Journal of the KSQC
Vol.21, No.1, June 1993

1993년 6월 韓國品質管理學會誌 제21권 제1호

선형계획법을 이용한 회귀분석 결과의 비교 연구

A Comparative Study of the Results of the Regression Analysis by Linear Programming

김 광 수*
정 지 안**
이 진 규***

ABSTRACT

This study attempts to present the linear regression analysis that involves more than one regressor variable, because regression analysis is the most widely used statistical technique for describing, predicting and estimating the relationships between given data.

The model of multiple linear regression may be solved directly by the two linear programming methods, i.e., to minimize the sum of the absolute deviation (*MSD*) and to minimize the maximum deviation (*MMD*). In addition, some results was compared to each techniques for accuracy and tested to the validity of statistical meaning.

* 충주산업대학교 산업공학과
** 동국대학교 대학원 산업공학과
*** 동국대학교 산업공학과

1. 서 론

공학문제나 경영문제 등을 다루면서 얻어진 자료에 대해서, 이 자료들의 관계를 설명하거나 예측(prediction) 또는 추정(estimation)을 하는 것은 매우 보편적인 일이며 중요한 문제 중의 하나이다. 따라서 이들 자료의 관계를 밝히는 기법의 개발은 통계학 분야에서 매우 활발히 이루어 졌으며, 그 중 하나인 회귀분석(regression analysis)이 널리 사용되고 있다.

이들 방법의 중요성은 무엇보다도 각 변수들 간의 관계를 충분히 반영하여 정확한 결과를 얻는데 있다. 그러나 일반적으로 회귀분석이 이들 자료들의 관계를 비교적 정확히 반영하여 설명이나 예측 등에 사용되고 있지만, 문제시 되는 것은 보다 정확한 방법을 찾아내는 일과 이상치에 의해 어떤 영향을 받을 것인가 하는 문제 등을 해명하는 일이다.

회귀분석외에 선형계획법을 적용하여 이와 같은 분석을 할 수 있다는 것은 이미 Wagner (1959), Ignizio(1982)등에서 볼 수 있고, 예측의 정확성 면에서도 회귀분석 보다 오히려 선형계획법이 우수하다고 하였다[5], [7].

회귀분석과는 달리 판별분석(discrimination analysis)에도 선형계획법을 적용하는 시도가 Freed와 Glover(1986), Glorfeld와 Gaither (1982)등에 의해 발표되었다[3], [4].

본 연구에서는 선형다중 회귀분석에 의해 얻은 결과와 선형계획법을 이용하여 얻은 결과들 간의 정확도를 알아보고, 해석적인 관점에서 몇가지 분석결과들을 상호 비교하여 평가해 보는 일이다. 참고로 분석에 사용된 도구로서는 비교적 흔히 사용되는 소프트웨어들로서 다중회귀 분석과 연구에 사용된 모든 통계적인 분석에는 Minitab을 사용하였고, 선형계획법에는 QSB + (quantitative systems for business plus)를 사용하였다.

2. 다중선형 회귀분석

두개 이상의 변수가 포함된 회귀모형을 다중회귀 모형이라 하며, 이때 변수들이 선형적인 관계에 있으면 이를 선형다중 회귀모형(multiple linear regression model)이라고 한다.

n 개의 변수 x_j($j = 1, 2, \cdots, n$)에 대한 관측치를 y라 하고, 다중회귀 모형을 일반식으로 나타내면,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

이 된다.

이때, β_j ($j = 0, 1, 2, \cdots, n$)는 회귀계수 (regression coefficients)로서 모수추정에 의해 주어진 자료로 부터 구해내야 하고 ε은 편차를 나타낸다.

이렇게 해서 구해진 예측식은,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_n x_n + \varepsilon$$

이 된다.

우선 혼란을 줄이기 위해 용어를 정리하면, 두 변수의 관계에 의해 주어진 값을 관측치라 하고, 모수 추정에 의해 구한 식을 예측식 그리고 예측식에 변수값을 대입하여 얻은 값을 예측치라 한다.

모수 추정에는 최소자승법(method of least squares)이 사용되며, 만약 i번째 관측자료에 대한 반응치를 y_i라고 하면, 이것은

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in} + \varepsilon_i$$
$$= \beta_0 + \sum_{j=1}^{n} \beta_j x_{ij} + \varepsilon_i,$$
$$i = 1, 2, \cdots, m(m > n)$$

이 된다.

이때 최소자승함수는

$$S(\beta_0, \beta_1, \cdots, \beta_n) = \sum_{i=1}^{m} \varepsilon_i^2$$
$$= \sum_{i=1}^{m} (y_i - \beta_0 - \sum_{j=1}^{n} \beta_j x_{ij})^2$$

으로 표시된다.

함수 S를 $\beta_0, \beta_1, \cdots, \beta_n$에 대해 최소화 시키면서, 이들의 최소자승 추정치를 구하기 위해서는 아래의 두 조건, 즉

$$\frac{\partial S}{\partial \beta_0} \mid \hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_n$$
$$= -2 \sum_{i=1}^{m} (y_i - \hat{\beta}_0 - \sum_{j=1}^{n} \hat{\beta}_j x_{ij}) = 0$$
$$\frac{\partial S}{\partial \beta_0} \mid \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_n$$
$$= -2 \sum_{i=1}^{m} (y_i - \hat{\beta}_0 - \sum_{j=1}^{n} \hat{\beta}_j x_{ij}) x_{ij} = 0,$$
$$j = 1, 2, \cdots, n$$

을 만족시켜야 한다. 이렇게 해서 최소자승 정규 방정식 (normal equations)을 얻게 되는데, 이 연립방정식에서 추정치를 구하여 앞서 말했던 예측식을 만들 수 있다. 이것을 보다 간단하게, $y = X\beta + \varepsilon$ 와 같이 행렬로도 나타낼 수 있는데,

이때 $y = [y_1, y_2, \cdots, y_n]'$,
$\beta = [\beta_0, \beta_1, \cdots, \beta_n]'$,
$\varepsilon = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]'$,

그리고

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

로 된다. 여기에서 최소자승 추정량 $\hat{\beta}$를 구하기 위해서는

$$S(\beta) = \sum_{i=1}^{m} \varepsilon_i^2 = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta)$$

를 최소화하는 것으로

$$\frac{\partial S}{\partial \beta} \mid \hat{\beta} = -2X'y + 2X'X\hat{\beta} = 0$$

을 만족해야 한다.

이때 β의 추정량 $\hat{\beta}$는 $\hat{\beta} = (X'X)^{-1}X'y$가 된다[8][9].

본 연구에서는 Montgomery와 Peck(1982)의 예에 있는 문제로서, 간단히 설명하면 자판기 운용업자가 음료의 운송과 자판기의 관리를 하기 위해서는 시간 (Y)이 중요한데, 이때 영향을 미치는 중요한 변수들로서 양(X_1)과 거리(X_2)를 고려하여 회귀분석을 실시하는 것이다. 그것은 두개의 변수에 대한 것으므로 앞서 나타낸 예측식으로 나타내면,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

이 된다[6].

이 문제에 대한 Montgomery와 Peck (1982)의 추정치 $\hat{\beta}_0 = 2.34123$, $\hat{\beta}_1 = 1.61591$,

$\hat{\beta}_2 = 0.01439$이고, Minitab을 이용했을 경우의 추정치는 $\hat{\beta}_0 = 2.341$, $\hat{\beta}_1 = 1.6159$, $\hat{\beta}_2 = 0.014385$로서 소수점 3자리 또는 4자리 이하에서 차이가 나는데, 본 연구에서는 소수점 이하 다섯자리까지 사용한다. 이때의 계산결과는

$$\hat{y} = 2.34123 + 1.61591x_1 + 0.01439x_2$$

이다.

두개의 변수값과 관측치는 4장에서 보이고 그리고 회귀식에 의해 계산된 결과와 선형계획법에 의해 계산된 결과의 비교 등도 역시 그때 다루게 된다.

3. 선형계획법의 적용

회귀분석에 선형계획법을 적용할 수 있다는 것은 서론 부분에서 언급되었고, 이들에 대한 연구는 Wagner(1959), Ignizio(1982)와 판별분석을 다룬 문헌에서도 그 개념이 설명되어 있다[1]. 이때, 두가지 개념으로 선형계획법을 적용하기 위한 모형화를 할 수 있는데, 한가지는 절대편차합의 최소화(minimize the sum of absolute deviation : *MSD*)이고, 다른 한가지는 최대편차의 최소화(minimize the maximum deviation : *MMD*)로서, 이들 두가지 방법을 적용한 결과를 서로 비교하고자 한다.

3-1. *MSD*에 의한 모형화

다수의 변수로 구성된 자료가 선형관계를 가지고 있으며 각 변수들과 예측치의 선형관계가

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_n x_{in},$$
$$i = 1, 2, \ldots, m$$

인 것으로 예측되었다.

이때, 관측치를 y_i라고 하고, 이들 예측치와 관측치 간의 편차(deviation)를 ε_i라고 하면,

$$\varepsilon_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}$$
$$+ \cdots + \hat{\beta}_n x_{in}, \quad i = 1, 2, \ldots, m$$

인 관계가 된다.

이들의 관계를 선형적으로 적합시키기 위한 방법으로 각각의 편차 ε_i의 절대편차(absolute deviation)의 합을 최소화 하면서 $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_n$,을 구하는 것을 생각 할 수 있다.

즉, 위 개념에 대한 수학적 모형은,

$$\text{Min} \sum_{i=1}^{m} |\varepsilon_i|$$

$$\text{s.t., } y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots$$
$$+ \hat{\beta}_n x_{in}) = 0, \quad \text{for all } i$$

이 된다. 이 모형은 목적함수 식이 절대값으로 되어 있고, $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_n$,는 비부수(non-negative)가 되어야 하나, 그 조건이 없기 때문에 선형계획법을 적용할 수 없다.

이를 해결하기 위해 선형변환(linear transformation) 즉, $\hat{\beta} = \hat{\beta}_{j+} - \hat{\beta}_{j-}$와 $\hat{\beta}_j \geq 0$. ($j = 0, 1, 2, \ldots, n$)인 조건으로 변환할 수 있다.

그리고 i 번째 자료의 제약조건에 부의 편차 n_i,와 양의 편차 p_i,를 도입하면,

$$y_i - (\hat{\beta}_0^+ - \hat{\beta}_0^- + \sum_{j=1}^{n} (\hat{\beta}_j^+ x_{ij} - \hat{\beta}_j^- x_{ij})) + n_i - p_i = 0,$$

이 되고, 이 제약조건을 만족시키기 위해서는 모든 i에 대해서 새로운 목적함수$(n_i - p_i)$를 최소화 하는 수학 모형이

$$\begin{aligned} &\text{Min } \sum_{i=1}^{m} (n_i - p_i) \\ &\text{s.t., } y_i - (\hat{\beta}_0^+ - \hat{\beta}_0^- + \sum_{j=1}^{n} (\hat{\beta}_j^+ x_{ij} - \hat{\beta}_j^- x_{ij})) + n_i - p_i = 0, \text{ for all } i \\ &\hat{\beta}_j^+, \hat{\beta}_j^- \geq 0 \quad \text{for all } j \\ &n_i, p_i \geq 0, \quad \text{for all } i \end{aligned}$$

이 되는 선형계획 모형이 만들어 진다. 이 선형계획 모형에 대해 2장에서 사용한 예를 적용하기 위해서 관측치 y_i 값은 이미 알고 있는 것이므로,

$$\begin{aligned} &\text{Min } \sum_{i=1}^{m} (n_i - p_i) \\ &\text{s.t., } (\hat{\beta}_0^+ - \hat{\beta}_0^- + \sum_{j=1}^{n} (\hat{\beta}_j^+ x_{ij} - \hat{\beta}_j^- x_{ij})) + n_i - p_i = y_i, \quad \text{for all } i \\ &\hat{\beta}_j^+, \hat{\beta}_j^- \geq 0 \quad \text{for all } j \\ &n_i, p_i \geq 0, \quad \text{for all } i \end{aligned}$$

와 같이 변환하여 적용한다. 이렇게 *MSD*를 예에 적용하여 얻은 예측식은

$$\hat{y} = 3.66208 + 1.42720 x_1 + 0.01426 x_2$$

가 된다.

3-2. *MMD*에 의한 모형화

*MSD*에 의한 모형화와는 달리, *MMD*에 의한 모형화의 개념은 관측치 y_i와 예측치 예측치 $\hat{y}_i$에 대해서, 최대편차 ε를 최소화 시키는 것으로 수학적 모형화는

$$\begin{aligned} &\text{Min } \varepsilon \\ &\text{s.t., } (y_i - \hat{y}_i) - \varepsilon \leq 0, \quad \text{for all } i \\ &(y_i - \hat{y}_i) + \varepsilon \geq 0, \quad \text{for all } i \\ &\varepsilon \geq 0 \end{aligned}$$

가 된다.

이에 대해 앞의 예를 적용하기 위해 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_n x_{in}$로 변환하고, y_i는 알고 있는 값이므로,

$$\begin{aligned} &\text{Min } \varepsilon \\ &\text{s.t., } \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_n x_{in} + \varepsilon \geq y_i, \quad \text{for all } i \\ &\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_n x_{in} - \varepsilon \leq y_i, \quad \text{for all } i \\ &\varepsilon \geq 0 \end{aligned}$$

가 되는 선형계획 모형을 적용하며, 이 방법으로 계산된 예측식은 $\hat{y} = 0.52756 + 1.86197 x_1 + 0.01156 x_2$이였다.

4. 계산결과의 분석

회귀분석 결과와 선형계획법을 적용한 결과

에 대하여 비교하면 크게 세가지로 설명할 수 있다.

첫째, 비교는 관측치와 각각의 예측치에 의한 계산의 정확성을 확인하는 것이다. 〈 표 1 〉에는 Montgomery와 Peck의 예에 대한 원래 자료와 관측치 그리고 각 방법에 따른 예측치와 편차를 나타내고 있다.

〈 표 1 〉을 통해서 설명할 수 있는 것은 예측의 정확도에 있어서는 원래의 관측치와 비교해서 *MSD*법이 25개중 17개의 결과가 회귀분석보다 정확한 것을 알 수 있고, 반면에 *MMD*는 6개 만이 정확했다. 그러므로 *MSD*가 예측의 정확성에서는 다른 방법들 보다 우수한 것으로 생각이 되지만 그렇다고 해서 전적으로 우수한 결과를 얻었다고는 할 수 없으며 추가로 통계적인 몇가지 분석을 실시한다.

둘째, 편차의 산포를 비교하는 것으로, 일반적으로 단순회귀에서는 자료의 산포로서 그들의 관계를 개략적으로 알 수 있지만, 다중회귀모형에 대해서는 거의 의미가 없다. 그렇지만 편차에 대한 산포로서는 분산의 안정성으로서 정확성을 검사할 수 있는데 아래의 〈 그림 1-1 〉

〈 표 1 〉 회귀분석과 선형계획법의 결과비교

관측치	운송량	운송거리	회귀분석		*MSD*		*MMD*	
Y	X_1	X_2	예측치	편차	예측치	편차	예측치	편차
16.68	7	560	11.7110	−5.0310	21.6381	−4.9581*	20.0350	−3.3549**
11.50	3	220	10.3548	1.1452	11.0809	0.4191*	8.6567	2.8433
12.03	3	340	12.0816	−0.0516	12.7921	−0.7621	10.0439	1.9861
14.88	4	80	9.9561	4.9239	10.5117	4.3683*	8.9002	5.9798
13.75	6	150	14.1952	−0.4452	14.3643	−0.6143	13.4334	0.3166**
18.11	7	330	18.4013	−0.2913	18.3583	−0.2483*	17.3762	0.7339
8.00	2	110	7.1560	0.8440	8.0851	−0.0851*	5.5231	2.4769
17.83	7	210	16.6745	1.1555	16.6471	1.1829	15.9889	1.8411
79.24	30	1460	71.8279	7.4121	67.2977	11.9423	73.2643	5.9757**
21.50	5	605	19.1267	2.3733	19.4254	2.0746*	16.8312	4.6688
40.33	16	688	38.0961	2.2339	36.3082	4.0218	38.2724	2.0576**
21.00	10	215	21.5942	−0.5942	21.0000	0.0000*	21.6327	−0.6327
13.50	4	255	12.4743	1.0257	13.0072	0.4928*	10.9232	2.5768
19.75	6	462	18.6849	1.0651	18.8134	0.9366*	17.0401	2.7099
24.00	9	448	23.3311	0.6689	22.8954	1.1046	22.4642	1.5358
29.00	10	776	29.6670	−0.6670	28.9998	0.0002*	28.1178	0.8822
15.35	6	200	14.9147	0.4353	15.0773	0.2727*	14.0114	1.3386
19.00	7	132	15.5521	3.4479	15.5348	3.4652	15.0873	3.9127
9.50	3	36	7.7070	1.7930	8.4570	1.0430*	6.5296	2.9704
35.10	17	770	40.8920	−5.7920	38.9047	−3.8047*	41.0822	−5.9823
17.90	10	140	20.5149	−2.6149	19.9305	−2.0305*	20.7657	−2.8657
52.32	26	810	56.0108	−3.6908	52.3199	0.0001*	58.3024	−5.9824
18.75	9	450	23.3599	−4.6099	22.9239	−4.1739*	22.4873	−3.7373**
19.83	8	635	24.4062	−4.5762	24.1348	−4.3048*	22.7639	−2.9339**
10.75	4	150	10.9634	−0.2134	11.5099	−0.7599	9.7094	1.0406
$\sum \lvert \varepsilon_i \rvert$	for all i			57.1012		53.0659		71.3359

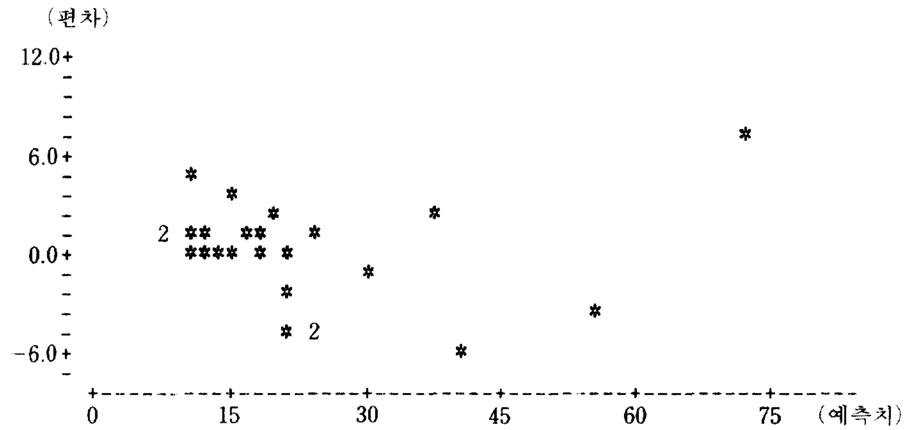
* *MSD*의 편차가 회귀분석의 편차보다 적은것(25개 관측치중 17개)

** *MMD*의 편차가 회귀분석의 편차보다 적은것(25개 관측치중 6개)

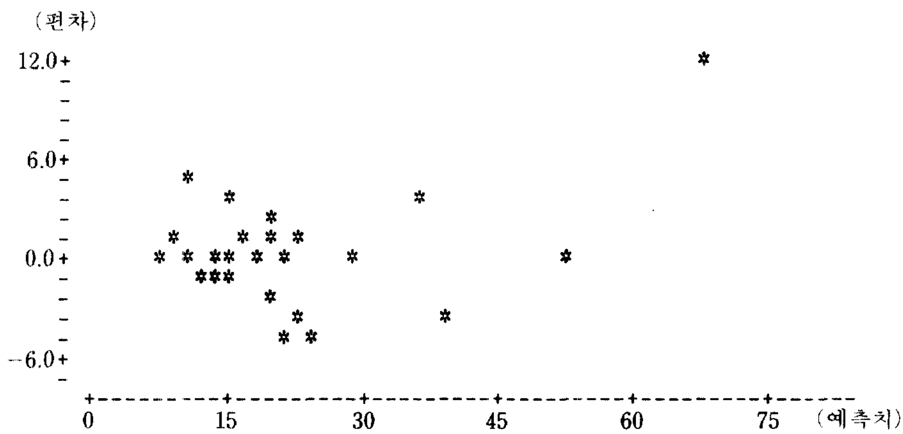
~〈그림 1-3〉에서는 예측치에 대한 각 방법들의 편차를 나타낸다. 이때 산포도(scatter diagram)를 그릴 때, 관측치를 사용치 않고 예측치를 사용하는데, 이것은 관측치와 편차 간에는 일반적으로 상관관계가 있기 때문이다.

〈그림 2〉에서는 이것들에 대한 다중 산포도를 나타낸 것이다.

〈그림 1-1〉~〈그림 1-3〉과 〈그림 2〉를 통해 설명할 수 있는 것은 명확하게 구별되는 것은 아니지만 ± 1.2 내에서는 회귀분석이 7점, *MSD*가 11점, *MMD*가 2점이 각각 위치하고, ± 3.6 내에서는 회귀분석이 18점, *MSD* 19점, *MMD* 19점이 각각 위치 하였다. 따라서 이것은 〈표 1〉에서 본 결과의 정확성과 같이 설명될 수 있다. 그렇지만 ± 3.6에서 보면 전체적으로 분산의 변동은 거의 비슷한 것으로 나타났다. 이러한 그림에서 관측할 수 있는 특징으로는 회귀분석 결과와 *MSD* 결과를 보면 이상



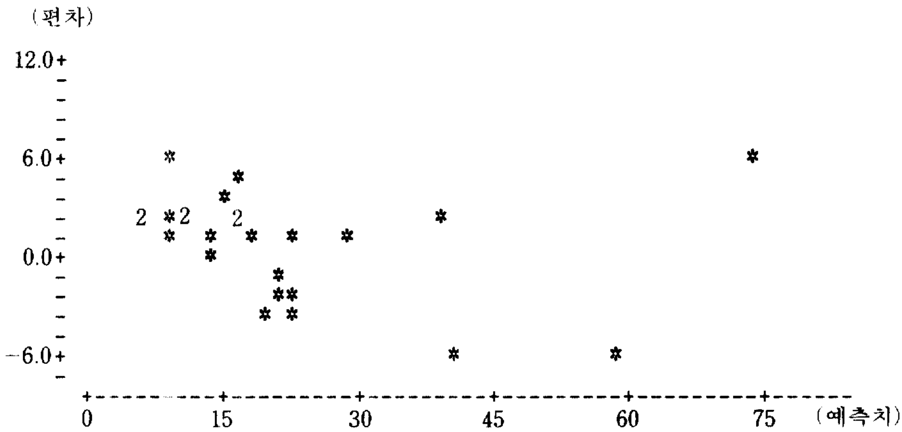
〈그림 1-1〉회귀분석의 예측치와 편차의 산포도



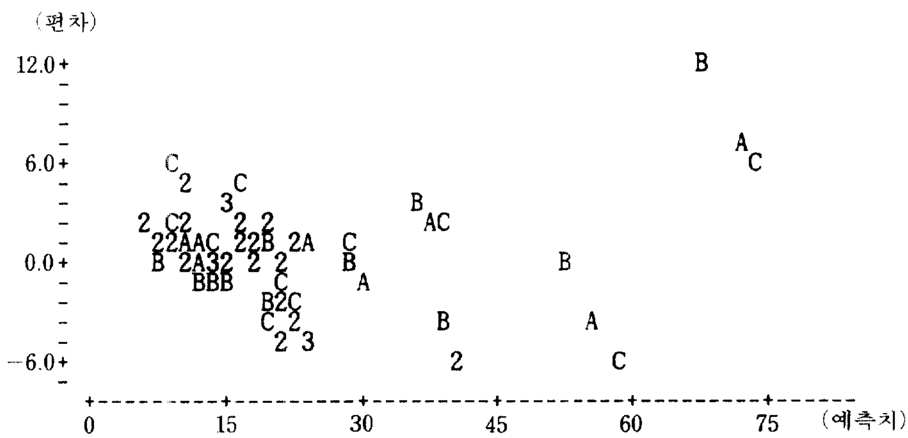
〈그림 1-2〉*MSD*의 예측치와 편차의 산포도

치로 생각되는 점이 보통의 편차보다 더 크게 나타나고 있다는 것을 알 수 있다.

끝으로 주어진 자료에 대해 모형의 적합성을 설명하기 위해서 분산분석(analysis of variance)에 의해서 F검정을 실시하고 결정계수로서 모형이 자료에 얼마나 적합한가를 설명하면 〈표 2〉와 같은 분산분석 결과를 얻을 수 있다.



〈 그림 1-3 〉 *MMD*의 예측치와 편차의 산포도



〈 그림 2 〉 예측치와 편차의 다중 산포도

A = 회귀분석의 예측치대 편차
B = MSD의 예측치대 편차
C = MMD의 예측치대 편차

〈 표 2 〉 각 방법의 분산분석 결과

	변동요인	제곱합	자유도	제곱평균	F통계량	$F(2, 22 : 0.01)$
회귀분석	회귀변동 편차변동 총 변 동	5552.57 231.97 5784.54	2 22 24	2776.29 10.54	263.41	5.72
MSD	회귀변동 편차변동 총 변 동	4856.08 928.46 5784.54	2 22 24	2428.04 42.20	57.54	5.72
MMD	회귀변동 편차변동 총 변 동	5398.53 386.01 5784.54	2 22 24	2699.57 17.55	153.82	5.72

분산분석은 변수들이 관측치에 유의한 영향을 미치는가를 판단하기 위한 것으로 F검정을 하게 되는데, 이때의 가설은,

$$H_0 : \beta_0 = \beta_1 = \beta_2 = 0$$
$$H_1 : \beta_0 = \beta_1 = \beta_2 \neq 0$$

이다. 검정결과는 세가지 방법 모두 기각으로 변수들이 관측치에 영향을 미쳤다는 것을 알 수 있다. 추가적인 설명으로 F통계량의 값이 크면 클수록 적합도가 높다는 것을 알 수 있다.

결정계수를 계산한 결과는 회귀분석에서는 0.96, *MSD*에서는 0.84, *MMD*에서는 0.93이었다. 〈 표 1 〉에서 계산의 정확성 면에서는 *MSD*에 의한 방법이 다른 방법에 비해 보다 좋은 결과를 얻었으나 적합정도에 있어서는 84%만이 모형에 의해 설명되고 16%는 오차에 의해 설명된다는 것을 알 수 있다. 반면 회귀분석에 의한 방법은 모형으로 설명할 수 있는 것이 96%가 됨으로서 앞서의 분산분석에 의한 F검정 결과와 마찬가지로 모형과 자료의 적합도가 상당히 높다는 것을 알 수 있다.

5. 결 론

일반적으로 회귀분석을 주로 사용한 사람들은 회귀분석을 가장 좋은 방법으로 생각하고, 선형계획법을 주로 사용한 사람들은 또한 선형계획법을 선호한다고 하였다 [4].

그러나 본 연구에서 접근한 세가지 방법에 대한 결과를 분석해 보면 예측의 정확성 면에서는 *MSD*, 회귀분석, *MMD*의 순으로 나타났다. 그러나 산포도로서 분석한 결과에서는 뚜렷한 특징을 찾기는 어려웠지만 다중 산포도에서 보듯 비슷한 산포 정도를 나타냈고, 회귀분석과 *MSD*법이 이상치에서 편차를 더 심하게 나타냈다.

통계적인 검정 결과를 보면 예측의 정확성과는 달리 모형의 적합성으로 볼 때 역시 회귀분석이 가장 적합한 것으로 나타났고 또한 *MSD*보다는 *MMD*가 더 적합성이 높은 것으로 나타났다.

앞으로의 연구는 현장 등 실제 문제에 대한 실증적 연구(empircal approach)를 통하여 정확성과 적합성을 높여줄 수 있는 방법에 관한 연구가 계속해서 이루어져야 할 것으로 생각된다.

參 考 文 獻

1. Cavalier, T. M., J. P., Ignizio, and A. L., Soyster, (1989), "*Discriminant analysis via mathematical programming : Certain problems and their causes*," Computers and Operation Research, Vol. 16, No. 4, pp 353 – 362.
2. Charnes, A., W. W., Cooper, and T., Sueyoshi, (1986), "*Least squares/ridge regression and goal programming / constrained regression alternatives*," European Journal of Operational Research, 27, pp. 146 – 157.
3. Freed, N., and F. Glover, (1986), "*Evaluating alternative linear programming models to solve the two-group discriminant problem*," Decision Sciences, Vol. 17, pp. 151 – 162.
4. Glorfeld, L. W., and N., Gaither, (1982), "*On using linear programming in discriminant problems*," Decision Sciences, Vol. 13, pp. 167 – 171.
5. Ignigio, J. P., (1982), "*Linear Programming in Single-& Multiple-Objective Systems*," Prentice-Hall, N. J., pp. 243 – 253.
6. Montgomery, D. C., and E. A., Peck, (1982), "*Introduction to Linear Regression Analysis*," Wiley, New York, pp. 109 – 176.
7. Wagner, H. M., (1959), "*Linear programming techniques for regression analysis*," Journal of American Statistical Association, 54, pp. 206 – 212.
8. Weisberg, S., (1980), "*Applied Linear Regression*," Wiley. pp. 31 – 57.
9. 박성현, (1983), 회귀분석, 대영사, pp. 171 – 187.