

310.6  
통계청  
5.1  
=2

제 1 회

『통계의 날』 기념 논문집

1995. 12

통 계 청

060000



## 머 리 말

국내적으로는 개인의 비밀보호의식의 강화와 맞벌이 부부등 부재자가구의 증가로 통계조사여건이 점차 어려워지고 있으며, 또한 지방자치제의 본격 실시를 계기로 보다 다양하고 지역별로 세분된 통계에 대한 수요도 증가하고 있습니다.

국제적으로는 OECD 가입을 위한 제반 준비로 국제기준에 맞는 통계의 생산이 요청되어 새로운 체계에 의한 통계생산이 절실히 요청되고 있습니다. 이와같이 통계환경의 급격한 변화는 통계업무 종사자들의 끊임없는 변신을 요청하고 있습니다.

이제까지의 통계생산이 경제발전을 뒷받침하는 데 주안점이 놓여 있었다면, 이제부터는 그 중요성이 급격히 증대되고 있는 환경이나 복지 등에 대한 통계 생산으로 이천해 가야함은 재론의 여지가 없다할 것입니다.

이러한 통계 주변의 급격한 환경변화에 대응하여 통계의 중요성에 대한 국민의 관심과 이해를 깊게 인식시켜 날로 악화되는 조사 환경속에서 통계조사에 대한 국민의 협조증진과 통계인의 자긍심 및 사기양양을 위해 호구조사규칙 제정일인 9월 1일을 「통계의 날」로 처음 제정한 바 있습니다.

따라서 지난 9월 19일에 대한통계협회의 주관과 통계청 및 한국통계학회의 후원으로 제1회 「통계의 날」 기념행사를 거행하게 되었습니다. 이 행사를 통하여 통계인들의 대화의 장을 마련하여 여러 의견을 결집하고, 통계발전을 기약하는 자리가 되었음은 뜻 깊은 일이라 할 것입니다. 아울러 그 동안 「통계의 날」을 제정하고 행사가 거행될 수 있도록 수고해 주신 관계자 여러분들의 노고에 진심으로 감사드립니다.

이 논문집에는 기념행사의 일환으로 실시한 통계심포지움에서 발표된 3편의 논문과 토론 그리고 전국 대학교 통계학과에 재학중인 대학생을 대상으로 한 통계경진대회에서 입상한 논문을 게재하였습니다. 이 논문집이 통계 발전을 위한 밑거름이 될 수 있도록 널리 활용하여 주시기 바랍니다. 또한, 이를 계기로 앞으로 더욱 좋은 논문들이 계속하여 나올 것을 여러분들과 함께 기약해 봅니다.

1995. 12

통계청 통계기획국장 김 일 현

# 목 차

I. 기 념 사/대한통계협회장 .....	3
II. 격 려 사/통계청장 .....	5
III. 축 사/한국통계학회장 .....	8
IV. 심포지움 발표 논문 및 토론 .....	11
1. 제1주제 : 통계학 교육과 학회발전에 대한 회고와 전망 .....	13
◆ 발 표 : 백 운 봉/전 고려대학교 교수	
◆ 토 론 : 김 종 호/동국대학교 교수	
구 자 홍/인하대학교 교수	
2. 제2주제 : 통계기구와 조사방법의 변천 .....	28
◆ 발 표 : 윤 기 중/연세대학교 교수	
◆ 토 론 : 남궁 평/성균관대학교 교수	
이 재 창/고려대학교 교수	
3. 제3주제 : 국가통계조직의 발전방향 .....	53
◆ 발 표 : 김 일 현/통계청 통계기획국장	
◆ 토 론 : 박 성 현/서울대학교 교수	
이 성 만/한국은행 부부장	
V. 통계경진대회 입상논문 .....	77
《1등-1편》	
1. 고려대학교 학생들의 출신지역별 의식비교/안태근, 오진승, 심성환, 강성민 ...	79
《2등-2편》	
2. 로짓모형을 이용한 '94년 한국프로야구의 구단별 특성분석	
/정진우, 김대환, 서두성 .....	98
3. 남녀간의 외모평가 차이검정/이의구.....	110
《3등-3편》	
4. GPA의 문제점 및 해결방안 연구/이철영 .....	122
5. 두 모집단의 모평균 비교시 발생하는 문제점/배은경.....	134
6. 수출, 수입액을 이용한 환율 예측/이혁중 .....	145

# 기 념 사

(대한통계협회회장 함 만 준)



9월 1일은 「통계의 날」입니다. 금년에는 8월 하순부터 9월 상순에 걸쳐 중국 북경에서 국제통계회의에 연이어 한·중·일의 통계회의 등이 개최되는 바람에 이렇게 대회 행사가 늦어지게 된 것입니다.

「통계의 날」은 통계의 중요성에 대한 국민의 관심과 이해를 깊게 하고 통계 조사에 대한 국민의 一層의 협력을 얻는 동시에 통계업무에 종사하는 사람들의 사기를 높이기 위하여 1994년 9월 「통계의 날」 선정을 위한 공청회를 열어 이곳에서 건의한 것을 통계위원회에서 양해함으로써 9월 1일로 정해졌습니다.

9월 1일을 「통계의 날」로 정하게 된 것은 가장 기본적인 통계조사인 호구조사를 체계적으로 실시할 수 있도록 별도의 칙령으로 규칙이 1896년 9월 1일에 처음으로 마련된 것이 유래가 된 것입니다.

매년 국가나 지방자치단체 또는 관련기관에서 작성하는 통계는 350개 이상이나 됩니다. 그리고 이러한 통계는 통계조사의 대상이 된 개인, 국가, 법인 및 사업체 등의 협력에 의하여 작성되고 있습니다.

이 「통계의 날」을 기점으로 관계정부부처, 지방자치단체 및 기타 관계기관은 국민 한 사람 한 사람에 대해 통계에 대한 올바른 이해와 협력을 얻고 동시에 통계관계자의 사기의 양양을 도모하기 위해 매년 통계대회, 통계공로자 표창식, 강연회, 체육대회, 전시회 등의 행사를 실시하게 되는 것입니다.

통계조사에 대한 국민의 협력을 어떻게 확보하는가는 통계의 중요성이나 통계조사의 필요성에 관하여 국민의 이해가 있어야 합니다. 즉 통계조사에 의한 정보수집이 통계조사의 목적에 비추어 볼 때 필요 최소한도의 것일까, 구체적으로 말하면 국민의 보고 부담의 경감의 관점에서 조사대상자의 숫적크기, 조사내용의 양적크기, 기입의 난이도를 엄정히 고려한 뒤에 통계조사에 대한 국민의 협력을 요구하여야 되지 않나 생각됩

니다. 그러므로 통계에 대한 국민의 인식이 올바른 방향으로 바뀌도록 유도하여야 합니다.

통계조사에 대한 국민의 적극적인 이해와 협조가 필요합니다. 기본적으로 양질의 통계는 통계조사에 대한 국민의 성실하고 정확한 응답에 달려있기 때문입니다.

한편 지금까지 우리 사회에서는 흔히들 통계를 庶政百般의 기초자료라고 일컬었고 또한 일국의 경제, 사회, 문화의 과거, 현재, 미래의 모습을 만들어 내는 통계인에 대해서는 기여한 만큼의 사회적 예우와 대접을 받지 못하였습니다만은 통계에 대한 사회적 인식이 고조되고 국민의 이해가 깊어질 때까지 우리 통계인은 자부심을 갖고 양질의 통계작성에 더욱 열을 다하여야 하겠습니다.

「통계의 날」을 기하여 이러한 우리의 행사가 해를 거듭할수록 국민의 깊은 이해와 협조를 기대하면서 우리들 자신도 자부심을 갖고 꾸준히 정진해 나갑시다.

감사합니다.

# 激 勵 辭

(統計廳長 李 康 雨)



오늘은 우리 統計人들이 오래전부터 念願해 오던 「統計의 날」을 制定하고 첫번째로 紀念하는 아주 뜻 깊은 날 입니다. 그동안 우리나라의 統計發展에 온갖 精誠을 다하신 統計元老 諸賢과 統計學者님들, 그리고 正確한 統計作成에 끊임 없이 努力하여 주신 우리 統計作成機關의 統計人들의 熱과 誠이 凝集되어 오늘 「統計의 날」이 誕生하게 된 것입니다.

올해는 國內外的으로 큰 變革을 겪고 있습니다. 對外的으로는 우루과이라운드의 結果로 WTO體制가 出帆하면서 世界는 開放化의 急速한 進展으로 經濟社會 모든 부문에서 無限競爭時代를 맞게 되었습니다.

對內的으로는 歷史的으로 意義가 깊은 光復 50週年을 맞는 해 이기도 하며, 世界의 開放化 물결속에서 이에 對應하기 위하여 世界化 推進을 적극 開放하고, 또한 6.27 地方自治團體長과 地方議會 議員選舉로 地方化時代가 開幕된 아주 뜻 깊은 해 이기도 합니다.

이러한 國內外的인 環境變化는 우리 統計人에게도 變革을 要求하고 있습니다. 先進國 進入을 위한 OECD 加入, 世界化와 同時에 推進되는 地方化를 支援하기 위해서는 새로운 統計의 發展은 물론 既存統計의 質을 改善 시키는 것이 우리의 重要한 課題가 되었습니다.

現在 統計法의 適用對象으로 作成되고 있는 統計는 98개 機關에 370種에 達하고 있으나, 內容的으로는 OECD加入에 따라 우리가 提供하여야 할 80余種의 統計要請에 크게 미치지 못하고 있는 實情입니다.

世界化, 開放化 進展에 따른 社會 各 部門의 無限競爭을 支援하기 위해서는 各國의 統計情報를 迅速히 入手하여 提供함은 물론, 우리나라의 統計 先進化도 時急히 要請되고 있습니다. 또한 地方化時代의 早期定着에 절대적으로 不足한 細分된 地域統計의 生産과 地方自治團體의 統計作成能力을 빠른 時日內에 涵養시키는 것도 우리의 時急한 當面課題인 것입니다.

이러한 緣由에서 오늘 이 자리는 우리의 오랜 宿願인 第1回 「統計의 날」을 紀念하는 祝賀의 자리임과 동시에, 世界化와 地方化 推進에 따라 우리에게 賦課된 統計의 先進化를 다짐하여야 하는 날이기도 합니다.

다시 말씀드리면 올바른 政策形成을 支援하고 企業經營의 指標로서, 더 나아가 國民 個個人의 日常生活속에 깊숙히 스며드는 信賴받는 統計로 改善發展시켜 나가야 할 무거운 義務와 責任을 다 함께 생각하여야 할 視點에 있습니다.

물론 統計의 開發과 改善은 어느 한 個人이나 한 機關의 獨白的인 노력만으로는 達成될 수 있는 것은 아닙니다. 統計를 책임맡고 있는 우리 統計人이 올바른 統計를 만들기 위한 모든 노력을 傾注해야 함은 물론 統計作成에 直接 情報를 提供해 주는 國民들의 統計에 대한 誠實한 協助가 절대 필요합니다.

民主化의 進展에 따른 國民들의 私生活 保護意識의 增大와 企業의 祕密保護 強化로 因하여 統計調査環境은 날로 惡化되고 있습니다. 또한 尖端情報産業의 發達로 統計情報 以外の 産業에 從事하고 있는 기관으로 부터 우리는 挑戰을 받고 있습니다.

이러한 時代的 變化에 安逸하게 대처하면 우리 統計人이 설 자리가 없어질 것입니다. 劣惡할 調査環境을 탓만 할 것이 아니라 國民의 統計에 대한 認識轉換을 위하여 良質의 統計를 생산함으로써 國民의 統計에 대한 信賴를 높여 나가는 한편, 우리가 生産한 統計를 解釋하고 分析하여 統計利用者의 요구에 부응할 수 있도록 積極서비스하고 弘報해 나가야 할 것입니다.

이러한 脈絡에서 「統計의 날」의 制定은 큰 意義를 갖는다 하겠습니다. 우리는 「統計의 날」 制定을 契機로 統計에 대한 國民의 관심과 이해를 넓혀나가는 한편, 統計業務에 從事하는 우리 統計人 모두의 自矜心을 높여 統計發展에 專念할 수 있도록 하여야 하겠습니다. 國家統計事務를 管掌하고 있는 中央統計機關의 책임자로서 본인은 第1回 「統計의 날」을 맞이하여 國家統計發展을 위한 몇가지 方案을 玆려하고자 합니다.

첫째, 國內外的으로 增加되는 統計需要에 效果的으로 對應할 수 있도록 統計廳의 統計 作成 機能을 계속 擴充해 나가도록 노력 하겠습니다.

둘째, 主要 中央部處 統計作成機關의 統計企劃能力을 強化하고, 地方政府의 統計作成 基盤의 造成을 적극 支援하겠습니다.

# 축 사

(한국통계학회장 송 문 섭)



오늘 제 1 회 통계의 날을 맞이하여 내외 귀빈 여러분과 함께 이 날을 경축하며, 특별히 오늘 유공자로 표창을 받으신 여러분들께 감사의 말씀과 함께 축하의 말씀을 드립니다. 또한, 이 경사스러운 자리에서 축사를 드리게 된 것을 영광으로 생각합니다.

국가의 발전과 더불어 사회구조나 경제구조, 산업구조 등이 모두 복잡한 형태로 바뀌어 감에 따라 국가기관의 정책결정과정이나 민간의 연구과정에서 통계의 수요가 급증하고 있는 것이 오늘의 현실입니다. 또한 모든 학문분야에서는 연구의 과학화와 객관화를 중시하면서 통계학의 응용이나 통계적 방법론의 적용도 일반화 되고 있는 것이 세계적인 추세입니다.

우리의 생활 주변에서도 통계자료를 통한 각종정보가 홍수를 이루고 있다고 할 수 있습니다. 예를 들어 여론조사의 결과와 각종 경제지표, 통계분석자료 등이 여러 언론매체를 통하여 발표되고 있으며, 특히 통계청을 비롯한 여러 통계관련기관에서는 많은 자료를 요약 발췌할 뿐만 아니라 재생산하여 유용한 고급 정보를 제공해주고 있습니다.

특히나 컴퓨터와 통신기술의 발달로 인하여 이와 같은 추세는 더욱 가속화 되고 있으며, 그만큼 통계의 중요성을 정부 당국이나 온 국민이 공감하게 되고, 이를 바탕으로 통계의 날이 제정된 것으로 이해합니다.

이를 계기로 통계업무에 종사하는 우리 통계인들은 커다란 자긍심을 느끼는 동시에, 통계인들의 위상이 한층 격상되어지기를 기대해봅니다.

이제 지방자치시대가 열림에 따라 국가의 통계도 전국단위의 통계뿐만 아니라 지역 통계와 사회복지통계의 필요성이 더욱 절실해지고 있으며, 이에 따라 국가기관의 통계업무도 급증하리라 생각됩니다. 그러나 통계관련기관의 행정조직이 이와 같은 요구에 따를만큼 충분하지 못하므로, 우리 통계인들은 우리나라의 공식통계가 정상궤도에 오를 수 있도록 합심하여 문제해결에 앞장서야 할 것입니다.

셋째, 國際的으로 우리나라의 統計位相을 높이기 위하여 國家間, 國際間에 統計交流 協力을 계속 확대해가는 한편, 2001년 國際統計機構(ISI)總會를 서울에서 開催하도록 하겠습니다.

넷째, 統計의 產學協同을 보다 強化하기 위하여 統計委員會의 活性化와 統計研究所의 設立을 적극 推進하도록 하겠습니다.

바쁘신 가운데서도 이 자리에 參席하여 紀念式을 빛내주신 여러분에게 감사를 드리며, 아울러 어려운 與件 속에서도 統計發展에 많은 寄與를 해주신 有功者와 統計學術 심포지움과 統計競進大會에 참여하여 주신 분들에게 진심으로 축하의 말씀을 드립니다.

끝으로 오늘 第1回「統計의 날」行事가 國家統計發展에 커다란 跳躍의 발판을 構築하는 記念碑的인 祝祭의 날이 되길 바라면서 激勵辭에 갈음하고자 합니다.

감사합니다.

제 1 회 『통계의 날』 기념

통계심포지움 발표

논문 및 토론

## 통계학 교육과 학회발전에 대한 회고와 전망

발표 : 전 고려대학교 통계학과 교수 백운봉

### 1. 통계학과의 창설과 통계학회의 발족

통계학은 17세기 부터의 오랜 역사를 가지고 있다. 미국에서 통계학회가 창립된 것이 1839년 이다. 미국에서 이보다 더 오래된 학회는 철학회 뿐이라고 한다. 그러나 새로운 학문으로 현대적인 통계학이 우리에게 소개된 것은 제2차대전 이후 즉 1945년 이후의 일이라고 할 수 있다. 우선 일본에서 2차대전 후 그동안 발전된 새로운 통계학이 충격적으로 받아들여져서 추측통계학 또는 추계학(推計學)으로 우리에게 소개(책을 통해서)되었고 미군-군정하의 농업시험장을 통해서 또 미국 유학생 특히 생각지도 않은 통계학으로 고생한 농·생물학 전공자들을 통해서 새로운 필수적인 학문으로 우리에게 전달 보급되었던 것이다. 내 자신도 농과대학 출신으로 농업학교 수학교사로서 이러한 영향을 받고 통계학을 전공하게 되었던 것이다.

그러나 전통적으로 통계학이라고 하면 경제통계학과 똑같은 개념으로 이해되어 온 것이 사실이다. 이와 같은 사정은 우리나라에서도 마찬가지로여서 경제학과 교수들에 의해서 경제학과의 인접학과로 즉 상경계열 학과로 1963년에 고려대에, 64년에 성균관대, 중앙대에 통계학과가 설치되고, 그리고 66년에 연세대에 응용통계학과가 역시 상경대학에 신설되었고, 자연계로서는 오직 하나 62년에 동국대에 통계학과가 설치되었던 것이다.

이와 같은 상황아래 1971년 12월에 한국통계학회가 창립되었고 이때에는 물론 경제학자들의 참여도 많았던 것이 사실이다.

### 2. 통계학회의 발전과 통계학교육

지난 '91년 5월로 통계학회 창립 20주년을 맞이하였었다. 통계학회가 창립된 1971년은, 이미 언급한바와 같이, 통계학은 각 대학에서 강의되고 통계학과가 서울시내 몇몇 대학에 설치되어 이미 졸업생도 배출하고 있는 그러한 때 이다. 통계학회만 해도 이미 '63년에 당시 서울대 의대 교수이던 권이혁교수와 문리대 교수이던 최지훈교수가 중심이 되어 창립발기가 된 일이 있었다.

우리나라에서 '70년대는 통계학에 대한 인식이 고조되고 통계학이 발전궤도에 오르기 위한 기반이 조성되던 시기였다고 할 수 있다. 이무렵 해외 그 중에서도 미국에서 통계학으로 박사학위를 취득한 분들이 속속 귀국하여 대학에서 자리를 잡고 학회지를 통해서 논문을 발표하기도 한 때이다. 이와 같이 해서 학회의 기반이 다져지고 발전궤도에 오르기 시작한 시기라고 할 수 있다. 이

와 같은 사정이 1982년에 발간된 통계학회 창립 10주년 기념호인 “통계학연구, 제 10권”에 잘 요약되어 나타나 있다.

1971년에는 5개 대학에 통계학과 또는 응용통계학과가 설치되어 있었던 것이 1981년 까지에는 전국적으로 27개 국립 또는 사립대학에 통계학과가 설치되었고, 이중에서도 서울대학교를 비롯한 몇몇 국립대학은 계산통계학과라는 이름을 사용하고 있는 것이 특색이다. 그리고 새로 설립된 통계학과의 대부분은 자연계에 속해 있다. 이와 같은 통계학과의 증가추세는 계속되어 1990년에는 통계학과가 설치된 대학수는 59개로 되어 있으며 이 이후에도 매년 3-4개 대학씩이 계속해서 통계학과를 신설하고 있어, 아마도 현재는 70여 대학에 이르고 있을 것으로 생각한다.

이와 같은 사정으로 통계학회 회원은 급속도로 증가하게 되었다. 더구나 이것이 일찍이 설치된 대학에서의 석·박사학위 취득자의 배출, 또 이들 대학출신들의 미국에서의 학위취득과 귀국, 여기에 수학과등 타 학부출신의 통계학전공자들 까지 합쳐져서 통계학회는 급속도로 팽창하고 발전한 것이다.

통계학회 회원수는 1993년 회원명부에 사진이 첨부되어 수록된 것만으로도 363명을 헤아릴 수가 있다. 이중에서 국내 대학 박사학위 소지자가 162명, 미국내 대학에서의 박사학위 취득자가 133명이며 기타 외국에서의 취득했거나 타 분야 학위 취득자가 14명으로 실로 회원 중 82%가 박사학위를 가지고 있다. 그리고 회원중 80%정도가 대학교수이다. 이와 같은 통계학회 회원은 매년 더 급속하게 증가하여 현재는 600여 명에 이르고 있다.

통계학회는 이와 같이 수준높은 학회로 성장하고 있는 것이다. 따라서 이와 같은 학회에서의 학회지 ‘통계학연구’의 학술적 수준도 급속도로 높아지고 세계 어느곳에서도 무시할 수 없는 학술지로 성장하고 있다. 학회에서는 1989년부터 ‘응용통계연구’를 추가적으로 발행하고 있으며 또 1994년도 부터는 ‘한국통계학회논문집’도 발행하고 있다.

여기에 또 한가지 중요한 발전이 있었다. 그것은 통계학회 산하에 여러개의 ‘연구회’가 조직되어 활동하고 있다는 것이다. 공업통계연구회, 생물통계연구회, 조사통계연구회, 통계계산연구회, 그리고 통계교육·상담연구회가 부설되어 있다. 또 최근에는 공식통계연구회도 발족한바 있다. 이들 각 연구회에 의한 워크샵 또는 학술논문 발표회등은 이와 관련있는 인접학회 또는 관련실무자들과의 회동기회를 마련할 수 있어 통계학적 사고(思考)의 보급과 상호이해에 크게 이바지할 수 있는 계기를 마련해 주고 있는 것이다. 예를 들면 공업통계연구회에서 통계적 품질관리에 관한 워크샵을 개최하면 통계학자, 품질관리학자, 품질관리실무자, 그리고 기타 관심있는 사람들이 같이 회동할 수 있는 기회를 제공하게 된다. 생물통계연구회에서의 모임에서는 생물통계학자와 생물학자, 의사, 독성학자, 제약회사 관계자들이 같이 모여 현안을 토의할 수 있는 기회를 얻을 수 있을 것이다.

여기에서, 우리는 통계학 주변에서 혁명이 진행되고 있다는 것을 깨달아야 한다. 우리는 컴퓨터의 발전이 통계학발전의 기폭제가 되고 있다는 인식을 넘어서 이제는 이것이 통계학 자체의 혁명

을 자극하고 있다는 사실을 심각하게 받아들여야 한다. 컴퓨터의 발전은 통계학연구의 방향에 지대한 영향을 끼치고 있는 것이 사실이다. '부스트랩(Bootstrap)' 방법과 같은 것은 컴퓨터의 발전된 기능없이 그 발상조차 불가능 하였을 것이다. 이와 같이 생각지도 못하던 방향으로의 통계학의 발전은 계속되고 있는 것이다.

모든 통계적 분석은 컴퓨터를 이용해서 이루어지고 있다. 아마도 앞으로는 간단한 평균치 계산일 지라도 펜을 들고 종이 위에다 계산하는 사람은 없을 것이다. 전에는 대학원에서나 배우고 이론적으로 이해하는 데 그쳤던 것이나 또는 규모가 큰 역행렬을 계산해야 하는 등 계산이 복잡하고 힘들어서 실제로는 이용하기가 어려워 실용면에서는 탁상공론에 그쳤던 통계적방법도 SAS나 SPSS와 같은 통계계산용 패키지를 사용해서 손쉽게 이용할 수 있게 되어 있다. 그리고 이와 같이 얻은 분석결과가 일반 이용자 앞에 꺼리낌 없이 출력되고 있는 것이 현 실정이다.

이러한 현실에서 통계학도는 어디에서 무엇을 할 수 있는가? 일반적인 통계업무 종사자가 컴퓨터에서의 출력결과를 이해하는 데 누군가의 조언이 필요한 것인가? 이러한 것이 반드시 통계학과 출신자에 의해서만 가능한 것인가? 과연 통계학과와의 존속이 필요한 것인가?

이러한 문제들은 현존 통계학과 교수들간에 심각하게 논의되고 해결책이 강구되어야 할 시기와 있다고 생각한다.

### 3. 통계전문가(Statistician)를 필요로 하는 분야

우리나라에서 통계조사업무의 최고기관이 통계청이고 농림수산부에도 또 국 규모의 농산물 통계조사기관이 있다. 이들이 통계조사 및 분석 그리고 결과를 공표하는 정부최고기관이다. 물론 이외의 각 행정부처에서도 통계청 기타 통계전문기관에서 제공하는 자료외에 필요에 따라 독자적으로 통계자료를 수집하여 적절한 의사결정을 하고 있을 것이다. 이러한 경우 우리나라에서 통계전문가가 이에 관여하고 있는지는 알 수 없으나, 미국의 경우를 예로 들면, 미국의 환경보호처에는 환경에 관한 통계와 정보를 수집하고 의사결정에 참여하고 있는 부서가 있다. 이 부서에서 통계학 전공자를 구하고 있다. 그런데 어떤 능력의 소지자를 구하고 있는가가 우리의 관심사이다. 여기에서는 “어떠한 문제이든 이것에 파고들어 해결할 수 있는 통계전문가(statisticians that are persons for all seasons)”를 원하고 있다는 것이고, 그리고 “비통계가인 팀동료들을 이해시켜 서로 협력하여 문제를 해결할 수 있는 사람”을 구한다는 것이다.

통계학은 산업분야에서도 널리 활용되고 있다. 예를 들면 공산품의 품질을 향상시키기 위한 통계적 방법, 새로운 의약품의 개발과 제조를 위한 통계적 방법, 구체적으로 말해서 Bioassay를 위한 실험계획과 분석등이 우리의 적극적인 참여를 기다리고 있는 통계학 응용분야인 것이다.

이 중에서 공산품의 품질향상을 위한 통계적방법과 신약개발과정에서의 통계적 방법의 활용에 대해서 좀더 자세히 살펴보기로 한다.

통계적 품질관리라고 하면 여기에는 3가지 측면이 있다. 즉, 생산공정에서의 품질관리(On-line

QC), 제품에 대한 발취검사(Acceptance sampling, Sampling inspection), 그리고 세번 째로 제품 성능의 변동이나 부작용을 줄여서 사회에 끼치는 총-손실을 최소화 하기 위해 수행되는 품질관리 활동(Off-line QC)으로 나누어 생각할 수 있다.

이 중에서 마지막 것, 이것은 한 마디로 품질고급화를 위한 개발노력이라고 할 수있을 것이다. 이에 대한 방법으로서 세계적으로 널리 알려져 있고, 우리나라에서도 보급되고 있는 것이 소위 '다구찌 방법'이라는 것이다.

미국의 통계학계 에서는 물론, 품질관리분야 에서도 전연 알려져 있지 않았던 일본인 Taguchi, Genichi(田口玄一)씨가 1980년 미국의 AT&T 의 Bell Lab 에 다녀온 후 십여년이 지난 오늘, 그의 업적은 미국의 통계적 품질공학자나 엔지니어들에 의해서 널리 수용 되어 실천되고 있다.

다구찌씨는 생산계획과 공정의 개선을 위하여 통계적 실험계획법을 사용해서 크게 성과를 올린 사람이다. 이와 같은 그의 방법은 엔지니어들이 오랫동안 기다렸던 것이라고 할 수 있다. 따라서 다구찌방법은 이들에 의해서 크게 환영받았던 것이며, 이로 인하여 '통계학이 통계전문가가 아닌 사람들에 의해서 이해되고 사용되게 되었다'는 것이다. 이것을 다구찌씨의 큰 공적중의 하나로 꼽고 있다.

그러나, 또 한편에서는 다구찌방법이 통계학자나 품질관리 이론가들에 의해서 상당히 비판되고 논쟁되어 온 것이 사실이다. 그의 방법중 어느것은 모호하고 또 오도(misleading) 되기 쉬운 것이 있다고 지적되고 있는 것이다. 그럼에도 불구하고 그의 방법은 세계적으로 널리 수용되어 실천되고 있다. 이것이 올바르게 실행되도록 하는 임무가 통계학 전공자에게 있다고 생각한다. 이와 같은 의미에서 다구찌방법은 학부과정에서도 특강형식등으로 상세하게 강의되고 비판될 필요가 있다고 생각한다.

Toxicology를 毒性學 또는 毒物學이라고 번역한다. 제약업계에서는 독성학적 연구가 필수적이다. 독성학분야에서의 연구에서는 그 원리와 실재를 이해하는데 통계적방법의 필요성이 크게 재고되고 있다. 그러므로 이러한 연구과정에서는 독성학자와 통계학자의 긴밀한 협조가 절대적으로 요청된다는 것이다. 이것을 통계학전공자이면서 독성학자이거나 독성학전공자이면서 통계학자인 연구자가 절실하게 필요하다고 표현하고 있다. 이것은 위에서 언급한바 미국 환경보호처에서의 통계전문가에 요구하고있는 자격과 유사한 것이라고 할 수 있다.

사실 실제분야에서 원하는 통계전문가는 이러한 사람들 이다. 이공학 계열의 어느연구소에서 통계학박사학위 소지자를 찾는 경우, 통계학박사학위를 가지고 있으며 이공학계의 학부출신자라야 한다는 조건을 내세우고 있는 구인광고와 같은 것도 이러한 예일 것이다.

한미간의 파이프라인 프로덕트(Pipeline Product, 未市販物質) 보호협정이 체결되므로서 우리나라에서도 신물질과 신약개발의 필요성이 절실하게 되었다고 한다. 그러나 신약개발에는 엄청난 돈과 시간과 노력이 투입되어야 한다. 이와 같은 사정을 이해하기 위하여 미국에서의 예를 들어본다.

한 제약회사에서 5년 내지 10년에 걸친 연구개발의 결과 하나의 새 약이 인간에게 안전하고 효과적이라는 판단에 도달하였다고 하자. 그러면 인간을 상대로 하는 실험을 거쳐 그 약품에 대한

최종적인 판단을 하기로 결정하게 될 것이다. 미국에서는 인간을 상대로 약효를 실험하기 위해서는 IND (Notice of Claimed Investigational New Drug Exemption)를 FDA(Food and Drug Administration)에 제출해야 한다.

인간을 상대로 하는 약효실험 (임상실험)은 3단계에 걸쳐서 실시된다. 제 1단계 실험(Phase 1 trial)은 탐색적인 실험이라고 할 수 있고, 제 2단계 실험은 제 3단계의 확정적인 실험에 이르는 중간 단계 실험이라고 할 수 있다.

Potency(효력), Dose Range(투약분량 범위), 그리고 Effective Doses(효과적인 투약량)와 같은 제약 파라미터(pharmaceutical parameters)는 제약개발의 여러 단계에 있어서 실험을 통해서 결정되는 것이므로 Bioassay를 위한 실험의 계획과 분석에 대한 통계학응용이 필수적인 것으로 된다. Biological assay (Bioassay)는 이제는 생물체에 대한 약의 효과를 연구하는데 필수적으로 사용되는 방법이라고 널리 받아들여지고 있다. 이와 같은 그 고유의 사용법으로 인하여 Bioassay Statistics는 Chemometrics, Econometrics, 또는 Psychometrics등과 더불어 응용통계학의 하나의 독립된 분과로 다루어지고 있는 것이다.

이와 같은 사정으므로 FDA에서는 신약개발에서 통계적방법을 필수적으로 적용할 것을 요구하고 있고, 이것이 제약회사에서 제출한 최종시판허가신청서(NDA, New Drug Application)의 처리과정에 잘 나타나고 있다.

일단 허가신청서가 FDA에 의해서 접수되면 신청서에 대한 평가(review)가 시작 된다. 이때 주요 Reviewer들은 다섯 분야의 사람들로 구성되어 있고 이중의 한 중요 분야의 담당자가 Statistical Reviewer 이다. 이 사람은 첫째로 적절하고 잘 제어된 통계적 실험계획아래 연구계획이 수립되어 실험이 실시되고 그 결과에 대한 통계적 분석이 합리적으로 이루어졌나를 평가한다. 그리고 더 자세한 정보가 필요하다면 허가신청서에 대한 추가적인 실험과 분석을 요구하기도 한다. 이러한 사정으므로 각 제약업체에서의 통계학에 대한 관심이 재고되고 결과적으로 통계전문가의 참여가 증대되고 있는 것이다.

여기에서 예로 들은 것은 산업분야에서 쉽게 우리 눈에 띄는 응용분야들 이다. 행정분야에서와 마찬가지로 산업분야에서도 통계적으로 해결해야할 여러 문제가 산적해 있다. 여기에는 재치있고 통찰력있는 통계전문가가 필요한 것이다. 우리나라에서는 이러한 일을 위하여 통계전문가를 채용하는 단계에 이르고 있다고는 할 수 없으나 곧 필요성을 느끼게 될 것이다.

#### 4. 통계학계의 전망과 통계학교육

많은 사람들이 통계학의 장래를 밝게 보고 있는 것이 사실이다. 그리고 실제로 이러한 통계학의 중요성과 그 응용의 다양함과 그 성과의 지대함이 여러 분야에서 인식되고 있는 것도 사실이다.

그런데 우리나라에서의 실정은 좀 다르다는 것이 유감이라고 아니할 수 없다. 우리 나라에서 통계학 전공자가 자리를 잡을 수 있는 곳이 얼마나 있는가? 통계학으로 박사학위를 취득한 사람이 갈 곳은 대학교수 자리 말고는 별로 없는 것이 사실이 아닌가? 이것은 통계전공자를 고용할 필요

성을 느끼고 있는 곳이 별로 없다는 우리 사회실정을 반영하고 있는 것이다.

통계분석과 통계적인 판단에 의해서 의사결정을 해야하는 곳에 종사하고 있는 사람들, 그들은 통계학 전공자가 아니면서도, 자신들이 통계학에 관해서 필요한 것은 충분히 소화하여 나름대로 이용하고 있다고 생각하고 있다. 이러한 주장에는 일리(一理)가 있다. 통계계산이 SAS나 SPSS와 같은 통계계산용 패키지를 이용해서 손쉽게 이루어지고 있고, 따라서 통계문제는 컴퓨터가 전부 해결하여 주는 시대가 온 것으로 생각하기 쉬운 것은 어쩔수 없는 일이다. 그러므로 예를 들어 신약개발 연구자나 품질공학도인 엔지니어들은 자기들이 필요로 하는 통계학은 충분히 습득하고 활용하고 있다고 자부하고 있는 것이다. 더 나아가서 문제되고 있는 것에 대한 전문지식도 없는 통계전문가의 개입이 문제해결에 아무런 도움도 되지않는다고 생각하고 있는 것이다.

그러면 통계전문가가 설 땅은 어디에 있는가? 통계전문가가 아니고는 할 수 없는 것은 무엇인가? 오늘날과 같이 컴퓨터가 발달한 세상에서 말이다. 여기에서 통계전문가를 양성하고 있는 통계학과와 존립의의를 생각하지 않을 수 없다.

불확정 법칙(Laws of uncertainty)이라는 것이 물리학적 원리나 화학법칙과 마찬가지로 자연현상을 지배하는 기본실체라는 것이고 통계학은 이러한 불확정적인 현상에 대한 법칙을 인지하고 탐구하는 학문이다. 그러므로 통계학은 사회현상이나 산업기술계에서의 실제현상을 올바르게 이해하는데 필요불가결한 학문인 것이다. 이러한 학문을 탐구하는 통계학과가 어찌 타학과와 비교해서 손색이 있을 수 있겠는가? 그러므로 폭넓은 기초교양을 쌓고 통계학적 사고력을 키우고 이것을 순발력있게 실제 문제에 응용할 수 있는 능력을 갖추도록 하는 것이 통계학과와 교육목표로 되어야 할 것이다.

한국통계학회는 수많은 인재를 거느리고 지속적인 발전을 이루어 나갈 것이다. 통계학을 전공한 사람이 활동할 수 있는 분야도 많이 늘어날 것이다. 아직은 그렇지 않다 할지라도 앞으로 많은 분야에서 통계학 전공자를 찾게 될 것이다. 이것이 문화발전의 추세인 것이다. 앞에서 통계학자의 도움을 원치않는 사람들도 있다고 하였으나, 그러나 그들도 문제를 옳게 이해하고 제대로 해결하자면, 결국은 통계전문가와와 협력이 필요하게 될 것이다. 우리나라에서도 다른 선진국에서 처럼 통계전문가들의 활동범위가 점차적으로 확대되어 갈 것이다.

통계학의 장래는 보증되어 있다 (The future of statistics is secure)고 주장되고 있다. 그러나 이것이 현재와 같은 통계학과와 존속을 보증하는 것은 아니다. 통계학에 관한 교육이 통계학과에서만 실시되고 있는 것이 아니다. 통계학과 출신자와 경쟁관계에 있는 타과출신자가 많이 있다. 산업공학과등 공학계 출신, MBA와 같은 학위를 가진 경영전문가, 그리고 전자계산 전공자들을 생각해 볼 수 있다.

통계학과에서는 통계학적 사고에 숙달되어 통계학의 도움을 필요로 하는 전문분야에 기여할 수 있는 능력을 갖춘 인재를 길러내야 한다. 통계학과 출신은 쓸모있는 유능한 통계전문가로 될 수 있는 바탕을 갖추고 있어야 한다. 그리고 통계학을 전공한 석사, 박사는 통계학 이론바탕 위에 어떤 실제문제에 대해서도 통계적으로 대응할 수 있는 전문가로 인정받게끔 자질을 갖추고 있어야

할 것이다. 그리고 이들의 도움이 얼마나 유효하고 적절한가를 보여주어야 할 것이다.

끝으로, 하바드 대학의 Carl N. Morris 교수의 견해를 들어보기로 한다. 그는, 이상적으로만 된다면 '수레바퀴의 중심축과 바퀴살'과 같은 비유로 통계학과와 대학내에서의 위치를 말할 수 있다는 것이다. ( Ideally, I view statistics departments through a "hub-and-spoke" metaphor.) 통계학과는 수레바퀴의 중심축으로서 대학(university)내의 중심에 위치하고 통계학을 응용·이용하는 다른 여러 학과들은 바퀴의 살에 해당하는 곳에 자리잡고 있다는 것이다. 다시 말하면, 통계학과는 다른 여러 학과에서의 통계문제에 대한 상담에 응하고 문의된 문제를 협의 해결해 주는 협의 중심체로 인식되어야 한다는 것이다. 이를 위하여 통계학과는 봉사하는 자세로 타과에서 개설하는 통계학 강의를 적극적으로 수용하여 담당하고, 이러한 강의를 통해서 또는 기타 모든 기회를 이용해서 타과에서의 통계문제에 관한 상담의욕에 대처할 수 있는 준비가 되어 있다는 과의 태세를 보여주어야 할 것이다. 이와 같이 통계학과는 대학내의 중심에 위치하고 있어 대학내 각 학과에서의 상담에 응하고, 이것을 해결할 수 있는 능력을 갖춘 필요불가결의 중요 기관이라는 것을 대학운영당국이 인식토록 하여야 한다. 우리는 이와 같이 통계학과는 대학구성상 빠질 수 없는 구성요소라는 것에 자부심을 가져야 할 것이다.

## 통계학 교육과 학회 발전에 대한 회고와 전망

토론1 : 동국대학교 통계학과 교수 김 종 호

통계학 교육과 학회 발전에 대한 회고와 전망에 대해서 백운봉 교수께서 역사적으로 그 발전 과정을 상세히 말씀해 주시고 또한 앞으로의 전망에 대해서 잘 설명해 주셨습니다. 본인은 발표자의 발표 내용에 공감하면서 몇 가지 지적과 입장을 말씀드리겠습니다.

### 1. 통계학과의 창설과 통계학회의 발족

통계학이 마치 농·생물학과와 경제학과를 통하여 도입되고 보급된 것 같이 말하고 있으나, 이미 1950년대에 심리학계나 교육학계에서도 통계학 강좌가 개설되어 있었고, 수학과에서도 추측통계학, 확률과 통계 등의 과목으로 개설되어 강의가 있었다. 타 학계 등에서의 통계학 강좌 현황도 더 자세히 소개되었으면 좋았을 것이다.

한편, 통계학이 경제통계학으로 인식되어 왔다는 사실에 동의할 수 없다. 60년대 초에 서울의 4개 사립대학에서 통계학과를 상경대에 신설하고 응용통계학과로서 발전시켜 온 것이 사실이다. 그렇다고 해서 통계학이 경제통계학이라는 인식은 통계학의 위상을 잘못 정립하고 있는데 기인한다고 보여진다. 당시의 사회상이 경제 주도에 경제학 교수가 사회에 많이 관여하는 때로서 통계학이 절대로 필요하였기에 통계학을 겸업으로 하고자 하는 의도에서 상경대학에 통계학과를 설치하였다고 볼 수도 있다. 60년대 초는 경제 발전 5개년 계획이 처음으로 수립되면서 통계학의 필요성에 통계학과 신설이 논의되던 때이다. 무엇을 전공한 교수가 통계학과를 만드느냐에 따라 통계학과 소속이 정해졌다고 볼 수 있다. 수학자 중에서 통계학을 한 사람은 미미하였고 사회에 참여도 없었다. 그러나, 경제학을 전공한 교수는 상대적으로 많았고 사회 참여도 활발한 가운데 통계학의 필요성이 요구되었다. 이러한 사실은 70년대 초에 창립된 통계학회 회원의 구성에서도 잘 나타나 있다. 이제는 통계학의 위상을 자연스럽게 정립해 보는 것이 통계학과와 학회 발전 차원에서 좋은 계기가 되지 않을까 생각할 수 있다.

## 2. 통계학회의 발전과 통계학 교육

통계학회의 발전을 소개하는 과정에서 주로 양적인 발전만 언급하고 있는 것 같다. 통계학회의 발전과 통계학 교육의 내용을 알아보기 위해서는 교과과정(curriculum)의 구성과 현재 상황이 구체적으로 다루어졌어야 되지 않을까 생각한다.

우리의 통계학의 역사가 짧은 것에 비해서 통계학과나 통계학회의 양적 팽창과 발전은 인정할 수 있지만 질적 발전에 대한 설명이 있어야 했다. 특히 통계학 교육은 통계학과의 특성있는 위상 정립과 더불어 통계학에 대한 교육목표가 정해지고, 그에 따른 교육 과정이 작성되어야 한다. 피교육자나 이용자에 합당한 교재의 개발과 교육 방법 등을 연구개발하여 수요자를 충족시켜 주어야 하는데 아직까지는 통계학이 필요하면서도 접근하여 이해하기 어렵고 응용하기가 쉽지 않다고 하는 인식 속에서 그리고 구태의연한 교육 속에서 탈피해야 한다는 지적이 있어야 할 것 같다. 앞으로의 통계학 교육에 대한 논의는 컴퓨터의 이용과 더불어 이용자가 이해하기 쉬운 방법을 진지하게 찾아야 할 것이다.

## 3. 통계 전문가를 필요로 하는 분야

통계적 사고와 지식, 응용기술은 모든 학문(인문, 사회, 자연, 기술)분야와 거의 모든 산업 분야에서 그리고, 정부 행정기관에서 필요하다. 발표자는 몇몇 산업 분야의 경우만 예로 들었다. 기타 다양한 분야의 예도 함께 들어주었으면 좋았을 것이다.

특히, 외국에서는 많은 분야에서 통계적 방법이 이용되고 있고, 많은 통계 전문가가 참여하고 있는데 반하여 우리의 현실은 모든 분야에서 통계학이 중요하고 통계적 방법의 필요성을 인정하고 있으면서도 통계 전문가의 현장 참여가 많이 부족하며, 제도적으로도 통계학 전문 인력이 뚜렷하게 일할 장소를 제공받지 못하고 있는 점을 지적하고 넘어가야 할 것이다.

정부 통계 작성 기관에서 정부 통계를 많이 생산하고 있으면서도 통계학 전문 인력을 확보하고 있지 못한 실정이다. 현재 통계를 작성하는 실무자나 통계학자들의 깊이 있는 성찰이 있어야 한다고 보여진다.

#### 4. 통계학계의 전망과 통계학 교육

현재 상황의 분석과 미래 사회의 전망을 토대로 통계학 및 통계학과의 위상을 재정립하고, 그 위상에 부합되는 통계 교육의 방향과 교과 과정을 제시해야 된다고 보고 싶다.

통계학의 중요성과 그 응용의 다양함과 그 성과의 지대함에 따라 통계학의 장래를 밝게 보는 것은 보편적으로 긍정할 수 있는 사실이다.

물론, 우리나라의 풍토에서는 과도기적으로 문제가 없는 바는 아니지만 양적으로 성장한 통계학과 통계학 전공자들의 노력에 의해서 통계인들의 활동 범위가 점차적으로 확대되어 가는 것을 예측하기는 어렵지 않다.

학회의 다양한 특성적 연구 활동과 같이 통계학 교육도 대학마다 특성을 갖고, 특성있는 교육으로 가는 교과 과정의 개발이 필요하다고 본다. 통계학의 모든 분야의 성격에 따른 교육을 한 대학의 한 학과가 모두 감당하기란 어려운 문제점이있다. 앞으로 통계학과가 그리고, 통계학이 모든 사회 분야에서 기여하기 위해서는 대학별로 특성화할 필요가 있다고 본다. 통계학과가 대학 구성상 필요 불가결한 요소라는 인식에 공감하고 자부심을 갖는 시기가 올 것이다.

## 통계학 교육과 학회발전에 대한 회고와전망

토론2 : 인하대학교 통계학과 교수 구 자 홍

### 1. 통계학과의 인가와 그 성장

1960년대 초기는 시기적으로 5.16 직후로서 한국국민의 남다른 고등교육에 대한 열의와 이에 편승한 사립대학의 난립(亂立)으로 열악한 교육환경에서 대학의 양적팽창이 국민경제에까지 위협하는 지경에 이르렀었다.

이에 당시의 국가재건최고회의는 소위 “ 대학정비령(大學整備令) ”을 내려 각 대학의 학부 혹은 학과들의 일부의 통폐합을 서두르던 시기였으며, 아울러 「경제5개년계획」을 수립하여 가는 시기였다.

이와같은 시점에 새로운 학문분야로 국가경영의 필요에따라 통계학과의 인가 신설된 것이다.

다른한편, 종전에는 내무부 통계국으로 그 명맥을 이어오던 중앙통계부서가 신설된 전 경제기획원 조사통계국으로 이관되었고, 1960년 「인구 Census」 등을 지원하기 위하여 내한한 주한통계고문단(駐韓統計顧問團)의 활동 중 “정부통계의 개선책” “통계간행물의 정비책” “표본조사의 권고” “통계용어집”의 발간등의 자문및 대정부건의는 간접적으로 통계전문가 양성의 필요성을 부각시켰다고 사료되며, 이와같은 분위기와 그 수요는 마침내 통계학과의 설치에 충분한 계기(契機)를 마련해 준 것으로 여겨진다.

그런데, 주제발표자께서도 지적하셨듯이 종전까지의 한국의통계교육은 통계학과의 아닌 타전공분야 예컨대, 농생물학분야, 경제학분야, 의학분야등에서 그들의 도구과목(道具科目)으로서 필요한 내용만 강의 되어온 것으로 사료된다. 당시 필자는 학부와 대학원에서 확률론과 수리통계학을 수강하였고, 또 세부전공분야를 정하기위하여 고심하던 시기였다. 이때 동국대학교 수학과 교수로서 서울대에 출강하시던 홍성해(洪性海)교수와 초대 조사통계국장 강오전(姜五全)교수 사이에 통계학과의 신설과 인가에 대한 자료정리와 그 인가신청서류의 작성과정에 도움을 청하시어 도움과 조력을 한것이 었그제 같이 새로와 진다.

이와때를 같이하여, 고려대학교에서도 통계학과의 신설을 준비 중 이라는 소문을 들었고, 당초에는 이공대의 박태삼 교수께서 통계학과 신청을 위한 자료, 특히 교과과정에 대한 문의와 요청을 홍교수에게 하여왔다는 이야기를 들었다. 그런데, 후일 최종적으로는 정경대학에 제시던 경제학과

김윤환 교수를 중심으로 정경대학에 통계학과가 설치되었고, 성균관대에는 강오전 교수님이 경상대학에 통계학과를 창설하시고, 또 중앙대에도 박찬계 교수님이 중심이 되어 정경대에 통계학과가 창설되었다. 그후 통계학과들의 증설과 발전은 주제발표에 모두 충분히 소개되었다고 사료된다.

그런데 초창기 통계학과들의 단과대 및 학부별 소속은 창설자들의 전공과 의지(意志)에 따라 학과들의 소속이 정해진 것으로 여겨지며, 통계학의 학문적 성격은 덜 고려된 것으로 여겨진다.

그러나 앞서 주제발표자께서도 지적하였듯이 초창기 자연계 통계학과로서는 동국대의 경우 뿐이었고, 모두 경상대와 정경대 등 경상계에 설치되었다. 특히, 연세대의 경우 상경대에 설치하고, 그 학과 명칭도 「응용통계학과」로 표방하였다. 그후 1970년대에 와서 「서울대학교 종합화계획」과 더불어 공대 응용수학과에서 자연계에 「계산통계학과」로 변경하였다.

이것이 계기가 되어 지방국립대들이 자연계 통계학과들의 증설과 발전의 전기가 마련되었다고 볼수있다.

## 2. 통계학회의 창립과 그 발전

1970년초 통계학과들로서는 고대, 동대, 중대, 성대 등 서울시내의 4개 통계학과들 뿐이었으며, 이들 통계학과가 중심이 되어 학생들간에는 매년 체육대회를 돌아가며 개최하고, 상호간의 유대와 친목을 도모하였으며, 이 대회에 격려차 함께참석한 교수들 간에는 학문적 교류와 교과운영 등 공동관심사를 논의 하다가 통계학회의 창립이 절실하여 마침내 1971년 「한국통계학회」를 창립하기에 이르렀다. 그 당시 창설 Member들로서는 주로 통계학과 교수들을 주축으로 경제학과, 경영학과 등의 통계담당교수들, 한국은행과 시중은행 조사부 간부들 그리고 대학부설 전자계산소 스태프(STAFF)들로서 모두 43명이 고작이었다.

그후 본 학회는 오늘 날 「사단법인 한국통계학회」로 장족의 성장과 발전을 거듭하여오고 있으며, 주제발표자도 이미 언급하셨듯이 당시 유일한 학술지 「통계학연구」 외에도 「응용통계연구」, 「통계학회논문집」, 그리고 PROCEEDING등을 연 2회씩 발행하여오고 있으며, 현재는 약 600명 회원의 학술발표 지면으로 통계학과 통계교육 발전에 명실상부(名實相符)한 학술지들로서 국내외의 인정을 받고있다.

### 3. 통계학 교육과 그 전망

이미 언급하였듯이 초창기의 한국 통계학교육을 담당하는 통계학교수와 그 조직인 통계학과는 다분히 창설자의 전공과 세부전공분야에 따라 소속되고, 그분들에 의하여 교육 발전되어 왔으며 또 그 당시로서는 당연한 발전경로였다고 생각된다. 주제 발표자도 언급하였듯이 그후 미국등 해외 또는 국내에서 통계학을 전공한 신진 통계학자들이 자리잡게됨에 따라 통계학교육의 본연의 모습과 합리적이고 균형잡힌 교육으로 그 진로(進路)를 다져오게 되었다.

회고해 보건대, 1970년대까지의 통계학교육과 교과내용은 각 대학 통계학과 나름대로의 커리큘럼(curriculum)을 통하여 강의되어 왔으며, 그 내용도 상당한 차이가 있었던 것으로 사료된다.

「통계학연구」에 투고된 연구논문도 통계 이론적 연구보다는 경제학 또는 경영학 계량경제학등의 연구논문이 상당수 투고되어 실려있는 것으로 미루어 볼 수 있다. 뿐만 아니라, 통계자료 처리와 정보생산및 통계 이용자를 위한 통계조직및 Service교육에 관한 통계직업교육이 체계적으로 이루어지지 않은 것으로 사료되며, 그렇다고 이론 통계학만 연구, 교육된 '시기'라고 할 수도 없겠다. 그 체제나 교육내용이 통계학교육의 과도기(過渡期)였다고 사료된다.

예컨데, 필자의 경우 1964년 동국대 통계학과에 출강하며 가르친 강좌명이, "최소자승법및 오차론", "정차방정식과 보간및 보외법"등 오늘날 통계학과 교과과정에서는 찾아 볼 수 없는 좀 긴 강좌명이었다.

그리고 주제 발표자께서 「통계학연구」 창간호에 투고하신 논문 "통계직업교육에 관한 조사연구"에서도 통계학과 그 역할 중 통계직업교육과 통계실무자 양성및 훈련의 필요성이 잘 언급되어 있다.

즉 "인도태생(印度胎生)의 많은 통계학자가 세계에서 크게 활동하고 존경을 받고 있는 것이 사실이지만, 인도는 통계조직이 빈약(貧弱)하고 인도의 경제,사회문제에 대한 통계학의 역할이 보잘 것 없는 것이 사실이다."라는 인도의 통계학자 Mahalanobis의 표현을 인용하였다. 이는 당시의 우리 나라 통계학교육과 통계조직및 통계학의 국가,사회에 대한 현황을 비유적으로 지적함에 부족함이 없다고 하겠다.

다른 한편, 통계학연구와 통계자료처리의 도구학문으로서의 Computer Science와 Computer교육 훈련도 필수적이라 할 수 있으며, 특히 통계직업교육에서 그 중요성이 절실하다고 하겠다. 회고해 보건대, 1960년대말 한국생산성본부(KPC) 부설 한국전자계산소(당시소장: 이주용)에 FACOM 222가 설치되고, 전산요원의 양성및 훈련실습을 처음으로 시작하여 필자가 동대,통계학과 상급학년생들을 모두 실습비로 입소시켜 같이 교육도 받고 실습도 하였다. 그들은 바로 졸업과 동시에 조사통계국에 전산요원으로 보내어 지금은 통계학 교수 또는 전산·통계전문인력으로 활동 중에 있다.

그후 1970년대에 들어와 이미 언급하였듯이 서울대 계산 통계학과(창설자:최지훈교수)의 창설이 계기가 되어 전산을 도구과목으로 하는 통계학 교육이 체계적으로 이루어지고, 오늘날에는 많은 통계데이터들이 유익한 통계정보로 가공되어 통계 이용자들에게 Service되므로써 정보화사회에 있어서 통계교육의 필요성과 통계인구의 저변확대(底邊擴大)에 그 역할을 유감없이 발휘하게 되어 오고 있다고 본다.

#### 4. 결론

오늘날 21세기를 향한 정부의 「교육개혁안」의 내용중 대학의 학제개편안에 따른 「계열화 통합」이 추진되고 있는 시점에서 통계학과의 학제와 바람직한 교육 방향을 전망하여 보기로 하자.

우선 자연계 통계학과의 이상적인 학제로는 수학과, 응용수학과, 통계학과, 응용통계학과, 전산학과 및 정보과학과 등을 담고 있는 소위 「수리과학부(Division of mathematical science)」 또는 주제발표자께서 인용하신 하버드 대학 Carl N.Morris교수의 견해와 같이, 통계학교육을 위한 학제는 어느 단과대학에 속 함이 없이 대학의 중심 교육조직으로서 「통계과학대학(College of statistical Sciences)」으로 발전시킬 것을 추천하고 싶다.

또한 통계학과의 교과내용(Curriculum)도 수리통계학(확률론포함)을 중심으로 하는 이론통계와 통계적품질관리(SQC),의학통계,조사통계(표본조사론 포함)등 응용통계등을 조화롭게 교육할 수 있는 내용이어야 하겠으며, 특히 전산통계와 전산실습을 통한 통계학의 이론에 바탕을 둔 전산요원의 양성도 가능하도록 교육되어야 한다. 즉 통계학자와 통계실무가 양성에 다같이 주력해야 하겠다.

새로운 통계교육 학제와 내용으로서는 학부 중심 통계학과와 대학원 중심 통계학과로 특성화하고, 통계학은 이론과 방법론을 모두 갖춘 학문이므로 교육내용도 이론적 학문교육과 방법론의 적용을 하는 직업교육으로 구성되어야 한다. 그리고 학부중심 통계학과는 이론보다는 직업교육과 통계실무자 양성이 강조되어야한다.

또 대학원 중심 통계학과에서는 학자와 대학교수의 양성이 그 목적인 만큼 이론통계학이 교육내용에서 강조되어야하며, 이런 인력이 통계실무가 보다는 수요가 적으므로 수요와 그 대학의 여건에 맞추어 선택하여야 한다.

이와 같은 통계학 교육이 이루어지기 위해서는 정부의 참된 교육 개혁의지의 실천과 국민의 새로운 교육의식 전환이 전제되어야 더이상은 표류가 없는 교육제도의 정착을 기대할 수 있을 것이다.

다른 한편 대학 외의 중앙 통계부서들과도 산학협동을 통하여 보다 양질의 통계자료와 통계정보가 작성될 수 있도록 자문과 교류를 추진해 나가야 하겠으며, 아울러 정부에서도 각 부처에 필요

에 따라서는 예컨대, 일본의 문부성 「통계수리연구소(統計數理研究所)」나 후생성 「인구문제연구소(人口問題研究所)」와 같은 연구소들을 만들고, 각 대학에서 양성되는 유능한 통계인력을 유치하여 고급인력으로 키우고, 연구도 시키고, 각종 정책 수립을 위한 통계정보도 생산하게 하므로서 장기적으로 국력신장의 밑거름(肥料)으로 삼았으면 한다.

## 통계기구와 조사방법의 변천

발표 : 연세대학교 응용통계학과 교수 윤 기 중

### 1. 서 론

통계의 정의는 일의적이 아니지만 한 설에 의하면 세 가지 범위로 규정하고 있다. 즉 그 첫째는, 병역, 부역 그리고 과세를 위한 호구조사령에 의해서 조사된 통계로서 대표적인 것이 기원전 13세기 모세가 홍해를 건너 시나이 반도에 상륙하여 인구조사를 실시한 통계(구약 민수기 1장), 그리고 영국의 William Augustin이 왕관을 쓰고 1085년에 전 영국에 재산조사령을 내려 *Domesday Book*을 작성한 통계, 그리고 우리나라의 경우는 이 태조가 등극하고 4년 후인 1395년에 호구조사령을 내려 식년제로 호구를 조사하여 얻은 통계를 들 수 있다. 두 번째는, 17세기에 독일의 Helmstadt대학에서 Conring,H(1606-1681)교수에 의해서 처음으로 국내외의 인구, 토지 그리고 산업 등 정치 및 행정 목적의 통계를 수집하여 강의한 국상학과(Staatenkunde)의 통계를 들고 있다. 세 번째는, Sussmilch,J.P.(1707-1767)와 William Petty(1623-1687)가 제기한 정치산술학과의 통계를 들고 있다. 이들이 현대적인 의미로는 비슷한 것 같으나 원천적으로는 위의 세 분류가 나름대로의 의미를 가지고 있다. 첫 번째 조사령에 의해서 얻은 통계는 부역과 과세를 위한 전수 조사의 통계를, 그리고 둘째의 통계는 단순한 과세나 부역 외에 국력이나 군사력비교를 위해서 유익한 통계인 동시에 통계가 단순한 인구의 개념을 넘어 보다 그 범위가 확대되었다 할 수 있다. 셋째의 정치산술학과의 통계는 대수의 법칙으로 대변될 수 있는 통계로서 통계조사에서 얻은 자료의 성질을 말해주는 동시에 자료의 분석 결과 얻어진 통계라 할 수 있다. 즉 통계조사의 기본법칙으로서 근대통계학의 근본이라 할 수 있다.

이상의 내용을 요약하여 근대적인 용어로 표현하자면 통계는 다음과 같이 두 가지로 요약될 수 있다. 첫째는 인구, 경제 그리고 사회현상에 대한 통계적 정보라 할 수 있고, 둘째는 통계자료의 수집, 분석 그리고 그 자료를 토대로 추측하는 기법이라 할 수 있다. 첫째는 일반적으로 말하는 통계를 가리키고 둘째는 통계의 수집과 분석 그리고 추측하는 기법인 동시에 이러한 기법을 생성해 내는 과학이라 할 수 있다. 이 과학이 통계학(the science of statistics)인데 이러한 학문의 발전도 통계의 발전에 포함시켜 간단히 설명하게 된다. 그러면 통계의 발전은 위의 두 가지의 발전을 기술하는 것이라 할 수 있다. 즉, 첫째의 통계정보라고 한 통계는 통계법에 의해서 지정통계와 일반통계로 분류되지만 통계연보등에는 인구, 노동, 농림어업, 산업, 기업, 사회, 가계, 물가, 금융, 재정, 국민소득, 국부, 산업연관표, 무역, 국제수지 등으로 분류되어 있다. 이상을 편의상 다음과 같이 나누어 설명하게 된다.

1. 인구통계
2. 산업통계

3. 사회통계
4. 가계 및 물가통계
5. 국민계정

위에서 인구통계는 인가와 주택을, 산업통계는 한국표준산업분류에 의해서 집계되는 통계를 말한다. 사회통계는 보건 위생, 주거생활, 환경 및 사회지표 등을 포함한다. 가계 및 물가통계는 도시가계, 농가계조사 그리고 각종 물가지수를 포함한다. 국민계정은 국민소득, 산업연관표, 자금순환표, 국민대차대조표 그리고 국제수지표를 포함한다.

위와 같은 통계의 수요와 공급측면을 보면, 통계의 가장 큰 수요자는 정부이며 그 다음이 민간으로서 기업을 중심으로한 각 산업분야이다. 공급측면에서도 통계를 가장 많이 조사·작성·공급하는 기관은 수요에서와 같이 정부기관이다. 통계를 조사·작성·공급하는 기관을 통계기구라 할 때 이 기구는 정부기구와 민간기구로 나누어진다. 따라서 통계의 발전은 통계기구의 발전과 통계공급방법의 발전으로 볼 수 있는데, 통계 공급의 방법이란 통계의 생산방법으로서 통계조사의 방법을 말한다. 이 통계의 조사는 전수조사와 표본조사로 나누어 볼 수 있다. 이상에서와 같이 통계의 발전은 통계기구의 발전과 통계조사 방법의 발전으로 나누어 과거 50년의 뒤를 돌아보기로 한다.

## 2. 통계기구의 역사

우리 나라 정부의 통계기구는 분산형이기는 하나 주요 일반목적 통계업무를 통계청에서 담당하고 있다는 점에서 분산형인 미국이나 일본에 비하여 집중형에 접근하는 절충형인 듯하다. 우리나라 통계업무의 중심기구는 통계청이다. 또 통계의 생산면에서도 통계청이 중심적 역할을 담당하고 있다. 즉 위에서 언급한 다섯 가지 통계 가운데 인구통계와 산업통계의 대부분 그리고 사회통계의 대부분을 생산공급하는 통계기구가 또한 현재의 통계청이다. 통계청은 통계생산공급의 주류를 점하고 있을 뿐만 아니라 우리 나라 통계의 작성의 기준설정, 조정, 작성의 승인 등을 담당하는 통계작성의 중심기구이다. 이렇듯 통계청의 역사를 돌이켜 보는 것이 우리 나라 통계의 발전역사의 한 부분이 될 것으로 여겨진다.

우리 나라 통계기구를 대표하는 관청통계의 허약성은 처음 출발점부터 몇 가지의 제약이 뒤따르고 있었다. 그것은 일제가 전쟁수행을 위해서 기존의 통계담당 요원을 전쟁수행을 위해 동원하였고 다른 한편에서는 통계가 전쟁수행능력의 평가대상이 된다고 하여 1942년부터 주요통계 특히 산업통계는 발표하지 않았다. 전쟁 이후에도 발표하지 않은 통계에 대한 조사 흔적도 찾아볼 수 없었다. 그것은 조사요원 자체가 없었기 때문이다. 이러한 현상은 우리 나라에서만이 아니고 일본 본토에서도 같은 현상이었다. 즉 중앙정부의 통계기구의 축소와 지방정부의 통계과가 서무과 문서계로 통합되고 그 이하의 관청에서도 통계담당직원이 없어지고 대신 서무계의 직원이 겸임하도록 되어 있었다. 1941년에서 1942년사이에 통제경제체제하에서 모든 물자가 통제대상이었으므로 세삼 통계의 조사나 작성이 필요하지도 않았을 것이다.

우리 나라 정부통계를 관장해온 주무부서는 소속변동에 따라 몇 단계로 나누어진다. 처음에는 미군정, 두 번째가 공보처, 세 번째는 내무부, 네 번째는 경제기획원 그리고 다섯 번째가 오늘의 통계청시대를 맞이한 셈이다. 즉,

1. 미군정, 1945년 8월 - 1948년 8월
2. 공보처 통계국, 1948년 8월 - 1955년 2월
3. 내무부 통계국, 1955년 2월 - 1961년 7월
4. 경제기획원 조사통계국, 1961년 7월 - 1991년 1월
5. 통계청, 1991년 1월 - 현재

우리 나라 통계의 생산은 분산형이므로 중앙정부 각 부처에서 각각 다른 통계를 생산·공급하고 있다. 비정부기관도 정부기관 못지 않게 중요한 통계를 생산·공급해 왔다. 여기서는 정부와 비정부기관의 주요 통계작성 기구의 변천을 보기로 한다. 여기서 논의하고자 하는 주요 통계작성 기구를 다음의 3개 기관만 들고 이들 기구의 통계생산·공급의 과정을 보기로 한다.

1. 농림수산부
2. 노동부
3. 한국은행

## 2-1. 미군정시대

1945년 미군정은 일본 총독부의 행정조직을 이어 받아 새로이 미국 행정조직 형식으로 개편하여, 통계업무부서도 확장하였다. 민정장관산하에 군사국, 경무국, 광공업국, 법무국, 체신국, 교육국, 농상공, 교통국, 보건후생국 그리고 재무국 등 6개의 국 이외에 민정장관 직속으로 비서국(Secretariat 일명 관방 또는 비서실이라 하기도 함)에 8개부서가 있는데 그것은 총무, 외무, 지방행정, 기획, 재산관리, 공보, 군정 그리고 민정과이다. 이 가운데 기획과(Planning Section)에 다시 8개의 담당관실(Sub-section)이 있는데 이 가운데 조사기획담당관실(Research & Plans Sub-section)과 통계담당관실(Census & Statistics Sub-section)이 있었다. 이들 담당관들 가운데 통계담당관은 조사통계업무를, 조사기획담당관은 보고통계업무를 관장하게 하였다.

이러한 기획과는 1946년 3월 군정기구가 개편되어 부처제로 되면서 비서국(일명 관방)을 서무처에 흡수, 통합하였다. 과거 기획과에 소속되어 있던 조사기획담당관실의 통계관실(census & statistics)은 통계서(統計署)로, 그리고 조사기획관실(research & plans)은 조사연구서(調查研究署)로 바뀌어 졌다. 이 통계서는 1실 5개과로 구성되었다. 즉 서무처 통계서에는 통계관실, 서무과, 인구정태과, 가계과, 인구동태과 그리고 노동력과로 직원은 약 800명정도였다고 한다. 1947년 2월에 안재홍을 민정장관에 임명하면서 미군정이 남조선과도정부로 개칭되고, 1947년 5월에는 남조선과도정부 서무처 통계서로 명칭이 바뀌어지고, 서장에 한국인 정철섭(鄭喆燮)이 임명되고 기구도 다소의 변동이 있었다. 과거에 1실 5개과가 1실 4개과로 되고 직원수도 800명이 500명으로 감소되었다. 그것은 인구동태과가 보건후생부로 이관되고 노동력조사과는 노동부로 이관되었기 때문이다. 통계서에는 통계관실, 서무과, 관리과, 인구정태과 그리고 가계조사과가 있었다. 이 통계서는 1947년부터 국민의 의식생활 필수품의 수급계획을 위해서 15세 이상을 국민등록표에 기재·등록케하고 이를 집계하는 일이었다. 이 때의 통계는 해외귀환동포의 수, 월남동포의 수 그리고 일본인의 귀환자 수가 기록되었다. 한편 농림수산통계는 1947년 6월에 군정청법령 143호로 농림통계위원회를 설치하고 그 통계의 조사관장기관으로 농림부에 조사통계과를 설치하여 각급 행정기관에 통계직원 1,928명을 배치하였다.

서무처 통계서의 인구동태과가 보건후생부로 이관되어 한때는 위생통계국으로 격상되기도 했지만 얼마 되지 않아 위생통계국은 생정국(生政局)으로 개칭되고 인구통계업무는 생정국 보건과 인구통계계가 담당하게 되었다. 여기서 관장하는 업무는 군정청 보건위생국의 규정과 양식에 따라 출생, 사망, 사산, 혼인과 이혼에 관한 사항을 신고케하고 이를 집계하는 일이었다. 이러한 업무는 군정청의 위생통계조사규칙과 인구동태조사규칙에 따라 등록 또는 신고케하여 이를 집계하는 것이었다. 1947년 11월에 보건후생부의 직제 개편에 따라 생정국은 예방의약국이 되었고, 보건과의 인구통계계는 예방의약국의 통계과로 격상되어 인구동태통계를 담당하게 되었다.

군정 하에 조선은행은 조사부에서 각종 경제통계를 조사·작성·공표하고 있었는데 이 가운데 특기할만한 일은 금융통계는 물론이고 도매물가지수와 서울, 대전 그리고 부산 3개도시의 소매물가지수를 작성·공표했고 또 재무부와 기획처와의 협력으로 정부수립 이전까지 국민소득도 추계·공표 하였다. 기획처는 생산국민소득을, 재무부는 분배국민소득을, 그리고 조선은행 조사부에서는 지출국민소득을 각각 나누어 추계하였다.

## 2-2. 공보처시대

1948년 8월 15일로 정부가 수립되면서 정부는 내무부, 외무부 등 11개부와 국무총리직속으로 총무처, 공보처, 법제처 그리고 기획처 등 4개처를 두기로 했다. 이 당시의 정부조직법에 따라 통계업무는 공보처소관으로 이관되고, 이해 11월 4일에 대통령령으로 공보처 직제가 공포됨에 따라 통계업무는 공보처 통계국에서 담당하게 되었다. 통계국에는 통계관실, 서무, 기획, 국세조사 그리고 인구조사과의 1실 4개과로 조직되었다. 미군정시대에 보건후생부로 이관되었던 인구동태업무도 통계국으로 되돌아왔다. 이 때의 직원은 약 1,000명 정도라고 한다. 건국 이후 통계수요의 폭주에 비하여 생산·공급은 그에 미치지 못하였다. 1948년 12월 대통령령으로 제1회 총인구조사령을 공포하고, 1949년 5월 1일 영시를 기점으로 총인구조사를 그리고 호구조사규정에 의거 연말상주 인구조사를 실시하였다. 1949년 5월 1일에 조사한 총인구조사는 1년뒤에 6·25동란이 발발했지만 당시의 인구는 20,166,756으로 또 1955년의 간이인구조사 결과는 21,502,386으로 공표 했다. 인구동태조사령에 의거 6·25로 중단되었다가 1952년부터 재개하게 되었다.

1950년 3월에는 통계국의 직제개정으로 통계관실과 서무과가 폐지되고 3개과 즉 기획과, 국세조사과 그리고 인구조사과의 3개과로 축소되고 또 인원도 1,000명에서 600명으로 감소되었다.

한국은행은 한국은행법에 따라 통화, 금융업무, 재정, 물가, 임금, 생산, 국제수지 등의 통계를 수집·작성·공표 하도록 되어 있다. 조사부의 직무수행을 위하여 정부와 자연인 그리고 법인에 대하여 필요한 자료와 정보를 요구할 수 있게 하였다. 그 결과 1948년에 최초로 조선경제연보를 그리고 1949년부터는 경제연감을 연간으로, 물가총람은 부정기적으로 발간 공표하고 1947년부터는 조사월보를 월간으로 공표해 왔다

## 2-3. 내무부 통계국시대

정부수립 이후 두 번째의 정부조직법이 1955년 2월에 개정되어 공보처 통계국은 내무부로 이관되었다. 이에 따라 공보처시대와 같이 기획과, 국세조사과 그리고 인구조사과의 3개과를 답습했으나 인원은 600명에서 400명으로 감소하였다. 그 후 1957년 6월 대통령령에 의한 내무부직제 개정으로 과거(1950년)에 없어진 통계관제를 부활시키고 과의 명칭변경과 더불어 한 개의 과를 늘

였다. 즉 통계관은 서기관과 사무관 2명으로 부활확대시키고 과는 통계기준과, 인구통계과, 경제통계과 그리고 제표과로 확대시켰다. 즉 제표과가 신설된 셈이다. 통계국의 기구는 확대되었으나 공무원수는 대폭 감축되었다. 즉 정규직원 58명과 임시직 195명 합계 253명으로 감소되어 경제기획원시대를 맞게 되었다.

#### 2-4. 경제기획원 조사통계국시대

1961년 7월 22일의 직제개편에 따라 내무부 통계국은 신설 경제기획원으로 소속이 변경되었고 편제는 전과 다를 바 없었다. 1962년 6월의 정부조직법개정으로 통계국에 조사분석과를 추가 신설하면서 통계국의 명칭도 조사통계국으로 개칭하게 된다. 1963년 12월 조사통계국의 경제분석과를 종합기획국으로 편입시키고 조사통계국은 경제기획원의 외국으로 분리된다. 경제의 확장기이기도 하지만 조사통계국도 경제기획원 조사통계국시대에는 양적으로나 체질 면에서 눈부신 변화를 거듭하였다.

우선 체질 면에서는 1962년 1월의 통계법 제정을 기점으로 모든 통계조사를 위한 법적 기반을 견고히 한 점을 들 수 있다. 경제개발 5개년 계획을 되풀이하면서 늘어나는 통계수요에 따라 통계의 생산·공급도 증가하지만 이러한 통계의 생산·공급을 위한 법적 뒷받침을 견고히 한 것으로 평가된다. 다시 말하면 통계법을 제정하여 통계의 종류를 규정한 점, 동시에 신고의 의무와 벌칙에 의해서 통계의 진실성을 확보한 점, 통계법 시행규칙, 통계자료 공표협의 규칙, 인구 및 주택센서스 시행규칙, 인구동태조사 시행규칙 등 통계조사의 법적 기반을 1960년대 이미 완성한 것이다. 종합적으로 지정통계시행규칙으로 19개의 규칙을 공포했고, 통계관계법령, 인구관계법령 기타 조사관계법령의 제정 개폐가 대체로 1970년경에 거의 완성된 듯하다.

한편 통계기구의 확장을 보면 1961년 처음 내무부에서 경제기획원으로 소속이 변경될 때는 통계기준과, 인구통계과, 경제통계과 그리고 편수과의 4개과이었으나 1968년 2월의 직제개편에서는 통계심의관의 신설 외에 소비통계과와 자료처리과가 증설되어 1실 6개과로 확장되었다. 1971년 5월의 직제개편에서는 6개과가 8개과로, 또 1975년 8월의 직제개편에서는 통계조사관과 자료관리관의 신설 외에 자료관리과 등 2개과와 8개의 지방사무소가 신설되었다. 1981년 11월에는 통계조사관과 통계심의관 그리고 자료관리관은 통계계획관과 통계관리관으로 통합하는 한편 지방사무소 8개소를 11개소로 확장하였다. 통계법에 의거 통계위원회를 두고 16개의 분과위원회를 구성하기도 하였다.

#### 2-5. 통계청시대

1991년 1월에 경제기획원 조사통계국은 경제기획원 외청으로 승격하여 통계청시대를 맞이하게 된다. 조직은 청장 밑에 기획국, 조사국 그리고 자료관리국과 1개의 연수원 1 담당관 16개과 그리고 11개 지방사무소와 5개의 지방출장소로 확장 발족하게 되었다. 1993년에는 경제기획원과 재무부가 통합되어 재정경제원으로 개칭됨과 동시에 자동적으로 재정원의 외청이 되었다.

## 2-6. 농림수산부

정부수립과 동시에 농수산부의 통계과는 처음 장관비서실에 소속시켰으나 1950년의 직제개편 때에 차관직속으로 변경되었다. 경지면적, 농가호수 등과 같은 농업기본통계는 통계과가, 식량작물생산통계는 농정과에서 그리고 가축통계는 축정과에서 관장하고 있었다. 1955년 2월의 직제개편에서는 통계과가 차관실소속에서 농정국소속으로 변경됨과 동시에 과명도 조사통계과로 개칭되었다. 그리고 조사통계과에 통계계와 조사계의 두 계를 신설하였다. 또 농정과에서 관장하고 있던 식량작물 생산량조사업무가 조사통계과로 이관되었다. 1958년부터 1960년에 실시예정인 농업센서스업무가 조사통계과 소관업무로 귀속되면서 조사통계과의 조직도 확대하게 될 조짐이 보였다. 이 무렵 조사통계과에서는 농업센서스 외에 농산물생산량 조사방법에 대하여 논의가 진지하게 논의되고 있을 때였다.

1961년 7월에는 종래의 조사원 33명을 해임하고 신규로 80명을 채용, 농가경제조사구에 배치하였다. 1962년 1월에는 다시 80명을 신규로 채용, 조사구에 배치하는 한편 1961년에 채용한 80명은 본부로 진출 발령하는 인사로 통계업무부서에 인원이 확충되는 경향이였다. 그 결과 1964년 10월에는 직제개편으로 조사통계과에 관리계, 조사계 그리고 통계계가 새로이 신설되었다. 이와 같이 조사통계과 규모가 확장됨에 따라 1968년에는 축산국에서 관장하던 가축통계업무를, 그리고 1970년 1월에는 농정과 소관업무이었던 양곡소비조사업무가 조사통계과로, 모두 이관되었다. 업무확장에 따라 1966년 1월에는 신규로 200여명을 채용, 시도에 농가조사요원으로 배치하고 그 후 1968년 1월에는 1,667명이나 신규로 채용하여 각 읍면과 시도에 배치하였다. 또 1971년에는 잡급적인 통계요원과 조사원 2,000여명을 모두 정규직으로 승진발령하여 농수산통계기구의 조직을 정비하게 된다.

이와 같이 인원도 확충하고 조직도 확대됨에 따라 1971년 11월의 직제개편에서는 조사통계과가 국장급수준의 조사통계관실로 격상되는 동시에 조사통계관 밑에 경제통계담당관과 작물통계담당관이라는 직제를 신설했다. 그 후 1973년 1월의 직제개편에서는 조사통계관을 농림통계관으로 개칭했다가 다시 농업통계관으로 변경하게 되었다. 1974년 1월의 직제개편에서는 9개 도에 농업통계업무소를 새로이 신설함과 동시에 시 군에는 출장소를 설치하였다. 오늘에는 통계관리담당관, 농업통계담당관, 유통경제통계담당관 그리고 수산통계담당관과 9개의 지방통계사무소와 141개의 출장소를 구비한 방대한 조직으로 확대되었고 인원도 통계관실에 80여명과 지방에 2,000여명의 방대한 조직을 가지게 되었다.

이상과 같이 통계조사기구를 정비하면서 1960년과 1970년에는 농업센서스가 실시되었고 또 1975년부터는 간이 인구센서스와 같이 농업에서도 간이센서스를 실시하기 시작하였다. 이와 같이 조사를 거듭하면서 조사대상도 확장되었는데 1960년의 센서스에서는 경지면적 300평이상의 농가만을 대상으로 한데 비하여 1970년에는 고등원예, 특용작물, 과수, 가축, 양잠 그리고 양봉까지 포함시켰다.

통계기구와 조사대상을 확대하여 오늘날 발표되고 있는 농수산통계는 대체로 다음과 같다.

농가경제와 농산물생산비  
농수산물유통통계  
농축산물생산통계  
수산통계

농산물무역통계  
농지기반통계  
농업자제통계  
입업통계

세분해보면 농림수산부에서 87종; 산림청에서 40종, 농촌진흥청에서 21종 그리고 수산청에서 17종의 통계를 생산 공급하고 있다.(한국통계발전사2 PP.237-243)

## 2-7. 노동부

노동통계는 노동청이 발족하기 전인 1963년 이전에는 보건복지부의 전신인 보건사회부에서 생산·공급해 오다가 1963년 9월의 정부조직법 개정 에 따라 노동청이 발족하게 되자 기획관리실에 노동통계담당을 두고 노동행정에 필요한 기초통계를 조사·작성하는 실험단계를 거쳐 1968년 4월부터는 사업체노동실태조사, 매월노동통계조사 그리고 직종별임금 등을 조사하여 노동행정에 이용해 오고 있다. 이와 같이 노동통계영역의 확대로 1968년 6월의 직제개편에는 통계담당이 노동통계담당관으로 승격되었다. 1970년의 직제개편에서는 과수준의 노동통계담당관실이 기획관리실 소속에서 신설된 인력개발관실 소속으로 변경되었고, 1973년의 직제개편에서는 고용통계과와 임금통계과로 확대 개편되었다. 또, 1976년 3월의 직제개편에서 고용통계과와 임금통계과가 통합되어 노동통계과로 축소된 듯 했으나, 다음 해인 1977년에는 인력개발관실에 과수준의 전산실을 신설하였다.

현재노동통계는 지정통계로서 다음과 같은 지정통계를 생산·공급하고 있다.

사업노동실태조사(연도별)  
매월노동통계조사(월)  
직종별임금조사(연)  
노동생산성통계조사(월)

일반통계로써는 노동력유동실태조사, 구인구직취업현황조사보고 등 15종이 있다.

## 2-8. 한국은행

한국은행은 중앙은행으로서 한국통계의 생산·공급에 중요한 역할을 담당해 왔다. 1910년부터 한국의 물가지수를 최초로 작성, 우리 나라 은행권의 가치변동을 평가해 왔고 일찍부터 국민소득을 정부와 같이 추계하기도 하였으며, 한국의 국민계정을 작성해 오기도 하였다. 한국은행은 한국은행법에 의해서 통화신용에 관한 적절한 정책 수립에 필요한 정보 제공을 위해 통화와 은행업무, 재정, 물가, 임금, 생산, 국제수지 등에 관한 통계자료를 수집하고 이를 통계월보로 발행해야 하는 의무가 있고, 또 국가의 통계업무 개량과 경제통화, 은행업무에 관한 대중교육과 계몽을 촉진하기 위해 정부기관과 협력할 의무가 있다. 한국은행은 이러한 통계업무를 위하여 필요한 자료와 정보를 정부와 자연인 또는 법인에게 요구할 수 있는 권리를 법률로써 보장받고 있다. 이와 같이 한국은행의 통계업무는 법에 의한 고유업무로서 그 통계 업무의 영역은 통화금융, 재정, 물

가, 경기, 외환, 경기 그리고 국민계정 등 광범위하다.

한국은행은 이러한 업무수행을 위하여 현재 조사2부에 금융통계과, 산업통계과, 물가조사과, 국민소득과, 연관분석과 등 5개과를 두고 있다. 금융통계과에서 관장하고 있는 통계업무는 각 개념의 통화, 본원통화, 국내여신, 통화서베이, 금융서베이, 각 금융기관의 예금과 대출통계 그리고 어음, 수표교환, 부도상황, 요구불 예금의 회전을 등의 통계를 작성하여 공급하는 한편, 자금순환표와 국제수지표도 작성하고 있다. 자금순환표는 한국은행에서 1963년 처음 작성하였는데 역사상 최초의 것은 1952년 미국의 자금순환표이다. 그로부터 11년이 지난 1963년부터 한국은행 금융통계과에서는 한국의 자금순환표를 매년 분기별로 작성하여 발표하고 있다. 국제수지표는 1955년 우리 나라가 국제통화기금(IMF)에 가입한 이후 매년 분기별로 국제수지표를 작성·공표하고 있다.

국민소득과에서는 한국의 국민소득을 생산, 분배, 그리고 지출의 3면에서 추계·발표하고 있다. 한국에서 처음 국민소득이 추계된 것은 1937년에 한 개인의 연구결과로 발표된 바가 있었다. 1941년부터는 식민지 정부의 사세당국에 의해서 분배측면에서 국민소득을 추계해왔다. 그에 의하면 다음에서 보는바와 같다.

한국의 국민소득 (단위: 억円)

연 도	국민소득
1937	27
1941	30
1942	35
1943	40
1944	55
1945	70

자료: 통계청, 한국통계발전사(Ⅱ) 1992, P.530

해방후 1947년과 1948년 기획처 경제기획국과 재무부 사세국 그리고 한국은행 조사부의 3자 합작으로 생산, 분배 그리고 지출면에서 처음으로 국민소득을 추계 하였다. 당시 기획처 경제기획국에서는 생산국민소득을, 재무부 사세국에서는 분배국민소득을, 그리고 한은 조사부에서는 지출 국민소득을 각각 추계 하였는데 그 결과는 다음과 같다.

남한의 국민소득추계 (단위: 백만圓)

연 도	기획처 경제기획국	재무부 사세국	한국은행 조사부
1947		91,251	444,423
1948	670,683	120,000	692,428

자료: 통계청, 한국통계발전사 (Ⅰ) 1992, P.194

정부는 국민소득추계를 위하여 1951년 국민소득 조사위원회를 구성하여 체계적인 추계방법을 연구하고, 또 국제연합한국부흥단(UNKRA)의 도움으로 1955년에는 기획처 경제기획국과 한국은행이 협력하여 1953년 국민총생산액을 국민계정방식에 따라 경상가격기준으로 추계 하였다. 이때 재무부 사세국에서도 예산설명자료로써 요소비용에 의한 국민소득을 추계 하였고, 또 미국의 원조기관이었던 경제조정관실(OEC)에서도 미국 달러화기준으로 한국국민소득을 추계 하였다. 한국의 국민소득을 3개 기관에서 각각 달리 추계 하는 것을 본 재무부 사세고문단이 3자의 국민소득의 추계방법을 검토하고 한국의 국민소득은 한국은행에서 후계할 것을 권고하였다. 이러한 과정을 거쳐 한국은행은 1958년에 1953년부터 1957년까지의 한국국민소득을 UN의 권유에 따른 통일방법으로 추계·발표하는 동시에 1955년 불변가격의 국민소득으로 연도별 발전상태를 가늠케 하였다. 1961년에는 한국은행이 국민소득의 추계업무를 담당하던 조사부 제1통계과를 확대·개편하고 국민소득과를 신설하여 국민소득추계업무를 전담케 하였다. 한국의 국민소득은 1937년부터라 할 수도 있지만 남북의 분단과 국민계정의 통일적인 방법으로 추계 되었다는 점에서 그 시계열은 1953년부터 유효하다 할 수 있다.

연관분석과에서는 1964년에 1963년도의 43개부문의 한국의 산업연관표를 작성·공표하고 그 후 1966년도, 1970년도, 1975년도 그리고 1980년도의 실측표와 1968년도, 1973년도, 1978년도, 1983년도의 연장표를 작성·공표 하였다. 1985년부터는 신국민계정체계(신 SNA)로 개편하게 되었다. 이것은 1941년 미국의 레온티프(W.W.Leontief)가 미국경제의 구조(The Structure of American Economics)를 발표한 이래 미국에서는 1947년도부터 산업연관표를 작성하였고, 한국에서는 1957년에 처음으로 부흥부 산업개발위원회에서 작성했지만 당시 계산능력의 제약으로 실용화에는 미흡하였다. 즉 처음으로 작성했을 뿐이지 한국경제의 분석과 경제발전계획에 실용화할 수 있는 산업연관표는 한국은행의 1963년도의 그것이 최초의 것이라 할 수 있다.

국부통계도 연관분석과에서 관장하고 있는데 최초의 국부조사는 1969년4월에 실시했고 기준은 1968년 12월말로 했으며, 당시의 조사계획은 경제기획원 조사통계국(지금의 통계청)이 주관은 하되, 각 기관이 나누어 조사하였다. 즉 조사통계국은 가계자산과 내구소비재를, 한국은행은 정부, 국영기업, 그리고 공기업의 고정자산을, 산업은행은 법인기업의 고정자산을, 그리고 중소기업은행은 중소기업의 고정자산을 각각 분담·조사하였다. 그 조사 결과는 1970년에 공표예정이었으나 오늘에 이르기까지 1968년도의 국부조사결과는 공표 되지 않고 있다. 그 후 1978년부터 10년마다 국부조사를 실시하고 그 결과가 공표 되고 있다. 이와 같이 한국은행에서는 국민소득, 산업연관표, 국제수지표, 자금순환표, 그리고 국부통계에 의한 국민대차대조표의 작성을 담당하고 있다.

이 외에 물가지수는 한국에서 최초로 중앙은행인 한국은행(1910-1947년까지는 조선은행)이 1910년부터 작성·공표하고 있다. 과거에 생계비와 서울소매물가를 조사하여 생계비지수와 서울소매물가지수를 작성·공표해오다 1965년부터 통계청으로 이관시켜 소비자물가지수로 개편하여 작성하고 있다. 현재 임금에 관련된 통계는 노동부에서 관장하고 있으나, 임금통계는 1910년부터 한국은행에서 조사해왔다. 즉 1910년부터 임금을 조사하기는 했으나, 1925년부터 월별평균노임지수를 작성하여 통계월보에 1942년까지 수록, 공표해 왔다. 또 1948년부터는 노임지수를 두 가지로 작성했는데 그 하나는 1910년 7월 기준 100으로 한 것이며, 다른 하나는 1936년의 연 평균치를 기준 100으로 하는 노임지수이다. 후자인 1936년을 기준으로 한 지수는 1942년 이후 중단되었던 것을 1948년에 소급하여 중단기간의 지수를 작성하여 시계열로 연결시켰다. 물가통계과에서는 도매물가지수의 연속계열인 생산자물가지수, 수출상품가격지수 그리고 수입상품가격지수를 월별로 작성·

공표하고, 산업통계과에서는 경기지수를 작성하는 한편, 기업경영을 분석하고 그 결과를 연도별로 공표하고 있다. 이상에서 본 바와 같이 한국은행은 우리 나라의 주요 경제통계의 조사와 분석의 일익을 담당하고 있다.

### 3. 통계조사

앞에서 우리 나라에서 작성 공표하고 있는 통계를 인구통계, 산업통계, 사회통계, 가계 및 물가통계, 그리고 국민계정으로 대별해 보았다. 여기서는 이들 통계가 어떻게 조사되어 왔는가를 살펴보고 통계조사방법상의 전환기를 찾아 보기로 한다. 그리고 통계조사방법의 변화 내용과 그 변화의 동인을 찾아 보기로 한다. 이들 통계를 얻기 위한 조사대상에 따라 나누어 보면, 첫째로는 가구를 대상으로 하는 가구조사, 둘째는 사업체를 대상으로 하는 사업체조사, 그리고 셋째로는 농가, 농지, 생산물을 조사대상으로 하는 복합적인 농업조사로 나누어 이들 조사의 방법의 변천을 보기로 한다. 여기서는 먼저 조사방법의 변천을, 그 다음으로 조사방법의 전환기와 그 내용 그리고 세 번째에는 변화의 동인을 찾아보기로 한다.

#### 3-1. 조사방법의 변천

##### I 가구조사

가구를 조사대상으로 하는 조사는 생계비조사, 도시가계조사, 농가계조사, 경제활동인구조사, 그리고 인구주택총조사가 있다. 인구주택총조사는 주기적으로 반복하는 것이므로 조사 방법상의 변화는 없고, 또 농가계조사는 농업조사에서 논의하게 되므로 이 두 가지 조사에 관해서는 여기서 제외시킨다.

A) 도시가계조사(생계비조사): 1950년 6·25동란시 부산에서 한국은행과 기획처 통계국의 협력으로 도시근로자가구의 생계비조사를 처음으로 실시하였다. 이 때 조사대상가구수는 봉급자가구 90가구와 근로자가구 30가구 등 모두 120가구를 대상으로 하였다. 이 조사대상 120가구는 비확률 표본 즉 비유의 표본으로 선택했다고 한다. 서울로 환도한 이후에도 서울과 부산의 도시가구를 대상으로 임의표본으로 조사해오다가 1959년에 조사명과 조사방법을 개편하였다. 즉 조사명은 생계비조사에서 도시가계조사로 바꾸고, 표본도 그 크기를 200으로 증가시키는 한편 임의표본추출법에서 확률표본추출법으로 전환하였다. 1963년에는 조사대상도시를 32개 도시로 확대하는 한편 표본크기도 1,700으로 증가시키고, 또 표본의 추출방법도 확률표본추출법에 의해서 조사가구를 선택하였다. 그 후 도시가계조사는 1965년도에 통계청(당시 경제기획원 조사통계국)으로 이관되고 표본도 다목적표본설계에 따라 조사가구를 선택하게 되었다. 1965년 이후 조사방법의 변화는 없었고 다만 표본크기가 변한 것과 다소의 조사항목의 변동, 그리고 집계결과표의 변화만 있을 뿐이다. 이러한 변화는 신SNA에 따르기 때문이다.

B) 경제활동인구조사: 경제활동인구조사는 본래 1957년 노동력조사로 출발하였다. 내무부 통

계국에서 1955년의 간이인구센서스자료에 의거 450개 표본구에서 1,000분의 1의 추출률로 확률표본 4,000가구를 대상으로 조사해 오다가 1961년에 노동력조사를 경제활동인구조사로 개칭하고, 476개의 표본구에서 5,100가구를 확률표본으로 추출하여 조사하였다. 통계청에서는 1969년에 다목적표본설계에 의해서 도시가계조사, 경제활동인구조사 그리고 인구동태조사 대상가구를 선정하고 있다.

## II 사업체조사

사업체를 조사대상으로 하는 조사는 주요기업체실태조사, 광공업동태조사 그리고 광공업총조사, 건설업통계조사, 도소매업통계조사 등 여러 가지 조사가 있으나 여기서는 1960년 이전부터 조사한 통계만을 대상으로 한다.

A) 광공업총조사: 1955년 한국은행조사부가 <광업 및 제조업사업체조사>라는 명칭으로 조사했으나 1958년에는 한국산업은행으로 이관되어 한국산업은행이 이 조사를 실시하였다. 그 후 1960년도의 조사는 통상산업부(당시의 상공부)와 산업은행이 공동으로, 그리고 1963년과 1966년, 1968년은 통계청(당시 경제기획원 조사통계국)과 한국산업은행이 공동으로 조사하였다. 그 후 1973년도의 조사는 통계청 단독으로 조사하는 동시에 그 명칭도 <광공업센서스>라고 개칭하였다. 이 명칭이 <산업센서스>로 했다가, 다시 <광공업총조사>라고 변경하였다. 1983년부터는 매 5년마다 정기적으로 조사하는 총조사로써 조사의 범위는 확대되고 있으나 그 방법은 아무런 변동이 없다.

B) 광공업동태조사 : 이 조사는 한국은행이 1954년 산업생산지수작성을 위해서 통상산업부(상공부)의 협력을 얻어 <주요기업체실태조사>명칭으로 1968년까지 조사하다가 1968년에 이 조사와 지수작성업무를 한국산업은행으로 이관시켰다. 조사대상업체의 선정은 생산액이 큰 업체부터 차례로 적당한 수의 사업체를 선택하여 조사하였다. 이와 같이 임의표본에 의해서 조사하다 1958년과 1960년의 광공업센서스조사자료에 의거하여 다단계확률표본추출방법을 설계하여 1962년 3월부터는 <광공업표본조사>로 개편하여 매월 조사하게 된다. 이것이 <광공업동태조사>로 개칭된 것은 1970년부터이다. 1970년도는 한국산업은행과 통계청(조사통계국)이 공동으로 조사했으나 그 다음부터는 통계청으로 완전히 이관되었다.

## III 농업조사

농업을 대상으로 실시하는 조사는 여러 가지가 있지만 여기서는 1960년 이전부터 조사하던 것으로 중요하다고 인정되는 몇 가지 조사만을 보기로 한다.

A) 농업총조사: 매 10년마다 실시하는 이 조사는 UN/FAO의 권고에 의해서 우리 나라는 1960년에 처음으로 실시하였다. 이 조사는 농경지 300평 이상을 직접 경작하는 가구를 대상으로 80여개 항목을 조사하는 전수조사이다. 조사할 때마다 조사항목수는 변경되는데 1990년의 경우에는 182개 항목이었다. 이 조사의 내용은 대체로 농업사업체조사와 농업지역조사로 나누어지는데 전자의 농업사업체조사는 농가, 농가인구, 농경지, 작물, 가축, 농기계를 그리고 농업지역조사는 지역현황과 지역경제여건에 관한 조사이다. 조사요원은 국민학교 교사로 하는 경우도 있고 이장, 새마을지도자와 독농가 20,000여명을 동원하여 조사한다.

농업총조사의 중간연도인 매 5년, 1965, 1975, 1985 그리고 1995년에는 <간이농업조사>라는

명칭으로 70여개 항목을 표본에 의해서 조사하고 있다. 이 조사의 표본은 <농업기본통계조사>의 표본조사구 약 5,000 조사구에 있는 농가가 된다.

B) 농업기본통계조사: 이 조사는 농가의 특성과 농가인구의 특성을 조사하는 것인데, 이 때 농가의 특성은 농업전업인가 또는 겸업인가 그리고 농경지의 규모는 어느 정도인가를 뜻한다. 농가인구의 특성도 조사대상인 그 특성은 성, 연령, 취업상태 등이다. 그 조사결과는 지대별 전업농가수, 겸업농가수, 경지규모별농가수, 성별, 연령별 그리고 취업상태별 농가인구로 집계된다. 이러한 통계는 정부의 농업정책수립을 위해서 중요한 불가결의 통계이므로 정부수립 이전부터 매년 12월말 기준으로 행정계통의 보고통계로 수집해 왔었다. 행정계통의 보고통계가 1974년부터 표본조사로 전환했다. 이 때의 표본은 1970년의 농업총조사결과에 의해서 설계되었는데 그것은 농업총조사의 조사구 25,762구 가운데 10분의 1의 추출률로 선택했기 때문에 2,576개구가 된다.

C) 농가경제조사: 1953년에 한국은행이 국민소득추계를 위한 기초자료를 얻기 위하여 7월 1일부터 표본의 유의선정방법으로 전국에서 300호의 농가를 선정하여 조사 해왔다. 조사의 방법은 표본농가에 일계부를 배부하고 자계식으로 기재케 하고 월말에 일계부를 회수하는 것이다. 조사원이 표본농가를 방문하여 기재의 방법을 지도하고 기재능력이 없는 농가는 타계식으로 기재해 준다. 조사항목은 표본농가의 특성, 수입지출, 노동시간 그리고 그 농가의 재산상태이다.

1954년에는 농수산부와 한국은행이 각각 조사해 오다 그 후에 한국은행의 조사는 중단하고 농수산부에서만 조사해 오고 있다. 1961년에는 1960년의 <농업총조사>자료를 이용하여 확률표본추출법을 설계하여 과거의 임의표본에서 확률표본에 의한 조사로 전환했다. 이러한 확률표본의 설계는 <농업총조사>와 <농업간이조사>를 실시할 때마다 변경한다. 표본도 변동하지만 모집단도 변하기 때문에 총조사와 간이조사가 실시되면 그 조사결과에 의해서 표본설계를 다시 하게 된다.

D) 양곡소비량조사: 이 조사는 양곡수급계획과 국민식생활개선을 위한 기초자료를 얻기 위하여 1962년부터 농림부 양정국에서 표본으로 농가 426가구와 비농가 294가구 모두 720가구를 대상으로 조사했다. 표본은 1960년도의 <인구 및 주택총조사>결과에 의해서 설계된 확률표본추출법에 의한 것이다. 이러한 표본설계는 5년마다 대체되고 또 표본크기도 사정에 따라 변경되어 왔다. 이 조사의 표본은 위에서 본 <농가경제조사>표본과 같다. 한 표본에 의해서 <농가경제조사>, <양곡소비량조사>, <농업관측조사> 그리고 <영농실태조사>를 병행 이용하고 있다.

E) 농산물생산비조사: 이 조사는 농산물의 적정가격결정과 영농개선책 등의 농업정책수립을 위해서 농수산부에서 실시해 오고 있다. 이 조사를 위한 표본은 <농가경제조사>표본과 같다.

F) 작물통계조사: 이 조사는 작물별 파종면적과 생산량을 파악하여 합리적 농업정책수립에 이용하기 위해서이다. 정부수립 이전부터 작물통계는 행정보고통계로 작성해오다 1974년에 일부의 작물통계조사를 표본조사로 전환했고 1979년에는 전 작물에 대하여 표본조사로 대체하게 되었다.

#### IV 조사방법의 변화

이상에서 1960년 이전부터 실시한 조사통계 가운데 중요하다고 판단되는 8개만을 들고 개관해 보았다. 다음 표에서 보는 바와 같이 확률표본에 의해서 조사되는 시점 8개 가운데 1962년이

4개로서 50%에 해당되고 1959년과 1961년이 각각 1개씩이고 나머지 2개는 모두 1974년이다. 이와 같이 1960년의 초가 우리 나라 조사통계의 한 전환점인 듯하다. 한가지 주목되는 점은 주요통계의 조사가 처음 한국은행에서 시작한 점이다. 아래 표에서 보는바와 같이 <도시가계조사>, <광공업동태조사> 그리고 <농가경제조사>가 처음 한국은행에서 조사하다 현행조사기관에 이관시킨 점이다.

### 조 사 개 요

조 사 명	조사개시년	과거조사방법	확률표본조사전환시점
도시가계조사	1951(한은)	유의 표본	1959
경제활동인구조사	1957(통계국)	확률표본	1962
광공업동태조사	1954(한은)	유의표본	1962
농업기본통계조사	정부수립전	보고통계	1974
농가경제조사	1953(한은)	유의표본	1961
양곡소비량조사	? (농림부)	?	1962
농산물생산비조사	? (농림부)	?	1962
작물통계조사	1947(농림부)	?	1974(1979)

주: 조사개시년 다음 괄호 속은 조사개시 당시의 조사담당기관

현재 한국의 조사통계 117개 가운데 연도별 출발시점을 보면 1960년대와 1970년대가 32와 33으로 가장 많다. 이들 조사통계 가운데 전수조사는 33개이고 표본에 의해서 조사되는 통계가 84나 된다.

### 조 사 통 계 현 황

출발시점	1950년 또는 그 이전	1960대	1970년대	1980년대	1990년대	합계
전수조사	11	9	13	2	2	37
표본조사	13	21	25	23	10	92
계	24	30	38	25	12	129

주: 위자료는 1994년도 <한국통계조사현황>에 의해서 작성

### 3-2. 통계의 조정

위에서 본 것과 같이 1860년대 초에 통계조사의 방법이 개선된 것으로 생각되는 1960년대 중엽에 통계의 시계열이, 발표시점에 따라 같지 않은 것이 두 곳에서 발견된다. 즉 그 하나는 농작물 생산통계이며 다른 하나는 국민소득계정이다. 이 두 통계는 1965년 이전에 발표된 통계연보와 그 이후에 발표된 연보의 수치가 같지 않다. 이들 각각을 나누어 비교해 보기로 한다.

#### I 주요농산물생산통계

농산물은 여러 가지가 있으나 우선 연도별 미곡생산량을 1964년, 1965년 그리고 1966년의 한국은행 <경제통계연보>의 통계로 비교해 본다. 비교연도는 1960년에서 1965년까지로 한다.

자료별 양곡생산량 비교

연	도	단위	1960	1961	1962	1963	1964	1965
A	1964년통계연보(153면)	천석	15,949	18,903	15,938	19,207		
B	1965년통계연보	천M/T	2,296.7	2,722.6	2,295.1	2,765.8	2,921.9	
C	1966년통계연보(168면)	천M/T	3,034.1	3,476.3	3,041.2	3,755.7	3,954.5	3,501.1
D	1966년농림통계(146면)	천M/T		3,462.5	3,014.9	3,758.0	3,954.5	3,501.1
A/B			6.94	6.94	6.94	6.94		
A/C			5.26	5.44	5.29	5.11		
A/D				5.43	5.29	5.11		

위에서 1964년도의 <경제통계연보>에서의 미곡생산량은 단위가 용량단위인 석(石)인 대하여, 1965년도와 1966년도의 <경제통계연보>에서의 그것은 중량단위인 톤(metric ton)에 대하여, 또 1966년도의 <농림통계연보>에서의 단위도 톤에 대하여 표현했다. 톤당 용량(환산율)을 알기 위해서 양자의 비율을 계산해 본 결과 1964년도의 미곡생산량과 1965년도 연보에 표시된 미곡생산량 간의 비율 즉 톤당 석(A/B)은 1960년에서 1963년까지 8.94석으로 일정한데 대하여 1964년도 연보의 생산량과 1966년도 연보의 그것과의 비율 (A/C)은 1960년에서 1963년까지 일정하지 않다. 또 1965년도의 연보와 1966년도 연보에서 다 같이 톤(metric ton)단위임에도 불구하고 그 생산량도 같지 않다. 1966년도의 <농림통계연보>에서의 생산량과 비교해 본 결과 약간의 차이는 있어도 1966년도의 <경제통계연보>의 통계와 같다. 다같이 중량으로 표시되었지만 <경제통계연보>에는 정곡(精穀)을 영어로 cleaned rice로 표현한데 대하여 <농림통계연보>에는 정곡을 polished rice로 표현하고 있다. 이것만으로도 1965년 이전의 통계와 1966년 이후의 통계간에는 반드시 통계수자의 조정이 있었다. 즉 1964년과 그 이전의 <경제통계연보>에 수록된 미곡생산량통계(1960년, 1961년, 1962년, 그리고 1963년도의 미곡생산량)와 또 1966년이후의 <경제통계연보>에 수록된 미곡생산량 통계는 동일한데 대하여 1965년과 1966년도의 그것은 같지 않다. 이것은 1965년과 1966년 2차에 걸친 조정이라 할 수 있다.

## II 국민소득계정

위에서 본 농업생산량통계의 조정은 분명히 국민소득계정에 영향을 주었을 것으로 보고 1960년, 1961년 그리고 1962년도의 <산업별국민총생산액>을 1964년도, 1965년도 그리고 1966년도의 <경제통계연보> 자료별로 비교해 보기로 한다. 산업은 농업·농림업 및 제조업만을 보고 그 외 산업은 기타로 묶어서 보기로 한다. 우선 경상시장가격의 소득을 보기로 한다.

다음표에서 보는 바와 같이 <경제통계연보>의 연도에 따라 같은 항목의 추계치가 같지 않은 점으로 보아 1965년과 1966년 사이에 통계의 추계치에 대한 큰 조정이 있었음을 알 수 있다. 이러한 조정의 근거와 내용에 대하여 <조사월보>등을 보았으나 찾아볼 수 없었다. 1964년 이전과 1967년 이후의 <경제통계연보>에는 통계수치의 일관성을 유지하고 있다. 한 예로써 1960년의 농림수산업의 총생산액은 위 표에서 보는 바와 같이 709억 6천만원이다. 이 추계치가 1961년, 1962년, 1963년 그리고 1964년도의 <경제통계연보>에서의 값은 일치한다. 그러나 1964년과 1965년도 <경제통계연보>에서의 값은 709억 6천만원과 719억 2천만원으로 같지 않고 또 1965년과 1966년의 그것이 각각 719억 2천만원과 857억 2천만원으로 같지 않다. 그러나 1966년도와 1967년 그리고 1968년도의 <경제통계연보>에서의 이 들 값은 일치한다. 이 하나의 추계치외의 모든 추계치가 같은 양상을 보이고 있다. 즉, 1964년과 그 이전에 발행된 <경제통계연보>에서의 해당 추계치는 일관성을 유지하고 있고 또 1966년과 그 이후에 발행된 <경제통계연보>에 기재된 당해 추계치도 일관성을 유지하고 있다. 이것으로 보아 조정의 시점은 1965년과 1966년 2개년의 추계에서 조정되었음을 알 수 있다.

연도별 자료별 산업별 국민 총생산액 (단위:10억원, 구성비 단위: %)

산 업	1960			1961			1962		
	64연보	65연보	66연보	64연보	65연보	66연보	64연보	65연보	66연보
농림수업	70.96	71.92	85.72	91.50	95.93	114.28	93.95	100.88	121.01
(구성비)	33.7	32.9	35.2	37.9	36.9	39.0	33.4	33.4	35.7
광 업	3.49	4.94	5.25	4.94	5.19	5.51	6.29	6.61	6.99
(구성비)	1.6	2.2	2.2	2.0	2.0	1.9	2.2	2.2	2.1
제조업	26.71	27.81	32.98	30.07	30.49	38.86	36.46	39.87	49.68
(구성비)	12.7	12.7	13.6	12.5	11.7	13.2	13.0	13.2	14.7
기 타	100.55	114.11	119.20	114.9	128.31	134.70	144.78	155.01	160.92
(구성비)	48.0	52.1	49.0	47.4	49.4	45.9	51.4	51.3	47.5
국민총생산	210.71	218.78	243.14	241.41	259.92	393.35	281.48	302.37	338.60

주:표에서 연도 하단에 64연보, 65연보, 66연보에서 64, 65, 65는 앞 19라는 두자리를 생략한 것이고 연보는 한국은행조사부 발행의 <경제통계연보>를 생략한 것임.

### III 통계고문단의 활동

미국의 국제협력처(ICA)의 도움으로 한국통계제도와 통계조사활동의 개선을 위해서 1958년 4월에 미국의 통계고문단이 내한하여 5년간이나 체재하게 된 적이 있었다. 초대 단장은 Stuart A.Rice이고, 2대 단장은 Charles B.Lawrence, 3대 단장은 초대단장이던 S.A.Rice박사이고, 4대 단장은 Benjamine J.Tepping이었다.

이들 고문단의 한국정부에 대한 건의는 다음과 같았다.

1. 일반목적 통계의 집중
2. 중앙표본조사기관의 창설
3. 한국통계자료처리소의 창설
4. 농림부소관통계의 개선
5. 통계기관 사무기능의 한계
6. 통계조정과 관리의 권한
7. 국제조사위원회를 통계위원회 개편
8. 민간의 통계사업참여
9. 통계법의 제정
10. 정부통계요원의 훈련

이러한 권고 외에도 1960년 실시하게 될 인구·주택총조사계획의 검토와 개선에 대한 방안 제시 등으로 한국통계발전에 크게 기여하였다. 가장 주목되는 것은 통계법을 제정하고, 통계의 조정권한을 통계국에 부여하고, 또 농업통계개선에서는 작물통계를 표본조사로 전환하게 한 점을 들 수 있다. 자료처리시설의 강화를 강조했기 때문에 1960년의 인가와 광공업에 대한 두 총조사결과를 신속하게 발표할 수 있었다.

이러한 권고에 따라 농림부의 통계개선과 광공업총조사결과에 의해서 농산물통계의 조정과 국민소득통계의 조정이 이루어진 것으로 추측된다.

## 4. 맺 음

이상을 다음과 같이 요약 정리할 수 있겠다.

1. 통계기구는 해방이후 통계수요의 폭증에 의해서 통계를 생산·공급하는 정부의 통계기구가 다소의 곡절은 있었으나 계속 확대·발전하여 오늘에 이르고 있다. 다만 통계수요에 따라 자연 발생적으로 정부 각 기관에 통계부서가 생기게 되어 오늘의 통계제도는 분산형의 특징을 갖게 되었다.

2. 통계의 생산방법은 과거에 인구총조사(센서스)와 행정업무의 부산물인 행정통계, 보고통계 그리고 유의표본조사에 의한 통계이었으나 1960년 이후는 유의표본이 확률표본추출법으로 전환되고 행정통계와 보고통계의 많은 부분이 확률표본에 의한 표본조사로 전환되었다. 총조사도 10년간격에서 중간 년도인 매 5년에도 실시하여 표본조사를 뒷받침하는 시대적 특징을 갖게하였다.

그 외에도 70년대와 80년대에 노동과 환경통계의 수요증가에 따라 통계의 생산·공급이 계속 증가하고 있는 점이다.

3. 위에서 본 바와 같이 통계의 생산·공급을 담당하고 있는 통계기구의 발전적 변화와 조사방법의 발전이 바로 한국통계발전을 대변하는 것이라 할 수 있다. 이러한 통계기구와 조사방법의 발전을 유도하게 한 요소는 여러 가지가 있지만 이 가운데 중요 요소는 1958년 4월에 내한하여 5개년간 체류한 미국통계고문단의 활동과 1962년에 제정하여 공포한 통계법을 들 수 있겠다. 통계법의 주요골자는 통계 정보 제공을 요구받으면 통계 정보 제공을 법률로 의무화한 점, 정부의 각종통계의 조정과 기준을 법률로 규정한 점 그리고 일반통계와 지정통계에 의해서 통계의 정확성을 정부가 보장하고 있는 점이다.

4. 통계발전의 한 근거는 위에서 본 바와 같이 1965년도에 농산물생산통계가 용량단위(곡물의 경우 석)에서 중량단위(metric ton)로 전환하는 동시에 생산량의 조정이 1965년과 1966년 2차에 걸쳐 있었던 점, 또 국민소득계정도 1965년과 1966년 2차에 걸쳐 조정된 점을 들 수 있다. 또 곡생산량 통계의 조정은 아마도 1960년의 농업총조사에 근거한 표본조사, 1959년부터 실시한 평예(plot cutting)법의 확률표본조사 결과와 과거의 보고통계 그리고 가계조사에 의한 소비통계간의 불일치에 근거한 것으로서 정확한 통계로의 발전을 의미하는 것으로 평가된다. 또 국민소득계정도 생산통계의 조정이 파급된 것으로서 정확한 통계를 위한 조정으로 평가된다. 또 국부통계의 경우는 1969년 4월에 각기관이 막대한 예산과 인력을 들여 처음으로 조사된 자료가 공표되지도 못했다. 그것은 자료의 부정확성 때문이었다. 이러한 경험이 10년후인 1978년의 국부통계조사를 성공시켰다.

5. 앞으로 통계의 발전을 거듭하기 위해서는 정확한 통계의 생산·공급의 효율화와 일반의 통계에 대한 인식과 합리적 사고의 제고가 요청된다. 일반행정도 각각의 전문지식이 요구되지만 통계업무 또한 특별한 전문지식이 요구된다. 여러 제약이 있겠지만 통계담당기관은 필요한 전문 인력을 얼마나 확보하고 있는지 세심 되새겨보아야 할 것 같다.

#### 參 考 文 獻

- 
- 통계청, 「韓國統計發展史 I」, 1992  
—— 「韓國統計發展史 II」, 1992  
—— 「韓國統計調查現況」, 1994  
—— 「94 통계간행물 총람」, 1994  
—— 「통계목록」, 1994  
한국은행 조사부, 「경제통계연보」, 1962  
\_\_\_\_\_, \_\_\_\_\_, 1964  
\_\_\_\_\_, \_\_\_\_\_, 1965  
\_\_\_\_\_, \_\_\_\_\_, 1966  
\_\_\_\_\_, \_\_\_\_\_, 1967  
\_\_\_\_\_, \_\_\_\_\_, 1968

- 한국은행 조사2부, '한국은행의 통계업무', 「통계」, 2권1호(1976).PP.42-47
- 金昶池, '우리나라의 勞動統計 現況', 「통계」, 3권3호(1977).PP.63-75
- Hong Nai Park, 'A Study on Sampling Methods in Rice Yield Survey' 「統計學研究」, 4권1호(1975년 6월).PP.19-32
- Bureau of Statistics, EPB, *Major Sample Survey of Korea*, 1972
- Grazia Arangio-Ruiz, 'The New Organization of Italian Statistics', *Bulletin of The International Institute, Proceedings of the 48th Session*, book 2 (Cairo) 1991, 2/13. PP.1-24
- Bruce Petrie, 'Social Statistics—Some Canadian Developments'-----  
-----, 3/13, PP.1-8
- Guido N.Rey, 'Official Statistics in Italy',-----,  
*Proceedings of the 49th Session*, Book 3(Firenze) 1993, PP.273-290
- Oomens, C.A., 'Some Notes on the Development of Statistics in The Netherlands',  
-----,-----, Book 4(Firenze) 1993, PP.326-330

## 통계기구와 조사방법의 변천

토론1 : 성균관대학교 통계학과 교수 남궁 평

제 1회 통계인의 날을 맞이하여 “ 통계기구와 조사방법의 변천”에 대하여 발표하여 주셔서 감사합니다. 통계는 개인이나 단체의 입장에서 볼 때 그 활동의 지표로서의 역할이 있으며 국가의 입장에서는 정책수립과 실적평가의 지표로서의 역할이, 국제사회의 입장에서는 국가간의 신용의 지표로서의 중요한 역할이 있다.

과거 50년간 한국의 통계기구(제도)와 조사방법의 변천과정을 재조명하여 봄으로써 통계에 대한 새로운 인식제고로 인하여 통계의 발전을 기할 수 있다고 본다.

### 1. 통계기구현황에 대한 보완

한국의 통계활동은 극히 분산되어 있어 각급기관 및 단체는 자기가 필요로 하는 통계를 일차적으로 각 기관 책임하에 작성하고 있다. 이는 통계전문요원의 집중적 활용곤란으로 통계작성능력이 제한되고 있으며, 통계상호간의 통일성과 일관성을 유지하기 위한 통계의 정비가 요구되는 실정이다. 통계기구의 역사에서는 통계업무주관기관을 중심으로 상세히 논술했으나 내용을 첨가하면 다음과 같다.

<표 1> 통계기구및 작성현황 (단위 : 기관,종)

	작성 기관수	계	승 인 통 계				
			종 류 별		작 성 방 법 별		
			지정 통계	일반 통계	조사 통계	보고 통계	가공 통계
계	96	374	37	337	176	169	29
정 부 기 관	40	266	31	235	102	141	23
- 중앙행정기관	25	216	31	185	82	126	8
- 지방자치기관	15	50	-	50	20	15	15
지 정 기 관	56	108	6	102	74	28	6
- 금융기관	8	31	2	29	20	8	3
- 정부투자기관	9	25	-	25	11	14	-
- 정부출연기관	11	7	2	5	7	-	-
- 공공법인	14	23	-	23	16	5	2
- 각종단체	14	22	2	20	20	1	1

< 자료 : 통계청, 통계목록 1995. 1 >

<표 1>에서 보듯이 통계작성 기관수는 96개기관이며 374종의 통계를 생산공급하고 있다. 정부 기관은 중앙행정기관과 지방자치단체를 포함하여 40개 기관에서 266종의 통계를 작성하고 있으며, 지정기관은 금융기관, 정부투자기관, 정부출연기관, 공공법인, 그리고 각종단체 등 56개기관에서 108종의 통계를 작성하고 있다.

이를 1984년의 통계작성현황과 비교해 보면 다음과 같다.

통계작성기관수는 79개 기관에서 96개 기관으로 17개 기관이 증가하였으며, 통계작성수는 337종에서 374종으로 9% 증가하였다. 또한 지정통계는 34종에서 37종으로 3종이 늘어났으며 일반통계는 303종에서 337종으로 9%증가하였다.

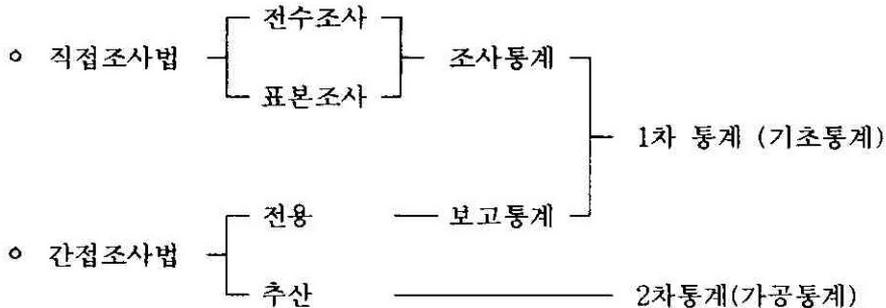
통계청에서 발행한 통계목록(1995. 1. )에 의하면 농림수산부에서 작성하는 통계는 27종으로써 지정통계 9종과 일반통계 18종을 생산공급하고 있다. 이는 한국통계발전사 II. pp.237-243과 상이하다.

또한 노동부의 경우는 지정통계 3종, 일반통계 13종을 생산하고 있는데 지정통계는 사업체노동 실태조사, 매월노동통계조사, 임금구조기본통계조사이다.

## 2. 조사방법

조사방법의 선택기준으로는 조사목적, 예산, 요구정도 그리고 조사항목 등을 고려하여 선택해야 한다. 조사방법은 자료의 수집방법에 의한 분류와 조사기법에 의한 분류 등으로 유형을 살펴볼 수 있다.

### 1) 자료수집방법에 의한 분류



### 2) 조사기법에 의한 분류



현재 실시하고 있는 조사방법을 통계작성기관별로 통계조사현황을 살펴보면 <표 2>와 같다. 총 374종의 통계 중 176종의 조사통계가 공급되고 있는 데 전수조사가 55종, 표본조사가 121종이다. 예를 들면, 현재 실시하고 있는 가계조사의 경우에는 표본조사와 유치조사를 병행하고 있으며 인구주택총조사는 전수조사와 표본조사 면접조사를 실시하고 있다. 또한 의식조사의 경우에는 표본조사와 면접조사, 그리고 신속한 결과를 요구하는 조사에서는 표본조사와 전화조사를 실시하고 있다.

<표 2> 조사통계현황(1995. 1월 현재)

	전수조사	표본조사	보 고	가 공	계
중 앙 기 관	26	56	126	8	216
지 방 자 치 단 체	15	5	15	15	50
금 융 기 관	-	20	8	3	31
정 부 투 자 기 관	4	7	14	-	25
정 부 출 연 기 관	-	7	-	-	7
공 공 법 인 체	2	14	5	2	23
각 종 단 체	8	12	1	1	22
계	55	121	169	29	374

참고로 현행 통계청에서 실시하고 있는 전수조사 중 대표적인 인구주택총조사에 대한 조사방법의 변천과정을 <표 3>에서 살펴보고, <표 4>에서는 표본조사현황을 제시하였다.

<표 3> 인구주택총조사의 변천과정(한국통계발전사II, p 87-88)

조사명칭	실시년도	감독 주관 기관	조사 기준시점	조사 방법	비 고
간이 국세조사	1944	조선총독부	5. 1	현 주 개 념	
제1회 총인구조사	1949	공 보 처 통 계 국	5. 1	현주개념으로 자계식	정부수립후 첫 인구센서스
제1회 간이 총인구조사	1955	내 무 부 통 계 국	9. 1	상 동	
인구.주택 국세조사	1960	내 무 부 통 계 국	12. 1	상주개념으로 면접조사식	주택에 관한 조사 실시(최초) 경제활동 및 출산력 사항 : 표본집계
인구센서스	1966	경제기획원 조사통계국	10. 1	상 동	10% 표본조사 병행 (경제활동 및 출산력)
총인구 및 주택조사	1970	"	"	"	10% 표본조사 병행 (경제활동, 출산력, 인구가동 및 일부주택에 대한 사항)
"	1975	"	"	"	5% 표본조사 병행 (경제활동, 출산력, 인구가동 및 일부주택에 대한 사항)
인구 및 주택센서스	1980	"	11. 1	"	5% 표본조사 병행 (경제활동, 출산력, 인구가동)
인구 및 주택센서스	1985	"	11. 1	"	모든 항목 전수조사
인구주택 총 조사	1990	"	11. 1	"	10% 표본조사 병행 (경제활동, 출산력, 인구가동)

<표 4> 통계청의 표본조사현황 (통계연수원, 중급통계과정 I, p 245)

조사명	표본추출방법	표본개편주기
<b>&lt;가 구 부 문 &gt;</b>		
1. 경제활동인구조사	층 화 계 통 추 출	5 년
2. 인구동태표본조사	"	5 년
3. 도시가계조사	"	5 년
4. 고용구조조사	층 화 추 출	3 년
5. 사회통계조사	층 화 계 통 추 출	5 년
6. 국부통계조사	층 화 추 출	10 년
7. 전국소비실태조사	층 화 추 출	5 년
<b>&lt;사 업 체 부 문 &gt;</b>		
1. 광공업 동태조사	절사 및 층화 계통 추출	5 년
2. 도소매업동태조사	"	5 년
3. 도소매업통계조사	"	5 년
4. 운수업 통계조사	층 화 추 출	1 년
5. 건설업 통계조사	층 화 추 출	1 년

### 3. 제 안

- 1) 노동통계조사의 경우 수준조사와 구조조사로 구분할 수 있는데 이 두가지 조사를 동일한 표본으로 조사하는 다목적 통계조사를 할 수 있도록 표본설계를 개선함으로써 조사경비의 절약과 효과적인 표본관리를 할 수 있다.
- 2) 지자체 실시로 인하여 지역별 경제사회개발계획의 수립을 위하여 지역별 통계자료의 수요가 급증하게 되므로 표본설계를 개편하여 추정값의 신뢰성을 제고해야 한다.
- 3) 통계전문 인력의 부족문제를 해소하기 위해 통계학을 전공한 인력을 보완하고 표본이론을 실제조사에 효율적으로 적용할 수 있는 방안과 조사기법의 개발을 위해 산학협동체제를 실시하여 정도 높은 통계를 생산해야 한다.
- 4) 무응답을 감소시킬 수 있는 다양한 조사방법을 개발해야 한다. 미국, 캐나다 경우 센서스 후 결과에 대한 사후분석을 실시하고 있는 바 우리나라의 경우에는 내부적으로 실시하고 있는지는 모르지만 일반인에는 미공개 상태이다.
- 5) 각종 통계조사의 종합적인 조정과 통계체계의 정비가 필요하다. 이는 각종 통계자료를 이용하는 데 있어서 통계상호간의 통일성및 일관성의 부족으로 인하여 통계의 정확성및 신뢰성이 낮아지기 때문이다.
- 6) 통계의 가공, 분석기능을 확대하여 통계의 활용을 극대화하기 위한 방안이 개발되어야 한다.

## 통계기구와 조사방법의 변천

토론2 : 고려대학교 통계학과 교수 이재창

1. 국가통계가 발전한 역사적 배경을 설명하고 우리나라의 통계생산기구의 변천과 공식통계의 발전 및 변모과정을 소상하게 정리하고 있다. 미 군정시대, 공보처 통계국, 내무부 통계국, 경제기획원 조사통계국을 거쳐 오늘의 통계청에 이르는 변천과 또한 중요 통계생산기구로 농림수산부, 노동부 및 한국은행의 통계생산활동을 시대적으로 고찰함으로써 우리나라의 통계발전상을 생산기관과 조사방법 그리고 이의 중요한 변화시점을 매우 잘 정리해 놓고 있다.

2. 한 나라의 통계수요는 그 나라의 경제발전단계에 따라 크게 다른 양상으로 나타나고 있으며 우리의 통계도 예외는 아닐 것이다. 일제치하에서의 통계의 불모로부터 국가경제발전의 초기단계에는 경제 개발계획에 필요한 통계생산에 혼신을 다해서 시작한 것이 역역히 엿보인다. 뿐만 아니라 권력구조와 정치적 갈등현상에 따라 통계기구는 그의 소속과 편제가 좌우되어 온 것도 알 수 있다. 다만 초창기의 통계는 통계전문가의 부족, 전문지식의 결여나 정치인의 통계에 대한 인식부족을 극복하여야 했고 이런 와중에서도 필요한 통계의 공급을 시도하였으므로 통계의 질적 향상을 효과적으로 이룩하기에는 역부족이었던 것 같다. 그러나 이렇게 열악한 여건에서 우리들의 선배 통계인들이 그늘에서 구축해온 통계는 크게 평가되어야 할 줄 믿으며 오늘의 통계청시대를 맞이하는 원동력이 되었음을 우리모두가 새롭게 인식해야겠다.

3. 과거의 역사로 현재를 파악하고 또한 미래에 대한 거울로 삼아야 한다는 교훈을 되새겨보면 현재의 미비한 점들은 있는 그대로 지적하고 겸허히 받아들여야 하며, 우리나라 통계의 미래를 준비하는데 길잡이로 삼아야 할 것이다.

지난 시대에는 국가주도의 계획경제체제로 통계가 거시적 계획에 기여해온 반면에, 현재와 미래는 국경없는 무한경쟁을 통해 민간주도적 시장경제체제로 가고있어 미시적 통계정보의 정확하고도 시의 적절한 공급이 민족번영에 일익을 하게 될것이다. 정부기관 뿐만 아니라 기업, 연구기관, 학계를 포함한 모든 국민에게 통계정보제공을 통해 경쟁에서 상대우위를 유지할 수 있도록 하는 사회간접자본의 역할을 수행해야할 변화시점에 와 있다고 생각되며 이러한 시대적 요구에 부응하는 체제로 탈바꿈하는 용기도 필요하겠다.

4. 윤교수의 발표내용에는 더 첨가할 것이 없을 정도로 잘 정리가 되어 있으므로 본 토론자는 과거를 토대로 앞을 대비하는 몇가지 제언을 통해 토론을 대신하고자 한다.

흔히들 정보화시대를 특징짓는 표현으로 정보의 「민주화」(Democratization of Information)라는 말을 하고 있다. 이는 초고속정보통신망을 통해 접근 가능한 통계데이터는 세계 어느곳에서나 누구든지 얻을 수 있다는 이야기이다. 따라서 이제는 누가 얼마나 빨리 부가가치를 더 많이 창출하는가가 관건이다. 즉 데이터로 부터 정보를 얻어내는 통계응용력의 경쟁시대가 된다.

통계의 생산, 가공 및 보급의 효율적 연계는 이러한 경쟁에 대비하는 것으로 통계활용의 경쟁력은 모든 경쟁에 우선하다고 생각한다. 이를 위해서는 통계기구뿐만 아니라 우수인력의 보급과 새로운 방법의 개발들이 뒤따라 주어야 할 것이며 통계교육은 현실적 적응력을 갖춘 실용성에 바탕을 두어야 할 것이다.

뿐만 아니라 정부의 통계에 대한 인식 역시 통계인의 노력여하에 있다고 보며 모든 통계인은 통계이용자에게 통계활용이 얼마나 유용한 것인지를 실제로 입증함으로써 우리의 위상을 높일 수가 있다.

정보화시대에는 경쟁의 압력으로 인해 빠른 발걸음없이 남에게 뒤떨어질수 밖에 없어, 정부나 공공 통계기구들도 선의의 경쟁을 피하기는 어렵게 된다. 이러한 경쟁은 이미 선진국에서 그 징후를 보이기 시작했으며 정부도 이윤추구의 목적으로 나타나는 사기업과의 통계생산·공급의 경쟁에서 예외는 될수가 없을 것이다. 따라서 통계기구도 역할과 위상에서 변화에 적응할 수 있는 유연성과 적응력을 배양해야 할 것 같다.

통계도 국제화시대를 맞아 국제적인 표준화에 대비해야 되겠으며, 외국의 통계를 우리의 정의와 분류로 바꾸어 사용자가 쉽게 이해할 수 있도록 공급해 주어야 할 것이다.

통계조사방법은 그동안 엄청난 발전을 이룩해 왔으며 우리통계도 이런 면에서는 상당한 수준에 도달 했다고 보여진다. 그러나 방법의 향상만으로는 해결 될수 없는 문제도 많이 있다. 「무응답」은 세계적인 추세가 되어 모든 나라가 걱정하는 문제가 되고 있다. 자기의 정보는 남에게 주기 싫지만 남의 정보는 알고 싶어하는 속성이 두드러 진다. 통계정보의 본원인 국민각자가 통계조사에 능동적으로 참여하게 하는 연구와 노력이 필요하다. 통계조사의 기피현상은 응답자의 인식전환이 선행되어야 하는데, 이를 위해서는 자기의 응답이 정확한 통계를 만들어주고, 이를 다시 자기에게 유형, 무형, 그리고 직접, 간접으로 이익이 되어 되돌아 온다는 확신에서만 가능하다. 통계생산기구는 이러한 면을 포함한 대민홍보와 신뢰구축에 힘써야 할 것으로, 이는 통계이용자확대라는 기반 구축이 우선해야 할 것이다. 이를 위한 정책과 서비스의 구상도 이제는 중요한 부문으로 생각된다.

## 國家統計組織의 發展方案

發表: 統計廳 統計企劃局長 金 日 炫

### I. 國家統計制度의 類型

國家統計制度의 類型은 일반적으로 分散型 統計制度和 集中型 統計制度 두가지로 구별되어 있다. 그러면 다음에서 이 두가지 統計制度의 長·短點과 國際的인 潮流에 대해서 살펴 보고자 한다.

#### 1. 分散型 統計制度

分散型 統計制度는 기본적으로 國家의 여러 기관들이 自己固有業務와 관련하여 별도의 統計組織을 가지고 필요한 통계를 自體生産하는 制度로 볼 수 있다. 이와 관련 各機關의 統計活動을 調整하는 責任과 權限은 별도로 독립기관이나 위원회가 設置되어 이 機關에서 附與 받게 된다. 分散形態를 취하고 있는 國家는 美國, 日本, 英國 등을 例示할 수 있다.

이 分散型 統計制度의 長·短點을 요약하면 다음과 같다.

##### 가. 分散型的 長點

- 自己部處가 수행하는 業務分野와 關聯, 필요한 자료의 需要把握이 容易하고, 그 분야에 대한 專門知識을 統計에 직접 반영할 수 있음
- 또한 각 部處가 필요한 時期에 신속히 統計를 작성하여 活用할 수 있는 長點도 있음.

##### 나. 分散型的 短點

- 國家統計의 均衡發展이 곤란하여 統計의 重複과 不一致의 위험이 있을 수 있음.
- 裝備 및 人力의 利用面에서도 중복과 낭비요인이 상존함
- 각 部處가 모두 自己部處의 목적에만 충실하다 보면 통계자료를 他機關에서 이용함에 다소 어려움이 있을 수 있음

## 2. 集中型 統計制度

集中型 統計制度는 기본적으로 한 機關이 모든 經濟社會分野의 統計情報를 作成하여 각 이용기관에 제공하는 제도로 볼 수 있다. 그리고 이 制度에서는 이 제도의 원활한 운영을 위해 각 部處의 代表者로 구성되는 統計連絡機構가 설치되어 각 부처에서 필요한 통계가 反映될 수 있도록 하고 있다.

集中形態를 취하고 있는 國家는 대표적으로 캐나다, 네델란드, 독일 등을 예시할 수 있다. 특히 캐나다의 경우는 모든 분야의 統計業務가 캐나다 통계청에 집중되어 있는, 완벽한 集中型 모델로 인식되고 있다.

이 集中型 統計制度의 長·短點은 分散型 統計制度의 長·短點과는 반대로서, 요약하면 다음과 같다.

### 가. 集中型의 長點

- 한 機關에서 모든 統計를 作成하므로 여러 統計分野에 우선순위를 부여하여 統計의 均衡的 發展과 體系化가 可能
- 熟練된 人的資源과 裝備가 부족한 국가에서는 集中에 따른 規模의 經濟가 있음
- 여러 分野의 統計資料를 單一機關에서 生産함으로서 資料의 利用이 便利하고 效率的임.
- 中央統計機構가 行政的으로 獨立되고 政治的으로 中立되어 統計의 客觀性 維持가 容易함.

### 나. 集中型의 短點

- 統計가 單一機關에서 作成됨으로서 자칫 統計作成機關과 統計利用機關이 遊離될 危險이 常存함
- 中央統計機關이 巨大해지면 利用機關의 變化要求에 充分히 副應하지 못하는 危險이 있음

### 3. 統計制度에 대한 國際的인 潮流

一般的으로 各 政府部處가 政策과 執行業務를 遂行·分析하기 위해서는 자기분야의 통계만이 아닌 여러 분야의 통계를 필요로 하기 때문에 分散型制度를 採擇하는 國家에서도 集中의 傾向을 나타내고 있다.

分散型 制度의 代表적인 國家인 美國의 경우도 聯邦統計活動이 4개 기관으로 集中되고 있다.

- 商務部 센서스局 (Bureau of the Census)
- 勞動部 勞動統計局 (BLS)
- 農務部 農業統計室 (SRS)
- 保健福祉部 國立保健統計센터 (NCHS)

그리고 점차 經濟·社會가 복잡해지면서 巨視經濟分析, 國民計定, 經濟社會指標의 出現으로 各 個別分野別로 統計를 作成하는 分散型 制度로는 대응이 쉽지 않음이 지적되고 있다. UN에서도 統計의 專門人力과 장비가 부족한 개발도상국이나 新生 國家에서는 集中型 制度를 採擇하도록 권장하고 있다.

그러나 다른 한편으로는 統計調查環境의 악화와 各種 尖端情報器機의 발달로 調查統計의 비중이 점차 감소되고 行政統計의 역할이 커짐에 따라, 先進國에서는 統計調整權限을 中央 統計廳에 附與 및 強化하면서 分散型制度를 취하는 형태가 증대될 展望이다.

즉, 統計制度의 國際的인 조류를 요약하면 集中型 制度를 採擇하고 있는 國家에서는 統計 作成機關과 統計利用機關間의 連繫 및 相互調和를 위해 常設的인 통계위원회를 두어 統計 政策을 조정하고 상호 협조를 도모하는 방향으로 進展되고 있다고 볼 수 있다.

그리고 分散型制度를 채택하고 있는 國家에서는 中央統計機關의 統計調整機能을 강화하여 類似統計의 作成 重複防止 및 質的改善을 도모하는 방향으로 進展되고 있다.

## II. 우리나라의 統計作成體系

<표1>에서 보듯이 우리나라의 統計制度는 外形적으로는 統計廳을 비롯한 92개 기관에서 353종의 통계를 작성하는 分散型 統計制度로 볼 수 있다. 그러나 실질적으로는 농림수산부의 統計組織을 除外하고 各 部處의 統計組織이 脆弱하여 대규모통계조사와 일반목적 통계 작성업무 및 統計綜合調整機能이 統計廳에 집중되어 있어 集中型 要素가 매우 강한 형태로 볼 수 있다.

<표 1> 우리나라의 統計作成機關 및 人力現況

( '94. 2. 1 현재 )

	작성기관수	인력 (명)	승인 통계
계	92	5,638	353
○ 정부 기관	41	4,697	248
- 중앙 행정기관	26	3,977	210
- 지방 자치단체	15	720'	38
○ 민간 지정 기관	51	941	105

주 : 1) 시·군 통계인력 포함.

統計專擔組織現況을 보면, 中央行政機關에서 통계청의 統計專擔組織을 보유하고 있는 機關이 농림수산부, 통상산업부, 환경부, 보건복지부, 노동부, 건설교통부 등 6개 부처로 增加現象을 나타내고 있다(표2 참조).

그러나 이 6개 기관중 自體統計活動을 수행하는 부처는 農林水産部 및 勞動부에 불과하며, 新設된 3개 부처(通商産業部, 環境部, 建設交通部)의 統計組織에서는 오히려 行政電算業務에 치중되어 있는 실정이다.

從事人力도 환경부의 경우 統計係에 4명, 건설교통부의 경우 統計調査係에 3명, 通商産業部の 경우는 全無한 실정으로 人力配置 側面에서도 매우 취약함을 보이고 있다. 이와 같이 中央行政機關의 경우 統計廳과 農林水産部를 제외하면 자체적으로 統計를 企劃하고 조사하는 기능을 遂行하기가 원천적으로 불가능한 것임을 알 수 있다.

광역 지방자치단체의 경우도 地方化時代에 對備하여 統計組織과 人力이 增大되어야 함에도 불구하고 오히려 機構가 縮小되는 비운을 맞기도 하였다. 1992년에는 서울시 統計擔當官室이 統計電算擔當官室 統計係로 축소되었고, 이어서 1994년에는 나머지 14개 市·道에서도 조직이 축소되었다.

마찬가지로 基礎自治團體인 市·郡에서도 68개 市 및 105개 郡에서 28.3%인 49개 市·郡의 統計係가 法務統計係 또는 庶務統計係 등으로 명칭이 바뀌어 지면서 축소되었다.

또한 같은 표2에서 統計를 專擔할 수 있는 組織인 통계청, 농림수산부 및 노동부의 경우도 人力의 상당수가 統計를 企劃하고 分析하는 분야가 아닌 現場業務를 담당하는 지방사무소에 근무하고 있음을 볼 수 있다.

물론 정확한 통계를 만들어 남에 있어 資料蒐集業務가 중요하지만, 이를 逆說的으로 表現하면 統計를 專擔하는 組織에서도 分析人力이 脆弱하다는 점을 지적할 수 있는 것이다.

<표 2> 7개 중앙행정기관 및 지방행정기관의 통계인력현황

( '95. 7. 1 현재)

	조직이름	인원	특징
7개 중앙 행정 기관		3,578	
· 통계청		1,334	· 지방사무소에 856명
· 농림수산부	통계정보관	2,080	· 지방사무소에 2,017명
· 통상산업부	전산통계담당관	9	· 전산업무만 전담
· 환경부	"	16	· 통계계에 4명
· 보건복지부	"	24	
· 노동부	통계담당관	89	· 지방사무소에 71명
· 건설교통부	전산통계담당관	26	· 통계조사계에 3명
지방 행정 기관		104	
· 서울시	전산통계담당관	7	
· 경기도	통계전산담당관	7	
· 경상북도	"	13	· 통계 1계·2계가 있음'
· 제주도	"	3	

주 : 1) 경북 및 경남의 경우만 2개계가 있음

### Ⅲ. 國家統計制度的 問題點

우리나라 統計作成制度에서의 問題點은 統計組織側面, 統計內容側面 그리고 統計環境側面 등 세가지로 구분하여 살펴 볼 수 있다. 이를 다음에서 보다 이해하기 쉽도록 個條式으로 살펴보기로 한다.

#### 1. 統計組織

##### 가. 中央政府

- 分散型 制度 採擇國家로서 中央政府의 統計組織이 크게 脆弱
  - 35개 中央 院·部·處·廳中 統計專擔組織을 가진 部處는 겨우 7個機關 (統計廳, 農林水産部, 勞動部, 保健福祉部, 通商産業部, 建設交通部, 環境部)
  - 自體統計作成活動을 遂行하는 部處는 3개기관에 不過 (統計廳, 農林水産部, 勞動部)
  - 新設된 3개 部處의 統計組織은 오히려 行政電算業務에 置中 (通商産業部, 建設交通部, 環境部)
- 中央政府의 統計組織間에 有機的인 連繫活動이 미흡하고, 統計開發 또는 改善에 消極的
  - 統計作成에 있어서 相互有機的인 協力體制가 未洽하여 부족한 人力으로 統計의 極大化 機會를 상실
  - 劃一的인 政府組織 運營으로 증가되는 統計需要에 統計組織의 신속적 대응이 어려워 適期에 새로운 統計開發이 困難
- 統計廳의 廳昇格 (90. 12. 27) 으로 國家統計의 綜合調整機能을 갖는 中央統計機關으로서 獨立性은 確保 되었으나 最小限의 機構改編으로 본래의 機能遂行에 蹉跌

- 1級 機關長의 位置로서 政府 各 部處 長次官 및 廳長을 상대로 統計作成 및 中止, 統計改善, 統計開發要請 등 統計調整活動을 遂行하는데는 限界
- 統計廳에서 실시하는 大規模 全國調查(人口住宅總調查, 産業總調查 등)는 市·道와 같은 地自體를 통하여 실시하여야 하나 市·道를 指揮하는데 困難
- 中央行政機關으로서 一般廳이 갖추어야 할 基本組織이 缺如

一 般 廳	統 計 廳
<ul style="list-style-type: none"> <li>· 企劃管理官               <ul style="list-style-type: none"> <li>— 企劃豫算擔當</li> <li>— 行政管理擔當</li> <li>— 法務擔當</li> </ul> </li> <li>· 公報官</li> <li>· 監查官</li> <li>· 非常計劃擔當官</li> </ul>	<ul style="list-style-type: none"> <li>· 企劃課長</li> </ul>

- 統計需要의 增加에 비해 統計關聯專門要員의 確保가 곤란하며, 統計人力이 오히려 減少 現象(1992년 1,388명 → 1995년 1,334명)

#### 나. 地方統計組織

- 地自制의 本格 實施에 따른 地域統計需要의 增加에 逆行하여 오히려 地方政府 統計組織을 縮小
  - 市·道 統計擔當官室을 電算擔當官室과 縮小 統合하여 統計係로서 命脈 維持
  - 一部 市郡 統計係를 他業務部署와 統合
  - 빈번한 人事移動으로 職員의 統計專門性和 經驗蓄積이 困難

○ 中央政府의 地域統計組織이 業務의 特殊性을 考慮하더라도 너무 獨立的이고 非效率的으로 運營

- 統計廳의 經常統計 資料蒐集機能을 수행하기 위한 獨立 地域統計組織 運營 (12個 事務所 856名)

- 産業部門에서 農業의 比重이 계속 줄어들고 있으나 農林水産部는 統計廳보다 더 큰 龐大한 地域統計組織을 繼續 保有 (9個 事務所 142개 出張所 2,017名)

- 統計廳과 同一한 事業體를 대상으로 統計資料蒐集을 擔當하기 위하여 勞動部도 별도의 資料蒐集人力을 확보 (地方勞動事務所 71名 配置)

## 2. 作成統計

### 가. 中央部處

○ 産業構造의 變化에 따라 比重이 계속 높아지고 있는 서비스業 部門의 통계가 매우 脆弱

- 金融, 保險, 證券 등 行政統計를 제외하고 情報産業, 不動産業, 社會서비스業 部門의 통계가 크게 不足

- 國際間에 서비스 交易의 증가로 人力, 技術, 知的財産權에 관한 交易情報가 主要 關心事로 부각

○ OECD 가입을 앞두고 제출하여야 할 약 80여종의 統計의 改善 또는 開發이 時急

- 既存에 生産되고 있는 많은 統計도 先進國 水準으로 基準과 質을 提高시키는 것이 필요

- 많은 統計, 特히 社會部門, 環境部門의 統計가 새로 開發이 要請

- 이제까지 經濟開發中心의 政策으로 經濟統計에 치우쳐 社會部門統計의 不足 등 國家 基本統計의 不均衡 發展
  - 우리나라의 全體承認統計中 經濟·産業部門의 統計가 75.1% 차지
  - 保健, 社會, 福祉部門統計는 주로 行政報告統計에 依存  
(전체 50種中 報告統計가 32種으로 65%)
- 政府統計機關에서 작성하여야 할 國家의 主要基本統計가 아직도 韓國銀行 등 非政府 統計組織에서 작성됨으로서 統計의 均衡的 發展이 未洽
  - 國民計定統計는 여러部門의 基礎統計를 가지고 推計하는 통계로서 基礎統計의 改善 發展과 直結
  - 國民計定統計를 擔當하는 韓國銀行이 中央統計機關에게 基礎統計改善을 要求하는 데는 제약

#### 나. 地方自治團體

- 이제까지 中央政府 依存의 行政慣行으로 地自體가 스스로 作成하는 統計는 極小
  - 住民登錄人口統計(年間), 事業體基礎統計(年間), 其他 各種 市民意識調查, 交通量 調查 등 小數의 年間調查 遂行
  - 月刊 經常調查는 1件도 수행치 못하며 주로 行政資料를 蒐集하여 統計年鑑 發刊에 置重
- 統計廳을 위시한 中央部處에서 地域統計를 作成하여 提供하는데는 限界
  - 全國規模의 센서스 등에서 末端行政單位의 統計生産은 可能하나 月別 經常標本調査로는 市·道 單位以下の 地域統計를 生産해서 提供하는 것은 無理
  - 地自體가 필요로 하는 經常統計는 地自體가 스스로 作成하는 것이 先進國의 慣行

### 3. 統計作成環境

- 中央·地方政府를 不問하고 公務員의 統計에 대한 認識이 歪曲되어 있고 豫算的 配慮가 매우 脆弱
  - 統計의 重要性을 論하면서도 事業의 우선순위 決定에서 統計는 恒常 뒷전
  - 統計의 뒷받침 없이 企劃 및 行政遂行의 經驗으로 統計의 필요성 또는 중요성을 看過
  - 統計部署에 良質의 人力配置 外面
  - 統計業務의 輕視로 大部分의 公務員이 統計部署에 配置되는 것을 忌避하고 配置되는 瞬間부터 업무보다도 他部署 移轉에 보다 神經
  
- 家口 또는 事業體를 대상으로 直接 訪問하여 작성하는 統計의 調查環境이 날로 惡化
  - 豫算制約으로 大規模 調查時 良質의 臨時調查要員 確保가 어렵고 (낮은 手當) 이들에 대한 充分한 教育訓練의 缺如로 蒐集된 統計의 質 低下
  - 正確한 統計를 작성하기 위해서는 個人 또는 事業體에서 正確한 情報 提供을 前提로 하나 私生活 侵害, 事業體 機密保護를 이유로 應答忌避 또는 拒否現象이 날로 增加
  - 맞벌이 부부의 增加, 核家族化 現象으로 應答者를 適期에 面接하기가 困難하여 再訪問으로 인한 調查業務量이 增加

## IV. 國家統計制度的 改善課題

우리나라 統計制度的 改善課題로서는 中央政府統計組織 및 機能의 再調整, 地方統計組織의 再編成, 作成統計의 內實化 등 크게 세가지로 요약할 수 있다. 이를 다음에서 개조식으로 살펴보기로 한다.

### 1. 中央政府統計組織 및 機能의 再調整

#### 가. 統計廳의 役割 및 機能 強化

○ 國家統計 事務管掌機關으로서 役割 遂行이 可能하도록 地位 格上

- 各 政府 行政機關을 상대로 統計調整活動을 圓滑히 수행할 수 있고 統計作成 協力を 求할 수 있도록 統計廳을 一般廳과 같은 次官級 廳으로 格上
- 中央行政機關으로써 갖추어야 할 基本組織 補強(企劃豫算, 行政管理, 公報, 法務, 監査, 非常企劃)

○ 統計調査의 專門性을 必要로 하는 調査統計를 統計廳으로 集中

- 家口나 事業體를 對象으로 面接調査하여 作成하는 一般 目的 調査統計의 集中 (農業總調査, 農村家計調査, 事業體賃金 및 雇傭實態調査등)
- 統計作成 能力이 없는 他部處가 必要로하는 調査統計의 開發 및 作成 提供 (海外投資業體, 外國人投資業體 産業動向調査 등)

○ 綜合統計의 性格을 갖는 統計의 作成體系 確立

- 巨視經濟分析, 各種 經濟·社會指標, 環境指標 등 綜合統計를 統計廳이 主管하여 基礎統計의 均衡的 發展 圖謀(國民所得統計, 地域所得統計, 景氣指標, 社會指標, 環境指標 등)
- 統計研究所를 設立하여 景氣豫測統計의 開發, 各種 統計結果의 分析 및 解釋, 統計 實務와 理論의 接木活動, 統計專門家養成 등을 專擔케 하여 次元 높은 統計 生産

○ 우수한 統計人力 確保를 위한 制度 및 裝置 마련

- 統計部署 勤務者의 人事 優待制度 導入(가점제, 수당지급 등)
- 統計廳의 研修院을 통하여 統計從事者에 대한 統計專門教育訓練을 持續적으로 實施

○ 世界化·國際化 추진에 積極적으로 부응할 수 있도록 國內外 統計情報의 綜合管理 및 對國民서비스 強化

- 國內 각 기관에서 생산되는 각종 통계정보를 蓄積 提供하는 통계 데이터베이스를 利用機關의 多様な 목적에 맞도록 개편하고 擴充
- 世界化를 추진하고 있는 政府 및 民間機關을 지원하기 위하여 國際機構(OECD, IMF, UNSTAT, ECE, ESCAP 등)와 주요 선진국으로부터 각종 統計情報를 신속히 입수하여 각종 媒體를 통하여 제공

나. 其他 中央部處 統計機關의 育成

○ 統計組織이 없거나 脆弱한 中央部處에 統計組織을 新設 또는 補強하여 固有業務와 관련된 統計作成 能力 強化

- 勞動部, 保健福祉部, 通商産業部, 建設交通部 등 각 부처는 業務量에 상응한 수준의 통계조직을 확보하고,
- 統計作成의 專門性에 비해 固有業務의 專門性이 보다 要請되는 觀測統計, 測定統計 등은 해당 部處에서 作成
- 해당 部處의 業務遂行에 부수되어 生産되는 行政統計 또는 報告統計를 積極 開發 하여 활용

#### 다. 中央政府 統計機關間 連繫性 強化

- 同一 對象處를 대상으로 하는 統計調査는 共同 또는 役割을 分擔하여 遂行하는 風土 造成

##### 《先進國의 例》

- 美 國 : 상무부 센서스국이 농업센서스를 실시하여 농무부 농업통계실에 주며, 매월 노동력조사(CPS)를 실시하고 그 결과를 노동부 노동통계국에서 분석 발표

- 各 部處間에 統計作成 協力 강화

- 他 利用機關이 이용할 수 있는 多目的 行政統計의 資料處理를 統計廳이 代행
- 統計廳에서 標本抽出을 위한 母集團 管理 및 資料提供의 圓滑化

## 2. 地方統計組織의 再編成

### 가. 市·道 統計組織의 擴大 補強

- 基本的으로 地域統計를 地自體가 스스로 作成할 수 있는 組織을 確保

- 中央部處가 의뢰하는 統計作成 뿐만 아니라 자체 필요한 統計의 作成能力 培養
- 專門人力이 統計部署에 安定的으로 勤務할 수 있는 與件造成

- 中央部處 統計機關과 有機的인 連繫性 維持

- 中央部處 統計作成業務의 代행을 위한 충분한 豫算 支援 確保

#### 나. 中央部處 地方統計組織의 再扁

- 一般目的 經常標本調查(月別調查)의 統計廳에 集中에 따른 資料蒐集 機能의 統合
- 農林水産部의 地方統計組織은 農業生産統計調查, 農事流通情報 蒐集 機能으로 국한하여 縮小 調整

#### 다. 地方統計組織間의 業務協力體制 構築

- 地方政府統計機關에 技術 支援
  - 調査員 訓練, 調査技術指導, 資料處理 內檢指導 등 相互 統計技術情報의 交流
- 統計廳과 農林水産部 地方統計組織의 共同作業 方案 講究

### 3. 作成統計의 內實化 및 世界化

- OECD 加入 등 先進國 進入에 對備 統計의 先進化·世界化 推進
  - 統計作成基準, 調査技法, 調査項目의 國際化로 資料의 國際間 比較性 提高
  - '93 新國民計定體系 등 國際機關에서 勸告하고 있는 新基準의 早速한 適用
- 多様な 統計生産으로 國家統計의 均衡發展 圖謀
  - 國民의 生活의 質을 測定할 수 있는 社會福祉統計의 開發
  - 쾌적한 生活과 地球環境保護에 필요한 基礎 環境統計 및 指標 開發
  - 一次 統計의 生産에서 加工統計·分析統計의 生産에 역점

○ 地方化時代に 必要な 地域統計의 體系的 作成

- 中央部處에서는 大規模 調査를 통한 細分된 地域統計 生産 提供
- 地方政府에서는 經常標本調査를 積極 開發하여 자체 活用

○ 統計調査 環境의 惡化에 對處하여 統計調査 方法의 改善努力 持續

- 全數調査를 止揚하고 小規模 精密標本調査로 代置하여 時宜性和 精度를 동시에 提高
- 電話調査, 郵便調査, 휴대용 컴퓨터를 利用한 調査 등 調査技法의 改善으로 克服

○ 行政統計作成技法의 開發과 活用の 極大化

- 各種 行政遂行의 結果에서 행정통계를 體系的이고 效率的으로 작성할 수 있도록 申告制度, 登錄制度 등 各種 政策立案時 統計作成을 위한 政策形成者의 配慮를 유도
- 최첨단 電算裝備를 活用하여 行政業務를 電算化하고 이를 利用하여 다양한 行政統計 生産
- 行政統計의 D/B를 各 部處에 構築하고 相互連結하여 各 機關의 다목적 利用에 提供

## V. 結 言

統計의 中要性에 대해서는 동서고금을 통해서 그 인식에 일치를 보이고 있다. 한 나라의 통계분야가 先進化되지 않은 상태에서 그 사회가 先進社會로 발전해 간 사례를 우리는 아직 보지 못하고 있는 것도 사실이다.

즉, 統計分野의 발전은 그 사회를 先進社會로 만들어 가는데 있어 없어서는 안될 중요한 일부인 것이다. 이제 우리사회도 여러 측면에서 先進社會로의 進入과 成熟이 期待되고 있는 현 시점에서 統計分野가 그에 상응하게 발전되지 못하고 있는 점은 우리들 통계인 모두의 覺醒을 촉구하고 있다.

우리나라가 명실상부한 先進國 隊列에 合流하고, 世界化에 앞장서기 위해서는 統計에 대한 政策的인 配慮와 과감한 투자가 요구되는 것이다. 先進國의 統計擔當機關이라 하면 많은 高級人力을 보유하면서, 질 높은 統計生産은 물론 研究分析에도 상당한 투자가 이루어지는 등 行政의 科學化에 앞장서 나가고 있는 실정이다.

따라서 오늘날 우리 사회를 先進社會로 이끌어 가는데 반드시 이루어야 할 統計分野의 先進化를 위하여 우리는 資料生産의 측면에서 統計組織을 확대하고 소요되는 豫算을 합리적으로 確保·活用 함은 물론 主要統計調査에 관련된 法令을 整備·強化하는 한편, 統計研究所의 設立 등을 통해 통계의 理論的·技術的 分野를 부단히 개발해 나가야 할 것이다.

또 資料의 活用側面에서는 통계자료의 분석활용을 極大化함으로써 통계생산과 이용의 양 부분이 相互 補完적으로 균형있게 발전해 가는 통계의 선진화를 이룩해야 할 것이다.

이러한 點과 관련하여 日本의 統計行政의 中長期構想과 新中長期構想은 우리에게 깨우치는 바가 많다고 본다. 이 두개의 中長期 構想은 각각 1985년 및 1995년에 學界와 統計作成機關 및 統計利用機關 代表가 함께 참여하여 작성한 것이다.

특히 新中長期構想에서는

- 1) 社會·經濟變化에 대응한 통계의 整備 및 開發
- 2) 主要統計調査의 實施時期的 調整
- 3) 應答者의 負擔輕減 및 調査方法의 改善
- 4) 調査結果의 利用 擴大
- 5) 統計調査의 효율적 實施와 正確性 確保
- 6) 國際統計協力推進 등

6개분야로 나누어 向後 推進方向을 종합적으로 제시하고 있다.

우리도 學界·統計作成機關·統計利用機關 삼자가 相互 有機的인 체계를 구축하여 우리나라의 統計制度나 統計內容이 명실상부한 선진화가 이룩될 수 있도록 하는 지참이 조속히 마련되어야 할 것이다.

< 參考 > 主要國의 統計組織 比較

	獨 逸	카나다	네덜란드	美 國	日 本	韓 國
<b>&lt; 統計組織 &gt;</b>						
○ 組 職 形 態	集 中 型	集 中 型	集 中 型	分 散 型	完 全 分 散 型	分 散 型 (集 中 型 加 味)
○ 中 央 統 計 機 關	聯 邦 統 計 廳  (10局38課)	統 計 廳  (6室, 16局 65課)	中 央 統 計 局  (1局, 4課 29下 部 組 織)	商 務 部 센서스局  (8局長補, 43課)	總 務 廳 統 計 局  (5部 17課)	統 計 廳  (3局 15課)
○ 地 方 組 織	16 州 政 府 統 計 廳	8 地 方 事 務 所	없 음	12 地 方 事 務 所	地 方 自 治 團 體 統 計 組 織	12 地 方 事 務 所
<b>&lt; 人 力 &gt;</b>						
○ 中 央 統 計 機 關 (名)	3, 226	4, 200	2, 770	3, 745	1, 743	475
(人 口 百 萬 名 當)	(52. 6)	(158. 4)	(182. 5)	(15. 0)	(14. 1)	(10. 8)
○ 他 部 處 및 中 央 銀 行	300	-	230	..	10, 517 <sup>1</sup>	2, 556 <sup>1</sup>
○ 地 方 組 織	13, 704	2, 200	-	3, 729	2, 704 <sup>2</sup>	1, 641 <sup>3</sup>
○ 統 計 人 力 合 計	17, 230	6, 400	3, 000	..	14, 964	4, 672
(人 口 百 萬 名 當)	(222. 3)	(241. 3)	(197. 6)		(121. 2)	(106. 0)
○ 人 口 (千 名)	61, 324	26, 521	15, 180	249, 224	123, 460	44, 056

주 : 1) 타부처의 지방사무소 인력이 포함되어 있음.

2) 47개 道·都·府·縣 인력으로서 市(區)·町·村의 통계인력(전임 2,000명, 겸임 10,000명)은 제외 되어 있음.

3) 통계청 지방사무소와 市·道 및 區·市·郡 통계인력임.

## 국가통계조직의 발전방향

토론1 : 서울대학교 통계학과 교수 박 성 현

제1회 「통계의 날」 기념 통계심포지움에서 우리나라 「국가통계조직의 발전방안」에 대하여 심도있게 논의하는 것은 매우 적절하다고 판단되며, 본인이 토론하게 된 것을 영광으로 생각합니다. 우선 이 발전방향에서 지적한 바와 같이 통계제도 운영의 방향을 통계청에 통계조정 권한을 강화하면서 집중형이 감미된 분산형제도를 취하는 것이 합당하다고 생각합니다. 우선 통계청의 통계조정권한을 강화하기 위해서는 통계청을 일반청의 수준으로 격상되어야 하고, 중앙정부의 통계조직이 자체통계조사기능을 가질 수 있도록 강화되어야 할 것이다. 예를 들면, 통상산업부, 환경부, 보건복지부, 건설교통부 등에도 통계조직이 강화되어 자체통계조사기능이 어느정도 갖추어 질 수 있도록 되어야 할 것이다. 또한 서울시 등의 지방자치단체에도 통계조직의 확대개편이 필요하다고 생각한다.

국가통계제도의 문제점에 대해서는 전적으로 발표자와 의견이 같으며, 국가통계제도의 개선과제에 대해서도 동의한다. 발표된 내용에 추가하여 국가통계제도의 개선방향에 대하여 다음의 몇가지 내용을 추가하고 싶다.

- (1) 먼저 통계청의 소속을 재정경제원에서 국무총리실로 바뀌는 것이 어떨까 제안하고 싶다. 아직까지 재정경제원 산하에 있으면서 경제통계는 엄청난 발전을 가져왔으나 사회통계, 보건 복지통계, 환경통계, 인구통계, 공업통계 등의 전반적인 통계발전은 상대적으로 저조하였던 것이 사실이다. 국가발전을 위해서는 국가사회의 모든 분야의 통계가 발전되어야 하며, 균형있는 발전을 촉진하기 위해서는 국무총리실에 소속되는 것이 타당하다.
- (2) 두번째로 통계의 질적향상과 통계기법의 개발을 지속적으로 수행하기 위해서 한국통계과학연구원(Korean Institute of Statistical Sciences)의 설립이 절실히 요구된다. 왜냐하면 통계자체를 연구하는 것을 중심으로 통계응용을 연구하는 econometrics, sociometrics, biometrics, technometrics, environmentrics 등을 포함시켜 연구하는 연구원이 바람직하며, 그러기 위해서는 좀 더 광범위한 의미를 갖는 통계과학이라는 용어를 쓰는 것이 적절해 보인다. 일본은 이미 1944년 전쟁을 수행하는 와중에서 통계수리연구소를 설립하였으며, 오늘날에는 연구원의 수만도 52명이나 되는 큰 조직으로 변모되어 일본의 통계연구의 본산이 되고 있다. 우리나라는 이미 상당히 늦은 감은 있으나, 조속히 통계과학연구원을 설립하여 체계적인 통계연구가 이루어졌으면 한다.

- (3) 세번째로 통계청을 비롯하여 국내 각 기관에서 생산되는 각종 통계정보를 축적 제공하는 통계 데이터 베이스를 이용자의 다양한 목적에 맞도록 개편하고 공급하기 위하여 “통계센타”의 설치가 필요하다. 통계생산자가 직접 통계정보를 공급하는 어려움이 많으며, 이용자의 다양한 요구를 순발력 있게 대응하는 것은 거의 불가능하다. 또한 세계화의 흐름속에서 국제기구(OECD, IMF, UNSTAT 등)와 주요 선진국으로부터 각종 통계정보를 신속히 입수하여 각종 매체를 통하여 제공하는 역할을 감당하기 위해서도 “통계센타”의 설립이 필요하다.
- (4) 네번째로 학계, 통계작성기관, 통계이용기관 간에 좀 더 유기적인 협조체계를 구축하여 통계선진화의 기틀을 마련하는 것이 매우 중요하다고 본다. 유기적인 협조체계를 만드는 방법으로 학계 중심으로 움직이는 한국통계학회의 운영을 통계작성기관이나 통계이용기관의 참여폭을 확대하여 모든 통계인들이 참여할 수 있도록 확대개편하여 운영하는 것이 하나의 방안이 될 수 있다. 또 하나의 방안은 통계청이 통계청이 좀 더 적극적인 자세로서 학계에 통계연구용역비도 제공하고, 학계·통계작성기관·통계이용기관이 공동으로 통계논문등을 연구할 수 있도록 연구모임의 장을 개발하는 것도 하나의 방법이 될 것이다. 이와 병행하여 통계협회에서는 이들 삼자간의 친목도모 행사를 다양하게 기획하여 실행하는 것도 바람직하다.
- (5) 마지막으로 통계작성, 통계연구, 통계이용등의 중요성을 정부가 인식하고 국민이 이해할 수 있도록 지속적인 홍보활동이 필요하고, 정보화시대에 즈음하여 통계자료의 가치를 높여가는 우리 스스로의 자구노력이 필요하다고 생각된다. 21세기에는 정보의 원천이 되는 통계자료의 수립, 축적, 가공, 분석등의 활동이 매우 중요한 자산가치를 가질 것이라고 믿어지며, 이러한 자산을 잘 관리하고 운영할 수 있는 것은 순전히 통계인의 역량에 달려있다고 생각된다.

## 국가통계조직의 발전방향

토론2 : 한국은행 부부장 이 성 만

통계인의 날이 제정되고 오늘 그 첫 기념식을 가지면서 우리나라의 통계가 당면하고 있는 문제점과 국가 백년대계를 염두에 두고 우리나라 통계의 앞날을 생각해 보는 것은 참으로 의의가 있고 뜻깊은 일이라고 하겠습니다.

특히 우리나라가 해방된 이후 건국과 경제발전과정에서 한국은행은 경제와 관련된 많은 통계를 개발, 작성하였으며 이중 상당부분이 해당기관에 이관된 이후 발전을 거듭하고 있는 모습을 보면서 가슴 뿌듯하게 느끼며 우리나라 경제발전을 위하여 지력과 정력을 아끼지 않았던 수 많은 선배제현의 크고 작은 노력과 희생에 감사할 수 있는 좋은 기회를 가지게 된 것을 통계의 날이 제정된 이날 통계인의 한사람으로서 기쁘게 생각합니다.

사실 통계는 사회간접자본의 하나로써 정보산업의 총아이며 정보의 생산, 전달, 공유에 따라서 경제적 가치가 크게 변화하는 정보화 사회에서 그 중요성은 두말할 필요도 없이 인정되고 있는 것입니다. 특히 경제발전단계가 낮고 경제규모가 작을 때에는 직관이나 추측에 입각한 의사결정이 가능했고 또한 실수가 있더라도 그 피해가 크지 않았고 어느 정도 수습할 수 있었습니다. 그러나 경제가 발전하여 양적 질적으로 확대 심화된 단계에서는 약간의 실수라도 크나큰 손실을 야기할 수 있기 때문에 의사결정에 있어서 매우 신중할 수 밖에 없으며 많은 자료와 확실한 근거 바탕으로 결정내리지 않으면 안되는 상황으로 전환되었습니다. 따라서 합리적 수리적 분석적인 즉 과학적인 사고방식을 갖지 않으면 안되는 것입니다.

이렇게 여건이 변화되는 상황에서 우리나라의 통계제도를 검토해 보는 것은 오늘이 「통계의 날」이라서 더욱 의미 있다고 하겠습니다. 김일현 통계국장께서 발표하신 국가통계제도의 발전방향은 우리나라가 당면하고 있는 통계제도의 문제점을 비교적 소상히 열거하고 날로 급변하는 통계환경을 고려하여 앞으로의 통계제도를 재정비해야할 방향과 우리나라의 통계제도 특히 통계청의 앞날에 대한 견해와 비전을 제시하고 있는 것으로 사료됩니다. 주제발표에서도 이미 충분히 언

급되었습니다만 일반적으로 일국의 통계제도유형을 분산형과 집중형으로 크게 구분하고 있는데 우리나라의 통계제도는 완전집중형도 아니고 분산형도 아닌 양자가 혼재된 혼합형이라고 할 수 있겠습니다. 따라서 우리나라의 통계체제로서는 어느 것이 바람직한가 하는 것은 이와 같은 우리나라의 현실 여건을 충분히 감안하여 볼 때 분산형 또는 집중형의 어느 특정한 제도로 일원화해 나가는 것보다는 각 제도가 가지고 있는 장점을 최대한 살리고 단점을 최소화 또는 보완해 나가면서 양자를 조화시키는 방향으로 개선해 나가야 할 것입니다. 이와 같이 양 제도를 조화시키는 방향에서 향후 국가통계제도의 거시적인 발전방안을 나름대로 제시해 본다면 다음과 같습니다.

우선 인구 및 광공업통계 등 전국적으로 광범위하게 이루어져야 하는 기초통계와 사회복지 및 환경통계 등 일반적으로 이용되는 국가통계는 업무효율상 규모의 경제효과가 크고 행정력을 이용할 수 있는 통계청과 같은 국가기관이 전담하여 편제하고 경제통계, 노동통계, 농림어업통계, 재정통계 등 각분야별 전문통계와 국민계정통계 등 기초통계를 토대로 한 2차적인 가공통계는 각 분야의 특수한 통계수요에 유연하게 부응할 수 있고 통계편제기술상 고도의 전문적인 기법이 필요한 경우가 많으므로 이들 통계를 직접적으로 이용하는 기관이나 전문적인 노하우가 축적되어 있는 기관이 담당토록 하는 것이 효율성면에서도 바람직하다 하겠습니다. 이와 같이 전문통계 및 가공통계의 경우 분산형체제를 유지함으로써 발생하는 통일성의 결여와 비능률적인 측면은 통계청과 통계작성기관의 대표 및 학계 등의 통계전문위원으로 구성되는 통계위원회에서 협의 조정토록 하면 충분할 것입니다.

또한 통계청이 현재 조사 편제하고 있는 기초통계 및 일반통계의 경우 통계에 대한 수요는 더욱 다양해지고 확대되고 있으나 조사여건은 날로 악화되고 있는 것이 사실이므로 통계청의 조직을 보다 강화하고 인적 물적 자원을 확충해 나가며 지방자치단체의 통계업무에 대한 지원기능을 보강하는 등 통계조사 여건을 보완하는 것도 시급히 추진해 나가야 할 것입니다. 이러한 통계제도라는 어느 의미에서 하드웨어적인 측면에서의 보완과 더불어 통계마인드 제고, 통계이용편의를 위한 DB구축, 대국민 통계협조분위기 확산 등 소프트웨어적이며 지적 심리적 측면에 대한 투자는 성과는 낮은 반면 힘들고 장기적이며 지루한 과업이지만 우리나라의 통계발전을 위해서는 먼저 국민의 지적변화가 우선되어야 할 것입니다.

끝으로 국민소득 및 산업연관표에 대한 당행의 견해를 말씀드리자면, 먼저 국민소득통계는 자금순환, 국제수지, 산업연관표 등과 함께 소위 국민계정을 구성하는 핵심통계이며 이들 국민계정통계는 당초 각각 별개로 발전하여 왔으나 1979년대 이후 UN 및 OECD 등 국제기구에서 국민경제의 움직임을 실물 및 금융면에서 보다 종합적으로 파악할 수 있도록 각 통계를 유기적으로 연결하여 작성하도록 권고함에 따라 미국, EU, 일본 등 많은 나라에서 이들 통계를 유기적으로 연결하여 작성하는 추세에 있다고 하겠습니다.

한편 이러한 국민계정통계중 자금순환, 국제수지는 미국, 일본을 비롯한 대부분의 나라에서 중앙은행이 작성하고 있는 것이 통계이나 국민소득, 산업연관표 등은 나라에 따라 중앙통계기관, 중앙은행 또는 여타 국가기관이 작성하고 있습니다. 그런데 우리나라의 경우에는 상기 통계들을 여러기관이 나누어 작성하지 않고 당행이 모두 종합하여 편제하고 있는데 이중 국민계정통계를 통계청으로 이관한다면 이는 국민계정통계의 연계성을 증시하는 국제적인 추세에 어긋날 뿐만 아니라 통계작성업무의 효율성이 떨어지는 등 국민계정통계의 발전에도 도움이 되지 않을 것입니다.

특히 1993년 SNA 개정으로 국민소득, 자금순환, 국제수지, 산업연관표 등의 작성 베이스를 일치시키는 작업을 진행시키고 있고 이 작업은 장기간이 소요된다는 점을 감안할 때 국민소득통계의 이관문제는 더 이상 논의할 필요성이 없다고 생각합니다. 더욱이 GNP 및 I-O 통계는 경제정책과 밀접한 관계에 있는 경제통계로 경제정책에 대한 조사분석기관이 담당하는 것이 보다 타당하며 통계적 처리가 중요시되는 업무가 아닌 점을 이해하셔야 할 것입니다.

제 1 회 『통계의 날』 기념

대학생 논문 경진대회

입상 논문

# 1등

## 고려대학교 학생들의 출신지역별 의식비교

안태근 · 오진승 · 김성환 · 강성민

(지도교수 : 허 명 회)

고려대학교 정경대학 통계학과

### 요 약

우리 나라 사람들은 예로부터 출신도별로 그 나름대로의 독특한 생활 방식과 사고방식을 지녀왔다고 생각한다. 따라서 개개인의 의식구조 및 사회현상에 대한 반응 역시 출신 지역에 따라 서로 다른 특징을 가질 것으로 생각되어 진다.

본 조사를 통해 고려대학교에 재학 중인 학생들이 각 출신도별로 생활 패턴과 의식구조에 있어 어떤 지역적 차별성을 가지고 있는지 알아보려고 한다. 각 출신도별 학생들의 정치적 성향 및 대학 생활에 있어서의 다양한 측면을 조사, 연구함으로써 앞으로 사회의 중심이 될 젊은이들의 의식구조를 이해하는데 도움이 되고자 한다.

### 1. 조사의 개요

#### 1.1 조사 목적

고려대학교에 재학 중인 학생들이 출신 지역별로 대학 생활 및 사회적, 정치적 견해에 있어 어떠한 차이가 있는 지를 살펴보고, 또 그러한 지역적 차이의 원인이 무엇인가를 알아보려고 한다.

- 소비 문화, 이성/결혼관, 대학 생활, 취업관, 정치/사회적 성향, 지역적 성향에 관한 각각의 설문 문항에 대하여 2차원 분할표를 작성하여 어떠한 지역적 차이가 있는 지를 알아보았다.
- 혼전 순결과 취업관에 대한 대응 분석을 통해 '성'과 '직업'에 대한 지역적 유사성과 차별성의 특징을 가시화 하였다.
- 출신 지역별로 자신이 선호하는 지역을 분할표 및 HISTOGRAM으로 나타내어 지역적 친밀도를 가시화 하였다.
- 지역적 성향에 대한 4점척도 10개 문항에 대해 수량화 2법, 군집분석, 다차원척도법(MDS)을 적용하여 지역간의 유사성과 차별성이 무엇인지를 알아보았다.

#### 1.2 조사 설계

- (1) 모집단 : 고대에 재학 중인 남녀 재학생
- (2) 표본 구성

- 출신지별/성별 분포

출신도	서울	경기	강원	충청	전라	경상	제주	계	남성	여성
인원수	25명	175명	151명	24명						

- 학년별 분포

학년	88	89	90	91	92	93	94	95	계
인원수	4명	7명	44명	17명	19명	32명	19명	33명	175명

- (3) 표본추출 방법 : 각 출신도별로 25명씩 층화추출하였고 출신도는 출신 고등학교 소재지를 기준으로 하였다. 조사 방법은 출구 조사를 이용하였고, 표본 확보가 어려운 제주도와 강원도 지역은 그 지역 출신 면접원을 통해 표본을 수집하였다.
- (4) 자료 수집 방법 : 자기 기입식 방법
- (5) 조사 도구 : 설문지 175부

1.3 주요 설문 내용

영역	문항내용	문항수
소비문화	용돈 규모, 용돈 출처, 용돈 지출, 슬값 부담 여부, 유행 등에 대한 관심	6 문항
이성 / 결혼관	이성 교제 및 결혼관에 대한 관련 사항	8 문항
대학생활	학과, 전공 및 대학생활 전반에 대한 관련사항	8 문항
취업관	선호 기업과 졸업 후 진로 등 취업에 대한 관련사항	6 문항
정치/사회적성향	정치/사회적 성향과 남북통일에 대한 관련사항	7 문항
지역성	지역적 성향에 대한 4점척도 문항	11 문항

2. 조사결과의 분석

2.1 소비문화에 대한 분석 결과

설문에서는 모두 6개 항목에 대한 질문을 통해 학생들의 소비 문화에 대한 지역적 차이가 나타나는가를 알아보고자 하였다. 그러나 모든 항목에서 지역적 차이가 두드러지게 나타나지는 않고 대부분의 학생들이 지역과는 관계없이 고른 소비 패턴을 보였다. <표 1.1>을 보면 용돈 규모에서 과반수 이상이 한 달에 15~25만원 정도의 용돈을 쓴다고 응답하였으나, 10만원 이하의 적은 돈을

쓰는 학생 수와 25~35만원, 심지어 35만원 이상의 많은 돈을 쓰는 학생 수도 무시할 수 없는 수치를 나타냈다. 이는 상당수의 학생들이 부모님으로부터 많은 용돈을 받거나 과외(전체의 약 25%)를 통해 학생수준이상의 돈을 쓰는 경우가 많아 학생들간의 경제적인 격차가 크다는 것을 보여준다. <표 1.2>를 보면 용돈 지출 규모의 순위가 식비-술값-책 구입-스포츠(당구,탁구,볼링 등) 등의 순으로 나타났다. 여기서 술값이 2위로 나타난 것은 표본에서의 남자의 비율(전체의 약 86%)이 높았고, 남학생들의 소비 문화에서 '술'이 차지하는 비중이 상당하다는 것을 보여준다. 여학생의 경우 '술값'이라는 문항에 답한 응답자는 없었다.

<표 1.1> 한 달에 쓰는 용돈 규모 (단,주거비는 제외) 유의확률 = 0.868

지역	응답	10 만원 이하	15 - 25 만원	25 - 35 만원	35 만원 이상
서울		2 (8%)	14 (56%)	5 (20%)	4 (16%)
경기		4 (16%)	13 (52%)	4 (16%)	4 (16%)
충청		3 (12%)	11 (44%)	8 (32%)	3 (12%)
강원		4 (16%)	11 (44%)	8 (32%)	2 (8%)
전라		3 (12%)	16 (64%)	6 (24%)	0 (0%)
경상		4 (16%)	11 (44%)	9 (36%)	1 (4%)
제주		4 (16%)	12 (48%)	7 (28%)	2 (8%)
계		24	88	47	16

<표 1.2> 용돈 지출 (3가지 선택)

☞ 분석을 위해 처음 2가지 응답만을 가지고 중복응답 처리한 결과임.

지출 결과	식비	술값	커피숍	의복구입	책 구입	연극, 영화 관람	스포츠
응답횟수	151	83	8	17	44	11	32
백분율	86%	47%	5%	10%	25%	6%	18%
순위	1	2	7	5	3	6	4

## 2.2 이성/결혼관에 대한 분석 결과

### (1) 이성 친구와 결혼 상대자 선택시의 고려 요건

<표 2.1>과 <표 2.2>는 이성 친구와 결혼 상대자 선택시 가장 중요하게 생각하는 점에 대한 2차원 분할표이다. 그 결과를 보면 대부분의 학생들이 이성 친구와 결혼 상대자에게 성격을 가장 중요한 선택 요건으로 택했다. 설문지를 작성할 때 결혼 상대자에게 가장 중요한 항목으로 대부분이 선택할 것으로 예상된 '사랑'을 제외시켰으나, '성격'이라는 항목도 제외시켰으면 더 흥미있는 결과를 얻을 수 있었을 것이라는 아쉬움이 있다. 이번 조사에서 차이를 보이는 점은 두가지이다. 첫째는 이성 친구를 사귄 때에는 중요하게 생각하지 않았던 '경제적인 능력'을 결혼 상대자를 고를 때 중요하다고 선택한 학생들이 있었는데, 지역별로는 경기도와 전라도의 경우가 상대적으로 많았다. 이는 두 지역이 결혼 문제에서 다소 현실적이라는 측면을 반영한다고 하겠다. 또한 경기

도의 학생은 이성 친구의 경우 '외모'를 중요시하는 특이함을 보였는데 이것은 경기도의 지역적 특성이라기 보다는 설문에 응한 학생들의 개인적 취향이라고 보여진다. 둘째는 충청도 출신의 경우 '집안 환경'을 결혼 상대자의 중요한 조건으로 비교적 많이 선택한 점이다. 이것은 유교적 전통이 강한 충청도의 지역성을 반영한 것으로 보인다.

<표 2.1> 이성 친구를 사귈 때 가장 중요하게 생각하는 점 유의확률 = 0.154

지역 \ 응답	경제적 능력	성 격	외 모	집안 환경	학 별
서울	0 (0%)	19 (79%)	5 (21%)	0 (0%)	0 (0%)
경기	0 (0%)	15 (60%)	10 (40%)	0 (0%)	0 (0%)
충청	0 (0%)	20 (80%)	3 (12%)	2 (8%)	0 (0%)
강원	0 (0%)	20 (80%)	4 (16%)	1 (4%)	0 (0%)
전라	0 (0%)	21 (88%)	3 (12%)	0 (0%)	0 (0%)
경상	0 (0%)	18 (72%)	7 (28%)	0 (0%)	0 (0%)
제주	0 (0%)	18 (72%)	7 (28%)	0 (0%)	0 (0%)
계	0	131	39	3	0

<표 2.2> 결혼상대자를 고를 때 사랑 외에 가장 중요하게 생각하는 점

유의확률 = 0.176

지역 \ 응답	경제적 능력	성 격	외 모	집안 환경	학 별
서울	1 (4%)	17 (71%)	3 (13%)	2 (8%)	1 (4%)
경기	5 (20%)	18 (72%)	0 (0%)	2 (8%)	0 (0%)
충청	2 (8%)	16 (64%)	2 (8%)	5 (20%)	0 (0%)
강원	2 (8%)	19 (76%)	2 (8%)	2 (8%)	0 (0%)
전라	4 (16%)	19 (76%)	1 (4%)	1 (4%)	0 (0%)
경상	0 (0%)	22 (88%)	2 (8%)	1 (4%)	0 (0%)
제주	0 (0%)	23 (92%)	2 (8%)	0 (0%)	0 (0%)
계	14	134	12	13	1

(2) 맞벌이에 대한 인식

<표 2.3>을 보면 많은 학생들이 맞벌이에 대해 긍정적인 대답을 하였다. 응답에서 차이를 보이는 지역은 서울·경기와 충청·제주이다. 여성의 결혼 후 취업을 경제적인 면과 여성의 사회 활동에 의한 자아실현이라는 두 가지 면으로 볼 때, 서울·경기의 경우 맞벌이를 '반드시 해야 한다'는 강한 필요성을 제기하고 있는데, 이는 경제적인 측면을 고려한 현실적이고 합리적인 선택으로 보여진다. 이에 반해 충청·제주는 '안 하는 것이 좋다'고 답함으로써 여성의 결혼 후 활동에 대해 전통적이고 보수적인 입장을 가지고 있음을 보여준다.

<표 2.3> 맞벌이에 대한 견해

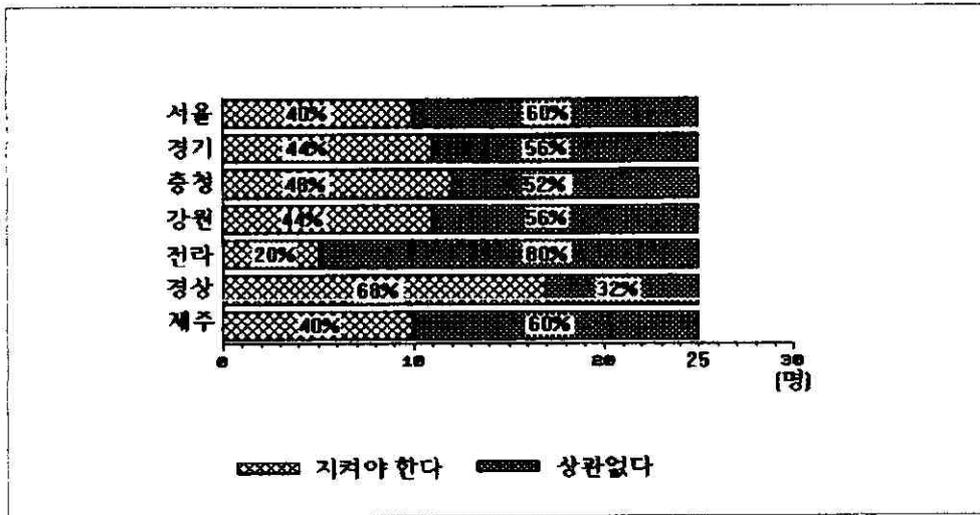
유의확률 = 0.156

지역	응답	반드시 해야한다	하는 것이 좋다	안하는것이 좋다	안 하여야 한다
	서울	5 (20%)	17 (68%)	3 (12%)	0 (0%)
경기	6 (24%)	17 (68%)	2 (8%)	0 (0%)	
충청	1 (4%)	15 (60%)	6 (32%)	1 (4%)	
강원	2 (8%)	20 (80%)	2 (8%)	1 (4%)	
전라	0 (0%)	21 (84%)	4 (16%)	0 (0%)	
경상	2 (8%)	17 (68%)	5 (20%)	1 (4%)	
제주	2 (2%)	16 (64%)	7 (28%)	0 (0%)	
계	18	123	31	3	

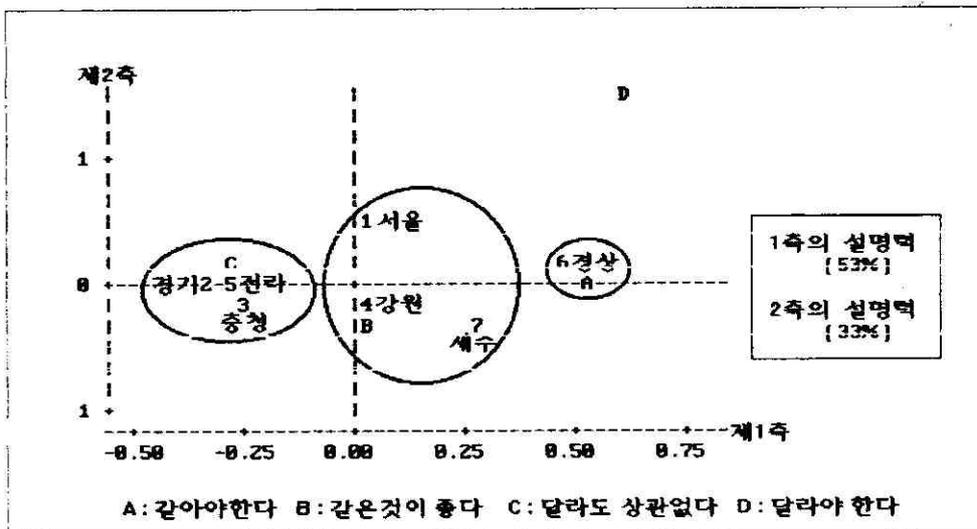
(3) 성(性)에 대한 인식

다음은 성에 대한 개방·보수성을 알아보는 질문에 대한 결과이다. <그림 2.1>을 보면 혼전 순결에 대한 질문에 대해 전체의 과반수 이상이 '상관없다'라고 답변하므로써, 결혼전 성관계가 결혼 생활에 지장이 되지는 않는다는 의견을 보였다. 이는 현재 젊은 층의 성에 대한 개방적인 시대 조류를 나타낸다. 하지만 경상도의 경우 68%의 학생들이 혼전 순결을 '지켜야 한다'고 응답하여 성에 대한 다소 보수적인 면을 보여준 데 반해, 전라도는 80%의 학생들이 '상관없다'고 응답하므로써 전라도와 경상도의 성에 대한 정반대의 입장을 보여주고 있다. 하지만 전라도의 편차가 타지역에 비해 너무 큰 것으로 보아 설문 응답자들이 주로 개방적인 학생들 위주로 우연히 추출되었을 가능성도 배제할 수 없다. 혼전 순결에 대한 결과는 이성 친구와 결혼 상대자가 같아야 하는지의 질문에 대한 대응분석 플롯에서도 그대로 나타난다. <그림 2.2>를 보면 경상도의 경우 '같아야 한다'는 응답과 가깝고, 전라·경기·충청 지역은 '달라도 상관없다'는 응답과 가까움을 알 수 있다. 따라서 이 결과를 통해서도 앞의 결과와 마찬가지로 성에 대하여 경상도 출신의 보수적인 면, 전라도 출신의 개방적인 면을 엿볼 수 있다.

<그림 2.1> 혼전 순결에 대한 지역별 응답



<그림 2.2> 이성친구와 결혼상대자의 동일성 여부에 대한 대응분석 플롯



### 2.3 대학생활에 대한 분석 결과

<표 3.1>을 보면 대부분의 지역이 학과를 선택할 때 '적성'을 가장 많이 고려했다고 응답했다. 그러나 경기도의 출신의 경우 '성적'을 가장 많이 고려했다고 응답하였는데, 이는 다른 지방의 학생들의 경우 대부분 일정 성적 이상의 학생들이 서울 소재 대학에 진학을 희망하므로 성적은 기본적인 조건으로 생각하는데 반해, 다른 지역에 비해 지역적으로 서울이 가까운 경기도의 경우 학과 선택에서 기본적인 성적을 가장 많이 고려하는 현상으로 이해된다. 또한 충청·강원 출신은 졸업후 전망을 비교적 많이 고려하는 면을 보이고 있다.

<표 3.2>를 보면 동문회·향우회 활동의 경우 경상도 학생의 경우 많은 수가 '적극적이다'라고 응답하였다. 이는 경상도 지역이 지연·학연의 연고 관계가 다른 지역에 비해 강하다는 것을 보여준다. 대부분의 학생들은 고교동문회 및 향우회 활동에 보통 이상의 관심을 가지고 있었지만, 충청도의 경우 '참여하지 않는다'는 의견이 비교적 많았다.

<표 3.3>을 보면 대학 생활의 목적에 대해서 '폭넓은 인간관계'가 '전문지식의 습득'보다 더 큰 비중을 차지한 것은 대학을 학문의 연구기관으로 보는 일반인들의 의식과는 다른 경향을 보였으나, 제주도의 경우는 예외적으로 '전문지식의 습득'을 제일 큰 목적으로 여기고 있다. 또한 소수이기는 하지만 강원도 출신은 '사회현실 참여'를 두번째로 많이 선택하여 다른 지역과 차별적인 경향을 보여준다.

<표 3.1> 학과 선택의 기준

유의확률 = 0.100

지역 \ 응답	적성을 고려하여	성적을 고려하여	사회적 지명도	향후 전망
서울	10 (40%)	7 (28%)	5 (20%)	3 (12%)
경기	7 (28%)	15 (60%)	1 (4%)	2 (8%)
충청	9 (36%)	6 (24%)	3 (12%)	7 (28%)
강원	7 (28%)	7 (28%)	4 (16%)	7 (28%)
전라	14 (56%)	8 (32%)	1 (4%)	2 (8%)
경상	13 (52%)	5 (20%)	2 (8%)	5 (20%)
제주	12 (52%)	6 (26%)	3 (13%)	2 (9%)
계	72	54	19	28

<표 3.2> 동문회 및 향우회 활동에 대한 참여 정도

유의확률 = 0.177

지역 \ 응답	적극적이다	보통이다	소극적이다	참여하지 않는다
서울	7 (28%)	12 (48%)	4 (16%)	2 (8%)
경기	6 (24%)	10 (40%)	8 (32%)	1 (4%)
충청	3 (12%)	12 (48%)	4 (16%)	6 (24%)
강원	8 (32%)	9 (36%)	6 (24%)	2 (8%)
전라	1 (4%)	13 (52%)	8 (32%)	3 (12%)
경상	11 (44%)	8 (32%)	3 (12%)	3 (12%)
제주	6 (24%)	9 (36%)	8 (32%)	2 (8%)
계	42	73	41	19

<표 3.3> 대학 생활의 목적

유의확률 = 0.043

지역 \ 응답	폭넓은 인간관계	전문지식의 습득	성공적인 취업	사회 현실 참여
서울	14 (56%)	10 (40%)	0 (0%)	1 (4%)
경기	15 (60%)	6 (24%)	3 (12%)	1 (4%)
충청	15 (60%)	4 (16%)	4 (16%)	2 (8%)
강원	10 (43%)	5 (22%)	2 (9%)	6 (26%)
전라	14 (58%)	7 (29%)	2 (9%)	1 (4%)
경상	12 (54%)	8 (36%)	1 (5%)	1 (5%)
제주	4 (18%)	12 (51%)	3 (13%)	4 (18%)
계	84	52	15	16

#### 2.4 취업관에 대한 분석 결과

<표 4.1>을 보면 취업 이유에 대하여 대부분의 학생들이 가장 큰 이유를 경제적 자립이라고 답하여 매우 현실적인 입장을 보이고 있다. 다만 경상도 출신의 경우 직장에서의 자아실현이라고 답변하였는데 이것은 경상도 출신이 자기 개발이나 자기 성취 욕구가 강하다는 증거라고 할 수 있다.

<표 4.2>를 보면 만약 자신과 연고가 없는 지방으로 발령이 났을 경우의 의견에 대한 응답으로 '어쩔 수 없으므로 따른다'나 '당연히 가야 한다'라는 답변이 많은 것을 알 수 있다. 이는 조직 사회에서 자신의 주장보다는 상급자의 명령을 수동적으로나마 존중하는 성향을 보인다. 다만 경기도 출신의 경우 타지역에서는 거의 나타나지 않은 '사직한다'는 의견이 20%를 차지해 자기 주장이 강한 모습을 보여준다.

<그림 4.1>은 입사시 기업 선택 요건에 대한 응답 결과의 대응분석 플롯이다. 그림을 보면 입사시 기업 선택 요건에 있어 출신 지역을 크게 두 집단으로 구분할 수 있었다. 서울과 경상도가 기업 선택 요건에 있어 기업의 '사회 기여도'를 선택한 것과 가까운 반면 다른 지역은 보수·안정성이라고 답변하므로써 보다 현실적인 성향을 보여주고 있다. '사원 복지'나 '개인의 시간적 여유'라는 응답자가 거의 없는 것은 위의 연고가 없는 지역으로의 발령에 대한 응답 결과에서 이미 이야기한 것처럼 학생들이 대체적으로 순종적이고 개인 생활의 보장에 대한 의식이 부족하다는 것을 반영한다.

<표 4.1> 졸업후 취업시 그 이유에 대한 견해

유의확률 = 0.175

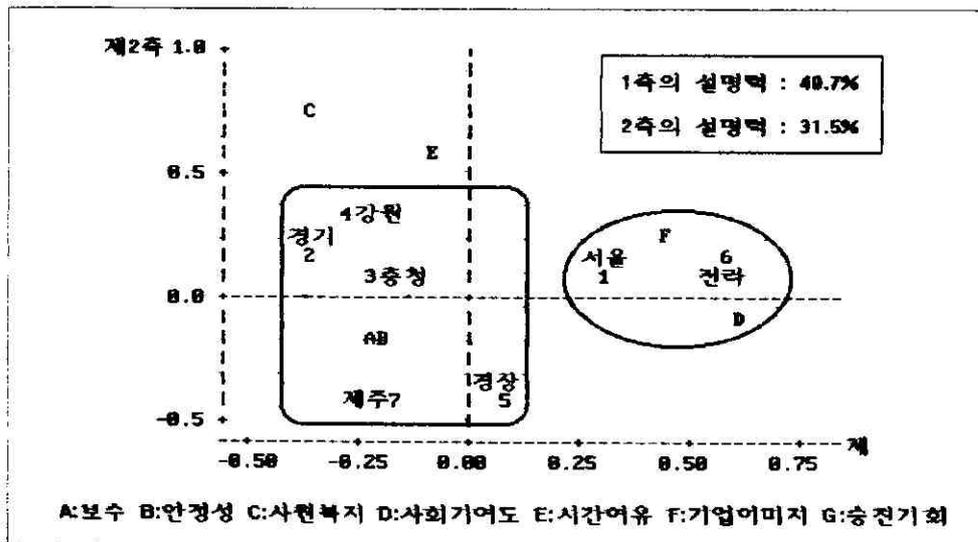
지역	응답	경제적 자립 때문에	직장에서의 자아실현	자영업을 위한 과도기	가계를 돕기 위해	기 타
서울	10 (43%)	8 (35%)	2 (9%)	3 (13%)	0 (0%)	
경기	14 (56%)	8 (32%)	1 (4%)	2 (8%)	0 (0%)	
충청	12 (57%)	5 (24%)	3 (14%)	1 (5%)	0 (0%)	
강원	15 (60%)	6 (24%)	3 (12%)	1 (4%)	0 (0%)	
전라	19 (79%)	3 (13%)	1 (4%)	0 (0%)	1 (4%)	
경상	7 (29%)	12 (50%)	3 (13%)	0 (0%)	2 (8%)	
제주	14 (64%)	5 (23%)	2 (9%)	0 (0%)	1 (4%)	
계	91	47	15	7	4	

<표 4.2> 자신과 연고가 없는 지방으로의 발령에 대한 견해

유의확률 = 0.015

지역	응답	당연히 가야한다	어쩔 수없으므로 따른다	불이익을 당하 더라도 가지않는 방법을 찾는다	사직 한다
서울	9 (36%)	13 (52%)	3 (12%)	0 (0%)	
경기	5 (20%)	9 (36%)	6 (24%)	5 (20%)	
충청	3 (12%)	15 (60%)	6 (24%)	1 (4%)	
강원	2 (8%)	17 (68%)	5 (20%)	1 (4%)	
전라	9 (38%)	12 (50%)	3 (12%)	0 (0%)	
경상	10 (40%)	10 (40%)	4 (16%)	1 (4%)	
제주	9 (36%)	14 (56%)	2 (8%)	0 (0%)	
계	47	90	29	8	

<그림 4.1> 입사시 기업선택 요건에 대한 대응분석 플롯



## 2.5 정치/사회적 성향에 대한 분석 결과

<표 5.1>과 <표 5.2>는 각각 국내 정치 문제에 대한 관심도와 선호 정당에 관한 응답의 2차원 분할표이다. 두 표를 보면 현재의 대학생들이 정치적인 문제에 대하여 관심이 적다는 것을 알 수 있다. 이것은 개인적인 사고방식이 학생들 사이에 널리 보편화되었기 때문이라고 본다. 아직은 국내 정치 문제에 대해 관심을 갖는 학생들의 수가 더 많았지만 '아니다'라고 답변한 학생도 상당수(63%)에 이르렀다. 특히 서울·충청 지역의 경우 무관심의 비율이 더 높은 비중을 차지하고 있으며, 선호 정당 역시 두 지역은 거의 없는 것으로 나타났다. 이에 비해 제주도는 '매우 그렇다'라고 답한 사람이 상대적으로 많아 정치적인 관심도가 타지역에 비해 높은 것으로 나타났다. 우리나라의 정당 구조는 지역 대표적인 정당 구조라고 할 수 있다. 그러나 정당선호에 있어서 실제로 지역적인 특성이 별로 나타나지 않았다는 것도 이번 조사에 있어서 특이할 만하다.

<표 5.3>을 보면 국회의원 선거시 인물 선택의 기준에 대한 물음에 거의 모든 지역이 '지역 발전의 기여도'라고 답변하였는데, 이는 대학생임에도 불구하고 국회의원을 중앙정치의 역할을 담당하는 인물이라기 보다는 단순히 지역개발의 역할을 담당하는 인물로 여기는 편협한 시각을 보여 주었다. 다만 충청도 지역은 '도덕성'을 중시하는 답변을 보여 유교적인 전통을 중요하게 하는 지역적인 특성이 이 응답에서도 나타났다. 정치에 대한 관심도와 마찬가지로 남북통일에 대한 응답에서도 대학생들의 변화된 의식이 드러났다. <표 5.4>를 보면 '민족 화합 차원에서 통일이 되어야 한다'는 통일의 당위성을 답한 수가 많았지만 '평화 상태가 유지된다면 현 상태도 무방하다'는 대답도 42명(약 24%)이나 되었다. 이것은 통일 후의 사회 혼란으로 인한 국민적인 부담 등을 우선 걱정하는 변화된 전후 세대의 통일관을 엿볼 수 있다.

<표 5.1> 국내 정치문제에 대한 관심

유의확률 = 0.003

지역 \ 응답	매우 그렇다	그렇다	아니다	결코 아니다
서울	2 (8%)	8 (32%)	14 (56%)	1 (4%)
경기	6 (24%)	12 (48%)	7 (28%)	0 (0%)
충청	2 (8%)	10 (40%)	12 (48%)	1 (4%)
강원	5 (20%)	19 (76%)	1 (4%)	0 (0%)
전라	3 (12%)	13 (52%)	9 (36%)	0 (0%)
경상	3 (12%)	12 (48%)	9 (36%)	1 (4%)
제주	11 (44%)	6 (24%)	8 (32%)	0 (0%)
계	32	80	60	3

<표 5.2> 선호 정당

유의확률 = 0.000

지역 \ 응답	민자당	민주당	자민련	신민당	재야 단체	없 다
서울	0 (0%)	2 (8%)	0 (0%)	0 (0%)	0 (0%)	23 (92%)
경기	1 (4%)	7 (28%)	0 (0%)	0 (0%)	2 (8%)	15 (60%)
충청	0 (0%)	0 (0%)	1 (4%)	0 (0%)	0 (0%)	24 (96%)
강원	3 (12%)	8 (32%)	1 (4%)	0 (0%)	3 (12%)	10 (40%)
전라	0 (0%)	6 (24%)	0 (0%)	0 (0%)	3 (12%)	16 (64%)
경상	0 (0%)	4 (16%)	0 (0%)	0 (0%)	1 (4%)	20 (80%)
제주	1 (4%)	5 (20%)	0 (0%)	0 (0%)	9 (36%)	10 (40%)
계	5	32	2	0	18	118

<표 5.3> 국회의원 선거시 인물 선택 기준

유의확률 = 0.412

지역 \ 응답	사회적 지명도	소속 정당	지역 발전 기여도	도덕성	혈연 및 학연
서울	1 (4%)	0 (0%)	13 (52%)	9 (36%)	2 (8%)
경기	1 (4%)	3 (12%)	13 (52%)	8 (32%)	0 (0%)
충청	1 (4%)	0 (0%)	10 (40%)	14 (56%)	0 (0%)
강원	3 (12%)	3 (12%)	15 (60%)	3 (12%)	1 (4%)
전라	0 (0%)	3 (13%)	13 (54%)	7 (29%)	1 (4%)
경상	2 (8%)	2 (8%)	11 (44%)	9 (36%)	1 (4%)
제주	1 (4%)	1 (4%)	15 (60%)	8 (32%)	0 (0%)
계	9	12	90	58	5

<표 5.4> 남북통일에 대한 견해

유의확률 = 0.160

지역 \ 응답	민족화합 차원에서 통일이 되어야한다	평화가 유지된다면 현상태도 무방하다	혼란이 예상되므로 되어서는 안된다
서울	17 (68%)	6 (24%)	2 (8%)
경기	22 (88%)	3 (12%)	0 (0%)
충청	16 (67%)	8 (33%)	0 (0%)
강원	15 (60%)	10 (40%)	0 (0%)
전라	18 (75%)	6 (25%)	0 (0%)
경상	20 (83%)	3 (13%)	1 (4%)
제주	19 (76%)	6 (24%)	0 (0%)
계	127	42	3

## 2.6 지역의 성향에 대한 분석 결과

### (1) 선호 지역에 대한 지역별 응답

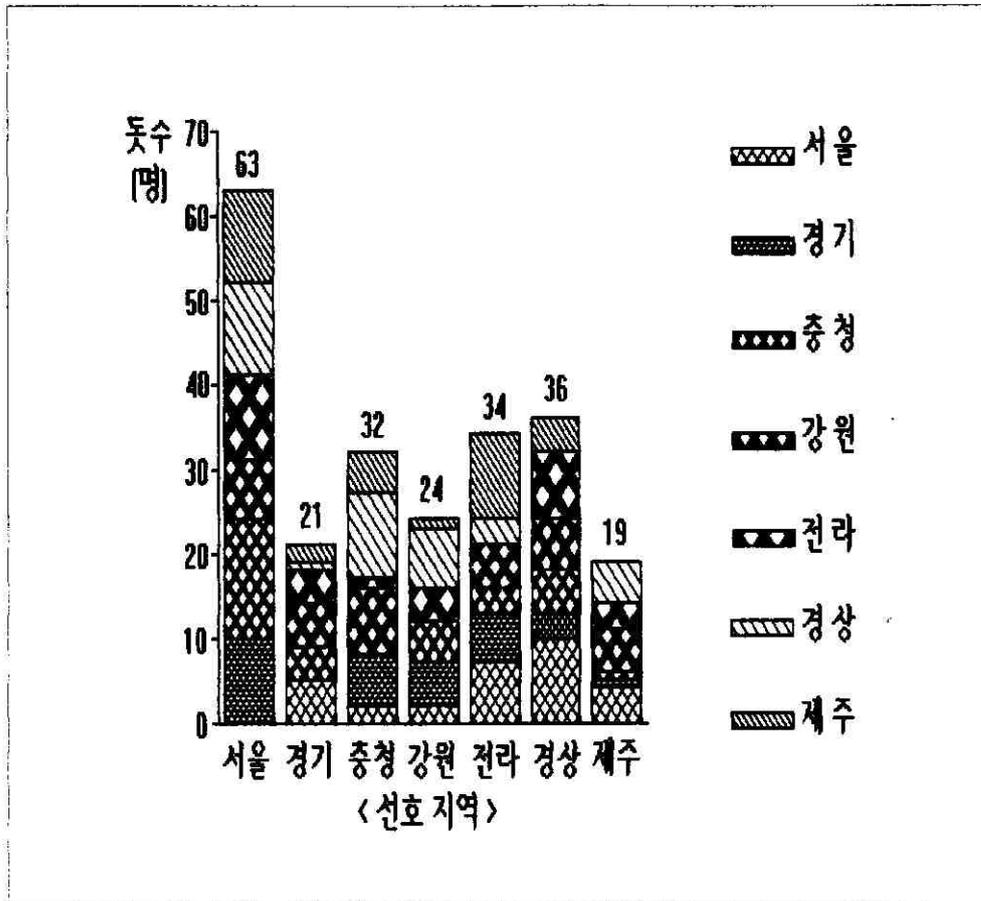
<표 6.1>과 <그림 6.1>을 보면 자신이 선호하는 지역을 두개만 적으라는 질문에 대한 응답에서 자신의 지역을 답한 것을 제외한 나머지 자료를 종합한 결과 서울이 63명으로 가장 많음을 알 수 있다. 이는 젊은 층의 서울 선호를 잘 반영하고 있다고 할 수 있다. 이밖에 경상·전라·충청이 근소한 차이로 뒤를 잇고 있고 이에 비해 강원·경기·제주에 대한 선호는 적은 수를 보였다. 이것은 강원·경기·제주에 비해 영호남에 대한 지역 인지도가 강하고, 선호 지역을 주변 친구와 연관시켜 생각하는 대학생들의 특성으로 생각하였을 때 경기·강원·제주 출신 지역의 학생수가 전라·경상에 비해 상대적으로 적기 때문일 것으로 생각된다.

지역 편견의 감정적 측면을 지역인간의 호오(好惡, like-dislike)의 차원에서 평가하여 응답자들로 하여금 각 지역인에 대해 얼마나 좋아하는지 혹은 싫어하는지를 척도(아주 싫어한다-아주 좋아한다)상에서 평가하러 하였으나 설문 시간과 응답자들의 거부감 등을 고려하여 위와 같은 질문을 하였다. 실제로 호오태도의 분석 결과에서도 영남인과 호남인간의 갈등이 두드러지지 않았으며 이는 우리 사회의 지역 감정의 문제는 영호남간의 감정적 갈등으로만 파악하려는 입장이 틀렸음을 보여주었다고 할 수 있다.

<표 6.1> 호감을 느끼는 지역

지역 선호지역	서울	경기	충청	강원	전라	경상	제주	지역별 응답두수
서울		10 (40%)	14 (4%)	7 (28%)	10 (40%)	11 (44%)	11 (44%)	63
경기	5 (20%)		4 (16%)	5 (20%)	4 (16%)	1 (4%)	2 (8%)	21
충청	2 (8%)	6 (24%)		8 (32%)	1 (4%)	10 (40%)	5 (20%)	32
강원	2 (8%)	5 (20%)	5 (20%)		4 (16%)	7 (28%)	1 (4%)	24
전라	7 (28%)	6 (24%)	3 (12%)	5 (20%)		3 (12%)	10 (40%)	34
경상	10 (40%)	3 (12%)	5 (20%)	6 (24%)	8 (32%)		4 (16%)	36
제주	4 (16%)	1 (4%)	1 (4%)	5 (20%)	3 (12%)	5 (20%)		19

<그림 6.1> 선호지역에 대한 지역별 응답



(2) 출신 지역에 대한 인식도

<표 6.2>를 보면 출신 지역의 개발정도에 대한 질문에서 지역적 차이가 분명함을 알 수 있다. 서울과 경상은 '개발이 잘 되었다'라고 대부분 응답한 반면 충청·강원·전라는 '개발되지 않았다'라고 응답했으며, 특히 전라의 경우 '개발되어 있다'라고 응답한 학생이 전혀 없어 거점성장전략의 채택으로 인한 서울-부산의 축선에서 제외된 전라도의 불만을 강하게 나타내고 있다.

반면에 경기와 제주는 '개발되었다'는 측과 '개발되지 않았다'는 측이 거의 반반을 이루고 있다. 경기도의 경우는 서울-인천의 집적 이익의 추구로 인한 기타 농촌 지역의 저개발, 제주도는 관광지로의 개발과 실제 거주 지역의 개발 차이 때문에 나타난 결과로 보인다.

<표 6.3>을 보면 출신 지역에 대해 애향심을 묻는 질문에 대해서 대부분의 학생들이 애향심이 가지고 있다고 대답을 했음을 알 수 있다. 그러나 서울·경기는 비교적 지역성이 약해 애향심이 없는 학생수가 상대적으로 많았고, 이에 반해 전라·제주는 애향심이 아주 강하다고 응답한 수가 많았다. 이는 지역간의 단결 정도의 차이를 보여준다고 할 수 있다.

<표 6.4>를 보면 주변 친구들의 출신 지역별로 다양한가의 질문에 경상도 출신은 '매우 그렇다'라고 답해 친구를 사귄 때 지역을 고려하지 않는 성향을 보였지만, 경기와 제주는 '아니다'라고 답

한 학생이 많아 특정 지역을 선호하는 모습을 보였다. 제주는 섬이라는 지리적 특성에 의한 설명이 가능하겠으나 경기는 일반적인 지역성이 미비한데도 다양하지 않은 경향을 보이고 있다.

<표 6.2> 나의 출신 지역은 타지역에 비해 개발되어 있다      유의확률 = 0.000

지역 \ 응답	매우 그렇다	그렇다	아니다	결코 아니다
서울	14 (58%)	9 (36%)	2 (8%)	0 (0%)
경기	0 (0%)	11 (44%)	12 (48%)	2 (8%)
충청	1 (4%)	7 (28%)	17 (66%)	0 (0%)
강원	1 (4%)	3 (12%)	10 (40%)	11 (44%)
전라	0 (0%)	0 (0%)	15 (60%)	10 (40%)
경상	1 (4%)	19 (76%)	4 (16%)	1 (4%)
제주	1 (4%)	11 (44%)	11 (44%)	2 (8%)
계	18	60	71	26

<표 6.3> 나는 출신 지역에 대한 애乡심을 가지고 있다      유의확률 = 0.002

지역 \ 응답	매우 그렇다	그렇다	아니다	결코 아니다
서울	0 (0%)	16 (64%)	9 (36%)	0 (0%)
경기	2 (8%)	16 (64%)	7 (28%)	0 (0%)
충청	3 (12%)	21 (84%)	1 (4%)	0 (0%)
강원	4 (16%)	19 (76%)	2 (8%)	0 (0%)
전라	9 (36%)	14 (56%)	2 (8%)	0 (0%)
경상	5 (20%)	17 (68%)	3 (12%)	0 (0%)
제주	9 (36%)	13 (52%)	3 (12%)	0 (0%)
계	32	116	27	0

<표 6.4> 내 주위의 절친한 친구들은 출신 지역이 다양하다      유의확률 = 0.000

지역 \ 응답	매우 그렇다	그렇다	아니다	결코 아니다
서울	4 (16%)	16 (64%)	4 (16%)	1 (4%)
경기	7 (28%)	7 (28%)	11 (44%)	0 (0%)
충청	2 (8%)	21 (84%)	2 (8%)	0 (0%)
강원	0 (0%)	17 (68%)	8 (32%)	0 (0%)
전라	1 (4%)	17 (68%)	7 (28%)	0 (0%)
경상	11 (44%)	9 (36%)	5 (20%)	0 (0%)
제주	1 (4%)	10 (40%)	12 (48%)	2 (8%)
계	26	97	49	3

### (3) 수량화 방법 II, 군집분석 및 다차원척도법에 의한 지역성향 분석결과

지역성향을 묻는 질문 중 첫번째 질문인 선호지역에 대한 질문을 제외하고 나머지 4점 척도 10개 문항에 대해 수량화 2법과 군집분석을 적용하였고, 다차원 척도법을 위해 4점 척도로 나타난 10개 문항의 결과를 다시 2점 척도의 응답 결과('예', '아니오')로 변환을 하여 각 출신지별로 '예'라고 답한 사람들 수의 합의 차를 각 출신지별 간의 비유사성의 거리로 계산하였다. (각 문항에 대하여 '매우 그렇다'와 '그렇다'의 응답자 수를 합하여 '예'의 응답자 수로 변환하고, 또 '아니다'와 '결코 아니다'의 응답자 수를 합하여 '아니오'의 응답자 수로 변환하였다.)

#### (가) 수량화 방법 II 에 의한 분석 결과

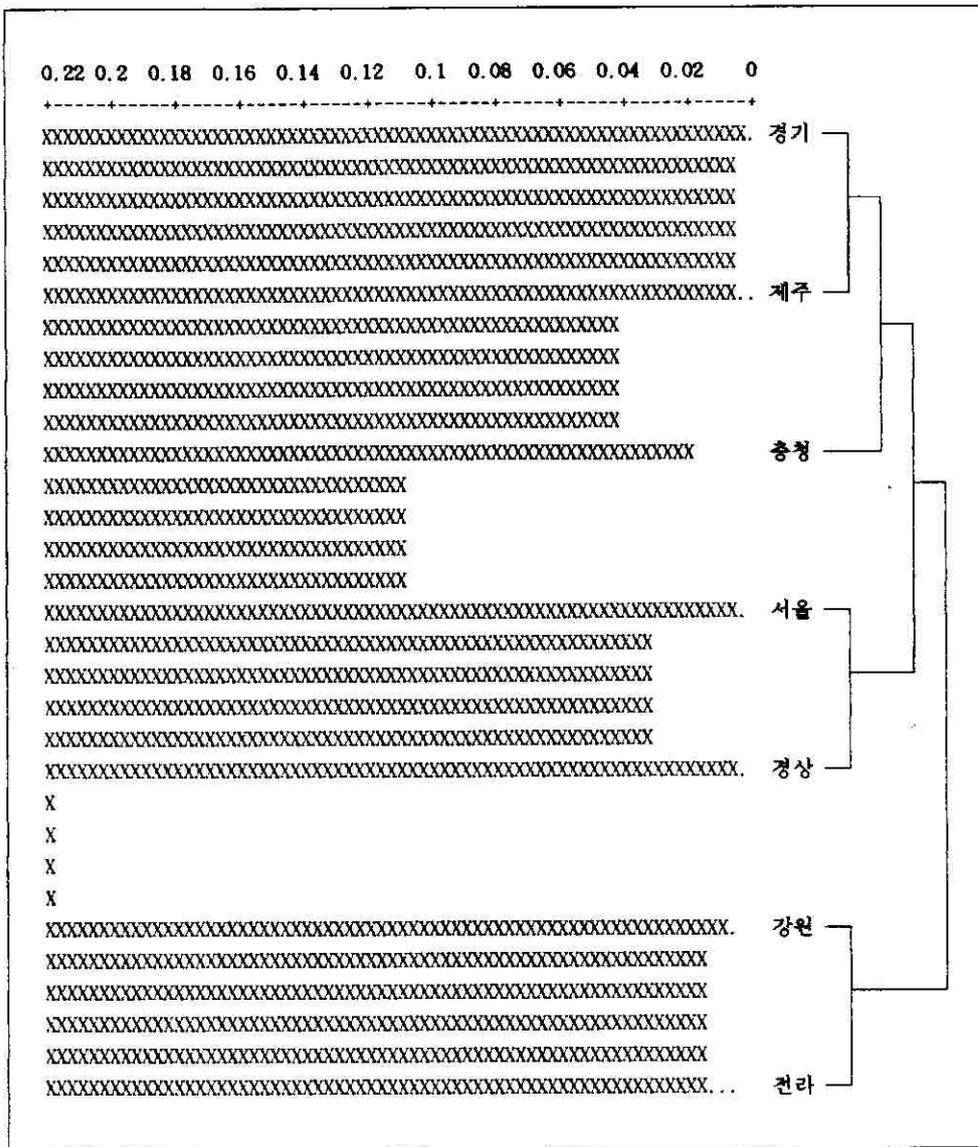
다음의 표는 정준상관분석에 의한 수량화 결과이다. 외적기준을 출신지역으로 하고 10개의 4점 척도 문항을 설명변량으로 했을 때 제 1축의 설명력은 40.66%, 제 2축의 설명력은 22.85%로 두 개의 축을 동시에 고려했을 때, 총 63.5%로 비교적 큰 설명력을 가진다고 볼 수 있다. 표에서 가장 큰 범위를 갖고 있는 것은 출신 지역의 개발 여부로 개발 되지 않은 지역은 음(-)의 부호를 가진다. 따라서 강원도와 전라도 지역이 개발이 잘 되지 않았다고 느끼는 지역임을 알 수 있다. 두 번째로 큰 범위를 갖는 지역감정의 존재 여부에 대한 질문과 세 번째로 큰 범위를 갖는 주변 친구들의 출신지역 다양성의 여부를 묻는 문항에서는 '그렇다'와 '아니다'가 음(-)의 부호를 동시에 가지고 있어서 해석을 하는 데 어려움이 있다.

문항	내용	1 측	2 측
외적기준 ( 지역 )	서울	1.849	1.438
	경기	0.059	-0.747
	충청	-0.061	0.023
	강원	-1.171 (3.16)	0.983 (2.63)
	전라	-1.311	0.633
	경상 제주	0.659 -0.021	-1.147 -1.237
지역감정은 존재한다	매우 그렇다	0.077	-0.926
	그렇다	-0.013 ②	0.174 ②
	아니다	-0.103 (1.82)	0.014 (2.82)
	결코 아니다	1.717	-1.876
자신의 출신지역은 개발되어 있다	매우 그렇다	1.719	2.105
	그렇다	0.579 ①	-0.715 ①
	아니다	-0.471 (2.96)	-0.195 (2.82)
	결코 아니다	-1.241	0.725
지역감정은 정치적인 목적으로 만들어졌다	매우 그렇다	-0.145	2.105
	그렇다	0.075	-0.715
	아니다	0.045 (0.45)	-0.195 (0.67)
	결코 아니다	-1.241	0.725
개인의 성격은 출신지역의 영향을 받는다	매우 그렇다	0.019	0.795
	그렇다	-0.061	-0.105
	아니다	0.099 (0.25)	0.035 (0.90)
	결코 아니다	-0.151	0.505
연고지 프로야구팀을 좋아한다	매우 그렇다	0.193	0.180
	그렇다	-0.197	-0.070
	아니다	0.083 (0.39)	-0.070 (0.29)
	결코 아니다	-0.007	0.220
애항심을 가지고 있다	매우 그렇다	-0.183	-0.179
	그렇다	0.017	-0.009
	아니다	0.147 (0.33)	0.251 (0.43)
	결코 아니다	.	.
특정 지역사람을 좋아하지 않는다	매우 그렇다	0.039	0.064
	그렇다	-0.161	0.174
	아니다	-0.021 (0.41)	-0.036 (0.31)
	결코 아니다	0.249	-0.136

국내 정치문제에 관심이 많다	매우 그렇다	-0.237		-0.397	
	그렇다	-0.067		0.243	
	아니다	0.203	(0.50)	-0.087	(0.75)
	결코 아니다	0.263		-0.507	
주위 친구들의 출신지역이 다양하다	매우 그렇다	0.737		-0.367	
	그렇다	-0.103	③	0.303	③
	아니다	-0.243	(1.16)	0.663	(1.87)
	결코 아니다	0.917		-1.117	
대학 생활에서 지역감정을 느낀다	매우 그렇다	-0.879		1.001	
	그렇다	-0.119	④	0.121	④
	아니다	0.181	(1.06)	0.011	(0.51)
	결코 아니다	-0.479		-0.509	
측의 설명력		40.66 %		22.85 %	

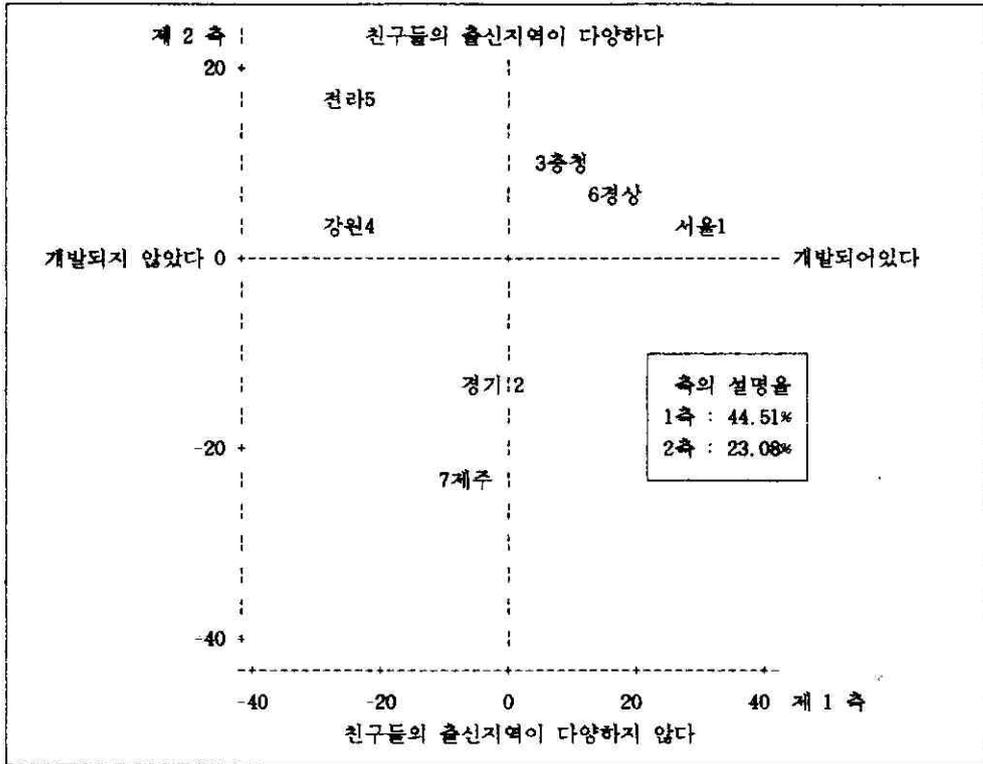
(나) 군집분석 의한 결과

4점 척도 10개 문항을 가지고 7개 출신지역에 대해 군집분석을 한 결과는 다음과 같다. 우선 주변 친구들의 출신지역의 비다양성에 의해 경기도와 제주도가 같은 군집으로 묶였고, 자기 지역에 대한 개발 인지 여부에 의해 잘 개발되어 있다는 지역으로 서울과 경상도가 같은 군집으로, 미개발지역인 강원도와 전라도가 같은 군집으로 나타났다. 그리고 충청도가 경기·제주 그룹에 묶인 것은 충청도의 개발 여부가 경기·제주와 비슷하기 때문이라고 생각 되어 진다.



(다) 다차원 척도법에 의한 분석 결과

아래의 그림을 보면 제 1축의 설명력은 44.51%, 제 2축의 설명력이 23.08%로 두 축에 의해 약 67.6% 정도가 설명되므로 다차원 척도법에 의한 지역성향 지도가 상당히 의미를 가진다고 할 수 있다. 우선 제 1축은 지역의 개발정도 여부의 축으로 서울과 경상도 지역이 개발이 잘 되었다고 볼 수 있고, 전라도와 강원도 지역이 개발이 잘 되지 않았다고 볼 수 있다. 제 2축은 주변 친구들의 다양성을 나타내는 축으로, 제주와 경기도 지역 출신 학생들이 주변 친구의 출신지역이 다양하지 못하고 전라도와 충청도 출신의 학생들이 주변 친구들의 출신지역이 다양하다고 볼 수 있다. 따라서 이것은 앞에서 이야기한 군집분석과 결과가 일치함을 알 수 있다.



### 3. 결 론

조사의 계획 단계에서는 소위 '지역 감정'이라는 단어에서 동기를 얻어 일반적으로 생각하는 특정한 사항에 대해 실제로 고대의 학생들 사이에서도 차이가 나타나는가를 알아보려고 하였다. 그러나 지역개발이라는 사실적인 질문 외에는 그다지 커다란 지역적 차이를 발견할 수가 없었다. 다만 경상도 지역이 개인적인 면에서 개방적인 반면에 성에 대한 가치관이 보수적이었으며, 충청도 지역은 아직도 유교적 전통이 강하게 남아 있는 경향을 보였다. 전라도 지역은 결혼·이성관에서 현실적인 면과 성에 대해 개방적인 면을 보였고, 경기도는 자기 주장이 강하고 현실적이며 서울에서 생활권을 형성하려는 경향을 보였다. 전체적인 응답에서 정치와 통일문제에 있어서 대학생들의 의식이 점차 방관적으로 바뀌고 있다는 것은 긍정적으로만 받아들일 수 없는 사항이라고 하겠다.

설문지 작성의 경험 부족으로 인해 작성할 때의 의도와는 전혀 다른 결과를 얻은 것도 있었고, 분석에는 부적당한 형태의 설문문항도 조사가 끝난 후에 발견하여 분석에 어려움이 따르기도 했다. 또 지역별로의 표본수가 그리 크지 않아 표본이 전체적인 지역별 의견을 대변한다고 보기는 힘들겠다. 앞으로 이와 같은 분석에서는 사전의 치밀한 준비와 계획이 선행되어야 할 것이다.

#### < 참고 문헌 >

- 1) 김종철·최장집 외 (1991), 지역 감정 연구, 학민사.
- 2) 허명희 (1992), 수량화 방법론의 이해, 자유아카데미.
- 3) 김충련 (1994), SAS라는 통계 상자, 데이타리서치.

## 로짓모형을 이용한 1994년 한국프로야구의 구단별 특성 분석

정진우, 김대환, 서두성

(지도교수 : 유성모)

고려대학교 경상대학 응용통계학과

### 요 약

야구경기에서 승·패에 관련된 여러가지 요인들중 팀의 승리와 패배에 결정적인 역할을 하는 요인을 찾기위해 로짓모형을 이용하였다. 요인들중 유의성이 있는 요인을 변수선택기법(BACKWARD)을 통해 추출해내었다. 이러한 변수선택 과정을 거쳐 선택된 변수와 선택되지 않은 변수들은 모두 나름대로 의미를 가지며 변수들의 표준화된 회귀계수로 로짓모형을 도출함으로써 한국프로야구 8개구단의 구단별 특성을 현실에 맞게 분석하였다.

### 1. 서 론

한국프로야구는 1982년 탄생된 이래 해를 거듭할수록 국민들의 관심이 모아지면서 6개구단으로 출발하여 현재는 8개구단에서 총 450명 이상의 프로선수들이 뛰고 있다. 인기가 상승함에 따라 각 구단에 대한 승·패를 놓고 관심이 집중되는 것은 당연한 결과라고 하겠다. 이에 본 논문에서는 1994년도 각 구단의 여러 가지 자료를 이용하여 로짓모형을 만들어 구단의 승·패 결정에 있어서의 주된 요인을 분석하는데 중점을 두었다. 단순히 승·패요인을 분석하기보다는 될 수 있으면 이용가능한 자료를 수집하여 로짓모형을 통해서 구단의 특성과 장·단점을 밝혀내어 1994년 각 구단의 특성을 한 눈에 알 수 있는 모형을 도출하기 위해 노력을 했다.

변수 선택과정에서 야구경기의 승·패에 영향을 미칠 수 있는 변수는 방어율, 장타율, 홈런수, 삼진수, 도루수, 사사구수, 실책수, 감독의 작전, 선수 개개인의 다양한 능력 등 무수히 많지만 본 논문에서는 구단별로 경기의 승·패에 결정적인 영향을 미치는 변수를 변수 선택과정을 거쳐서 선택하였다.

### 2. 본 론

#### 2.1 변수의 정의

전통적인 회귀모형에서는 종속변수와 독립변수가 모두 연속형이다. 하지만 팀의 승·패처럼 이항형(binary-type) 종속변수일 때에는 일반적인 회귀모형을 적합시킬 수가 없다. 따라서 이러한 모형을 적합시키기 위해서 본 논문에서는 1994년 페넨트레이스의 총경기(구단별 126경기)의 자료를 사용하였으며 로짓 모형의 적용을 위하여 무승부가 된 경기는 제외시켰다. (무승부 경기 수는 총 18 경기이다)

본 논문에서 쓰인 변수들에 대하여 살펴보면 다음과 같다.

TEAM : 한국 프로야구의 8개구단 (순서는 1994년도 순위에 따른 것임)

1. LG 트윈스 (LG TWINS)
2. 태평양 돌핀스 (TAEPYUNGYANG DOLPHINS)
3. 한화 이글스 (HANHWA EAGLES)
4. 해태 타이거즈 (HAI TAI TIGERS)
5. 삼성 라이온즈 (SAMSUNG LIONS)
6. 롯데 자이언츠 (LOTTE GIANTS)
7. OB 베어스 (OB BEARS)
8. 쌍방울 레이더스 (SSANGBANGWOOL RAIDERS)

BATTING1 : 자기팀의 총 루타수 (1루타+2루타\*2+3루타\*3+홈런\*4 : 경기당)

BATTING2 : 상대팀의 총 루타수

STEAL1 : 자기팀의 도루수 (경기당)

STEAL2 : 상대팀의 도루수

BALLS1 : 자기팀이 얻은 사사구수 (경기당)

BALLS2 : 상대팀에게 내준 사사구수

ERROR1 : 자기팀이 범한 실책수 (경기당)

ERROR2 : 상대팀이 범한 실책수

CONCENT1 : 자기팀의 집중력 (한 회에 얻은 점수가 1점일 경우 : 0

2-3점일 경우 : 1

4-5점일 경우 : 2

6-7점일 경우 : 3

8점 이상일 경우 : 4)

CONCENT2 : 상대팀의 집중력

HW : 자기팀의 경기가 HOME 경기이면 1, AWAY 경기이면 0

위에서 정의된 변수들을 설명하면, BATTING은 팀의 타격실력, STEAL은 팀의 도루실력, BALLS는 팀의 선구능력, ERROR1은 상대팀의 행운, ERROR2는 자기팀의 행운, CONCENT는 팀의 집중력 즉, 찬스에서 연속 득점능력을 뜻하며, HW는 홈 경기에서의 이점을 나타낸다. 이는 확증적 요인분석(Confirmatory Factor Analysis)에 의해서도 뒷받침됨을 밝혀둔다.

## 2.2 구단별 로짓모형

로짓모형을 분석함에 있어서 본 논문에서는 각 구단별로 유의성이 있는 변수들에 대해서만 언급을 하였다. 선택되지 않은 변수는 승·패에 관계없는 변수가 아니라 다른 변수들이 승·패여부에 대한 설명을 하고 난 후에 추가적인 설명을 하지 못하는 변수이다. 선택된 변수(유의수준 5%이내의 변수)들은 표준화된 회귀계수의 추정치(Standardized Coefficient Estimate 이하 SCE로 표기)로 분석을 했다. 또한 SCE의 부호가 양인 것은 자기팀의 승리에 영향을 준 변수이며 부호

가 음인 것은 상대팀의 승리에 영향을 준 변수들임을 말한다. 참고로 <표 1>은 구단별 성적에 관한 자료를 나타내고 있다.

<표 1> 1994년도 구단별 성적표

구단명(순위)	승률	방어율	타율	총득점	총실점
L G (1)	0.643	3.14	0.282	655	452
태평양 (2)	0.552	3.50	0.244	485	471
한화 (3)	0.524	3.52	0.247	482	510
해태 (4)	0.524	3.60	0.271	563	509
삼성 (5)	0.484	3.47	0.260	541	499
롯데 (6)	0.456	4.44	0.257	489	604
O B (7)	0.425	3.46	0.246	477	507
쌍방울 (8)	0.393	4.74	0.248	494	634

## 2.2.1 LG 트윈스

<표 2> LG 트윈스에 대한 로짓모형 적합결과

Analysis of Maximum Likelihood Estimates							
Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-1.8612	1.5946	1.3623	0.2431	.	0.155
BATTING1	1	0.4709	0.1801	6.8392	0.0089	1.593008	1.601
BALLS1	1	0.5832	0.2593	5.0586	0.0245	0.794233	1.792
CONCENT1	1	3.0328	0.9380	10.4551	0.0012	2.638634	20.756
BATTING2	1	-0.4837	0.1489	10.5521	0.0012	-1.297281	0.616
CONCENT2	1	-3.0277	0.9186	10.8630	0.0010	-2.129797	0.048

Association of Predicted Probabilities and Observed Responses			
Concordant	= 98.5%	Somers' D	= 0.970
Discordant	= 1.5%	Gamma	= 0.970
Tied	= 0.0%	Tau-a	= 0.449
(3645 pairs)		c	= 0.985

LG 트윈스는 <표 2>에서 보듯이 타력(BATTING1, SCE=1.593008)이 상대팀의 타력(BATTING2, SCE=-1.297281)에 비해 앞섰으며 BALLS1(SCE=0.794233)이 선택된 것으로 보아 선수들의 선구안이 좋은 것으로 보인다. 또한 집중력(CONCENT1, SCE=2.638634)이 상대팀의 집중력(CONCENT2, SCE=-2.129797)보다 높은 것도 눈에 띈다. LG 트윈스는 주자가 있는 기회가 있으면 적시타를 쳐서 대량득점으로 기회를 살렸으며 BATTING2와 CONCENT2로 보아 상대팀들에게 어느 정도 많은 안타와 집중적인 득점을 허용하였음을 알 수 있는데, 이는 LG의 경기가 재미있었음을 말해준다고 볼 수 있다. 여러 가지 변수들 중에서 타력과 집중력이 승·패에 큰 영향을 준 것을 LG 트윈스의 분석을 통해서 알 수 있다.

## 2.2.3 태평양 돌핀스

<표 3> 태평양 돌핀스에 대한 로짓모형 적합결과

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Odds Ratio
INTERCPT	1	-3.7071	1.6131	5.2812	0.0216	
BATTING1	1	0.3455	0.1246	7.6903	0.0056	0.997220
BALLS1	1	0.5528	0.2692	4.2162	0.0400	0.663158
CONCENT1	1	2.5410	0.6971	13.2875	0.0003	1.701319
STEAL2	1	-1.3374	0.5215	6.5753	0.0103	-0.749561
CONCENT2	1	-3.2632	0.7505	18.9075	0.0001	-2.289747

Association of Predicted Probabilities and Observed Responses			
Concordant	= 98.1%	Somers' D	= 0.963
Discordant	= 1.8%	Gamma	= 0.964
Tied	= 0.0%	Tau-a	= 0.480
(3618 pairs)		c	= 0.982

태평양 돌핀스는 타력(BATTING1, SCE=0.997220)이 승리의 역할을 했음을 쉽게 <표 3>에서 볼 수 있다. 하지만 태평양이 2위를 할 수 있었던 뚜렷한 이유는 <표 3>에서 보듯이 상대팀의 타력(BATTING2)이 변수로써 제외되었는데 그것은 상대팀의 타력을 봉쇄하였다는 것을 알 수 있다. 이것은 간접적으로 태평양의 투수들이 뛰어난다는 것을 알 수 있다. 실제로 1994년도 태평양은 '투수들의 왕국'이라고 불릴 만큼 뛰어난 투수들이 다수 있었다. 또한 태평양 선수들 역시 뛰어난 선구안(BALLS1, SCE=0.663158)으로 많은 사사구를 얻어내어 공격에 일익을 담당했음을 알 수 있다. 하지만 태평양은 도루허용 변수(STEAL2, SCE= -0.749561)가 선택된 것으로 보아 포수가 2루로 송구하는 능력이 다소 부족하다는 것을 알 수 있다. 그러므로 앞으로 포수의 송구능력을 향상시키는데 주력해야 할 것이다.

## 2.2.4 한화 이글스

<표4> 한화 이글스에 대한 로짓모형 적합결과

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Odds Ratio
INTERCPT	1	3.6048	1.1531	9.7731	0.0018	
CONCENT1	1	3.7646	0.9001	17.4938	0.0001	2.556882
BATTING2	1	-0.3898	0.1334	8.5373	0.0035	-1.231221
CONCENT2	1	-2.7364	0.8077	11.4787	0.0007	-1.952232

Association of Predicted Probabilities and Observed Responses			
Concordant	= 97.4%	Somers' D	= 0.951
Discordant	= 2.3%	Gamma	= 0.953
Tied	= 0.2%	Tau-a	= 0.478
(3835 pairs)		c	= 0.975

한화 이글스는 <표 1>에서 보면 해태 타이거즈의 타율과 득점력면에서 뒤지지만 승률이 같음을 알 수 있다. 그 이유는 뚜렷한 타력을 보이지 않으나 상대팀에 비해서 집중력(CONCENT1, SCE=2.556882)이 훨씬 높으며 또한 상대팀에게 안타(BATTING2, SCE=-1.231221)를 허용한 것에 비하여 상대팀의 집중력(CONCENT2, SCE=-1.952232)이 자기팀의 집중력(CONCENT1)보다 낮아 대량실점을 하지 않았기 때문이다. 이점은 한화가 효율적인 야구를 했음을 시사한다. 따라서 한화는 8개구단 중 집중력(CONCENT1)면에서 가장 뛰어난 구단으로 평가된다. 하지만 한화는 공격적 측면(BATTING1, STEAL1, BALLS1)이 부족하므로 이를 향상시키는데 주력해야 할 것이다.

## 2.2.5 해태 타이거즈

<표 5> 해태 타이거즈에 대한 로짓모형 적합결과

Analysis of Maximum Likelihood Estimates							
Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	0.1600	0.7821	0.0419	0.8379	.	1.174
STEAL1	1	0.9757	0.4322	5.0962	0.0240	0.618672	2.653
CONCENT1	1	2.8554	0.7236	15.5714	0.0001	2.103991	17.381
CONCENT2	1	-2.8774	0.6274	21.0366	0.0001	-2.145785	0.056
HW	1	1.9327	0.7699	6.3017	0.0121	0.534913	6.908

Association of Predicted Probabilities and Observed Responses	
Concordant = 96.4%	Somers' D = 0.935
Discordant = 2.9%	Gamma = 0.941
Tied = 0.6%	Tau-a = 0.469
(3705 pairs)	c = 0.967

해태 타이거즈는 도루변수(STEAL1, SCE=0.618672)로 보아 기동성 있는 공격을 펼쳐 왔음을 볼 수 있다. <표 1>에서 보듯이 타율이 높음에도 불구하고 타력(BATTING1)의 변수가 선택되지 않은 것은 팀승리에 개인선수들의 능력이 작용을 하지 못했기 때문으로 분석된다. 따라서 좀 더 효과적인 작전과 기회가 있으면 무산시키지 않는 선수들의 승부근성을 향상시켜야 하겠다. 또한 상대팀의 타력(BATTING2)이 선택되지 않은 것으로 보아 해태의 투수력이 우수했음을 알 수 있다. 이 구단의 특이한 점으로는 8개 구단 중 유일하게 홈 경기(HW, SCE=0.534913)에 강한 모습을 보여 주었다. 이것은 열광적인 관중들 또한 해태의 승리에 일익을 담당했음을 알 수 있다. 그러나 한화보다 우수한 능력을 보유하고도 집중력의 부족으로 인해 승률에 있어서 한화와 동률을 이루게 된 것 또한 특이한 점이라 볼 수 있다.

## 2.2.6 삼성 라이온즈

<표 6> 삼성 라이온즈에 대한 로짓모형 적합결과

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Odds Ratio
INTERCPT	1	-4.5097	1.3228	11.6224	0.0007	0.011
BATTING1	1	0.4440	0.1403	10.0164	0.0016	1.515439
CONCENT1	1	3.6211	0.8804	16.9183	0.0001	3.089402
CONCENT2	1	-4.1433	1.0713	14.9586	0.0001	-2.816127

Association of Predicted Probabilities and Observed Responses	
Concordant = 98.1%	Somers' D = 0.963
Discordant = 1.8%	Gamma = 0.964
Tied = 0.1%	Tau-a = 0.485
(3776 pairs)	c = 0.981

삼성 라이온즈는 승리의 요인으로 타격(BATTING1, SCE=1.515439)이 매우 중요한 역할을 했음을 알 수 있다. 또한 이 구단은 뚜렷한 투수진이 없는 대신 득점할 수 있는 좋은 기회가 오면 연속적인 점수를 얻는 팀의 집중력(CONCENT1, SCE=3.089402)이 매우 높으므로 찬스에 강한 구단으로 분석된다. 투수들의 능력이 상대팀의 타력(BATTING2)을 제압한 것이 승리로 작용하였다. 객관적인 자료로는 훌륭한 구단으로 분석된다. 그러나 집중력(CONCENT1)이 높음에도 불구하고 성적이 좋지 않았던 것은 기회가 있을 때에는 몰아쳐서 대량득점을 하지만 그러한 기회를 많이 만들지 못했다는 것을 말한다. 또한 상대팀에게 높은 집중력(CONCENT2, SCE=-2.816127)을 허용함으로써 대량실점으로 진 경기가 많았음을 알 수 있다.

## 2.2.7 롯데 자이언츠

<표 7> 롯데 자이언츠에 대한 로짓모형 적합결과

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Odds Ratio
INTERCPT	1	-0.4420	1.4813	0.0890	0.7654	0.643
BATTING1	1	0.3906	0.1156	11.4135	0.0007	0.905170
STEAL1	1	1.0393	0.3772	7.5923	0.0059	0.698605
CONCENT1	1	2.2269	0.7601	8.5844	0.0034	1.233700
BATTING2	1	-0.3525	0.1353	6.7916	0.0092	-1.130663
CONCENT2	1	-2.7721	0.7299	14.4253	0.0001	-2.186349

Association of Predicted Probabilities and Observed Responses	
Concordant = 97.8%	Somers' D = 0.957
Discordant = 2.1%	Gamma = 0.957
Tied = 0.0%	Tau-a = 0.479
(3752 pairs)	c = 0.979

롯데 자이언츠는 팀의 타력(BATTING1, SCE=0.905170)과 주자의 도루 (STEAL1,

SCE=0.698605)가 승리의 요인으로 작용하였다. 그러나 자기팀의 집중력(CONCENT1, SCE=1.2337)이 상대팀의 집중력(CONCENT2, SCE=-2.186349) 보다 매우 낮음을 알 수 있다. 이것은 타력(BATTING1)과 도루(STEAL1)가 연속적인 점수를 내는데 많은 영향을 주지 못하며, 또한 한번 실점을 내주면 대량실점으로 이어지는 위기대처능력이 부족함을 알 수 있다. 타력과 도루능력이 빛을 발휘하도록 감독의 재치 있는 작전능력이 더욱 필요하다. 그리고 안타(BATTING2, SCE=-1.130663)를 많이 내준 것이 패배의 요인으로 분석된다. 이것은 이 팀의 투수진이 매우 약함을 뜻한다.

## 2.2.8 OB 베이스

<표 8> OB 베이스에 대한 로짓모형 적합결과

Analysis of Maximum Likelihood Estimates							
Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	2.8935	1.3298	4.7345	0.0296		18.056
ERROR2	1	1.2342	0.5432	5.1622	0.0231	0.627908	3.436
CONCENT1	1	3.7174	0.8918	17.3746	0.0001	2.753181	41.157
BATTING2	1	-0.3055	0.1141	7.1720	0.0074	-0.791981	0.737
BALLS2	1	-0.4083	0.2083	3.8419	0.0500	-0.484402	0.665
CONCENT2	1	-3.6395	1.0466	12.0930	0.0005	-2.484321	0.026

Association of Predicted Probabilities and Observed Responses	
Concordant = 97.3%	Somers' D = 0.947
Discordant = 2.6%	Gamma = 0.948
Tied = 0.1%	Tau-a = 0.465
(3744 pairs)	c = 0.974

OB 베이스는 <표 8>에서 보면 상대방의 실책(ERROR2, SCE=0.627908)으로 인한 행운이 승리의 요인으로 작용한 경기가 많았다는 것이 특이할 만하다. 또한 승리의 요인으로 작용한 변수가 하나도 선택되지 않은 것으로 보아 전반적으로 공격력이 부진했음을 알 수 있다. 이것은 <표 1>에서 총득점이 8개 구단 중 가장 낮은 순위를 기록한 것에서도 볼 수 있다. 또 변수중 BATTING2 (SCE=-0.791981)와 BALL2 (SCE=-0.484402)가 선택되었다는 것은 계구력 등 전반적인 투수력이 약하다는 것을 알 수 있다.

## 2.2.9 쌍방울 레이더스

<표 9> 쌍방울 레이더스에 대한 로짓모형 적합결과

Analysis of Maximum Likelihood Estimates							
Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-3.2743	1.7470	3.5128	0.0609	.	0.038
BATTING1	1	0.5903	0.2110	7.8248	0.0052	1.433808	1.805
BALLS1	1	0.9020	0.3665	6.0553	0.0139	1.100864	2.464
CONCENT1	1	2.9281	0.9080	10.3988	0.0013	2.016537	18.692
STEAL2	1	-1.4551	0.5938	6.0044	0.0143	-0.852395	0.233
BALLS2	1	-0.4694	0.2206	4.5262	0.0334	-0.631818	0.625
CONCENT2	1	-6.1596	1.8723	10.8229	0.0010	-5.073025	0.002

Association of Predicted Probabilities and Observed Responses			
Concordant	=	94.3%	Somers' D = 0.927
Discordant	=	1.6%	Gamma = 0.966
Tied	=	4.1%	Tau-a = 0.444
(3478 pairs)			c = 0.963

쌍방울 레이더스는 <표 1>에서 보듯이 타력에 있어서 우수한 팀이다. 따라서 타격(BATTING1 SCE=1.433808)과 선수들의 선구능력(BALL1 SCE=1.100864)이 승리의 요인으로 작용했다. 그러나 패배의 요인이 STEAL2(SCE=-0.852395), BALL2(SCE=-0.631818)인 것으로 보아 상대팀의 진루가 타격이외에 사사구, 도루 등으로 이루어져 포수의 송구능력, 투수의 제구력이 떨어짐을 알 수 있다. 특히 CONCENT2(SCE=-5.073025)에서 보면 한번 실점을 내어주면 연속적으로 상대방의 집중포화를 받음으로써 구단의 수비능력이 매우 떨어짐을 보여주고 있다. 이것이 팀의 순위가 최하위에 머무르게 된 결정적인 요인으로 작용한 것임을 알 수 있다. 따라서 쌍방울은 타력에 있어서는 우수한 구단이지만 수비력에서 뒤떨어져 있다. 투수력, 포수의 송구능력, 완벽한 수비를 위해 노력한다면 우수한 구단이 될 수 있을 것이다.

## 2.3 구단별 로짓모형의 선택된 변수의 표준화된 회귀계수표

구단별 로짓모형을 전체적으로 비교하기 위하여 <표 10>을 작성할 수 있다.

<표 10> 구단별 로짓모형의 선택된 변수의 표준화된 회귀계수표

변수 구단명(순위)	BAT1	STEA1	BALL1	ERR2	CON1	BAT2	STEA2	BALL2	ERR1	CON2	HW
L G (1)	1.593		0.974		2.639	-1.297				-2.130	
태평양 (2)	0.997		0.663		1.701		-0.750			-2.290	
한 화 (3)					2.557	-1.231				-1.952	
해 태 (4)		0.618			2.104					-2.146	0.535
삼 성 (5)	1.515				3.089					-2.816	
롯데 (6)	0.905	0.699			1.234	-1.131				-2.186	
O B (7)				0.628	2.753	-0.792		-0.484		-2.484	
쌍방울 (8)	1.434		1.101		2.017		-0.852	-0.632		-5.073	

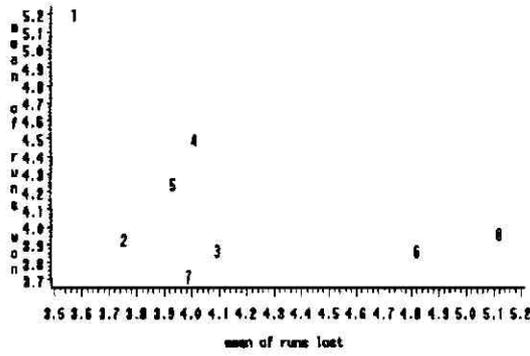
<표 10>을 보면 집중력 변수 CONCENT1, 2가 모두 선택되어 팀의 승패에 많은 영향을 준 것을 알 수 있으며 BATTING1은 타력, BATTING2는 수비력을 나타낸다. 해태 타이거즈와 롯데 자이언츠에서의 STEAL1(각각 SCE=0.618, SCE=0.699)이 선택된 것은 두 구단에 발빠른 선수들이 풍부하다는 것이며 해태 타이거즈의 HW(SCE=0.535)가 선택된 것은 홈 구장인 광주구장에서의 승률이 높다는 것이다. OB 베어스는 상대팀 범실 ERROR2(SCE=0.628)가 행운으로 작용한 것이 특이한 점이며 태평양 돌핀스와 쌍방울 레이더스는 STEAL2(각각 SCE=-0.750, SCE=-0.852)에서 포수의 송구능력이 타 구단에 비해 상대적으로 약함을 알 수 있다. 또한 사사구 허용에 있어서도 OB 베어스와 쌍방울 레이더스의 BALL2(각각 SCE=-0.484, SCE=-0.632)가 선택되어 투수진의 제구력이 뛰어나지 못했다는 것을 보여준다.

<표 11> 각 구단별 집중력

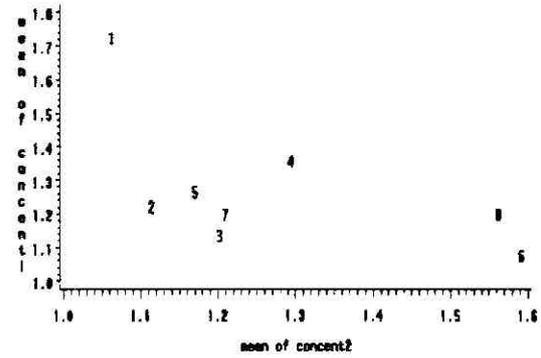
구 단 명	SCE (CON1) - SCE (CON2)
한화 이글스	0.605
L G 트윈스	0.509
삼성 라이온즈	0.273
O B 베어즈	0.269
해태 타이거즈	-0.042
태평양 돌핀스	-0.589
롯데 자이언츠	-0.952
쌍방울 레이더스	-3.056

<표 11>에서 보면 선수들이 경기에 임하는 집중력과 감독의 작전능력을 간접적으로 가늠해 볼 수 있는 척도로 사용되어 질 수 있다. 예를 들면 한화 이글스는 집중력이 타 구단에 비해 뛰어나고 쌍방울 레이더스는 집중력이 타 구단에 비해 떨어지며 해태 타이거즈는 중간정도로 볼 수 있다.

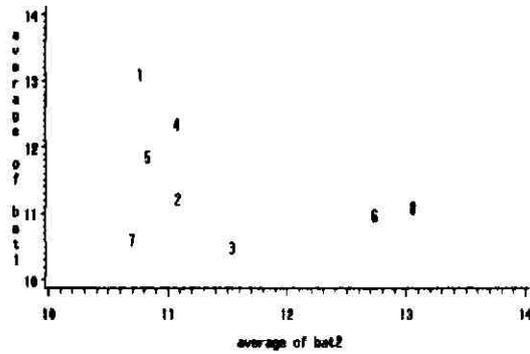
## 2.4 부문 성적을 통한 구단비교



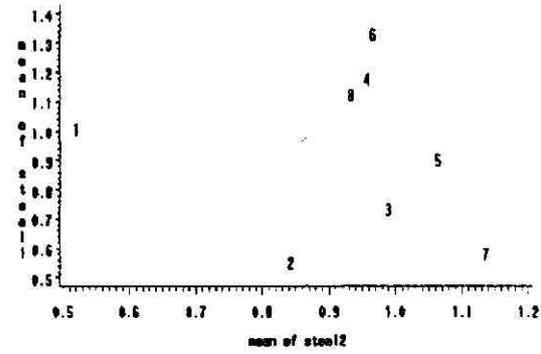
<그림 1> 경기당 평균 득·실점



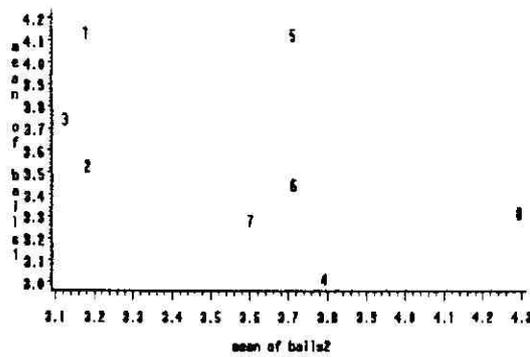
<그림 2> 경기당 집중력(득·실)



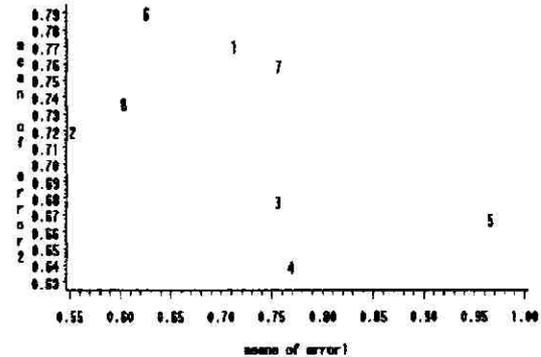
<그림 3> 경기당 루타수(득·실)



<그림 4> 경기당 도루수(득·실)



<그림 5> 경기당 사사구수(득·실)



<그림 6> 경기당 상대팀 실책수(득·실)

위 그림을 보면 구단의 부문간 득·실을 비교한 것이다. 세로축이 자기팀의 성적이고 가로축은 상대팀의 성적이다. 따라서 세로축은 높고 가로축은 낮은 팀이 그 부분에서 뛰어난 것이 되며, 실책 부문은 자기팀의 실책이 상대팀에 이득이 되므로 세로축과 가로축을 바꾸어 놓았다. 상위 팀들이 주로 좌측 상단 가까이 분포되어 있음을 알 수 있다. 또한 실책을 나타낸 <그림 6>의 경우에는 순위와는 관계없이 분포가 나타난 것을 보아 야구경기는 상대팀의 실책보다는 역시 자기팀의 순수한 실력이 승·패에 작용한다는 것이 나타나있다.

### 3. 결 론

본 논문에서는 1994년 한국 프로야구의 8개팀 17개 부문의 기록에 대한 통계적 자료분석을 위하여 로지스틱 회귀분석을 활용하였다.

프로야구 자료의 전체적 분석보다는 각 구단별 특성을 도출해 내기 위하여 구단별 승·패에 대한 로짓모형을 설정하였으며 그 결과 유의하다고 생각되는 결과를 얻었다. 여러 요소가 팀의 승·패에 중요한 역할을 하지만 각 팀마다 강한 부문이 있는가하면 약한 부문이 있었고 상위에 위치한 팀일수록 일반적으로 공격, 수비 등 전반적인 부문에서 뛰어나지만 각 구단별로의 팀 컬러가 존재하는 것을 확인할 수 있었다. 따라서 야구경기는 타율, 방어율의 순위가 꼭 팀의 순위와 일치 하지는 않은 것이다. 그러나 그 문제는 집중력의 정도가 타당한 설명을 가능케 하여 주었다. 누가 안타를 많이 치고 적게 치느냐의 문제보다는 누가 효율적으로 루상에 계속해서 주자를 내보내서 찬스마다 점수를 몰아 내는가가 팀의 승·패에 직접적인 영향을 주는 요인인 것이다. 실제로 상위 팀일수록 상대팀과의 집중력 차이가 더 높은 양의 값을 가진다. 예로 한화와 해태를 들 수 있는데 객관적인 성적으로는 해태가 훨씬 앞서있지만 한화와 승률이 같으며, 이 결과는 1994년 한 해 해태 타이거즈 감독의 작전능력과 선수들의 승부근성이 부족했던 것을 나타낸다.

이번 분석을 통해서 구단별 승리의 요인으로는 여러 가지 객관적인 자료도 중요하지만 본 논문에서 정의한 집중력과 같이 눈에 보이지 않는 감독의 작전능력, 선수들의 승부근성 등이 팀의 승·패에 적지 않은 영향을 주는 것을 알 수 있었다. 1994년도 구단별 분석 내용을 각 구단마다 참조한다면 미약하나마 단점보완에 도움이 되리라 생각한다.

### 4. 감사의 글

본 논문을 지도하여 주신 유성모 교수님과 조언을 아끼지 않으신 임성수 교수님께 깊은 감사를 드립니다.

### 참고 문헌

1. KBO (1994) 「 한국 프로야구 연감 」
2. 허명희, 이태림, 임성수 「 통계적 자료분석 I · SAS 」 (1994) -방송통신대학
3. 김충련 「 SAS라는 통계상자 」 (1994) -데이타리서치

Characterization of Korean Professional Baseball Teams with 1994 Records  
Using Logit Model

Jin-woo Chung, Dae-whan Kim, Doo-seong Seo<sup>2)</sup>

Abstract

Logistic analysis are used to find factors playing an important role to game outcomes of 1994 Korean baseball season for each team. Significant factors for game outcomes are selected for each team through variable selection procedures. Interpretations for the characteristics of each team are also provided.

---

2) Department of Applied Statistics, Korea Univ., 208 Seochang-dong, Chochiwon, Yeongi-gun, Chungnam 339-700, KOREA

## 남녀간의 외모평가 차이검정

이 의 구

(지도교수 : 권 세 혁)

한남대학교 경상대학 응용통계학과

### 요 약

본 논문은 사람의 외모를 평가할 때, 남자와 여자사이에 시각차이가 존재하는가의 여부를 검정하고자 하는 목적에 따라, 조사자들을 통하여 자료를 수집하고, 이에 대한 신뢰성을 검토한 후 실제 검정기법으로 모수적 방법과 비모수적 방법을 병행하여 그 결과를 통해 '차이'의 존재를 확인하고, 그 원인에 대해서 고찰해 보았다.

### 1. 서 론

흔히들 사람의 외모를 볼 때 '남자가 보는 눈과 여자가 보는 눈은 틀리다'라는 말을 하곤 한다. (물론 여기서 '눈'이라 함은 사물을 인식, 판단 혹은 평가할 수 있는 힘을 뜻하는 것이다.) 그리고 거의 대부분의 사람들이, 이 말을 그냥 당연한 것으로 인정해 버리고 만다. 원래 그렇다는 듯이... 누구나 '그냥 그럴 것 같다'는 것뿐이지 증명된 것은 아니다.

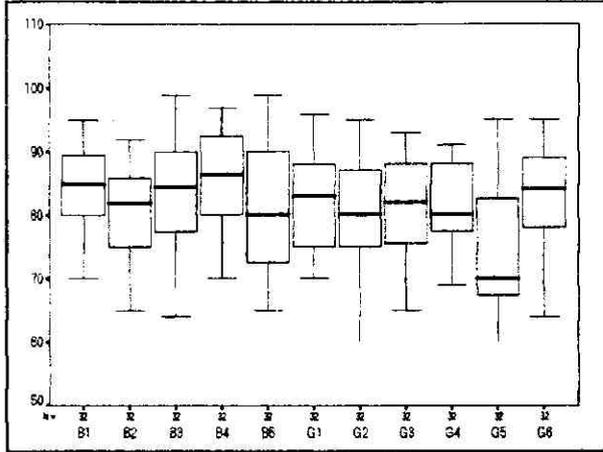
'과연 정말로 그런가?' 간단한 통계적 기법을 이용하여, 이 문제를 검정해 보도록 하겠다. 이와 함께 정말 차이가 있다면, 그 차이는 어떤 것인가? 즉, 남자(여자)가 동성(同性)과 이성(異性)의 외모를 평가함에 있어서 어느 쪽에서 더 까다롭다고 할 수 있는가에 대해서도 알아보도록 하겠다. 자료는 청주대학교 예술대학앞을 지나는 남,녀 36명(남자=18, 여자=18명)을 대상으로, 11명의 조사자들(남자=5, 여자=6)이 대상자들의 외모를 주관적으로 측정한 점수중에서, 결측치를 제거한 32명(남자=17, 여자=15)에 대한 점수만을 자료로 사용하였다. 조사자료의 특성상, 자료가 가질 수 있는 측정오류를 줄이기 위하여, 남,녀 조사자들에 대한 신뢰도를 조사하고 그 결과로 제거 여부를 판단한 후, 검정자료로 사용한다.

두 집단간의 차이를 검정(두 집단의 평균치 차이가 표본오차에 의한 것인지, 두 집단의 속성에 의한 것인지)하는 것이 목적이므로, T-test(paired)를 사용한다. 두 집단의 차이는 각각의 집단을 대표할 수 있는 『대표값』의 차이로 해석하여, 이 대표값이 서로 차이가 없으면 두 집단이 차이가 없는 것으로 결론을 내린다. 아울러, 정규성을 따르지 않으면서 표본의 수가 적은 경우에 사용되는 비모수적 기법(Wilcoxon Signed Ranks Test)을 병행하여, 각각의 결과를 비교하도록 한다.(두 검정방법 모두 대표값으로는 중앙값을 사용한다.)

### 2. 신뢰도 분석

신뢰도 분석의 목적은, 자료를 얻는 조사과정에서 생길 수 있는 측정오류를 줄이기 위함이다. 일반적으로 알파계수가 0.6 혹은 0.7 이상 이면, 신뢰도가 높다고 할 수 있으므로, 만일 이보다 신뢰도가 떨어질 경우, 문제가 되는 항목(변수로 사용된 조사자)을 제거함으로써, 이를 해결하도록 하겠다.

2.1 전체자료(남,녀 = 32명)를 대상으로



<결과 2.1> 기초통계량-전체

변 수 (조사자)	기초통계량		
	Mean	Std Dev	
남자	B1	83.0313	7.3154
	B2	80.8125	7.6134
	B3	82.8125	10.0111
	B4	85.6250	8.0713
	B5	81.1250	10.0635
여자	G1	81.7813	6.9872
	G2	80.4375	8.6171
	G3	81.2188	8.0191
	G4	81.5625	6.2110
	G5	74.1250	10.0635
	G6	82.1563	7.8210

<결과 2.1>은 신뢰도분석 이전에, 분석의 대상이 되는 남자조사자와 여자조사자에 대한 정보를 간단히 살펴본 것이다. 기초 통계량의 결과에 의하면 남자조사자의 경우 B4의 평균이 가장 크다는 것과 여자의 경우에는 G5의 평균이 가장 작고 편차는 가장 큰 것을 알 수 있으며, 전체적으로 남자조사자들 점수의 편차가 여자조사자들의 그것보다 더 크다는 것을 알 수 있다. 그러나, 이런 정보가 신뢰도 분석에 큰 영향을 미친다고 볼 수는 없는 것이 신뢰도분석은 변수간의 상관관계를 통해서 얻어지는 것이므로, 위의 결과만을 가지고 설블리 속단하는 것은 위험한 일이다. 즉, 위의 결과는 단지 각 변수들을 살펴보는 데 사용되는 참고자료일 뿐이다.

2.1-1 남자조사자들의 신뢰도

<결과 2.1-1>은 다섯명의 남자조사자들의 신뢰도에 관한 결과이다. 이 결과에 의하면, 크론바하 알파모델에 의하여 남자조사자 5명(B1-B5)을 하나의 가상변수(스케일)로 신뢰도 검사를 실시한 결과, 그 값이 0.7507로 신뢰성이 있다고 할 수 있다는 결론을 얻어낼 수 있다.

다음은 <결과 1.1-1>의 [ALPHA IF ITEM DELETED]의 결과를 이용하여, 남자조사자들의 신뢰도를 높인 경우이다. 이 결과는 검정과정에서, '신뢰도의 정도차이에 의한 검정결과의 차이'를 살펴보기 위하여 참고적으로 얻은 것이다.

<결과 2.1-1 남자조사자 5명의 신뢰도-전체>

Statistics for	Mean	Variance	Std Dev	N of Variables
SCALE	413.4063	946.6361	30.7675	5
Item-total Statistics				
Scale Mean if Item Deleted		Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
B1	330.3750	761.1452	.3270	.7647
B2	332.5938	657.7329	.5914	.6843
B3	330.5938	563.6038	.5950	.6759
B4	327.7813	625.2732	.6348	.6659
B5	332.2813	610.2732	.4728	.7285
Reliability Coefficients				
N of Cases =	32.0			N of Items = 5
Alpha =	.7507			

▶ 조사자 B1과 B5을 제거한 경우

Reliability Coefficients		
N of Cases =	32.0	N of Items = 3
Alpha =	.7833	

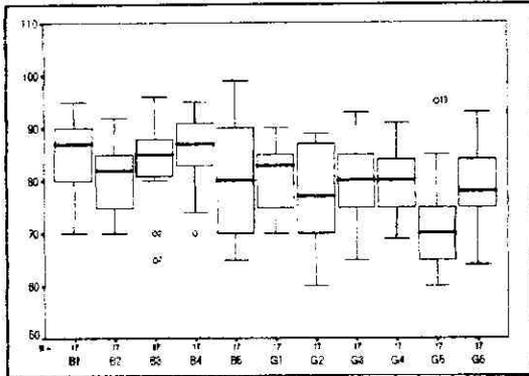
2.1-2 여자조사자들의 신뢰도

<결과 2.1-2>에 의하면, 크론바하 알파값이 0.8791로 신뢰성에는 문제가 없는 것으로 보여진다. 게다가 이 경우에는 조사자 6명(G1-G6)을 하나의 스케일로 신뢰도를 검사했을 때, 얻어지는 알파값이 가장 크므로, 신뢰도를 높이기 위한 변수의 제거는 필요치 않다.

<결과 2.1-2 여자조사자 6명의 신뢰도-전체>

Statistics for Scale	Mean	Variance	Std Dev	N of Variables	
	481.2813	1452.2087	38.1079	6	
Item-total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
G1	399.5000	1066.7742	.7375	.6058	.8521
G2	400.8438	995.8135	.7027	.6411	.8557
G3	400.0625	1003.7379	.7561	.6996	.8464
G4	399.7188	1097.0474	.7695	.6181	.8514
G5	407.1563	963.4264	.6203	.4973	.8775
G6	399.1250	1070.4355	.6265	.5139	.8679
Reliability Coefficients		N of Items = 6			
N of Cases = 32.0					
Alpha = .8791					

2.2 남자모델 (N=17) 만을 대상으로



<결과 2.2> 기초통계량-남자모델

변수 (조사자)	기초통계량	
	Mean	Std Dev
남자 B1	84.3529	8.1390
B2	81.0588	7.0929
B3	84.0588	8.2346
B4	85.7647	7.8941
B5	81.0000	10.5948
여자 G1	81.0588	5.6731
G2	77.4706	8.8326
G3	79.4706	8.4418
G4	79.8235	6.4443
G5	72.6471	9.7014
G6	78.5882	8.3745

2.2-1 남자조사자들의 신뢰도

<결과2.2-1>은 남자모델들에 대한, 남자 조사자들의 점수의 신뢰성에 문제가 있다고 할 수 있다. 그러므로, 이 경우에는 불가피하게 크론바하 알파값을 높이기 위한 변수의 제거가 필요할 것으로 보인다. 0.7이상의 알파값을 얻기 위해 남자관찰자 B1과 B5를 제거하였다.

<결과2.2-1 남자조사자들의 신뢰도-남자모델>

Statistics for	Mean	Variance	Std Dev	N of Variables
SCALE	416.2353	561.1912	23.6895	5
Item-total Statistics	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
B1	331.8824	493.2353	.0047	.5421
B2	335.1765	389.5294	.4334	.2770
B3	332.1765	369.6544	.3907	.2833
B4	330.4706	386.1397	.3634	.3091
B5	335.2353	403.9412	.1057	.5191
Reliability Coefficients			N of Items = 5	
N of Cases = 17.0			Alpha = .4505	

남자 조사자 B1,B5 제거

남자조사자 5명의 자료중 2명의 자료를 제거함으로써, 어느정도 만족할 만한 알파값을 얻을 수 있었다. 따라서, 남자모델에 대한 남,녀 조사자들의 점수 차이검정에 있어 남자조사자의 자료는 3명(B2-B4)만을 사용하도록 하겠다.

Reliability Coefficients	N of Items = 3
N of Cases = 17.0	Alpha = .7351

2.2-2 여자 조사자들의 신뢰도

남자모델들만을 대상으로 한 경우에도, 여자조사자들의 신뢰성에는 큰 문제가 없는 것으로 보인다. 크론바하 알파값은 0.8422, 충분히 신뢰성을 인정할 수 있지만, G5를 제거할 경우 더 높은 신뢰도를 얻을 수 있으므로, 일단 가능한 높은 신뢰도를 얻을 수 있게끔 변수를 제거해 보도록 하겠다.

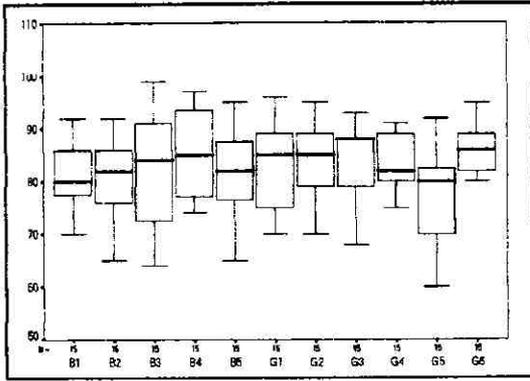
<결과 2.2-2 여자 조사자들의 신뢰도-남자모델>

Statistics for	Mean	Variance	Std Dev	N of Variables
SCALE	469.0588	1298.6838	36.0373	6
Item-total Statistics	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
G1	388.0000	1004.0000	.7302	.8079
G2	391.5882	844.1324	.7336	.7921
G3	389.5882	873.5074	.7092	.7978
G4	389.2353	960.1912	.7436	.7999
G5	396.4118	943.7574	.4375	.8618
G6	390.4706	956.3897	.5254	.8355
Reliability Coefficients			N of Items = 6	
N of Cases = 17.0			Alpha = .8422	

여자조사자 G5,G6 제거

Reliability Coefficients	N of Items = 4
N of Cases = 17.0	Alpha = .8719

### 2.3 여자모델 (N=15)만을 대상으로



<결과 2.3> 기초통계량-여자모델

범 수 (조사자)	기초통계량		
	Mean	Std Dev	
남자	B1	81.5333	6.1860
	B2	80.5333	8.4080
	B3	81.4000	11.8491
	B4	85.4667	8.5429
	B5	81.2667	9.7941
여자	G1	82.6000	8.3649
	G2	83.8000	7.2427
	G3	83.2000	7.2821
	G4	83.5333	5.4885
	G5	75.8000	10.5370
	G6	86.2000	4.7238

#### 2.3-1 남자조사자의 신뢰도

남자조사자들의 신뢰도는 여자모델만을 대상으로 한 경우, 가장 높은 신뢰도를 얻을 수 있었다. 위의 결과에서도 볼 수 있듯이, 5명 모두를 하나의 스케일을 구성한 경우의 크론바하 알파값이 가장 크므로, 더 이상 신뢰도를 높이기 위한 작업도 불필요하다.

<결과 2.3-1 남자 조사자들의 신뢰도-여자모델>

Statistics for	Mean	Variance	Std Dev	N of Variables
SCALE	410.2000	1434.0286	37.8686	5
Item-total Statistics				
Scale		Scale	Corrected	Alpha
Mean		Variance	Item-	if Item
if Item		if Item	Total	Deleted
Deleted		Deleted	Correlation	
B1	328.6667	1115.8095	.6774	.8793
B2	329.6667	993.9524	.6967	.8671
B3	328.8000	819.0286	.6998	.8810
B4	324.7333	924.4952	.8403	.8354
B5	328.9333	867.0667	.8167	.8377
Reliability Coefficients				
N of Cases =	15.0		N of Items =	5
Alpha =	.8854			

#### 2.3-2 여자조사자들의 신뢰도

여자조사자들의 경우에도, 남자조사자들의 결과와 마찬가지로 신뢰성이 있다는 결론을 얻을 수는 있지만, 6명 모두를 스케일로 구성한 경우보다, G2를 제거한 경우에 더 높은 신뢰도를 구할 수 있다.

<결과 2.3-2 여자조사자들의 신뢰도-여자모델>

Statistics for SCALE		Mean	Variance	Std Dev	N of Variables
		495.1333	1344.4095	36.6662	6
Item-total Statistics	Scale	Scale	Corrected	Alpha	
	Mean if Item Deleted	Variance if Item Deleted	Item-Total Correlation	if Item Deleted	
G1	412.5333	872.1238	.8143	.8645	
G2	411.3333	1018.3810	.5918	.8984	
G3	411.9333	940.0667	.7867	.8698	
G4	411.6000	1047.1143	.7522	.8814	
G5	419.3333	755.6667	.8246	.8730	
G6	408.9333	1083.2095	.7683	.8846	
Reliability Coefficients				N of Items = 6	
N of Cases = 15.0					
Alpha = .8975					

여자 조사자 G2 제거

Reliability Coefficients		N of Items = 5
N of Cases = 15.0		
Alpha = .8984		

2.4 신뢰도 분석의 결과 및 요약

지금까지 분석대상을 세집단(전체자료, 남자모델, 여자모델)으로 나눈 후, 각각에 대한 남, 여 조사자들의 신뢰도를 검사해 보았다. 신뢰성이 있다고 판단되는 기준을, 크론바하 알파값이 0.7이상 이 되는 경우로 제한한 후, 이후의 검정에 가능한 형태로 정리해 보면 다음과 같다.

1) 신뢰성이 있다고 판단되어지는 경우(ALPHA값이 0.7 이상)

대상집단	변수	조사자	ALPHA	조사자	ALPHA
	남자(명/5)	남자(명/5)		여자(명/6)	
전체자료 (N=32)	5	5	.7507	6	.8791
남자모델 (N=17)	3(B2-B4)	3(B2-B4)	.7351	6	.8422
여자모델 (N=15)	5	5	.8854	6	.8975

2) 신뢰성이 가장 높은 경우

대상집단	변수	조사자	ALPHA	조사자	ALPHA
	남자(명/5)	남자(명/5)		여자(명/6)	
전체자료 (N=32)	3(B2-B4)	3(B2-B4)	.7833	6	.8791
남자모델 (N=17)	3(B2-B4)	3(B2-B4)	.7351	4(G1-G4)	.8719
여자모델 (N=15)	5	5	.8854	5(G1, G3-G6)	.8984

3. 남,녀 조사자들간의 차이검정

3.1 전체자료를 대상으로 한 경우 (N=32)

<결과 3.1>에 의하면, 우선 t-검정 결과 두 평균의 차이가 2.8750이고, p-값이 0.016으로 유의수

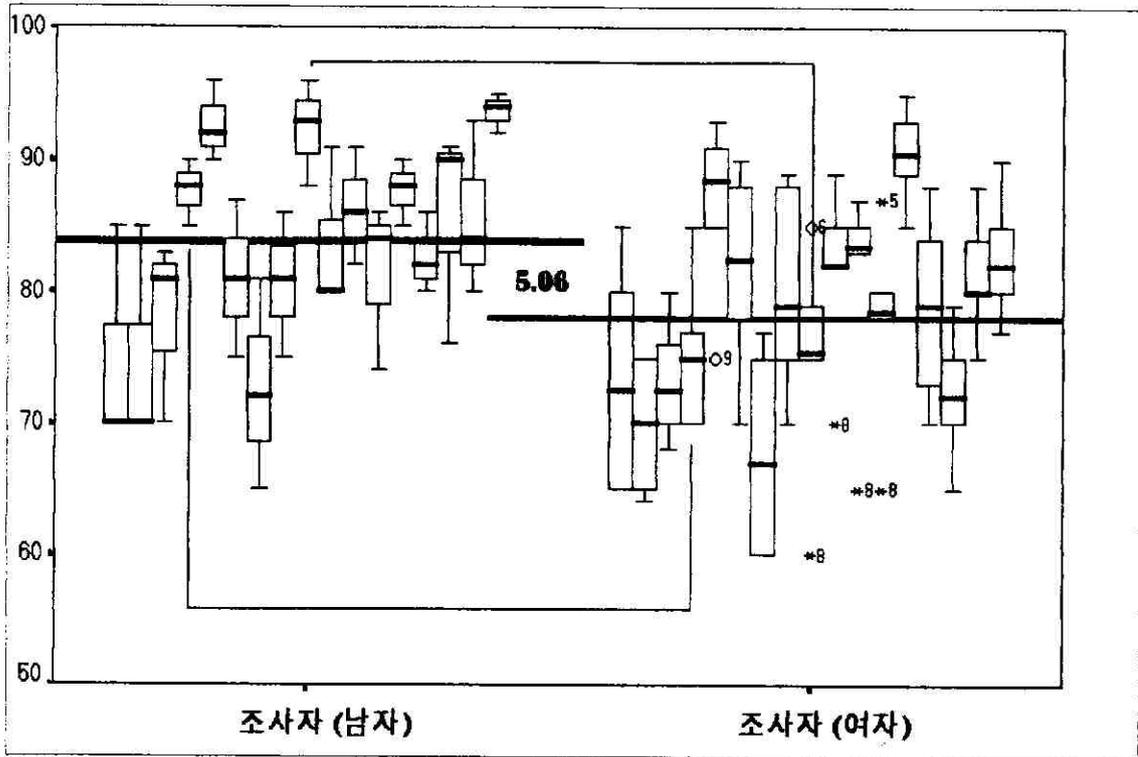
준 0.05에서 귀무가설을 기각할 수 있다. 즉, 평균차이 2.8750은 통계적으로 유의하다고 말할 수 있다. 윌콕슨 검정의 결과를 보아도, 여자조사자의 점수가 남자조사자의 점수보다 작은 경우의 평균순위가 18.29이고, 큰 경우의 평균순위는 10.33으로, 유의수준 0.05에서 귀무가설 '남,녀 조사자들의 점수에는 차이가 없다'를 기각할 수 있다. 이 차이는, '조사된 자료에 의하면, 전체적으로 남자조사자들의 점수가 여자조사자들의 점수보다 높다'는 것을 말한다고 볼 수 있다.

<결과 3.1 남자조사자 5명 (B1-B5) : 여자조사자 6명 (G1-G6)>

* t-tests for paired samples			t-value	df	2-tail Sig
Mean	Paired Differences SD	SE of Mean			
2.8750	6.356	1.124	2.56	31	.016
95% CI (.583, 5.167)					
* Wilcoxon Matched-Pairs Signed-Ranks Test					
Mean Rank	Cases				
18.29	17	- Ranks (G1R6 LT BOY5)			
10.33	12	+ Ranks (G1R6 GT BOY5)			
	3	Ties (G1R6 EQ BOY5)			
	32	Total			
Z =	-2.0218		2-Tailed P =	.0432	

### 3.2 남자모델만을 대상으로 한 경우(N=17)

남자모델만을 대상으로 한 경우, 남녀간의 시각차이는 존재하는 것으로 보여진다. 이 경우 남녀 조사자들의 점수 차이를 아주 뚜렷하게 볼 수 있는 데, 그림에 표시된 두개의 BOX-PLOT을 보면 남자조사자들에게 높은 점수를 받았던 모델들이 여자조사자들에게는 평균에도 못받는 점수를 얻고 있음을 알 수 있다. 이를 통해서도 남녀조사자들간에는 시각차가 존재하는 것으로 보여진다. 그리고, 이는 p-값의 크기로도 비교가 가능한데, 결과3.2에서 출력된 p-값이, 결과3.1의 p-값보다 현저하게 작다는 것을 알 수 있고, 또한 부호순위검정의 결과에서도 평균순위의 차이가 결과3.1의 그것보다 훨씬 크다는 것을 알 수 있다.



<결과 3.2 남자조사자 3명(B2-B4) : 여자조사자 6명(G1-G6) >

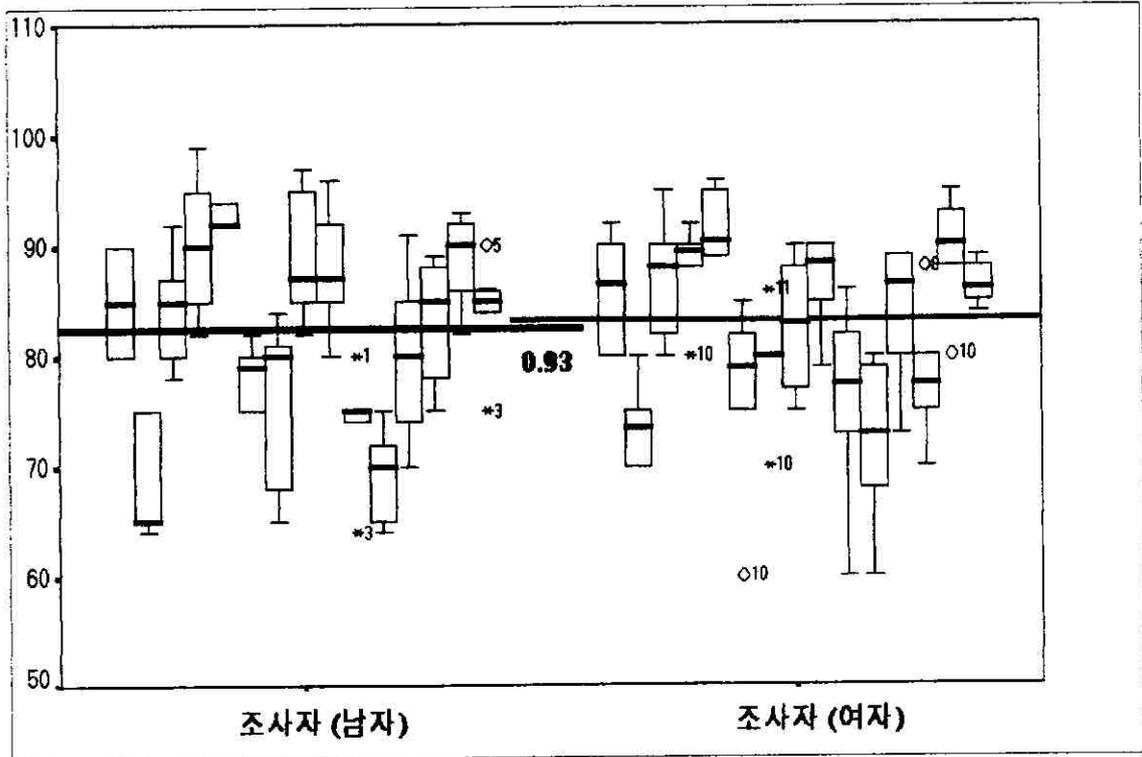
* t-tests for paired samples			t-value	df	2-tail Sig
Mean	SD	SE of Mean			
5.0588	6.624	1.606	3.15	16	.006
95% CI (1.652, 8.465)					

* Wilcoxon Matched-Pairs Signed-Ranks Test		
Mean Rank	Cases	
10.21	12	- Ranks (GIRL6 LT BOY3)
3.38	4	+ Ranks (GIRL6 GT BOY3)
	1	Ties (GIRL6 EQ BOY3)
	17	Total
Z =	-2.8181	2-Tailed P = .0048

### 3.3 여자모델만을 대상으로 한 경우(N=15)

앞서 검정한 결과3.1, 결과3.2와는 대조적으로, 여자모델만을 대상으로 한 경우에는 귀무가설(남녀의 외모평가는 차이가 없다)을 기각할 수 없는 결과를 보인다. t-검정의 경우, 평균차이가 -0.9333으로 상당히 작고, 부호순위검정에 있어서는 cases상으로는 여자조사자들이 같은 여자모델들에게 남성들에 비해서 더 좋은 점수를 주었다고는 하나, 그 평균순위의 차이는 크지 않기 때문에, '남,녀 조사자들의 점수차이는 없다'라는 결론을 얻을 수 있었던 것으로 보인다.



<결과 3.3 남자조사자 5명(B1-B5) : 여자조사자 6명(G1-G6)>

* t-tests for paired samples				t-value	df	2-tail Sig
Mean	Paired Differences SD	SE of Mean				
- .9333	3.835	.990		-.94	14	.362
95% CI (-3.058, 1.191)						
* Wilcoxon Matched-Pairs Signed-Ranks Test						
Mean Rank	Cases					
6.25	4	- Ranks (G1R6 LT B0Y5)				
6.63	8	+ Ranks (G1R6 GT B0Y5)				
	3	Ties (G1R6 EQ B0Y5)				
	15	Total				
Z =	-1.0983	2-Tailed P = .2721				

\* 신뢰성이 가장 높은 경우<참고>

1) 전체자료를 대상으로 한 경우 (N=32)

* t-tests for paired samples						
Mean	Paired Differences SD	SE of Mean	t-value	df	2-tail Sig	
2.5938	6.596	1.166	2.22	31	.034	
95% CI (.215, 4.972)						
- - - - Wilcoxon Matched-Pairs Signed-Ranks Test						
Mean Rank	Cases					
16.87	19	- Ranks	(GIRL6 LT BOY3)			
11.45	10	+ Ranks	(GIRL6 GT BOY3)			
	3	Ties	(GIRL6 EQ BOY3)			
	32	Total				
Z = -2.2272                      2-Tailed P = .0259						

2) 남자모델만을 대상으로 한 경우(N=17)

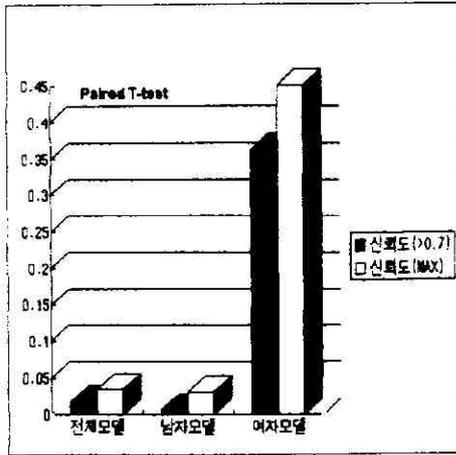
* t-tests for paired samples						
Mean	Paired Differences SD	SE of Mean	t-value	df	2-tail Sig	
4.0588	7.015	1.701	2.39	16	.030	
95% CI (.451, 7.667)						
- - - - Wilcoxon Matched-Pairs Signed-Ranks Test						
Mean Rank	Cases					
10.82	11	- Ranks	(GIRL4 LT BOY3)			
5.67	6	+ Ranks	(GIRL4 GT BOY3)			
	0	Ties	(GIRL4 EQ BOY3)			
	17	Total				
Z = -2.0119                      2-Tailed P = .0442						

3) 여자모델만을 대상으로 한 경우(N=15)

* t-tests for paired samples						
Mean	Paired Differences SD	SE of Mean	t-value	df	2-tail Sig	
- .7333	3.654	.943	-.78	14	.450	
95% CI (-2.757, 1.291)						
* Wilcoxon Matched-Pairs Signed-Ranks Test						
Mean Rank	Cases					
4.20	5	- Ranks	(GIRL2 LT BOY5)			
6.80	5	+ Ranks	(GIRL2 GT BOY5)			
	5	Ties	(GIRL2 EQ BOY5)			
	15	Total				
Z = -.6625                      2-Tailed P = .5076						

비교

신뢰도의 정도가 과연 검정결과에 영향을 미치는가? 신뢰계수가 0.7이상인 경우의 검정결과와 가장 높은 신뢰계수를 가지는 경우의 검정결과를 비교했을 때, 신뢰도의 정도에 따라 그결과가 달라지는 경우는 찾아볼 수 없었다. 즉, 이 자료에 의하면 신뢰도의 정도가 검정결과에 큰 영향을 미친다고 볼 수 없다는 것이다. 그러나, 신뢰도가 높아질수록, p-값이 커지는 현상을 볼 때, 경우에 따라서는 주관적으로 설정하는 신뢰도의 기준이 검정결과도 바꿀수 있으리라고 생각된다.



대상집단	Alpha > 0.7		Alpha 최대	
	T-TEST	RANK	T-TEST	RANK
전체자료	.016	.0432	.034	.0259
남자모델	.006	.0048	.030	.0442
여자모델	.362	.2721	.450	.5076

#### 4. 결 론

지금까지 사람의 외모를 평가함에 있어서, 남자와 여자간에 과연 차이가 있는가를 검증해보기 위해서, 조사된 자료를 바탕으로 분석해 보았다. 그 결과를 요약해 보면,

- 1) 전체적으로는, 즉 대상(모델)을 남,녀로 구분하지 않은 경우(n=32)에는 '외모 평가기준'은 남,녀 간에 차이가 있다는 결과를 얻을 수 있었다.
- 2) 대상을 남자로만 제한한 경우(n=17) 역시 남,녀간의 외모평가의 차이는 유의하다는 결과를 얻을 수 있었다.
- 3) 대상을 여자로만 제한한 경우(n=15)에는, 1,2와는 달리 남,녀간의 차이가 유의하지 않다는 결과를 얻을 수 있었다.

이 결과들을 종합해 볼 때, 2)의 결과(대상이 남자인 경우)에서 나타나는 뚜렷한 '평가기준의 차이'가 전체자료의 분석결과에 영향을 미쳐 동일한 결과를 보이는 것으로 생각된다.

결국 대상이 남자인 경우에는 남녀간에 시각차이가 존재하지만, 대상이 여자인 경우에는 그렇지 않다는 것인데, 왜 여자의 경우에는 시각의 차이가 없는 걸까? 아마도 그이유는 남녀 모두가 여자의 외모를 평가하는데 이미 익숙해져 있기 때문인 것으로 생각된다. 따라서 좋은 점수를 받을 수 있는, 즉 이쁜 여자의 외모는 어떤 것인가? 의 답이 남녀조사자 모두에게 어느정도 굳어져 있기 때문이라고 보여진다.

좀더 나아가 분석과정을 살펴보면, 남자들은 같은 남자에게, 여자들 역시 같은 여자들에게 상대적으로 더 높은 점수를 주고 있다는 것을 알 수 있다. 그런데, 여자들이 同性에게 준 점수에 비해서 남자들이, 同性에게 준 점수가 훨씬 후하다는 것을 알 수 있다. 이는 바꿔 말하면, 남자들이 異性에게 준 점수보다 여자들이 異性에게 준 점수가 훨씬 적었다고 볼 수도 있다. 異性의 외모를 평가함에 있어 남자조사자들보다 여자조사자들의 눈(평가기준)이 높기 때문에 혹은, 여자가 남자의 외모를 평가할 때 고려하는 사항이, 남자가 여자의 외모를 평가할 때보다 더 많기 때문이 아닐까 생각된다.

이상의 분석을 마치며 보완하여야 할 사항으로는, 분석결과에 영향을 미쳤을 것이라 여겨지는 남녀 조사자들의 연령차의 해소이다. 이 약점을 보완하고, 그밖의 사항(표본의 연령층의 다양화 등...)들을 고려한다면 좀 더 나은 결과를 얻어낼 수 있을 것으로 기대된다.

#### <참고문헌>

- [1] 허명희(1991), 설문지 시험지문항의 신뢰성 분석, 「응용통계연구」 제4권 제 1호.
- [2] 서울대학교 자연과학대학 계산통계학과 편저 (1988). 「현대통계학」, 영지문화사.

## GPA의 문제점 및 해결방안 연구

이 철 영

(지도교수 : 권 세 혁)

한남대학교 경상대학 응용통계학과

## 요 약

본 연구는 가장 일반적인 순위결정방법의 도구로 사용되는 GPA가, 경우에 따라서 가질수 있는 심각한 문제점을 지적하고, 이를 해결하기 위하여 인자분석과 비모수적인 방법 등을 응용하는 방법을 통하여 좀 더 공정한 순위를 만들어내는, GPA의 문제점을 해결하는 새로운 순위방법을 제시하였다.

## 1. 서 론

## 연구목적

모든 사람들은 수년간에 걸쳐 학교에서 교육을 받게 되는데, 그 배움의 수준이나 숙달의 정도를 평가받기 위해 일정한 절차에 따라서 시험을 치루게 된다. 그리고 그 결과에 의해서, 제각기 '순위'라는 것을 배정받게 된다.

현재 어디서나 이 '순위'를 판정하는 도구로, 'GPA( Grade-point average )'방법이 쓰인다. 그 이유로는 먼저 가장 손쉬운 방법이라는 것과, 이를 대체할 만한 마땅한 계산방법이 없기 때문이라고 생각된다. 그러나 과연 이 방법이 타당하다고 할 수 있는가? 특히, 대학의 경우, 시험이란 것이, 이전(초.중.고)의 그것과 성격을 달리하고 있을 뿐더러, 교수님들의 주관적인 배점방식이 허용되고 있고, 학생들이 수강과목을 선택하여 들을 수 있다. 이런 상황에서 모든 성적을 절대적으로 평가해버리는 'GPA'방식은 그리 효율적이라 할 수 없는 것이다. GPA방식에서는 각 과목의 점수들의 '분산'이 전혀 고려되지 않기 때문이다. 서로 다른 분산을 가지고 있는 과목들의 점수를, 단순히 그들의 합만으로 순위를 만들어 학생들의 배움의 수준을 측정한다는 것은 쉽게 납득하기 어려운 일이지만, 특히 성적처리에 있어서는 거의 모든 경우에, GPA 방식이 쓰이고 있고, 또 그 대상자들 역시 무감각하게 그 결과를 받아들이고 있는 것이 현실이다.

'순위'라는 것이, 만약 불가피하게 꼭 필요한 것이라면, 그것은 공정하게 결정되어야 하는 것이다. 그러나, 이 공정한 순위결정 도구로서의 GPA방식은 문제점을 지니고 있는 것이 사실이다. 이에 탐색적 분석방법을 사용하여 성적을 구성하고 있는 과목들을 대상으로, 새로운 '순위'를 만들어 보고자 한다. 즉, GPA방식의 문제점을 해결할 수 있는 방법들로 재순위를 만드는 것이다. 그리고, 이와 더불어 GPA방식에 의한 결과에는 어떤 원인들이 영향을 주고 있는가를 측정이 가능한 자료를 이용하여 알아보고, 그 원인이 어떤 의미를 가지는 가도 함께 알아보고자 한다.

## 2. 본 론

### 자료설명

분석에 사용된 자료는 95년에 졸업한 응용통계학과 학생 20명(91학번)의 성적자료를 이용하여 작성되었다. 대학강의의 특성(선택가능-)상 상당히 많은(98개)의 과목수가 측정되었으나, 대부분의 선택과목들이 극히 일부학생들만이 수강한 것들이었으므로, 연구에 모든 과목을 변수로 사용하는 것은 어려운 일이었다. 이런 이유로, 자료로 사용한 과목들은 모든 학생이 반드시 수강하여야만 하는 필수과목들로 제한하였다. 그리고, 분석에 사용된 자료로 쓰인 과목들은 모두 취득학점이 3 학점이다.(자료출처--한남대학교 응용통계학과 성적기록부)

사용된 변수는 23개의 교과목중, 2학기 동안 수강해야만 하는(동일한 교수님과 동일한 교재를 사용하는) 과목들, 또는 강의 내용이 비슷하다고 여겨지는 과목들을, 그들의 상관관계를 고려해본 후에 하나의 변수들로 만들었다. 23개 교과목을 이용해 만들어진 변수는 총 16개로 다음과 같다.

[ 통계수학 통계방법 전산통계 수리통계 품질관리 회귀분석 표본론 실험계획  
국 어 영 어 컴퓨터 역 사 경영 학 경제 학 통계 학 회 계 학 ]

### 3. 인자분석을 이용하는 방법

#### 1) 기초통계량

<표3.1>은 각 과목별 기초통계량을 정리한 것이다. 우선 눈에 띄는 부분은 전공과목들의 평균과 편차에 대한 것인데, 통계수학을 제외한 전공과목들의 평균이 다른 과목(교양, 계열기초)들에 비해 높다는 것을 볼 수 있다. 아마도 이 결과는 자료가 졸업생들의 성적기록부에서 발췌된 것이므로, 재수강의 영향을 받았다고 볼 수 있을 것 같다. 아울러 다른 과목들이 주로 1,2학년 때 수강하는 필수과목임을 감안할 때, 전공과목들은 비교적 고학년때 수강한 것이 그 이유가 될 수도 있을 것이다. 이외에, 편차가 상대적으로 큰 과목들은 계열기초과목집단에 속하는 것들로, GPA순위에 가장 큰 영향을 미칠 것으로 예상된다.

<표3.1> 각 변수별 기초통계량

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
통계수학	20	2.95000	0.68948	59.00000	1.93333	4.33333
통계방법	20	3.85500	0.45938	77.10000	3.00000	4.50000
전산통계	20	3.17750	0.73601	63.55000	2.25000	4.40000
수리통계	20	3.34000	0.76856	66.80000	1.75000	4.50000
품질관리	20	3.56750	0.57612	71.35000	2.25000	4.40000
회귀분석	20	3.36500	0.77885	67.30000	2.00000	4.50000
표본론	20	3.75000	0.72001	75.00000	2.00000	4.50000
실험계획	20	3.77000	0.61823	75.40000	2.30000	4.50000
국어	20	3.19250	0.69798	63.85000	1.75000	4.00000
영어	20	3.06750	0.79675	61.35000	1.15000	4.15000
컴퓨터	20	2.67500	0.83153	53.50000	1.50000	4.30000
역사	20	3.51000	0.58300	70.20000	3.00000	4.50000
경영학	20	2.66500	1.08690	53.30000	1.00000	4.30000
경제학	20	2.83500	0.84497	56.70000	1.50000	4.30000
통계학	20	3.80000	0.75044	76.00000	2.00000	4.50000
회계학	20	3.13500	0.80608	62.70000	1.50000	4.50000

2) 인자분석

모형 및 가정

$x' = (x_1, x_2, \dots, x_m)$  평균  $\mu$ , 공분산행렬  $\Sigma$ 를 갖는 다변량분포를 따르는 확률벡터.

인자모형은  $x$ 를 적은 개수 ( $m < p$ )의 관측이 안되는 공통인자  $F_1, F_2, \dots, F_m$ 과 특정인자  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ 의 선형식으로, 그 형태는 다음과 같다.

$$X_i - \mu_i = \sum_{k=1}^m l_{ik} F_k + \varepsilon_i, \quad i=1, 2, \dots, p$$

$F_k, k=1, 2, \dots, m$  (공통인자)  $\varepsilon_i, i=1, 2, \dots, p$  (특정인자).  $l_{ik}, i=1, \dots, p, k=1, \dots, m$

가정) 1)  $F_i$  들은 서로 독립이고 평균 0, 분산 1인 관측이 안되는 가설적인 확률변수이다.

2)  $\varepsilon_i$  들은 서로 독립이고 평균 0, 분산  $\psi_i$  인 관측이 안되는 확률변수이다.

3)  $F$  와  $\varepsilon$ 은 서로 독립이다.

교과목의 차원 축소

인자분석의 첫단계는 우선 몇개의 인자가 있어야 할 것인가를 결정하는 것이다. 이를 판단하는 방법으로 먼저 고유값의 크기로 결정하는 방법, 그리고 인자의 공헌도에 의한 방법을 사용한 후, 그 결과로 타당한 보유인자를 결정하겠다.

분석은 세계의 집단(전공과목, 교양과목, 계열기초과목)별로 각각 실시하도록 하겠다. 이유는 집단의 구분없이 인자분석을 실시할 경우에, 본래의 분석목적을 벗어나는, 과목성격차이나 혹은 그밖의 이유에 의해서 만들어지는 인자들을 애초에 제거하기 위함이다.

인자분석에 앞서 변수들간의 상관관계를 살펴보기 위해, 전공과목으로 분류되는 8개 과목(변수) 사이의 상관행렬을 계산한 결과가 아래 <표3.2>에 주어져 있다. 상관행렬을 살펴보면, (통계수학, 수리통계)가 가장 높은 상관관계를 가지고 있고, (전산통계, 수리통계), (표본론, 회귀분석)순으로 높은 상관관계를 볼 수 있다.

<표3.2> 표본상관행렬

	통계수학	통계방법	전산통계	수리통계	품질관리	회귀분석	표본론	실험계획
통계수학	1.00000 0.0	0.51430 0.0203	0.69187 0.0007	0.86013 0.0001	0.58333 0.0069	0.58627 0.0066	0.61420 0.0040	0.53135 0.0159
통계방법	0.51430 0.0203	1.00000 0.0	0.37901 0.0994	0.46749 0.0377	0.40584 0.0758	0.50876 0.0220	0.49726 0.0257	0.42124 0.0644
전산통계	0.69187 0.0007	0.37901 0.0994	1.00000 0.0	0.77673 0.0001	0.25078 0.2862	0.56000 0.0102	0.29969 0.1992	0.29339 0.2093
수리통계	0.86013 0.0001	0.46749 0.0377	0.77673 0.0001	1.00000 0.0	0.42090 0.0646	0.57794 0.0076	0.55212 0.0116	0.63570 0.0026
품질관리	0.58333 0.0069	0.40584 0.0758	0.25078 0.2862	0.42090 0.0646	1.00000 0.0	0.42253 0.0635	0.68230 0.0009	0.56898 0.0088
회귀분석	0.58627 0.0066	0.50876 0.0220	0.56000 0.0102	0.57794 0.0076	0.42253 0.0635	1.00000 0.0	0.79353 0.0001	0.38902 0.0900
표본론	0.61420 0.0040	0.49726 0.0257	0.29969 0.1992	0.55212 0.0116	0.68230 0.0009	0.79353 0.0001	1.00000 0.0	0.51906 0.0190
실험계획	0.53135 0.0159	0.42124 0.0644	0.29339 0.2093	0.63570 0.0026	0.56898 0.0088	0.38902 0.0900	0.51906 0.0190	1.00000 0.0

표본상관행렬에 기초한 인자분석을 실시했을 때 얻어지는 결과가 <표3.3>에 주어져 있다. 고유값의 크기에 따라 고유인자를 결정하는 방법은, 표본상관행렬로부터 인자를 유도할 경우, 상관행렬의 고유값이 1 보다 큰 것들의 갯수에 해당하는 만큼의 인자를 보유하는 방법이다. 결과를 보면, 이 방법에 의해서 보유되는 인자는 첫번째 인자뿐이다. 그리고 인자패턴행렬의 값들을 살펴보면, 8개의 과목이 모두 양의 적재를 가짐을 알 수 있다. 그중 높은 적재를 보이는 것이 '통계수학'과 '수리통계'이고, 상대적으로 가장 낮은 적재를 보이는 과목은 '통계방법'이다.

<표3.3> 상관행렬의 고유값(전공과목)

Kaiser's Measure of Sampling Adequacy: Over-all MSA = 0.66400785							
통계수학	통계방법	전산통계	수리통계	품질관리	회귀분석	표본론	실험계획
0.811500	0.950567	0.568789	0.633053	0.589516	0.641934	0.617204	0.644418
Prior Communality Estimates: ONE							
Preliminary Eigenvalues: Total = 8 Average = 1							
	1	2	3	4			
Eigenvalue	4.7472	1.0713	0.7465	0.5959			
Cumulative	0.5934	0.7273	0.8206	0.8951			
1 factors will be retained by the NFACTOR criterion.							
Factor Pattern							
	FACTOR1			FACTOR1			
통계수학	89 *	품질관리		64 *			
통계방법	61 *	회귀분석		75 *			
전산통계	65 *	표본론		78 *			
수리통계	87 *	실험계획		65 *			

교양과목의 경우, 표본상관행렬을 살펴본 결과 (컴퓨터, 영어)가 가장 높은 상관관계를 보이고 있는 반면, 이와는 대조적으로 (컴퓨터, 국어)가 가장 낮은 상관관계를 보이고 있다. 그리고, 교양과목 역시 앞서 분석한 전공과목(표2.1-2)과 마찬가지로 한 개의 고유인자만을 가지는 것을 알 수 있다. 그러나 전공과목의 고유인자가 전체분산의 59.34%(Proportion)를 설명하고 있는데 반해, 교양과목의 고유인자는 38.1%만큼의 설명력밖에 가지지 못하고 있다. 이 원인은 보유되는 첫번째 인자와 나머지 세개의 인자들의 고유값차이가 크지 않다는 것이다. 그렇다고, 만약 고유인자로 두개의 인자를 선택할 경우, 고유인자수에 비해 변수의 수가 너무 작고, 또한 두번째 인자의 패턴행렬에서 음의 적재가 생기는 문제등이 발생한다. 이 문제점들을 고려해볼 때, 고유인자수는 하나로 사용함이 적당할 것으로 생각되어지므로, 고유인자수를 하나로 하여 계속 분석을 진행하겠다.

계열기초과목의 상관관계를 살펴보면, 먼저 실시했던 전공과목과 교양과목에 비해서 전체적으로 낮은 상관관계를 보이고 있다. 그중 높은 상관관계를 보이는 것은 (경제학, 회계학)이고, (경제학, 경영학)이 낮은 상관관계를 가진다. 계열기초과목의 경우에도 전공과목, 교양과목과 마찬가지로 고유값이 1보다 큰 하나의 인자만을 고유인자로 결정하였다. 이 고유인자는 전체분산의 52.31%를 설명하고 있다. 인자패턴행렬을 살펴보면, '회계학'과 '경제학'이 상대적으로 높은 인자적재를 가지고 있다.

## 요약 및 인자의 해석

지금까지 각 집단별로 보유인자수를 결정하기 위하여 인자분석을 실시해 보았다. 그 결과로, 상관행렬의 고유값의 크기에 의한 방법에 의해서, 집단별로 각각 한 개씩의 보유인자를 유도할 수 있었다. 그리고, 보유되는 인자들은 모두가 대응되는 각 과목들과 고르게 연관되어 있는 것으로 볼 수 있다. 따라서 인자들은 각각의 집단을 대표하는 성격을 띄게 된다고 볼 수 있는데, 구지 이름을 붙이자면, 해당집단 과목들의 '학력'정도가 적당할 듯 싶다. 예를 들어, 첫번째 집단(전공과목)의 경우에 인자의 의미는 '전공과목의 학력' 이 될 것이다.

### 3) 인자점수의 계산

인자점수를 필요로 하는 이유로 다음의 두가지를 들 수 있다. 그 첫째는 원래 변수들의 차원에서 보다는, 변수들 간의 구조적 관계에서 의해 요약된 인자에 의해 표현되는 더 낮은, 공통인자의 차원에서 각 객체를 검토하기 위함이고, 둘째는 각 객체에 유도된 인자점수를 후속적인 통계분석에서 구체적인 개념을 가진 새로운 변수, 혹은 지표로 이용하기 위함이다.

여기서 의미하는 인자의 점수는 곧 관찰할 수 없는, 혹은 하나의 확률변수로 고려되었으므로, 인자점수는 확률적 인자에 대한 예측값의 의미를 가진다. 그러나 이 값은 인자패턴행렬의 비유일성에 의해서 유일하지 않게 된다.

다음은 회귀적인 방법에 의해서 인자점수를 계산할 때, 이용되는 인자점수계수들을 각 집단의 인자별로 정리한 것이다.

통계수학	0.30014
통계방법	0.08790
전산통계	-0.09217
수리통계	0.38900
품질관리	0.12004
회귀분석	0.23685
표본론	0.07549
실험계획	0.03099

전공 과목의 학력

국 어	0.39581
영 어	0.30424
컴퓨터	0.05915
역사	0.29390

교양과목의 학력

경영학	0.16711
경제학	0.29264
통계학	0.22084
회계학	0.45877

계열기초과목의 학력

인자의 점수를 평가하기 위해서 먼저, 회귀적 방법에 의한 인자점수를 구해 보았다. 이 결과는 표준화된 인자점수계수와 원래의 자료행렬의 곱으로 계산되어지는데, 이 방법으로 세 집단 각각의 인자점수를 구하였다. (이 인자점수들은 평균이 0, 분산이 1로 표준화된 점수들이다 <표3.4>)

<표3.4> 인자점수(표준화)

집단 학생	전공 과목	교양 과목	계열 기초
1	0.40977	-0.18376	-0.33374
2	-0.80366	-1.12719	-0.08149
3	1.60102	0.72589	1.28148
4	0.36002	0.43362	0.67531
5	0.91715	0.33401	1.04566
6	1.17298	-1.01703	0.63256
7	-0.85365	1.12885	-0.51351
8	-1.44790	-0.46152	-1.56401
9	0.19704	0.07056	-0.11775
10	1.32409	0.48249	0.43371
11	0.04881	-0.40448	-0.26910
12	1.46459	1.02799	1.68278
13	-1.17006	-0.54135	-0.64785
14	-0.20709	0.66917	-0.50752
15	-1.42847	-1.13004	-1.04784
16	-0.85651	-0.28882	-1.18526
17	-0.60790	0.84340	-0.14777
18	-0.62550	-0.82737	0.99580
19	0.50824	0.33205	-0.43253
20	-0.00297	-0.06647	0.10106

<표3.5> GPA순위와 각 집단별 인자점수순위

학생	GPA 순위	인 자 점 수 순 위		
		전공과목	교양과목	계열기초
1	11	7	12	13
2	16	15	19	9
3	2	1	4	2
4	5	8	7	5
5	4	5	8	3
6	6	4	18	6
7	14	16	1	16
8	19	20	15	20
9	9	9	10	10
10	3	3	6	7
11	13	10	14	12
12	1	2	2	1
13	17	18	16	17
14	10	12	5	15
15	20	19	20	18
16	18	17	13	19
17	12	13	3	11
18	15	14	17	4
19	7	6	9	14
20	8	11	11	8

<표3.5>는 GPA에 의한 순위와 <표3.4>의 인자점수를 이용한 순위들을 비교한 것이다. 즉, 이 비교의 목적은 GPA와 가장 깊은 연관을 가지고 있는 인자를 찾기 위한 것이다. 이를 위한 방법으로 GPA순위와 각 집단별순위의 상관관계를 살펴보았다.

그 결과를 보면, GPA순위와 가장 큰 상관관계를 보이는 과목집단은 전공과목과목집단이다. 이를 통해, 연구에 사용된 자료에 의한 GPA순위는 전공과목들의 영향을 많이 받고 있다고 생각할 수 있을 것이다.

<표3.6> GPA순위와 각 집단별 순위의 상관관계

Pearson Correlation Coefficients / Prob >  R  under Ho: Rho=0 / N = 20			
	전공과목	교양과목	계열기초
GPA	0.95038	0.60602	0.79098
RANK FOR VARIABLE GPA	0.0001	0.0046	0.0001

4) 가중치를 고려한 인자점수의 순위

가중치를 고려할 경우에 가장 손쉬운 방법은 집단(전공, 교양, 계열기초)간의 모평균차이를 비교함으로써 그 차이에 따라 가중치를 주는 방법이다. 그런데, <표3.4>에서 구한 인자점수들은 표준화된 값이므로, 이 점수에 가중치를 줄 경우 문제가 있다고 본다. 따라서 NOSTD -OPTION을 사용하여 비표준 인자점수를 계산하였다.

<표3.7> 인자점수 (비표준화)

학생	인자 점수			학생	인자 점수		
	전공과목	교양과목	계열기초		전공과목	교양과목	계열기초
1	4.16010	3.26869	3.28191	11	3.86313	3.12973	3.32566
2	3.26386	2.62145	3.46096	12	4.86144	4.11973	4.99103
3	4.98523	3.86788	4.65081	13	3.04269	2.99466	2.99994
4	4.09490	3.70668	4.13576	14	3.72069	3.89735	3.13799
5	4.52913	3.56836	4.38379	15	2.83512	2.59117	2.65818
6	4.72696	2.70701	4.07726	16	3.30032	3.09139	2.53922
7	3.18821	4.16044	3.18785	17	3.40025	3.95973	3.45969
8	2.83382	3.11352	2.21350	18	3.45569	2.76479	4.30447
9	3.97784	3.48581	3.47567	19	4.22707	3.69103	3.16238
10	4.80184	3.68680	3.94102	20	3.84807	3.30777	3.66172

집단으로 분류된 과목들의 성적을 이용하여, 다음의 가설을 검정할 수 있다. 먼저  $Pr > F$ , 즉 유의확률값을 보면 0.0004로 '각 과목집단들의 점수는 서로 같지 않다'고 결론을 내릴 수 있다. <표3.8> 따라서, 이 모평균의 차이를 '가중치'로 이용해 보았다. 즉, 각 집단별 과목들의 평균점수들은 각 집단에 속한 과목들의 난이도(?)를 가늠할 수 있는 척도로 생각한 것이다.

<표3.8> 분산분석의 결과와 각 집단의 평균

Analysis of Variance Procedure					
Dependent Variable: GRADE					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	10.47652039	5.23826020	8.07	0.0004
Error	317	205.83443092	0.64931997		
Corrected Total	319	216.31095132			
	R-Square	C. V.	Root MSE	GRADE Mean	
	0.048433	24.48554	0.8058039	3.2909374	
CLASS	N	Mean	Std Dev	Minimum	Maximum
전공집단	160	3.4718748	0.7255476	1.7500000	4.5000000
교양집단	80	3.1112500	0.7800509	1.1500000	4.5000000
계열기초	80	3.1087500	0.9682548	1.0000000	4.5000000

<표3.8>의 검정에 의해, 모평균의 차이로 얻어진 모평균의 차를 가중치로 이용한 순위와 기존의 GPA순위를 비교해보았다. <표3.9>를 살펴보면, 8명의 학생의 순위에 차이가 있음을 알 수 있다. 이 차이의 해석을 돕기위해 '과대평가'와 '과소평가'의 개념을 사용해보도록 하겠다. 물론 이 개념들은 각 과목들의 분산을 고려하지 못하는 GPA방법의 문제점에 기인한 것이다. 즉, '과대평가'로 보이는 학생들은 편차가 큰 과목에서 높은 점수를 받았으리라 예상할 수 있다.

GPA 에 의해 과대평가된 학생 : 학생 = 1, 10, 11, 14, 20 (5명)

GPA 에 의해 과소평가된 학생 : 학생 = 6, 9, 17, 18, (4명)

<표3.9> GPA순위와 가중치가 고려된 인자점수의 순위

학생	GPA 순위	가중치를 고려한 인자점수				인자점수 순위
		전공 과목	교양 과목	계열 기초	합	
1	11	1.33439	1.04930	1.17566	3.55935	12
2	16	1.04691	0.84153	1.23981	3.12824	16
3	2	1.59905	1.24165	1.66604	4.50674	2
4	5	1.31347	1.18990	1.48153	3.98491	5
5	4	1.45276	1.14550	1.57039	4.16865	3
6	6	1.51621	0.86899	1.46058	3.84579	6
7	14	1.02265	1.33557	1.14197	3.50018	14
8	19	0.90897	0.99949	0.79293	2.70139	19
9	9	1.27592	1.11900	1.24507	3.64000	8
10	3	1.54023	1.18352	1.41177	4.13553	4
11	13	1.23913	1.00469	1.19134	3.43516	15
12	1	1.55935	1.32250	1.78791	4.66976	1
13	17	0.97597	0.96133	1.07466	3.01196	17
14	10	1.19344	1.25111	1.12411	3.56866	11
15	20	0.90939	0.83181	0.95223	2.69343	20
16	18	1.05860	0.99239	0.90961	2.96060	18
17	12	1.09066	1.27114	1.23935	3.60115	10
18	15	1.10844	0.88754	1.54197	3.53795	13
19	7	1.35587	1.18488	1.13284	3.67360	7
20	8	1.23430	1.06185	1.31172	3.60787	9

#### 4. 각 과목별 순위를 이용하는 방법

각 과목들의 원래 성적이 없는 관계로, 각 과목별 순위를 얻기가 곤란하므로 의미상 연관이 있다고 여겨지는 과목들을 묶어 변수들을 재조정후 순위를 얻을 수 있었다.<표4.1> 이는 과목별 순위를 얻기위해 불가피한 과정이었으며, 이에 따라 비교대상인 GPA 역시 새롭게 만들어진 변수들에 의해서 새로 구한 것이다.

과목 1=통계수학

과목 3=프로그래밍실습+전산통계+교육컴퓨터개론

과목 5=통계적품질관리+실험계획법

과목 7=대학국어

과목 9=경제학원론+회계원리

과목 2=통계적 방법+통계학

과목 4=기초확률론+수리통계

과목 6=회귀분석+표본론

과목 8=대학영어

과목10=경영학원론+한국의 전통문화

<표4.1 각 과목별 순위>

학생	과목1	과목2	과목3	과목4	과목5	과목6	과목7	과목8	과목9	과목10
1	7	8	16	12	13	5	10	7	19	15
2	20	12	12	19	10	14	17	20	11	14
3	3	2	2	2	2	1	4	12	3	2
4	11	6	6	12	5	5	2	2	2	14
5	5	14	9	5	4	7	6	5	5	14
6	6	12	7	2	8	2	10	19	9	3
7	15	3	16	15	12	18	19	3	13	9
8	17	20	19	20	18	16	17	16	16	17
9	9	8	18	12	8	8	10	12	7	4
10	2	4	1	3	11	3	4	13	11	14
11	9	10	17	13	10	10	10	14	15	9
12	1	2	4	5	1	6	13	4	1	1
13	16	16	13	18	17	17	20	10	20	19
14	15	13	8	9	15	10	6	7	5	9
15	19	19	20	18	16	20	11	18	18	17
16	18	19	10	14	19	11	17	17	17	20
17	13	6	16	18	15	15	19	2	9	9
18	10	17	5	7	20	19	14	10	15	18
19	4	15	3	6	6	14	2	10	7	9
20	12	9	12	9	4	12	13	15	13	14

<표4.1>의 순위를 자료로 이용한 주성분분석의 결과에서 공헌도가 56%인 첫번째 주성분을 유

도할 수 있었다. 나머지 주성분들은 설명력이 떨어지므로, 첫번째 주성분만을 분석에 이용하였다.

<표4.2 주성분 분석결과>

	Eigenvalue	Difference	Proportion	Cumulative
PRIN1	5.65489	4.23416	0.565489	0.56549
PRIN2	1.42073	0.52390	0.142073	0.70756
:	:	:	:	:
PRIN10	0.03945	:	0.003945	1.00000

<표4.3>은 <표4.2>의 결과로 얻어진 첫번째 주성분의 크기에 의해서 얻어진 순위와 기존의 GPA를 비교한 것이다. 가장 눈에 띄는 변화가 1등과 2등이 순위결정방법에 의해서, 뒤바뀐 것이다. 그이유로 <표4.1>- 각 과목별 순위-를 보면 쉽게 알 수 있는데, 결과적으로 학생2의 경우, 전체적으로 학점이 후한 과목에서 열세를 보임으로서, 실제 등수와는 관계없이 점수상으로 득(?)을 보게 된 것이라 할 수 있다.

GPA 에 의해 과대평가된 학생 : 학생 = 1, 8, 12, 13, 19 (5명)

GPA 에 의해 과소평가된 학생 : 학생 = 3, 6, 15, 16, 20, (5명)

<표4.3 주성분순위와 GPA순위>

순위	주성분	GPA	순위	주성분	GPA
학생	순위	순위	학생	순위	순위
1	11	10	11	12	12
2	16	16	12	2	1
3	1	2	13	18	17
4	4	4	14	9	9
5	6	6	15	19	20
6	5	7	16	17	18
7	14	14	17	13	13
8	20	19	18	15	15
9	8	8	19	7	5
10	3	3	20	10	11

## 5. 각 과목들의 점수를 표준화 하는 방법

이 방법은 각 과목들의 점수를 공통평균과 공통분산으로 통일시킴으로써, 그 점수를 재조정하여, 그 결과를 순위결정에 이용하고자 하는 것이다. 원래의 자료를 처리하는(표준화) 과정에는 기존의 GPA방법과 맥락을 함께 하는 것으로, '표준화 GPA방법'이라 할 수 있을 것이다. (편의상, 4. '각 과목들의 순위를 이용하는 방법'에서 사용되었던, 조정된 10개의 과목을 이용하였고, 새로운 학점부여에 있어서는 '상대평가'방법을 취하였다.)

다음은 방법 2.에서 사용된 10개의 과목을 Mean=B-(3.0), STD =1.0으로 표준화시킨 결과이다.  
 <표5.1 평균=B0, 분산=1로 표준화된 각 과목별 점수>

과목	과목 1	과목 2	과목 3	과목 4	과목 5	과목 6	과목 7	과목 8	과목 9	과목 10
학생 1	3.32241	3.33536	2.33761	2.87986	2.73827	4.00103	3.07763	3.67249	1.92164	2.29132
2	1.48715	2.98603	2.62436	1.87865	3.06139	2.77233	2.04259	0.87964	2.77196	2.75603
3	4.46325	4.59297	4.68894	4.54853	4.28925	4.36241	4.37143	3.30501	4.08609	4.61486
4	2.97520	3.54496	3.65665	2.87986	4.03076	4.00103	4.63019	4.18696	4.85911	2.75603
5	3.86803	2.49696	2.91111	4.08130	4.09538	3.63965	3.98329	3.81948	3.85419	2.75603
6	3.52082	2.98603	3.19786	4.54853	3.70763	4.21786	3.07763	1.10012	3.08117	4.26633
7	2.23117	4.45323	2.33761	2.21239	2.80290	1.47135	1.78383	4.03997	2.69466	3.33691
8	1.98317	1.23935	1.76412	0.87745	1.83353	2.19412	2.04259	1.98208	2.30815	2.17514
9	3.12400	3.33536	2.10822	2.87986	3.70763	3.35054	3.07763	3.30501	3.46768	3.45309
10	4.61205	4.38336	4.80364	4.41503	2.93214	4.07331	4.37143	3.08452	2.77196	2.75603
11	3.12400	3.05589	2.22291	2.54612	3.06139	3.27826	3.07763	2.71704	2.54005	3.33691
12	5.05847	4.59297	4.40219	4.08130	4.41850	3.85648	2.68949	3.89298	5.09102	5.66045
13	2.08237	2.14762	2.50966	2.07889	2.09203	2.04956	1.39569	3.45200	0.76211	1.59426
14	2.23117	2.84629	2.96846	3.08010	2.41515	3.27826	3.98329	3.67249	3.85419	3.33691
15	1.73516	1.58869	1.19062	2.07889	2.28590	0.74858	2.94825	1.61460	1.99894	2.17514
16	1.83436	1.58869	2.79641	2.27913	1.12267	3.13371	2.04259	1.83509	2.15354	1.36190
17	2.37998	3.54496	2.33761	2.07889	2.41515	2.55550	1.78383	4.18696	3.08117	3.33691
18	3.07440	1.79829	3.94340	3.61407	0.99342	1.32680	2.30135	3.45200	2.54005	1.94279
19	4.06644	2.28736	4.57424	3.88106	3.90151	2.77233	4.63019	3.45200	3.46768	3.33691
20	2.82639	3.19563	2.62436	3.08010	4.09538	2.91688	2.68949	2.34956	2.69466	2.75603

<표5.1>의 결과를 이용하여 학생들에게 새로운 학점을 부여[ A+(2):A0(1):A-(1), B=2:2:2, C=2:2:2, D=2:1:1 ]한 뒤, 이를 수치로 환산한 수 순위를 구해보았다.<표5.2>

<표5.2 환산된 과목별 점수>

과목	과목 1	과목 2	과목 3	과목 4	과목 5	과목 6	과목 7	과목 8	과목 9	과목 10	sum	rank
1	3.3	3.3	2.0	2.5	2.3	3.5	3.0	3.3	1.3	2.0	26.5	10
2	1.0	2.5	2.5	1.3	3.0	2.3	1.5	1.0	2.5	2.3	19.9	16
3	4.3	4.5	4.5	4.5	4.5	4.5	4.0	2.5	4.3	4.5	42.1	1
4	2.5	3.5	3.5	2.5	3.5	3.5	4.5	4.5	2.3	4.5	34.8	4
5	3.5	2.3	3.0	3.5	4.0	3.3	3.5	3.5	2.3	3.5	32.4	7
6	3.5	2.5	3.3	4.5	3.3	4.5	3.0	1.3	3.0	4.3	33.2	6
7	2.0	4.3	2.0	2.0	2.5	1.5	1.3	4.3	2.3	3.0	25.2	13
8	1.5	1.0	1.3	1.0	1.5	2.0	1.5	2.0	2.0	1.5	15.3	19
9	3.0	3.3	1.5	2.5	3.3	3.3	3.0	2.5	3.3	4.0	29.7	8
10	4.5	4.0	4.5	4.3	2.5	4.3	4.0	2.3	2.5	2.3	35.2	3
11	3.0	3.0	1.5	2.3	3.0	3.0	3.0	2.3	2.0	3.0	26.1	12
12	4.5	4.5	4.0	3.5	4.5	3.5	2.3	4.0	4.5	4.5	39.8	2
13	2.0	2.0	2.3	1.5	1.5	1.5	1.0	3.0	1.0	1.3	17.1	18
14	2.0	2.3	3.3	3.0	2.0	3.0	3.5	3.3	3.5	3.0	28.9	9
15	1.3	1.3	1.0	1.5	2.0	1.0	2.5	1.5	1.5	1.5	15.1	20
16	1.5	1.3	3.0	2.3	1.3	2.5	1.5	1.5	1.5	1.0	17.4	17
17	2.3	3.5	2.0	1.5	2.0	2.0	1.3	4.5	3.0	3.0	25.1	14
18	3.0	1.5	3.5	3.3	1.0	1.3	2.3	3.0	2.0	1.5	22.4	15
19	4.0	2.0	4.3	3.5	3.5	2.3	4.5	3.0	3.3	3.0	33.4	5
20	2.5	3.0	2.5	3.0	4.0	2.5	2.3	2.0	2.3	2.3	26.4	11

<표5.2>에서 구한 순위와 기존의 GPA순위의 비교 <표5.3>는 '4.각과목들의 순위를 이용한 방법'과 어느 정도 비슷한 경향을 보이고는 있지만, 변화를 보이는 학생의 수가 줄어들었음을 알 수 있다. 즉, 학점부여과정에서 발생하는 가산점의 문제때문으로 여겨진다.(예, B0 ⇨ B+ : 0.3 가산, B+ ⇨ A- : 0.5 가산)

GPA 에 의해 과대평가된 학생 : 학생= 5, 12, 13, 17 ( 4명)

GPA 에 의해 과소평가된 학생 : 학생= 3, 6, 16 ( 3명)

<표5.3 GPA와의 비교>

학생	기존 GPA점수	기존 GPA순위	표준화 GPA순위	표준화 GPA점수
1	3.31667	10	10	2.65
2	2.88667	16	16	1.99
3	4.11500	2	1	4.21
4	3.75667	4	4	3.48
5	3.65167	6	7	3.24
6	3.53000	7	6	3.32
7	3.17167	14	13	2.52
8	2.58667	19	19	1.53
9	3.42000	8	8	2.97
10	3.83667	3	3	3.52
11	3.23833	12	12	2.61
12	4.16000	1	2	3.98
13	2.74500	17	18	1.71
14	3.39833	9	9	2.89
15	2.56500	20	20	1.51
16	2.74000	18	17	1.74
17	3.19667	13	14	2.51
18	3.04333	15	15	2.24
19	3.66167	5	5	3.34
20	3.26500	11	11	2.64

## 6. 결 론

지금까지 GPA방식의 순위결정에 대한 문제점을 해결해보려는 의도에서 성적자료를 이용하여, 인자분석, 비모수적 방법, 그리고 기존의 GPA를 응용하는 방법들에 의한 새로운 순위를 만들어 보는 과정을 보았다. 조사된 자료에 대하여 각 방법들을 실시하여 분석한 결과를 요약해 보면 다음과 같다.

먼저 인자분석을 이용한 방법의 결과를 살펴보면,

1. 세집단(전공과목, 교양과목, 계열기초)으로 나누어진 자료에 대하여, 집단별로 각각 인자분석을 실시한 결과, 집단별로 하나의 인자를 보유할 수 있었고, 보유한 인자들은 해당집단에 속해있는 과목(변수로 사용된)의 성격을 대표할 수 있는 것으로 볼 수 있기에, 이들의 의미는 각 집단의 과목들에 대한 '학력'으로 줄 수 있었다.

2. 각 집단별로, 앞서 보유한 인자를 이용하여, 회귀적인 방법을 통하여 표준화된 인자점수들을 계산한 후, 집단별로 이 인자점수의 크기순으로 순위를 정하고, 기존의 GPA방식에 의해 결정된 순위와 비교하였다. 그 결과로 전공 과목집단의 인자가 GPA와 깊은 연관성을 가지고 있음을 알 수 있었다.

3. 가중치를 고려한 인자점수의 순위를 위해, 세집단에서 보유한 인자들의 점수(비표준화)를 분산분석을 통해서 각 집단간의 차이를 찾을 수 있었고, 이 차이를 이용하여 세집단에 가중치를 적용할 수 있었다.

이 가중치를 고려한 인자점수의 순위와 GPA순위를 비교한 결과, 20명의 학생중에서 9명의 경우에 순위가 틀려지는 것을 볼 수 있었다. 그중 5명이 인자점수의 순위보다 GPA순위가 더 높은 경우인데, 이들은 분산이 큰 과목에서 좋은 점수를 받았을 것이라는 추측할 수 있다.

두번째로 각 과목의 순위를 이용한 방법에서는

결과적으로 기존의 GPA방식에 의한 순위와 비교했을 때, 10명의 학생들의 순위가 달라지는 것

을 볼 수 있었다. 그러나, 이 결과와 앞서의 인자분석의 결과와 비교하는 것은 옳지 않다(각기 사용목적에 따라 자료가 다르게 사용되었고, GPA순위도 다르다). 어쨌든 가장 대표적인 '순위바뀔'의 예인 1등과 2등을 보더라도 과목간의 배점차이에서 발생하는 문제로, 과목간의 순위와는 관계없이 전체순위가 뒤바뀌는 것이 현 GPA의 문제점이라는 것을 다시 한번 확인할 수 있었다.

마지막으로 기존의 GPA를 응용한 표준화 GPA방법은, 앞서의 두 방법이 기존의 GPA와 전혀 별개의 방법으로 진행된 반면, 일부 GPA의 방법을 수용함으로써, GPA의 가장 큰 문제점인 과목점수들의 분산차이에만 중점을 둔 것이다. 결과 역시 격차는 크지 않으나, 기존 GPA순위와는 많은(8명) 차이를 보이고 있다.

이상으로 요약되는 분석을 마치면서, 하나의 질문을 제기해 본다. 과연 이 새로운 방법들에 의해서 만들어지는 순위가, GPA순위보다 더 정확한 결과라 말할수 있겠는가? 먼저 인자분석의 경우를 생각해 보자. 이 과정의 전공과목집단을 예로 들자면, '전공학력'은 8개의 과목을 통해서 얻어진 것인데, 차원축소를 행한후 1개로 축소가 가능했다면 그 인자에 대한 현실적인 의미를 고려해서 새로운 평가척도를 만들어 그것으로 대체해도 될 것인가를 생각하게 된다. 진짜 문제는 "만약 분산비가 80%정도면 감수하겠는데 60%정도로는 감수하기 힘들다?"

인자분석의 결과에 있어서 뭘 수 없는 문제가 바로 이런 해석상의 문제이다. 따라서, 이 방법은 단지 GPA방법의 문제점을 해결해 보고자 시도한, 분산을 고려할 수 있는 순위결정방법의 '체식'일 뿐이지, 그 이상의 의미를 두는 것은 곤란하다고 생각된다.

그러나, 그 외의 두가지 방법(비모수적, 표준화GPA)은 어느 정도 타당성을 갖추고 있다고 여겨진다. 분석과정에 있어 원래 과목별 성적이 없는 관계로 발생된 불가피한 문제가 없진 않았으나, 만일 원래 점수자료에 이용한다면, 기존의 GPA방식에 의한 순위보다는 더욱 설득력있는 순위가 되지않을까 한다.

순위를 주는 사람이나 그것을 받는 사람이나, 그들이 그 과정을 인정하건 않하건, 학생시절의 순위는 개인의 학교생활을 평가하는 객관성있는 줄자로 사용되고 있음(사실 그외에는 존재하는 것도 없지만...)은 부정할 수 없다. 이런 상황에서 현재 사용되고 있는 GPA가 좀더 설득력있는 순위매김으로 인정받으려면, 많은 수정과 보완이 필요하지 않은가 싶다.

끝으로, 이 연구에서 특히 아쉬웠던 부분은 주로 자료에 관한 것이었다. 공개된 자료가 아니었으므로, 개개인의 성적들을 일일이 찾아서 정리하고, 분석가능한 형태로 자료를 가공하는 과정에서 많은 정보의 손실이 발생했으며, 무엇보다 분석에 사용된 각 과목의 원래점수를 구할 수 없어 학점을 숫자로 환산함으로써, 과목들의 정확한 평균과 편차를 이용하지 못했다는 점은, 자료정리에 투자한 시간을 생각할 때 끝내 아쉬움으로 남는다. 그밖에 발생하는 통계학상의 문제에 대한 처리 역시 미숙한 상태에서 실시한 분석이므로, 결과가 본래 목적했던 바에 어느정도 접근했다는 것만으로 만족해야만 했다.

### < 참고문헌 >

- [1] 김기영 · 전명식( 1990 ). 「SAS 인자분석」, 자유아카데미
- [2] 성내경 ( 1991 ). 「SAS 분산분석」, 자유아카데미
- [3] 최병현 ( 1991 ). 「PC SAS 입문」, 박 영 사

## 두 모집단의 모평균 비교시 발생하는 문제점

배 은 경

(지도교수 : 권세혁)

한남대학교 경상대학 응용통계학과

### 요 약

우리가 일반적으로 두 모집단 비교시 사용하는 검정방법은 모집단을 정규분포로 가정하고 만들어졌다. 가정이 깨어진 경우 발생할 수 있는 문제를 인식하기위해 평균과 분산이 동일하게 고정시키고 가정이 만족된 경우(모집단 정규분포사용)과 가정이 깨어진 경우(모집단으로 카이분포 사용)에 대하여 두 모집단의 모평균의 동일성 검정을 실시하고 이를 비교하여 보았다.

### 1. 서 론

한 제약회사에서, 새로운 진통제를 개발한 연구원들이 기존의 진통제보다 새로운 진통제의 진통효과가 더 빠르게 나타난다고 주장하고 있다. 회사 경영진은 이를 확인하기 위해 100명의 환자를 랜덤추출하여 각각 50명씩 기존의 진통제와 새로운 진통제를 복용시킨 후 진통효과가 나타나는 시간의 평균  $\bar{X}$ 를 관찰, 비교하여 연구원들의 주장을 판단하기로 하였다. 이로써 새로운 진통제의 진통효과가 더 빠르다는게 사실로 인정되어지면, 생산공정을 변경하여 새로운 진통제를 생산하려고 한다. 이때 어떠한 방법으로 결정을 내려야하는가?

기존의 진통제와 새로운 진통제에 대해서 전수 조사가 가능하다면 어떤 오류도 없이 정확하고 올바른 결정을 내릴 수 있을 것이다. 그러나 전수 조사는 현실적으로 불가능한 얘기다. 우리가 할 수 있는 것은 표본을 랜덤하게 추출하여 기존 진통제의 모집단과 새로운 진통제의 모집단을 추측하고 이를 바탕으로 두 모집단을 비교 검정하는 것이다. 그러나, 무한대에 가까운 모집단을 수소의 표본만을 가지고 파악한다는 것에는 우리가 따른다.

우리가 알고자 하는 모수  $\mu$ 가 미지수이므로 두 가설, 귀무가설과 대립가설 중 어느 하나가 옳은가를 오류없이 판단할 수 없다. 즉 모수를 표본에 의존하여 검정, 판단하다보면 두가지 오류가 발생하게 된다. 여기서 말하는 두가지 오류란 제1종오류(귀무가설이 참인데도 귀무가설을 기각하는 오류)와 제2종 오류(귀무가설이 거짓인데 귀무가설을 채택하는 오류)를 말하는 것이다.

두 모집단의 모평균을 비교하는 방법을 생각해 보자. 우리는 일반적으로 귀무가설과 대립가설을 정하고 랜덤하게 추출한 표본으로부터 검정통계량  $\bar{x} - \bar{y}$ 을 구한다. 이를 바탕으로 기각역을 정하고 검정통계량과 비교하거나 검정통계량의 유의확률을 계산하여 유의수준과 비교함으로써 귀무가설의 기각여부를 결정한다. 여기서 간과하고 넘어가기 쉬운 사항이 있는데, 이는 기각역이나 유

의확률을 계산하기 위해서는 검정통계량의 분포를 알아야 한다는 사실이다.

검정통계량의 분포를 알기 위해서는 모집단의 분포를 알아야 한다. 그러나 대부분 경우 모집단의 분포를 알 수 없으므로 모집단이 정규분포라는 가정을 하거나 표본의 크기가 어느 정도 클 경우 중심극한 정리에 의해 검정통계량의 분포가 정규분포를 따른다는 사실을 바탕으로 모집단이 정규분포라고 가정하고 검정 이론을 만들었으며 우리는 이를 기초로 검정을 실시한다. 그러나 실제로 사용되는 자료 중에서 어느 정도의 자료나 가정을 만족하는 정규분포를 따를까? 만약 모평균을 비교하는 두 모집단이 정규분포와는 동떨어진 치우친 분포라면 모집단이 정규분포를 따른다는 가정하에서 만들어진 이론에 그대로 적용해서 얻어진 결과를 어느 정도나 믿을 수 있나? 또한 모집단이 정규분포를 따르지 않으나 표본의 크기가 어느 정도 크므로 가정이 만족된 경우와 동일한 결과가 나올 것이라고 생각할 수 있나? 모집단이 어떤 분포를 따르느냐에 따라 검정 결과가 차이가 날 것이라고 생각되어진다. 이를 확인하는 것이 목적의 하나이다.

우리는 모집단에 관한 가정이 만족되었는지 깨어졌는지를 고려하지 않고 무턱대고 검정을 실시하여 성급한 판단을 내리고 쉽게 믿는 경우가 허다하다. 모의 실험을 통해 가정이 만족된 경우와 깨어진 경우의 결과를 비교해보고 그 차이를 실험으로 원인을 파악해 보는 동시에 성급한 판단으로 인한 위험을 조금이라도 인식하고 가능하다면 해결방법을 생각해 보는 것이 이 논문의 목적이다.

## 2. 시뮬레이션을 통한 분석

분석에서 사용하는 자료는 simulation 기법, 즉 분석을 위해 가정한 모집단에 대하여 난수함수를 이용 랜덤 추출한 것을 사용한다. 여기에서 사용되는 분포는 정규분포와 카이 분포인데 sas에서 제공되어지는 난수함수 rannor과 rangam을 사용하여 가정한 분포에서 난수를 발생시켰다. 추출하는 표본의 크기는 10과 30으로 정했다. 이는 일반적으로 중심극한 정리가 적용되는 표본의 크기를 25이상인 경우로 봄으로 가정이 깨어진 경우와 만족된 경우에 대해서 중심극한 정리가 적용되는 경우와 만족되지 않은 경우를 비교하기 위하여 표본의 크기를 10과 30으로 정한 것이다.

모든 경우에 대해서 분석을 할 수 없으므로 몇가지 경우의 분포에 대하여 다음의 검정방법을 이용 분석하고자 한다.

이론적으로 두 모집단이 정규분포를 따르고 모분산은 모르지만 같을 경우 검정통계량  $(\bar{X} - \bar{Y})$  은 다음과 같이 표준화를 하였을때 T분포를 따른다.

$\sigma_1^2$ 과  $\sigma_2^2$ 가 같다고 가정할 때

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

$$\text{여기서 } S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad \text{이다}$$

랜덤추출된 자료를 이용, 위의 분포를 바탕으로 검정통계량의 유의확률을 계산하고 이를 유의수준과 비교하여 귀무가설의 기각여부를 판정하는 방법을 분석에 이용하였다.

## 2.1 기각률과 채택률 비교 분석

가정이 깨어진 경우의 기준 분포를 자유도가 2인  $\chi^2$  분포로 정하고 1종 오류를 범할 확률과 검출력을 살피기 위해 두 모평균의 차이가 0인 경우와 0.5인 경우의 두가지를 중심으로 모의실험을 실시하였다. 이와 가정이 만족된 경우의 모의실험을 실시 하여 나온 결과를 비교하여, 만약 같은 결과가 나온다면 문제가 없지만 아니라며 가정의 파괴를 염두에 두지 않은 기존의 검정법을 그대로 사용하여 결론을 내리기에 문제가 있다고 생각된다. 어떤점에서 차이가 있는지를 알고자 아래의 모집단을 사용하여 모의실험을 하고 가정이 만족된 경우와 깨어진 경우의 검정 결과를 비교해 보았다. 표본의 크기는 10과 30으로 정하고 20번을 반복하여 실험했다.

### 2.1.1 두 모평균 차이가 없을 경우

- ┌ 가정이 만족된 경우: 두 모집단  $X \sim N(2,4)$ 와  $Y \sim N(2,4)$ 의 모평균에 관한 검정
- └ 가정이 깨어진 경우: 두 모집단  $X \sim \chi^2(2)$ 와  $Y \sim \chi^2(2)$ 의 모평균에 관한 검정

<표 1> 20번 반복 모의실험시 기각횟수

		가정이 만족된 경우	가정이 깨어진 경우
n = 10	H0 기각	1 ( 5%)	0 ( 0%)
	채택	19 ( 95%)	20 ( 100%)
n = 30	기각	1 ( 5%)	3 ( 15%)
	채택	19 ( 95%)	17 ( 85%)

유의수준 0.05에서 두 모집단의 모평균의 동일성 여부를 검정해 본 결과 가정이 만족된 경우는 표본의 크기가 10인 경우와 30인 경우가 동일하게 5%의 기각률을 보였다. 귀무가설이 참일때 귀무가설이 기각될 확률의 최대치를 5%로 정하고 검정을 실시하였으므로 이와 같은 결과는 이론적으로 당연한 결과이다. 가정이 깨어진 경우를 생각해 보면 표본의 크기가 작은 경우에는 1종오류의 위험률이 0%이다. 이것만 봐서는 위험률이 5%인 가정이 만족된 경우보다는 1종 오류의 위험률 적으므로 가정이 깨진 경우가 더 좋은 결과를 보이는 것처럼 보인다. 이론적으로 생각하면 표본이 커짐으로써 모집단은 치우친 분포를 따르지만 검정통계량은 정규분포에 근사해감으로 가정이 만족된 경우와 유사하게 검정결과가 나와야 할 것 같다. 그러나 표본이 큰 경우는 1종 오류의 위험률이 15%로 증가한다. 이는 표본의 수가 커짐으로써 해서 추출되는 표본이 모집단의 치우친 특성을 나타내는 속도가 정규분포로 근사해 가는 속도보다 빠르기 때문이라고 생각되어진다. 즉 분석에 사용한 카이 분포가 정규분포에 비해서 중심에 더 많이 모여 있고 오른쪽으로 이산치라 할 만한 큰 값을 가진다는 특성때문에 이와 같은 결과가 나왔다고 본다.

〈표 2〉 20번 반복 모의실험시 최대, 최소 유의확률

		가정이 만족된 경우	가정이 깨어진 경우
n = 10	최소값	0.0107	0.0749
	최대값	0.9660	0.9760
n = 30	최소값	0.0237	0.0129
	최대값	0.9747	0.9782

가정이 만족된 경우, 표본의 크기가 커짐으로 유의확률의 값이 더 커짐을 알 수 있다. 이는 표본의 크기가 커짐으로 해서 마치 분산이 작아지는 효과를 가져오고 이로해서 검정통계량을 중심 쪽으로 끌어당겨지는 것과 같은 결과를 가져온 것이다. 그러나 이와 같은 특징도 가정이 깨어진 경우에는 나타나지 않는다. 성급하게 판단하기가 어렵다.

2.1.2) 두 모평균의 차이가 0.5인 경우

- ┌ 가정이 만족된 경우: 두 모집단  $X \sim N(2,4)$ 와  $Y \sim N(2.5,5)$ 의 모평균에 관한 검정
- └ 가정이 깨어진 경우: 두 모집단  $X \sim \chi^2(2)$ 와  $Y \sim \chi^2(2.5)$ 의 모평균에 관한 검정

〈표 3〉 20번 반복 모의실험시 기각횟수

		가정이 만족된 경우	가정이 깨어진 경우
n = 10	H0 기각	4 (20%)	6 (30%)
	채택	16 (80%)	14 (70%)
n = 30	기각	3 (15%)	6 (30%)
	채택	17 (85%)	14 (70%)

가정이 만족된 경우의 검출력보다 가정이 깨어진 경우의 검출력이 좋다. 그러나 만약에 0.5의 차이를 실제의 차이로 인정할 수 없다면 귀무가설을 되도록 기각하고 싶지 않을 것이다. 가정이 만족된 경우의 기각률을 최대 위험수준으로 감수하고 나머지 위험들은 대비를 안했다고 생각한다면 모집단의 가정이 깨어진 경우는 크게 손해를 볼 수 있다.

〈표 4〉 20번 반복 모의실험시 최대, 최소유의확률

		가정이 만족된 경우	가정이 깨어진 경우
n = 10	최소값	0.0251	0.0161
	최대값	0.9643	0.9095
n = 30	최소값	0.0013	0.0042
	최대값	0.9853	0.8675

유의 확률의 최소값의 경우 모평균의 차이가 없을 때에는 표본의 크기가 커짐으로써 유의확률의 값이 더 커졌는데 반하여 모평균의 차이가 0.5가 생기니까 유의확률의 값이 더 작아졌다. 이는 다음의 사실을 말해 준다. 두 모집단이 정규분포를 따르는 경우는 표본의 크기가 커질때 더 정확

한 결과로 검정결과를 유도한다. 즉 표본의 크기가 큰 경우 모집단의 평균의 차이가 없을 경우는 유의확률이 표본이 작은 경우에 비해 크게 나오고 차이가 있을 경우는 더 작은 유의확률을 나타낸다. 그러나 이러한 특징을 가정이 깨어진 경우에 적용하기가 어렵다. 즉 가정이 만족된 경우는 이론적인 근거에 따라 그 결과를 예측할 수 있지만 가정이 깨어진 경우는 불가능하다.

이 분석에는 다음과 같은 문제점이 있다. 우리가 모집단의 설정한 분포의 분산이 너무 커서 기각되어지는 수가 작아지는 것이 아닌가 하는 것과 두 모집단이 동일한 분포를 따름으로 분명한 차이를 확인할 수 없는 것이 아닌가 하는 문제와 기각률이나 채택률은 확률의 개념이다. 20번 반복하고 비교한다는 것에는 무리가 따를 수 있다. 이러한 문제들을 고려하여 다음의 분석을 실시하였다.

## 2.2 P-값에 의한 분석

확률은 반복 실험을 하여 수렴하는 값이다. 20번 반복 실험하는 것 만으로는 가정이 만족된 경우와 깨어진 경우의 기각률과 채택률을 비교한다는 것에는 무리가 따를 수 있다. 좀 더 정확한 분석을 위해 반복수를 늘렸다.

모평균을 비교하는 두 모집단이 서로 다를때(가정이 깨어진 경우), 즉 하나의 모집단은 정규분포를 따르고 다른 하나의 모집단은 카이 분포를 따를 경우와 두 모집단이 정규분포를 따르는 경우(가정이 만족되는 경우)의 유의확률을 각각 계산하여 유의확률이 0.05를 간격으로 빈도수를 계산하여 이를 겹쳐서 plot을 그려봄으로 가정이 만족된 경우와 깨어진 경우에 대해서 비교해보고자 한다

### 2.2.1 모집단의 평균이 차이가 없을 때

두 모집단이 아래에 제시한 분포일때에 표본을 10으로 하여 랜덤추출하고 이 표본을 가지고 두 모평균의 동일성 여부를 검정한다. 모의실험으로 얻어진 유의확률을 data로 받아들여 이를 plot으로 그려 비교하고자 했다. 표본을 10개씩 뽑아 유의확률을 계산하는 작업을 1000번 반복하여 가정이 만족된 경우와 깨어진 경우의 dots에 관한 유의확률을 겹쳐그림으로 그 차이를 살펴려 하였다.

#### ① 표본의 크기가 10인 경우

① 분포에 관한 가정이 만족된 경우  
(굵은 선)

┌모집단 1: 평균 1, 분산 2인 정규분포  
└모집단 2: 평균 1, 분산 2인 정규분포

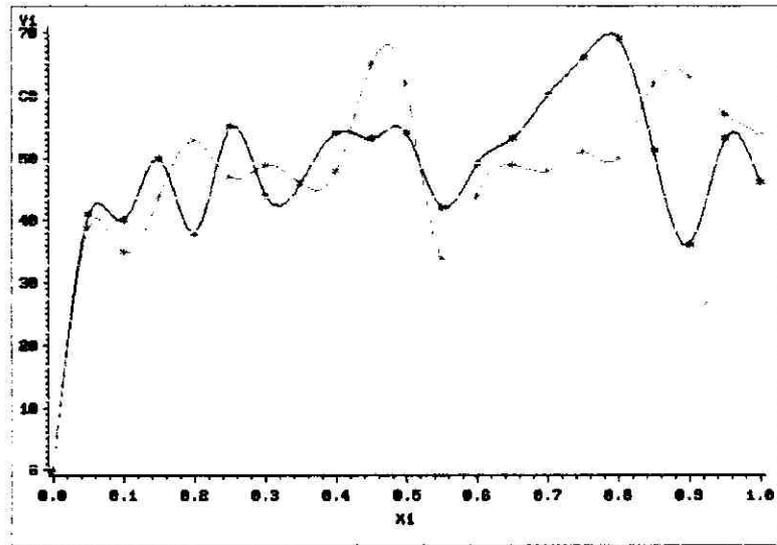
② 분포에 관한 가정이 깨어진 경우  
(가는 선)

┌모집단 1: 평균 1, 분산 2인 정규분포  
└모집단 2: 자유도가 1인 카이 분포

각 모의실험결과 얻어진 유의확률을 계산하여 전체 범위(0에서 1까지)를 0.05씩 20구간으로 나누어 구간내에 있는 유의확률의 빈도수를 y축으로 구간의 상한 0.05, 0.10, ..... ,0.95, 1을 x축으로 하여 plot을 그렸다. 굵은 선은 가정이 만족한 경우이며 가는 선은 가정이 깨어진 경우이다.

<그림 2.1>을 살펴보면 유의확률이 0.05, 0.10, 0.15보다 작은 경우에는 가정이 만족된 경우에 비해 가정이 깨어진 경우가 더 적은 빈도수를 나타낸다. 유의수준 0.05와 0.10에서 유의수준보다

작게 나타날 유의확률의 개수가, 즉 귀무가설이 기각될 확률이 가정이 만족된 경우에 비해 가정이 깨어진 경우가 더 작다는 의미가 되고, 이는 평균의 차이가 없는 경우 모의실험 결과 가정이 깨어진 경우가 더 작은 빈도로 오류를 범한다는 의미로 이 한가지 사실만을 살펴보면 가정이 깨어진 경우가 가정을 만족한 경우에 비해 더 사실에 접근해 있다고 생각할 수 있다. 이와 같은 결과는 두 번째 모집단에 포함된 카이 분포의 특성(정규분포에 비해서 대부분의 자료가 중앙값을 중심으로 모여있는 특성)때문이라고 생각되어 진다. 그러면 치우친 꼬리에서 표본의 추출시 이산점으로 불린 만큼 큰 값이 나올수 있다는 사실은 그냥 무시해도 되나? 아니다. 꼬리 부분의 자료는 뽑힐 확률이 매우 작기 때문에 10개의 표본을 뽑을시에는 그 표본에 포함될 확률이 무시해도될만큼 작아진다. 그러나 표본의 크기가 커진 경우에는 이를 그냥 무시할 수 없다. 표본의 크기가 커짐으로 꼬리 부분에서 뽑힐 확률도 늘어나기 때문이다. 표본의 크기가 30인 경우를 살펴보자.



<그림 2.1> 모평균의 차이가 없는 경우 p값 산점도(n=10)

② 표본의 크기가 30인 경우

㉠ 분포에 관한 가정이 만족된 경우  
(굵은 선)

┌모집단 1: 평균 1, 분산 2인 정규분포  
└모집단 2: 평균 1, 분산 2인 정규분포

㉡ 분포에 관한 가정이 깨어진 경우  
(가는 선)

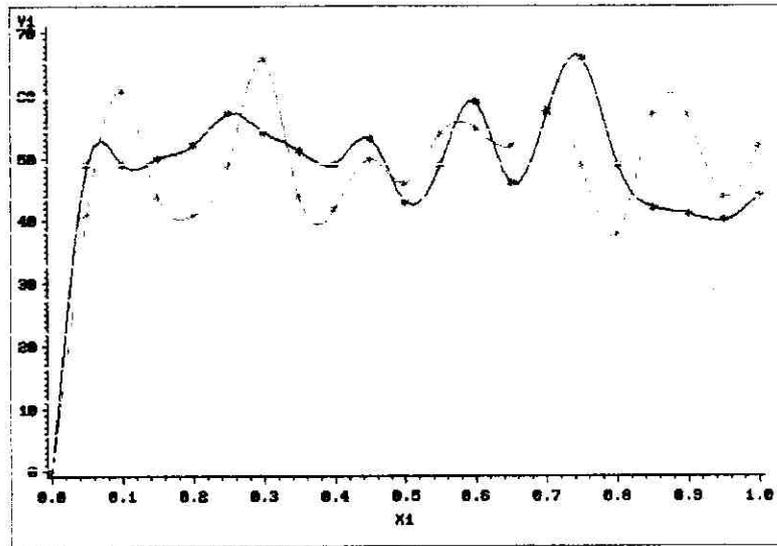
┌모집단 1: 평균 1, 분산 2인 정규분포  
└모집단 2: 자유도가 1인 카이 분포

<그림 2.2>를 살펴보면 표본의 크기가 10인 경우와 동일하게 유의확률이 0.05미만인 부분에서 가정이 깨어진 경우의 빈도수가 작으나 0.05에서 0.10미만에서는 가정이 깨어진 경우의 빈도수가 큰 차이로 증가한다.

표본의 크기가 커지면 일반적으로 정규근사를 생각할 수 있는데도 불구하고 가정이 깨어진 경우와 만족된 경우의 plot에는 상당한 차이가 있는 것같이 보여진다. 이는 표본의 크기가 커짐으로 모집단 카이 분포의 치우침의 특성을 가지는(이산점이 있는) 표본이 추출될 확률이 높아지는데 반하여 정규근사하는 속도가 느리기 때문이라고 보여진다.

그러나 아쉬운 것은 가정이 만족된 경우에서도 유의확률과 빈도수 사이에서 어떤 패턴도 발견

할 수 없다는 것이다. 이 점에 대해서는 기회가 있으면 더 연구가 필요하다고 느꼈다.



<그림 2.2> 모평균의 차이가 없는 경우 p값 산점도(n=30)

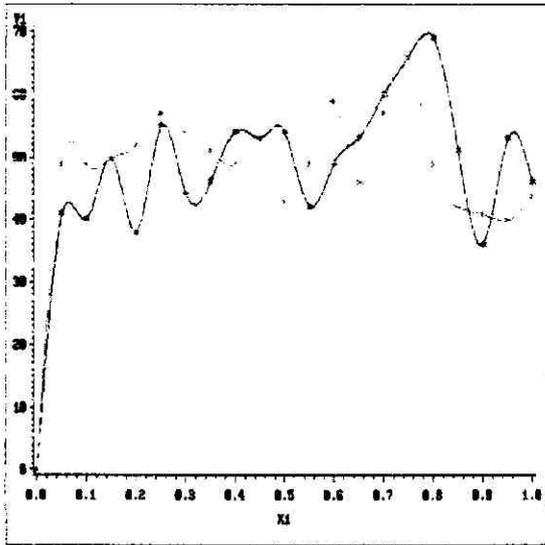
③ 참고

표본의 크기가 10인 경우와 표본의 크기가 30인 경우에 대하여 두 모집단이 정규분포를 따르는 경우의 유의확률의 변화와 그렇지 않은 경우의 유의확률의 변화를 살펴보고자 하였다.

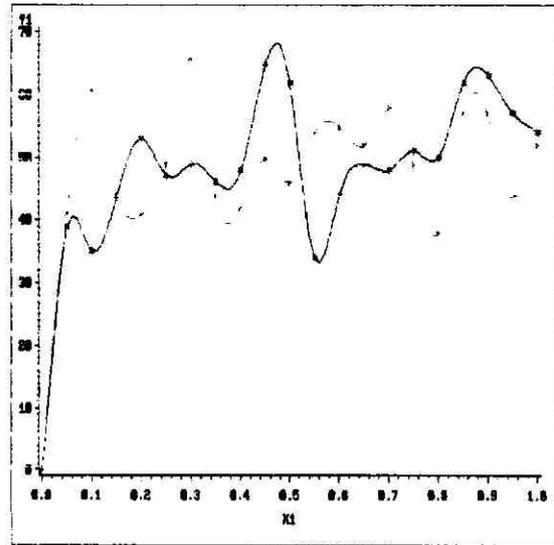
- |      |   |                         |   |                         |
|------|---|-------------------------|---|-------------------------|
| n=10 | ① | 모집단 1: 평균 1, 분산 2인 정규분포 | ① | 모집단 1: 평균 1, 분산 2인 정규분포 |
|      |   | 모집단 2: 평균 1, 분산 2인 정규분포 | ② | 모집단 2: 자유도가 1인 카이 분포    |
| n=30 | ① | 모집단 1: 평균 1, 분산 2인 정규분포 | ① | 모집단 1: 평균 1, 분산 2인 정규분포 |
|      |   | 모집단 2: 평균 1, 분산 2인 정규분포 | ② | 모집단 2: 자유도가 1인 카이 분포    |

아래의 <그림2.3>과 <그림2.4>를 비교해 살펴보았다. 이론적으로 표본의 크기가 커지면 귀무가설을 기각할 확률이 높아진다고 말한다. 반복수가 20인 모의실험에서는 이러한 특징이 명확하게 나타나지 않았는데 반해서 반복수를 늘린 경우에는 확률 0.10미만에서 유의확률이 표본의 크기가 10인 경우에 비해 30인 경우의 빈도수가 높다. 이는 유의수준 0.05, 0.10에서 표본의 크기가 30일 경우가 귀무가설이 기각될 확률이 높다는 의미이다.

가정이 만족된 경우와 깨어진 경우를 비교해 보면 가정이 만족된 경우에 비해서 깨어진 경우의 표본에 따른 빈도수의 차이가 더 크고 불규칙해 보인다.



<그림 2.3> 두분포-정규분포의 경우



<그림 2.4> 서로 다른 분포의 경우

### 2.2.2 두 모집단의 평균이 차이가 0.5일 경우

두 모집단의 평균 차이가 0.5인 경우 카이 분포의 중앙값의 위치를 생각하여 검정결과를 분석해야 한다. 모집단1이 카이 분포의 오른쪽으로 0.5만큼 이동했다면 카이 분포의 중앙값(표본의 추출 시 확률이 가장 높은 곳)과 정규분포의 중앙값(평균과 동일하다.)과의 거리가 커지는데 반해서 모집단1이 카이분포의 왼쪽으로 0.5만큼 이동한다면 정규분포의 중앙값과 카이 분포의 중앙값은 평균의 차이가 없을 때보다 더 가깝게 위치하게 된다. 이는 검정 결과에 크게 영향을 미칠 것으로 보여진다. 이를 모의실험을 통해 비교해 보자.

#### (1) 카이 분포 오른쪽으로 0.5이동한 경우

##### ① 표본의 크기가 10인 경우

㉠ 분포에 관한 가정이 만족된 경우  
(굵은 선)

㉡ 분포에 관한 가정이 깨어진 경우  
(가는 선)

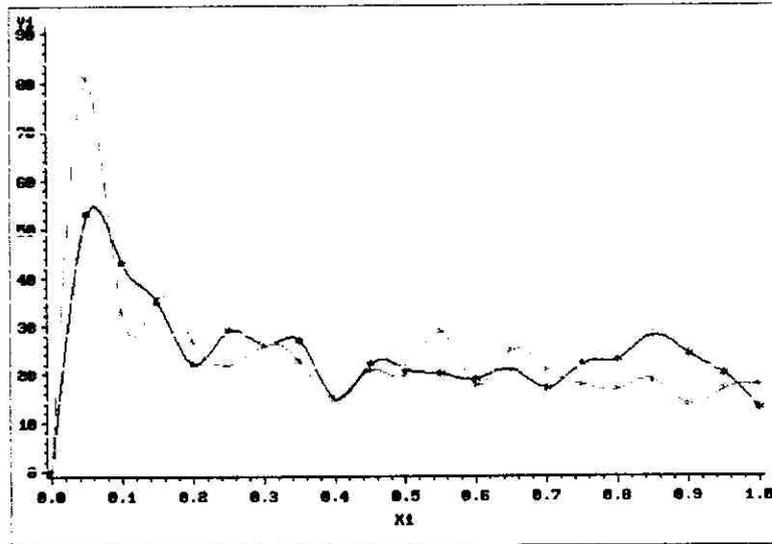
┌모집단 1: 평균 1.5, 분산 2인 정규분포

└모집단 2: 평균 1, 분산 2인 정규분포

┌모집단 1: 평균 1.5, 분산 2인 정규분포

└모집단 2: 자유도가 1인 카이 분포

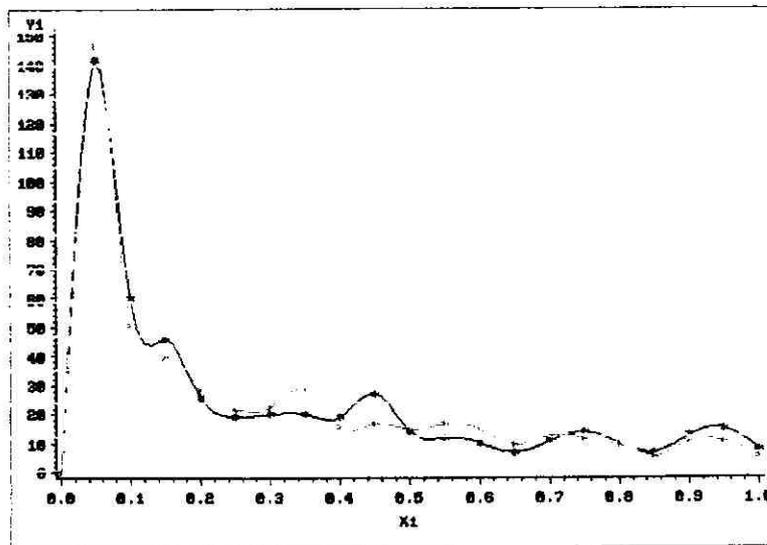
<그림2.5> 을 살펴보면, 가정이 만족된 경우에 비해 가정이 깨어진 경우에 유의수준 0.05에서 기각률이 높다. 이는 카이 분포에서 중앙값을 중심으로 표본이 뽑힐 경우에 정규분포의 중심으로 부터 뽑힌 표본과의 차이가 가정이 만족 경우에 비해 크기 때문이라고 생각되어진다.



<그림 2.5> 모평균의 차이가 0.5인 경우 p값 산점도(n=10)

② 표본의 크기가 30인 경우

<그림2.6>을 보면, 다른 산점도에 비해 모집단이 카이 분포의 오른쪽으로 0.5이동하고 표본의 크기가 30인 경우 가정이 만족된 경우와 깨어진 경우 비슷한 패턴을 보인다. 이는 표본이 작은 경우에 비해 중심뿐만 아니라 분포의 전반에 걸쳐 표본이 뽑힐 확률이 커지고 중심극한의 정리에 의해 정규분포로 검정통계량이 근사해 감으로 가정이 만족된 경우와 비슷한 결과가 나오는 것으로 보여진다. 그리고 평균의 차이가 없는 경우와는 달리 차이가 있는 경우의 유의확률의 발생 빈도수가 패턴을 보인다는 사실을 하나더 알 수 있었다.



<그림 2.6> 모평균의 차이가 0.5인 경우 (n=30)

(2)카이 분포 왼쪽으로 0.5 이동한 경우

표본의 크기가 10인 경우에 가정이 만족된 경우와 모집단이 카이 분포의 왼쪽으로 이동한 경

우의 산점도를 겹쳐그려서 비교하고자한다<그림7>. 모집단1이 왼쪽으로 이동한 경우와 오른쪽으로 이동한 경우에 차이를 <그림8>을 통해 비교해 보자.

① 표본의 크기가 10인 경우

㉠ 분포에 관한 가정이 만족된 경우  
(굵은 선)

┌모집단 1: 평균 0.5, 분산 2인 정규분포  
└모집단 2: 평균 1, 분산 2인 정규분포

㉡ 분포에 관한 가정이 깨어진 경우  
(가는 선)

┌모집단 1: 평균 0.5, 분산 2인 정규분포  
└모집단 2: 자유도가 1인 카이 분포

② 참고

모집단이 오른쪽으로 0.5이동한 경우와

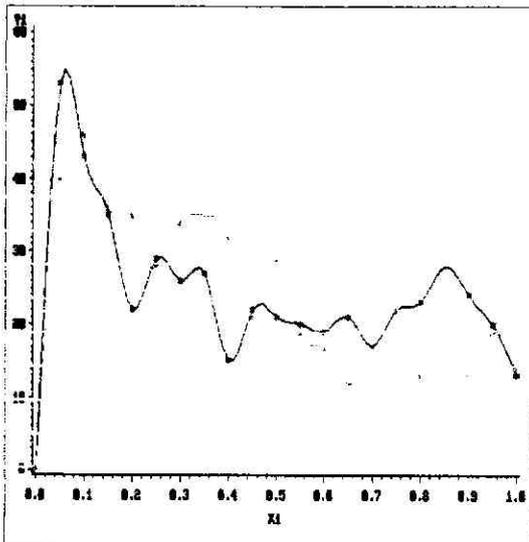
㉠ 분포에 관한 가정이 만족된 경우  
(굵은 선)

┌모집단 1: 평균 1.5, 분산 2인 정규분포  
└모집단 2: 자유도가 1인 카이 분포

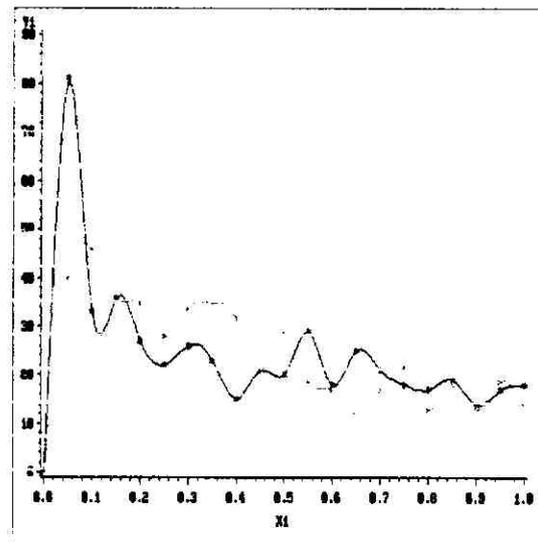
㉡ 분포에 관한 가정이 깨어진 경우  
(가는 선)

┌모집단 1: 평균 0.5, 분산 2인 정규분포  
└모집단 2: 자유도가 1인 카이 분포

오른쪽으로 이동한 경우에 반하여 왼쪽으로 이동한 경우는 가정이 깨어진 경우에 비해 더 작은 빈도수를 보인다. 실제로 오른쪽으로 이동한 경우와 왼쪽으로 이동한 경우가 상당한 차이가 있음을 확인할 수 있다.



<그림 2.7> 표본의 크기 10



<그림 2.8> 참고

### 3. 결 론

표본을 이용하여 모수를 판단하다보면 모집단에 관한 가정이 필요하다. 일반적으로 모집단을 정규분포라고 생각하고 검정을 실시하지만 실제로 정규분포를 따르는 자료는 흔하지 않다. 그러나 우리는 이런 가정은 생각하지 않고 무턱대고 검정을 실시하고 쉽게 그 결과를 믿고 판단을 내린다. 여기서는 실제로 가정이 만족된 경우와 깨어진 경우를 모의 실험을 통해 비교하고 거기에서 생길 수 있는 차이를 확인하고 그 원인을 찾아 가능하다면 가정이 깨진 경우에 생길 수 있는 위험률을 최소화하는 방법을 찾고자하는 것이 목적이었다.

앞에서의 분석 내용을 간추려 보면 검정의 결과는 표본이 어떻게 추출되는냐에 의해 크게 영향을 받는다. 그리고 이 표본 추출은 모집단의 분포의 특성에 영향을 받는다. 평균의 차이가 없는 경우를 살펴보면 표본의 크기가 작은 경우에는 확률이 높은 중앙값을 중심으로 표본이 추출되므로 분산과는 상관없이 중앙값 가까이에 어느 정도의 확률을 가지고 밀집되어 있느냐에 따라서 모집단이 정규분포를 따르는 경우보다 가정이 깨어진 경우의 1종 오류의 위험률이 줄어든다는 사실을 <그림2.1>을 통해 확인했다. 그러나 표본의 크기가 커지면 중심극한 정리를 적용할 수 있지만 이에 반해 모집단 분포와 근사한 특성을 표본이 가지게 되고 분포가 상당히 치우친 경우에는 중심극한의 정리 보다도 분포의 영향을 크게 받는다는 것을 알 수 있었다<그림2.2>. 표본의 크기에 따라 가정이 만족된 경우와 깨어진 경우를 각각 비교해본 것은 이론적인 내용과 비슷한 결과를 보였으나 가정이 깨진 경우의 표본에 따른 유의확률의 차이는 만족된 경우에 비해 크고 불규칙함을 <그림2.3>와 <그림2.4>를 통해 보았다. 그 외의 산점도를 살펴보면 검정결과가 모집단의 분포에 영향을 받아 가정이 만족된 경우와 가정이 깨어진 경우는 상당히 다른 결과를 보이고 모평균이 실제로는 동일하게 0.5의 차이를 가지는데도 확률이 높은 중앙값이 어디에 위치하고 있느냐에 따라서도 검정 결과에 영향을 미쳤다.

즉 모집단이 어떠한 분포를 따르느냐에 따라서 검정결과가 크게 좌우되고 모집단에 대한 아무런 정보 없이 모집단은 정규분포를 따른다고 가정하고 검정하는 것은 매우 위험한 행동이다 해결 방법이라고 하기는 어렵지만 검정이나 분석을 실시하기 전에 한번쯤 가정에 관한 의심을 해보고 사전 분석을 통해 모집단에 대해서 얻을수 있는 정보를 얻고, 이를 최대한 활용하여 위험률을 최대한으로 줄여야겠다.

### 참고문헌

- [1] 서울대학교 자연과학대학 계산통계학과 편저 (1988). [제3 개정판 현대통계학], 영지문화사.
- [2] 전명식 (1991). [수리통계학], 자유아카데미.
- [3] 윤배현·최종후 (1992). [minitab을 이용한 통계적 방법], 자유아카데미.
- [4] 성내경 (1992). [sas system과 sas언어], 자유아카데미.

## 수출·수입액을 이용한 환율 예측

이 혁종

(지도교수 : 박 유성)

고려대학교 정경대학 통계학과

## 1. 문제 제기

우리 나라는 산업 구조의 대부분을 수출과 수입에 매우 많이 의존하는 경제 구조를 가지고 있다. 따라서 수출·입 업무에서 많은 영향을 미치는 환율의 정확한 예측이 필요하며, 실제 현장에서는 그 예측에 있어서의 부정확함으로 인해 매년 커다란 손실을 보고 있는 상태이다.

이에 이번 분석에서는 1994년도 2학기 수업이었던 “시계열 자료 분석”에서 배웠던 여러 가지 기법들을 이용하여, 환율을 예측하고자 한다. 모든 경제 변수들이 그러하듯이 환율은 많은 변수들에 의해 영향을 받는다. 그 중에서 가장 큰 영향을 미치는 수출·입액을 가지고 환율을 예측하도록 하여, 분석의 신뢰도를 높이고자 한다. 물론 월간 수출·입액을 제외한 다른 변수들- 자본 수지, 금융 수지들이 환율 변화에 미치는 영향은 차후에 반드시 고려해 볼 문제라고 생각된다.

또한, 여러 경제 분야에서, 이러한 통계 기법을 이용한 환율 변화의 확실한 예측이 부정확한 환율 예측 때문에 발생하는 불필요한 손해를 감소시키는데 도움이 되었으면 하는 것이 이 보고서의 목적이다.

## 2. 분석의 방법

경제 이론에 따르면 환율은 주로 외환의 공급과 수요에 의해 결정된다고 볼 수 있다. 예를 들어, 외환의 공급이 많으면 외화에 대한 원화의 환율(이하 환율)이 떨어지게 된다. 이러한 외환의 공급은 수출에 의해서 발생하기도 하고, 자본거래에 의한 외의 유입으로 발생하기도 한다. 또, 외환의 수요는 수입과 자본거래에 의한 유출로 발생하게 된다. 이때, 이번 분석에서는 외환의 유출과 유입에 가장 커다란 영향을 미치는 월간 수출액과 수입액의 다중 전이 함수 모형을 이용하여 미래의 환율을 예측하고자 한다.

분석의 준비 작업 중에서, 분석에 사용될 원자료(raw data)가 다중 전이 함수형을 적용시키기에 부적당한 경우, 자료의 정상성 가정을 만족하기 위한 분산 안정화를 기하기 위해 자료의 변환을 필요로 하게 된다. 이때, 자료 분석에 있어 널리 사용되고 있는 박스-콕스 변환(Box-Cox transform)을 이용한다. 또, 이 변환을 실시할 때 사용되는 변환 계수  $p$ 는 산포의 균일화에 관한 수리적 이론에 의거하여 찾게 되어, 자료 변환에서 변환 계수  $p$ 를 분석자의 임의로 시행착오를 거친 후, 확정지어 버리는 오류를 범하지 않게 되었다.

덧붙여서, 변환 계수를  $p$ 를 구하기 위해, 자료의 2차원 산점도에서의 기울기를 구하는 작업에서도, 특이점에 영향을 많이 받는 최소제곱법(least squares method)을 사용하지 않고, 자료의 탐색

적 분석에 용이한 세 그룹 저항성 직선(three-group resistant line: Rline)을 사용하였다. 이상의 내용은 통계분석 패키지 MINITAB으로 수행하였으며, 분석 내용은 후미에 첨가되어 있다.

이상의 자료 정리를 통해 다중 전이 함수 모형에 적합시키도록 하며, 분석의 대부분은 또다른 통계 패키지 SAS(Statistical Analysis System 6.04)을 이용하였다. 한가지 아쉬운 점은, 통계 전 분야에 걸쳐 사용되는 SAS가 텍스트 모드 상에서의 그래픽 기능이 좋지 않아서, 원자료의 시도표를 나타내는 그림은 STATGRAPH의 우수한 그래픽 기능을 이용하여 원자료의 자료 분포를 직관적으로 알아낼 수 있도록 노력하였다.

### 3. 자료 설명

이 분석에 이용된 자료는 1985년부터 1993년까지의 우리 나라의 월별 수출액 및 수입액, 그리고 매 월말 환율 자료이며, 『경제통계연보(한국은행)』 1986년부터 1994년호에 걸쳐 기재된 통계 수치를 이용하였다. 그 자료는 다음과 같다.

[표 1] 수출액(단위:백만불)

85년	1642.9	1804.4	2347.1	2366.0	2466.4	2689.1	2475.7	2355.8	2597.2	2663.4	3069.4	3805.7
86년	2026.6	2302.7	2565.0	2558.0	2993.4	3190.2	2930.3	2950.0	3183.5	3268.5	3257.9	3488.4
87년	2898.2	2932.2	3653.4	3673.0	4028.9	4241.2	4224.4	3478.6	4440.7	4041.7	4423.8	5244.8
88년	3958.5	4121.1	4791.5	4626.5	4773.4	5052.4	5202.8	5309.9	5392.9	5256.2	5810.5	6400.7
89년	4400.8	4336.8	5329.7	4783.7	4950.0	5427.4	5232.7	5268.3	5474.6	5304.2	5654.8	6214.0
90년	3961.1	4679.9	5271.3	4949.8	5214.8	5688.7	5489.5	5321.2	6111.8	5284.9	6019.9	7022.8
91년	4639.4	4865.5	5725.9	5838.7	6146.2	6660.4	5540.0	5762.6	5680.2	6636.7	6677.4	7697.1
92년	5392.3	5144.2	6428.8	6360.9	6337.1	7053.4	6351.9	6176.6	6625.5	7125.7	6628.7	7006.6
93년	5328.9	5935.3	6914.8	6750.7	6743.3	7248.4	6662.0	6543.8	7185.5	7480.8	7344.6	8097.8

[표 2] 수입액(단위:백만불)

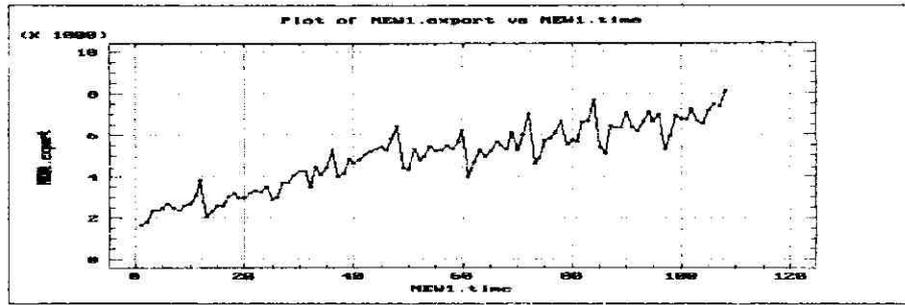
85년	2400.9	2063.2	2264.5	2499.1	2758.5	2440.5	2435.0	2440.3	2412.8	2612.7	3103.9	3704.3
86년	1619.6	1977.4	2200.5	2033.2	2028.8	2165.6	2147.8	1975.5	1910.1	2202.4	2334.4	2763.6
87년	2308.9	2741.9	3136.2	3345.2	3634.1	3539.9	3623.9	3489.3	3671.0	3392.6	3836.4	4300.4
88년	3688.1	3620.4	4209.4	4206.0	4224.1	4287.3	4494.1	4834.3	4317.2	4455.5	4598.9	4875.3
89년	4421.0	4286.8	5300.7	4802.8	5012.0	5325.0	5152.4	5768.8	4952.6	5357.8	5479.8	5605.0
90년	4624.4	5252.4	5941.5	5321.5	5917.3	5464.6	5412.7	5632.5	6128.0	6161.5	7522.2	6465.1
91년	6043.2	6701.3	6822.7	6907.8	7116.9	6575.6	7126.7	6536.9	6434.4	7085.9	7363.1	6810.4
92년	7336.7	6364.2	7197.8	6831.4	6926.6	7246.9	7158.8	6074.9	6126.3	7242.9	6403.8	6865.1
93년	6300.7	6487.1	7177.1	6852.5	6948.2	7532.1	7193.6	6570.0	7020.4	7008.0	7256.2	7454.4

[표 3] 환율(₩/\$)

85년	830.6	842.8	852.3	865.9	871.0	873.8	876.3	886.8	891.7	892.2	889.1	890.2
86년	888.7	883.8	885.2	885.1	889.8	886.6	885.5	880.2	877.0	873.2	865.0	861.4
87년	875.2	854.8	846.9	834.1	822.7	808.9	808.0	807.0	805.8	801.4	796.4	792.3
88년	781.6	760.8	746.2	740.0	732.9	728.3	723.8	722.0	719.0	701.4	687.5	684.1
89년	680.6	673.1	671.9	666.3	666.7	667.2	667.4	669.2	670.0	671.6	672.7	679.6
90년	686.3	694.0	702.1	707.0	712.3	716.0	715.1	714.0	712.9	713.8	713.1	716.4
91년	719.0	724.4	724.7	725.1	723.0	723.1	726.1	735.6	741.5	750.3	754.5	760.8
92년	788.4	762.0	768.2	775.1	778.8	783.5	790.2	788.1	787.6	782.4	785.1	788.4
93년	794.0	794.9	794.0	795.9	801.1	803.7	806.6	808.4	808.8	808.2	807.6	808.1

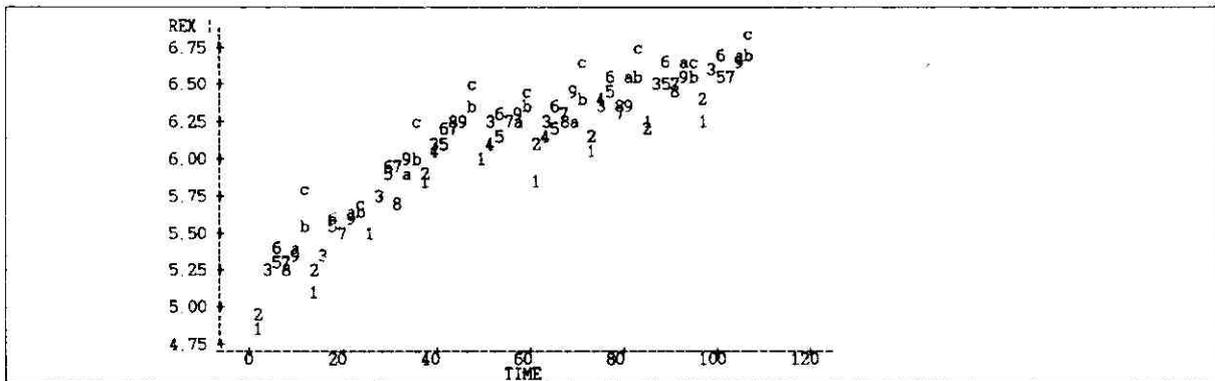
## 4. 수출액의 시도표

[그림 1] 수출액의 시도표(원자료)



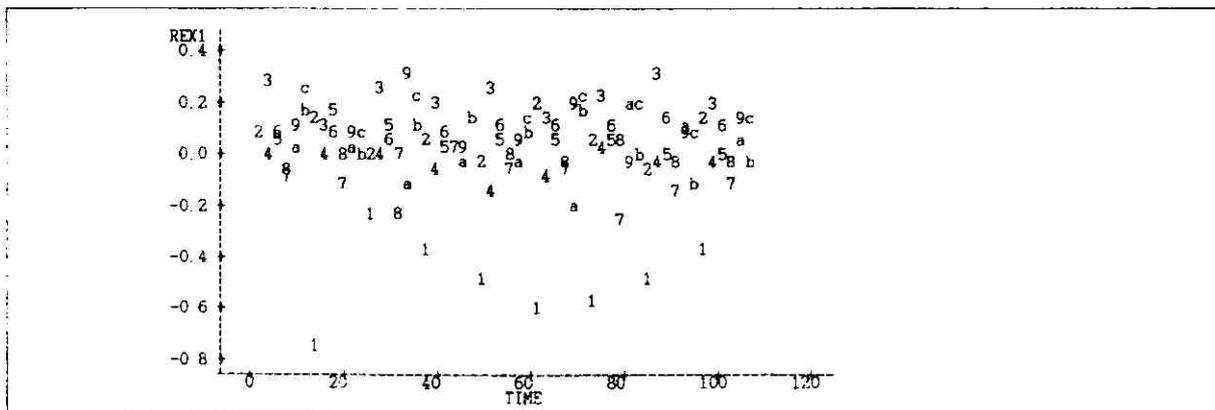
위에서 보면 수출액이 시간에 따라 증가하고 있으며 그 홀이진 정도(분산)가 점점 커지고 있는 것을 알 수 있다. 이는 수출액 자료를 다중 적어 함수 모형에 이용하기 위해서는 어떠한 변환을 해 주어야 한다는 것을 뜻하며 앞에서 언급했던 Rline에 의하여 구한 변환 계수( $P=0.2136$ )를 적용한  $y=x^{0.2136}$  변환을 실행하였다. 그 결과는 다음과 같다.

[그림 2]  $y=x^{0.2136}$  변환 후의 수출액의 시도표



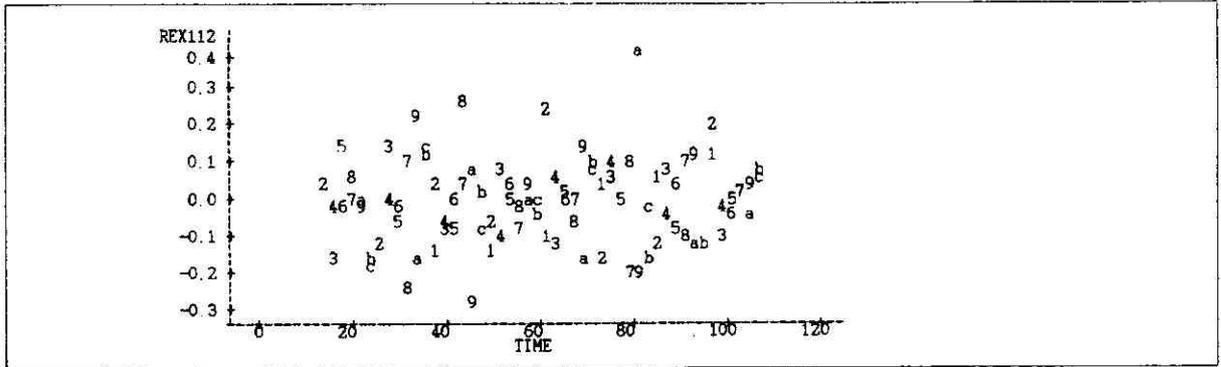
위에서 볼 수 있듯이 수출액은 점점 증가하고 있으며 따라서 차분을 통하여 그 추세를 제거시키야 할 필요성을 느끼게 된다. 그 결과는 아래와 같다.

[그림 3] 전체적 추세를 제거한 수출액의 시도표(1차 차분)



위와 같은 변환과 차분을 거친 결과가 여전히 어떠한 경향을 띠고 있음을 볼 수 있다. 1월달의 자료는 아래쪽으로, 그리고 3,12월달의 자료는 위쪽으로 몰려 있음에 따라 우리는 12차분을 시행하여 그 월별 추세를 제거시켰다.

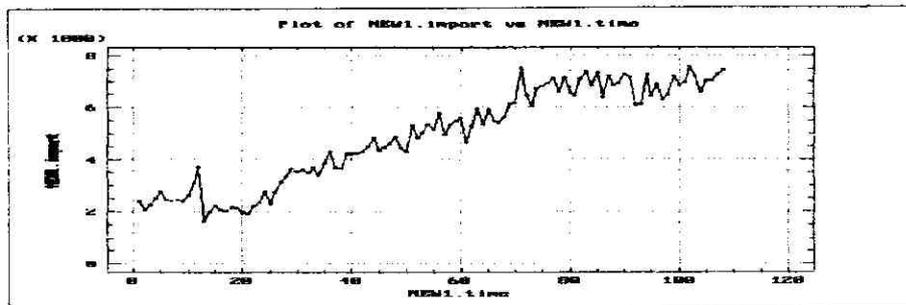
그림 4) 월별 추세를 제거한 수출액의 시도표(12차분)



이제 위에서 볼 수 있듯이 우리는 더이상 어떠한 경향성도 발견할 수 없다. 따라서 위의 자료는 우리가 실시한 변환과 차분에 의해 정상성 가정을 만족시킨다고 할 수 있겠다. 이는 이 자료가 Arima Procedure에 이용될 수 있다는 것을 의미한다.

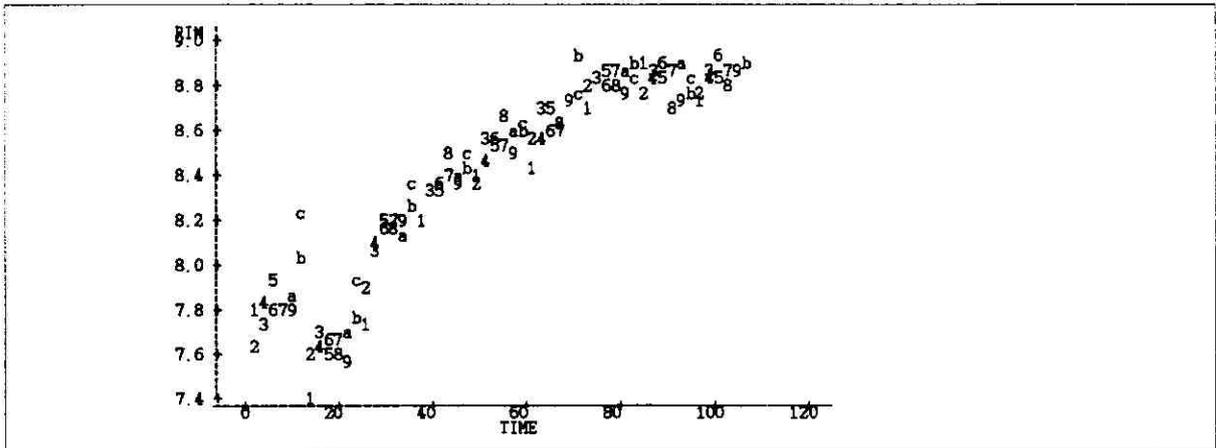
## 5. 수입액의 시도표

[그림 5] 수입액의 시도표(원자료)



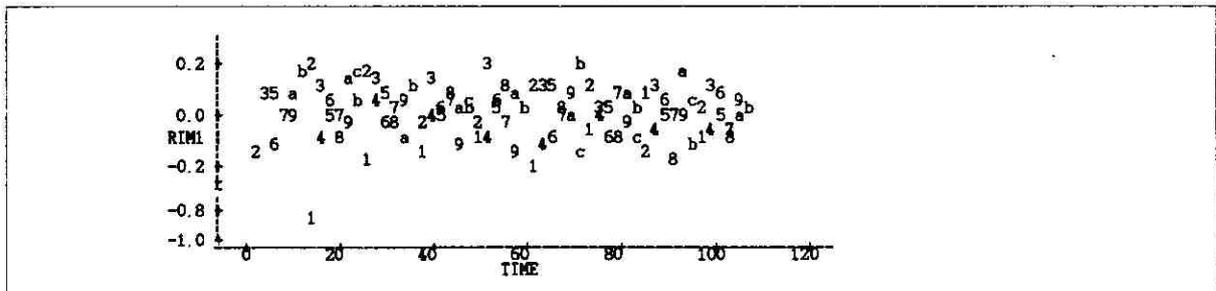
수출액의 시도표와 마찬가지로 그 흠어지는 정도(분산)가 증가하고 있다. 이는 수입액 자료를 이용하기 위해서는 어떠한 변환을 해 주어야 한다는 것을 뜻하며 앞에서 언급했던 Rline에 의하여 구한 변환 계수( $P=0$ )를 적용한  $y = \log x$  변환을 실행하였다. 그 결과는 다음과 같다.

[그림 6]  $y = \log_e x$  변환을 실행한 수입액의 시도표

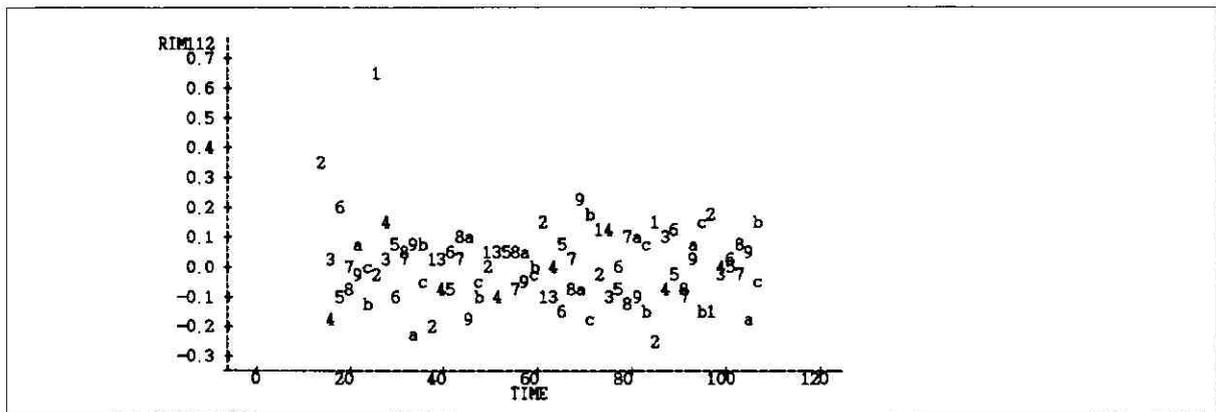


위에서 보는 바와 같이 전체 추세가 눈에 띄므로 1차 차분을 하여 그 추세를 제거하였다. 그 결과는 다음 페이지와 같으며 이를 보면 여전히 월별 추세를 갖고 있어서 12차 차분을 한번 더 시행하여야 할 필요성이 느껴진다. 그 결과가 아래에 나와 있으며 그 결과를 보면 더이상 어떠한 추세도 갖지 않아서 정상성의 가정을 만족시키고 있다. 따라서 이 자료도 Arima Procedure에 이용할 수 있음을 알 수 있다.

[그림 7] 전체 추세를 제거한 수입액의 시도표(1차 차분)

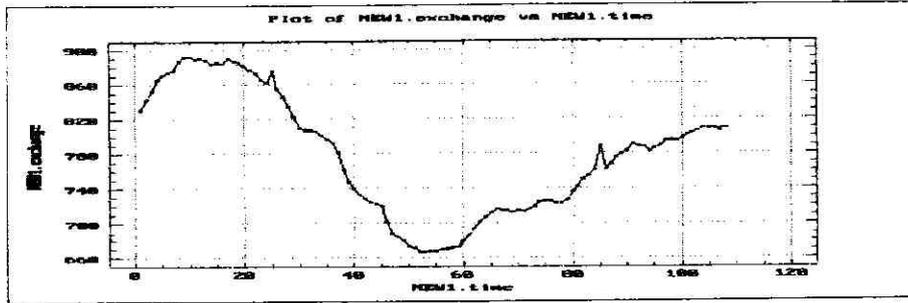


[그림 8] 월별 추세를 제거한 수입액의 시도표(12차분)



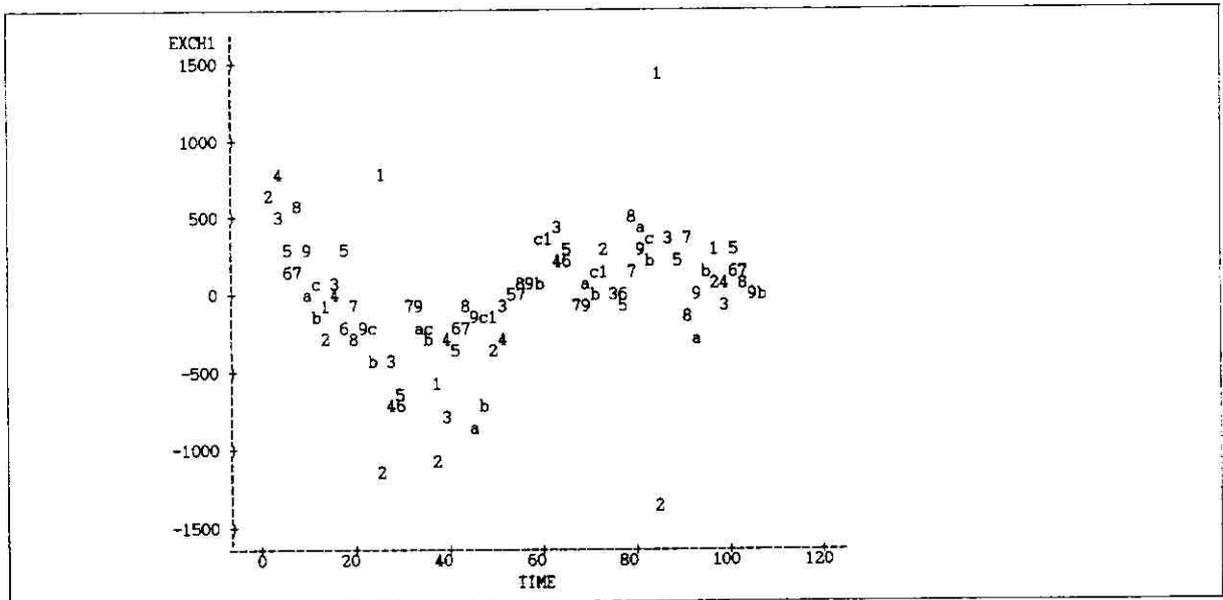
## 6. 환율의 시도표

그림 9) 환율의 시도표(원자료)



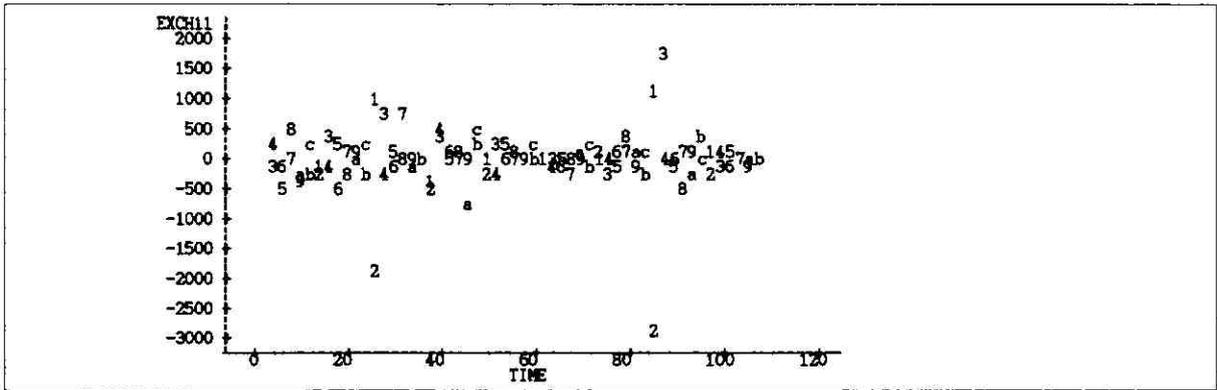
위에서 보면 환율은 시간에 따라 감소한 후, 다시 증가하고 있으며 그 흠어진 정도(분산)가 점점 작아지는 것을 알 수 있다. 이는 환율 자료를 이용하기 위해서는 어떠한 변환을 해 주어야 한다는 것을 뜻하며, 앞에서 언급했던 Rline에 의하여 구한 변환 계수( $P=1.5319$ )를 적용한  $y=x^{1.5319}$  변환을 실행하였다. 그 후 변환된 자료의 추세가 사인 커브와 유사하게 변화하고 있어서 차분을 통하여 추세를 제거하고자 한다. 아래는 1차 차분한 결과이다.

그림 10) 1차 차분하여 변환된 환율의 시도표



위에서 보면 아직까지도 추세가 남아 있는 것이 보이며 따라서 한 번 더 1차 차분을 취하는 것이 타당하다고 생각되었다.

[그림 11] 再次 1차 차분을 취하여 변환된 환율의 시도표

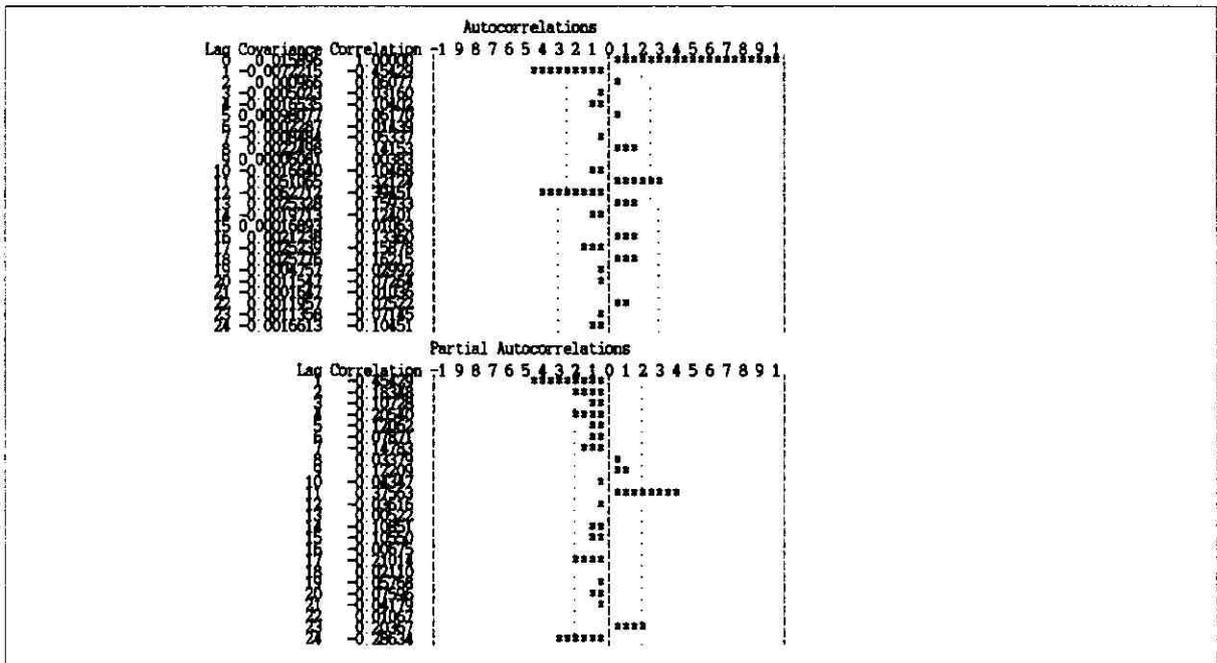


이제 위에서 볼 수 있듯이 우리는 더이상 어떠한 경향성도 발견할 수 없다. 따라서 위의 자료는 두번 실시한 차분을 통해 정상성 가정을 만족시킨다고 할 수 있겠다. 이는 이제 이 자료가 Arima Procedure에 이용될 수 있다는 것을 의미한다.

### 7. 수출액의 사전백색화 모수

먼저 수출액의 ACF와 PACF를 보도록 하자.

[그림 12] 수출액의 ACF와 PACF



[표 4] 수출액의 백색 잡음(White Noise)

Autocorrelation Check for White Noise									
To Lag	Chi Square	DF	Prob	Autocorrelations					
6	22.20	6	0.001	-0.454	0.061	-0.032	-0.104	0.062	-0.014
12	54.41	12	0.000	-0.053	0.142	0.004	-0.105	0.321	-0.395
18	67.24	18	0.000	0.159	-0.124	0.011	0.134	-0.159	0.162
24	70.79	24	0.000	-0.030	-0.073	-0.010	0.075	-0.071	-0.105

ACF와 PACF를 보기 전에 우선 ARIMA 모형의 적용을 위하여 백색 잡음의 P-Value를 보자. 위에서 볼 수 있듯이 P-value가 상당히 작아서 문제가 없다는 것을 알 수 있다. 또, ACF와 PACF를 보면 ACF의 경우, 계절형 부분의 LAG=12에서 CUT-OFF되는 형태를 띄고 있으며 비계절형 부분의 LAG=1에서 CUT-OFF되고 있다. 또 PACF를 보면, 계절형 부분은 LAG=12, 24에서 CUT-OFF되고 비계절형 부분은 LAG=1에서 CUT-OFF된 것으로 모형을 만들어 볼 수 있다. 따라서 우리가 만들어 본 모형과 그에 대한 AIC값과 SBC값은 다음과 같다.

[표 5] 예상 모형들의 AIC와 SBC의 비교

MODEL	AIC	SBC
ARIMA (0, 1, 1) × (0, 1, 1) <sub>12</sub>	-178.78972	-171.12809
ARIMA (1, 1, 0) × (2, 1, 0) <sub>12</sub>	-161.14974	-150.93423
ARIMA (1, 1, 1) × (2, 1, 1) <sub>12</sub>	-176.48135	-161.15809

위의 세 모형中 AIC값과 SBC값이 작은 모형을 선택해야 하므로 ARIMA (0, 1, 1) × (0, 1, 1)<sub>12</sub> 가 가장 좋다는 것을 알 수 있다.

이제 ARIMA (0, 1, 1) × (0, 1, 1)<sub>12</sub> 을 이용한 모수 추정치를 보자.

[표 6] ARIMA (0, 1, 1) × (0, 1, 1)<sub>12</sub> 모형의 모수 추정치

Conditional Least Squares Estimation				
Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MU	-0.0018601	0.0012553	-1.48	0
MA1,1	0.66929	0.07756	8.63	1
MA2,1	0.71098	0.07999	8.89	12

각 추정치에 대한 T-RATIO를 보면 MU가 유의하지 않아서 NOCONSTANT 옵션이 필요하다는 것을 알 수 있다. 이제 그 옵션을 이용한 최종 추정을 위하여 다음과 같이 수정하였다.

**ESTIMATE q=(1)(12) NOCONSTANT;**

그 결과는 다음과 같다.

[표 7] 상수를 제외한 ARIMA (0, 1, 1) × (0, 1, 1)<sub>12</sub> 모형의 모수 추정치

Conditional Least Squares Estimation				
Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MA1,1	0.63555	0.08007	7.94	1
MA2,1	0.68171	0.08189	8.32	12

잔차의 자기 상관성을 확인하기 위하여 다음 페이지에 있는 OUTPUT을 보면 P-value가 대체로 커서 잔차가 독립이라는 귀무가설을 기각하지 못하므로 우리가 설정한 모형이 적절하다고 할 수 있다.

[표 8] ARIMA (0, 1, 1) × (0, 1, 1)<sub>12</sub> 모형에서 잔차의 자기 상관성 체크

Autocorrelation Check of Residuals										
To Lag	Chi Square	DF	Prob	Autocorrelations						
6	4.00	4	0.407	0.063	-0.060	-0.116	-0.029	0.027	0.130	
12	9.36	10	0.498	0.064	0.156	0.007	-0.031	0.122	0.074	
18	14.81	16	0.539	-0.025	-0.133	-0.076	0.071	0.003	0.133	
24	19.64	22	0.605	0.007	-0.064	-0.083	-0.041	-0.103	-0.120	

사전백색화 모수는 아래의 수치를 이용하여 구해 볼 수 있을 것이다. 이 수치는 이후에 환율을 추정하기 위해 사전백색화 작업에 사용된다.

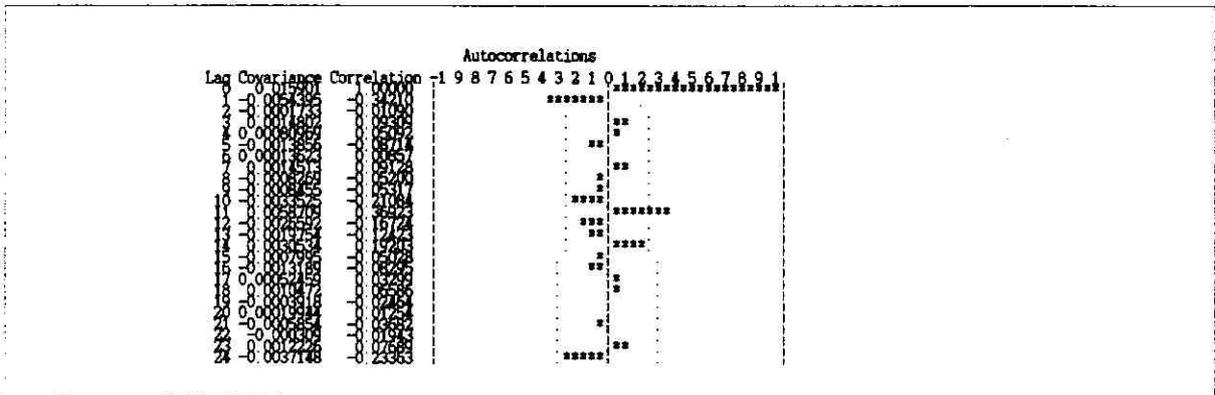
[표 9] 수출액의 사전백색화 모수

Moving Average Factors	
Factor 1:	1 - 0.63555 B**(1)
Factor 2:	1 - 0.68171 B**(12)

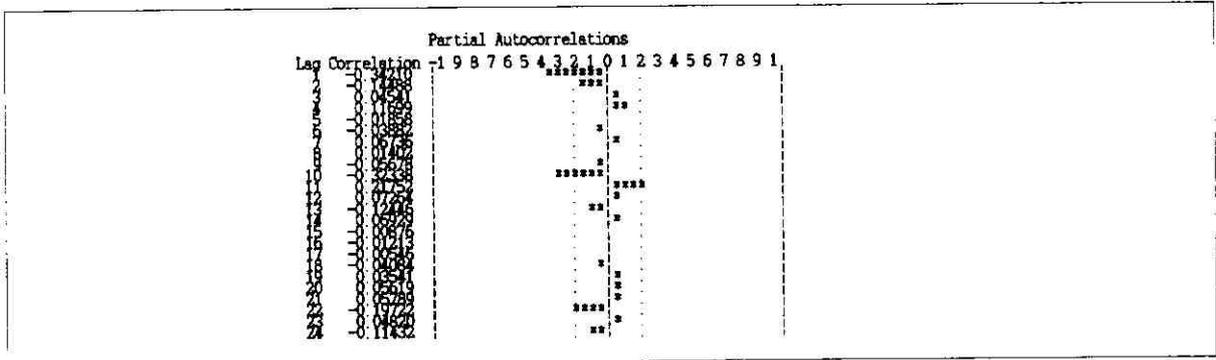
## 8. 수입액의 사전백색화 모수

먼저 수입액의 ACF와 PACF를 보도록 하자.

[그림 13] 수입액의 ACF와 PACF



[그림 13] 수입액의 ACF와 PACF (계속)



[표 10] 수입액의 백색 잡음(White Noise)

Autocorrelation Check for White Noise									
To Lag	Chi Square	DF	Prob	Autocorrelations					
6	13.40	6	0.037	-0.342	-0.011	0.093	0.051	-0.087	0.009
12	37.74	12	0.000	0.091	-0.052	-0.053	-0.211	0.369	-0.167
18	45.41	18	0.000	-0.124	0.192	-0.050	-0.083	0.033	0.066
24	53.56	24	0.000	-0.025	0.013	-0.037	-0.019	0.077	-0.234

수출액의 경우와 마찬가지로, ACF와 PACF를 보기 전에 우선 ARIMA 모형의 적용을 위하여 백색 잡음의 P-Value를 보자. 위에서 볼 수 있듯이 P-value가 상당히 작아서 문제가 없다는 것을 알 수 있다. ACF에서 계절부분은 LAG=12,24에서 CUT-OFF되고 있고 비계절 부분은 LAG=1에서 CUT-OFF되고 있다. PACF에서 계절 부분은 역시 기각역을 벗어난 수치가 없고 비계절 부분은 LAG=1에서 CUT-OFF되고 있다. 따라서 우리는 다음의 3가지 모형을 생각해 볼 수 있다. 아래에는 각 모형과 그에 대한 AIC값과 SBC값을 같이 기재하였다.)

[표 11] 예상 모형들의 AIC와 SBC의 비교

MODEL	AIC	SBC
ARIMA (0, 1, 1) × (0, 1, 2) <sub>12</sub>	-132.64799	-122.43248
ARIMA (1, 1, 0) × (0, 1, 0) <sub>12</sub>	-131.68139	-126.57364
ARIMA (1, 1, 1) × (0, 1, 2) <sub>12</sub>	-131.73386	-118.96447

위의 세 모형中 AIC값과 SBC값이 작은 모형을 선택해야 하는데, ARIMA (0, 1, 1) × (0, 1, 2)<sub>12</sub> 모형이 SBC가 조금 크지만, 이후 계속되는 모형적합에서 가장 좋은 모형이라는 것을 알 수 있다.

이제 ARIMA (0, 1, 1) × (0, 1, 2)<sub>12</sub> 을 이용한 모수 추정치를 보자.

1. ACF에서는 LAG=10,11, PACF에서는 LAG=10에서 CUT-OFF한 것을 볼 수 있다. 그러나, 모형 설정에서는 12의 배수에 근접한 CUT-OFF는 12의 배수로 간주하는 것이 편리하다. 단, 오차의 자기 상관성에서 문제가 발생할 수 있으므로, 그 문제는 잠시 후에 다시 언급하기로 한다.

[표 12] ARIMA (0, 1, 1) × (0, 1, 2)<sub>12</sub> 모형의 모수 추정치

Conditional Least Squares Estimation				
Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MU	0.0042863	0.0080408	0.53	0
MA1,1	0.32449	0.10115	3.21	1
MA2,1	0.15853	0.10587	1.50	12
MA2,2	-0.13811	0.10544	-1.31	14

각 추정치에 대한 T-RATIO를 보면 MU가 유의하지 않아서 NOCONSTANT 옵션이 필요하다는 것을 알 수 있다. 이제 그 옵션을 이용한 최종 추정을 위하여 다음과 같이 수정하였다.

ESTIMATE q=(1)(12 24) NOCONSTANT;

그 결과는 다음과 같다.

[표 13] 상수를 제외한 ARIMA (0, 1, 1) × (0, 1, 2)<sub>12</sub> 모형의 모수 추정치

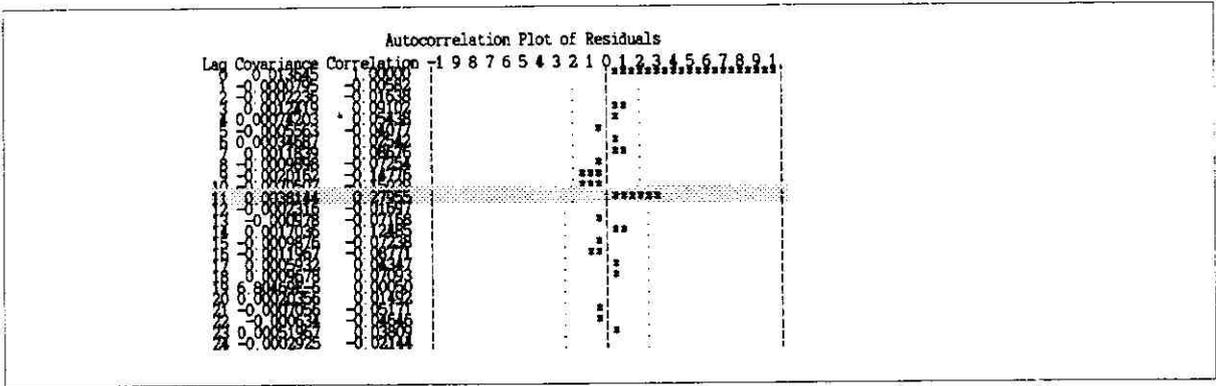
Conditional Least Squares Estimation				
Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MA1,1	0.32342	0.10044	3.22	1
MA2,1	0.11836	0.10642	1.11	12
MA2,2	0.18260	0.10892	1.68	24

오차의 자기 상관성을 확인하기 위하여 다음 페이지에 있는 OUTPUT을 보면 P-value가 유의하지 않은 수치가 몇 개 존재하므로 잔차가 독립이라는 귀무가설을 기각할 수도 있다. 따라서 우리가 설정한 모형이 적절하다고 확신할 수는 없다. 그러므로 잔차의 ACF, PACF를 보도록 하자.

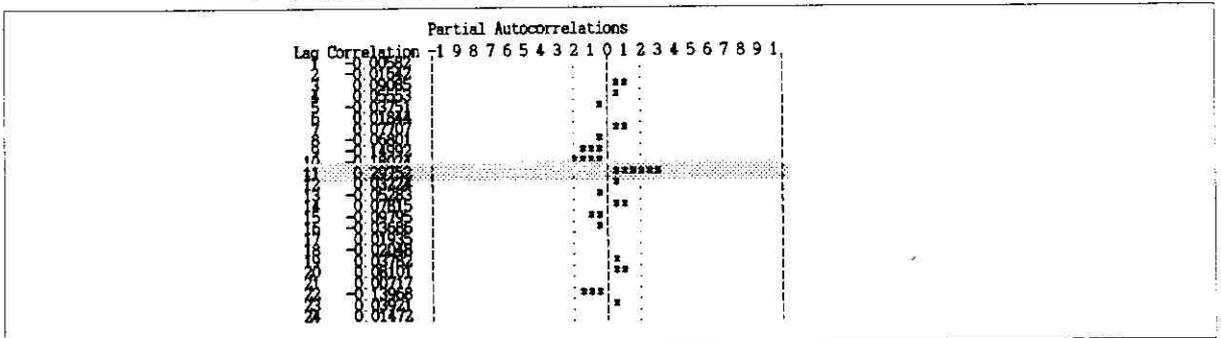
[표 14] ARIMA (0, 1, 1) × (0, 1, 2)<sub>12</sub> 모형에서 잔차의 자기상관성 체크

Autocorrelation Check of Residuals									
To Lag	Chi Square	DF	Prob	Autocorrelations					
6	1.40	3	0.706	-0.006	-0.016	0.091	0.054	-0.041	0.025
12	16.14	9	0.064	0.087	-0.073	-0.148	-0.150	0.280	-0.017
18	20.81	15	0.143	-0.072	0.125	-0.072	-0.088	0.043	0.071
24	21.69	21	0.418	0.000	0.015	-0.052	-0.046	0.038	-0.021

[그림 14] ARIMA (0, 1, 1) × (0, 1, 2)<sub>12</sub> 모형에서 잔차의 ACF



[그림 15] ARIMA (0, 1, 1) × (0, 1, 2)<sub>12</sub> 모형에서 잔차의 PACF



앞 페이지의 ACF, PACF를 보면 LAG=11에서 상당히 유의적인 값을 가지고 있음을 볼 수 있다. 이것은 앞에서 언급했던 것과 같이, 모형설정시 CUT-OFF된 부분을 고려하지 않았기 때문이다. 따라서 모형을 다음과 같이 수정하였다.

**ESTIMATE q=(1)(11 12) NOCONSTANT;2)**

위의 모형을 이용한 모수 추정치는 다음과 같으며, 잔차의 자기 상관 문제는 상당히 향상되었음을 알 수 있다.

[표 15] 상수를 제외한 q=(1)(11 12) 모형의 모수 추정치

Conditional Least Squares Estimation					
Parameter	Estimate	Approx. Std Error	T Ratio	Lag	
MA1,1	0.26952	0.10175	2.65	1	
MA2,1	-0.34607	0.09810	-3.53	11	
MA2,2	0.17251	0.09698	1.78	12	

2. ESTIMATE q=(1)(11 24) noconstant: 와 같은 모형도 고려해 보았으나, AIC와 SBC에서 덜 적합한 것으로 밝혀졌다.

[표 16] q=(1)(11 12) 모형에서 잔차의 자기 상관성 체크

Autocorrelation Check of Residuals										
To Lag	Chi Square	DF	Prob	Autocorrelations						
6	1.62	3	0.656	-0.022	0.056	0.091	0.032	-0.044	0.033	
12	8.08	9	0.526	0.065	-0.103	-0.145	-0.154	0.022	-0.002	
18	10.57	15	0.783	-0.081	0.094	-0.019	-0.069	0.020	0.030	
24	13.57	21	0.887	0.065	0.025	-0.053	-0.017	-0.021	-0.123	

사전백색화 모수는 아래의 수치를 이용하여 구해 볼 수 있을 것이다. 이 수치는 이후에 환율을 추정하기 위해 사전백색화 작업에 사용된다.

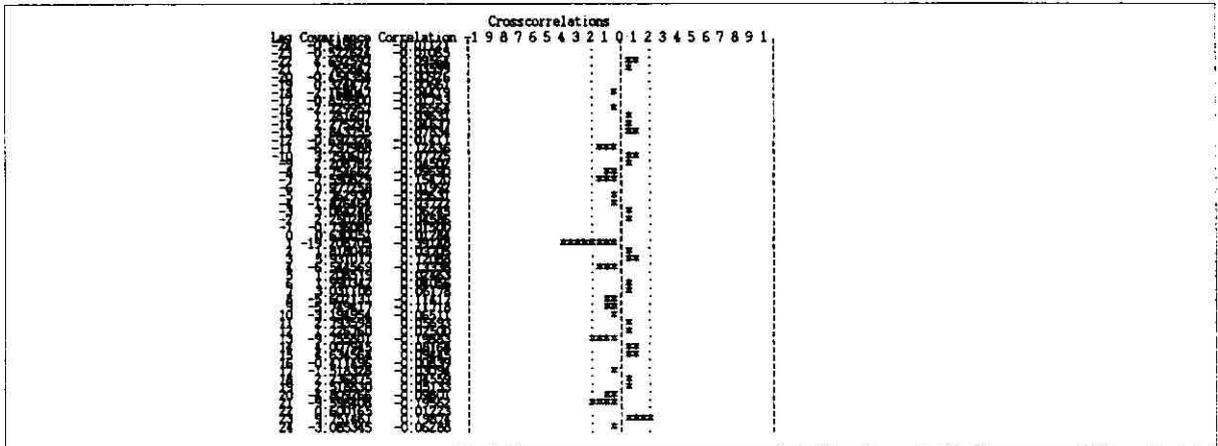
[표 17] 수입액의 사전백색화 모수

Moving Average Factors	
Factor 1:	1 - 0.26952 B**(1)
Factor 2:	1 + 0.34607 B**(11) - 0.17251 B**(12)

### 9. 사전백색화 모수를 이용한 사전백색화 작업

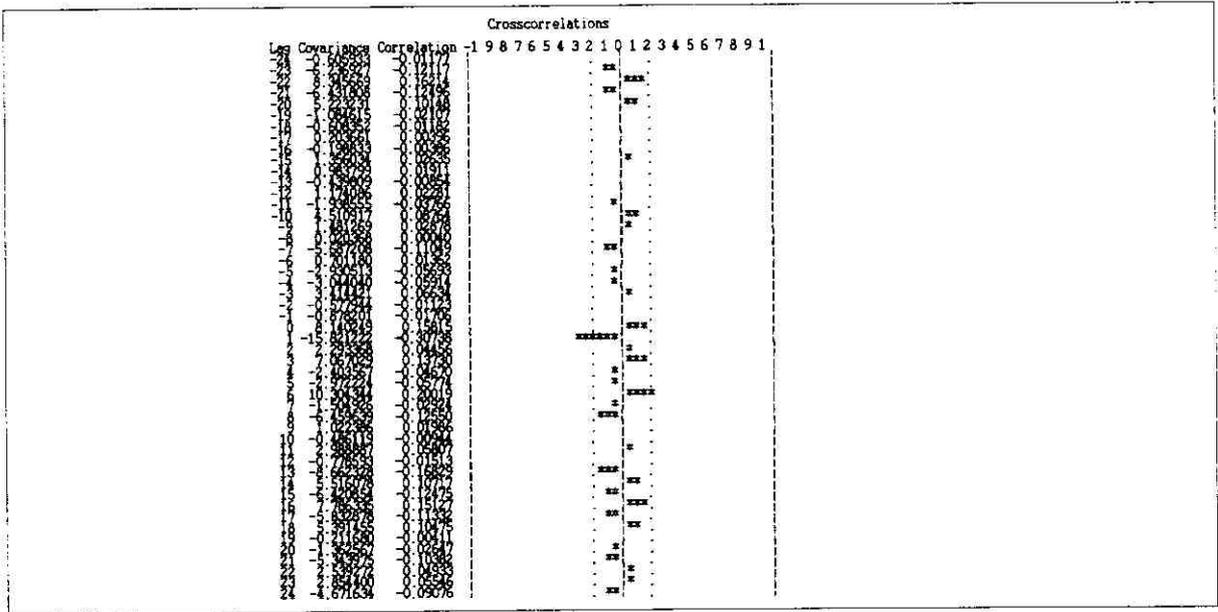
다음의 수출액과 환율의 CCF를 보면, 음의 LAG에서는 기각역을 벗어난 값이 없어 전이함수 모형의 적용이 적절하다는 것을 알 수 있다. 또 S, D, R을 결정하기 위하여 양의 LAG를 보면 D1=1, S1=3, R1=1이 됨을 알 수 있다.

[그림 16] 수출액과 환율의 CCF



다음의 수입액과 환율의 CCF를 보면, 음의 LAG에서는 기각역을 벗어난 값이 없어 전이함수 모형의 적용이 적절하다는 것을 알 수 있다. 또 S, D, R을 결정하기 위하여 양의 LAG를 보면 D2=1, S2=5, R2=0이 됨을 알 수 있다.

[그림 17] 수입액과 환율의 CCF



이제 앞에서 구한 S, D, R을 이용하여  $W_{S1}(B)$ ,  $W_{S2}(B)$ ,  $\delta_{R1}(B)$ ,  $\delta_{R2}(B)$ 를 구해 보면 아래와 같다.

[표 18] D1=1, S1=3, R1=1, D2=1, S2=5, R2=0으로 부터의  $W_{S1}(B)$ ,  $W_{S2}(B)$ ,  $\delta_{R1}(B)$ ,  $\delta_{R2}(B)$ 의 추정치

Conditional Least Squares Estimation						
Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	-1.90414	46.49324	-0.04	0	REXCH	0
NUM1	-924.77164	489.06545	-1.89	0	REX	1
NUM1,1	906.89054	742.21903	1.22	1	REX	1
NUM1,2	460.05572	740.15384	0.62	2	REX	1
NUM1,3	912.21569	541.18658	1.69	3	REX	1
DEN1,1	-0.67016	0.17617	-3.80	1	REX	1
NUM2	-1055.4	494.85330	-2.13	0	RIM	1
NUM1,1	412.35622	523.28697	0.79	1	RIM	1
NUM1,2	-764.05159	508.57515	-1.50	2	RIM	1
NUM1,3	-312.07153	484.76822	-0.64	3	RIM	1
NUM1,4	-26.25212	461.08785	-0.06	4	RIM	1
NUM1,5	-659.06094	423.00728	-1.56	5	RIM	1

위에서  $\mu$ 로 표시된 계수 이외에는 T-RATIO가 유의하지 못하므로 가장 유의하지 않은 계수 부터 제외해 나가면, 결국  $\mu$ 로 표시된 계수만이 유의한 것으로 아래에 나타난다.

[표 19] D1=1, S1=3, R1=1, D2=1, S2=5, R2=0으로 부터의 유의한 추정치

Conditional Least Squares Estimation						
Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
NUM1	-1064.1	433.55377	-2.45	0	REX	1
NUM1,1	940.27741	521.19261	1.80	1	REX	1
NUM1,2	410.68627	359.29699	1.14 <sup>3)</sup>	3	REX	1
DEN1,1	-0.67849	0.22498	-3.02	1	REX	1
NUM2	-809.43925	435.32277	-1.86	0	RIM	1
NUM1,1	-671.79947	412.65799	-1.63	2	RIM	1

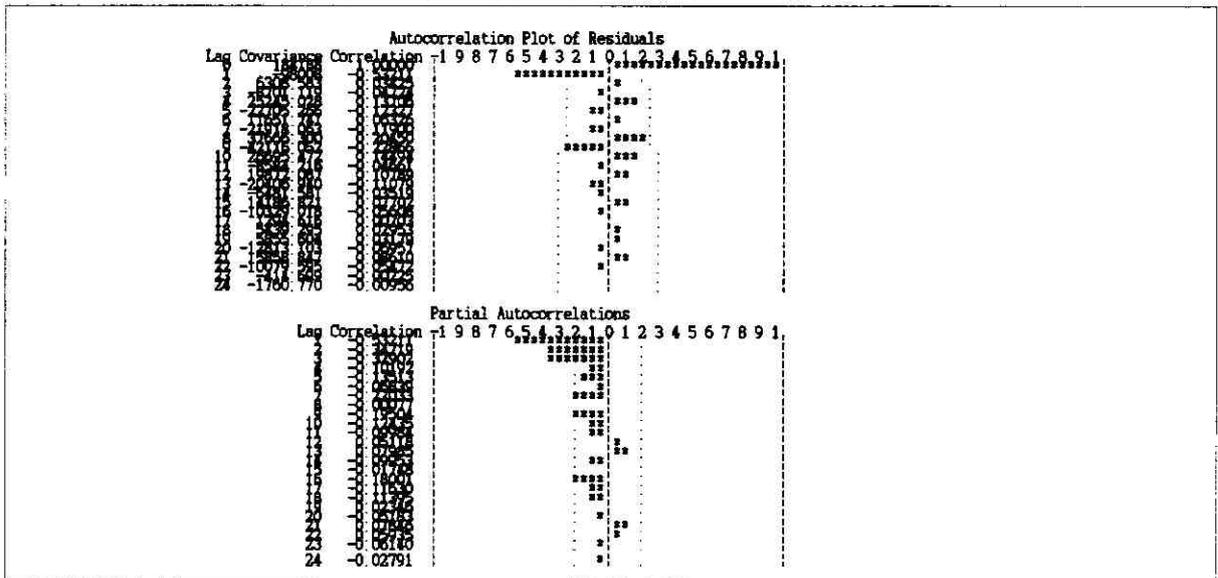
그리고 잔차의 P-value를 보면 그 값이 상당히 작으므로 Nt에 대한 모형을 필요로 한다.

[표 20] Nt의 모형화 판단 포토만토 검정량

Autocorrelation Check of Residuals									
To Lag	Chi Square	DF	Prob	Autocorrelations					
6	30.67	6	0.000	-0.532	0.034	-0.047	0.137	-0.123	0.063
12	45.43	12	0.000	-0.119	0.204	-0.229	0.145	-0.047	0.108
18	48.02	18	0.000	-0.111	-0.035	0.077	-0.056	0.007	0.030
24	49.99	24	0.001	0.032	-0.070	0.086	-0.055	-0.002	-0.010

이제 잔차의 ACF, PACF를 보도록 하자.

[그림 18] 잔차의 모형 설정의 위한 ACF와 PACF



3. T Ratio가 1.14로써 조금 작은 것 같지만, 이 항을 제거하면, 연속적으로 다른 항들도, 유의하지 않은 것으로 되게 된다. 그러한 이유로, 이 항은 유의 하다고 채택한다.

최종적으로 잔차까지 고려한 모형은 다음과 같다.

ESTIMATE q=1 INPUT=(1\$(1,3)/(1)REX 1\$(2)/RIM) NOCONSTANT;

이렇게 하여 모형을 추정한 최종 결과가 다음과 같다.

[표 21] 모형의 최종 추정치

Conditional Least Squares Estimation						
Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MA1,1	0.93295	0.04575	20.39	1	REXCH	0
NUM1	-1291.8	325.80294	-3.97	0	REX	1
NUM1,1	1152.6	472.03044	2.44	1	REX	1
NUM1,2	442.74382	352.49993	1.26	3	REX	1
DEN1,1	-0.60923	0.25670	-2.37	1	REX	1
NUM2	-671.36576	279.78955	-2.40	0	RIM	1
NUM1,1	-762.33727	255.21637	-2.99	2	RIM	1

이제 최종적인 잔차의 독립성을 보면 P-value가 상당히 커서 독립이라는 귀무가설을 기각할 수 없다. 또한 입력 계열(REX, RIM의 추정된 변수)과 Nt사이에도 상관관계 검정을 위한 포트만토(Portmanteau)값이 만족스러운 수준이므로 설정된 모형이 상당히 적합하다고 할 수 있겠다.

[표 22] 모형 진단을 위한 잔차의 포트만토 검정 통계치

Autocorrelation Check of Residuals									
To Lag	Chi Square	DF	Prob	Autocorrelations					
6	2.82	5	0.727	-0.077	-0.009	0.004	0.144	-0.049	-0.001
12	10.38	11	0.497	-0.055	0.128	-0.148	0.138	0.073	0.085
18	13.10	17	0.729	-0.105	-0.065	0.042	-0.026	0.016	0.081
24	15.69	23	0.868	0.052	-0.046	0.070	-0.062	-0.072	-0.047

[표 23] 입력 계열과 잔차의 상관관계 검정을 위한 포트만토 검정 통계치

Crosscorrelation Check of Residuals with Input REX									
To Lag	Chi Square	DF	Prob	Crosscorrelations					
5	1.30	2	0.522	-0.046	0.009	-0.039	-0.014	0.011	0.102
11	11.28	8	0.186	0.181	-0.055	-0.149	-0.228	-0.022	0.022
17	15.89	14	0.320	-0.205	-0.033	0.086	0.012	0.003	-0.025
23	29.97	20	0.070	0.037	-0.109	-0.293	-0.049	0.125	-0.199

Crosscorrelation Check of Residuals with Input RIM									
To Lag	Chi Square	DF	Prob	Crosscorrelations					
5	4.52	4	0.340	-0.007	0.045	-0.079	0.083	-0.056	0.179
11	5.71	10	0.839	0.012	0.012	0.004	-0.100	-0.022	0.049
17	11.78	16	0.759	-0.204	0.070	-0.025	0.137	-0.040	0.006
23	18.52	22	0.675	0.033	0.060	-0.179	-0.087	0.089	-0.150

## 10. 모형의 이해

[표 24] 다중 전이 함수 모형의 최종 모형 결과

<p>Model for variable REXCH</p> <p>No mean term in this model. Period(s) of Differencing = 1,1.</p> <p>Moving Average Factors Factor 1: 1 - 0.93295 B**(1)</p> <p>Input Number 1 is REX with a shift of 1. Period(s) of Differencing = 1,12.</p> <p>The Numerator Factors are Factor 1: -1292 - 1152.6 B**(1) - 442.74 B**(3)</p> <p>The Denominator Factors are Factor 1: 1 + 0.60923 B**(1)</p> <p>Input Number 2 is RIM with a shift of 1. Period(s) of Differencing = 1,12.</p> <p>The Numerator Factors are Factor 1: -671.4 + 762.34 B**(2)</p>
---

이 자료를 이용하여 환율 예측에 이용되는 수식을 아래와 같이 정리할 수 있다.

$$\begin{aligned}
 1. \nabla^2 EXCHANGE^{1.5319} &= \frac{(-1292 - 1152.6B^1 - 442.74B^3)}{(1 + 0.60923B^1)} B^1 \nabla_{12} \nabla EXPORT^{0.2136} \\
 &+ (-671.4 + 762.34B^2) B^1 \nabla_{12} \nabla \log IMPORT + (1 - 0.93295B^1) a_t
 \end{aligned}$$

위의 식을 보면 수출액의 계수는 전부 음수인데, 이는 수출을 많이 하면 할수록 외환 보유고가 많이 되고, 그로 인해서 환율이 평가절상-1\$당 원화의 화폐가치가 늘어나게 된다는 객관적 사실과 일치한다. 또 수입액의 계수가 양수 값을 가지고 있다는 사실에서, 수입이 많으면 많을수록 환율은 평가절하-원화의 화폐가치가 줄어들게 된다는 학술적 사실과 일치함을 볼 수 있다. 이상에서와 같이 주어진 자료에 적합한 모형을 설정했음을 직관적으로 알 수 있게 된다.

$$\text{※ } rexch = EXPORT^{1.5319}, \quad rex = EXPORT^{0.2136}, \quad rim = \log IMPORT$$

앞 페이지의 모형을 추정된 계수를 대입하여, 전개한 식이 다음의 식이다.

$$\begin{aligned}
2. \text{rexch}_t &= 1.39\text{rexch}_{t-1} - 0.22\text{rexch}_{t-2} + 0.61\text{rexch}_{t-3} \\
&\quad - 1292\text{rex}_{t-1} + 139.4\text{rex}_{t-2} + 11252.6\text{rex}_{t-3} - 442.74_{t-4} \\
&\quad + 442.74\text{rex}_{t-5} + 1292\text{rex}_{t-13} + 139.4\text{rex}_{t-14} - 11252.6\text{rex}_{t-15} \\
&\quad + 442.74\text{rex}_{t-16} - 442.74\text{rex}_{t-17} \\
&\quad - 671.4\text{rim}_{t-1} + 262.36\text{rim}_{t-2} + 1171.38\text{rim}_{t-3} - 297.9\text{rim}_{t-4} \\
&\quad - 464.44\text{rim}_{t-5} + 671.4\text{rim}_{t-13} - 262.36\text{rim}_{t-14} - 1171.38\text{rim}_{t-15} \\
&\quad + 297.9\text{rim}_{t-16} + 464.44\text{rim}_{t-17} \\
&\quad + a_t - 0.32a_{t-1} - 0.57a_{t-2}
\end{aligned}$$

위에서 볼 때, 우리 나라의 원화 환율은 끼치는 영향 중에, 환율은 1달전 환율이 가장 큰 영향을 미치고, 수출액이 미치는 영향 중에, 1달전, 4달전 수출액은 환율을 감소시키고 있다. 여기서, 주의해서 보아야 할 부분은 수입액이 환율에 미치는 영향인데, 1달전 수입액은 이론에 따르지 않고, 환율을 감소시키는데(평가절상), 3달전 수입액이 아주 큰 영향으로 환율을 증가(평가절하)시켜, 결국엔 수입액이 환율을 증가시키게 된다.

이는 경제학 이론이 주장하는 바와 일치하며, 참고로, 3달전 수출액이 수입액보다 원화 환율에 미치는 영향이 작은가는 우리 나라가 원화 환율에 대한 국제 경제에서의 영향력이 약하다는 추측을 한편 해볼 수 있다.

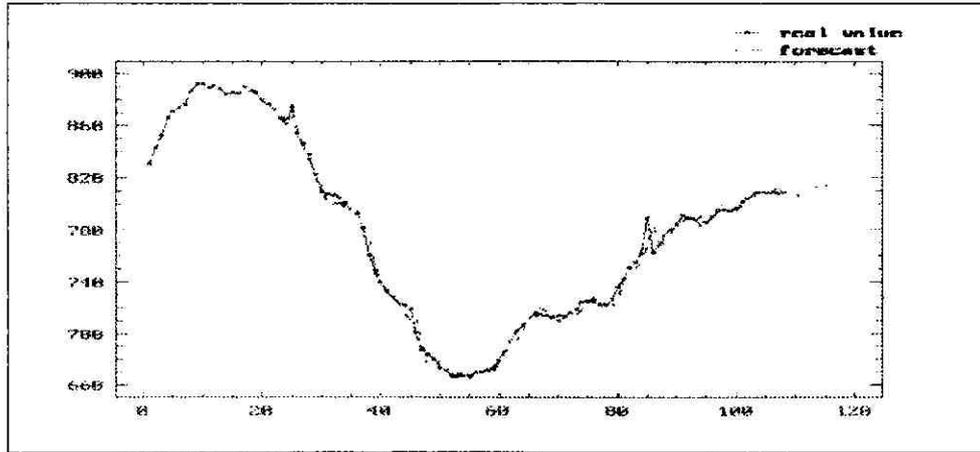
## 11. 예 측

[표 25] 실제값과 예측치와의 비교

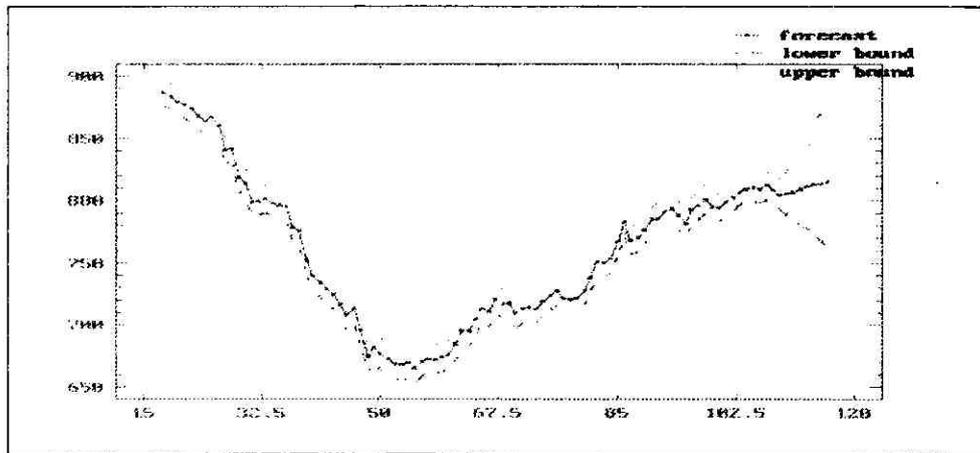
연도	관측치	예측치	예측 오차	95% 신뢰구간
94년 1월	808.1	804.660	3.440	793.486 ~ 815.752
94년 2월	808.0	805.464	2.536	788.292 ~ 822.444
94년 3월	806.5	807.209	-0.709	784.135 ~ 829.937
94년 4월	807.5	809.345	-1.845	781.043 ~ 837.129
94년 5월	806.1	811.625	-5.525	777.446 ~ 845.055
94년 6월	805.5	813.214	-7.714	773.319 ~ 852.094
94년 7월	802.6	813.862	-11.262	767.749 ~ 858.625
94년 8월	801.1	814.918	-13.818	762.448 ~ 865.648

위의 표는 94년 1월부터 8월까지 예측된 값을 실제 관측치와 비교한 것이다. 예측 오차가 2~3 원 정도이고, 이는 수출액과 수입액의 예측치로 환율을 계산하므로 발생한 것이라고 추측된다. 또, 장기간의 예측은 상당히 부정확한 것으로 나타나는 데, 이는 최근의 자료에 가장 큰 영향을 받는 시계열자료를 예측하는데, 예측값을 바탕으로 또다른 예측을 하는 것이기 때문에, 장기간이 될 수록, 예측의 신뢰도가 많이 감소되게 된다.

[그림 19] 실제값과 예측값과의 비교



[그림 20] 실제값과 예측값의 95% 신뢰구간



## 12. Program

[표 26] 최종 SAS 프로그램

```
option nodate ps=60;

data exch; infile "c:\yi\rate.dat"; input exchange @@; time=_n_;
    rexch=exchange**(1.5319); exch1=dif(rexch); exch11=dif(exch1);
data export; infile "c:\yi\ex.dat"; input export @@;
    rex=export**(0.2136); rex1=dif(rex); rex112=dif12(rex1);
data import; infile "c:\yi\im.dat"; input import @@;
    rim=log(import); rim1=dif(rim); rim112=dif12(rim1);
data total; merge exch export import;
    s1=mod(time,12); if s1=10 then s='a'; else if s1=11 then s='b';
    else if s1=0 then s='c'; else s=s1; drop s1;

proc plot;
    plot (rexch exch1 exch11)*time=s;
    plot (export rex rex1 rex112)*time=s;
    plot (import rim rim1 rim112)*time=s;
run;

proc arima data=total;
    identify var=rex(1,12);
    estimate q=(1)(12) noconstant;
    identify var=rim(1,12);
    estimate q=(1)(11 12) noconstant;
    identify var=rexch(1,1) crosscor=(rex(1,12) rim(1,12));
    estimate input=(1$(1,2,3)/(1)rex 1$(1,2,3,4,5)/rim);
    estimate input=(1$(1,3)/(1)rex 1$(2)/rim) noconstant plot;
    estimate q=1 input=(1$(1,3)/(1)rex 1$(2)/rim) noconstant;
    forecast lead=8;
run;
```

### 13. Minitab 이용 내용

[표 27] Minitab을 사용한 산포 균일화 분석 내용

```

MTB > read 'c:\y1\y.dat' c1
108 ROWS READ

C1
  1    2    3    4    . . . .

MTB > read 'c:\y1\ex.dat' c2
108 ROWS READ

C2
1642.9  1804.4  2347.1  2366.0  . . . .

MTB > read 'c:\y1\im.dat' c3
108 ROWS READ

C3
2400.9  2063.2  2264.5  2499.1  . . . .

MTB > read 'c:\y1\rate' c4
108 ROWS READ

C2
 830.6  842.8  852.3  865.9  . . . .

MTB > describe c2

C2          N    MEAN  MEDIAN  TRMEAN  STDEV  SEMEAN
          108    4918   5239    4930    1561    150

C2          MIN    MAX    Q1     Q3
          1643   8098   3658   6169

MTB > let c5=c1*(8098-1643)/107
MTB > rline c2 using c5
Slope = 0.7836 Level = 2266.2717 Half-slope ratio = 0.572

MTB > describe c3

C3          N    MEAN  MEDIAN  TRMEAN  STDEV  SEMEAN
          108    4886   5082    4908    1829    176

C3          MIN    MAX    Q1     Q3
          1620   7532   3357   6574

MTB > let c6=c1*(7532-1620)/107
MTB > rline c3 using c6
Slope = 1.0473 Level = 1803.6670 Half-slope ratio = 0.695

MTB > describe c4

C2          N    MEAN  MEDIAN  TRMEAN  STDEV  SEMEAN
          108    777.61  784.30  777.47   70.76   6.81

C2          MIN    MAX    Q1     Q3
          666.30  892.20  717.05  833.22

MTB > let c7=c1*(892.2-666.3)/107
MTB > rline c4 using c7
Slope = -0.5319 Level = 843.9889 Half-slope ratio = -0.476
    
```

<참고 문헌>

- 김한수 著 『환율의 경제학』 육법사, 1985.  
 박진근 著 『국제경제』 박영사, 1986.  
 허명희 著 『탐색적 방법에 의한 통계자료분석론』 자유아카데미, 1993.  
 허명희·박유성 著 『시계열 자료분석』 자유아카데미, 1994.  
 한국은행 『조사통계연보』 1986~1994.

<표 목 차>

[표 1] 수출액(단위:백만불) .....	146
[표 2] 수입액(단위:백만불) .....	146
[표 3] 환율(₩/\$) .....	146
[표 4] 수출액의 백색 잡음(White Noise) .....	152
[표 5] 예상 모형들의 AIC와 SBC의 비교 .....	152
[표 6] ARIMA (0, 1, 1) × (0, 1, 1) <sub>12</sub> 모형의 모수 추정치 .....	152
[표 7] 상수를 제외한 ARIMA (0, 1, 1) × (0, 1, 1) <sub>12</sub> 모형의 모수 추정치 .....	153
[표 8] ARIMA (0, 1, 1) × (0, 1, 1) <sub>12</sub> 모형에서 잔차의 자기 상관성 체크 .....	153
[표 9] 수출액의 사전백색화 모수 .....	153
[표 10] 수입액의 백색 잡음(White Noise) .....	154
[표 11] 예상 모형들의 AIC와 SBC의 비교 .....	154
[표 12] ARIMA (0, 1, 1) × (0, 1, 2) <sub>12</sub> 모형의 모수 추정치 .....	155
[표 13] 상수를 제외한 ARIMA (0, 1, 1) × (0, 1, 2) <sub>12</sub> 모형의 모수 추정치 .....	155
[표 14] ARIMA (0, 1, 1) × (0, 1, 2) <sub>12</sub> 모형에서 잔차의 자기상관성 체크 .....	155
[표 15] 상수를 제외한 q=(1)(11 12) 모형의 모수 추정치 .....	156
[표 16] q=(1)(11 12) 모형에서 잔차의 자기 상관성 체크 .....	157
[표 17] 수입액의 사전백색화 모수 .....	157
[표 18] D1=1, S1=3, R1=1, D2=1, S2=5, R2=0으로 부터의 $W_{S_1}(B), W_{S_2}(B), \delta_{R_1}(B), \delta_{R_2}(B)$ 의 추정치 .....	158
[표 19] D1=1, S1=3, R1=1, D2=1, S2=5, R2=0으로 부터의 유의한 추정치 .....	159
[표 20] Nt의 모형화 판단 포트만토 검정량 .....	159
[표 21] 모형의 최종 추정치 .....	160
[표 22] 모형 진단을 위한 잔차의 포트만토 검정 통계치 .....	160
[표 23] 입력 계열과 잔차의 상관관계 검정을 위한 포트만토 검정 통계치 .....	160
[표 24] 다중 전이 함수 모형의 최종 모형 결과 .....	161
[표 25] 실제값과 예측치와의 비교 .....	162
[표 26] 최종 SAS 프로그램 .....	164
[표 27] Minitab을 사용한 산포 균일화 분석 내용 .....	165

<그림 목차>

[그림 1] 수출액의 시도표(원자료) .....	147
[그림 2] $y=x^{0.2136}$ 변환 후의 수출액의 시도표 .....	147
[그림 3] 전체적 추세를 제거한 수출액의 시도표(1차 차분) .....	147
[그림 4] 월별 추세를 제거한 수출액의 시도표(12차분) .....	148
[그림 5] 수입액의 시도표(원자료) .....	148
[그림 6] $y=\log x$ 변환을 실행한 수입액의 시도표 .....	149
[그림 7] 전체 추세를 제거한 수입액의 시도표(1차 차분) .....	149
[그림 8] 월별 추세를 제거한 수입액의 시도표(12차분) .....	149
[그림 9] 환율의 시도표(원자료) .....	150
[그림 10] 1차 차분하여 변환된 환율의 시도표 .....	150
[그림 11] 再次 1차 차분을 취하여 변환된 환율의 시도표 .....	151
[그림 12] 수출액의 ACF와 PACF .....	151
[그림 13] 수입액의 ACF와 PACF .....	153
[그림 14] ARIMA (0, 1, 1) × (0, 1, 2) <sub>12</sub> 모형에서 잔차의 ACF .....	156
[그림 15] ARIMA (0, 1, 1) × (0, 1, 2) <sub>12</sub> 모형에서 잔차의 PACF .....	156
[그림 16] 수출액과 환율의 CCF .....	157
[그림 17] 수입액과 환율의 CCF .....	158
[그림 18] 잔차의 모형 설정의 위한 ACF와 PACF .....	159
[그림 19] 실제값과 예측값과의 비교 .....	163
[그림 20] 실제값과 예측값의 95% 신뢰구간 .....	163

29

Handwritten mark

---

제1회 「통계의 날」기념논문집

발행일 : 1995. 12.

발행인 : 이 강 우

발행처 : 통 계 청

우) 136-723

서울특별시 강남구 역삼동 647-15

TEL : 222-1823~4

FAX : 538-5343

---