응용통계연구 제 11권 2호, 1998년, pp. 451-459

다중반응치 자료에 대한 순차적 BIPLOT활용에 대한 연구 *

장대홍[1)]

요 약

반응표면분석에서 다반응값의 최적화 문제는 단반응값 최적화문제보다 복잡하다. 이런 다반응값 문제에서 반응변수들이나 설명변수 상호간의 관계나 중요성 등을 평가하는 것은 중요하다. 이러한 평가를 위하여 biplot를 이용할 수 있는데, 1차 회귀모형이 적합치 않은 경우, 2차 회귀모형을 위한 순차적 실험계획을 이용하여 2차 회귀모형에 대응되는 biplot를 그려 선형 및 비선형효과를 알 수 있게 된다.

1. 서 론

반응표면분석에서 다반응값(multiresponse)의 최적화 문제는 단반응값(single response)의 최적화 문제보다 복잡하다. 이런 다반응값 문제에서 반응변수들이나 설명변수들 상호간의 관계나 중요성 등을 평가하는 것은 중요하다. 이러한 평가를 위하여 biplot가 유용한 그림도구로 쓰일 수 있다. biplot는 자료행렬을 비정치값분해를 이용하여 행표시자와 열표시자로 나타내어 저차원(2차원 내지 3차원) 행렬로 근사시켜 2차원 평면이나 3차원 공간상에 그림으로 나타내어 자료행렬 상의 각 변수와 관측값들 상호간의 관련성을 한 눈에 알아 볼 수 있게 하는 그림도구이다. 이 biplot는 Gabriel(1971)이 제안한 이후로 최근까지도 여러 학자들에 의하여 연구되고 있다 (Gower와 Hand(1996)을 보라.). Smith와 Cornell(1993)은 다반응값 혼합물 실험자료에 이 biplot를 이용하여 반응변수들이나 설명변수들 상호간의 관계나 중요성 등을 연구하였다. 장대홍(1996)은, 이 biplot가 다반응값 혼합물 실험에서 뿐만이 아니라 다반응값 반응표면분석, 크게 보면 다변량회귀분석에서도 쓰일 수 있음을 보였다. 그러나, 두 논문 모두 1차 회귀모형만을 가정하고 내용을 전개하였다. 즉, 반응변수들에 대한 설명변수들의 효과를 선형인 경우로 한정하였다.

본 논문을 통하여 1차모형이 적합치 않은 경우에 2차 회귀모형을 위한 순차적 실험계획을 이용하여 2차 회귀모형에 대응되는 biplot를 제시함으로써 반응변수들에 대한 설명변수들의 효과를 선형뿐만이아니라, 비선형인 경우로 확대 해석하는데 사용하고자 한다.

2. 순차적 BIPLOT

m개의 반응변수들이 각각 k개의 설명변수들로 설명된다고 가정하고, k개의 코드화된 설명변수들로 이루어진 l번째 반응변수에 대한 1차 회귀모형을

* 이 연구는 96년도 학술진흥재단 지방대 육성과제 연구비 지원에 의한 결과임.
1) (608-737) 부산광역시 남구 대연3동 599-1, 부경대학교 자연과학대학 수리과학부, 교수.

$$y_l = \beta_{0l} + \beta_{1l}x_1 + \beta_{2l}x_2 + \cdots + \beta_{kl}x_k + \varepsilon,\ l = 1, 2, \cdots, m \tag{2.1}$$

라 하고, 2차회귀모형을

$$y_l = \beta_{0l} + \sum_{i=1}^{k} \beta_{il}x_i + \sum_{i \le j}^{k} \beta_{ijl}x_ix_j + \varepsilon,\ l = 1, 2, \cdots, m \tag{2.2}$$

라 하자. 여기서, $\varepsilon \sim (0, \sigma^2)$ 이고, 서로 독립이다. 우선적으로 1차회귀모형을 적합시키기 위하여 n_1개의 실험점으로 이루어진 실험계획을 시행하여 반응값들을 측정하였을때, y_{il}를 l번째 반응변수의 i번째 관측값 $(i = 1, 2, \cdots, n_1; l = 1, 2, \cdots, m)$이라고 하자. y_{il}를 표준화 시키면,

$$Z_{il} = \frac{y_{il} - \overline{y_l}}{\sqrt{\sum_{i=1}^{n_1} (y_{il} - \overline{y_l})^2}} \tag{2.3}$$

이 된다. 여기서, $\overline{y_l} = \sum_{i=1}^{n_1} y_{il}/n_1$ 이다. 그러면, l번째 표준화된 반응변수에 대한 회귀식을 행렬로 표시하면,

$$\underline{Z_l} = X_1\underline{\beta_l} + \underline{\varepsilon_l},\ l = 1, 2, \cdots, m \tag{2.4}$$

으로 표시할 수 있다. 여기서, X_1는 1차 모형행렬이고, $\underline{\beta_l}$은 l번째 반응변수에 대응되는 $k \times 1$ 모수 벡터이다. 그래서, $n_1 \times m$ 표준화반응행렬 Z_1이 정의될수 있는데, Z_1의 l번째 열이 $\underline{Z_l}$이다. 식 (2.4)을 이용하여 $\underline{\beta_l}$에 대한 추정값 $\underline{b_l}$을 구하면, 다음과 같은 $k \times m$ 회귀계수 행렬

$$\begin{aligned} B_1 &= [\underline{b_1}, \underline{b_2}, \cdots, \underline{b_m}] \\ &= (X_1'X_1)^{-1}X_1'Z_1 \end{aligned} \tag{2.5}$$

이 구하여 진다. l번째 표준화된 반응변수의 추정값은

$$\hat{Z}_l(\underline{x}) = b_{1l}x_1 + b_{2l}x_2 + \cdots + b_{kl}x_k \tag{2.6}$$

로 표시할 수 있는데, l번째 반응변수에 대한 i번째 설명변수$(i = 1, 2, \cdots, k)$ 의 1차효과는 b_{il}로 나타내어진다. 그리하여, B_1을 비정칙값분해를 행하여 biplot를 작성하면 반응변수들이나 설명변수들 상호간의 관계나 중요성 등을 평가할 수 있게 된다.

B_1의 계수(rank)를 r이라 할때 $(r \le min(k, m))$ B_1에 대해 비정칙값분해를 행하여 rank-two 행렬 B_1^*를 만들어 GH^T biplot와 j plot를 구하면, GH^T biplot의 열표시자를 통하여 반응변수들 간의 상관관계를 알 수 있고, 행표시자들과 열표시자들을 통하여는 반응변수들과 설명변수들 간의 관계를 규명할 수 있게 되고, j plot을 통하여는 설명변수들의 중요도를 알

아낼 수 있게 된다 (GH^T biplot와 j plot를 만드는 방법과 해석은 Smith와 Cornell(1993)과 장대홍(1996)을 참조하시오.).

1차모형이 적합치 않은 경우 2차 모형을 적합시키기 위하여 n_2개의 실험점을 더 추가할 때, 2가지 방법이 있다. 첫번째 방법은 n_1개의 실험점들로 구성된 기존의 실험계획을 첫번째 블록으로 보고, 새로 추가된 n_2개의 실험계획을 두번째 블록으로 보아 블록효과를 고려한 2차모형을 생각하여 볼 수 있고, 두번째 방법은 $n = n_1 + n_2$개의 실험점으로 이루어진 실험계획을 재시행하여 2차모형을 적합시켜 볼 수 있다. 두번째 방법으로 실험을 시행하여 반응값을 측정하였을 때, y_{il}을 l번째 반응변수의 i번째 관측값($i = 1, 2, \cdots, n; l = 1, 2, \cdots, m$)이라 하자. y_{il}을 표준화시키면,

$$Z_{il} = \frac{y_{il} - \overline{y_l}}{\sqrt{\sum_{i=1}^{n}(y_{il} - \overline{y_l})^2}} \tag{2.7}$$

이 된다. 여기서, $\overline{y_l} = \sum_{i=1}^{n} y_{il}/n$이다. 그러면, l번째 표준화된 반응변수에 대한 회귀식을 행렬로 표시하면,

$$\underline{Z_l} = X_2\underline{\beta_l} + \underline{\varepsilon_l}, \ l = 1, 2, \cdots, m \tag{2.8}$$

로 표시할 수 있다. 여기서, X_2는 2차 모형행렬이고, $\underline{\beta_l}$은 $p \times 1$ 모수벡터이다 ($p = 2k + \binom{k}{2}$). 그래서, $n \times m$표준화반응행렬 Z_2가 정의될 수 있는데, Z_2의 l번째 열이 $\underline{Z_l}$이다. 식 (2.8)을 이용하여 $\underline{\beta_l}$ 에 대한 추정값 $\underline{b_l}$을 구하면, 다음과 같은 $p \times m$ 회귀계수행렬

$$\begin{aligned} B_2 \quad &= [\underline{b_1}, \underline{b_2}, \cdots, \underline{b_m}] \\ &= (X_2'X_2)^{-1}X_2'Z_2 \end{aligned} \tag{2.9}$$

이 구하여 진다. l번째 표준화된 반응변수의 추정값은

$$\hat{Z}_l(\underline{x}) = \sum_{i=1}^{k} b_{il}x_i + \sum_{i \le j}^{k} b_{ijl}x_ix_j, \ i = 1, 2, \cdots, k \tag{2.10}$$

로 표시할 수 있는데, l번째 반응변수에 대한 i번째 설명변수의 1차효과는 b_{il}로, 2차효과는 b_{ijl}로 표시된다. B_2에 대하여 비정칙값분해를 행하여 rank-two행렬 B_2^*를 만들어 GH^T biplot와 j plot을 구하면, GH^T biplot을 통하여서는 반응변수들 간의 상관관계와 반응변수들과 설명변수들 간의 관계를 알 수 있게 되고, j plot을 통하여서는 설명변수들의 중요도를 알 수 있게 된다.

3. 수치 예

Montgomery 와 Voth(1994)에 나타나는 항공기 승무원 피난시스템에서 쓰이는 발사화약 연구에서 쓰이는 혼합물 성분과 반응변수는 다음 표 3.1과 같고, 실험결과는 표 3.2와 같았다.

1차회귀모형이 적합한 지를 보기 위하여 실험점 1-9로 구성된 실험점군들을 가지고 다음 표 3.3같이 회귀분석을 하여 보았다. 표 3.3에서 알 수 있듯이 y_1에서는 1차모형을 적용하기가 어렵다. 그러므로, 2차회귀모형으로 적합하는 것이 타당하다. 여러가지 변수선택판정 통계량들(결정계수, 수정결정계수, 잔차제곱평균, C_p 등의, SAS가 제공하는 12가지 정도의 통계량)의 결과나 순차제곱합의 계산결과를 보아도 2차회귀모형을 적합시키는 것이 타당하다. 다른 확인방법으로서, 실험이 꼭지점들에서 2반복, 중심점에서 3반복되므로 1차회귀모형에 대한 적합된 검정을 행하여 보면 2차회귀모형을 사용하는 것이 타당함을 알 수 있다. 15개 실험점 모두를 사용하여 2차회귀모형을 적합하니 다음 표 3.4와 같았다.

표 3.4를 사용하여 biplot을 그려보니 그림 3.1과 그림 3.2와 같았다. 2차모형을 이용한 biplot과 1차모형을 이용한 biplot가 그림 3.1과 그림 3.3에서 보는 바와 같이 아주 다른 모습을 나타내기 때문에 2차모형을 이용한 biplot를 이용하여 자료분석을 하여야 한다. 이 그림들을 통하여 다음과 같은 사실을 알 수 있었다.

1. 반응변수인 연소율과 연소율의 표준편차와는 음의 상관관계가 강한 반면 연소율과 제조지수와는 음의 상관관계가 약하다. 연소율의 표준편차와 제조지수와는 상관관계가 거의 없다.

2. 설명변수인 혼합연료와 제조지수와는 강한 양의 상관관계가 있고, 설명변수인 산화제는 연소율의 표준편차와 강한 양의 상관관계를 갖는 한편, 연소율과는 강한 음의 상관관계를 갖는다. 설명변수인 접합제는 제조지수와 강한 음의 상관관계를 갖는다.

3. 혼합연료와 산화제의 2차효과는 제조지수와 음의 상관관계가 있고, 혼합연료와 접합제의 2차효과는 연소율과 양의 상관관계가 있다. 산화제와 접합제의 2차효과는 연소율과 양의 상관관계를, 연소율의 표준편차와는 음의 상관관계를 갖는다.

4. 2차효과들이 1차효과들보다 반응변수에 더 크게 작용한다. 설명변수들의 복합적인 효과는 산화제와 접합제의 2차효과(2.9983)>혼합연료와 접합제의 2차효과(1.8131)>혼합연료와 산화제의 2차효과(1.1221)> 혼합연료의 1차효과(0.6497)>접합제의 1차효과(0.6059)>산화제의 1차효과(0.3508) 순이다. 여기서, 주의해야 할 사항은 이 실험은 혼합물 실험이므로

표 3.1: 발사화약 연구에서의 변수들

설명변수	x_1	혼합연료
	x_2	산화제
	x_3	접합제
반응변수	y_1	연소율 (cm/sec)
	y_2	연소율의 표준편차
	y_3	제조지수

표 3.2: 실험결과

실험점	계 획			반응값		
	x_1	x_2	x_3	y_1	y_2	y_3
1	1.0	0.0	0.0	32.5	4.1	32
2	1.0	0.0	0.0	37.9	3.7	25
3	0.0	1.0	0.0	54.5	8.9	18
4	0.0	1.0	0.0	32.5	9.2	21
5	0.0	0.0	1.0	64.0	14.0	14
6	0.0	0.0	1.0	78.5	13.0	16
7	0.333333333	0.333333333	0.333333333	112.5	4.6	19
8	0.333333333	0.333333333	0.333333333	98.5	3.5	20
9	0.333333333	0.333333333	0.333333333	103.6	3.0	18
10	0.5	0.5	0.0	44.0	6.8	20
11	0.5	0.0	0.5	63.2	4.7	18
12	0.0	0.5	0.5	94.0	4.5	17
13	0.666666666	0.166666666	0.166666666	67.1	3.5	20
14	0.166666666	0.666666666	0.166666666	73.0	5.2	22
15	0.166666666	0.166666666	0.666666666	87.5	7.0	17

순수2차 효과는 없고, 혼합2차 효과만 있게 된다.

표 3.3: 1차회귀모형에 대한 R^2과 p-값

	y_1	y_2	y_3
R^2	0.1802	0.6296	0.8217
p-값	0.5510	0.0508	0.0057

표 3.4: 2차회귀모형을 위한 B_2행렬, R^2과 p-값

	y_1	y_2	y_3
x_1	-0.3669	-0.1927	0.5165
x_2	-0.2924	0.2051	0.0020
x_3	-0.0102	0.5595	-0.3043
x_1x_2	0.8579	-0.0147	-0.7554
x_1x_3	1.0657	-1.2819	-0.8446
x_2x_3	2.0936	-2.1349	0.3152
R^2	0.8247	0.9830	0.8106
p-값	0.0032	0.0001	0.0045

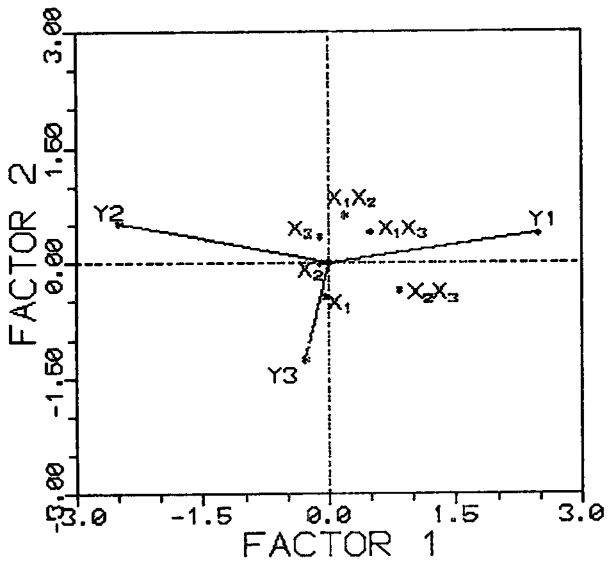

그림 3.1: 2차회귀모형에 대한 GH^T biplot

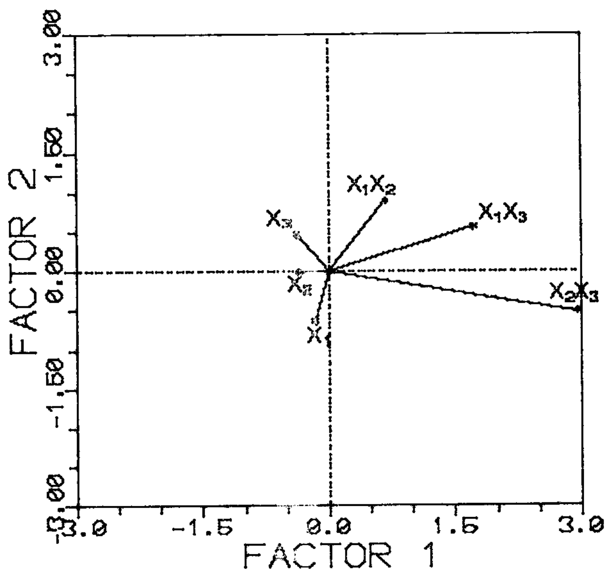

그림 3.2: 2차회귀모형에 대한 j plot

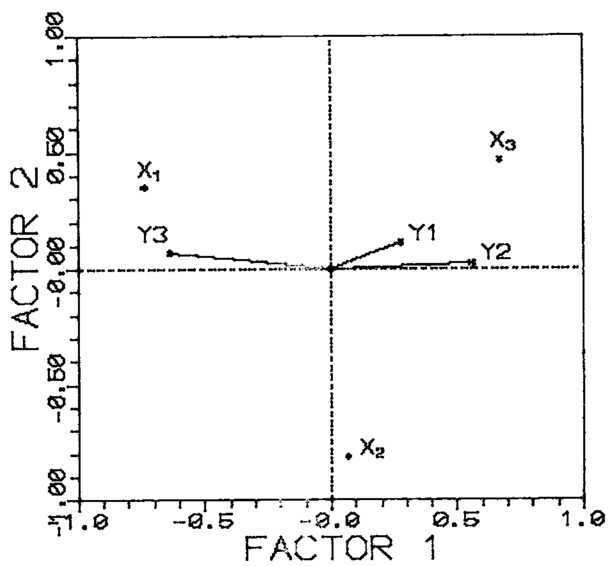

그림 3.3: 1차회귀모형에 대한 GH^T biplot

4. 결 론

다반응값 반응표면분석에서 반응변수들이나 설명변수들 상호간의 관계나 중요성 등을 평가하는 것은 자료분석시 매우 중요하다. biplot은 이러한 다반응값 반응표면분석에서 반응변수들 간의 상관관계나 설명변수들의 효과를 하나의 그림으로 나타낼 수 있는 유용한 도구이다. GH^T biplot을 통하여는 반응변수들 간의 상관관계나 반응변수들과 설명변수들 간의 관계를 규명할 수 있고, j plot을 통하여는 설명변수들의 중요도를 알아낼 수 있다. 1차회귀모형이 적합한 경우 이를 이용한 biplot을 작성하면 되고, 1차회귀모형이 적합치 않은 경우 실험점들을 첨가시켜 2차회귀모형을 적합시켜 이를 이용한 biplot을 작성한다.

참고문헌

[1] 장대홍(1996). 다반응값 자료에 대한 biplot 활용에 관한 연구, 한국통계학회 논문집, 제3권 제1호, 1–9.

[2] Gabriel K. R.(1971). The Biplots–Graphics Display of Matrices with Application to Principle Component Analysis, *Biometrika*, Vol. 58, 453–467.

[3] Gower, J. C. and Hand, D. J.(1996). *Biplots*, Chapman and Hall, London.

[4] Montgomery, D. C. and Voth, S. R.(1994). Multicollinearity and Leverage in Mixture Experiments, *Journal of Quality Technology*, Vol. 26, 96–108.

[5] Smith, W. F. and Cornell, J. A.(1993). Biplots Display for Looking at Multiple Response Data in Mixture Experiments, *Technometrics*, Vol. 35, No. 4, 337–350.

[1998년 4월 접수, 1998년 7월 최종수정]

A Study of Applications of Sequential Biplots in Multiresponse Data *

Dae-Heung Jang[1)]

ABSTRACT

The analysis of data from a multiresponse experiment requires careful consideration of the multivariate nature of the data. In a multiresponse situation, the optimization problem is more complex than in the single response case. The biplot is a graphical tool which make the analyst to understand the correlation of the response variables, the relation of the response variables and the explanatory variables and the relative importance of the explanatory variables. In case of good fitting of the first order model, we can draw the biplot with the first order experimetal design. Otherwise, we can make the biplot with the second order experimetal design by adding other experimental points.

* This research was supported by the Korean Research Foundation, 1996.

1) Division of Mathematical Science, Pukyong National University, Pusan, 608-737 Korea.