

## 일반화선형모형에서의 3차원 CERES그림\*

강명욱<sup>1)</sup> 김부용<sup>2)</sup> 진진영<sup>3)</sup>

### 요약

일반화선형모형에서 추가되는 설명변수의 비선형성의 존재와 형태를 파악하는데 사용되는 3차원 CERES그림의 구조와 유용성에 대해 알아본다. CERES그림은 설명변수들 사이에 비선형의 관계가 존재하는 경우 편잔차그림으로는 알아낼 수 없는 비선형성에 대한 탐지가 가능하다. 이를 생성된 자료를 이용하여 확인해 본다.

주요용어: 일반화선형모형, 편잔차그림, CERES그림, LOESS.

### 1. 서론

정규이론에 근거한 선형모형에서는 설명변수의 선형결합(linear predictor)이 직접 반응변수를 설명한다. 이러한 선형모형은 분산분석, 선형회귀분석 등에서 매우 다양한 모형의 틀을 제공하는 것이 사실이지만 모든 상황에서 충분한 것은 아니다. 반응변수가 이항반응자료처럼 이산형인 경우에는 일반 선형모형은 적절하지 않고 정규이론에 의한 고전적 선형모형을 확장한 일반화선형모형(generalized linear models)으로 해결이 가능하다.

회귀분석에서 기존에 사용되고 있는 모형에 새로운 설명변수를 추가하고자 할 때 추가변수가 회귀모형에 어떤 영향을 미치는지를 알아보아야 하고 추가될 때의 형태에 대한 판단이 필요하며 선형이 아닌 형태로 추가되어야 한다면 어떤 형태로 변환되어야 하는지를 진단하여야 한다. 이러한 진단은 그림을 통한 방법으로 가능하며 지금까지 활발히 연구가 진행되고 있는 방법으로 추가변수그림(added variable plot)과 편잔차그림(partial residual plot), 그리고 편잔차그림을 확장시킨 덧편잔차그림(augmented partial residual plot)과 CERES그림(combining conditional expectations and residuals plot)이 있다.

편잔차그림은 Ezekiel (1924)에 의해 처음 제시되었고 Cook과 Weisberg (1982) 등에 의해 회귀진단의 도구로 사용되었다. 편잔차그림을 개선한 덧편잔차그림은 Mallows (1986)에 의해 소개되었고 Cook (1993)은 더욱 일반화시킨 CERES그림을 새로운 진단방법으로 제시하였다. 이러한 그림을 이용한 비선형성의 탐색은 Atkinson (1985), Chatterjee와 Hadi

\* 본 연구는 숙명여자대학교 2006년도 교내연구비 지원에 의해 수행되었음.

1) (140-742) 교신저자. 서울특별시 용산구 청파동 2가, 숙명여자대학교 통계학과, 교수.

E-mail: mwkahng@sm.ac.kr

2) (140-742) 서울특별시 용산구 청파동 2가, 숙명여자대학교 통계학과, 교수.

E-mail: buykim@sm.ac.kr

3) (140-742) 서울특별시 용산구 청파동 2가, 숙명여자대학교 통계학과, 석사과정.

E-mail: syoory@nate.com

(1988), Cook과 Weisberg (1999) 등에 의해 회귀진단의 도구로 많은 연구가 있어 왔다. 최근에는 Kahng과 Jeong (2000)이 비선형모형에서의 CERES그림에 대한 연구로 발전시켰고, Kahng과 Lee (2004)는 2차원 CERES그림을 통해 일반화선형모형에서 추가변수의 유의성 및 비선형성에 대한 진단으로 확장하였다.

본 연구에서는 일반화선형모형에서 CERES그림을 통한 추가변수의 필요성과 함수의 형태에 대해 연구하고자 한다. 2절에서는 일반화선형모형을 소개하고 3절에서는 3차원 CERES그림을 일반화선형모형에 적용시켜본다. 4절에서는 일반화선형모형의 하나인 이항회귀모형(binomial regression model)에서 CERES그림을 통하여 추가변수의 형태를 알 수 있는지를 확인해본다.

## 2. 일반화선형모형

Nelder와 Wedderburn (1972)에 의해 체계화된 일반화선형모형은 지수족(exponential family) 분포와 연결함수(link function)를 이용하여 정규이론에 의한 선형모형을 다음과 같은 두 가지 과정을 통해 일반화한 것이다. 첫째, 오차의 분포는 정규분포를 포함하는 지수족의 여러 가지 분포를 사용한다. 둘째, 반응변수의 기대값과 설명변수의 선형결합을 연결시키는 연결함수를 설정한다. 고전적 선형모형은 반응변수가 서로 독립적이며 정규분포를 따르고 연결함수가 항등함수(identity function)인 일반화선형모형의 특수한 형태라고 할 수 있다.

일반화선형모형에서 반응변수  $Y$ 는 설명변수  $\mathbf{x}^T = (x_1, \dots, x_p)$ 의 선형결합  $\eta$ 를 통하여 결정된다. 이때 선형결합  $\eta$ 는 연결함수  $g(\cdot)$ 에 의하여  $\mu = E(Y)$ 와 연결된다. 설명변수  $\mathbf{x}^T$ 의  $n$ 개의 관측값에 대하여 선형결합은

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta} = \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad i = 1, \dots, n \quad (2.1)$$

이다. 선형결합 (2.1)에 추가시키려고 하는 새로운 설명변수  $z$ 가 선형 또는 비선형함수  $h(z)$ 의 형태로 추가된다고 하면 선형결합을 다음과 같이 나타낼 수 있고

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}^* + h(z_i), \quad i = 1, \dots, n \quad (2.2)$$

이 모형 (2.2)를 설명변수와 반응변수의 관계를 나타내는 참모형이라고 하자. 여기서  $h(\cdot)$ 는 선형 또는 형태가 알려지지 않은 비선형함수이다.

## 3. CERES그림

### 3.1. 2차원 CERES그림

선형모형에서와 마찬가지로 일반화선형모형에서도 추가변수  $z$ 에 대한 변환의 필요성을 확인하기 위하여 잔차산점도와 잔차-설명변수산점도를 그려보고 만약 비선형의 형태가 나타나면  $z$ 가 어떠한 형태의 비선형함수  $h(z)$ 로 변환되어야 하는지를 알아보기 위해 편잔

차그림을 사용할 수 있다. 만약 모형 (2.2)에서  $h(z)$ 의 함수형태가  $z$ 의 선형함수라면 이 모형은 다음과 같은 모형이 된다.

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha z_i, \quad i = 1, \dots, n. \quad (3.1)$$

추가변수  $z$ 의 편잔차그림은 모형 (3.1)에 적합시킨 후 얻어지는 잔차  $e$ 에  $\hat{\alpha}z$ 를 더한 편잔차를 세로축으로 하고 추가변수  $z$ 를 가로축으로 하는  $\{e_i + \hat{\alpha}z_i, z_i\}$ 의 산점도이다. 이 때  $\hat{\alpha}$ 은 모형 (3.1)에서 구한  $z$ 의 회귀계수추정값이다.

만약  $h(z)$ 가  $z$ 에 대한 선형함수라면 편잔차그림은  $h(z)$ 의 형태를 잘 나타내게 된다.  $h(z)$ 가  $z$ 에 대한 비선형인 경우에도 Cook (1993)에 의하면 모형 (2.2)가 참모형이고 조건부기대값  $E(\mathbf{x}|z)$ 이  $z$ 와 연관성이 약하거나  $z$ 에 대해 선형이면 모형 (3.1)에서의 추정량은  $\boldsymbol{\beta}^*$ 의 일치추정량이 되므로 편의가 없어져서 편잔차그림은  $h(z)$ 의 형태를 나타낼 수 있다. 그러나  $E(\mathbf{x}|z)$ 가  $z$ 에 대해 선형이 아니면 편잔차그림은  $h(z)$ 의 형태에 대한 올바른 정보를 주지 못하며 비선형성을 감지할 수 없다.

일반화선형모형에서 다음과 같은 선형결합을 생각해보자.

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + \boldsymbol{\alpha}^T \mathbf{m}(z_i), \quad i = 1, \dots, n. \quad (3.2)$$

여기서  $\mathbf{m}(z) = \{m_j(z)\}$ 는  $p \times 1$ 벡터이고 조건부기대값  $E(\mathbf{x}|z)$ 로 정의하며  $\mathbf{m}(z)$ 의 값이 사전에 알려지지 않았기 때문에  $m_j(z)$ ,  $j = 1, \dots, p$ 는 각각  $x_j$ 를 반응변수로하고  $z$ 에 의해 비모수회귀를 시켜 구한 적합값으로 다음과 같이 추정한다.

$$\hat{\mathbf{m}}(z) = \{\hat{m}_j(z)\} = \{\hat{E}(x_j|z)\}, \quad j = 1, \dots, p. \quad (3.3)$$

비모수회귀적합은 이미 알려져 있는 여러 가지 방법 중 어느 것을 사용하여도 그 결과에 거의 차이가 없다 (Cook, 1993). 여기서는 그 중 LOESS방법을 사용한다. 모형 (3.2)에서  $\mathbf{m}(z_i)$ 를 식 (3.3)의  $\hat{\mathbf{m}}(z_i)$ 로 바꾸면 다음과 같은 모형이 된다.

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + \boldsymbol{\alpha}^T \hat{\mathbf{m}}(z_i), \quad i = 1, \dots, n. \quad (3.4)$$

추가변수  $z$ 의 CERES그림은 모형 (3.4)에 적합시킨 후 얻어지는 잔차  $e$ 에  $\hat{\boldsymbol{\alpha}}^T \hat{\mathbf{m}}(z)$ 를 더한 CERES를 세로축으로 하고 추가변수  $z$ 를 가로축으로 하는  $\{e_i + \hat{\boldsymbol{\alpha}}^T \hat{\mathbf{m}}(z_i), z_i\}$ 의 산점도이다. 이 때  $\hat{\boldsymbol{\alpha}}$ 은 모형 (3.4)에서 구한  $\hat{\mathbf{m}}(z)$ 의 회귀계수추정값벡터이다. 모형 (2.2)가 참모형이라면 CERES그림의 세로축을 이루는 CERES의 기대값은 다음과 같다.

$$E(e_i + \hat{\boldsymbol{\alpha}}^T \hat{\mathbf{m}}(z_i)|z_i) = E(\mathbf{x}_i|z_i)^T (\boldsymbol{\beta}^* - \boldsymbol{\beta}) + h(z_i). \quad (3.5)$$

모형 (3.4)의  $\boldsymbol{\beta}$ 가 참모형 (2.2)에서의 실제모수  $\boldsymbol{\beta}^*$ 에 근접하면 식 (3.5)에 의해 CERES그림은  $h(z)$ 의 형태를 잘 묘사할 수 있다. 모형 (3.2)와 (3.4)에서  $\mathbf{m}(z)$ 를  $E(\mathbf{x}|z)$ 로 정의하고  $\hat{E}(\mathbf{x}|z)$ 로 추정하였으므로  $E(\mathbf{x}|z)$ 는 당연히  $m_j(z)$ 의 선형함수로 볼 수 있으며 모형 (3.4)에서  $\boldsymbol{\beta}$ 의 추정량은 실제모수  $\boldsymbol{\beta}^*$ 의 일치추정량이라 할 수 있고, 따라서 식 (3.5)에서 편의가 없어지므로 CERES그림이  $h(z)$ 의 형태를 잘 묘사한다는 사실을 Kahng과 Lee (2004)에서 확인하였다.

### 3.2. 3차원 CERES그림

모형 (2.1)에 두 개의 설명변수  $z_1$ 과  $z_2$ 가 선형 또는 비선형함수  $h(z_1, z_2)$ 의 형태로 추가된다고 하면 추가변수의 효과를 나타내는 다음의 모형을 생각할 수 있고 이 모형을 참모형이라고 하자.

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}^* + h(z_{i1}, z_{i2}), \quad i = 1, \dots, n. \quad (3.6)$$

여기서  $h(\cdot, \cdot)$ 는 선형 또는 형태가 알려지지 않은 비선형함수이다.

조건부기대값  $E(\mathbf{x}|z_1, z_2)$ 가  $z_1$ 과  $z_2$ 의 선형함수로 표현되지 않을 경우 3차원 편잔차그림은  $h(\cdot, \cdot)$ 에 대한 정보를 제대로 제공하지 못한다. 이러한 문제점을 해결하기 위해서 보다 완화된 조건에서도 적절한 정보를 제공해 주는 3차원 CERES그림을 적용해 볼 수 있다. 다음의 선형결합을 생각해보자.

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + \boldsymbol{\alpha}_1^T \mathbf{m}_1(z_{i1}) + \boldsymbol{\alpha}_2^T \mathbf{m}_2(z_{i2}), \quad i = 1, \dots, n. \quad (3.7)$$

여기서  $\mathbf{m}_1(z_1) = \{m_{1j}(z_1)\}$ 과  $\mathbf{m}_2(z_2) = \{m_{2j}(z_2)\}$ 는  $p \times 1$ 벡터이고 각각 조건부기대값  $E(\mathbf{x}|z_1)$ 과  $E(\mathbf{x}|z_2)$ 로 정의하며  $\mathbf{m}_1(z_1)$ 과  $\mathbf{m}_2(z_2)$ 의 값들은 사전에 알려져 있지 않기 때문에 이들을 추정해야 한다.  $m_{1j}(z_1), j = 1, \dots, p$ 는 각각  $x_j$ 를 반응변수로하고  $z_1$ 에 의해 비모수회귀를 시켜 구한 적합값으로 추정한다. 또한  $\mathbf{m}_2(z_2)$ 도 같은 방법으로 추정한다. 즉,

$$\begin{aligned} \hat{\mathbf{m}}_1(z_1) &= \{\hat{m}_{1j}(z_1)\} = \{\hat{E}(x_j|z_1)\}, \\ \hat{\mathbf{m}}_2(z_2) &= \{\hat{m}_{2j}(z_2)\} = \{\hat{E}(x_j|z_2)\}, \quad j = 1, \dots, p \end{aligned} \quad (3.8)$$

이다. 모형 (3.7)에서  $\mathbf{m}_1(z_{i1})$ 과  $\mathbf{m}_2(z_{i2})$ 를 식 (3.8)의  $\hat{\mathbf{m}}_1(z_{i1})$ 과  $\hat{\mathbf{m}}_2(z_{i2})$ 로 바꾸면 다음과 같은 모형이 된다.

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + \boldsymbol{\alpha}_1^T \hat{\mathbf{m}}_1(z_{i1}) + \boldsymbol{\alpha}_2^T \hat{\mathbf{m}}_2(z_{i2}), \quad i = 1, \dots, n. \quad (3.9)$$

CERES그림은 모형 (3.9)에 적합시킨 후 얻어지는 잔차  $e$ 에  $\hat{\boldsymbol{\alpha}}_1^T \hat{\mathbf{m}}_1(z_1)$ 과  $\hat{\boldsymbol{\alpha}}_2^T \hat{\mathbf{m}}_2(z_2)$ 를 더한 CERES를 수직축으로 하고 추가변수  $z_1$ 과  $z_2$ 를 서로 직각을 이루는 두 개의 수평축으로 하는  $\{e_i + \hat{\boldsymbol{\alpha}}_1^T \hat{\mathbf{m}}_1(z_{i1}) + \hat{\boldsymbol{\alpha}}_2^T \hat{\mathbf{m}}_2(z_{i2}), z_{i1}, z_{i2}\}$ 의 3차원 산점도이다. 이 때  $\hat{\boldsymbol{\alpha}}_1^T, \hat{\boldsymbol{\alpha}}_2^T$ 는 모형 (3.9)에서 구한  $\hat{\mathbf{m}}_1(z_{i1})$ 과  $\hat{\mathbf{m}}_2(z_{i2})$ 에 대한 회귀계수추정값벡터이다.

모형 (3.6)이 참모형이라면 3차원 CERES그림의 수직축을 이루는 CERES의 기대값은 다음과 같다.

$$E(e_i + \hat{\boldsymbol{\alpha}}_1^T \hat{\mathbf{m}}_1(z_{i1}) + \hat{\boldsymbol{\alpha}}_2^T \hat{\mathbf{m}}_2(z_{i2})|z_{i1}, z_{i2}) = E(\mathbf{x}_i|z_{i1}, z_{i2})^T (\boldsymbol{\beta}^* - \boldsymbol{\beta}) + h(z_{i1}, z_{i2}) \quad (3.10)$$

3차원 CERES그림이  $h(z_1, z_2)$ 의 형태를 잘 나타내기 위해서는 모형 (3.9)의  $\boldsymbol{\beta}$ 가 참모형 (3.6)의 실제모수  $\boldsymbol{\beta}^*$ 에 근접해야 한다.  $h(z_1, z_2)$ 가  $m_{1j}(z_1), m_{2j}(z_2), j = 1, \dots, p$ 의 선형함수로 나타나는 경우  $h(z_1, z_2)$ 는  $\boldsymbol{\alpha}_1^T \mathbf{m}_1(z_1) + \boldsymbol{\alpha}_2^T \mathbf{m}_2(z_2)$ 로 근사가 가능하며 이 때 모형 (3.9)는 모형 (3.6)의 근사모형이 되고 모형 (3.9)에서의  $\boldsymbol{\beta}$ 는 모형 (3.6)에서의 실제모수  $\boldsymbol{\beta}^*$ 에 근접하게 된다. 따라서 3차원 CERES그림은  $h(z_1, z_2)$ 의 형태를 잘 묘사하게 된다. 그

러나  $h(z_1, z_2)$ 가  $\alpha_1^T \hat{m}_1(z_1) + \alpha_2^T \hat{m}_2(z_2)$ 로 근사가 되지 않으면  $\beta^*$ 와  $\beta$ 는 근접한다고 하기 어렵고 편이가 발생하므로 3차원 CERES그림은  $h(z_1, z_2)$ 의 형태를 제대로 나타내지 못할 것이라고 생각할 수도 있다.

하지만 모형 (3.9)에서  $m_1(z_1)$ 과  $m_2(z_2)$ 는 각각  $\hat{E}(x|z_1)$ 과  $\hat{E}(x|z_2)$ 로 추정하였으므로  $E(x|z_1)$ 과  $E(x|z_2)$ 는 당연히  $m_{1j}(z_1)$ 과  $m_{2j}(z_2)$ 의 선형함수로 볼 수 있다. 따라서 이 모형에서 구한  $\beta$ 의 최소제곱추정량  $\hat{\beta}$ 은  $\beta^*$ 의 일치추정량이라 할 수 있고 식 (3.10)에서의 편이는 없어지므로 3차원 CERES그림이  $h(z_1, z_2)$ 의 형태를 잘 나타내게 된다.

3차원 CERES그림의 수직축을 중심으로 회전할 때 보여지는 연속된 2차원의 산점도 중에서 비선형의 형태를 가장 잘 나타내는 산점도를 생각해 보자. 이 산점도의 가로축은 회전각도인  $\omega$ 의 함수인  $z(\omega)$ 라 표현할 수 있다. 즉  $z(\omega)$ 는 3차원 CERES그림을 수직축으로 회전할 때 나타나는 2차원 산점도 중 CERES를 세로축으로 하고  $z_1$ 을 가로축으로 하는 산점도를 시작으로 하여  $\omega$ 만큼 회전한 후에 나타나는 산점도에서의 가로축으로 다음과 같이  $z_1$ 과  $z_2$ 의 선형결합의 형태로 표현할 수 있다.

$$z(\omega) = z_1 \cos(\omega) + z_2 \sin(\omega).$$

두 개의 설명변수  $z_1$ 과  $z_2$ 의 어떠한 비선형함수  $h(z_1, z_2)$ 를 두 변수의 선형결합에 대한 비선형함수  $h^*(z(\omega))$ 로 근사할 수 있다면 다음의 모형을 참모형 (3.6)의 근사모형이라고 할 수 있다.

$$\eta_i = \mathbf{x}_i^T \beta^* + h^*(z(\omega)), \quad i = 1, \dots, n.$$

따라서  $h^*(\cdot)$ 의 형태는  $z(\omega)$ 의 2차원 CERES그림을 통해 얻을 수 있고 이는 3차원 CERES그림의 수직축을 중심으로 하는 회전에서 비선형형태를 가장 잘 나타내는 산점도와 일치한다.

#### 4. 예제

일반화선형모형에서  $E(x|z_1, z_2)$ 가  $z_1$ 과  $z_2$ 에 대해 선형이 아닌 경우 CERES그림이 추가변수의 비선형함수의 형태를 잘 나타내 주는지를 알아보려고 한다. 다음을 실제 선형결합이라 하자.

$$\eta = -3 + 2x + e^{-(.5z_1 + .9z_2)}. \tag{4.1}$$

설명변수  $x$ 는  $U(0, 2)$ 에서  $n (= 50)$ 개의 값을 생성시키고 추가변수는  $z_1 = 1/x_1 + \delta$ ,  $z_2 = .5x_1 + \delta$ 가 되도록 생성하여  $E(x|z_1, z_2)$ 가  $z_1$ 과  $z_2$ 에 대해 비선형관계가 되도록 하였다.  $\delta$ 는  $N(0, 1)$ 에서 생성한다. 이와같이 결정한  $x, z_1, z_2$ 를 이용하여 모형 (4.1)에 적용시켜 50개의  $\eta$ 를 계산하였다.

이때  $\eta = \log[\pi(x)/(1 - \pi(x))]$ 라고 하면  $\pi(x) = \exp(\eta)/(1 + \exp(\eta))$ 로 50개의 비율형태의 값이 된다.  $\pi(x)$ 가  $m (= 50)$ 번의 시행에서 특정사건이 발생할 확률이라고 생각하고  $y$ 는  $m$ 번의 시행에서 특정사건이 발생한 도수로 보고  $B(m, \pi(x))$ 인 이항분포에 의해 생성하였다. 따라서 반응변수  $y$ 는 이항회귀모형의 반응변수로 볼 수 있다. 식 (4.1)의 선형결합  $\eta$ 는

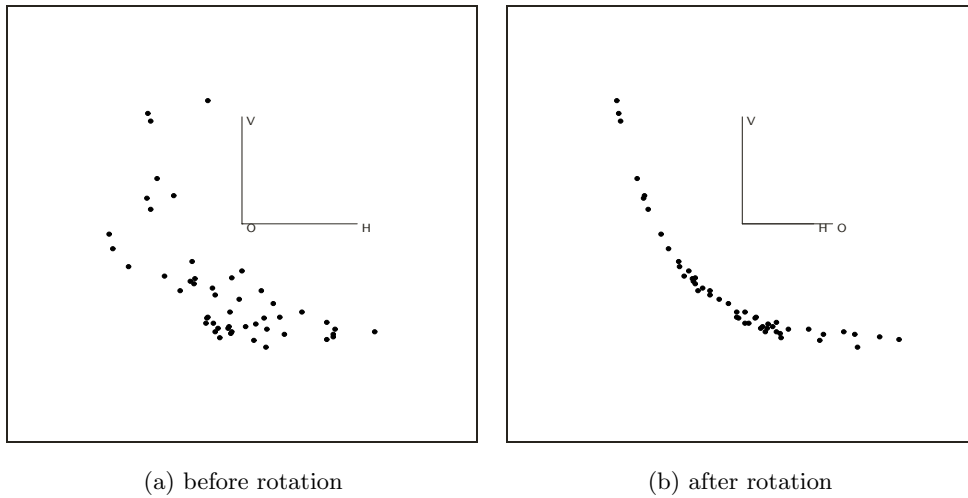


그림 4.1: 3차원 CERES그림 (V: CERES, H:  $z_1$ , O:  $z_2$ )

$-3 + 2x$ 에 추가변수  $z_1$ 과  $z_2$ 가  $h(z_1, z_2) = e^{-(.5z_1 + .9z_2)} = e^{-z^*} = h(z^*)$ 의 형태로 추가된다고 볼 수 있다.

추가변수가 어떠한 형태로 변환되어 모형에 추가되어야 하는지 알아보기 위하여 편잔차그림을 적용해 볼 수 있는데  $E(\mathbf{x}|z_1, z_2)$ 가  $z_1$ 과  $z_2$ 에 대해 선형이 아니므로 편잔차그림이 추가변수  $z_1$ 과  $z_2$ 의 함수 형태를 올바르게 나타낼 것으로 기대할 수 없다.

3차원 CERES그림을 그려보면 그림 4.1의 (a)와 같고 이 그림에서 수직축을 중심으로 하는 회전에서 보여지는 연속된 2차원 산점도 중 가장 명확한 곡선이 보이는 산점도가 그림 4.1의 (b)이다. 이것은 CERES축을 세로축으로 하고  $z_1$ 를 가로축으로 하는 그림 4.1의 (a)에서  $60^\circ$  회전된 상태이다. 이 회전 각도를 이용하면 이 산점도의 가로축은  $z(60^\circ) \cong .5z_1 + .87z_2$ 가 된다. 이 그림에서 나타나는 비선형곡선은 지수함수형태로 나타남을 알 수 있다. 따라서  $E(\mathbf{x}|z_1, z_2)$ 가  $z_1$ 과  $z_2$ 에 대해 선형이 아닌 비선형함수일 경우에도 CERES그림이 변환의 형태를 잘 나타내 주는 것을 확인할 수 있다.

## 참고문헌

- Atkinson, A. C. (1985). *Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Oxford University Press, Oxford.
- Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*, John Wiley & Sons, New York.
- Cook, R. D. (1993). Exploring partial residual plots, *Technometrics*, **35**, 351-362.
- Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman & Hall/CRC, New York.

- Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*, John Wiley & Sons, New York.
- Ezekiel, M. (1924). A method of handling curvilinear correlation for any number of variables, *Journal of the American Statistical Association*, **19**, 431–453.
- Kahng, M. and Jeong, H. (2000). CERES plot in nonlinear regression, *The Korean Communications in Statistics*, **7**, 1–11.
- Kahng, M. and Lee, E. (2004). CERES plot in generalized linear models, *The Korean Communications in Statistics*, **11**, 575–582.
- Mallows, C. L. (1986). Augmented partial residuals, *Technometrics*, **28**, 313–319.
- Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models, *Journal of the Royal Statistical Society, Ser. A*, **135**, 370–384.

[ 2007년 5월 접수, 2007년 10월 채택 ]

## Three Dimensional CERES Plot in Generalized Linear Models\*

Myung Wook Kahng<sup>1)</sup> Bu-Yong Kim<sup>2)</sup> Jin-Young Jeon<sup>3)</sup>

### ABSTRACT

We explore the structure and usefulness of three dimensional CERES plot as a basic tool for dealing with curvature as a function of the new predictors in generalized linear models. If predictors have nonlinear effects and there are nonlinear relationships among the predictors, the partial residual plot is not able to display the correct functional form of the predictors. Unlike this plots, the CERES plot can show the correct form. This is illustrated by simulated data.

*Keywords:* Generalized linear model, partial residual plot, CERES plot, LOESS.

---

\* This Research was supported by the Sookmyung Women's University Research Grants 2006.

1) Corresponding author. Professor, Dept. of Statistics, Sookmyung Women's University, Chungpa-dong 2-ga, Yongsan-gu, Seoul 140-742, Korea.

E-mail: mwkahng@sm.ac.kr

2) Professor, Dept. of Statistics, Sookmyung Women's University, Chungpa-dong 2-ga, Yongsan-gu, Seoul 140-742, Korea.

E-mail: buykim@sm.ac.kr

3) Graduate Student, Dept. of Statistics, Sookmyung Women's University, Chungpa-dong 2-ga, Yongsan-gu, Seoul 140-742, Korea.

E-mail: syoory@nate.com