

## 한국 경찰 인력 증가요인의 인과성 추론을 위한 회귀모형 적용에 관한 연구

이 병 철

행정학과

(1988. 4. 30 접수)

### <要 約>

본 연구는 관련 문헌 분석을 통해 그리고 회귀모형의 통계적 이론들의 규명과 정책자료의 컴퓨터 처리 분석을 통해, 한국 경찰 인력 증가요인의 인과성 추론을 위한 회귀모형 적용상의 특성들에 대해 분석해 보려고 했다. 이러한 노력은 회귀모형의 활용에 의해 한국 경찰 인력 증가 경향의 분석을 통해, 정책학적 관점에서 한국 행정의 개별적 이론화를 규명하려 했으며 더 나아가선 미래 한국 경찰의 인력 증가 요인을 예측하려는 데 경주되었다.

한국 행정의 이론적 구성요소는 한국 행정 현상에 대한 설명이나 이해의 영역뿐 아니라 정책 분석적 맥락에서 행정적 문제를 해결하는 범주까지 확산되어져 있다. 이런 이론들에서도 출된 방법들이나 접근법들은 규범적이고 거시적인 차원으로만 국한되는 것이 아니라 분석의 실증적이고 현상론적인 양태로 활용되어 지고 있다.

이것이 의미하는 바는 명확한 규범적 분석의 논리를 강조하면서도 행정현상의 분석에 실증적이고 계량적인 방법론을 심도있게 적용한다는 것이다.

회귀모형의 한국경찰 인력 사례에 대한 적용은 이러한 계량적 접근이 행정이론의 발전에 도움이 될 수 있고 정책결정, 집행 및 평가의 문제를 해결하는 데도 조력할 수 있다는 점을 보여주고 있다.

본 연구 분석의 결과는 미래에 있어 한국 경찰 인력 증가에 영향을 끼칠 수 있는 영향력 있는 요인들로 다음과 같은 것을 제시하고 있다.

1. 교통사고수
2. 도시수
3. 예산
4. 경찰관서수
5. 자동차대수

본 연구는 또한 이러한 요인들 중 가장 주된 영향 요인은 교통사고수라는 것을 밝혀냈는데 이것이 의미하는 바는 교통사고수가 경찰인력 증가와 깊은 상관관계를 지닐 뿐 아니라 인과성 역시 지니고 있음을 의미한다. 따라서 본 연구는 치안당국이 경찰 인력 증가 요인들—특히 교통사고수—에 주된 관심을 경주할 것과 이러한 상관성과 인과성을 이해해주길 촉구하고 있으며, 관련 정책 결정자는 이 분석에 근거해 경찰 인력의 적정규모를 심사숙고해 결정하여야 됨을 권고하고 있다.

## **A Study of Increasing factors in the police manpower of Korea.**

Lee, Byung-Chul  
 Dept. of Public Administration  
 (Received April 30, 1988)

### **<Abstract>**

The purpose of this study is to analyze the characteristics of the application of regression model to causal relationship inferences of police manpower increasing factors in Korea through a survey of the literature, leading to identification of statistical theories of regression model, and to computer processing analysis of policy raw data. This effort to identify an ideographic theory of Korean public administration, especially, in the view of the policy sciences, through the analysis of Korean police manpower increasing trends, by utilizing regression model, extends its focus to forecasting the future Korean police manpower increasing factor.

The theoretical component of public administration in Korea deals not only with Korean public administration phenomena but also with solving some administrative problems in the context of policy analysis.

The methods or approaches to derive such theories are limited to a normative dimension but utilize empirical and phenomenological modes of analysis. It means that empirical quantitative methods are heavily applied in analyzing administrative phenomena along with emphasizing clear logic of normative analysis.

An application of regression model to Korean police manpower case shows that these quantitative approaches will be helpful to develop the theory of public administration and solve the problem of policy making, implementation, and evaluation.

The results of this analysis indicate that the influential factors expected to affect Korean police manpower increasing in the future are:

- 1 . the number of traffic accidents
- 2 . the number of cities
- 3 . police budget
- 4 . the number of police boxes
- 5 . the number of autocars

The study also predicts that the most powerful influential factor is the number of traffic accidents.

This means that the number of traffic accidents deeply correlates not only the police manpower increase but also causally relates it.

Therefore, This study recommends that police authority concerns the police manpower increasing factors, especially, the number of traffic accidents, and understand this correlation and causal relationship.

The study also prescribes that policy maker decides sincerely the optimal size of police manpower according to this analysis.

## 1. 연구의 필요성

허만 칸(Herman Kahn)<sup>1)</sup>이나 다니엘 벨(Daniel Bell)<sup>2)</sup> 등이 서기 2000년대 이후에 도래할 후기산업사회모형(post-industrial society model)을 제시한 바 있고 토플러(Alvin Toffler)<sup>3)</sup>나 네스비츠(John Naisbitt)<sup>4)</sup> 등도 미래에 닥쳐 올 사회적 변화와 기술혁신<sup>5)</sup>에 대해 언급하고 있듯이, 앞으로 도래될 사회적 변화는 필연적이며, 이에 인류는 적극적으로 능동적인 입장에서 이의 변화와 도전을 받아 들이고 이에 슬기롭게 응전해 나가야 되리라 믿는다.<sup>6)</sup>

이미 옛적에도 미래에 대한 인간적 호기심과 외경심은 주술, 점성술, 천문학들의 발전을 촉발시켰고 그들이 예언한 내용은 비록 주먹구구식의 직관이나 영감 또는 감각에 의한 것이긴 했으나 일면 과학적인 조사방법과 분석기법이 사용된 면도 없지 않았다.<sup>7)</sup> 특히 예를 들어 점성술에서 달이 만월이 되면 사람들이 평상시보다 더 흥분이 되고 감정이 격앙이 되어 다툼이나 살인이 일어날 수 있다는 지금까지의 경험적, 실증적 자료에 근거해, 만월시에 재앙이 닥칠거라는 예언들은 비록 현대적 감각에선 우스꽝스럽지만 그래도 그들 나름의 체계적인 경험에 입각한 통계적 확률의 원시적 적용이란 측면에선 과학적 접근의 시도였다고 이해할 수 있는 것이다.

이렇듯 인간은 과거와 현재의 집적된 경험과 지식을 근거로 하여 앞으로 도래할 미지의 세계를 예견하려고 노력해 왔으며 이것은 현대와 같은 불확실성의 시대<sup>8)</sup> 또는 비연속적 사회<sup>9)</sup>에선 더욱 그 필요성이 절실해 지고 있다고 볼 수 있다.

보다 치열해지는 국제간의 정치, 경제, 군사적 경쟁과 국내의 정치적 소요, 경제불안, 사

1) Herman Kahn, *The Next 200years*(New York: William Morrow & Co., 1976), pp. 49-57.

2) Daniel Bell, *The Coming of Post-Industrial Society: A Venture in Social Forecasting*(New York: Basic Books Inc., 1976), pp. 3-6.

3) Alvin Toffler, *Future Shock*(New York: Bantam Book, 1970).

4) John Naisbitt, *Megatrends*(New York: William Morrow and Company, Inc. 1980).

5) Thomas E Jones, *Megatrends*(New York: Warner Books, Inc., 1982), p. 1.

6) Zbigniew Brzezinski, *Between Two Ages: America's Role in the Technetronic Era*(New York: The Viking Press, 1970).

7) Arnold J. Toynbee, *A Study of History*(London: Oxford University Press, 1973).

8) Thomas E Jones, *Options for the future*(New York: Praeger Publishers, 1980)

Richard D. Bingham, et. al.,(eds), *Reaching Decisions in Public Policy and Administration;Methods and Applications*(New York: Longman Inc., 1982).

9) John Kenneth Galbraith, *The Age of Uncertainty*(Boston, Mass.: Houghton Mifflin Co.,1977).

9) Peter F. Drucker, *The Age of Discontinuity: Guidelines to Our Changing Society*(N. Y.: Harper and Row, 1969)

회적 동요는 바로 현시대의 위기관리 능력을 시험하는 시금석이 되고 있고 바로 이러한 맥락에서 미래예측에 대한 간절한 회구는 그 존재의 의의를 갖게 되는 것이다.

폭발적으로 증대되는 사회적 제반문제와 욕구를 해결하기 위해 정부기능은 비대화해지고 이에 따른 구조적 분화는 행정부를 위시한 정부 통치기구의 확대를 부채질시켰고, 이는 행정 및 관료제의 비대화와 역기능을 수반케 했다.<sup>10)</sup> 이러한 시점에서 세계적 추세이기도 하지만 한국적 특수상황에 근거해 도시화가 급속히 산업, 경제발전에 힘입어 진행되었고,<sup>11)</sup> 이러한 도시화가 지니는 야누스적 양면성으로 인해 도시화의 어두운 면이 환경 및 생태계를 파괴시키고 전통적 가족제도를 붕괴시키고, 인구문제, 주택문제, 교통문제들을 유발시켰으며 급기야는 인간적 소외와 몰인정적 사회로 탈바꿈하게 됨으로 해서 사회적 범죄나 비행이 그 어느 때보다 기승을 부리게 되었다.<sup>12)</sup>

국민에 대한 봉사를 행정이념으로 하고 있는 국가는 이를 더 이상 방치할 수 없게 되었으며 적극적인 개입이 불가피해 졌고 이러한 복지국가적 관점에서의 정부의 개입은 치안력의 강화를 수반케 했다. 치안력은 경찰인력, 물자, 예산, 관리능력, 정보처리, 통제능력 등이 함께 어우러져서 만들어 낼 수 있는 합작품이긴 하지만, 여기에서 가장 주가 되는 것은 물자, 예산, 관리, 정보처리, 통제의 주체인 경찰인력인 것이다.

사회, 경제적 발전과 도시화의 가속화는 따라서 유능한 경찰인력을 필요로 하게 되고 범위의 광역화는 양적, 물적인 면에서 경찰인력을 소요하고 있는 것이다. 그렇다고 무턱대고 경찰인력을 증대하는 것은 국가예산의 낭비일 뿐 아니라 치안국가<sup>13)</sup>로의 전략을 의미하기도 하기 때문에 적절한 수준의 경찰인력을 산정하는 것이 무엇보다 필요하며 이를 위해선 미래의 경찰인력수급상황을 예측하고 인력증가요인의 상호관계 그리고 인력증가와 그 요인과의 인과관계 파악이 선행되어야 하는 것이다.

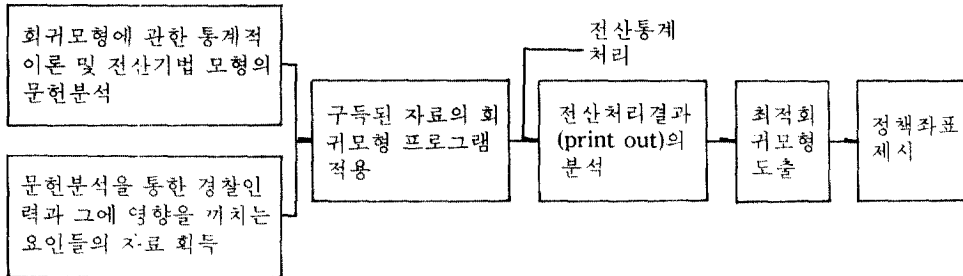
- 
- 10) Robert K. Merton, *Social Theory and Social Structure*, (Glencoe, Ill: The Free Press, 1957), pp. 195-206.  
 Philip Selznick, *TVA and the Grass Roots* (Berkeley, CA.:University of California Press, 1968.)  
 Alvin W. Gouldner, *Patterns of Industrial Bureaucracy*(Glencoe, Ill.: Free press, 1954).  
 Peter M. Blau, *The Dynamics of Bureaucracy* (Chicago, Ill.: University of Chicago Press, 1955).
- 11) 이성복, 「도시행정론 : 한국의 도시를 중심으로」 (서울 : 법문사, 1987), pp. 55-66.
- 12) 이성복, 「상계서」, pp. 72-75.  
 P. Sorokin and C. C. Zimmerman, *Principles of Rural-Urban Sociology*(New York: Holt, 1929).  
 김대환, “도시개발과 시민의식구조,” 최상철 외(공편), 「한국도시개발론」 (서울 : 일지사, 1981), p. 68.  
 R. J. Johnston, *The American Urban System: A Geographical Perspective* (New York, London: St. Martin's Press, Inc., 1982), pp. 304-322.
- 13) Amos Perlmutter, *The Military and Politics in Modern Times*(New Haven: Yale University Press, 1977)  
 Harold D. Lasswell, “The Garrison State,” *American Journal of Sociology*, 46(January 1941), pp. 455-468.  
 Samuel A. Stouffer et al., *The American Soldier* (Princeton, N. J.: Princeton University Press, 1949)  
 Samuel P. Huntington, *The Soldier and the State*(Cambridge, Mass.: Harvard Univ. Press), 1957)  
 Morris Janowitz, *The Professional Soldier* (Chicago: Free Press, 1960)  
 Carl J. Friedrich & Z. K. Brzezinski, *Totalitarian Dictatorship and Autocracy*(Cambridge, Mass: Harvard Univ. Press, 1956)

## 2. 연구의 목적

지금까지 논의한 사실에 의해 경찰인력의 증가가 일단 필요시된다는 논리적 연상선위에서 과연 경찰인력의 증가요인은 무엇이고 증가는 어떠한 방법으로 실현시키는 것이 가장 바람직할 것인가에 대한 해답을 구하는 것이 본 연구의 기본적 목적이라 하겠다. 경찰인력의 증가요인으론 전국의 도시수, 예산, 경찰관서수, 교통사고수, 차량대수, 범죄율 등과 같은 것들을 들 수 있는데 이러한 요인들은 서로간에 영향을 주고 있으며 또한 복합적인 작용을 통해 경찰인력증가에 영향을 주고 있기 때문이다.<sup>14)</sup> 따라서 본 연구는 이러한 각개요인들이 경찰인력 증가에 어떻게 영향을 끼치는가를 알고, 각 요인과 경찰인력 증가와의 상관관계 및 인과관계를 지득키 위해 회귀모형(regression model)을 사용할 것이며 여기에서 도출된 결과를 미래예측적 관점에서 분석, 평가하여 향후 경찰인력 수급에 관한 정책적 좌표로 활용될 수 있도록 하는 정책제언을 하려는데 부차적인 연구목적이 있는 것이다.

## 3. 연구방법

경찰인력증가에 영향을 주는 각 요인들에 대한 통계자료와 분석기법에 대한 통계적 이론 및 전산기법 모형은 관련문헌분석을 통해 획득토록 하며 전체의 연구 분석들은 회귀모형에 대한 이론적 접근을 기존의 학설과 기법에 근거해 개관하고 이러한 이론적 준거들에 근거하여 한국적 사례인 한국경찰인력의 증가추세와 그에 영향을 끼치는 요인들과의 관계성, 인과성을 연구설계에 따라 회귀분석을 통해 얻도록 하고 도출된 최적의 회귀모형에 입각하여 정책좌표를 제시한다. 이를 알기 쉽게 도해하면 다음과 같다.



(그림 1) 연구분석틀

## 4. 회귀모형의 이론적 접근

### (1) 회귀모형의 목적과 내용

한 변인을 다른 변인으로부터 예언하는 문제를 회귀(regression)의 문제라고 부른다. 이 용

14) David Nachmias, *Public Policy Evaluation: Approaches and Methods*(New York: St. Martin's Press; 1979), pp. 9-12.

어는 갈튼(Francis Galton)경의 연구로부터 유래하는데 그는 19세기에 여러가지 신체 특징과 행동 특징사이의 관계성을 연구하였다. 한 특성을 다른 것으로부터 예언하려는 노력에서, 그는 키 큰 아버지의 아들들은 키가 컸지만, 아버지들의 평균보다 더 크지는 않았다는 것을 관찰하였다. 역시 키 작은 아버지의 아들들은 작지만, 아버지들 보다는 평균에 더 가까웠다. 그러한 평균으로의 경향을 평균치로의 회귀(regression toward the mean)라고 부르는데, 그것은 예언하는데서 극단으로 떨어져 있는 사람은 예언되는 특성에 있어서 덜 극단적으로 될 것이라고 보는 짐작이 맞게 된다는 사실을 가리킨다. 어떤 특징들은 이 원리를 따르고 다른 것들은 그렇지 않지만, 대부분의 경우에 그것은 훌륭한 예언전략이 된다.<sup>15)</sup> 왜냐하면 한 변인의 흐름을 이해하고 이에 따른 다른 변인의 상태를 예측할 수 있기 때문인데, 이것은 특히 과거와 현재의 흐름을 가지고 미래를 예견해 볼 수 있다는 데서 훌륭한 미래예측의 한 방법이 될 수 있는 것이다.

예를 들어 소득(X)이 올라가게 되면 자동차 보유대수(Y)가 그만큼 늘어날 것인가를 알기 위해 회귀모형<sup>16)</sup>을 사용할 수 있는 것이다.  $t_0$ 의 시점에서 소득이 2,000불이고 자동차보유대수가 100대였다면  $t_1$ 의 시점에서 소득이 3,000불로 늘어났다면 자동차 보유대수는 얼마나 될까? 바로 이와같은 논리적 추론은 미래예측의 한 접근법이 될 수 있는 것이다.

이러한 두 변인간의 관계가 상호적인 의미를 갖기 위해선 선형적 관계성(linear relationship)을 가져야 한다. 이것의 의미는 도표에 일직선으로 정확하게 표시될 수 있는 두 변인들 사이의 관계성을 의미한다. 물론 회귀모형에서 다루는 모든 관계성이 직선적인 것만은 아니다. 일부는 비직선적(non-linear)이며 경우에 따라선 곡선적(curvilinear)관계성을 지닐 수 있다.<sup>17)</sup> 이것을 구하는 방법은 대수(logarithm)를 사용하기도 하지만 일반적인 회귀모형은 선형적 관계성에 그 기반을 둔다.

그 다음 고려해야 할 것은 이 직선의 기울기인데 이것은 -1에서 1까지의 범위 내에서 정해진다. -1에서 0까지는 음의 기울기를 갖는 것으로 부정(負的: negative) 혹은 반비례적 관계성(inverse relationship)을 지니는 것이 되고 0에서 1까지가 되면 양의 기울기를 갖는 정적(正的: positive) 혹은 정비례관계성(direct relationship)을 지니는 것이 된다.<sup>18)</sup>

### 1) 단순회귀모형

일반적으로 단순회귀의 경우 회귀방정식은 다음과 같이 표시된다.

$$\hat{Y}_i = \alpha + \beta x_i + \varepsilon_i, \quad i=1, 2, \dots, n \quad \varepsilon_i \overset{i.i.d.}{\sim} (0, \sigma^2)$$

변인  $\hat{Y}$ 는 예언된 Y(Y predicted)라고 불리우며  $\alpha$  값(Y절편)과  $\beta$  값(기울기)은 회귀상수

15) Robert B. McCall, *Fundamental Statistics for Psychology*(New York: Harcourt Brace Jovanovich, Inc. 1975)

이관용·김기중(역), 「기초심리 통계학」(서울:법문사, 1984), pp. 109-110.

16) 회귀모형은 확정적 모형(deterministic model)이 아닌 확률적 모형 또는 통계적 모형(statistical model)이다. 전자는 독립변수에 의해 종속변수가 완전히 설명되어지는 모형을 말하며, 후자는 독립변수로서 설명이 되지 않는 종속변수가 일부 존재하는 모형이다. 따라서 회귀모형은 미래를 확률적이나 통계적으로 예측하는 것이지 예언이나 단언을 하는 것은 아니다.

17) 이관용·김기중(역), 「상계서」, p. 111.

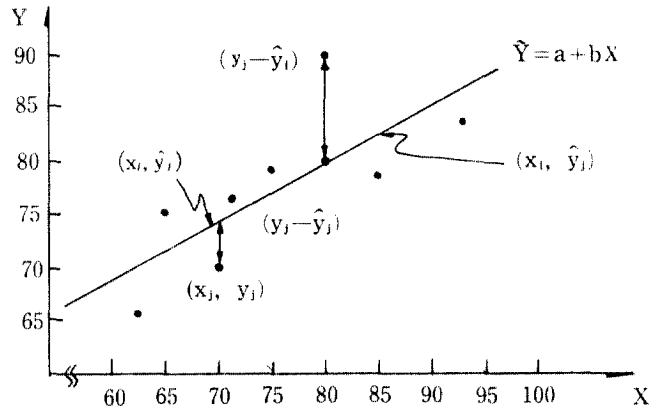
18) 「상계서」, p. 114.

(regression constant)라고 부른다.<sup>19)</sup> 보다 부연하면  $\alpha$  값은 회귀모수(parameter)<sup>20)</sup>,  $\beta$  값은 회귀계수(regression coefficient),  $\varepsilon$  는 오차항(error term)이라 한다.

회귀모형분석에서 중요한 과제는 그 관계성을 가장 잘 기술하는 회귀선의 회귀상수들을 상술하는 것이다. 이를 다른 말로 하자면 선형모형을 추정하는 것이다. 이것이 추정되면 추정된  $\alpha$  와  $\beta$  에 대한 통계적 추론을 위하여 이를 각각의 분포와 특히 표준오차(standard error of coefficient)를 구해야 하는 동시에 추정된 단순 회귀모형을 종합적으로 평가하기 위하여 예측치의 표준오차(standard error of estimate), 결정계수(coefficient of determination) 그리고 분산분석(analysis of variance)을 위한 표의 작성 이외에도 표준화변수에 의한 회귀계수 B와 탄력도(elasticity) 등을 계산한다.<sup>21)</sup>

선형모형의 모수  $\alpha$  와  $\beta$  를 주어진 자료에 의해서 추정하는 방법은 최방추정법(maximum likelihood estimation)과 최소자승법(least square method)이 있으나 양자의 추정결과는 같다.<sup>22)</sup> 다만 방법적용에 있어 이론적 배경만 달리 할 뿐이다.

최소자승법은 (그림 2)에서와 같은 점산도위에서 각점을 가장 가깝게 통과하는 직선방정식을 추정하는 방법을 말한다. 그 직선이 각점을 가장 가깝게 통과하는 직선이라면 그 직선과 각점과의 수직적 편차의 제곱의 합계가 최소이어야 하는 조건이 충족되어야 한다.



(그림 2) 회귀선

19) 「상계서」, p. 118.

20)  $\alpha$  값은 절편(intercept)으로도 불려진다. 김병수 외, 「SPSS를 이용한 통계자료분석」(서울: 박영사, 1987), p. 208. 이관용 외(역), 「기초심리통계학」(서울: 법문사, 1984), p. 118.

21) 김병수 외, 「상계서」, p. 209.

22) 관측치  $Y_i$ 와 표본자료를 통해 얻어진 표본회귀방정식의 추정치  $\hat{Y}_i$ 간에 오차가 생기게 되는데 이를 잔차(residual)  $e_i$ 라고 하며  $e_i = Y_i - \hat{Y}_i$ 로 표시한다. 회귀모형의 목적은  $e_i$ 를 가장 작게 하는데 있으므로,  $e_i$ 를 최소화하는 방안으로 다음의 두가지 방법이 거론된다. 첫째 방법은  $e_i$ 의 절대값을 극소화하는 기준으로서,  $e_i$ 의 부호를 무시하고 그 절대값들의 총합을 가장 적게 하는 추정선을 구하는 방법이다.

즉 점산도위의 각 점을  $(X_i, Y_i)$ 라 하고 우리가 찾으려는 직선상에서  $X_i$ 에 대응하는 점을  $(X_i, \hat{Y}_i)$ 라 할 때 가장 공평한 직선  $\hat{Y}_i = a + bX_i$ 는 다음 조건을 충족시키는  $a$ 와  $b$ 에 의해서 얻어진다.

그 조건은 최소자승법에 따라  $\sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - a - bX_i)^2$ 가 최소가 되는 것이다.

수직적 편차의 제곱의 합계가 최소로 하는 조건에서의  $a$ 와  $b$ 값은 다음과 같이 얻어진다.<sup>23)</sup>

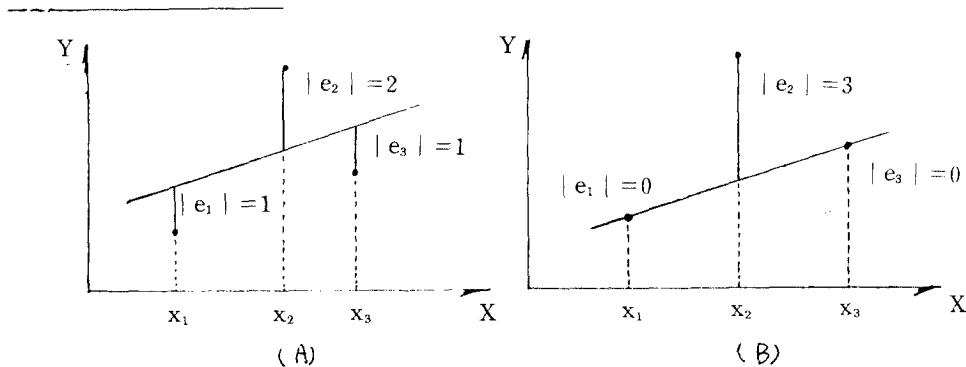
$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

$$a = \frac{\sum Y}{n} - b \frac{\sum X}{n}$$

$X$  및  $Y$ 의 평균을  $\bar{X}$  및  $\bar{Y}$ 로 하고 변수를 편차로 바꾸면  $a$ 와  $b$ 는 다음과 같이 얻을 수도 있다.

$$b = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

$$a = \bar{Y} - b\bar{X}$$



(A)의 경우  $\sum |e_i| = 4$ 이고, (B)의 경우는  $\sum |e_i| = 3$ 이다. 즉 (B)의 경우가 보다 나은 추정선을 제시하고 있지만, (B)도  $X_2$ 의 정보를 제대로 활용하고 있지 못하다. 이상을  $\text{Minimize } \sum |Y_i - \hat{Y}_i|$  또는  $\text{Minimize } \sum |e_i|$ 로 나타낼 수 있다. 또 한가지 방법으로  $e_i$ 의 제곱의 총합을 극소화하는 추정선을 구하는 이른바 최소자승법이 있는데 이의 수식은  $\text{Minimize } \sum (Y_i - \hat{Y}_i)^2$  또는  $\text{Minimize } \sum e_i^2$ 이다. 자세한 설명은 본문에 되어 있으므로 여기서는 생략한다.

23) 오차폭인  $\sum e_i^2$ 를  $E$ 라 하면  $E$ 는  $\sum (Y - a - bx)^2$ 이 되고  $E$ 를 최소화시키는  $a$ ,  $b$ 의 값을 구하기 위해  $E$ 를  $a$ ,  $b$ 로 미분하면  $a$ 와  $b$ 는 본문의 공식처럼 얻을 수 있게 된다. 이를 이해의 편의를 위해 실제자료에 넣고 계산으로 유도해 보자.

(표 A)

소득(천원)	312	164	280	196	200	288	146	361	149	252
연령(세)	48	32	40	34	30	50	26	50	22	43

위 (표 A)에서 종속변수는 소득이 되고 독립변수는 연령이 된다. 공식에 넣어 계산을 용이하게 하기 위해 위의 (표 A)를 조작해 보면 (표 B)와 같이 된다. 이를 앞의  $a$ ,  $b$  공식에 대입하면  $b = \frac{10 \cdot 94448 - 375 \cdot 2348}{10 \cdot 14993 - (375)^2} = 6.88$ ,  $a = \frac{2348 - 6.88 \cdot 375}{10} = -23.2$ 이 나오며 따라서 구하는 회귀방정식은  $Y = -23.2 + 6.88X$ 가 된다.



위의 회귀계수는 상관계수와 직접적인 관계가 있다. 만약에 상관계수가 0이라면 회귀계수도 0이 될 것이고 또 상관계수의 부호는 회귀계수의 부호와 같아진다. 그것은 다음과 같은 관계 때문이다.

$$r_{xy} = b \frac{S_x}{S_y} \quad \therefore b = r_{xy} \frac{S_y}{S_x}$$

식  $Y_i = \alpha + \beta x_i + \epsilon_i$ 에서  $S_x$ 와  $S_y$ 는 X와 Y 각각의 표준편차이다.  $S_x$ 와  $S_y$ 는 비부의 값이므로 b의 부호는  $r_{xy}$ 의 부호에 의존하며 또 그 크기도 영향을 받는다.

## 2) 다중회귀모형<sup>24)</sup>

앞의 경우는 설명변수가 하나일 때의 회귀방정식의 도출인데 설명변수가 두개이거나 그 이상인 경우가 많다. 이를 위해서 사용되는 기법이 다중회귀분석(multiple regression analysis)이다.

이것은 종속변수에 영향을 미치는 설명변수들이 2개이상인데 이것의 관계를 함수형태로 표시하면 다음과 같다.

$$Y = f(d, w, t)$$

(표 B)

Y	X	XY	X <sup>2</sup>	Y <sup>2</sup>
312	48	14976	2304	97344
164	32	5248	1024	26896
280	40	11200	1600	78400
196	34	6664	1156	38416
200	30	6000	900	40000
288	50	14400	2500	82944
146	26	3796	676	21316
361	50	18050	2500	130321
149	22	3278	484	22201
252	43	10836	1849	63504
Σ 2348	Σ 375	Σ 94448	Σ 14993	Σ 601342

24) 이것을 이해하기 위해서 필요한 참고서적은 다음과 같은 것들이 있다.

Norman H. Nie, et al., *Statistical Package for the Social Sciences*, 2nd ed.(New York: McGraw-Hill, 1976).  
Hadlai C. Hull, et al., *SPSS update, New Procedure and Facilities for Releases 7 and 8*(New York: McGraw-Hill, 1979).

J. P. Guilford and B. Fruchter, *Fundamental Statistics in Psychology and Education*, 5th ed(New York: McGraw-Hill, 1973), ch. 15.

Fred N. Kerlinger, *Foundations of Behavioral Research*, 2nd ed(New York: Holt, Rinehart and Winston, 1973), ch. 35.

N. R. Draper and H. Smith, *Applied Regression Analysis*(New York: Wiley, 1966).

R. A. Gordon, "Issues in Multiple Regression," *American Journal of Sociology*, Vol. 73, 1968, pp. 592-616.

Fred N. Kerlinger and E. J. Pedhazur, *Multiple Regression in Behavioral Research*(New York: Holt, Rinehart and Winston, 1973).

J. E. Overall and C. Klett, *Applied Multivariate Analysis*(New York: McGraw-Hill, 1973)

따라서 이것도 단순회귀와 같은 원리로 궁극적으로는 Y와 설명변수 d, w, t간의 관계를 나타내주는 함수 f를 찾는 것이 목적이 된다. 그러나 이러한 함수의 형태는 예외적인 몇가지를 제외하고는 대부분 우리에게 알려져 있지 않고 있다. 반면에 많은 경우 경험적으로 선형 함수가 알려져 있지 않은 함수 형태에 근사한 함수가 된다는 것을 알고 있다. 따라서 Y=f(d, w, t)의 관계에서 Y를 d, w, t의 선형함수로 표시할 수 있는 부분과 그러한 선형함수로서 표시할 수 없는 부분으로 나누어 보고 후자를 오차항으로 부른다. 즉 다음의 관계를 가정해 본다.

$$Y = \beta_0 + \beta_1 d + \beta_2 w + \beta_3 t + \epsilon$$

$\underbrace{\hspace{10em}}_{\text{d, w, t의 선형함수 } Y=f(d, w, t)\text{에서 선형함수가 설명하지 못하는 부분}}$

위 식과 같은 경우가 다변수 회귀분석모형(multiple linear regression)의 한 예가 된다. 즉 일반적으로 설명변수가 p-1개 있고 반응변수 Y는 p-1개 설명변수의 선형함수와 오차항의 합으로 표시된다. 또한 오차항  $\epsilon$ 은 직접 관찰하지 못하는 확률항(stochastic term)으로서 평균이 0이고 분산  $\sigma^2$ 를 가지고 있다고 가정하여 다음의 모형을 세운다.<sup>25)</sup>

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1} + \epsilon_i \quad \epsilon_i \sim (0, \sigma^2)$$

$X_1, \dots, X_{p-1}$ 은 비확률항

위 식처럼 표현될 수 있는 모형을 다변수 회귀분석모형이라 부른다.

여기에서 논의의 편의상 설명변수가 두개인 것만을 상정하여 다중회귀방정식을 구하는 공식을 제시하면 다음과 같다.<sup>26)</sup>

25) 다중회귀모형에서 모수인  $\beta_0, \beta_1, \dots, \beta_{p-1}$ 에 대한 추정은 단순회귀분석에서와 같이 최소자승법을 사용할 수 있다. 즉 오차항의 제곱의 합인  $\sum_{i=1}^n \epsilon_i^2$ 를 최소로 해주는  $\beta_0, \beta_1, \dots, \beta_{p-1}$ 을 구하는 방법인데, 최소자승법에 의한  $\beta_0, \beta_1, \dots, \beta_{p-1}$ 의 추정량  $b_0, b_1, \dots, b_{p-1}$ 은 다음 연립방정식의 해로서 얻어진다.

$$\begin{aligned} \sum Y_i &= n\beta_0 + \beta_1 \sum x_{i1} + \dots + \beta_{p-1} \sum X_{i,p-1} \\ \sum X_{i1} Y_i &= \beta_0 \sum X_{i1} + \beta_1 \sum x_{i1}^2 + \dots + \beta_{p-1} \sum X_{i1} X_{i,p-1} \\ \sum X_{i2} Y_i &= \beta_0 \sum X_{i2} + \beta_1 \sum X_{i1} X_{i2} + \dots + \beta_{p-1} \sum X_{i2} X_{i,p-1} \\ &\vdots \\ \sum X_{ip} Y_i &= \beta_0 \sum X_{ip} + \beta_1 \sum X_{i1} X_{ip} + \dots + \beta_{p-1} \sum X_{i,p-1} X_{ip} \end{aligned}$$

이를 통해 최소자승추정량인  $b_0, b_1, \dots, b_{p-1}$ 이 얻어지면 (표본)회귀방정식은 다음과 같이 얻어진다.

$$\hat{Y}_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \dots + b_{p-1} X_{i,p-1}$$

26) 이의 이해를 위해 실제자료를 넣고 계산으로 유도해 보자.

(표 C)

자동차보유 대수(Y)	2	3	2	5	5	7	5	6	7	8
인구(X <sub>1</sub> ) (10만)	1	2	2	3	4	5	6	8	9	10
종합운동장 존재여부(X <sub>2</sub> )	0	0	0	1	0	1	0	1	1	1

공식에 넣어 계산을 용이하게 하기 위해 (표 C)를 조작해 보면 (표 D)와 같이 된다.

$$Y = a + b_1 X_1 + b_2 X_2$$

$$b_1 = \frac{(\sum X_1 y)(\sum X_2^2) - (\sum X_1 X_2)(\sum X_2 y)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2}$$

$$b_2 = \frac{(\sum X_1^2)(\sum X_2 y) - (\sum X_1 X_2)(\sum X_1 y)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2}$$

$$a = \frac{\sum Y - b_1 \sum X_1 - b_2 \sum X_2}{N}$$

여기에서  $X$ ,  $Y$ 는 실제값을 의미하고  $x$ ,  $y$ 는 계산되어 얻은 값을 의미한다.

$$\sum x_1^2 = \sum X_1^2 - \frac{(\sum X_1)^2}{N}$$

$$\sum x_2^2 = \sum X_2^2 - \frac{(\sum X_2)^2}{N}$$

$$\sum x_1 x_2 = \sum X_1 X_2 - \frac{(\sum X_1)(\sum X_2)}{N}$$

$$\sum x_1 y = \sum X_1 Y - \frac{(\sum X_1)(\sum Y)}{N}$$

$$\sum x_2 y = \sum X_2 Y - \frac{(\sum X_2)(\sum Y)}{N}$$

$$\sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{N}$$

## (2) 회귀모형의 평가기준

앞에서 제시된 절차에 따라 회귀모형이 얻어지면 이것을 4가지 관점에서 평가하게 되는데, 그 자세한 통계적 이유나 이론적 의미는 논의의 성격상 생략하고 내용만을 간단히 서술하면

(표 D)

Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>1</sub> <sup>2</sup>	X <sub>2</sub> <sup>2</sup>	X <sub>1</sub> X <sub>2</sub>	X <sub>1</sub> Y	X <sub>2</sub> Y	Y <sup>2</sup>
2	1	0	1	0	0	2	0	4
3	2	0	4	0	0	6	0	9
2	2	0	4	0	0	4	0	4
5	3	1	9	1	3	15	5	25
5	4	0	16	0	0	20	0	25
7	5	1	25	1	5	35	7	49
5	6	0	36	0	0	30	0	25
6	8	1	64	1	8	48	6	36
7	9	1	81	1	9	63	7	49
8	10	1	100	1	10	80	8	64
Σ 50	Σ 50	Σ 5	Σ 340	Σ 5	Σ 35	Σ 303	Σ 33	Σ 290

이를  $a$ ,  $b_1$ ,  $b_2$  공식에 대입하면 (이때  $\sum X_1^2=90$ ,  $\sum X_2^2=2.5$ ,  $\sum X_1 X_2=10$ ,  $\sum X_1 Y=53$ ,  $\sum X_2 Y=8$ ,  $\sum Y^2=290$ 도 공식에 따라 먼저 계산한다.)  $b_1 = \frac{(53)(2.5) - (10)(8)}{90 \times (2.5) - (10)^2} = 0.422$ ,  $b_2 = 1.52$ ,  $a = 2.14$ 가 되어 결국 다중회귀방정식은  $Y = 2.14 + 0.42X_1 + 1.52X_2$ 가 되는 것이다.

다음과 같다.

### 1) 회귀모형의 설명도(the model's explanatory ability)

말 그대로 도출된 회귀모형이 얼마나 한 설명도를 지니고 있는지를 평가해 보는 것이다. 이를 위해선 처음의 실제값과 공식을 통해 나온 값과의 차이인 실제오차를 알기 위해 상관계수(correlation coefficient)  $\gamma$ 를 구하게 되고 이에 제곱을 함으로써 결정계수(correlation determination)  $\gamma^2$ 를 구하게 된다.  $\gamma^2$ 은 X라는 독립변수에 의해 Y라는 종속변수가 얼마나 잘 설명되고 있는지를 판별해 주는 역할을 한다.

여기서 상관계수  $\gamma$ 를 구하는 공식은 다음과 같다.

$$\gamma = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2} \sqrt{n\sum y^2 - (\sum y)^2}} \quad \text{그리고 다중회귀방정식의 결정계수}(\gamma^2)\text{의 공식은}$$

$$\gamma^2 = \frac{b_1\sum X_1Y + b_2\sum X_2Y}{\sum Y^2} \text{이다.}$$

예를 들어 상관계수가 0.94라면<sup>27)</sup> 결정계수는 그의 제곱값인 0.88이 된다.

이것이 의미하는 바는 X라는 독립변수에 의해 Y라는 종속변수가 88%만큼 설명이 되며, 설명이 안되는 영역은 12%라는 것을 뜻한다. 여기에서 하나 부연 할 것은 위에서 정의된  $\gamma^2$ 는 설명변수의 수가 증가하면 증가할 수록 큰 값을 가지게 되며 비록 Y와 전혀 관계가 없는 변수를 설명변수에 추가시켜도  $\gamma^2$ 는 증가하게 되는 맹점을 가지게 된다.

이를 보완키 위해 수정  $R^2$ (adjusted  $R^2$ )를 쓰며 이것은 다음과 같이 정의된다.

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-p}\right) (1-R^2)$$

수정  $R^2$ 이외에  $R^2$ 의 맹점을 보완하는 것으로는 맬로우의  $C_p$ 기준(Mallow's  $C_p$  Criteria) 등이 있으나 구체적 내용은 생략한다.

다음으로 추정치의 표준오차(standard error of estimate : SEE)를 구하고 이것을 변형시키면 표준비율오차(standard error of ratio : SER)가 나오는데 이것은 수치가 적을수록 설명능력이 높음을 의미한다.

참고적으로 SEE의 공식과 SER의 공식은 다음과 같다.

$$SEE = f_y \sqrt{1 - \gamma^2} \quad (f_y = \sqrt{\frac{SS_y}{df}} \text{ 이고 } s_{yy} = \sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{N} \text{ 이며 } df \text{는 } n - \text{독립변수의 개수를 의미함.})$$

$$SER = \frac{SEE}{\bar{Y}} \times 100 = \%(\bar{Y} \text{는 관찰치의 평균을 의미})$$

### 2) 회귀모형의 타당도(the model's substantive suitability)

도출된 모형이 얼마나한 타당도를 지니고 있는가를 평가하는 것으로 이를 위해선 이 모형의 유의수준(level of significance)이 얼마나 되느냐를 알아보기 위해 상관계수검증(R-Test)을 해야 하는데 이것은 가설검증의 방식을 통해 진행한다. 귀무가설( $H_0$ )과 대립가설( $H_a$ )을 만들어 가설이 틀렸음에도 불구하고 귀무가설을 받아들이는 제1종과오(Type I error)를 찾

27)  $\gamma = \frac{10.94448 - (375)(.2348)}{\sqrt{10.1499 - (375)^2} \sqrt{10.601342 - (2348)^2}} = 0.9373 \approx 0.94$

아내는 것으로 이러한 과오를 저지를 확률을 유의수준으로 하며 이에 대한 가설검증을 통해 회귀모형의 타당도를 판정하는 것이다. 다음은 베타값(Beta value)을 가지고 모형의 타당도를 측정하는데 이때  $\beta$  값은 옳은 가설을 기각할 위험도를 의미한다.

$$\text{공식은 } \beta = \frac{\text{독립변수의 표준편차}}{\text{종속변수의 표준편차}} \times \text{회귀계수이며 } \beta \text{ 값이 클수록 독립변수}$$

의 표준편차가 크기때문에  $\beta$  값이 적을수록 타당도가 높으며  $1-\beta$  는 검출력으로 사용된다.

3) 회귀모형의 신뢰도(the model's structural reliability)

도출된 모형이 얼마나한 신뢰도를 지니고 있나를 평가하는 것으로 T통계량과 F통계량에 대한 고려를 함께 하게 된다. 여기에서 F통계량은 T통계량을 제공한 형태이므로 양측검정(two tailed test)<sup>28)</sup>만 할 수 있게 된다.  $H_0$ 하에서는 F통계량이 작은 값을 취하게 되고<sup>29)</sup>  $H_a$ 하에서는 F통계량이 큰 값을 취하게 된다.

F통계량은  $H_0$ 하에서  $F(1, n-2)$ 의 분포를 하므로 주어진  $\alpha$  수준에서 다음과 같은 의사결정원칙(decision making rule)을 세울 수 있다.

$F > F(1, n-2 ; 1-\alpha)$ 이면  $H_0$ 을 기각하고

$F < F(1, n-2 ; 1-\alpha)$ 이면  $H_0$ 을 채택한다.

이상의 F검정은 다음과 같은 이론을 바탕으로 하고 있다. 즉 추정된 회귀계수  $b$ 는 평균  $\beta$  그리고 표준오차  $\sigma / \sqrt{\sum(x_i - \bar{x})^2}$ 의 정규분포확률변수라는 것과 또한  $SSE / \sigma^2$ 는 자유도  $n-2$ 의  $\chi^2$ 분포확률변수라는 두가지 가정에서 출발한다. 회귀계수  $b$ 가 정규분포확률변수이면 그의 표준화 변수의 제곱의 합계는  $\chi^2$ 분포의 확률변수가 된다. 즉

$$X_1^2 = \left[ \frac{b - \beta}{\sigma / \sqrt{\sum(x_i - \bar{x})^2}} \right]^2 = \frac{(b - \beta)^2 \sum (X_i - \bar{X})^2}{\sigma^2} \sim X^2(1) \text{이 되며 } X_1^2 \text{의 자유도는 } 1 \text{이 된다. 그리고}$$

$$\text{고 } MSE / \sigma^2 \text{는 } X_2^2 = \frac{(n-2)MSE}{\sigma^2} = \frac{SSE}{\sigma^2} = \frac{\sum e_i^2}{\sigma^2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{\sigma^2} \sim X^2(n-2) \text{가 된다. } X_2^2$$

의 자유도는  $n-2$ 가 되며  $X_1^2$ 와  $X_2^2$ 는 통계적으로 독립임을 보일 수 있다. 위 식을 각각의 자유도로 나누어 비율을 취하면 다음과 같이 F통계량을 얻게 된다. 위 식에서 모회귀계수  $\beta$  는 귀무가설이  $H_0 : \beta = 0.0$ 이므로  $\beta = 0.0$ 으로 생각하면 이들의 비율은

$$F = \frac{X_1^2 / 1}{X_2^2 / (n-2)} = \frac{\frac{b^2 \sum (x_i - \bar{x})^2}{\sigma^2} / 1}{\frac{\sum (Y_i - \hat{Y}_i)^2}{\sigma^2} / (n-2)} = \frac{b^2 \sum (X_i - \bar{X})^2 / 1}{\sum (Y_i - \hat{Y}_i)^2 / (n-2)} \sim F(1, n-2)_{H_0}$$

28) 양측검정이란 확률분포의 양단을 포함시켜 유의수준( $H_0$ 의 수용여부를 결정하기 위한 기준으로 사용되며  $H_0$ 을 받아들일 확률이 이러한 유의수준 이하로 떨어질 때  $H_0$ 은 부정되고  $H_a$ 가 받아들여짐)을 결정짓는 무방향적 가설검정 방법이며 단측검정(one tailed test)은 주어진 통계치들을 사용할 때 확률분포의 한쪽 방향에만 부정한계 영역을 설정하는 방향적 가설검정 방법이다.

29)  $H_0$ 하에서, 즉  $\beta = 0$ 이라면 단순회귀모형  $Y_i = \alpha + \beta X_i + \epsilon_i$ 는  $Y_i = \alpha + \epsilon_i$ 로 된다. 즉  $X$ 가 설명변수로서 의미가 없으므로  $X$ 를 가지고 회귀분석을 한다고 해서 회귀변동량인 SSR을 크게 얻을 수는 없을 것이다. 그러므로  $F = \frac{SSR / 1}{SSE / (n-2)}$ 은 작게 계산될 것이다.

가 된다.

$H_0 : \beta = 0.0$ 이 참일 때 위의 F값은 자유도 1과  $n-2$ 의 F분포확률변수가 된다는 이론에 의해서 F검정을 하게 되는 것이다. 즉 F값이 F분포도의 값보다 크면 귀무가설은 기각되고 대립가설이 채택되는 것이다. T통계량은 F통계량에 제곱을 하면 그 값이 나오지만 공식에 의해 구하면 다음과 같다.

$$T = \frac{\text{회귀계수 (regression coefficient)}}{\text{계수의 표준오차 (standard error of coefficient)}}$$

4) 회귀모형의 적절도(the model's plausibility)

도출된 모형의 적절도를 평가하는 것으로 주로 다공선성(multicollinearity)을 찾는 것으로 일명 대수부호검정(algebraic sign test)이라고도 한다.<sup>30)</sup> 다공선성이란 독립변수와 독립변수간의 관계가 독립변수와 종속변수 간의 관계보다 더 높아서 독립변수와 종속변수 간의 관계를 왜곡시키는 현상을 의미한다. 즉  $Y = a + bX_1 + cX_2$ 라는 회귀방정식에서 Y를 자동차보유대수라 하고  $X_1$ 을 소득,  $X_2$ 를 교육수준이라고 한다면 위 방정식의 이해를 다음과 같이 할 수 있다.

교육수준이 높고 그래서 소득이 올라가고 그래서 자동차보유대수가 증가된다고 이해할 수 있는데 이때 소득과 교육수준은 서로 영향을 주어 자동차보유대수(Y)값을 바라던 값이 아닌 다른 값으로 도출해 내게 되는 것이다. 특히 결정계수가 높을 때 다공선성이 많이 나타날 수 있으므로 이 점에 유의해야 한다. 예를 들어 상관관계행렬식(correlation matrix)이 다음과 같다고 하자((표 1 참고))

(표 1) 상관관계 행렬식

	Y	$X_1$	$X_2$	$X_3$	$X_4$
Y	1				
$X_1$	0.957	1			
$X_2$	0.884	0.818	1		
$X_3$	0.934	0.970	0.829	1	
$X_4$	0.262	0.217	0.285	0.266	1

이 행렬식을 보게 되면 Y(종속변수)와  $X_1$ (독립변수)의 관계가 0.957로 나와있고  $X_1$ 와  $X_3$  관계가 0.97로 나와있다. 0.97이 0.957보다 크므로 이것도 다공선성이 존재한다고 볼 수 있다. 이렇게 다공선성이 존재할 때는 하나의 변수를 버려야 한다. 즉  $Y = a + bX_1 + cX_2$ 에서

30)  $\begin{cases} x+y=2 \\ 2x+y=3 \end{cases}$ 이라는 방정식이 있을 때 이의 근(즉, x와 y의 값)은 하나로 결정이 된다. 그러나  $x+y=2$ 라고 할 때 여기서 x와 y의 근을 찾는 것은 무리이며 사실상  $x+y=2$ 의 x, y 값을 만족시키는 근은 무한대로 존재한다고 할 수 있다. 이와같이 하나의 직선을 공유하는 값이 무수히 존재할 수 있을 때 회귀모형에선 다공선성이 존재한다고 하며 이 경우 x와 y값은 무수히 존재하므로 각 값과 이 직선의 상관성은 높게 나타나지만 실제로는 x, y의 값을 구할 수 없기 때문에 이러한 관계성은 무시할 수 밖에 없는 것이다. 특히 사례수가 적은 때에 이러한 문제가 발생하면 PCM(principal component method) 방법을 사용하거나 sign test 등을 통해 이를 보완, 수정하여 적용한다.

$X_2$ 를 버리게 되는 것이다.

5) 회귀모형의 종합적 평가

이러한 일련의 작업이 끝나면 지금까지 평가한 설명도, 타당도, 신뢰도, 적절도 등을 가장 잘 대변해 주고 있는 회귀모형을 선택하는 것이다. 바로 이 회귀모형이 찾고자 했던 최적모형인 것이다. 일목요연하게 비교분석을 하기 위해 아래와 같이 표를 작성해 최적 회귀모형을 찾는 것도 한 방법일 수 있다.

(표 2) 회귀모형의 종합평가

	$\gamma^2$	SEE	SER	R-Test	$\beta$	F	t
Model 1							
Model 2							
Model 3							
Model 4							
Model 5							

5. 한국 경찰 인력 증가요인의 인과성추론을 위한 회귀모형의 적용

(1) 연구설계

첫째, 본 연구는 분석도구(analytic tool)로써 향후 경찰인력이 어떠한 추세로 증가되어 나갈 것이며, 이러한 경찰인력증가를 유발시키는 요인들에 대한 상호관계와 인과성을 추론키 위해 미래예측적 관점에서 회귀모형을 사용할 것이다.

둘째, 본 연구에서 적용될 회귀모형은 다음과 같은 독립변수와 종속변수로 구성된다. 종속변수(Y)는 경찰인력이 되고 독립변수는 전국 도시수( $X_1$ ), 예산( $X_2$ ), 경찰관서수( $X_3$ ), 교통사고수( $X_4$ ), 차량대수( $X_5$ )이며 본 연구에선 가변수(Dummy Variable)<sup>31)</sup>는 없다.

여기에서 선정된 변수는 연구자가 연구의 성격과 가정된 분석결과에 입각해 선별적으로 채택한 것이며 여기에서 선정되지 않은 변수가 선정된 변수들 보다 더 강한 상관성, 인과성을 설명할 수도 있다는 점을 부인하진 않는다. 이런 맥락에서 본 연구의 변수설정엔 제한적임을 전제하며 연구결과의 한계성도 이런 점에서 찾을 수 있다.

셋째, 본 연구의 가정(hypothetical assumption)은 전국의 도시수, 예산, 경찰관서수, 교통사고수, 차량대수의 증감에 따라 경찰인력도 증감할 것이라는 관계성의 입장에 서며, 각 요

31) 이원변수(binary variables)라고도 불려지는 가변수는 유목변수의 각 유목을 「0」 또는 「1」로 코딩하여 일종의 새로운 연속적 변수로 만들어진 변수를 말하는 것이다. 예를 들어 성별이라는 유목변수를 가변수로 만들려면, 우선 성별이라는 변수를 「남성」변수와 「여성」변수로 나누어서 남자인 경우에는 「남성」변수의 값은 「1」, 「여성」변수 값은 「0」으로 처리하고 여자인 경우에는 「남성」변수값은 「0」, 「여성」변수 값은 「1」로 처리한다.

인간의 영향의 비중도 판별 될 것이다. 이것은 회귀분석의 결과 이러한 관계성, 영향성이 보다 명료해 질 것이다.

네째, 각 변수들의 자료(raw data)는 다음과 같다. ((표 3) 참고)

(표 3) 각 변수들의 통계자료

변수 년도	Dep. Var (Y) 경찰인력(명)	Ind. Var 도시수 (X <sub>1</sub> ) (곳)	예산 (X <sub>2</sub> ) (백만원)	경찰관서수 (X <sub>3</sub> ) (개)	교통사고수 (X <sub>4</sub> ) (건)	차량대수 (X <sub>5</sub> ) (대)
71	43,516	32	17,779	2,481	42,060	144,337
72		32	20,585	2,563	43,715	150,135
73		35	21,784	2,600	42,488	170,714
74		35	28,790	2,601	42,476	183,549
75		35	44,905	2,609	58,323	200,521
76	44,740	35	87,326	2,609	70,241	226,320
77	45,860	35	115,259	2,707	78,863	275,312
78		36	131,502	2,780	94,316	384,536
79		36	167,839	2,864	113,927	494,378
80	56,003	40	215,029	3,268	120,182	527,721
81		50	297,344	3,319	123,373	571,754
82		50	381,709	3,333	141,218	646,996
83	61,052	50	455,017	3,333	170,026	735,316

이것은 1971년부터 1983년까지의 경찰인력, 도시수, 예산, 경찰관서수, 교통사고수, 차량대수에 관한 통계자료이며, 일부년도의 경찰인력은 통계의 불비로 제시치 못했다.<sup>32)</sup>

## (2) 분석<sup>33)</sup>

우선 분석의 편의를 위해 다공선성의 존재여부부터를 검토해 보기로 한다.

(표 4)는 대칭행렬을 이루는 변수 간의 상관관계를 나타내는 것이다.

다공선성이란 이미 앞에서 설명되었듯이 독립변수와 독립변수의 관계가 독립변수와 종속변수와의 관계보다 더 높기때문에, 독립변수와 종속변수와의 관계를 왜곡시키는 것으로써, 이 때는 독립변수들 중에 하나를 제외시키는 것이 필요로 되는 것이다. 그런데 위의 상관관계에서 볼 때 독립변수와 종속변수의 관계보다 독립변수와 독립변수의 관계가 상당히 높게 나타

32) 본고에서 참고한 관련 통계자료는 다음과 같다.

총무처, 「공무원통계」, 1983.  
 경제기획원, 「주요경제지표」, 1984.  
 경제기획원, 「한국통계연감」, 1984.  
 경제기획원, 「예산개요」, 1971-1984.  
 치안본부, 「경찰통계연보」, 1983.  
 치안본부, 「교통통계」, 1983.

33) 본 분석은 앞의 연구설계에서 제시한 자료들을 연세대학교 컴퓨터 분석실에서 SPSS 회귀분석 프로그램에 의거 산출한 결과물(print out)에 입각하고 있다.



(표 4) 대칭행렬 : 변수간의 상관관계

	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>
Y	1.00					
X <sub>1</sub>	0.08354	1.00				
X <sub>2</sub>	0.25323	0.94394	1.00			
X <sub>3</sub>	0.18464	0.92990	0.94023	1.00		
X <sub>4</sub>	0.28665	0.86440	0.97071	0.92946	1.00	
X <sub>5</sub>	0.19238	0.89487	0.97091	0.95760	0.98984	1.00

나므로 다공선성이 모든 독립변수간에 존재한다고 볼 수 있다. 따라서 회귀분석자료 처리결과물(print out)에서 추출한 가능한 회귀모형은 다음과 같다.

모형 1 :  $Y = a_1 + b_1 X_1 = 7114.8099 + 316.72350 X_1$

모형 2 :  $Y = a_2 + b_2 X_2 = 12465.547 + 4.4899152 X_2$

모형 3 :  $Y = a_3 + b_3 X_3 = -21291.211 + 14.243300 X_3$

모형 4 :  $Y = a_4 + b_4 X_4 = 3936.3526 + 0.17448154 X_4$

모형 5 :  $Y = a_5 + b_5 X_5 = 10691.188 + 2.3811061 X_5$

여기서 제시된 5개의 회귀모형이 가능성을 지닌 모형으로써 평가의 대상이 되는 것이다.

우선 설명능력부터 살펴보면 결정계수, 추정의 표준오차(SEE) 그리고 SER의 세가지 차원에서 검토해 볼 수 있다.

(표 5) R<sup>2</sup>;

(Total Coefficient of Determination)

모형	R <sup>2</sup>
1	0.00698
2	0.06413
3	0.03409
4	0.08217
5	0.03701

(표 6) SEE

(Standard Error of Estimate)

모형	SEE
1	26889.33
2	26104.14840
3	26519.692
4	25851.337
5	26479.624

(표 5)에서 보는 바와 같이 결정계수는 모형의 설명능력에 대한 직접적인 척도로서 사용될 수 있기 때문에 우선 먼저 살펴보기로 한다. R<sup>2</sup>의 값이 클수록 그 모형의 설명능력이 크기 때문에 이러한 관점에서 보면 모형 4가 가장 크며 모형 2, 5, 3, 1의 순서임을 알 수 있다.

다음 추정의 표준오차(SEE)는 종속변수에 대한 추정을 함에 있어서 존재하는 오차의 크기를 측정하는데 이용된다.

따라서 표준오차의 크기가 작을수록 모형의 설명능력이 좋다고 평가되는 것이다. (표 6)에 보면 5가지 모형이 거의 비슷한 바 그 오차가 각 모형이 비슷한 수준이라고 볼 수 있으므로 이에 의한 모형의 적격판별은 그렇게 커다란 의미는 없다고 볼 수 있다.

다음 SER은 SEE와 마찬가지로 도출된 수치가 적을수록 정확하며 설명능력이 높다고 볼 수 있다. 따라서 (표 7)에서 보면 SEE와 마찬가지로 그 값이 5개의 모형이 서로 비슷함을 알 수 있다.

(표 7) SER(Standard Error of Ratio)

모형	SER
1	139.2
2	135.1
3	137.3
4	133.8
5	137.1

(표 8) R:

(The Multiple Correlation of Coefficient)

모형	R
1	0.08354
2	0.2532
3	0.18464
4	0.28665
5	0.19238

따라서 이것도 모형의 적격판별을 하는데 커다란 의미를 부여한다고 볼 수는 없다. 설명능력의 차원에서 이상을 종합해 보면 모형 4가 가장 설명능력이 높다고 할 수 있다.

두번째로 모형의 타당도를 살펴보기로 한다. 이를 위해 우선 계수의 다중상관관계를 알기 위해 R-Test를 해보고 다음  $\beta$  통계량의 값을 보아 평가하도록 한다. R-Test는 모형의 종속변수와 독립변수간의 상관관계를 나타내는 것이다. 따라서 R의 값이 클수록 상관관계의 타당도가 높다고 볼 수 있다.

(표 8)에서 보면 모형 4가 가장 높고 모형 2, 5, 3, 1의 순서임을 알 수 있다.

다음  $\beta$  통계량을 살펴보면  $\beta$ 는 제2종 과오를 범할 확률로써 적으면 적을수록 좋다. 따라서 (표 9)에서 보면 모형 1이 가장 낮으나 이는 가변수를 사용하지 않은 탓으로 사료되므로 모형 1이 최적의 회귀모형이라 볼 수는 없는 것이다. 모형의 타당도라는 입장에서 이상을 종합 평가해 볼 때 모형 4가 가장 타당하다고 볼 수 있다.

다음 세번째로 모형의 구조적 신뢰도를 살펴보기로 한다. 이를 위해 우선 T검정과 F검정을 해 보아야 하는데 우선 T검정을 해 보면 다음과 같다.

(표 9)  $\beta$ -Value

모형	$\beta$ -Value
1	0.08354
2	0.25323
3	0.18464
4	0.28665
5	0.19238

(표 10) t-test

모형	t
1	0.278
2	0.863
3	0.623
4	0.992
5	0.192

(표 10)에서 보는 바와 같이 자유도(여기에서는 11임)를 90%로 보고 계산한 것이다. 모든 모형들이 90%의 유의수준을 만족 못 시키고 있음을 알 수 있다. 이는 여러가지 이유가 있겠지만 여기서는 너무 사례의 수(number of cases)가 적은 탓으로 기인된 결과가 아닌가 사료되는 것이다.

다음 F통계량은 종속변수의 실제치와 추정치와의 관계를 나타내는 것이다. (표 11)에서 보듯이 모든 모형이 90%수준의 유의수준을 만족시키지 못하고 있음을 알 수 있다.

(표 11) F 통계량

모 형	F
1	0.077
2	0.75373
3	0.38826
4	0.98473
5	0.42275

이상의 결과를 종합해 보면 구조적 신뢰도는 지극히 저조하다고 판단할 수 있다. 이렇게 되는 이유로는 첫째, 사례수의 부족, 둘째, 모형설정시 변수설정예외의 문제, 셋째, 가변수의 비존재등의 여러가지 요인이 복합적으로 영향을 끼친 탓으로 사료되는 것이다.

이상의 회귀모형평가분석을 종합적으로 일관성있게 제시키 위해 표를 만든 것이 (표 12)이다.

(표 12) 회귀모형 종합평가 (표)

Model	t	Beta	R	R <sup>2</sup>	adj R <sup>2</sup>	SEE	F	SER
1	0.278	0.08354	0.08354	0.00698	0	26889.33	0.077	139.2
2	0.868	0.25323	0.2532	0.06413	0	26104.14	0.75372	135.1
3	0.623	0.18464	0.18464	0.03409	0	26579.69	0.38826	137.3
4	0.992	0.28665	0.28665	0.08217	0	25851.33	0.98473	133.8
5	0.192	0.19438	0.19238	0.03701	0	26479.62	0.42275	137.1

### (3) 분석 결과

T값과 F값으로 나타나는 구조적 신뢰도가 유의적 수준을 만족시키지 못해 최적회귀모형 선정의 제한점을 제기하고 있으나 적절도로 대변되는 다공선성의 존재여부와 설명능력, 타당성, 구조적 신뢰도 등을 종합적으로 고려해 볼 때, 회귀모형 4가 가장 적합한 것으로 판명되는 것이다. 모형 4는  $Y = 3936.3526 + 0.17448154X_4$ 이다. 따라서 경찰인력증가에 영향을 가장 많이 주는 변수는 독립변수  $X_4$ 로 대변되는 교통사고수이며 전국도시수, 예산, 경찰관서수, 차량대수는 교통사고 수만큼 높은 상관성이 없음이 밝혀졌다고 할 수 있다.

이러한 결과가 의미하는 바는 교통사고건수가 계속 늘어나면 경찰인력도 위의 계수와 비례하여 증가할 것이란 점을 의미하며 교통사고건수가 경찰인력증가의 인과성을 가장 잘 시현시켜 주고 있다고 결론지을 수 있는 것이다.

## 6. 요약 및 결론

이상으로 한국에 있어서 경찰인력의 증가요인의 인과성을 추론키 위한 회귀모형의 적용에

대해 고찰해 보았다.

이미 앞의 연구설계에서 제시했듯이 본 연구에 적용된 독립변수들이 연구자의 자의적 선택에 의해 선정되어졌고 자료의 내용과 범위도 제한적이며 사례수가 적다는 것이 분석결과 제약점으로 드러났으나 공공정책결정을 기존의 규범적이고 기술적(descriptive)인 차원에서 탈피해 보다 객관적이고 계량적인 접근방법을 통해 접근해 보았다는 것은 이 분야에 있어서 비록 시론적인 시도이긴 하나 의미는 있다고 사료된다.

본 연구분석의 결과 도출된 사실은 다음과 같다.

첫째, 가치지향적인 공공정책결정에 필요한 정책분석에 계량적 기법, 특히 회귀모형이 적용가능하고 미래예측적 관점과 진단 및 처방적 관점에서 유용성이 크다는 것이 밝혀졌다.

둘째, 경찰인력증가는 계속 될 것이란 점이다. 도시수, 예산, 경찰관서수, 교통사고수 및 차량대수도 계속하여 늘어날 것이며 특히 경찰인력의 증가는 보다 증가율이 제고될 것이다.

셋째, 경찰인력증가의 요인으로 도시수, 예산, 경찰관서수, 교통사고수 및 차량대수의 다섯가지의 변수를 고려해 보았는데 이중 교통사고 수의 증가와 경찰인력의 증가의 상관성과 인과성이 가장 높다는 사실이 밝혀졌다. 따라서 교통사고수가 계속 증가하는 추세에 있는 한 경찰인력의 증가는 필연적이라는 사실을 알 수 있는 것이다.

넷째, 경찰인력증가에 영향을 끼치는 요인들간의 관계는 서로 밀접한 관련성을 지니고 있으며 이러한 사실은 다섯개의 변수 중 어느 하나에만 치중된 처방은 바람직하지 않으며 주요 변수에 대한 처방못지 않게 기타변수에 대한 고려도 있어야 함을 의미하는 것이라 하겠다.

이상의 분석결과를 고려해 볼 때 경찰인력의 증가는 기정사실화 되는 바 이에 대한 적절성과 그 효과성, 능률성에 대한 세심한 배려가 정책적 차원에서 요구되며 기 제시된 다섯가지 요인들에 대한 경찰인력증가에 대한 영향을 인식하고 이에 대한 적극적 처방이 강구되어야 하며 특히 교통사고수의 증가가 경찰인력의 증가에 깊은 인과성을 두고 있다는 본 연구결과를 인식하고 교통사고 수를 가능한 억제할 수 있는 방안을 치안적 차원에서 뿐 아니라 사회적 차원에서 고려해 보아야 되리라 믿는다. 바로 이러한 노력이 국가의 인적, 물적 자원의 낭비를 줄일 수 있다는 맥락에서 볼 때 이에 대한 본질적 처방이 필요시 된다고 사료되는 바이다.

## 참 고 문 헌(Bibliography)

1. 강석호, 「Operations Research—계량경영분석을 중심으로」, 서울 : 영지문화사, 1983.
2. 광수일·이경환, 「경영통계학」, 서울 : 박영사, 1983.
3. 경제기획원, 「주요경제지표」, 서울 : 경제기획원, 1984.
4. \_\_\_\_\_, 「한국통계년감」, 서울 : 경제기획원, 1984.
5. \_\_\_\_\_, 「예산개요」, 서울 : 경제기획원, 1971~1984.
6. 김기영·곽노균, 「계량의사결정론」, 서울 : 법문사, 1984.
7. 김대환, “도시개발과 시민의식 구조,” 최상철 외(공편), 「한국도시개발론」, 서울 : 일지사, 1981.
8. 김병수·윤기중 외(공저), 「SPSS를 이용한 통계자료분석」, 서울 : 박영사, 1987.
9. 김정년, 「통계학」, 서울 : 경문사, 1986.

10. 김우철 외 (공편), 「현대통계학」, 서울 : 영지문화사, 1984.
11. 김해식, 「SPSS—컴퓨터분석기법」, 서울 : 박영사, 1986.
12. 노화준, 「행정계량분석」, 서울 : 법문사, 1985.
13. \_\_\_\_\_, 「계량분석개론」, 서울 : 법문사, 1986.
14. 박성현, 「현대실험계획법」, 서울 : 대영사, 1983.
15. 소영일·이종민 외, 「SPSS를 활용한 현대통계학」, 서울 : 법문사, 1986.
16. 연세대학교 대학원 편, 「사회과학의 계량적 접근」, 서울 : 연세대학교 출판부, 1976.
17. 오택섭, 「사회과학데이터분석법」, 서울 : 나남출판사, 1986.
18. 윤기중, 「통계학」, 서울 : 법문사, 1986.
19. 윤기중, 「수리통계학」, 서울 : 박영사, 1986.
20. 이관용·김기중 역, Robert B. McCall 저, 「기초심리통계학 (Fundamental Statistics for Psychology)」, 서울 : 법문사, 1984.
21. 이성복, 「도시행정론—한국의 도시를 중심으로」, 서울 : 법문사, 1987.
22. 이효구, 「새통계학」, 서울 : 박영사, 1983.
23. 임인재, 「교육심리사회연구를 위한 통계방법」, 서울 : 박영사, 1976.
24. 총무처, 「공무원통계」, 서울 : 총무처, 1983.
25. 치안본부, 「경찰통계년보」, 서울 : 치안본부, 1983.
26. \_\_\_\_\_, 「교통통계」, 서울 : 치안본부, 1983.
27. Bell, Daniel, *The Coming of Post Industrial Society: A Venture in Social Forecasting*, New York: Basic Books, Inc., 1976.
28. Bingham, Richard D., et al(eds), *Reaching Decisions in Public Policy and Administration: Methods and Applications*, N. Y.: Longman Inc., 1982.
29. Blau, Pater M., *The Dynamics of Bureaucracy*, Chicago, Ill.: University of Chicago Press, 1955.
30. Brzezinski, Zbigniew, *Between Two Ages: American's Role in the Technetronic Era*, N. Y.: The Viking Press, 1970.
31. Draper, N. R. & H. Smith, *Applied Regression Analysis, 2nd. ed.*, N. Y.: John Wiley & Sons, Inc., 1981.
32. Drucker, Peter F., *The Age of Discontinuity: Guidelines to Our Changing Society*, N. Y.: Harper & Row, 1969.
33. Friedrich, Carl J. & Z. K. Brzezinski, *Totalitarian Dictatorship and Autocracy*, Cambridge, Mass.: Harvard University Press, 1956.
34. Galbraith, John K., *The Age of Uncertainty*, Boston, Mass.: Houghton Mifflin Co., 1977.
35. Gordon, K. A., "Issues in Multiple Regression," *American Journal of Sociology*, Vol. 73, 1968.
36. Gouldner, Alvin W., *Patterns of Industrial Bureaucracy*, Glencoe, Ill.: Free Press, 1954.
37. Guilford, J. P. and B. Fruchter, *Fundamental Statistics in Psychology and Education*, 5th ed., N. Y.: McGraw-Hill, 1973.
38. Hull, Hadlai C., et al., *SPSS Update, New Procedures and Facilities for Release 7 and 8*, N. Y.: McGraw-Hill, 1979.
39. Huntington, Samuel P., *The Soldier and the State*, Cambridge, Mass.: Harvard University Press, 1957.
40. Janowitz Morris, *The Professional Soldier*, Chicago: Free press, 1960.
41. Johnston, R. J., *The American Urban System: A Geographical Perspective*, N. Y.: St. Martin's Press, Inc., 1982.
42. Jones, Thomas E., *Options for the Future*, N. Y., Praeger Publishers, 1980.

43. Kahn, Herman, *The Next 200 Years*, N. Y.: William Morrow & Co., 1976.
44. Kerlinger, Fred N., *Foundations of Behavioral Research*, 2nd ed., N. Y.: Holt, Rinehart and Winston, 1973.  
Kerlinger, Fred N. and E. J. Pedhazur, *Multiple Regression in Behavioral Research*, N. Y.: Holt, Rinehart & Winston, 1973.
45. Lasswell, Harold D., "The Garrison State", *American Journal of Sociology*, Vol. 46, January, 1941.
46. McCall, Robert B., *Fundamental Statistics for Psychology*, N. Y.: Harcourt Brace Jovanovich, Inc., 1975.
47. Merton, Robert K., *Social Theory and Social Structure*, Glencoe, Ill.: The Free Press, 1957.
48. Nachmias, David, *Public Policy Evaluation: Approaches and Methods*, N. Y.: St. Martin's Press, 1979.
49. Naisbitt, John, *Megatrends*, N. Y.: Warner Books, Inc., 1982.
50. Neter, J. W., Wasserman & M. H. Kutner, *Applied Linear Regression Models*, Homewood: Richard D. Irwin, 1983.
51. Nie, Norman H., et al., *Statistical Package for the Social Sciences*, 2nd, ed., N. Y.: McGraw-Hill, 1976.
52. Overall, J. E. & C. Klett, *Applied Multivariate Analysis*, N. Y.: McGraw-Hill, 1973.
53. Permuter, Amos, *The Military and Politics in Modern Times*, New Haven: Yale University Press, 1977.
54. Selznick, Philip, *TVA and the Grass Roots*, Berkeley, CA.: University of California Press, 1968.
55. Smith, George W., *Quantitative Methods of Research in Education*, Washington, D. C.: College and University Press, 1975.
56. Sorokin, P. and C. C. Zimmerman, *Principles of Rural-Urban Sociology*, N. Y.: Holt, 1929.
57. Stouffer, Samuel et al., *The American Soldier*, Princeton, N. J.: Princeton University Press, 1949.
58. Toffler, Alvin, *Future Shock*, New York: Bantam Book, 1970.
59. \_\_\_\_\_, *The Third Wave*, New York: William Morrow & Company, Inc., 1980.
60. Toynbee, Arnold J., *A Study of History*, London: Oxford University Press, 1973.
61. Weisberg, S., *Applied Linear Regression*, N. Y.: Wiley, 1980.
62. Younger, M. S., *A Handbook for Linear Regression*, North Scituate, Mass.: Duxbury, 1979.