 부 민	5.48
4. 총콜레스테롤	2.44
5. A G	1.99
6. B U N	1.18
7. 총 단 백	1.11

<표 5>와 [그림 2]에 나타난 바와 같이 B형 간염 감염여부를 판단하는 모형에는 3개의 변수 SGPT, SGOT 그리고 AG비가 포함되어 나머지 변수는 기여하는 바가 유의하지 않음을 알수 있었다. 특히 SGPT는 일반적으로 우리가 알고 있듯이 매우 중요한 진단변수임을 확인할 수 있었다.

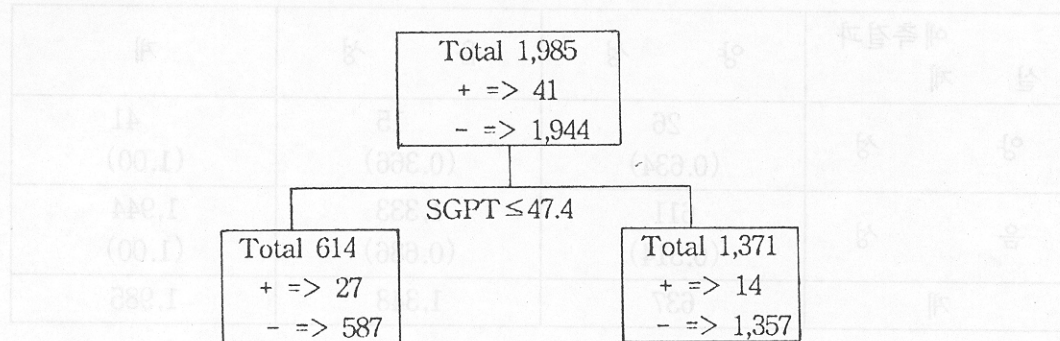
전체 자료를 사후 점검한 cross-validation결과는 다음과 같은 <표 6>의 진단 분류결과표를 얻을 수 있었다.

<표 6> B형간염 감염 진단 분류결과표

예측결과 실 제	양 성	음 성	계
양 성	569 (0.818)	127 (0.182)	696 (1.00)
음 성	37 (0.029)	1,252 (0.971)	1,289 (1.00)
계	6.06	1,379	1,985

(2) C형 간염 진단모형

B형 간염과 같이 조사전 변수로 C형 간염 감염 여부진단을 위한 나무구조 모형을 적합하여 다음 [그림3]과 같이 하나의 변수SGPT에 의한 나무구조 진단모형을 얻었다.



[그림 3] C형 간염 감염진단을 위한 나무구조 분류모형

[그림 3]의 C형 간염 나무구조 진단모형의 민감도는 0.634 특이도는 0.686으로 별로 유효한 모형이 아니었다. C형 간염의 정확진단을 위해서는 다른 변수들이 개발되어야 함을 제시하는 좋은 근거였다. 변수의 중요도 지수는 다음 <표 7>과 같았으나 CART에 의한 자동 단계별 적합에 의해 SGPT를 제외한 변수들은 C형간염 진단에 유의하게 기여하지 못함을 나타내주고 있다.

<표 7>

C형 간염 감염 진단모형 변수 중요도

변 수	중요도 지수
1. S G P T	100.00
2. S G O T	9.02
3. 알 부 민	0.00
4. 총콜레스테롤	16.72
5. A G	2.84
6. B U N	14.01
7. 총 단 백	0.00

적합된 모형에 의해 원자료를 cross-validation하여 다음의 <표 8>과 같은 진단분류결과표를 얻었다.

<표 8> C형 간염 감염 진단 분류 결과표

예측결과 실 제	양 성	음 성	계
양 성	26 (0.634)	15 (0.366)	41 (1.00)
음 성	611 (0.314)	1,333 (0.686)	1,944 (1.00)
계	637	1,348	1,985

4. 결 론

현대인의 건강을 위협하는 질환중 매우 중요하게 연구되고 있는 간염의 감염진단 방법으로 쓰이고 있는 각종 혈액성분의 항목을 대구혈액원의 95년 헌혈자 자료를 중심으로 CART에 의한 나무구조 간염진단 모형을 검토한 결과 다음의 결론을 얻을 수 있었다.

1) SGOT 45이상으로 제한한 조사대상자의 %가 B형간염에 %가 C형간염에 감염되어 헌혈에 의한 간염감염 연구에 중요한 자료가 되었다.

2) B형간염 감염진단을 위한 나무구조 모형을 적합시켜 3개의 조사항목 변수 즉 SGPT(기준치 1.55), SGOT(기준치 98.5), AG비(기준치 1.55)에 의한 91.7%의 진단력, 민감도 81.8%, 특이도 97.6%의 진단모형을 얻었다.

3) C형 간염 감염진단 모형으로 하나의 변수 SGPT(기준치 47.4)에 의한 나무구조 진단모형으로 68.5%진단력을 갖는 모형을 얻었다.

본 논문의 분석결과로 지금까지 사용되어온 간염 감염 진단변수의

기준치의 타당성과 일반성을 확인 할 수 있었고, 하나의 변수만에 의한 진단보다는 여러변수의 복합적인 모형에 의한 진단결과가 진단력을 오류율을 높여줄수 있고 이에따라 간염의 조기발견및 조기치료 또한 헌혈자에 의한 간염 감염여부의 정확한 진단과 발견으로 헌혈자를 통한 간염전염을 예방하여 국민건강증진에 기여함을 본 연구의 의의로 한다.

〈참고문헌〉

1. Breiman et al. (1984) Classification and Regression Trees, Wadsworth CA : Belmont.
2. W. H. Wolberg et. all (1986) Statistical Approach to Fine Needle Aspiration Diagnosis of Breast Masses, Acta Cytologica, Vol.31, pp.737~741.
3. W. Yi Loh, L Vanichsctakal (1988) Tree Structured Classification via Generalized Discriminant Analysis, JASA, Vol.83, No.403.
4. I. G. Vlachonikolis (1990) Predictive discrimination and Classification with mixed binary and continuous variables, Biometrika, Vol. 77, No.3, pp.657~662.
5. T. R. Lee (1991) Differential Diagnosis of Colitis Using Kernel Discriminant, Journal of Korean Air & Correspondence University, Vol. 13, pp.413~431.
6. D. J. Hand (1981) Discrimination and Classification, John Wiley & Sons.
7. M. H. Huh et al. (1989) Discriminant analysis on Bone Marrow Metastasis of Cancer using CART, The Applied Statistics, Vol. 4, No. 1.
8. B. S. Kim et al. (1987) Discriminatory Variable Selection

on Bone Marrow Metastasis of Cancer Using Logistic Regression, The Applied Statistics, Vol. 1, No. 2, pp.45~60.

9. T. R. Lee (1993) Tree structured Modeling Via Kernel Discriminant Modeling in Discrete Multivariate Data.

<참고문헌>

1. Breiman et al. (1984) Classification and Regression Trees, Wadsworth CA: Belmont.
2. W. H. Wolberg et al. (1986) Statistical Approach to Fine Needle Aspiration Diagnosis of Breast Masses, Acta Cytologica, Vol.31, pp.737~741.
3. W. Yi Loh, I. Vanichastakul (1988) Tree Structured Classification via Generalized Discriminant Analysis, JASA, Vol.83, No.403.
4. I. G. Vischnionikolis (1990) Predictive discrimination and Classification with mixed binary and continuous variables, Biometrika, Vol. 77, No.3, pp.657~662.
5. T. R. Lee (1991) Differential Diagnosis of Colitis Using Kernel Discriminant, Journal of Korean Air & Correspondence University, Vol. 13, pp.413~431.
6. D. J. Hand (1981) Discrimination and Classification, John Wiley & Sons.
7. M. H. Huh et al. (1989) Discriminant analysis on Bone Marrow Metastasis of Cancer using CART, The Applied Statistics, Vol. 4, No. 1.
8. B. S. Kim et al. (1987) Discriminatory Variable Selection

원그리기 자료에 대한 통계분석

김 범 준

(지도교수 : 이광진)

목원대학교 응용통계학과

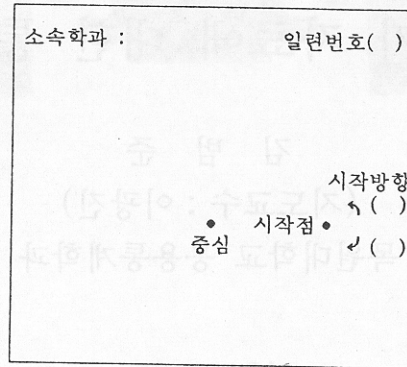
〈요 약〉

본 연구에서는 목원대학교의 영어영문학과와 응용통계학과의 두 학과에서 각각 30명씩의 학생들을 임의로 추출하여 반지름이 5cm인 원을 그리게 하고, 이들로부터 얻어진 60장의 원그림들을 자료로 하여 두 집단간에 원을 그리는 능력의 차이가 있는지의 여부와 원그리기에 어떤 패턴등이 있는지를 통계적으로 분석해 보았다.

1. 본 연구를 시작하며.

본 연구는 『직접 손으로 원을 자주 그려보는 집단과 그렇지 못한 집단간에는 원을 그리는 정밀도에 있어 차이가 있을 것이라는 일반적인 상식을 통계적으로 검정해 볼 수는 없을까?』라는 통계학도로서의 막연한 호기심 또는 의기로부터 출발하였다.

이를 위해 우선 원을 자주 그려보는 집단이라고 여겨질 수 있는 응용통계학과와 그렇지 못한 집단으로 여겨질 수 있는 영어영문학과의 두 학과를 대상으로, 각각 30명씩의 학생들을 임의표본으로 선정하였다. 물론 영어영문학과와 응용통계학과는 전공의 속성상 문과적인 두뇌를 가진 집단과 이과적인 두뇌를 가진 집단으로 볼 수도 있다. 선정된 학생들에게 중심과 시작점(실제로는 반지름이 5cm가 되도록 시작점을 잡았음)만이 주어진 아래 [그림 1]과 같은 시험지에 한붓그리기 방법으로 반지름이 50mm인 표준원을 그리게 하고, 그려나가는 방향도 기록하게 하였다.



[그림 1] 원그리기 시험지

통계적 분석을 위해 각 얻어진 원그림들은 다음과 같은 과정을 거쳐 숫자화되어 입력되었다.

반지름이 1mm, 각도가 3도 간격으로 눈금이 그려진 방사형 모눈 필름을 각 원그림들 위에 겹쳐두고, 시작방향으로부터 3도간격으로 원점으로부터의 거리들을 모두 재었다. 만일 원그림이 눈금과 눈금 사이를 지나가는 경우에는 0.5mm를 기준으로 반올림하였다. 이는, 예를 들어, 눈대중으로 거리가 47.4mm정도이면 47.5mm로, 53.8mm 정도이면 54.0mm로 기록하였다는 것이다.

참고로 본문에서 사용되는 편차, 절대편차, 표준원이라는 용어들의 정의는 다음과 같다.

편 차 = 원점으로부터의 거리 - 50

절대편차 = 편차의 절대값

표 준 원 = 반지름이 50mm인 원

여기에서, 편차에서 50을 뺀 이유는 학생들에게 그리게 한 표준원의 반지름이 50mm이기 때문이다. 그러나 실제분석에 사용된 자료는 각 각도별 학생들의 편차들로 구성된 크기가 120×60인 행렬자료로서 이는 부록에 주어져 있다(부록에 제시된 자료는 편집의 편의상 단위가 0.1mm임). 자료에서, 예를 들어, S12는 응용통계학과의 12번째 학생, E8은 영어영문학과 8번째 학생을 나타낸다.

분석에 앞서 우리는 상식에 바탕을 둔 다음과 같은 견해를 가지고

있었다. 영어영문학과 학생들과 응용통계학과 학생들간에는 원 그리기의 능력에 있어서 차이가 분명히 있을 것이고, 차이가 존재한다면 이는 당연히 두 집단간의 속성의 차이로부터 기인된 것임이 분명하다. 한편 원그리기의 능력이란 원그리기 경험과 아울러 어떤 성격적 또는 기질적인 요인에 크게 영향을 받는 것이기 때문에, 원을 잘 그리는 학생들만으로 두 학과를 비교하면 별로 차이를 보이지 않을 것이지만, 원을 잘 못그리는 학생들만으로 두 학과를 비교하면 학과간의 차이가 분명히 존재할 것이다.

2. 잘 그린 원그림이란?

원의 수학적 정의는 다음과 같다.

‘원이란 주어진 한 정점으로부터의 거리가 일정한 점들의 집합이다.’

엄밀한 수학적 정의에 입각하면 ‘원이면 원이지 원과 닮은 원’이란 있을 수 없다. 하지만 우리의 일상에서는 ‘원과 비슷하다’라는 개념은 얼마든지 있을 수 있다. 따라서 원과 비슷하다는 개념(즉 원과의 닮음의 정도)을 어떻게 계측화할 수 있느냐에 따라 잘 그린 원그림을 판단할 수 있을 것이다. 본 연구에서는 ‘반지름을 얼마나 잘 유지하고 있느냐’와 ‘얼마나 매끄럽게 곡선을 그리느냐’라는 두가지를 잘 그린 원그림의 판단기준으로 삼고, 이를 계측할 수 있는 측도로서 ‘편차의 표준편차’와 ‘절대편차의 평균’을 사용하기로 정하였다.

이들을 사용한 이유는 다음과 같다. 우선 ‘편차의 표준편차가 0’이라는 것과 ‘완전한 원’이라는 것은 필요충분조건의 관계이고, 편차의 표준편차가 클수록 반지름의 유지능력이 약함을 의미하기 때문에 편차의 표준편차가 반지름 유지능력의 한 측도로 사용된 것이다. 마찬가지로 ‘절대편차의 평균이 0’이라는 것과 ‘완전한 표준원’이라는 것도 필요충분조건의 관계이고, 절대편차의 평균이 커질수록 매끄러움의 측면에서 표준원으로 부터의 이탈이 심함을 의미하기 때문에 절대편차의 평균도 잘 그린 원그림을 판단할 수 있는 한 측도로 사용되

어질 수 있는 것이다. 참고로, 이들 두가지의 판단기준 또는 두가지의 측도들은 상호 독립적이 아니라 서로 보완적인 관계임을 알아야 할 것이다.

물론 본 연구에서 제시한 잘 그린 원그림의 측도들 2개 이외에도 얼마든지 좋은 측도들이 존재할 수 있을 것이라는 점을 부정하지는 않는다. 그러나 일반인들도 즉각 생각해 낼 수 있는 측도인 ‘표준원과의 편차의 제곱합’은 본 연구에서 사용되어지는 절대편차의 평균이라는 측도와 일치하는 개념이라는 점도 밝힌다.

3. 어떤 학생들이 원을 잘 그렸는가?

각 학생별 원그리기 능력을 판단하기 위하여 우선 편차와 절대편차의 각각에 대해 평균과 표준편차들을 구하여 보았다. 이는 다음의 <표 1>에 제시되어 있다.

<표 1> 편차의 평균과 표준편차, 절대편차의 평균과 표준편차

기 체 번호	소 속 학 과	편차의 평균	편차의 표준편차	절대편차의 평균	절대편차의 표준편차	기 체 번호	소 속 학 과	편차의 평균	편차의 표준편차	절대편차의 평균	절대편차의 표준편차
1	물리	-2.61	2.851	2.99	2.442	31	영문	2.80	5.418	5.40	2.796
2	물리	-2.85	3.736	3.21	3.423	32	영문	-1.71	7.364	6.01	4.552
3	물리	3.03	5.791	5.74	3.094	33	영문	-7.11	6.965	7.95	5.977
4	물리	-8.15	5.354	8.47	4.818	34	영문	-5.63	9.520	8.50	7.045
5	물리	-4.39	5.175	5.54	3.905	35	영문	-2.85	9.346	8.52	4.728
6	물리	-7.00	3.379	7.05	3.273	36	영문	3.68	3.393	3.94	3.086
7	물리	6.05	3.020	6.05	3.020	37	영문	-4.34	7.864	8.05	3.933
8	물리	-5.37	2.660	5.37	2.660	38	영문	-0.46	2.893	2.19	1.930
9	물리	-1.84	2.267	2.12	2.001	39	영문	-5.55	7.888	7.72	5.754
10	물리	-3.11	1.701	3.11	1.701	40	영문	-2.75	1.984	2.81	1.899
11	물리	2.62	2.856	3.22	2.161	41	영문	-0.03	3.989	3.33	2.170
12	물리	14.93	5.627	14.96	5.537	42	영문	-0.85	7.198	6.35	3.434
13	물리	4.83	8.459	7.55	6.124	43	영문	-2.17	2.415	2.61	1.923
14	물리	-7.32	5.019	7.32	5.019	44	영문	-4.23	7.983	7.36	5.204
15	물리	-2.54	1.476	2.54	1.476	45	영문	-2.85	6.753	6.28	3.739
16	물리	0.64	1.982	1.46	1.481	46	영문	-1.79	3.361	3.32	1.852
17	물리	-1.01	2.747	2.33	1.761	47	영문	1.59	5.979	4.75	3.941
18	물리	-10.84	5.854	10.84	5.854	48	영문	-7.58	4.294	7.58	4.294
19	물리	0.88	4.270	3.65	2.351	49	영문	2.15	3.097	3.09	2.149
20	물리	-1.82	1.966	2.16	1.578	50	영문	-0.21	3.117	2.52	1.832
21	물리	0.06	5.497	5.02	2.172	51	영문	-1.91	4.452	3.88	2.876
22	물리	-0.55	3.051	2.36	1.997	52	영문	-5.75	7.782	7.80	5.702
23	물리	-5.17	3.263	5.18	3.237	53	영문	-0.51	2.648	2.22	1.514
24	물리	-7.22	4.694	7.43	4.354	54	영문	-4.05	2.997	4.07	2.963
25	물리	1.26	3.275	2.99	1.812	55	영문	-10.20	8.242	10.66	7.623
26	물리	4.64	2.514	4.64	2.514	56	영문	-0.27	1.858	1.47	1.155
27	물리	-2.71	5.618	4.90	3.836	57	영문	5.09	6.768	6.21	5.742
28	물리	-3.40	3.141	3.45	3.095	58	영문	12.90	4.896	12.90	4.896
29	물리	-1.84	3.414	3.09	2.330	59	영문	-1.82	3.308	3.12	2.121
30	물리	-4.88	7.808	7.48	5.342	60	영문	-0.43	7.698	6.68	3.789

〈표 1〉에서 특징적인 것들을 설명하면 다음과 같다.

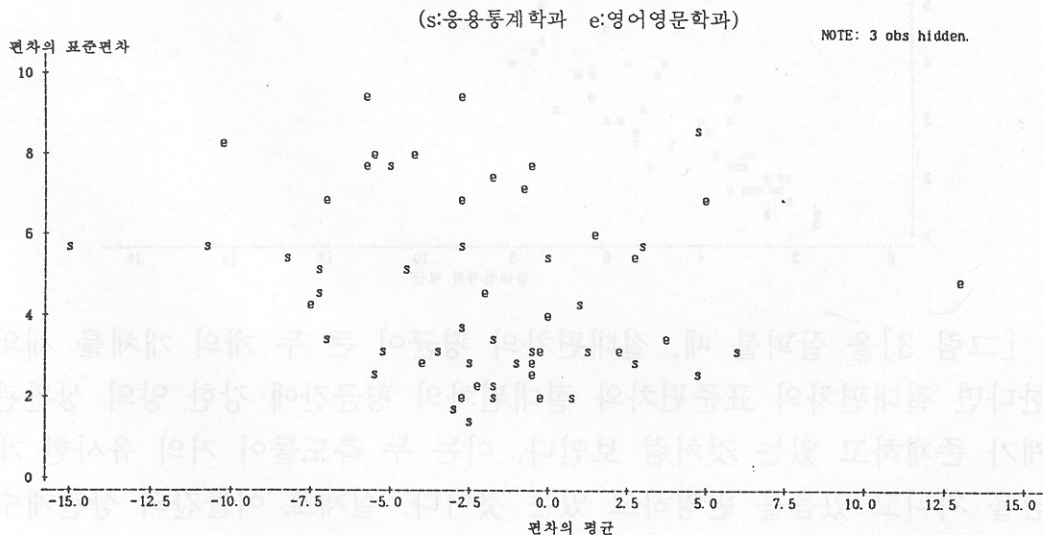
첫째, 12번의 응용통계학과 학생과 58번의 영어영문학과 학생은 편차의 평균과 절대편차의 평균의 입장에서 다른 학생들에 비해 큰 차이를 보이고 있는 것으로 보아 특이치로 간주될 수 있겠다. 따라서 이들 두 측도를 사용한 학과간의 비교에서는 이들 학생들을 제외시켜야 좋을 것이다.

둘째, 21, 35, 42번의 학생들은 편차의 평균의 절대값과 절대편차의 평균의 차이가 다른 학생들에 비해 월등한 차이를 보이고 있다. 이는 표준원을 기준으로 여러번 들숙날숙하게 원을 그렸음을 의미한다고 하겠다.

셋째, 편차의 평균이 0보다 작을수록 표준원보다는 작게, 0보다 클수록 표준원보다는 상대적으로 큰 원그림을 그린 것으로 볼 수 있기 때문에, 12, 18, 55번의 학생들은 작은 원그림을, 58번 학생은 큰 원그림을 그린 것으로 나타났다.

〈표 1〉을 바탕으로 편차의 평균과 편차의 표준편차를 이용한 개체의 산점도([그림 2]), 절대편차의 평균과 절대편차의 표준편차를 이

[그림 2] 편차의 평균과 편차의 표준편차를 이용한 개체의 산점도

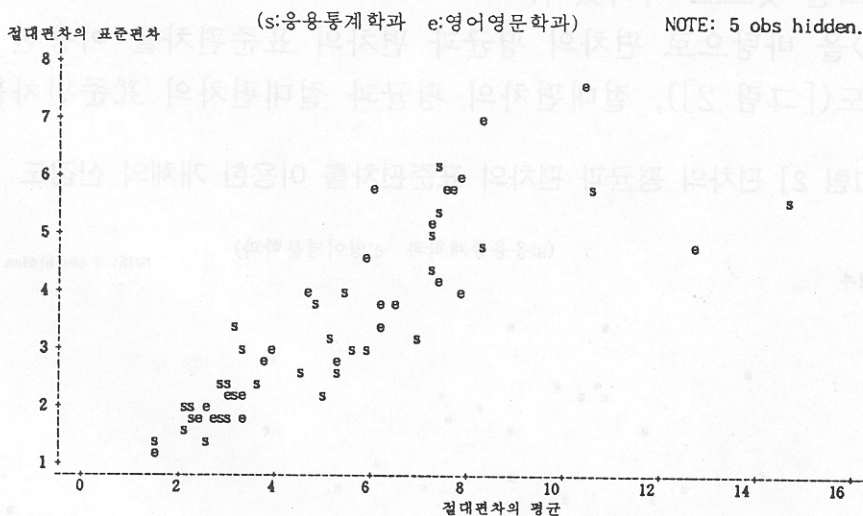


용한 개체의 산점도([그림 3]), 그리고 절대편차의 평균과 편차의 표준편차를 이용한 개체의 산점도([그림 4])를 그려보았다.

[그림 2]의 산점도에서 점들의 흩어진 양상으로 보아, 편차의 표준편차와 편차의 평균간에는 상관관계가 거의 없는 것처럼 보인다. 실제로 이들간의 상관계수는 -0.207 이다.

편차의 평균이 0에 가까울수록 크기적인 면에서는 표준원에 가깝게 그렸다는 것을 의미할 수는 있지만, 편차의 평균이 0이라고 해서 표준원을 의미하지는 못하기 때문에 이는 잘 그린 원그림의 판단기준으로 사용하기에는 부적절하다. 그러나 편차의 표준편차가 0에 가까울수록 매끄럽게 그렸다는 것을 의미하기 때문에 산점도에서 하단에 위치한 학생들이 일단 다른 학생들에 비해 원그림을 잘 그렸다고 할 수 있겠다.

[그림 3] 절대편차의 평균과 절대편차의 표준편차를 이용한 개체의 산점도



[그림 3]을 살펴볼 때, 절대편차의 평균이 큰 두 개의 개체를 제외한다면 절대편차의 표준편차와 절대편차의 평균간에 강한 양의 상관관계가 존재하고 있는 것처럼 보인다. 이는 두 측도들이 거의 유사한 개념을 지니고 있음을 반영하고 있는 것이다. 실제로 이들간의 상관계수는 0.842 로서, 특이치 두 개의 개체를 제외하면 더 큰 값을 가진다.

절대편차의 평균이 작다는 것은 표준원에 가까우면서 매끄럽게 그렸다는 것을 의미하지만, 절대편차의 표준편차가 0이라고 해서 매끄럽게 그렸다는 것을 의미하지는 않기 때문에 이는 매끄러움의 측도로서는 사용되어질 수 없는 것이다.

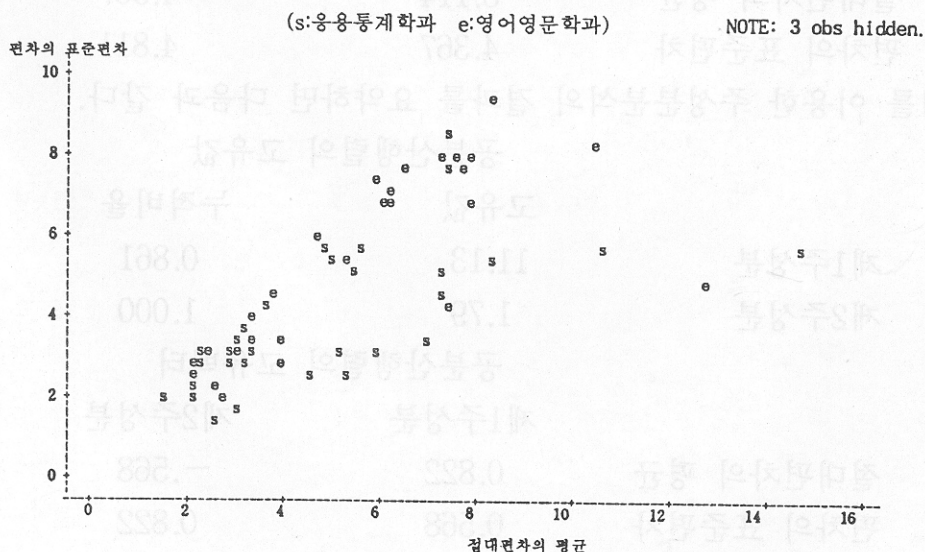
[그림 3]에서는 학과간의 어떤 구별현상이 나타나고 있지는 않은 도 볼 수 있다.

[그림 4]는 [그림 2]와 [그림 3]의 해석을 통해 잘 그린 원그림의 판단측도로 확정된 절대편차의 평균과 편차의 표준편차를 이용하여 각 학생들의 산포형태를 나타낸 그림으로 여기에서 다음과 같은 몇가지 결과를 도출해 낼 수 있다.

편차의 표준편차와 절대편차의 평균간에도 어느 정도의 큰 양의 상관관계가 있는 것처럼 보인다. 실제로 이들간의 상관계수는 0.670이었다. 물론 절대편차의 평균이 큰 두 명의 학생을 제외시킨다면 더 큰 상관계수를 가지게 된다.

편차의 표준편차값이 6이상인 학생들의 대부분이 영어영문학과 학생들인 것이 특이한 점이다. 그리고 이들은 절대편차의 평균이 일정 수준 이상 이라는 것도 특이하다.

[그림 4] 절대편차의 평균과 편차의 표준편차를 이용한 개체의 산점도



[그림 4]에서의 좌측 하단에 위치하는 그런대로 원그림을 잘 그린 집단(즉 절대편차의 평균의 값도 작고 편차의 표준편차의 값도 작은 집단)에는 응용통계학과 학생들과 영어영문학과 학생들이 고루 섞여 있지만, 그렇지 못한 집단에서는 2명의 특이치들을 제외한다면 편차의 표준편차의 입장에서 응용통계학과 학생들의 집단과 영어영문학과 학생들의 집단으로 확연히 구분되는 현상이 나타난다. 이는, 1절의 말미에서 언급된, 연구초기에 가졌던 우리의 감을 확인시켜주는 것이라고 할 수 있다.

4. 어느 학과의 학생들이 원을 더 잘 그렸는가?

이제 두 학과간의 비교문제를 다루고자 한다. 이를 위해서는 전체적인 입장에서 잘 그린 원그림에 대한 단일측도가 필요하다. 본 연구에서는 선정된 두 측도의 공분산행렬을 이용하여 얻어진 제1주성분을 사용하기로 하였다.

우선 절대편차의 평균과 편차의 표준편차간의 공분산행렬은 다음과 같다.

공분산행렬	절대편차의 평균	편차의 표준편차
절대편차의 평균	8.114	4.367
편차의 표준편차	4.367	4.811

이를 이용한 주성분분석의 결과를 요약하면 다음과 같다.

공분산행렬의 고유값		
	고유값	누적비율
제1주성분	11.13	0.861
제2주성분	1.79	1.000
공분산행렬의 고유벡터		
	제1주성분	제2주성분
절대편차의 평균	0.822	-.568
편차의 표준편차	0.568	0.822

이 결과로부터 우리는 다음과 같은 사실을 얻을 수 있다. 제1주성분의 설명력의 값이 86.1%로서, 제1주성분만으로 [그림 4]의 산표 형태를 약 86%정도 설명할 수 있다. 따라서 제1주성분이 잘 그린 원그림에 대한 단일측도로서 충분히 사용될 수 있는 근거를 얻었다고 하겠다.

제1주성분점수는 아래의 식을 이용하여 구할 수 있고, 이를 이용하여 얻어진 주성분점수들을 구하여 정리하면 아래 <표 2>와 같다.

$$\text{제1주성분점수} = 0.822 \times (\text{절대편차의 평균}) + 0.568 \times (\text{편차의 표준편차})$$

<표 2> 각 개체에 대한 제1주성분점수와 제1주성분점수의 순위

개체 번호	소속 학과	편차의 표준편차	절대편차 의 평균	제1주성분 점수	제1주성분 점수의 순위	개체 번호	소속 학과	편차의 표준편차	절대편차 의 평균	제1주성분 점수	제1주성분 점수의 순위
1	s	2.851	2.99	-2.95	14	31	e	5.418	5.40	0.49	36
2	s	3.736	3.21	-2.26	22	32	e	7.364	6.01	2.09	44
3	s	5.791	5.74	0.97	38	33	e	6.965	7.95	3.46	48
4	s	5.354	8.47	2.98	47	34	e	9.520	8.50	5.37	57
5	s	5.175	5.54	0.46	35	35	e	9.346	8.52	5.28	56
6	s	3.379	7.05	0.69	37	36	e	3.393	3.94	-1.86	25
7	s	3.020	6.05	-0.33	31	37	e	7.864	8.05	4.05	54
8	s	2.660	5.37	-1.10	29	38	e	2.893	2.19	-3.58	8
9	s	2.267	2.12	-3.99	5	39	e	7.888	7.72	3.80	51
10	s	1.701	3.11	-3.50	11	40	e	1.984	2.81	-3.59	7
11	s	2.856	3.22	-2.76	15	41	e	3.989	3.33	-2.02	23
12	s	5.627	14.96	8.47	60	42	e	7.198	6.35	2.28	45
13	s	8.459	7.55	3.98	53	43	e	2.415	2.61	-3.51	10
14	s	5.019	7.32	1.84	41	44	e	7.983	7.36	3.56	50
15	s	1.476	2.54	-4.10	4	45	e	6.753	6.28	1.96	43
16	s	1.982	1.46	-4.70	2	46	e	3.361	3.32	-2.39	21
17	s	2.747	2.33	-3.55	9	47	e	5.979	4.75	0.27	34
18	s	5.854	10.84	5.21	55	48	e	4.294	7.58	1.64	39
19	s	4.270	3.65	-1.60	27	49	e	3.097	3.09	-2.73	16
20	s	1.966	2.16	-4.13	3	50	e	3.117	2.52	-3.19	13
21	s	5.497	5.02	0.22	33	51	e	4.452	3.88	-1.30	28
22	s	3.051	2.36	-3.36	12	52	e	7.782	7.80	3.80	52
23	s	3.263	5.18	-0.91	30	53	e	2.648	2.22	-3.70	6
24	s	4.694	7.43	1.74	40	54	e	2.997	4.07	-1.98	24
25	s	3.275	2.99	-2.71	17	55	e	8.242	10.66	6.42	59
26	s	2.514	4.64	-1.78	26	56	e	1.858	1.47	-4.76	1
27	s	5.618	4.90	0.19	32	57	e	6.768	6.21	1.92	42
28	s	3.141	3.45	-2.41	20	58	e	4.896	12.90	6.35	58
29	s	3.414	3.09	-2.55	19	59	e	3.308	3.12	-2.59	18
30	s	7.808	7.48	3.55	49	60	e	7.698	6.68	2.83	46

각 측도들에 대한 학과별 기초통계량들은 다음과 같다.

학 과	학생의 수	측 도	최소값	최대값	평 균	표준편차
영어영문학과	29	절대편차의 평균	1.47	10.66	5.32	2.456
		편차의 표준편차	1.85	9.52	5.40	2.420
		제1주성분점수	-4.76	6.42	0.41	3.323
응용통계학과	29	절대편차의 평균	1.46	10.84	4.74	2.310
		편차의 표준편차	1.47	8.45	3.89	1.751
		제1주성분점수	-4.70	5.21	-0.92	2.708

T-검정을 통한 각 측도들에 대한 두 학과간의 비교결과는 다음과 같다.

변수 : 절대편차의 평균

학 과	개체의 수	평 균	표준편차	표준오차	최 소 값	최 대 값
영어영문학과	29	5.32	2.456	0.456	1.47	10.66
응용통계학과	29	4.74	2.310	0.429	1.46	10.84
분산들이		T-통계량값	자 유 도	P-값		
같지 않은 경우		0.9421	55.8	0.3502		
같은 경우		0.9421	56.0	0.3502		

두 집단의 모분산들이 동일하다는 귀무가설에 대한

F-통계량값 = 1.13, 자유도 = (28,28), P-값 = 0.7485

원그리기의 능력을 절대편차의 평균의 입장(즉 매끄러움의 입장)에서 볼 때는 10%의 유의수준하에서 두 학과간에는 차이가 없는 것으로 나타났다.

변수 : 편차의 표준편차

학 과	개체의 수	평 균	표준편차	표준오차	최 소 값	최 대 값
영어영문학과	29	5.40	2.420	0.449	1.86	9.52
응용통계학과	29	3.89	1.751	0.325	1.48	8.46
분산들이		T-통계량값	자 유 도	P-값		
같지 않은 경우		2.7183	51.0	0.0089		
같은 경우		2.7183	56.0	0.0087		

두 집단의 모분산들이 동일하다는 귀무가설에 대한

F-통계량값 = 1.91, 자유도 = (28,28), P-값 = 0.0925

그러나 원그리기의 능력을 편차의 표준편차의 입장에서 볼 때에는 10%의 유의수준하에서 두 학과간에는 큰 차이가 있는 것으로 나타났다. 편차의 표준편차라는 측도는 값이 작을수록 반지름을 잘 유지하면서 원을 그렸다는 것을 의미하므로 응용통계학과 학생들이 영어영문학과 학생들보다 반지름 유지라는 측면에서는 원을 더 잘 그린다고 할 수 있겠다.

이제 매끄러움과 반지름 유지라는 두 기준을 결합한 총체적 개념으로서의 제1주성분을 기준으로 두 학과간의 원그리기의 능력을 비교하면 어떤 결과가 나타날까? 이를 분석한 결과를 정리하면 다음과 같다.

T-검정의 결과

변수 : 제1주성분점수

학 과	개체의 수	평균	표준편차	표준오차	최 소 값	최 대 값
영어영문학과	29	0.41	3.323	0.617	-4.76	6.42
응용통계학과	29	-0.92	2.708	0.502	-4.70	5.21
분산들이		T-통계량값	자 유 도	P-값		
같지 않은 경우		1.6865	53.8	0.0975		
같은 경우		1.6865	56.0	0.0973		

두 집단의 모분산들이 동일하다는 귀무가설에 대한

F-통계량값 = 1.51, 자유도 = (28,28), P-값 = 0.2843

절대편차의 평균과 편차의 표준편차라는 두 가지의 기준을 결합하여 얻은 제1주성분점수의 입장에서 볼 때 두 학과간에는 10%의 유의수준하에서 통계적으로 유의한 차이가 있는 것으로 나타났다. 따라서 응용통계학과 학생들의 영어영문학과 학생들에 비해 원을 잘 그린다고 할 수 있다.

5. 전체, 그리고 각 학과별 대표원그림은?

총 60명의 학생들이 그린 원그림을 평균적인 개념으로 볼 때 어떤 패턴이 있는 지를 분석하기 위해 각 각도별로 편차의 평균과 표준편

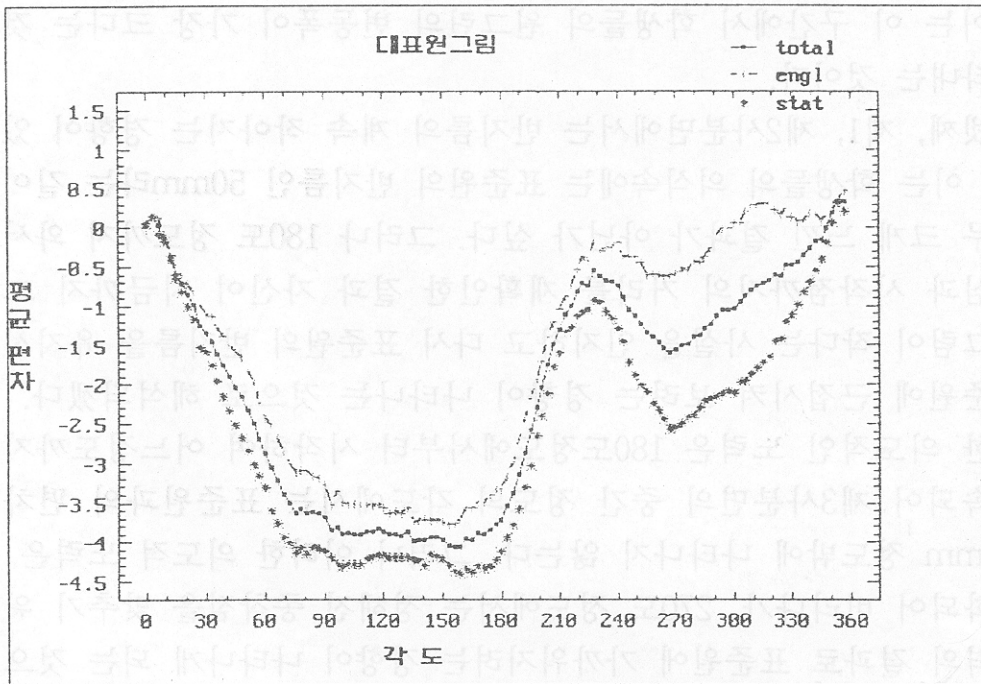
〈표 3〉 각도별 편차의 평균과 표준편차(전체, 그리고 학과별)

전					영어영문학과					융합통계학과					전					영어영문학과					융합통계학과				
각도	최소값	최대값	평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차	각도	최소값	최대값	평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차								
0	-1.0	2.0	0.025	0.297	-0.016	0.206	0.066	0.365	180	-20.0	15.0	-3.725	8.222	-3.333	9.337	-4.116	7.074												
3	-2.0	2.0	0.147	0.736	0.144	0.819	0.150	0.658	183	-20.0	15.5	-3.550	8.129	-3.166	9.209	-3.933	7.023												
6	-2.0	2.5	0.025	1.143	0.133	1.181	-0.083	1.114	186	-19.5	15.5	-3.383	8.041	-2.983	9.090	-3.783	6.971												
9	-2.5	3.5	-0.133	1.463	-0.083	1.559	-0.183	1.386	189	-19.0	16.0	-3.141	8.009	-2.700	9.031	-3.583	6.968												
12	-4.0	4.0	-0.371	1.888	-0.333	2.027	-0.416	1.771	192	-19.0	16.0	-2.875	7.945	-2.433	8.848	-3.316	7.053												
15	-4.5	4.5	-0.600	2.216	-0.550	2.478	-0.650	1.961	195	-18.5	16.0	-2.591	7.835	-2.150	8.673	-3.033	7.020												
18	-5.5	5.5	-0.708	2.596	-0.600	2.910	-0.816	2.283	198	-18.5	16.0	-2.300	7.840	-1.933	8.633	-2.666	7.088												
21	-7.0	6.0	-0.866	3.016	-0.733	3.339	-1.000	2.706	201	-18.0	16.5	-2.050	7.785	-1.700	8.525	-2.400	7.099												
24	-8.0	7.0	-0.991	3.380	-0.816	3.684	-1.166	3.099	204	-18.0	17.0	-1.758	7.771	-1.400	8.359	-2.116	7.261												
27	-9.5	8.0	-1.258	3.717	-0.983	4.088	-1.533	3.352	207	-17.5	17.5	-1.491	7.757	-1.200	8.282	-1.783	7.324												
30	-10.0	8.5	-1.341	3.978	-1.100	4.335	-1.583	3.643	210	-17.5	18.0	-1.316	7.654	-1.033	8.060	-1.600	7.352												
33	-10.5	9.5	-1.450	4.405	-1.133	4.890	-1.766	3.918	213	-17.0	18.0	-1.108	7.603	-0.883	7.896	-1.333	7.425												
36	-11.5	8.1	-1.615	4.733	-1.280	5.216	-1.950	4.259	216	-17.0	18.0	-1.000	7.538	-0.733	7.721	-1.266	7.472												
39	-12.0	9.0	-1.766	5.045	-1.383	5.579	-2.150	4.510	219	-17.0	18.0	-0.833	7.410	-0.533	7.479	-1.133	7.456												
42	-12.5	8.5	-1.908	5.245	-1.483	5.802	-2.333	4.683	222	-17.0	18.5	-0.725	7.363	-0.366	7.354	-1.083	7.479												
45	-13.5	8.0	-2.050	5.506	-1.583	6.023	-2.516	4.995	225	-17.0	18.5	-0.750	7.322	-0.433	7.220	-1.066	7.532												
48	-14.0	8.5	-2.165	5.724	-1.646	6.259	-2.683	5.190	228	-17.0	18.5	-0.558	7.258	-0.183	7.154	-0.933	7.463												
51	-15.0	9.5	-2.333	5.912	-1.816	6.376	-2.850	5.469	231	-17.5	18.0	-0.625	7.242	-0.300	7.160	-0.950	7.431												
54	-15.5	10.0	-2.500	6.154	-1.983	6.571	-3.016	5.773	234	-18.0	18.0	-0.675	7.196	-0.266	7.064	-1.083	7.424												
57	-16.0	10.0	-2.700	6.332	-2.233	6.724	-3.166	5.993	237	-18.0	18.0	-0.706	7.175	-0.246	7.063	-1.166	7.378												
60	-16.5	10.0	-2.875	6.421	-2.433	6.806	-3.316	6.096	240	-18.0	18.0	-0.750	7.160	-0.200	7.103	-1.300	7.295												
63	-16.5	10.0	-3.058	6.541	-2.516	7.015	-3.600	6.103	243	-18.0	18.0	-0.900	7.197	-0.350	7.153	-1.450	7.320												
66	-17.0	10.0	-3.225	6.733	-2.683	7.237	-3.766	6.265	246	-18.0	18.0	-1.058	7.193	-0.383	7.162	-1.733	7.282												
69	-17.5	10.0	-3.400	6.878	-2.866	7.395	-3.933	6.402	249	-18.0	18.0	-1.133	7.294	-0.383	7.317	-1.883	7.317												
72	-18.0	10.0	-3.508	6.964	-3.000	7.538	-4.016	6.427	252	-18.5	18.5	-1.233	7.330	-0.450	7.395	-2.016	7.305												
75	-18.0	10.0	-3.625	7.014	-3.116	7.610	-4.133	6.453	255	-18.5	19.0	-1.350	7.391	-0.600	7.483	-2.100	7.346												
78	-18.0	10.5	-3.575	7.091	-3.083	7.641	-4.066	6.589	258	-18.5	20.0	-1.408	7.415	-0.566	7.573	-2.250	7.284												
81	-19.0	11.0	-3.625	7.208	-3.100	7.769	-4.150	6.691	261	-18.5	20.0	-1.466	7.440	-0.600	7.558	-2.333	7.344												
84	-18.5	11.0	-3.616	7.282	-3.166	7.891	-4.066	6.723	264	-18.5	20.0	-1.541	7.451	-0.633	7.636	-2.450	7.276												
87	-18.5	11.0	-3.633	7.328	-3.183	7.907	-4.083	6.806	267	-18.5	20.0	-1.591	7.472	-0.616	7.621	-2.566	7.317												
90	-19.0	10.5	-3.691	7.387	-3.266	7.904	-4.116	6.941	270	-18.5	20.0	-1.558	7.477	-0.566	7.608	-2.550	7.337												
93	-19.0	11.0	-3.858	7.365	-3.533	7.817	-4.183	7.001	273	-18.5	20.0	-1.508	7.430	-0.516	7.542	-2.500	7.308												
96	-19.0	11.0	-3.808	7.336	-3.333	7.781	-4.283	6.963	276	-18.0	20.0	-1.491	7.381	-0.550	7.507	-2.433	7.256												
99	-19.5	11.0	-3.900	7.393	-3.483	7.854	-4.316	7.010	279	-18.0	20.0	-1.408	7.331	-0.433	7.429	-2.383	7.223												
102	-19.5	11.5	-3.891	7.452	-3.500	7.926	-4.283	7.060	282	-18.0	19.0	-1.375	7.251	-0.450	7.365	-2.300	7.139												
105	-20.0	11.5	-3.900	7.458	-3.516	7.947	-4.283	7.051	285	-17.5	18.0	-1.300	7.150	-0.383	7.289	-2.216	7.010												
108	-20.0	10.5	-3.900	7.422	-3.516	7.842	-4.283	7.091	288	-17.0	18.0	-1.233	7.102	-0.266	7.349	-2.200	6.832												
111	-20.0	11.0	-3.866	7.424	-3.516	7.806	-4.216	7.138	291	-16.5	17.0	-1.125	6.968	-0.100	7.263	-2.150	6.623												
114	-20.0	11.5	-3.866	7.488	-3.550	7.834	-4.183	7.246	294	-16.0	16.5	-1.066	6.830	-0.016	7.091	-2.116	6.508												
117	-20.0	10.5	-3.831	7.557	-3.500	8.018	-4.163	7.189	297	-15.5	16.0	-1.050	6.725	-0.016	7.050	-2.083	6.331												
120	-20.0	10.5	-3.866	7.642	-3.566	8.162	-4.166	7.212	300	-15.0	15.5	-1.033	6.577	0.050	7.016	-2.116	6.029												
123	-20.0	10.0	-3.908	7.688	-3.633	8.253	-4.183	7.209	303	-14.5	15.0	-0.958	6.469	0.066	7.000	-1.983	5.828												
126	-20.0	10.0	-3.900	7.788	-3.600	8.392	-4.200	7.265	306	-14.0	15.0	-0.875	6.264	0.200	6.847	-1.950	5.527												
129	-20.0	10.5	-3.916	7.862	-3.633	8.476	-4.200	7.332	309	-13.5	15.0	-0.800	6.111	0.266	6.709	-1.866	5.349												
132	-20.0	11.5	-3.876	7.950	-3.570	8.610	-4.183	7.366	312	-12.5	15.0	-0.758	5.884	0.283	6.567	-1.800	5.006												
135	-20.5	12.0	-3.991	8.087	-3.750	8.850	-4.233	7.391	315	-12.0	15.0	-0.708	5.698	0.316	6.454	-1.733	4.715												
138	-20.0	12.0	-3.975	8.191	-3.716	8.950	-4.233	7.500	318	-11.5	15.0	-0.683	5.524	0.283	6.360	-1.650	4.439												
141	-20.5	12.0	-3.983	8.287	-3.666	9.049	-4.300	7.592	321	-11.0	15.0	-0.616	5.234	0.216	6.140	-1.450	4.075												
144	-20.0	12.0	-3.933	8.305	-3.716	9.069	-4.150	7.615	324	-10.5	15.0	-0.631	4.967	0.153	5.893	-1.416	3.767												
147	-20.0	12.0	-3.941	8.367	-3.700	9.207	-4.183	7.585	327	-10.0	15.0	-0.483	4.704	0.233	5.642	-1.200	3.482												
150	-20.0	12.0	-3.983	8.386	-3.750	9.270	-4.216	7.552	330	-10.0	14.5	-0.491	4.396	0.133	5.304	-1.116	3.220												
153	-20.0	13.0	-4.041	8.417	-3.783	9.375	-4.300	7.490	333	-9.5	14.0	-0.450	4.158	0.083	5.097	-0.983	2.934												
156	-20.0	14.0	-4.066	8.482	-3.766	9.531	-4.366	7.439	336	-8.5	14.0	-0.358	3.920	0.100	4.853	-0.816	2.698												
159	-20.0	16.0	-4.058	8.597	-3.700	9.721	-4.416	7.458	339	-7.5	13.0	-0.258	3.577	0.183	4.453	-0.700	2.409												
162	-20.0	16.0	-3.950	8.602	-3.566	9.711	-4.333	7.480	342	-6.0	12.0	-0.183	3.174	0.183	4.018	-0.550	2.018												
165	-20.0	16.0	-3.975	8.573	-3.566	9.757	-4.383	7.349	345	-6.0	11.0	-0.166	2.770	0.083	3.506	-0.416	1.786												
168	-20.0	16.0	-3.891	8.532	-3.483	9.709	-4.300	7.313	348	-5.0	10.5	-0.008	2.482	0.183	3.144	-0.200	1.606												
171	-20.0	16.0	-3.908	8.440	-3.483	9.612	-4.333	7.223	351	-5.0	10.0	0.175	2.152	0.283	2.747	0.066	1.362												
174	-20.0	16.0	-3.85																										

차를 계산하여 보았다. 물론 학과별 패턴과 전체의 것을 비교하기 위해 학과별의 값들도 계산하였다. 이를 정리하면 <표 3>과 같다.

이 표에서 각 각도별로 살펴보면 알 수 있듯이, 학과별 편차의 평균의 차이에 비해 표준편차의 값들이 월등히 크다는 사실을 알 수 있다. 이러한 사실로부터 각도별 학과간의 유의한 차이가 있는 지에 관해서는 통계적 검정절차를 따라 거치지 않더라도 각도별로는 학과간에 차이가 없다라고 결론을 내릴 수 있다.

<표 3>의 각도별 편차의 평균을 이용하여 전체 학생들의 원그림들을 대표할 수 있는 대표원그림과 각 학과별 대표원그림을 직교좌표상에서 그려보면 아래의 [그림 5]와 같다.



[그림 5] 직교좌표에서 표현된 전체와 각 학과의 대표원그림

이 그림을 저자의 주관을 가미하여 해석하면 다음과 같다.

첫째, 평균적으로 학생들은 표준원보다 작게 그리는 경향이 있다.

그러나 학과별로는 영어영문학과와 대표원그림이 응용통계학과와 대표원그림보다 표준원에 가까운 것으로 나타났다. 이는 개별학생들을 기준으로 학과별 비교분석에서 얻은 결론과 상치되는 것으로 영어영문학과와 학생들 중에는 아주 큰 원그림, 아주 작은 원그림, 또는 들쭉날쭉한 원그림들을 그린 학생들이 응용통계학과와 학생들에 비해 상대적으로 많아서 상쇄효과에 의해 발생한 것으로 이해될 수 있겠다.

둘째, 제2사분면에서 편차의 평균이 가장 심하게 나타난다.

이는 제2사분면에서 표준원과의 차이가 가장 심함을 나타내는 것이다.

셋째, 표준편차의 값이 가장 큰 부분이 150도에서 180도 정도의 사이인 것으로 나타났다.

이는 이 구간에서 학생들의 원그림의 변동폭이 가장 크다는 것을 나타내는 것이다.

넷째, 제1, 제2사분면에서는 반지름의 계속 작아지는 경향이 있는데, 이는 학생들의 의식속에는 표준원의 반지름인 50mm라는 길이를 너무 크게 느낀 결과가 아닌가 싶다. 그러나 180도 정도까지 와서는 중심과 시작점까지의 거리를 재확인한 결과 자신이 지금까지 그린 원그림이 작다는 사실을 인지하고 다시 표준원의 반지름을 유지시키 표준원에 근접시켜 보려는 경향이 나타나는 것으로 해석되겠다. 이러한 의도적인 노력은 180도정도에서부터 시작하여 어느정도까지는 지속되어 제3사분면의 중간 정도의 각도에서는 표준원과의 편차가 0.5mm 정도밖에 나타나지 않는다. 그러나 이러한 의도적 노력은 둔약화되어 버리다가, 270도 정도에서는 정해진 종착점을 맞추기 위한 노력의 결과로 표준원에 가까워지려는 경향이 나타나게 되는 것으로 해석되겠다.

다섯째, 패턴이 변화하는 지점의 각도가 90도, 150도에서 180도 사이, 225도, 270도 정도로서 이 지점들이 바로 학생들이 자신이 그때까지 그린 원그림을 의도적으로 확인해 보고 조치를 취하는 시점이 아닌가 싶다.

6. 본 연구를 끝내며

본 연구에서는 우리 목원대학교의 영어영문학과와 응용통계학과의 두 학과에서 각각 30명씩의 학생들을 임의로 추출하여 반지름이 5cm인 원을 그리게하고, 이들로부터 얻어진 60장의 원그림들을 자료로하여 두 집단간에 원을 그리는 능력의 차이가 있는지의 여부와 원그리기에 어떤 패턴 등이 있는지를 통계적으로 분석해 보았다.

본 연구를 통해 얻은 주요 결론들은 아래의 몇가지로 요약될 수 있지만, 이보다도 그림자료들도 적절한 측도들만 개발될 수 있다면 객관적으로 비교분석이 가능할 수 있다는 자신감을 가진 것이 가장 큰 수확임을 밝힌다.

첫째, 본 연구에서 설정한 잘 그린 원그림의 판단측도들(절대편차의 평균, 편차의 표준편차, 이들로부터 유도되어진 단일측도인 제1주성분)에 대해 그런대로 측도로서의 타당성이 있음을 확인할 수 있었다.

둘째, 응용통계학과의 학생들이 영어영문학과 학생들보다 원그리기의 능력에 있어서 우수함을 통계적으로 확인할 수 있었고, 우수함의 근원은 매끄러움보다는 반지름의 유지능력에 있음도 알 수 있었다.

셋째, 60명의 학생들을 전체로하여 볼 때 전체적으로 표준원보다 작게 그리는 경향이 있음과 두 학과의 대표원그림의 패턴이 비슷함을 확인할 수 있었다. 그리고 원그리기 과정에서 몇 차례 정도 표준원으로의 복원노력이 일어남도 확인되었다.

그러나 응용통계학과 학생들과 영어영문학과 학생들간에 원그리기의 능력에 있어서 차이를 보이게 한 원인, 그리고 학과간 대표원그림의 패턴이 비슷한 이유에 관한 궁금함은 여전히 남는다.

[illegible]

서울특별시 구별 특성에 따른 쓰레기 배출량 분석

서두성, 이종현, 이규학

(지도교수 : 유성모)

고려대학교 응용통계학과

1. 서 론

산업의 발달과 경제 성장의 부작용으로 인한 환경오염이 날로 심각해지고 있다. 인구의 증가, 이농 현상에 의한 도시 집중화 현상, 도시 공동화 현상, 소득의 증가와 산업화로 인하여 발생하는 폐기물은 환경오염의 심각성을 가중시키고 있다. 폐기물 처리 문제는 지역이기주의와 맞물려 심각한 사회문제로 대두되고 있다.

특히 폐기물 처리 문제는 쓰레기 처리 문제가 큰 비중을 차지하므로 본 논문에서는 인구가 가장 밀집되어 있는 서울특별시 구별 쓰레기 배출량과 각 구별 특성을 바탕으로, 통계적 기법을 이용하여 쓰레기 배출량에 미치는 주된 요인을 분석하는데 중점을 두고 있다.

통계적 기법으로 변수 선택 과정을 통한 쓰레기 배출량에 관한 적절한 회귀 모형을 도출하고 모형에 대한 진단을 중점적으로 분석하였다.

2. 본 론

2.1. 변수의 정의

본 논문에 이용된 자료는 서울특별시 구별 자료가 나와 있는 제 35회 서울특별시 통계 연보 자료(1994)를 이용하였다.

변수에 대한 설명 및 자료¹⁾는 <표 1>, <표 2>와 같다.

본 논문에서 변수는 '변수 설명(변수명)'으로 통일하여 사용하기로 한다. 예로서 쓰레기 배출량을 나타내는 GARBAGE는 쓰레기 배출량(GARBAGE)으로 사용하기로 한다.

<표 1> 변수의 설명

변 수 명	설 명	비 고
GARBAGE	쓰레기 배출량	1993 ; 톤 / 일
POPULATN	특정 폐기물(산업 폐기물)을 제외한 일반 폐기물	
AREA	인구	1993 ; 명
AGRICALT	면적	1993 ; km ²
WORKER	농지 면적(1차 산업 형태 구조)	1993 ; m ²
MARKET	전, 답, 과수원, 목양지	
RESTAURN	제조업의 월평균 종사자(2차 산업 형태 구조)	1992 ; 명
INFLOW	시장 분포 합계(3차 산업 형태 구조)	1993 ; 개
RELIF	사설 시장, 백화점, 쇼핑센터	
PARK	음식점 수(3차 산업 형태 구조)	1993 ; 개
WELFARE	일반 음식점, 유흥 음식점, 휴게 음식점	
OWNHOUSE	유입 인구(주택단지, 공동화 현상)	1990 ; 명
TAX	주간 인구-상주(야간) 인구	
	생활 보호 대상자	1993 ; 명
	거택 보호 대상자, 자활 보호 대상자, 시설 보호 대상자	
	위락 시설	1993 ; m ²
	공원, 체육 용지, 유원지	
	복지시설 수용 인원	1993 ; 명
	아동 복지시설, 성인 복지시설	
	자기집의 비율(자기집 / 점유 형태별 가구 합계)	1990 ; %
	점유 형태별 가구 합계(자기 집, 전세, 보증부 월세, 월세, 무상)	
	세금	1993 ; 백만원
	취득세, 등록세, 재산세, 자동차세	

1) 유입 인구(INFLOW) 및 자기집의 비율(TOWNHOUSE)의 자료는 1990년에 실시된 인구 센서스 자료를 바탕으로 『인구 주택 총조사 보고서』를 이용하였다.

<표 2>

선택된 변수

ID	구 명	GARBAGE	POPULATN	AREA	AGRICALT	WORKER	MARKET	RESTAURN	INFLOW	RELIEF	PARK	WELFARE	OWNHOUSE	TAX
A	중 로	557	222423	24.01	1014000	1093	19	5881	252742	2038	202402	880	0.36599	56101
B	중 구	769	167383	9.99	0	2731	47	5202	370360	1469	342918	113	0.29447	60558
C	용 산	461	276439	21.86	108052	8117	18	3700	74923	2119	15683	264	0.36224	45264
D	성 동	1252	767976	33.94	1144382	5222	27	7139	-26429	4491	842755	143	0.29398	86407
E	동대문	785	459000	14.77	288248	2050	24	4442	66955	2781	81968	84	0.34539	41423
F	중 양	519	460157	18.13	2395168	1200	12	3457	-71187	4482	34083	58	0.30000	40090
G	성 북	663	514240	24.29	240543	1127	17	3744	-40357	3445	26976	187	0.36954	44824
H	도 불	763	780472	44.55	1533991	1770	26	5043	-121938	12526	129588	339	0.39802	80253
I	노 원	528	588188	35.60	2697320	4824	14	2632	-65312	25767	360179	466	0.48306	78125
J	은 평	460	512196	29.68	1650052	5326	21	3511	-77190	4667	42316	4173	0.43067	50756
K	시대문	552	381905	17.69	200046	4939	10	4029	23416	1926	44253	184	0.39435	47102
L	마 로	805	422406	23.86	3089868	1329	18	3307	-13023	6631	47064	306	0.39664	53326
M	양 천	489	499848	17.57	796967	9011	11	2735	-96376	1582	501997	333	0.41057	61490
N	강 시	665	474434	41.42	10705836	1511	21	2895	-9852	7320	82141	649	0.37692	78783
O	구 로	980	724230	32.80	2375991	9527	33	6517	-312	5899	121400	547	0.30659	79950
P	영동포	909	445221	24.44	210065	3624	35	5625	156229	2579	129543	373	0.35203	95412
Q	동 직	462	440361	16.37	145029	4980	19	3157	-27783	3120	443077	1614	0.35687	53394
R	관 약	636	590442	29.65	284130	1093	22	4185	-99050	6442	123348	408	0.33643	54818
S	서 초	654	410719	47.32	6152496	7062	39	6009	39466	694	562764	401	0.51213	134902
T	강 남	849	566379	39.26	4580116	7469	40	10225	125418	12245	680583	453	0.52581	225563
U	송 파	1413	685260	33.74	3525371	5905	18	6574	-79855	1967	1927082	521	0.43045	96667
V	강 동	850	535785	24.46	3528772	8169	25	4186	-60982	2590	459615	579	0.37670	63463

2.2. 기초 분석

(1) 상관 분석

쓰레기 배출량(GARBAGE)의 정규성 검정 결과 p-value가 0.0094로 회귀 모형에 대한 정규성이 만족되지 않았으나 로그 변환을 통하여 정규성(p-value=0.2364)을 만족시켰다. 이후에 사용되는 쓰레기 배출량(GARBAGE)은 원래의 쓰레기 배출량을 로그 변환한 값이며 변수명을 그대로 사용한다.

다음은 회귀 분석 실시에 앞서서 각 변수들간의 상관관계를 유의수준 $\alpha=0.1$ 에서 유의한 계수들을 도표로 작성하였다.

〈표 3〉 변수들간의 상관 계수

	GARBAGE	POPULATN	AREA	AGRICULT	WORKER	MARKET	RESTAURN	INFLOW	RELIEF	PARK	WELFARE	OWNHOUSE	TAX
GARBAGE		0.506 **				0.446 **	0.635 ***			0.584 ***			0.384 **
POPULATN			0.575 ***					-0.705 ***	0.404 **				
AREA				0.651 ***			0.378 *		0.450 **			0.548 **	0.609 ***
AGRICULT												0.418 *	0.444 **
WORKER							0.652 ***						0.363 *
MARKET								0.584 ***					0.579 ***
RESTAURN								0.367 *		0.453 **			0.790 ***
INFLOW													
RELIEF												0.399 *	
PARK													0.430 **
WELFARE													
OWNHOUSE													0.628 ***
TAX													

〈표 3〉에서 쓰레기 배출량(GARBAGE)과 상관관계가 유의하게 나타난 변수는 인구(POPULATN), 시장 분포(MARKET), 음식점 수(RESTAURN), 위락 시설(PARK), 세금(TAX)이다. 이로부터 회귀 모형에 이들 변수가 적용될 것이라고 예상을 할 수 있다. 그러나, 세금(TAX)의 경우 다른 변수와의 상관관계가 높으므로 다른 변수들에 의해 설명되거나 다른 변수들을 설명할 수 있을 거라 예측할 수 있다. 또한 시장 분포(MARKET)는 음식점 수(RESTAURN)와 상관관계가 0.652로 유의하므로 서로 연관을 갖고 관찰을 해 볼 필요가 있다.

쓰레기 배출량과 상관관계가 있는 인구(POPULATN), 시장 분포(MARKET), 음식점 수(RESTAURN), 위락 시설(PARK), 세금(TAX) 변수에 더욱 관심을 가지고 회귀 모형을 구축하고자 한다.

(2) 설명 변수들의 다중공선성 영향력

〈표 4〉는 회귀 모형에서 각 설명 변수들의 다중공선성(MULTICOLLINEARITY)이 각 회귀 계수 추정치의 정도에 미치는 영향을

SAS의 PROC REG에서 VIF(Variance Inflation Factor) 옵션을 사용하여 알아보았다. 전체 회귀 모형에서 상관관계 및 VIF값을 고려하여 제거하는 방법을 선택하였다.

〈표 4〉

VIF 값

변 수 명	1차 VIF값	2차 VIF값	3차 VIF값
POPULATN	17.38995825	3.56300664	3.36437476
AREA	5.97803153	5.92741393	5.39048907
AGRICALT	3.37474132	3.35959937	2.29217281
WORKER	1.49512024	1.49478835	1.44253046
MARKET	4.74710374	2.89324085	2.32157781
RESTAURN	15.46826186	8.45719957	2.95819994
INFLOW	18.00162697		
RELIF	3.71438987	2.12241306	1.92705512
PARK	2.00985875	1.92256846	1.87102631
WELFARE	1.26813555	1.26142623	1.17815638
OWNHOUSE	6.58934354	6.14153389	2.91973840
TAX	16.88339102	15.49376685	

1차 VIF값 중 유입 인구(INFLOW)의 값이 18.00으로 가장 높으며 이것은 〈표 2〉의 변수간의 상관 계수에서 유입 인구(INFLOW)의 설명 변수가 시장 분포(MARKET)와 인구(POPULATN)간의 상관 계수가 각각 0.584, -0.705로 상관관계가 높아 시장 분포(MARKET)와 인구(POPULATN)가 설명할 수 있으며 쓰레기 배출량(GARBAGE)과의 상관 계수도 작으므로 유입 인구(INFLOW)를 제거하였다.

다음으로 2차 VIF값 중 세금(TAX)의 값이 15.49로 가장 높으며 이것은 〈표 2〉의 변수간의 상관 계수에서 TAX의 설명 변수가 자기집의 비율(OWNHOUSE), 음식점 수(RESTAURN)간의 상관 계수가 각각 0.628, 0.790으로 높은 상관관계가 있어 자기집의 비율

(OWNHOUSE)과 음식점 수(RESTAURN)가 설명할 수 있으므로 TAX를 제거하였다.

3차 VIF값은 1~5로 높은 VIF값이 없다. 따라서 회귀 계수 추정치의 정도에 미치는 영향이 없는 10개의 변수로 회귀 모형의 변수 선택을 실시하였다.

2.3 회귀 모형의 선택 및 진단

(1) 회귀 모형의 선택

다중공선성으로 회귀 계수 추정치에 영향을 미치는 변수를 제외한 10개의 변수에서 변수 선택을 실시하였다. 여러 변수 선택 방법 중 STEPWISE와 BACKWARD 방법에 의해 선택된 변수로 회귀 모형을 만들었다. 또한 각 변수들간의 비교 분석을 위하여 표준화된 회귀 계수의 추정치(Standardized Coefficient Estimate)를 이용하였다.

〈표 5〉 STEPWISE 방법으로 선택된 회귀 모형

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F	R-SQUARE
MODEL	4	1.447	0.362	9.661	0.0003	0.6945
ERROR	17	0.637	0.037			
C TOTAL	21	2.084				

Parameter Estimates						
Variable	DF	Standardized Estimate	Parameter Estimate	Variance Inflation	T for H0: Parameter=0	Prob > T
INTERCEP	1	0.0000000	6.431321	0.0000000	21.548	0.0001
POPULATN	1	0.29088717	0.00000575	1.15429781	2.020	0.0595
RESTAURN	1	0.45728303	0.000079331	1.29604970	2.996	0.0081
PARK	1	0.36977214	0.00000271	1.44770697	2.292	0.0349
OWNHOUSE	1	-0.34321479	-1.673885	1.09840377	-2.443	0.0258

<표 6>

BACKWARD 방법에 의한 회귀 모형

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F	R-SQUARE
MODEL	4	1.463	0.366	10.016	0.0002	0.7021
ERROR	17	0.621	0.037			
C TOTAL	21	2.084				

Parameter Estimates						
Variable	DF	Standardized Estimate	Parameter Estimate	Variance Inflation	T for H0: Parameter=0	Prob > T
INTERCEP	1	0.00000000	5.822963	0.00000000	32.657	0.0001
POPULATN	1	0.39672130	0.00000784	1.15150688	2.793	0.0125
MARKET	1	0.40247335	0.012808	1.02667816	3.001	0.0080
PARK	1	0.37811413	0.00000277	1.16789349	2.643	0.0171
WELFARE	1	-0.29043465	-0.000106	1.01203872	-2.181	0.0435

<표 5>와 <표 6>에서 회귀 모형에 해당되는 F값은 9.661, 10.016이며, 매우 유의함(P-value ; 0.03%, 0.02%)을 알 수 있다. 회귀에 의하여 설명된 분산(R-SQUARE)은 각각 전체 분산의 69.5%, 70.1%로 서로 차이가 거의 없다.

그러나 쓰레기 배출량의 상당 부분이 음식물 쓰레기²⁾이며, 음식점 수(RESTAURN)와 쓰레기 배출량(GARBAGE)의 상관 계수가 시장 분포(MARKET)와 쓰레기 배출량(GARBAGE)의 상관 계수보다 높아 회귀 모형을 설명하기 적합하므로 STEPWISE 방법에 의해 선택된 회귀 모형을 토대로 분석하고자 한다.

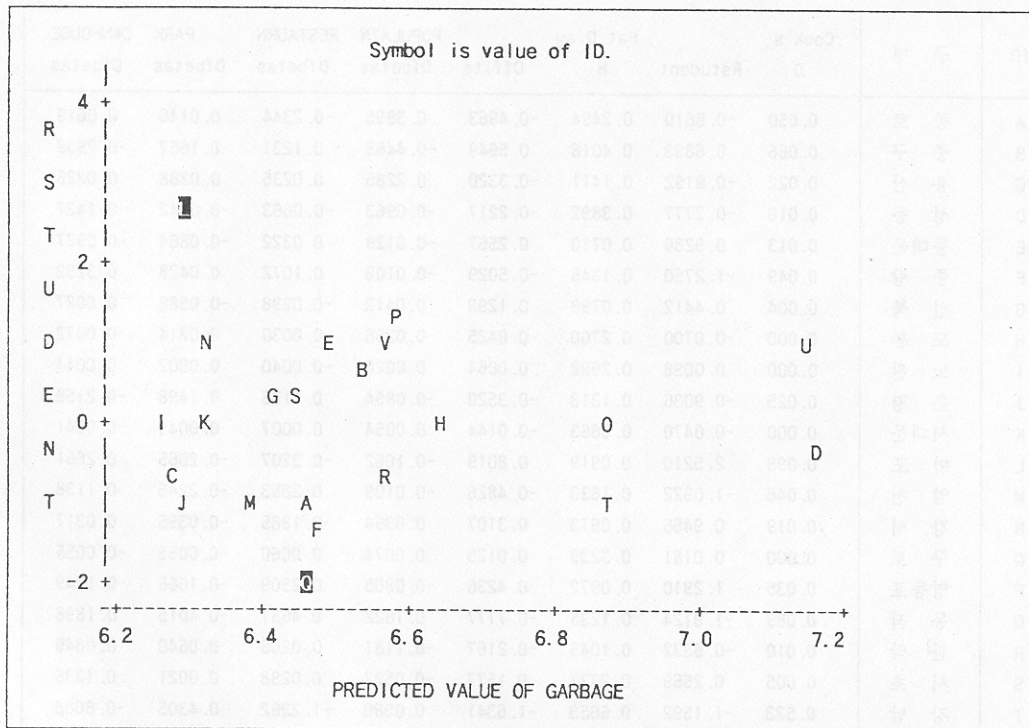
<표 5>에서는 선택된 변수들의 표준화된 회귀 계수의 추정치를 나타내며 각각의 회귀 계수 추정치가 0.291(인구 ; POPULATN), 0.457(음식점 수 ; RESTAURN), 0.370(위락 시설 ; PARK), -0.343(자기집의 비율 ; OWNHOUSE)이다.

2) 전국적으로 하루 4만7천 톤의 쓰레기가 발생되며, 그중 38%가 음식물 쓰레기이다.
문화 일보 1996년 7월 1일 21면 참조.

이러한 결과로 추정된 회귀식에 대한 분석은 2.4에서 보다 자세히 다루기로 한다.

(2) 회귀 모형의 진단

추정된 회귀 모형에 대해 오차항의 독립성 검정을 실시하였다. 오차항의 독립성을 검정하는 Durbin-Watson 값은 2.689로 4보다 2에 더 가까우므로 오차항들 간에 1차 자기 상관이 없다고 볼 수 있으므로 독립성을 만족한다.



[그림 1] 잔차플롯

[그림 1]은 외표준 잔차(RSTUDENT)와 예측치(PREDICTED VALUE)의 PLOT이다. 외표준 잔차는 i 번째 관측치를 제외하고 회귀식에 적합하였을 때의 표준 오차로 i 번째 잔차를 나누어 표준화시킨 값으로 자유도 15인 t 분포를 따른다. 일반적으로 외표준 잔차의

절대값이 2 이상인 관측치를 이상치로 간주한다. 잔차플롯의 경향성을 통해 등분산성, 독립성, 정규성 및 이상치의 유무 등을 시각적으로 확인할 수 있다.

[그림 1]에서 잔차의 특이한 경향이 없는 것으로 보이므로 등분산성을 만족하며 외표준 잔차의 절대값이 2에 가까운 L(마포구 ; 2.5210)과 Q(동작구 ; -1.9124)가 잠재적 이상치임을 알 수 있다.

〈표 7〉 회귀 진단 통계량

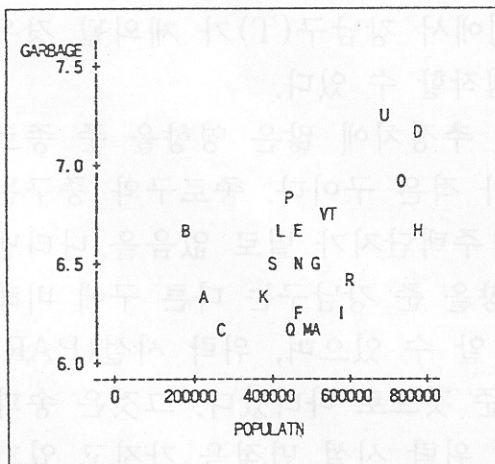
ID	구 명	Cook's D		Hat Diag H		POPULATN Dfbetas	RESTAURN Dfbetas	PARK Dfbetas	GWN-HOUSE Dfbetas
		D	Rstudent	H	Dffits	Dfbetas	Dfbetas	Dfbetas	Dfbetas
A	종 로	0.050	-0.8610	0.2494	-0.4963	0.3995	-0.2344	0.0146	0.0619
B	중 구	0.066	0.6893	0.4018	0.5649	-0.4465	0.1231	0.1657	-0.2938
C	용 산	0.022	-0.8192	0.1411	-0.3320	0.2285	0.0236	0.0288	0.0226
D	성 동	0.010	-0.2777	0.3892	-0.2217	-0.0963	-0.0663	-0.0542	0.1427
E	동대문	0.013	0.9286	0.0710	0.2567	-0.0129	0.0322	-0.0864	-0.0927
F	중 량	0.049	-1.2750	0.1346	-0.5029	-0.0108	0.1072	0.0428	0.3292
G	성 북	0.004	0.4412	0.0798	0.1299	0.0412	-0.0298	-0.0588	0.0027
H	도 봉	0.000	-0.0700	0.2700	-0.0425	-0.0366	-0.0030	0.0214	-0.0072
I	노 원	0.000	0.0098	0.2992	0.0064	0.0020	-0.0040	0.0002	0.0044
J	은 평	0.025	-0.9036	0.1318	-0.3520	-0.0894	0.1126	0.1498	-0.2158
K	서대문	0.000	-0.0470	0.0863	-0.0144	0.0054	0.0007	0.0049	-0.0041
L	마 포	0.098	2.5210	0.0919	0.8019	-0.1062	-0.3207	-0.2065	0.2661
M	양 천	0.046	-1.0922	0.1633	-0.4826	-0.0109	0.3853	-0.2246	-0.1138
N	강 서	0.019	0.9466	0.0973	0.3107	0.0354	-0.1885	-0.0396	0.0317
O	구 로	0.000	0.0181	0.3239	0.0125	0.0074	0.0060	-0.0055	-0.0055
P	영등포	0.035	1.2910	0.0972	0.4236	-0.0805	0.2509	-0.1666	-0.1249
Q	동 작	0.089	-1.9124	0.1235	-0.7177	0.1622	0.4637	-0.4015	0.1898
R	관 악	0.010	-0.6332	0.1049	-0.2167	-0.1181	0.0206	0.0640	0.0849
S	서 초	0.005	0.2569	0.2737	0.1577	-0.0527	0.0298	0.0021	0.1236
T	강 남	0.523	-1.1592	0.6653	-1.6341	0.0580	-1.2392	0.4305	-0.8088
U	송 파	0.642	1.0662	0.7401	1.7993	-0.0197	-0.3505	1.6060	-0.1364
V	강 동	0.013	0.9579	0.0650	0.2527	0.0436	-0.1101	0.0981	-0.0323

〈표 7〉에서 Cook's D 통계량은 i 번째 관측치가 제외됐을 때 최소 제곱 추정치 벡터를 표준화한 값으로 관측치의 영향력을 측정할

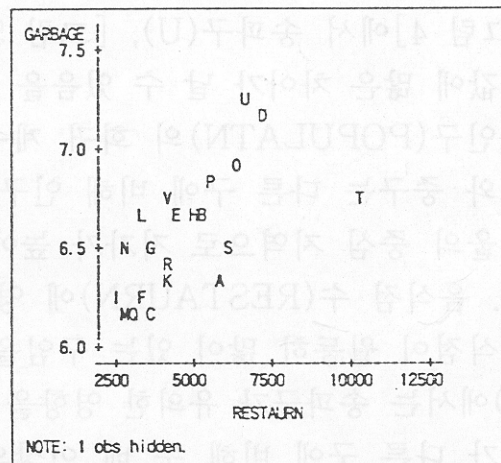
수 있다. 보통 1보다 큰 경우나 다른 관찰치 보다 특별히 큰 경우 영향력이 크다고 본다. 강남구(T)와 송파구(U)가 각각 0.523, 0.642로 1보다 크지는 않으나 다른 변수에 비해 상당히 크므로 잠재적으로 영향을 줄 수 있다.

Hat diag H 는 해당 관측치의 설명 변수들이 평균에서 떨어져 있는 정도를 나타내며, Diffits는 i 번째 관측치를 제외하고 회귀식을 적합한 뒤 추정된 값의 변화를 표준화한 척도이다. 이 두개의 관측치에서는 강남구(T ; Hat diag H = 0.6653, Diffits = -1.6341)와 송파구(U ; Hat diag H = 0.7401, Diffits = 1.7993)가 높은 값을 보여 회귀 계수 추정에 영향을 준 것으로 나타난다.

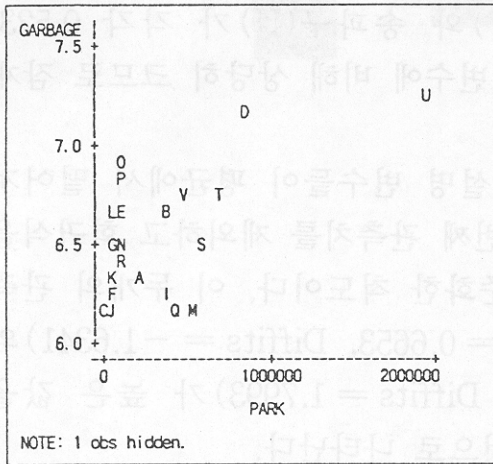
Dfbetas 값은 i 번째 관측치가 제외됐을 때 회귀 계수 β_k 의 추정치의 변화를 표준화한 척도이며, 절대값이 크면 영향력이 크다고 할 수 있다. <표 8>을 통해 인구(POPULATN)는 종로구(A ; 0.3995)와 중구(B ; -0.4465), 음식점 수(RESTAURN)는 강남구(T ; -1.2392), 위락 시설(PARK)은 송파구(U ; 1.6060), 자기 집의 비율(OWNHOUSE)은 강남구(T ; -0.8088)가 각각의 회귀 계수를 추정하는데 큰 영향을 준 것으로 생각된다.



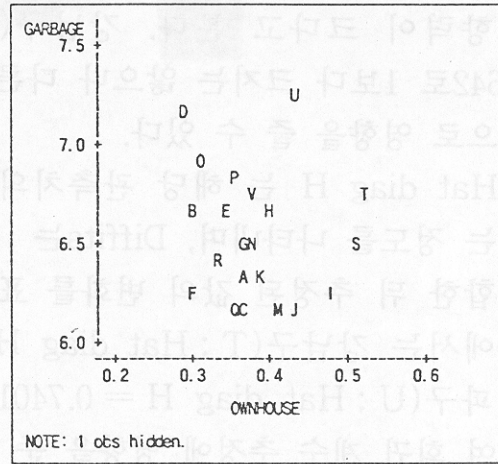
[그림 2] POPULATN



[그림 3] RESTAURN



[그림 4] PARK



[그림 5] OWNHOUSE

[그림 2]에서 [그림 5]는 쓰레기 배출량과 인구(POPULATN), 음식점 수(RESTAURN), 위락 시설(PARK), 자기집의 비율(OWNHOUSE)과의 PLOT이다. 이 PLOT을 통해 각 변수의 일반적인 경향과 상관 관계를 시각적으로 살펴볼 수 있다.

[그림 2]에서 종로구(A)와 중구(B), [그림 3]에서 강남구(T), [그림 4]에서 송파구(U), [그림 5]에서 강남구(T)가 제외될 경우 β_k 값에 많은 차이가 날 수 있음을 짐작할 수 있다.

인구(POPULATN)의 회귀 계수 추정치에 많은 영향을 준 종로구와 중구는 다른 구에 비해 인구가 적은 구이다. 종로구와 중구는 서울의 중심 지역으로 지가가 높아 주택단지가 별로 없음을 나타낸다. 음식점 수(RESTAURN)에 영향을 준 강남구는 다른 구에 비해 음식점이 월등히 많이 있는 구임을 알 수 있으며, 위락 시설(PARK)에서는 송파구가 유의한 영향을 준 것으로 나타났다. 그것은 송파구가 다른 구에 비해 두 배 이상의 위락 시설 면적을 가지고 있기 때문에 β_k 값을 추정하는데 큰 영향을 준다.

위에서 영향을 줄 수 있는 관측치를 살펴보았다. 위락시설(PARK)

에서 Dfbeta 값이 1.606으로 월등히 크고 다른 관측치에 비해 두 배 이상의 값을 갖는 송파구를 제외할 경우 선택되는 변수가 다를 것으로 예상된다. 이를 확인해 보기 위해 송파구를 제외한 후 STEP-WISE방법으로 변수 선택한 결과 인구(POPULATN), 시장 분포(MARKET), 복지 시설 수용 인원(WELFARE), 자기집의 비율(OWNHOUSE) 변수가 선택되었다. 이것은 송파구의 위락 시설(PARK)값이 상당한 영향을 준다는 것을 알 수 있다.

2.4 구별 특성에 따른 쓰레기 배출량 분석과 제안

서울 특별시 구별 쓰레기 배출량은 다음과 같은 회귀 모형으로 나타낼 수 있다.

$$\text{GARBAGE} = 0.291 \cdot \text{POPULATN} + 0.457 \cdot \text{RESTAURN} + 0.37 \cdot \text{PARK} - 0.343 \cdot \text{OWNHOUSE}$$

선택되지 않은 변수들은 서울 특별시 구별 쓰레기 배출량에 관계 없는 변수가 아니라 다른 변수들이 쓰레기 배출량에 대한 설명을 하고 난 후에 추가적인 설명을 하지 못하는 변수들임에 유의해야 할 것이다. 참고로 <표 8>은 선택된 변수들의 순위 자료이다.

추정된 회귀 모형을 보아 인구(POPULATN), 음식점 수(RESTAURN), 위락 시설(PARK)이 클수록 구별 쓰레기 배출량이 많이 나옴을 알 수 있으며, 자기집의 비율(OWNHOUSE)이 클수록 구별 쓰레기 배출량이 적게 나옴을 알 수 있다.

〈표 8〉

선택된 변수의 순위 자료

순위	구 명	GARBAGE	구 명	POPULATN	구 명	RESTAURN	구 명	PARK	구 명	OWNHOUSE
1	송 파	1413	도 봉	780472	강 남	10225	송 파	1927082	강 남	0.52581
2	성 동	1252	성 동	767976	성 동	7139	성 동	842755	서 초	0.51213
3	구 로	980	구 로	724230	송 파	6574	강 남	680583	노 원	0.48306
4	영등포	909	송 파	685260	구 로	6517	서 초	562764	은 평	0.43067
5	강 동	850	관 악	590442	서 초	6009	양 천	501997	송 파	0.43045
6	강 남	849	노 원	588188	중 로	5881	강 동	459615	양 천	0.41057
7	마 포	805	강 남	566379	영등포	5625	동 작	443077	도 봉	0.39802
8	동대문	785	강 동	535785	중 구	5202	노 원	360179	마 포	0.39664
9	중 구	769	성 북	514240	도 봉	5043	중 구	342918	서대문	0.39435
10	도 봉	763	은 평	512196	동대문	4442	중 로	202402	강 서	0.37692
11	강 서	665	양 천	499848	강 동	4186	도 봉	129588	강 동	0.37670
12	성 북	663	강 서	474434	관 악	4185	영등포	129543	성 북	0.36954
13	서 초	654	중 랑	460157	서대문	4029	관 악	123348	중 로	0.36599
14	관 악	636	동대문	459000	성 북	3744	구 로	121400	용 산	0.36224
15	중 로	557	영등포	445221	용 산	3700	강 서	82141	동 작	0.35687
16	서대문	552	동 작	440361	은 평	3511	동대문	81968	영등포	0.35203
17	노 원	528	마 포	422406	중 랑	3457	마 포	47064	동대문	0.34539
18	중 랑	519	서 초	410719	마 포	3307	서대문	44253	관 악	0.33643
19	양 천	489	서대문	381905	동 작	3157	은 평	42316	구 로	0.30659
20	동 작	462	용 산	276439	강 서	2895	중 랑	34083	중 랑	0.30000
21	용 산	461	중 로	222423	양 천	2735	성 북	26976	중 구	0.29447
22	은 평	460	중 구	167383	노 원	2632	용 산	15683	성 동	0.29398

인구가 많은 도봉구, 성동구, 구로구, 송파구는 주택 및 공업단지
로 인구가 밀집되어 있는 형태로 다량의 쓰레기가 배출됨을 알 수
있다. 인구의 증가는 쓰레기 배출량의 증가의 원인이 되고 그 밖에
도 여러 가지 원인이 있는데 이를 해결하기 위하여 인구의 분산 정
책과 지속적인 계몽 활동이 필요하다.

음식점 수(RESTAURN)는 구별 산업 형태 중 3차 산업의 형태
로 특히 서비스업의 특성을 갖는 변수이다. 〈표 8〉에서 음식점 수
(RESTAURN)가 많은 구인 강남구, 성동구, 송파구, 구로구, 서초
구, 종로구는 대체로 음식물 쓰레기 배출량이 많은 곳이라 할 수 있
다. 이러한 구들은 주문 식단체, 음식물 쓰레기의 분리 수거 등의 방

법을 통해 배출량을 줄이고 비료나 가축 사료로 재활용해서 음식물 쓰레기를 감소시킬 수 있으리라 생각된다.

송파구, 성동구, 강남구, 서초구는 이른바 강남 지역이라 부르는 곳으로 위락 시설이 밀집되어 있는 곳임을 알 수 있다. 이런 지역은 일회용품 쓰레기와 가연성 종이류가 대부분을 차지하므로 재활용에 노력을 기울여야 할 것이다. 또한 <표 3>에서 보면 위락시설(PARK)과 음식점 수(RESTAURN)와의 상관 계수가 0.463으로 위락 시설이 많이 분포되어 있는 곳에 음식점들이 많이 분포함을 알 수 있다.

성동구, 중구, 중랑구, 구로구는 자기집의 비율이 낮은 지역임을 알 수 있다. 이러한 지역은 주택 보급률이 낮은 지역으로 의식구조가 쓰레기 배출량에 영향을 끼침을 알 수 있다.

그러나 중랑구에서 쓰레기 배출량이 적은 까닭은 인구(POPULATN), 음식점 수(RESTAURN), 위락 시설(PARK)과 같이 쓰레기 배출량에 영향을 미치는 다른 변수들이 상대적으로 작기 때문인 것으로 분석된다. <표 8>에서 보면 종로구와 중구는 인구가 비슷하지만 쓰레기 배출량에 차이를 보이는 것은 인구(POPULATN), 음식점 수(RESTAURN), 위락 시설(PARK) 변수에는 별 차이가 없는데 자기집의 비율(OWNHOUSE)의 값에서 많은 차이를 보인다. 종로구와 중구간의 쓰레기 배출량에 차이가 나는 이유를 종로구가 주인 의식에 의한 의식의 차이 때문에 쓰레기 배출량이 중구보다 적다고 설명할 수 있겠다.

3. 결 론

지금까지 회귀 모형으로 쓰레기 배출량을 예측하고 회귀 모형에 영향력을 받는 구를 찾아 구별 특성을 구별짓고 그에 따른 쓰레기 배출량 분석을 하였다. 그리고 쓰레기 배출량이 많은 구와 적은 구를 선택된 변수들과의 관계를 고려하여 구별 특성에 맞게 분석하였다.

같은 특성을 갖는 구들은 서로 인접해 있음을 알 수 있는데 이러한 특징은 서울 특별시의 편향된 지역 개발에 따른 결과라 할 수 있다. 서울 특별시와 같은 거대 도시는 다핵화 되어야 도시 기능이 원활해질 수 있다. 그러나 강남 지역과 같은 일부 지역의 편향된 개발로 인하여 이러한 도시 기능이 무력화되어 가고 있다. 쓰레기 배출 또한 심각한 문제를 안고 있음을 알 수 있다. 도시 기능의 적정 배치는 거시적이고 종합적인 관점에서 해결 방안을 모색해야 될 것이다. 특히 음식점 수와 위락 시설이 각 구의 쓰레기 배출량에 영향을 끼치는 것을 알 수 있다. 이로부터 서울 특별시는 각 구의 산업 구조 형태 중 3차 산업 구조 형태에서 발생하는 쓰레기 배출량을 중점적으로 처리 및 관리하는데 역점을 두어야 하겠다. 쓰레기 배출량을 줄이기 위해서 1995년부터 본격적으로 실시되고 있는 쓰레기 종량제의 정착과 쓰레기 분리 수거의 제도적 보완이 있어야 겠다. 쓰레기의 적절한 재활용과 재활용 업체에 대한 지원책 마련 등을 통해 자원 재활용률을 높인다면 국가 경쟁력 향상에 도모할 수 있겠다. 그리고 쓰레기 문제는 다른 사람의 문제가 아닌 나 자신의 문제임을 일깨울 수 있는 지속적인 계몽 활동도 중요한 몫을 할 것이다.

본 논문에 쓰인 자료는 1994년 서울 특별시 통계 연보를 이용하여 분석하였다. 쓰레기 배출량을 산출하는 방법도 주먹구구식이 아닌 실질적이고 현실적인 통계량을 산출하여 정책에 올바르게 반영하는 것이 좋겠다. 또한 좀더 나은 분석을 위해 최근의 자료를 이용한다면 더 좋은 분석이 되었을 것이다. 본 논문에 사용된 변수들은 각각의 변수들과 상관 관계가 있어 간접적으로 쓰레기 배출량에 영향을 주었다고 생각한다. 따라서 경로 분석을 통하여 설명되지 않은 변수들을 추가로 분석할 수 있다. 이번 분석을 토대로 쓰레기 종량제가 시행되기 전인 94년과 시행된 후의 95년의 자료를 비교한다면 쓰레기 종량제의 효과를 알 수 있을 것이다.

4. 참고문헌

1. 서울 특별시 『서울 특별시 통계 연보』(1994).
2. 허명희, 이태림, 임성수 『통계적 자료 분석 I · SAS』(1994) - 방송 통신대학.
3. 허명희, 서혜선 『SAS 회귀 분석』(1994) - 자유 아카데미.
4. 김충련 『SAS라는 통계 상자』(1994) - 데이터 리서치.
5. John Neter, William Wasserman, Michael H. Kutner (1989) "Applied Linear Regression Model's" IRWIN, BOSTON.

고려대학교 학생들의 사회현상 및 시사에 대한 인지도 조사

최원혁 · 장우석 · 김영덕

(지도교수 : 박미라)

고려대학교 통계학과

〈요 약〉

현대의 사회는 정보의 홍수로 대변되는 빠른 변화를 거듭하고 있다. 이와 같은 변화 속에서 고려대학교 학생의 사회현상 및 시사정보에 대한 관심 및 인지정도를 가늠해 보고 이러한 정보수용의 태도에 성별, 소속단과대, 학년 등이 영향을 미치는지 알아보고자 한다.

본 조사가 대학인의 바람직한 정보수용 태도에 대해 숙고할 수 있는 계기를 제공할 수 있으면 한다.

1. 조 사 개 요

1.1. 조사목적

정보의 홍수로 대변되는 현재의 사회현상 및 시사에 대한 고려대학교 학생의 관심 및 인지정도를 알아보고, 이러한 정보수용의 태도나 결과가 성별이나 학년, 소속 단과대등의 변수별로 차이가 있는지 알아보고자 한다. 이와 더불어 사회현상을 정치, 경제, 문화, 과학, 스포츠의 5개 범주로 분류하여 이들 각각의 분야에 대해 학생들의 관심의 차이를 알아보며 정보수용 경로 및 이러한 시사지식의 필요성에 대하여 어떻게 생각하고 있는지 알아보고자 한다.

1.2. 조사설계

- (1) 모 집 단 : 고려대학교 재학생
- (2) 표본크기 : 209명
- (3) 표집방법 : 각 단과대별로 22~24명씩 층화추출
- (4) 조사기간 : 1996년 5월 6일~11일
- (5) 자료수집 : 자기 기입식 설문방식

1.3. 주요설문내용

〈표 1〉 주요설문내용

영역	설문내용	문항수
일반적 관심도	관심여부, 정보수용매체, 관심분야, 매체접근시간, 필요성 인지정도	5 문항
청	신문의 섹션을 기준으로 기사에 대한 인지 정도를 묻는 질문 각 6개항 기입식 질문 각 1 개항	7 문항
경		7 문항
과		7 문항
스		7 문항
포		7 문항
츠		7 문항
문		7 문항
화		7 문항
인 적 사 항	학년, 소속단과대, 성별, 출신지, 희망 직업	5 문항
계		45 문항

1.4. 조사대상 인적분포

표본으로 추출된 조사대상 학생들의 인적구성을 살펴보면 학년별 분포에 있어서 1, 2, 3, 4학년이 각각 54명(25.84%), 53명(25.36%), 45명(21.53%), 57명(27.27%)으로 거의 고른 분포를 나타내고 있다. 각 단과대에서는 22~24명 정도로 고르게 추출하였으며 성별구성은 남자가 143명으로 전체의 68.42%를 차지하며 여자가 66명으로 31.58%를 차지한다. 출신지 및 희망 직업별 분포는 아래

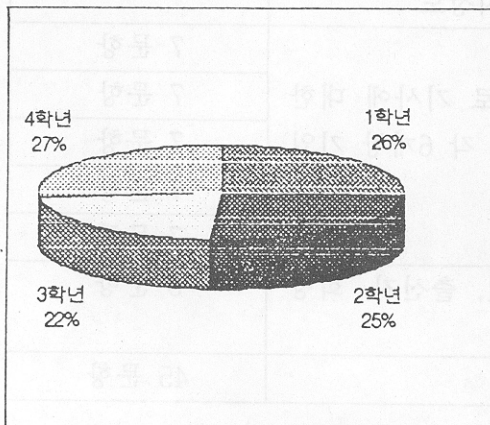
〈표 2〉와 〈표3〉에 나타나 있다.

〈표 2〉 출신지별 분포

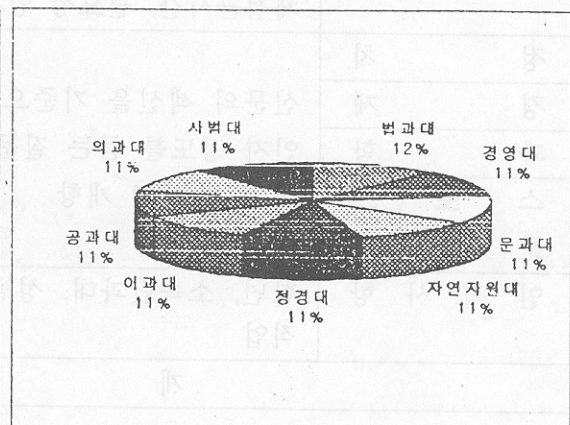
출신지	수도권	충청도	전라도	경상도	강원도	제주도	기타	계
인원수	134명	17명	16명	30명	6명	3명	3명	209명
(%)	64.11	8.13	7.66	14.35	2.87	1.44	1.44	100

〈표 3〉 희망 직업별 분포

희망직종	공무원	일반사무	기술사무	전문직	언론인	교육계	기타직	계
인원수	18명	22명	27명	74명	5명	34명	29명	209명
(%)	8.61	10.53	12.92	35.41	2.39	16.27	13.88	100



[그림 1] 학년별 분포



[그림 2] 소속 단과대별 분포

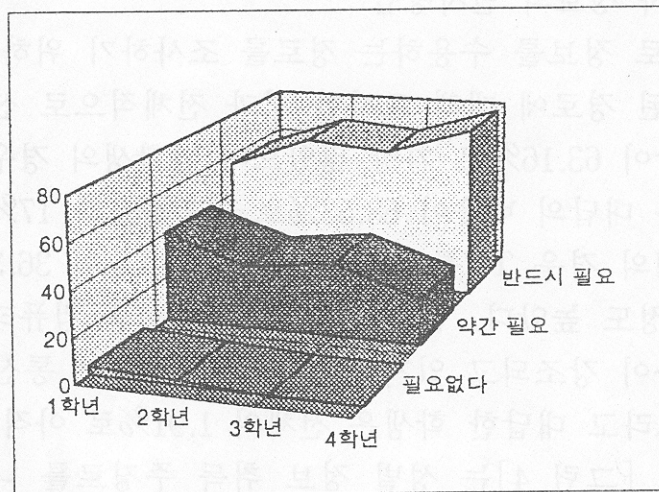
2. 조사 결과 분석

2.1. 일반 관심도에 대한 분석결과

(1) 사회현상, 시사정보의 필요성 및 수용태도

설문을 통해 각 개인이 느끼는 필요성을 알아보고자 하였다. 전체

적으로 98.56%의 학생들이 살아가는데 있어서 사회현상 및 시사정보에 대한 지식이 필요하다고 대답했으나 성별, 학년별 비교에서는 유의적인 차이를 보이고 있었다. [그림 3]을 보면 예상했던 결과대로 고학년으로 갈수록 반드시 필요하다는 반응이 많았고, 2학년과 3학년의 차이에 비하여 1학년과 4학년생의 필요성 인식은 확실히 차이를 보이고 있다. 특히 취업을 앞둔 4학년의 경우 약 79% 정도의 학생이 사회현상 및 시사정보가 살아가는데 반드시 필요하다고 답했다. <표 4>을 보면 성별 비교에 있어서 여학생 중 3명(4.55%)이 필요없다는 대답도 있었다는 점이 눈에 띈다. 4.55%인 점이 눈에 띈다.



[그림 3] 사회현상 및 시사에 대한 지식의 필요성

<표 4> 사회현상 및 시사에 대한 지식의 필요성 (성별비교)

p-값 : 0.085

구 분	전혀필요없다	별 필요없다	약간 필요	반드시 필요	계
남	0	0	44 (30.77%)	99 (69.23%)	143(68.42%)
여	1 (1.52%)	2 (3.03%)	20 (30.30%)	43 (65.15%)	66 (31.58%)
계	1 (0.48%)	2 (0.96%)	64 (30.62%)	142(67.94%)	209

‘사회현상, 시사에 관심이 있고 능동적으로 접근하는 편이십니까?’라는 질문을 통해 개인이 느끼는 필요성과 실질적인 태도가 같은 결과를 보이는지 알아보고자 했다. <표 5>에서 알 수 있듯이 필요성은 대다수가 인정하면서도 실제 자신에 대해 느끼는 정보 수용 태도는 미흡함을 보였다.

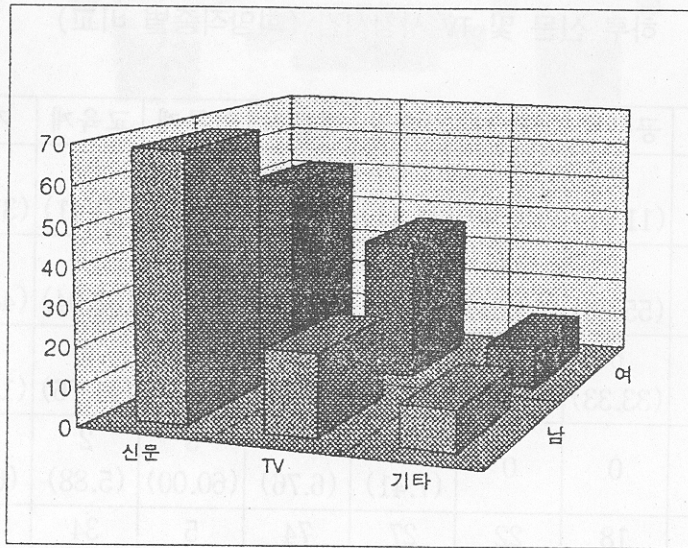
<표 5> 개인의 정보수용 태도

전혀 관심없다	조금 관심있다	관심있는 편이다	매우 관심있다	계
6 (2.87%)	95 (45.45%)	90 (43.06%)	18 (8.61%)	209 (100%)

(2) 시사 정보의 인지방법

구체적으로 정보를 수용하는 경로를 조사하기 위하여, 시사 정보를 얻는 주된 경로에 대해 조사한 결과 전체적으로 신문을 통해 얻는다는 대답이 63.16%로 가장 많았으며 남학생의 경우에 신문을 통해 얻는다는 대답의 비율이 68.53%로 여학생보다 17% 정도 높았고 반면 여학생의 경우 TV를 통해 얻는다는 비율이 36.36%로 남학생보다 16% 정도 높았다. 인터넷 정보사냥대회등 컴퓨터 통신을 통한 정보의 습득이 강조되고 있지만 인터넷등 컴퓨터 통신을 주된 정보 취득의 경로라고 대답한 학생은 전체의 1.91%로 아직까지는 미미한 수준이었다. [그림 4]는 성별 정보 취득 주경로를 나타낸 것이다. 신문과 TV를 제외한 다른 경로에 대해선 성별 차이가 거의 없어서 기타로 합치고 신문과 TV를 중심으로 그려보았다.

전체학생 중 뉴스시청 및 신문을 보는데 1시간 이하를 투자한다는 응답이 72.25%로 대다수였으며 4학년의 경우 2시간 이상 투자하는 비율이 12.28%로 타학년에 비해 높았다. 이 역시 필요성 인식정도와 맥락을 같이 한다고 볼 수 있다. 3학년도 1시간 이상 투자하는 비율이 1, 2학년보다 높고 30분 이하의 시간을 투자한다는 대답이 1, 2학년에 더욱 많아 대체로 고학년이 시사 및 사회현상에 관심이 많다는 것을 알 수 있다.



[그림 4] 시사정보를 얻는 주된 경로

〈표 7〉를 보면 장래 희망직종별 투자시간을 비교할 때 언론계, 기술직, 전문직 종사를 희망하는 학생들의 투자시간이 많았으며, 이에 반해 교육계, 관리직 사무원을 희망하는 학생들의 투자시간이 상대적으로 적었다.

〈표 6〉 하루 신문 및 TV 시청시간 (학년별 비교)

p-값 : 0.046

구 분	1 학년	2 학년	3 학년	4 학년	계
30 분 이 하	15명	15명	7명	7명	44(21.05%)
30분~1시간	24명	27명	23명	33명	107(51.20%)
1~2 시 간	12명	7명	15명	10명	44(21.05%)
2시 간 이상	3명	4명	0명	7명	14(6.70%)
계	54(25.84%)	53(25.36%)	45(21.53%)	57(27.27%)	209(100%)

〈표 7〉

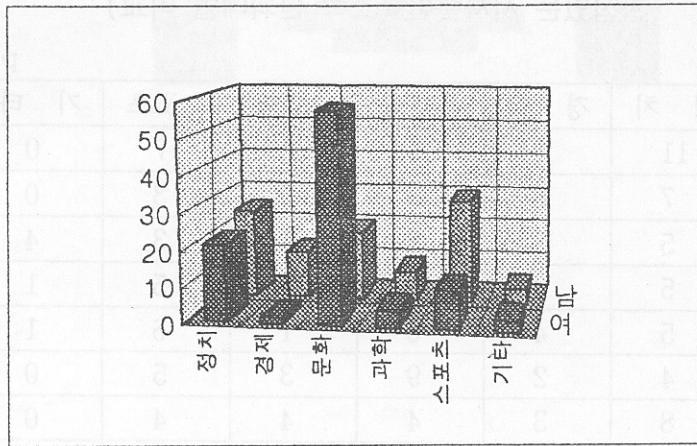
하루 신문 및 TV 시청시간 (희망직종별 비교)

p-값 : 0.003

구 분	공무원	관리직	기술직	전문직	언론계	교육계	기타	계
30 분 이 하 (%)	2 (11.11)	3 (13.64)	3 (11.11)	16 (21.62)	1 (20.00)	10 (29.41)	9 (31.03)	44 (21.05)
30분~1시간 (%)	10 (55.36)	14 (63.64)	12 (44.44)	38 (51.35)	1 (20.00)	18 (52.94)	14 (48.28)	107 (51.20)
1~2 시 간 (%)	6 (33.33)	5 (22.73)	10 (37.04)	15 (20.27)	0	4 (11.76)	4 (13.79)	44 (21.05)
2시 간 이상 (%)	0	0	2 (7.41)	5 (6.76)	3 (60.00)	2 (5.88)	2 (6.90)	14 (6.70)
계	18 (8.61)	22 (10.53)	27 (12.92)	74 (35.41)	5 (2.39)	34 (16.27)	29 (13.88)	209 (100)

(3) 관심있는 시사분야

가장 관심있는 시사분야를 조사한 결과, [그림 5]과 〈표 8〉에서와 같이 남녀별 그리고 소속단과대별로 관심분야가 다르다는 것을 알 수 있었다. 우선 가장 큰 차이가 나는 것이 문화면으로 여학생의 58.09%가 관심을 보인 반면 남학생은 19.58%만이 관심을 보였다. 스포츠와 경제는 반대로 남학생 관심 비율이 각각 29.37%, 13.29%이지만 여학생은 10.61%, 1.52%만이 관심을 보였다. 특히 경제 분야를 관심있다고 대답한 여학생은 66명중 단 1명이었다. 전체적으로 문화, 정치, 스포츠, 과학의 순으로 관심이 있었으나, 남학생의 경우는 스포츠, 정치, 문화, 경제, 과학 순의 관심을 보였으며 여학생은 문화, 정치, 스포츠, 과학, 경제 순의 관심을 보여 남학생과 대조를 이루었다.



[그림 5] 가장 관심있는 시사분야

법과대(45.83%) 학생들은 주로 정치분야에, 경영대(22.0%) 학생들은 주로 경제분야에 관심이 많았으며, 과학분야에 관심을 보인 것은 주로 자연계열 단과대 학생들이 많았다. 이는 단과대별 특성으로 사료되며, 사범대의 경우 문화 쪽에 가장 관심을 보이는 이유는 타 대학에 비해 사범대에 여학생들의 비율이 높은 까닭으로 생각된다.

특이한 것은 과학분야에 관심이 많을 것으로 예상되었던 의과대의 경우 과학에 관심이 있다고 대답한 사람은 아무도 없었으며 오히려 타단과대에 비해 스포츠, 문화 비중(86.36%)이 높은 것으로 나타났다. 이는 상대적으로 과중한 학사일정으로 인해 사회현상 및 시사에 대한 관심이 가벼운 주제 쪽으로 편중된 결과라고 생각된다.

〈표 8〉

관심있는 시사분야 (소속 단과대별 비교)

p-값 : 0.000

구 분	정 치	경 제	문 화	과 학	스포츠	기 타	계
법 과 대	11	1	5	0	7	0	24
경 영 대	7	6	8	0	3	0	24
문 과 대	5	3	8	1	3	4	24
자 대	5	0	6	6	5	1	23
정 경 대	5	4	6	1	6	1	23
이 과 대	4	2	9	3	5	0	23
공 과 대	8	3	4	4	4	0	23
의 과 대	2	1	8	0	11	0	22
사범대	3	0	13	0	5	2	23
계	50	20	67	15	49	8	209

2.2 최근 시사상식에 대한 인지도 조사결과

(1) 각 문항별 인지도 점수

각 문항에 대한 대답은 (전혀 모른다, 잘 모른다, 대충 안다, 잘 알고있다)의 4점 척도로 구성하였으며 각 문항의 평균값이 클수록 그 문항의 인지도가 높다는 것을 의미한다. 〈표 9〉의 인지도 점수를 보면 스포츠가 2.89로 가장 높았으며, 정치(2.597), 문화(2.535), 경제(2.102) 순이었고 과학은 1.908로 가장 낮은 점수를 보였다. 30개의 문항중 가장 인지도가 높은 문항은 ‘박찬호, 선동열의 현지 활약상에 대해 아는가’를 묻은 스포츠 섹션의 Q7이며, 가장 인지도가 낮은 문항은 ‘마이크로 머신의 활용에 및 활용방안에 대해 아는가’를 묻은 과학 섹션의 Q13이었다.

각 문항의 평균값을 2차 입력자료로 이용하여 5개 섹션의 평균이 같은지에 대한 다중비교를 실시한 결과는 〈표 10〉과 같다. TUKEY의 다중비교에 의하면 5% 유의수준에서 정치, 문화, 경제간에는 평균의 차이가 유의하지 않았으나, 스포츠와 과학은 뚜렷한 차이를 보이고 있다.

〈표 9〉

전 문항 기초통계

구 분	문 항	문항 평균	표준 편차	섹션 평균
정 치	Q 4 : 4.11총선결과 각 정당의 확보의석 수	3.00	0.72	2.60
	Q 8 : 4.11총선당시 각 정당 선대위 대변인의 이름	2.65	0.95	
	Q10 : 미국 공화당 대통령선거 후보지명전의 추이 및 결과	2.37	0.91	
	Q17 : 12.12 및 5.18 관련 법정 피고인의 이름	3.04	0.69	
	Q19 : 한미행정협정(SOFA)의 불평등 논란조약 내용	2.33	0.98	
	Q26 : 4.16 한미 제주도 정상회담시 제안된 한반도평 화 보장안의 내용	2.19	0.88	
경 제	Q 1 : PCS 사업권을 둘러싼 재계의 컨소시엄 구성현황	2.31	0.88	2.10
	Q 5 : '96년도 대규모 기업군 신규지정 및 제외업체	2.04	0.79	
	Q14 : 경제신문을 읽는 지의 여부	1.86	0.84	
	Q21 : M&A, R&D 등의 경제용어	2.06	1.10	
	Q22 : 조기퇴직제, 명예퇴직제 도입	2.46	0.94	
	Q27 : 증권시장 안정화기금의 해체	1.88	0.85	
문 화	Q 6 : 최근의 가요 인기순위	2.33	0.82	2.54
	Q15 : 싱가포르의 바이올리니스트 '바네사 메이'	2.77	0.96	
	Q16 : 아카데미 영화상의 시상내용	2.42	0.83	
	Q23 : R&B, 힙합, 레이브, 소울 등 음악장르에 속한 곡	2.69	0.78	
	Q25 : 각종 전시장, 공연장 이용 여부	2.21	0.72	
	Q29 : 현재 상영, 공연 중인 영화나 연극	2.79	0.65	
과 학	Q 3 : CDMA와 TDMA방식의 차이	1.70	0.88	1.91
	Q 9 : 인터넷 관련용어-Telnet, FTP, HTTP, WWW 등	2.16	1.02	
	Q11 : 멀티미디어 관련용어-PDA, VOD 등	1.73	0.80	
	Q13 : 마이크로머신의 활용예 및 활용방안	1.64	0.75	
	Q14 : INTRANET에 대하여	1.86	0.84	
	Q30 : 윈도우 95와 그 이전 버전의 차이점	2.44	0.97	
스�포츠	Q 2 : 애틀란타올림픽 남자마라톤 대표 3인	2.45	0.98	2.89
	Q 7 : 박찬호, 선동열의 현지 활약상	3.34	0.68	
	Q12 : 해외스포츠-NBA, PGA, MBL 등	2.59	0.85	
	Q18 : 농구 코리아리그의 신생팀	3.17	0.99	
	Q24 : 2002 월드컵 개최지 결정기구, 시기	3.25	0.74	
	Q28 : 애틀란타 올림픽 결선진출이 확정된 구기종목	2.55	0.93	
계	30 문항	2.41	0.45	

〈표 10〉

색선별 다중비교

Tukey Grouping		Mean	N	색선
	A	2.892	6	스포츠
B	A	2.597	6	정치
B	A	2.535	6	문화
B	C	2.102	6	경제
	C	1.908	6	과학

(2) 색선별 문항의 신뢰도

〈표 11〉에 제시된 것은 각 색선별로 6개 문항에 대한 신뢰계수(Cronbach's Alpha)를 나타낸다. 정치, 경제, 스포츠, 과학 색선은 그 신뢰도가 모두 80%를 상회하지만 문화 색선의 문항은 약 66%의 신뢰도를 보이고 있다. 이는 25번 문항의 경우 다른 문항이 대중예술(연예분야)의 인지도에 관한 질문인데 비해 순수예술에 연관된 능동적 참여여부를 묻는 질문이기에 상관관계가 낮은 것으로 판단된다.

〈표 11〉

신뢰계수 (Cronbach's Alpha)

구 분	정 치	경 제	과 학	스 포 츠	문 화
신뢰계수	0.811	0.825	0.855	0.829	0.655

(3) 각 분류변수에 따른 인지도 점수

각 개인의 인적사항에 관련된 변수 즉, 학년, 소속단과대, 성별, 출신지, 희망직종 등이 개인의 인지도에 미치는 영향을 알아보기 위해 분산분석을 실시한 결과는 〈표 12〉와 같다. 즉, 학년, 성별, 소속단과대가 인지도에 영향을 미치는 것으로 나타났으며 출신지, 희망직종은 별 영향이 없는 것으로 나타났다.

학년간의 다중비교 결과 2, 3학년 간에는 별 차이가 없었으며 1학년과 4학년 사이에는 인지도의 차이가 유의수준 5%에서 인정됐다.

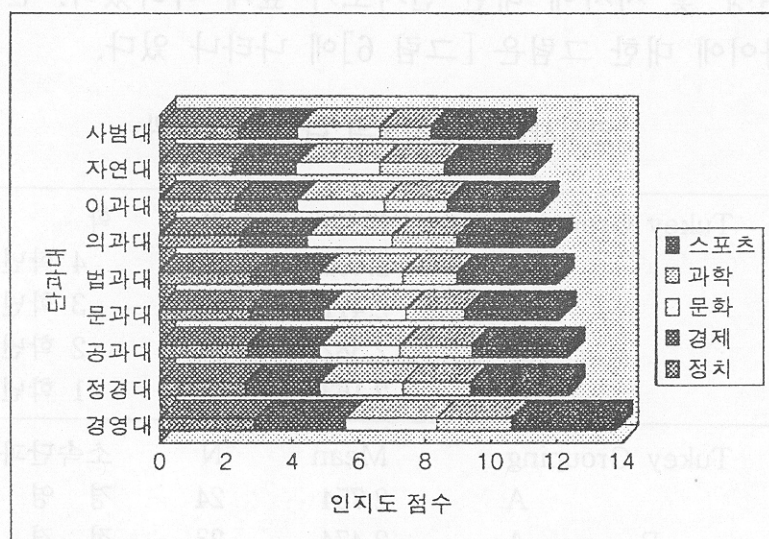
4학년의 인지도 평균이 약 2.66으로 1학년의 2.16 보다 더 높아 고학년일수록 사회현상 및 시사정보에 대한 관심 뿐 아니라 인지도도 높다는 것을 알 수 있다. 성별비교에 있어서는 남학생의 경우 약 2.51의 인지도로 여학생의 2.19 보다 높게 나타나 남학생이 여학생보다 사회현상 및 시사에 대한 인지도가 높게 나타났다. 단과대별 인지도의 차이에 대한 그림은 [그림 6]에 나타나 있다.

〈표 12〉 인적사항별 인지도의 다중비교 결과

Tukey Grouping		Mean	N	학 년
	A	2.6591	57	4 학년
	B	2.4111	45	3 학년
	B	2.3824	53	2 학년
	C	2.1636	54	1 학년
Tukey Grouping		Mean	N	소속단과대
	A	2.754	24	경 영 대
B	A	2.474	23	정 경 대
B	A	2.467	23	공 과 대
B	A	2.412	24	문 과 대
B	A	2.406	24	법 과 대
B		2.392	22	의 과 대
B		2.313	23	이 과 대
B		2.262	23	자연자원대
B		2.171	23	사 범 대
Tukey Grouping		Mean	N	성 별
	A	2.5079	143	남
	B	2.1899	66	여

단과대별 섹션 인지도를 비교한 결과 거의 모든 단과대에 있어서 스포츠, 문화, 정치의 인지도가 높았으며, 경제, 과학의 인지도가 낮게 나타났다. 경영대가 가장 인지도가 높았으며, 사범대가 가장 낮은 인지도를 보였다. 법과대와 의과대의 경우는 전반적 인지도는 중간

수준인데 비해 문화와 스포츠 섹션에 대한 인지도는 상대적으로 높은 수준을 나타내고 있는데, 이는 전체적으로 앞의 관심분야에 대한 설문조사결과와 거의 비슷한 경향을 보여 관심의 수준이 어느정도 인지수준과 관계가 있다고 생각된다.



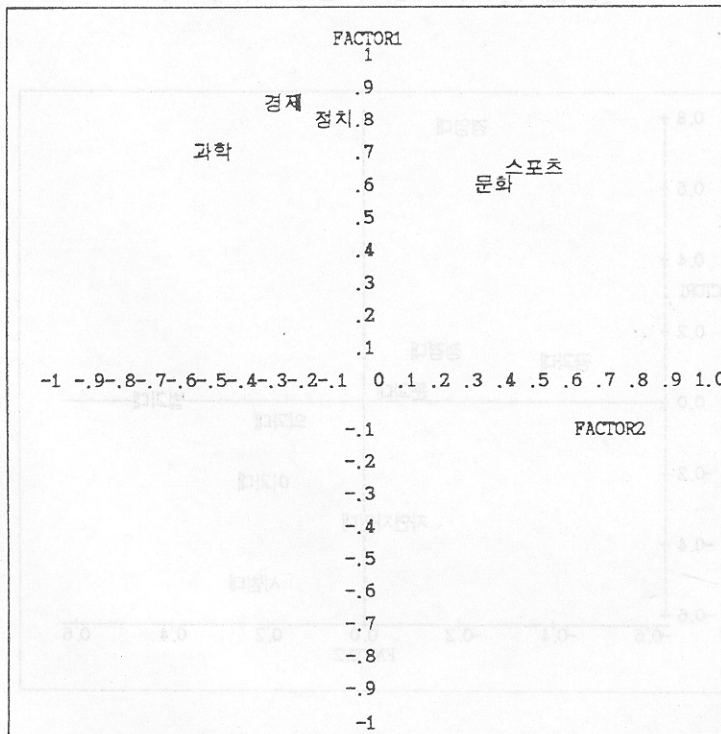
[그림 6] 단과대별 섹션 인지도

(4) 행렬도(biplot)를 이용한 분석

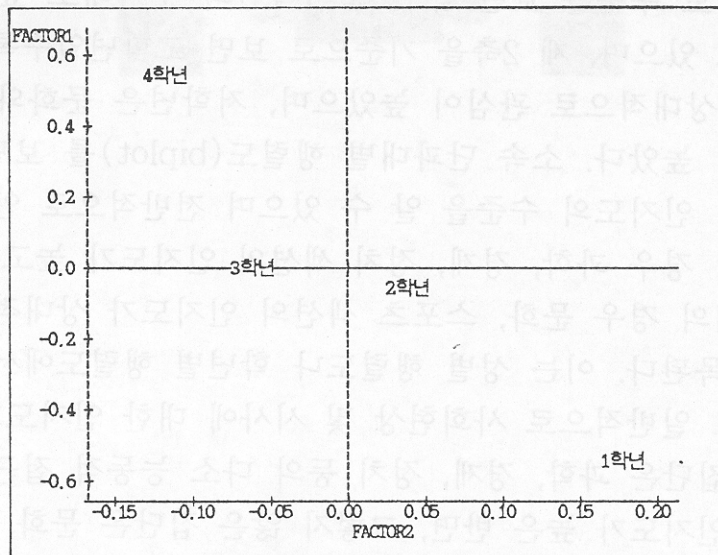
각 섹션들과 개체들간의 관계를 그림으로 표현하기 위해 행렬도(biplot)를 이용하여 그림을 그려보았다. 여기서는 인자분석행렬(factor analysis biplot)을 이용한 그림으로, [그림 7]은 변수, 즉 문항의 플롯이다. 제 1축은 모든 문항들이 양의 축에 있어서 인지도의 전반적인 수준을 의미한다고 할 수 있다. 반면, 제 2축은 스포츠와 문화가 양의 축에, 과학, 경제, 정치가 음의 축에 위치하여 스포츠, 문화와 과학, 경제, 정치의 대비를 나타낸다고 할 수 있겠다. 여기서 얻어진 각 개체의 인자점수를 이용하여 각 변수의 행렬도를 그린 결과가 [그림 8], [그림 9], [그림 10]에 나타나 있다.

학년별 행렬도(biplot)에서 보는 바와 같이 제 1축을 기준으로한

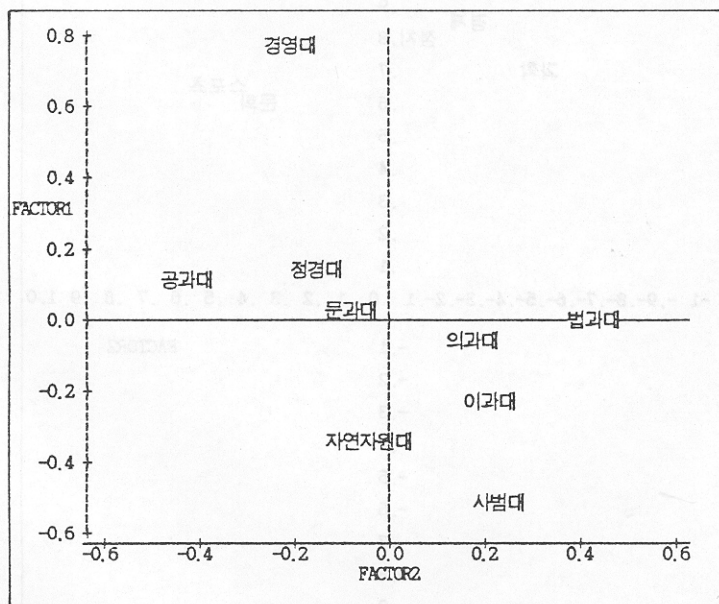
전반적 인지도 수준을 보면 4, 3, 2, 1학년의 순서대로 높은 인지수준을 보이고 있으며, 제 2축을 기준으로 보면 고학년일수록 과학, 경제, 정치에 상대적으로 관심이 높았으며, 저학년은 문화와 스포츠에 대한 관심이 높았다. 소속 단과대별 행렬도(biplot)를 보면 각 단과대의 전반적 인지도의 수준을 알 수 있으며 전반적으로 인지도가 높은 단과대의 경우 과학, 경제, 정치 섹션의 인지도가 높고, 인지도가 낮은 단과대의 경우 문화, 스포츠 섹션의 인지도가 상대적으로 높다는 점이 주목된다. 이는 성별 행렬도나 학년별 행렬도에서도 나타나는 경향으로 일반적으로 사회현상 및 시사에 대한 인지도가 높은 특성을 가진 집단은 과학, 경제, 정치 등의 다소 능동적 접근을 요구하는 섹션의 인지도가 높은 반면, 그렇지 않은 집단은 문화, 스포츠 등 쉽게 접할 수 있는 섹션의 인지도가 높았다.



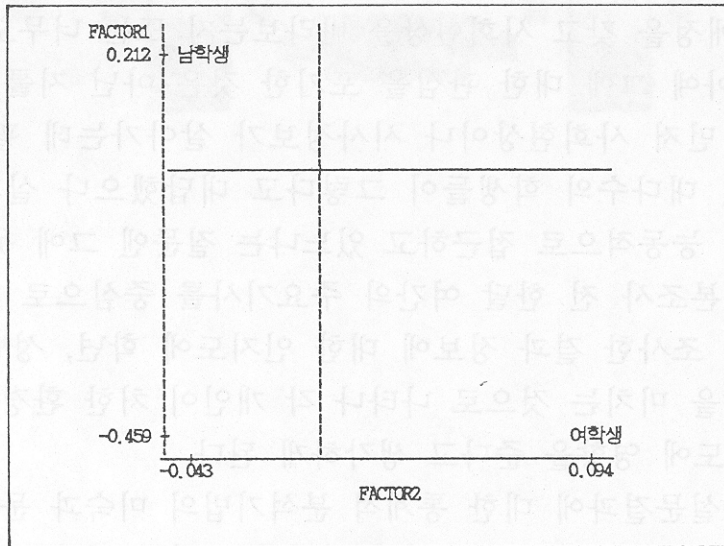
[그림 7] 각 섹션의 행렬도(biplot)



[그림 8] 각 학년의 행렬도(biplot)



[그림 9] 각 단과대의 행렬도(biplot)



[그림 10] 성별의 행렬도

3. 결 론

대학인은 최고의 학부생으로 곧 사회에 나아가 여러 분야에서 사회의 일원으로 생활해야 할 시기를 앞두고 있다. 그런데 사회는 지금 엄청난 속도로 변해 가고 있다. 정보의 홍수로 표현되는 지금의 사회의 발전속도는 너무나도 빨라서 개인이 그 정보를 모두 체득하기란 불가능할 지경에 이르렀다. 현대 SF의 대가 아이작 아시모프는 이러한 엄청난 변화에 대해 일반인들이 가져야 할 태도로 숲을 바라보는 자세를 권했다. 즉, 숲에 들어가 과일 하나하나를 살피면서는 길을 잃게 될지언정 필요로 하는 과일을 딸 수가 없다. 그러므로, 좀 떨어져서 숲을 바라봐 전체 숲의 모양이나 나무의 위치를 파악하는 태도가 필요하다는 것이다. 지금의 대학인에게 필요한 말일듯 싶다.

지금의 대학인(특히 고려대학교 학생들)이 이런 엄청난 정보의 홍수 속에서 과연 숲을 바라보는 자세로 관심있게 자신들이 살아갈 사

회에 대해 애정을 갖고 사회현상을 바라보는지 또는 너무도 많은 정보에 질려 아예 그에 대한 관심을 포기한 것은 아닌 지를 알아보기로 하였다. 먼저 사회현상이나 시사정보가 살아가는데 필요한가를 묻는 질문엔 대다수의 학생들이 그렇다고 대답했으나 실제 자신이 관심을 갖고 능동적으로 접근하고 있느냐는 질문엔 그에 못 미친 응답을 했다. 본조사 전 한달 여간의 주요기사를 중심으로 시사에 대한 인지도를 조사한 결과 정보에 대한 인지도에 학년, 성별, 소속단과대가 영향을 미치는 것으로 나타나 각 개인이 처한 환경이 개인의 정보수용 태도에 영향을 준다고 생각하게 된다.

본조사의 설문결과에 대한 통계적 분석기법의 미숙과 문항의 정도나 타당성에 대한 고려가 필요하다 하더라도 위의 결과를 통해 고대생의 정보수용 태도에 대한 개략적인 이해가 가능하리라 본다. 끝으로 21세기의 주역이 될 대학인의 폭넓은 사회현상에 대한 관심이 필요함을 밝히고자 한다.

〈참고문헌〉

- 허명희(1993), 통계상담의 이해, 자유 아카데미
- 허명희(1993), 통계조사 방법론, 자유 아카데미
- 허명희(1988), SAS 분산분석, 자유 아카데미
- 오인환(1992), 사회조사 방법론, 나남 출판
- 송문섭 외 3인(1993), SAS를 이용한 통계자료 분석, 자유 아카데미
- 김충련(1994), SAS라는 통계상자, 데이터 리서치
- 중앙일보, 경향신문, 문화일보(1996. 4. 1~1996. 4. 30)

부록 1> 설문지 양식

♣ 다음은 사회현상 및 시사(時事)에 대한 관심을 알아보는 설문입니다.

1. 사회현상, 시사에 관심이 있고 능동적으로 접근하는 편이십니까?
① 전혀 관심없다 ② 조금 관심있다 ③ 관심있는 편이다 ④ 매우 관심있다
2. 시사 정보는 주로 무엇을 통해 얻고 있습니까?
① 신문 ② TV ③ 기타 인쇄매체(책, 잡지 등) ④ 컴퓨터 통신(인터넷 등)
⑤ 그밖의 경로()
3. 가장 관심있는 시사 분야는?
① 정치 ② 경제 ③ 문화 ④ 과학 ⑤ 스포츠 ⑥ 기타()
4. 하루에 신문을 읽거나 TV뉴스를 시청하는 시간은 얼마입니까?
① 30분 이하 ② 30분~1시간 ③ 1시간~2시간 ④ 2시간 이상
5. 살아가는데 있어서 사회현상이나 시사정보에 대한 지식이 필요하다고 생각하십니까?
① 전혀 필요없다 ② 별 필요없다 ③ 약간 필요하다 ④ 반드시 필요하다

♣ 다음은 최근의 사회현상 및 시사에 대한 인지도를 알아보는 설문입니다.

1. 정부의 차세대 개인휴대통신(PCS) 사업권자 선정을 둘러싼 재계의 컨소시엄 구성 현황을 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
2. 지난 동아마라톤을 통해 아틀랜타 올림픽 출전자격을 획득한 선수들의 이름을 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
3. 코드분할다중접속방식(CDMA)과 시분할다중접속방식(TDMA)의 차이를 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
4. 4.11 총선 결과 각 정당이 확보한 국회의석 수를 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다

5. 공정거래위원회가 4월 12일 지정한 '96년도 대규모 기업군'의 30대 재벌에 신규 지정된 기업이나 제외된 기업을 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
6. 최근의 가요 인기순위를 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
7. 박찬호, 선동열등 해외로 진출한 우리나라 선수의 현지 활약상을 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
8. 4.11 총선 당시 각 정당의 선거대책위원회 대변인 이름을 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
9. 인터넷 관련용어(telnet, ftp, http, gopher, tcp/ip, www 등)를 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
10. 미국 공화당의 올 대통령선거 후보지명전의 추이 및 결과를 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
11. 멀티미디어 관련용어(PDA, VOD 등)을 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
12. 해외 스포츠(NBA, NFL, NHL, PGA GOLF, TENNIS 등)에 대한 기사 및 뉴스를 자주 접하십니까?
① 아니다 ② 아닌 편이다 ③ 그런 편이다 ④ 그렇다
13. 마이크로 머신(micro machine)의 활용에 및 활용방안에 대해 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
14. 매일, 한국, 서울경제신문 등 경제 신문을 자주 읽으십니까?
① 아니다 ② 아닌 편이다 ③ 그런 편이다 ④ 그렇다
15. 싱가포르 출신의 바이올리니스트 '바네사 메이'에 대해 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다

16. 지난 아카데미 영화상에서 남·여주연, 남·여조연, 작품상, 감독상, 각색상 등의 주요부문 수상작을 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
17. 12.12 및 5.18 관련 법정 비고인의 이름을 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
18. 코리안 리그 농구대회에 올해 처음 뛰고 있는 신생팀을 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
19. 한미행정협정(SOFA)의 불평등 논란이 되고 있는 조항에 대해 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
20. 인트라 넷(INTRANET)에 대해 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
21. M&A, R&D 등의 경제 용어를 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
22. 근간에 국내기업들에 추진되려 하고 있는 조기퇴직제나 명예퇴직제에 대해 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
23. 리듬 앤 블루스, 힙합, 얼터너티브, 슬래쉬, 레ιβ, 스캇, 소울, 갱스터랩 등의 음악 장르에 속하는 곡을 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
24. 2002년 월드컵 개최지가 결정되는 시기 및 결정기구를 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
25. 각종 공연장 및 전시관을 자주 이용하십니까?
① 아니다 ② 아닌 편이다 ③ 그런 편이다 ④ 그렇다
26. 4월 16일 있었던 한·미 제주도 정상회담에서 제안된 한반도 평화보장안의 내용에 대하여 알고 계십니까?
① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다

27. 증권시장 안정화 기금의 해체 예정 사실을 알고 계십니까?
 ① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
28. 구기종목중 아틀란트 올림픽 결선 진출이 확정된 종목에 대해 알고 계십니까?
 ① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
29. 현재 상영 또는 공연중인 영화나 연극들을 알고 계십니까?
 ① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다
30. 마이크로 소프트사의 '윈도우 95'가 이전의 윈도우와 다른 점에 대하여 알고 계십니까?
 ① 전혀 모른다 ② 잘 모른다 ③ 대충 안다 ④ 잘 알고 있다

♣ 다음은 기입식 질문입니다.

1. 15대 국회 의석 확보수 순으로 각 정당 및 무소속을 나열해 주십시오.
 ① 자민련 ② 새정치국민회의 ③ 무소속 ④ 신한국당 ⑤ 민주당
2. PCS 사업권을 둘러싼 동신장비 제조업체군의 경쟁구도를 적어 주십시오.
 보기 > 현대, 대우, 삼성, LG
 ()과 ()의 연합 대 ()
3. 선동렬, 박찬호의 소속팀을 적어 주십시오.
 (,)
4. 세계 최초로 국내에서 상용화에 성공한 휴대통신 기술방식을 적어 주십시오.
 ① TDMA ② CDMA ③ TRS ④ PCS ⑤ PDS
5. 가요순위 10위권 내의 가요를 3개 이상 적어 주십시오.
 ()

♣ 마지막으로 인적사항을 묻는 질문입니다.

1. 당신은 몇 학년 이십니까?
 ① 1학년 ② 2학년 ③ 3학년 ④ 4학년

2. 당신의 소속 단과대학은?

- ① 법과대 ② 경영대 ③ 문과대 ④ 자연자원대 ⑤ 정경대
⑥ 이과대 ⑦ 공과대 ⑧ 의과대 ⑨ 사범대

3. 당신의 성별은? ① 남자 ② 여자

4. 당신의 출신고교 소재지는?

- ① 서울, 경인의 수도권 ② 충청도 ③ 전라도 ④ 경상도
⑤ 강원도 ⑥ 제주도 ⑦ 해외 등 기타()

5. 당신의 장래 희망 직업군은?

- ① 공무원 ② 일반직 사무원 ③ 기술직 사무원(Engineer, Programmer 등)
④ 전문자격 종사자(고시 및 자격시험 준비) ⑤ 언론계
⑥ 교육계(학계) ⑦ 기타 직업()

설문에 응해 주셔서 감사합니다.

밖에서 본 5·18*

김재춘, 박장호

(지도교수 : 정영해)

동신대학교 통계학과

〈요 약〉

이 연구는 1996년 5월에 조사한 자료를 이용하여 호남지역 밖에서 보는 5·18에 대한 인식과 5·18 진상 규명 정도에 대한 견해를 분석한 것이다. 각 변수와 사회인구학적 요인의 관계를 카이제곱검정과 로짓을 이용한 모형(Logistic regression model)으로 분석하였다.

응답자의 대부분은 5·18을 민주화운동 또는 자기방어적 투쟁이라고 생각하고 있다. 로짓모형에 의하면, 젊은 층(20대~30대)에서는 학력에 따라 5·18의 인식에 별 차이가 없었으나, 장년층에서는 학력이 낮은 사람일수록 폭동이나 체제전복이었다고 생각하는 비율이 더 증가하였다. 학력에 따른 차이는 연령이 높아질수록 심화되었다. 5·18에 대한 인식은 지역에 따른 차이는 없는 것으로 나타났다.

5·18에 대한 진상 규명 정도에 대하여는 남자보다는 여자가, 연령이 높아질수록 더 밝혀졌다고 생각하는 것으로 나타났다. 부모가 한 명이라도 호남출신인 사람은 그렇지 않은 사람보다 밝혀지지 않았다고 생각하는 사람이 더 많은 것으로 나타났다. 대구·경북과 부산·경남에 거주하는 사람들은 다른 지역에 비하여 밝혀졌다고 생각하는 사람이 더 많다.

부산·경남, 대구·경북 응답자들은 5·18을 폭동 또는 체제전복을 위한 민중혁명이 아니라고 하는데 있어서는 다른 지역의 사람들과 의견을 같이하면서도, 진상규명 정도에 있어서는 다른 지역보다 “밝혀졌다”는 견해를 더 나타내, 지역적 차이를 보였다.

* 이 연구는 사단법인 광주사회조사연구소의 일부 지원(KSRC-KJC / PJH-P9607-26)으로 이루어졌다.

자료를 이용할 수 있게 허락하여 주신 광주방송과 사단법인 광주사회조사연구소에 감사드린다. 결과의 사회학적인 해석에 도움말을 주신 광주대학교 김순홍교수께 감사드린다.

1. 서 론

정부의 발표에 따라 1980년 사건 당시부터 많은 사람들에게 폭동으로 인식되어 오던 5·18이 제6공화국에서는 민주화운동으로 규정되고, 문민정부 들어서는 그 진상을 규명하기 위한 노력이 보다 활발히 이루어지고 있다. 특히 ‘모래시계’, ‘제4공화국’, ‘코리아게이트’ 등 텔레비전 프로그램의 방영은 전 국민에게 5·18의 실상을 알리는 계기가 되었다. 이 시점에서 호남지역 밖에 거주하는 국민들의 5·18에 대한 인식과 진상규명 정도에 대한 견해가 거주지역과 연령, 교육정도, 성별등의 사회인구학적 요인에 따라 차이가 있는지 알아보는 것은 의미있는 일이다.

1.1. 조사방법

조사대상은 광주·전남과 제주를 제외한 전국에 거주하는 20세 이상 60세미만의 성인 남·녀로서 총 800명을 조사하였다. 이 가운데 전북거주자(37명)는 광주·전남과 의식 성향이 비슷할 것으로 생각하고 분석에 포함시키지 않아서 분석의 대상은 총 763명이다.

표본의 선정은 비례할당 추출(quota sampling)로 하였으며, 그 방법은 다음과 같다. 전국을 서울·경기와 충청·강원, 부산·경남, 대구·경북의 네 개 지역으로 나누어¹⁾, 각 시·도별 1995년 통계연보에 근거하여 인구비율에 따라 표본 크기를 할당한 뒤, 각 시·도에서 다시 성별, 연령별 인구의 구성에 따라 표본크기를 할당하였다²⁾. 할당된 지역에서 전화번호를 이용한 표본추출을 하였는데, 그 지역에서 사용하는 국번 중에서 컴퓨터를 이용하여 무작위로 국번을 선정하고, 나머지 전화번호는 random digit dialing을 이용하였다. 조사는 전화 조사(telephone survey)로 이루어 졌다.

1) 의식 성향이 비슷하다고 생각되는 지역끼리 묶었다.

2) 조사된 사람의 수가 할당된 수와 한두명 차이가 나는 경우, 지역별 인구 비율과 전체 표본의 크기가 정해진 수와 같아지도록 적절한 가중치를 주어 조절하였다.

1.2. 조사내용과 변수의 정의

사회인구학적 변수와 5·18에 대한 인식과 진상규명 정도에 대한 견해를 조사하였다. 5·18에 대한 인식과 진상규명 정도에 대한 견해를 알아보기 위한 질문은 다음과 같다.

1. 5·18은 다음중 어느 것이라고 생각하십니까?

- ① 불순불자들의 사주에 의한 폭동이었다.
- ② 신군부의 집권에 반대하는 민주화운동이었다.
- ③ 체제 전복을 위한 민중 혁명이었다.
- ④ 살아남기 위한 자기 방어적 투쟁이었다.

기타(구체적으로 말씀해 주세요).

(88) 관심없다.

(99) 모르겠다.

2. ○○남께서는 현재까지 5·18의 진상이 어느정도 밝혀졌다고 보십니까?

- ① 모두 밝혀졌다.
- ② 대부분 밝혀졌다.
- ③ 아직까지는 미흡하다.
- ④ 밝혀진 것은 하나도 없다.

(8) 관심 없다.

(9) 모르겠다.

분석을 위하여 종속변수는 의미가 비슷한 범주끼리 묶었다. 5·18에 대한 인식의 응답항목 중 “① 불순불자들의 사주에 의한 폭동이었다”와 “③ 체제 전복을 위한 민중 혁명이었다”를 ‘폭동 및 체제전복’으로, “② 신군부의 집권에 반대하는 민주화운동이었다”와 “④ 살아남기 위한 자기 방어적 투쟁이었다”를 ‘민주화운동 및 자기방어’로 재분류하였다. 또한 5·18 진상규명 정도에 대한 견해는 “① 모두 밝혀졌다”와 “② 대부분 밝혀졌다”를 ‘밝혀졌다’로, “③ 아직까지는 미흡하다”와 “④ 밝혀진 것은 하나도 없다”는 ‘밝혀지지 않았다’로 재분류하였다.

사회인구학적 변수는 다음과 같다.

성별(SEX) : 0=여, 1=남.

연령(AGEG) : 0=20대, 1=30대, 2=40대, 3=50대.

교육정도(EDU) : 0=무학 /국졸, 1=중졸, 2=고졸, 3=대학 이상.

거주지(REGION) : 0=서울·경기, 1=충청·강원,

2=부산·경남, 3=대구·경북.

부모 출신연고(HOME) : 0=모두 호남이 아님,

1=한 명이라도 호남임.

연령은 10세 단위로 묶어 교육과 함께 순서식 변수(ordinal variable)로 분석하였다.

1.3. 분석방법

이원분할표의 카이제곱분석과 로짓모형분석을 실시하였다. 로짓모형은, 5·18에 대한 인식은 ‘폭동 및 체제전복을 위한 민중혁명’의 비율을, 진상규명에 대해서는 ‘밝혀지지 않았다’의 비율을 로짓으로 만들었다. 각 변수에 응답하지 않았거나 주어진 범주 이외의 응답(모르겠다, 관심없다 등)을 한 사람들은 분석에서 제외하였다. 로짓모형에 포함시키는 변수의 갯수는 Stokes, Davis and Koch(1995 ; p191)의 지침을 참고하였다. 즉, 응답자 수가 적은 쪽의 응답자 수를 5로 나눈 것을 최대 가능한 갯수로 결정하였다. 응답자 수가 많은 경우에는 조금더 엄격한 지침에 따라 응답자의 수를 10으로 나눈 것을 최대 가능한 갯수로 정하였다.

로짓모형은 로짓(또는 로그 오드 ; log odd)이나 승비(odds ratio)로 해석하게 되는데, 이를 설명하기와 이해하기는 모두 어렵다. 여기서는 설명을 쉽게하고 이해를 돕기 위하여 비율(백분율)로 환산하여 해석하기로 한다. 로짓모형에서는 선형의 관계가 있는 변수들도 비율에 있어서는 비선형의 관계가 있는 것으로 해석되게 된다³⁾.

유의수준은 0.05를 이용하였다.

3) 관심이 있는 범주에 응답한 사람의 비율을 p 라고 할 때, 로짓모형은 $\log \frac{p}{1-p} = Xb$ 이다. 이를 비율에 대한 모형으로 바꾸면 $P = \frac{e^{Xb}}{1+e^{Xb}}$ 가 된다.

2. 결 과

응답자의 사회인구학적 특성은 <표 1>과 같다. 성별과 연령, 거주지는 할당에 사용한 변수이기 때문에 그 분포는 할당된 분포와 같다. 교육정도에 있어서는 응답자의 약 77.0%가 고졸이상의 학력을 보였다. 부모의 출신연고에 있어서는 응답자의 81.0%가 호남과는 관련이 없었다.

<표 2>는 응답자들의 5·18에 대한 인식과 진상규명 정도에 대한 견해이다. 5·18이 민주화운동 또는 자기방어적 투쟁이었다고 응답한 사람은 전체의 77.5%(민주화운동 62.1%; 자기방어적 투쟁 15.4%)이다. 나머지는 불순분자에 의한 폭동(7.0%) 및 체제전복을 위한 민중혁명(8.8%)이라고 생각하는 것으로 나타났다.

진상규명 정도에 대해서는 64.3%가 아직 미흡하거나 전혀 밝혀지지 않았다고 생각하고 있었다. 밝혀졌다고 생각하는 사람은 모두 33.3% (모두 밝혀졌다 2.1%; 대부분 밝혀졌다 31.2%)로 나타났다.

2.1. 5·18에 대한 인식

<표 3>은 5·18을 '폭동 및 체제전복' 또는 '민주화운동 및 자기방어'라고 규정한 사람들(712명)만 이용하여 작성한 표이다. 사건이 일어나고 16년이 지난 1996년 현재, 거의 모든 응답자들이 5·18을 '민주화운동'이라고 응답할 것으로 예상하였으나, 아직도 '폭동' 또는 '체제전복을 위한 민중혁명'이라고 생각하는 사람이 15.8%에 이르는 것으로 나타났다. 지역적으로는 대구·경북과 충청·강원에 거주하는 사람들이 5·18을 '폭동 및 체제전복'이라고 생각하는 사람이 상대적으로 많았다.

카이제곱분석에 의하면 연령에 따라 폭동 및 체제전복이라고 생각하는 사람의 비율이 유의하게 다른것을 알 수 있다($p < .05$). 20대~30대 중에는 폭동 및 체제전복이라고 생각하는 사람의 비율이 낮지만, 연령이 높아 질수록 그 비율이 높아진다.

폭동 및 체제전복이라고 생각하는 사람의 비율은 또한 교육정도에 따라서도 유의하게 다르다($p < .01$). 학력이 높을수록 폭동 및 체제전복이라고 생각하는 사람의 비율이 낮았다. 성별이나 부모의 출신 연고, 거주지역은 5·18에 대한 인식과의 관계는 유의하지 않는 것으로 나타났다(모두 $p > .10$).

전체적으로 6.7%가 '기타' 및 '모르겠다'는 말로 응답을 유보하였는데, 연령이 높아지거나 교육정도가 낮아질수록 그 비율이 높아진 것으로 나타났다.(<표 3>의 맨 오른쪽 열).

로짓모형은 앞에서 언급한 사회인구학적 변인을 모두 독립변수로 이용하였다. 두 가지 가능한 응답 중에서 비율이 적은 쪽(폭동 및 체제전복)의 응답자 수가 121명이므로, 독립변수의 수를 24개 이내로 결정하고, 모든 주요인과 2변수 교호작용을 모형에 포함시켰다. 모형의 결정과정은 다음과 같다. 먼저 주요인을 포함시킨 상태에서 단계적 선택방법(stepwise selection)을 이용하여 유의한 교호작용을 선택하였다. 유의한 교호작용과 그 주요인을 포함시킨 상태에서 다시 다른 주요인들의 유의성을 검정하여 선택하였다.

최종 모형의 -2 Loglikelihood값은 5.85로서 매우 유의한 모형이다(χ^2 분포에서 $p = 0.0012$). 이 모형은 다음과 같다. $ageg \times edu$ 는 연령과 교육의 교호작용을 나타낸다.

$$\text{logit}(p) = -2.08 + .54 * ageg + .14 * edu - .23 * (ageg \times edu)$$

모형에는 교육과 연령의 교호작용이 있기 때문에, 연령과 교육의 로짓에 대한 영향을 따로 해석할 수 없다. 해석을 쉽게하고 이해를 돕기위해서 로짓모형을 비율로 환산하여 그래프로 나타내었다([그림 1]).

먼저, 같은 교육정도에서 연령에 따른 차이를 보기로 하자. 교육정도가 중졸이하인 사람들은 연령이 높아질수록 폭동 및 체제전복이라고 생각하는 사람의 비율이 높아지는 반면, 대학이상의 학력인 사람들은 연령이 높아질수록 폭동 및 체제전복이라고 생각하는 비율이

다소 낮아진다. 예를들면, 국졸이하인 사람들은 20대에서 11.1%였던 것이 50대에서는 39%로 증가한 반면, 대졸인 경우에는 20대에서 16.0%였던 것이 50대에서는 10.5%로 오히려 감소하였다. 고졸인 사람들은 연령에 따라 거의 차이가 없다.

같은 연령에서 교육정도에 따라 비교해보면, 20대에서는 대학이상이 가장 폭동이나 체제전복이라고 생각하고, 교육정도가 낮아질수록 그 비율이 다소 줄어드는 것같이 나타났다. 이러한 결과는 다른 연령층의 결과와 다르고, 젊고 교육을 많이 받은 사람일수록 5·18을 민주화운동이라고 생각할 것이라는 카이제곱분석의 결과와도 상당히 다르다. 그 이유를 알아보기 위하여 교육정도에 따라 폭동 또는 체제전복이라고 생각하는 사람의 비율에 차이가 있는지를 95% 신뢰구간을 통해서 알아보았다(<표 4>). 20대에서는 각 교육정도에서의 95% 신뢰구간에 모든 교육정도에서의 비율이 포함되어, 교육정도에 따른 차이가 없는 것으로 나타났다. 30대에서도 역시 차이가 없었으므로 20대와 30대에서는 교육의 효과를 무시할 수 있다 하겠다.

40대 연령층에서는 교육정도가 낮아질수록 폭동이나 체제전복이라고 생각하는 사람의 비율이 높아진다. 이러한 경향은 연령이 높아지면 더욱 두드러져서, 50대에는 대학이상의 교육을 받은 사람(10.5%)과 국졸이하의 교육을 받은 사람(38.8%)간의 차이가 28%나 된다.

5·18에 대한 인식을 나타내는 모형에 거주지역이 포함되지 않은 것도 매우 의미있는 일이다. 아직 법적으로 밝혀지지 않았지만, 대구·경북은 5·18의 가해자들의 출신지역이기 때문에, 또 충청·강원은 보수적인 지역이기 때문에 이 지역에 거주하는 사람들이 다른 지역에 비해서 5·18을 더 나쁘게 인식하고 있을 것으로 예상하였다. 거주지역이 모형에 포함되지 않는 것은, 폭동이 아니었다는 사실이 이미 여러 텔레비전이나, 언론을 통해 어느 정도 알려졌기 때문인 것으로 생각된다. 이 모형에 의하면, 같은 연령과 교육정도의 사람이라면, 출신지역에 따라서는 폭동 또는 체제전복이라고 생각하는

비율에 별 차이가 없다고 할 수 있다.

2.2. 5·18 진상 규명 정도에 대한 견해

응답자 3명중 2명은 5·18에 대한 진상이 ‘밝혀지지 않았다’고 생각하고 있는 것으로 나타났다(<표 5>). 5·18에 대한 진상에 대하여 모르겠다 및 무응답을 제외한 사람들(744명)만을 이용하여 카이제곱 분석을 한 결과, 연령($p<.05$)과 교육정도($p<.01$), 부모의 출신연고($p<.05$), 거주지역($p<.001$)이 유의한 것으로 나타났다.

연령별로 보면 40대~50대는 진상이 밝혀지지 않았다고 생각하는 비율이 낮고, 연령이 낮아질수록 그 비율이 높아진다. 50대는 약 44.6%가 밝혀지지 않았다고 생각하고, 연령이 낮을수록 그 비율이 높아져서 20대에는 밝혀지지 않았다고 생각하는 사람의 비율이 77.0%에 달한다.

교육정도가 높은 사람들은 교육정도가 낮은 사람들에 비해 밝혀지지 않았다고 생각하는 사람이 더 많다. 이는 연령과도 무관하지 않을 것으로 보이는데, 나이가 많은 사람 중에는 교육정도가 낮은 사람이, 나이가 적은 사람 중에는 교육정도가 높은 사람이 많기때문일 것으로 생각된다. 실제로 교육은 연령과 높은 역상관관계를 보였다($r=-.46, p<.001$).

부모의 출신연고가 호남일 경우 그렇지 않은 경우보다 훨씬 많은 응답자가 미흡하거나 밝혀지지 않았다고 대답했다($p<.05$).

지역에 따라서는 대구·경북, 부산·경남에 거주하는 사람들 중에 밝혀졌다고 생각하는 사람이 더 많았고, 서울·경기, 충청·강원의 순으로 ‘밝혀졌다’고 생각하는 사람의 비율이 적어지는 것으로 나타났다($p<.001$).

밝혀지지 않았다고 생각하는 사람의 비율의 로짓을 종속변수로한 모형을 구하였다. 로짓모형에 포함 시킬 수 있는 변수의 수는, 대답한 사람의 크기가 적은 쪽(‘밝혀졌다’)에 대답한 사람이 265명이므로 Stokes 등의 엄격한 지침(응답자수/10)을 따르더라도 26개까지 포

함시킬 수 있다(Stokes et al, 1995 ; p191). 따라서 5개의 주요인과 모든 2변수간 교호작용을 포함한 모형으로부터 단계적 선택방법(stepwise selection)으로 모형을 선정하였다. 유의수준 0.05에서 유의한 변수만 선택하였을 때, 모든 교호작용과 교육정도는 유의하지 않았다. 카이제곱분석에서는 유의하게 나온 교육정도가 유의하지 않게 나온 것은 연령과의 높은 역상관관계 때문으로 보인다. 선택된 모형의 -2 Loglikelihood값은 90.9로서 매우 유의한 모형이다(χ^2 분포에서 $p=0.0001$). 모형에 이용된 변수 중 지역을 나타내는 더미(dummy)변수는, 충청·강원이면 D1을 1로, 부산·경남이면 D2를 1로, 대구·경북이면 D3를 1로 하고, 그 외의 경우는 0으로 지정하였다. 최종 모형은 다음과 같다.

$$\text{logit}(p) = 1.13 + 0.11 \cdot d1 - 0.54 \cdot d2 - 0.66 \cdot d3 \\ + 0.38 \cdot \text{sex} - 0.50 \cdot \text{age} + 1.46 \cdot \text{home}$$

모형은 먼저 로짓으로 설명하고, 이어서 비율을 해석하기로 한다. 여기서는 ‘밝혀지지 않았다’의 비율을 로짓으로 만들었기 때문에, 로짓(종속변수)이 커질수록 밝혀지지 않았다고 생각하는 비율이 높은 것이다. 즉, 모형에 포함된 변수의 계수가 양수이면, 그 변수의 값이 커짐에 따라 ‘밝혀지지 않았다’고 하는 비율이 더 높아지는 것이다.

모형을 보면, 지역에 따라 종속변수인 로짓값이 다른 것을 알 수 있다. 지역의 차이는 절편의 차이로 나타낼 수 있다. 전체 절편 1.13과 각 지역을 나타내는 더미변수의 계수와 합이 곧 그 지역의 절편이 되는 것이다. 따라서 충청·강원의 절편이 가장 높고, 서울·경기, 부산·경남, 대구·경북의 순으로 절편이 낮아진다.

성별(sex)은 남자는 1, 여자는 0이므로, 남자의 ‘밝혀지지 않았다’고 하는 로짓값이 여자보다 0.38 더 높고 연령은 높아질수록 로짓값이 낮아진다. 즉, 20대보다 30대의 ‘밝혀지지 않았다’고 하는 로짓값이 0.5 감소하고, 30대보다 40대가 또 0.5 감소한다.

부모 중 한 명이라도 호남출신인 사람들의 경우, 그렇지 않은 사

람들보다 ‘밝혀지지 않았다’고 하는 로짓값이 1.46만큼 높아진다. 다른 변수의 계수와 비교할 때, 1.46이라는 차이는 아주 큰 것이다.

[그림 2-A]에서 [그림 2-D]는 각 지역의 연령에 따른 성별, 부모 출신연고별 ‘밝혀지지 않았다’고 생각하는 사람의 비율이다. 먼저 지역을 비교해보자. 각 지역에서 20대의 밝혀지지 않았다고 생각하는 비율을 보면 충청·강원, 서울·경기는 거의 비슷한 수준에서 가장 높으며, 부산·경남과 대구·경북의 순으로 낮아진다. 이러한 경향은 모든 연령층에서 유사하게 나타난다. 즉 영남지역의 사람들은 “그만하면 됐다”는 생각을 가지고 있는 것으로 보인다.

연령을 보면, 연령이 높아질수록 밝혀지지 않았다고 생각하는 비율이 낮아진다. 성별이나 부모의 출신연고에 따라 그 낮아지는 정도가 다소 차이가 있으며, 모든 지역에서 유사한 경향을 보이지만 대구·경북, 부산·경남은 50대에 이르면 밝혀지지 않았다고 생각하는 비율이 서울·경기, 충청·강원보다 훨씬 낮다.

남성과 여성을 비교해 보면, 모든 지역에서 남성이 여성보다 더 밝혀지지 않았다고 생각하는 것을 알 수 있다. 그 차이는 부모의 출신연고가 호남이 아닌 경우에는 남·여 모두 비슷하게 감소하지만, 호남 출신이 한 명이라도 있는 경우, 여자들의 감소하는 속도가 남자들보다 다소 빠르다. 따라서 20대에 비해 50대의 성별, 부모의 출신 연고별, 비율의 차가 훨씬 크게 나타난다.

또한 부모가 한 명이라도 호남출신인 사람이 그렇지 않은 사람보다 밝혀지지 않았다고 생각하는 사람의 비율이 훨씬 더 높다. 이 또한 연령이 높아짐에 따라 차이가 더 크게 나타난다.

거의 모든 지역에서 출신 연고가 호남인 50대의 밝혀지지 않았다고 생각하는 비율의 수준이, 호남과 연고가 없는 20대의 비율의 수준과 비슷한 것을 볼때, 아직도 5·18의 규명에 대한 관심이 출신 연고따라 매우 차이가 나는 것을 알 수 있다. 현 거주지역에 따른 차이와 함께, 5·18규명에 큰 걸림돌이 될 것으로 생각된다.

3. 토의 및 결론

1980년 5·18이 일어난지 16년이 지난 지금, 5·18의 진상규명과 피해자들의 명예회복을 위한 특별법 제정되고 5·18관련자들이 법정에 서는 등 진상규명과 사건의 재조명이 활발히 진행되고 있다. 분할표의 카이제곱분석과 로짓모형을 이용하여 5·18에 대한 타지역민들의 인식 및 진상 규명 정도에 관한 견해와, 성별, 연령, 교육정도, 거주 지역, 부모의 출신 연고의 관계를 분석해 보았다.

5·18에 대하여 거의 모든 응답자가 민주화운동이라고 응답할 것으로 기대했으나, 17.0%가 폭동 또는 체제전복을 위한 민중혁명이었다고 생각하고, 15.4%는 자기방어를 위한 투쟁이었다고 생각하는 것으로 나타났다. 응답의 비율은 교육정도에 따라 유의하게 달랐으나, 일반적인 선입견과는 달리, 응답자의 거주지역에 따라서는 유의한 차이를 보이지 않았다. 이러한 결과는 TV 드라마와 기타 언론에 의해 진상이 많이 알려졌기 때문인 것으로 보인다. 로짓모형을 통해서 보면, 20대와 30대에서는 학력에 따른 차이가 없었으나, 40대 이상에서는 학력이 낮을수록 폭동 또는 체제전복을 위한 민중혁명이었다고 부정적으로 보는 견해가 많은 것으로 나타났다.

진상 규명정도에 대한 견해에 있어서는 과반수가 밝혀지지 않았다고 생각하는 것으로 나타났다. 특히 부모의 출신 연고가 호남인 사람들은 그렇지 않은 사람들에 비하여 진상이 규명되지 않았다고 생각하는 사람의 비율이 월등히 높다. 이 비율은 남자가 여자보다 더 높고, 연령이 낮아질수록 더 높다.

특히 호남과 연고가 없는 사람들은 연령이 높아짐에 따라 5·18은 이제 규명 됐다고 생각하는 비율이 급속히 증가하는 반면, 연고가 호남인 사람들은 연령이 높아져도 아직 규명이 덜 됐다고 생각하는 사람이 많다.

이 연구결과에서 주목할만한 점은 지역적 특성이다. 부산·경남, 대구·경북 응답자들은 5·18을 폭동 또는 체제전복을 위한 민중혁

명이 아니라고 하는데 있어서는 다른 지역의 사람들과 의견을 같이 하면서도, 진상규명 정도에 있어서는 ‘밝혀졌다’는 견해를 다른 지역보다 더 많이 나타내, 지역적 차이를 보인다. 폭동이 아니라고 생각하니 이제 그 얘기는 그만 하자는 생각을 가진 사람이 더 많은 것으로 풀이 될수 있겠다. 또 다른 지역적 특성은 호남연고이다. 호남에 연고를 둔 사람들은 5·18진상 규명에 훨씬 더 관심이 많은 것으로 보인다.

이 연구에는 몇가지 제한점과 아쉬운점이 있다. 첫째는 표본의 선정이 확률추출이 아니고 비례할당에 의한 비확률추출이었기 때문에 적절한 신뢰성의 산정에 문제가 있는 것이다. 이러한 제한점은 조사대상이 넓은 지역에 퍼져있고 시간과 경비의 제약이 있었기 때문에 불가피하였다. 그러나 random digit dialing등의 방법을 동원하여 할당추출의 제약을 극복하기 위해 노력하였으므로 큰 문제는 없을 것으로 생각한다.

두번째는 시간과 능력의 한계이다. 많은 정보가 원자료에 들어있었지만, 시간과 능력의 한계로 두 개의 변수에만 초점을 맞추었다. 또한 분석을 용이하게 하기 위하여, 종속변수를 하나로 만드는 과정에서 성격이 다소 다른 응답들, 예를 들면 ‘민주화운동’이라는 응답과 ‘자기방어’라는 응답을 하나로 묶었다. 이러한 범주들을 구분하여 multiple logit으로 분석하면 더욱 흥미있는 결과를 얻게될 것으로 생각한다.

〈참고문헌〉

- [1] 김순홍, 양철호, 신선호, 마지현, 김신영 (1996). 「광주 5·18에 대한 전국민 설문조사 보고서」. 사단법인 광주사회조사연구소.
- [2] Choe, S.M., D.W. Kim, S.H. Kim and C. Yang (1996). "Overcoming Silenced Voices: Views from the Inside". Presented at the 48th Annual Meeting of The Association for

Asian Studies, Hawaii.

[3] Stokes, Davis, and Koch (1995). Categorical Data Analysis Using the SAS System. SAS Institute. Cary, NC.

〈참고문헌〉

- [1] 김순홍, 양철호, 신상호, 마경환, 김진영 (1996). 『광주 5·18에 대한 전국인 설문조사 보고서』, 시민법원 광주사회조사연구소.
- [2] Choe, S.M., D.W. Kim, S.H. Kim and C. Yang (1995). "Overcoming Silenced Voices: Views from the Inside," Presented at the 18th Annual Meeting of The Association for

〈표 1〉

응답자의 사회인구학적 특성

변 인				n	%
총 성	응	답	자	763	100.0
		남	별	383	50.2
		여		330	49.8
연			령		
20			대	247	32.4
30			대	231	30.3
40			대	161	21.1
50			대	124	16.2
교	육	정	국		
무	학 /		졸	84	11.0
중			졸	84	11.0
고			졸	331	43.4
대	학	이	상	253	33.2
무		응	답	11	1.4
거	주	지	역		
서	울	·	기	395	51.8
충	청	·	원	116	15.2
부	산	·	남	151	19.8
대	구	·	북	101	13.2
부	호	남	출		
한	분	이	신	145	19.0
두	분	다	아	618	81.0

〈표 2〉

5·18에 대한 인식과 진상규명에 대한 견해

변 인				n	%
총	응	답	자	763	100.0
5·18	에	대	인		
폭	주	화	운	53	7.0
민			동	474	62.1
체	제	전	복	67	8.8
자	기	방	어	118	15.4
기	타	및	모	51	6.7
진	상	규	명		
모	두	밝	혀	16	2.1
대	부	분	밝	238	31.2
미	흡	하	혀	474	62.1
전	혀	밝	혀	17	2.2
기	타	및	모	18	2.4

<표 3>

5 · 18에 대한 인식

구	분	합 계 + % (N)	폭동 및 체제전복	민주화운동 및 자기방어	기타 및 모르겠다 ++
전	체	100 (712)	17.0	83.0	6.7
성	별				
	남	100 (363)	16.6	83.4	5.2
	여	100 (349)	17.4	82.6	8.1
연	령*				
20	대	100 (240)	15.9	84.1	2.9
30	대	100 (216)	13.9	86.1	6.6
40	대	100 (151)	17.5	82.6	6.5
50	대	100 (106)	25.2	74.8	14.4
교	육				
정	도**				
무	학 / 국	100 (68)	28.0	72.0	19.1
중	졸	100 (77)	25.4	74.6	8.0
고	졸	100 (318)	14.8	85.2	4.1
대	학 이 상	100 (239)	14.0	86.0	5.6
부	모				
호	남				
출	신				
한	명 이 라 도	100 (108)	13.4	86.6	1.0
두	명 다 아 님	100 (604)	17.7	82.3	7.6
거	주 지 역				
서	울 · 경 기	100 (375)	15.8	84.2	5.0
충	청 · 강 원	100 (105)	22.0	78.0	9.5
부	산 · 경 남	100 (143)	14.0	86.0	5.3
대	구 · 경 북	100 (89)	21.4	78.7	11.9

+ 합계 712명은 기타 및 모르겠다는 제외된 사례수.

++ 기타 및 모르겠다는 전체 763명 중의 백분율. 분석에서 제외.

* p<.05, ** p<.01

<표 4>

로짓모형에 의한 연령 · 교육정도별

‘폭동 및 체제전복’의 비율(p)과 95% 신뢰구간

교육정도	20대	30대	40대	50대
	p (하한, 상한)	p (하한, 상한)	p (하한, 상한)	p (하한, 상한)
무학/국졸	.11 (.04, .26)	.18 (.11, .29)	.27 (.21, .34)	.39 (.29, .49)
중 졸	.13 (.07, .22)	.16 (.11, .23)	.21 (.17, .25)	.27 (.21, .33)
고 졸	.14 (.10, .20)	.15 (.12, .19)	.16 (.13, .20)	.17 (.12, .24)
대 학 이 상	.16 (.10, .25)	.14 (.10, .19)	.12 (.08, .18)	.11 (.05, .19)

〈표 5〉

5.18 진상규명 정도에 대한 견해

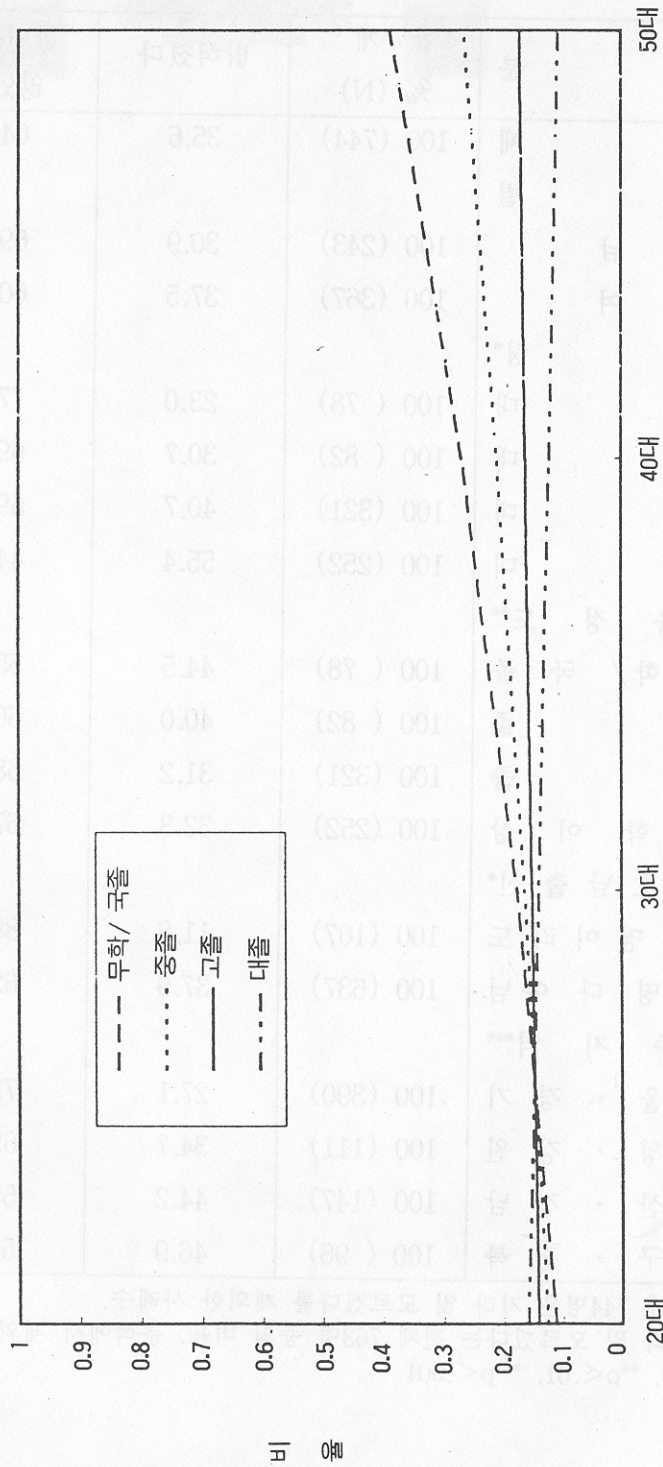
구	분	합 계 ⁺ % (N)	밝혀졌다	밝혀지지 않았다	기타 및 모르겠다 ⁺⁺
전	체	100 (744)	35.6	64.4	2.4
성	별				
	남	100 (243)	30.9	69.1	1.6
	여	100 (367)	37.5	60.4	3.3
연	령*				
20	대	100 (78)	23.0	77.0	1.6
30	대	100 (82)	30.7	69.3	2.2
40	대	100 (321)	40.7	59.3	0.5
50	대	100 (252)	55.4	44.6	7.2
교	육	정	도**		
무	학/	국	졸		
중		졸			
고		졸			
대	학	이	상		
부	모	호	남	출	신*
한	명	이	라	도	
두	명	다	아	님	
거	주	지	역***		
서	울	·	경	기	
충	청	·	강	원	
부	산	·	경	남	
대	구	·	경	북	

† 합계 744명은 기타 및 모르겠다를 제외한 사례수.

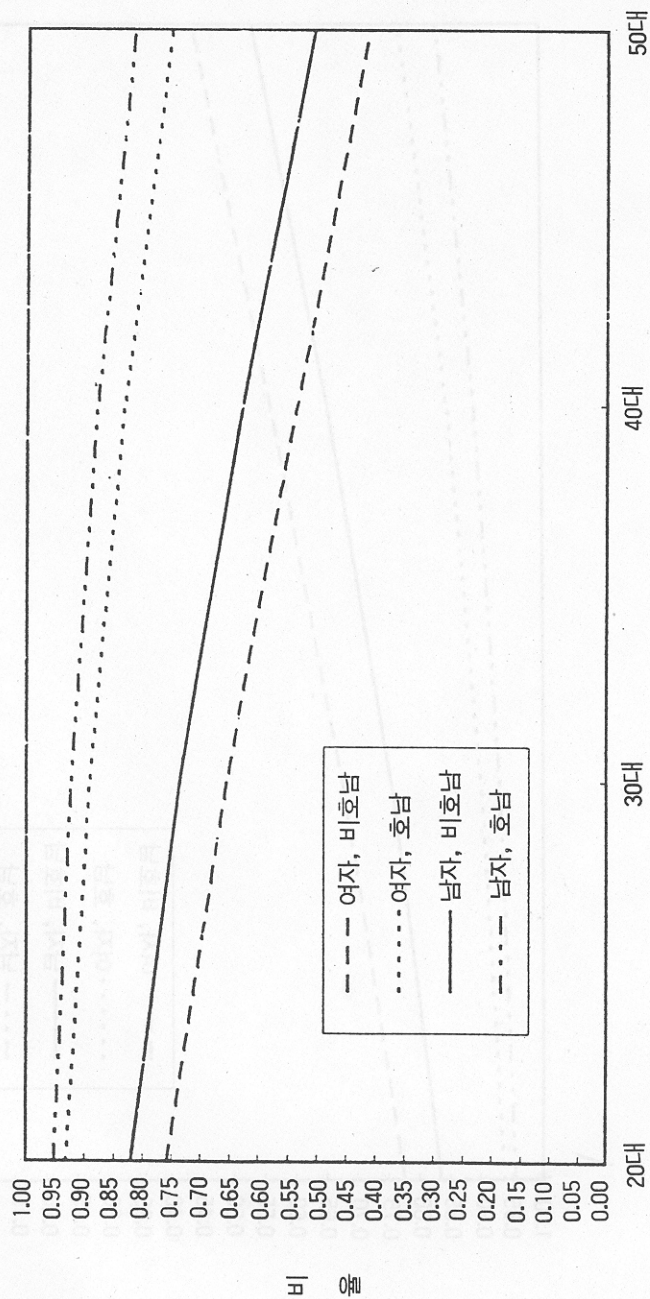
†† 기타 및 모르겠다는 전체 763명 중의 비율. 분석에서 제외.

* p<.05, **p<.01, ***p<.001

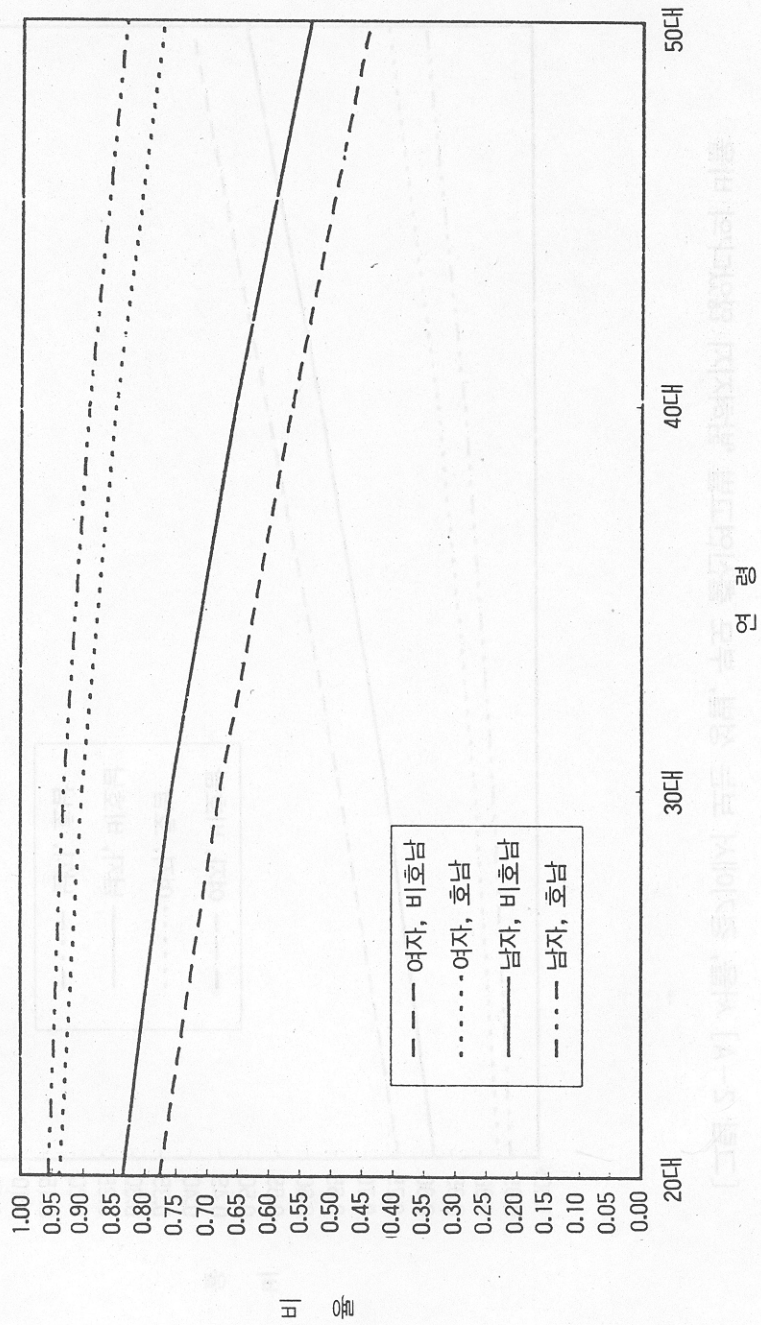
[그림 1] 교육수준별 '폭동 및 체제전복'의 비율



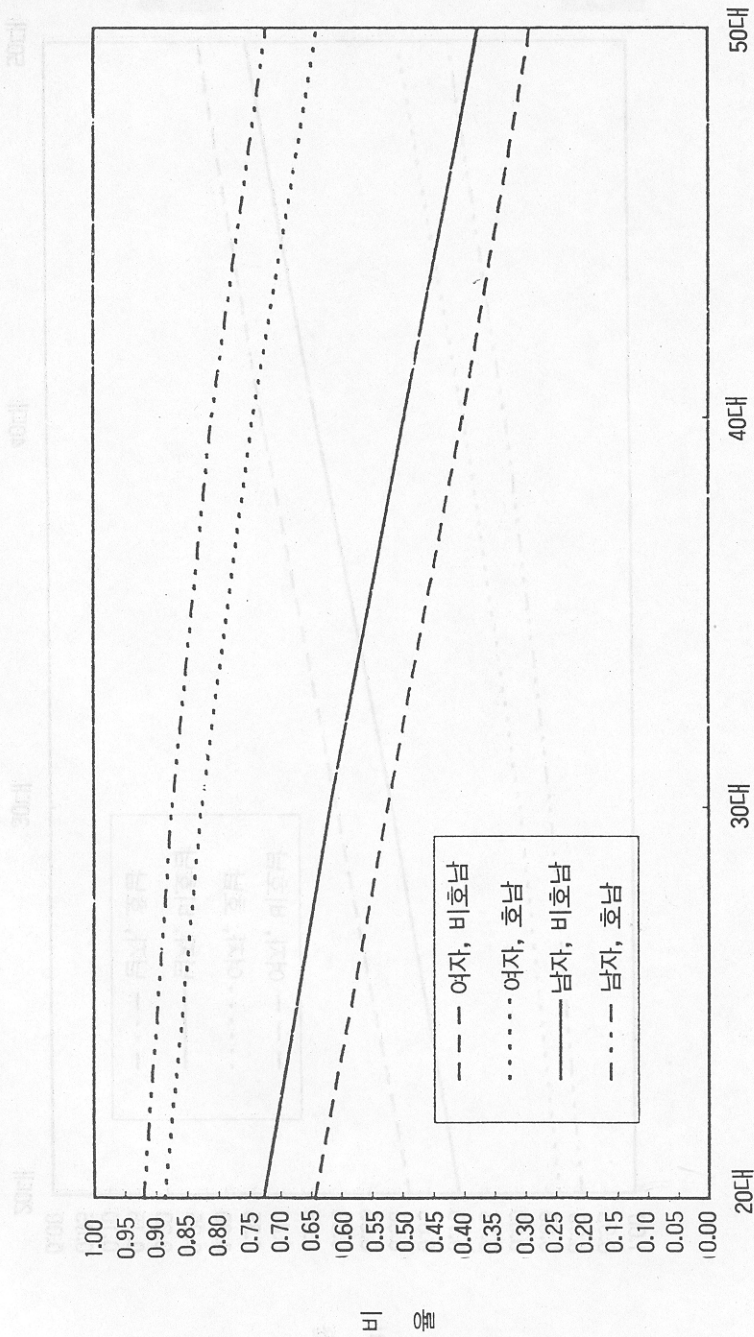
[그림 2-A] 서울, 경기에서 보는 성별, 부모 출신연고별 '밝혀지지 않았다'의 비율



[그림 2-B] 충청, 강원에서 보는 성별, 부모 출신연고별 '밝혀지지 않았다'의 비율



[그림 2-C] 부산, 경남에서 보는 성별, 부모 출신연고별 '밝혀지지 않았다'의 비율



[그림 2-D] 대구, 경북에서 보는 성별, 부모 출신연고별 '밝혀지지 않았다'의 비율

