Thesis for the Degree of Master of Science

Firework Plot as a Graphical Exploratory Data Analysis Tool for Evaluating the Impact of Outliers in Mixture Experiments

by

So Jin Ahn

Department of Statistics

The Graduate School

Pukyong National University

August 2014

Firework Plot as a Graphical Exploratory Data Analysis Tool for Evaluating the Impact of Outliers in Mixture Experiments
(혼합물 실험에서 특이점의 영향을 평가하기 위한 그래픽 탐색적 자료 분석 도구로서의 불꽃 그림)

Advisor: Prof. Dae-Heung Jang

by
So Jin Ahn

A thesis submitted in partial fulfillment of the requirements
for the degree of

Master of Science

in Department of Statistics, The Graduate School,
Pukyong National University

August 2014

Firework Plot as a Graphical Exploratory Data Analysis Tool for Evaluating the Impact of Outliers in Mixture Experiments

A dissertation

by

So Jin Ahn

Approved by:

(Seongbaek Yi)

(MaengSeok Noh)

(Dae-Heung Jang)

August 22, 2014

Contents

List of Tables

List of Figures

혼합물 실험에서 특이점의 영향을 평가하기 위한

그래픽 탐색적 자료 분석 도구로서의 불꽃 그림

안 소 진

부 경 대 학 교 대 학 원 통 계 학 과

요 약

일반적으로 회귀진단(잔차분석 및 영향진단)은 회귀 분석에서 특이점 또는 영향력 관찰점을 검출하기 위한 모델의 타당성 검사 도구로 실행된다. 특이점은 회귀 분석에서 많은 측정치를 왜곡 할 수 있으며, Jang 과 Anderson-Cook (2013)은 회귀분석에서 특이점과 영향력 관찰점의 영향을 평가하기 위한 간단한 도구로 불꽃 그림을 제안했다.

3-D 불꽃 그림과 쌍-불꽃 그림행렬은 자료분석에서 개별 관찰 가중치를 1에서 0으로 변화시켜 수치적 측도들에 상대적으로 미치는 영향을 곡선으로 나타낸다. 가중치 1은 모든 관측치의 기여도가 동일하게 대응하는 것을 의미하는 반면, 가중치 0은 특정한 관측치가 제거되었음을 의미한다. 3–D 불꽃 그림과 쌍-불꽃 그림행렬에서 보이는 변화량은 추정된 회귀계수와 잔차제곱합에 미치는 영향을 요약하므로 회귀 진단을 실행하기 위한 그래픽 탐색적 자료 분석 도구로서 3-D 불꽃그림과 쌍-불꽃 그림행렬을 혼합물 실험에서 사용할 수 있다.

1. INTRODUCTION

In mixture experiments, the measured response is assumed to depend only on the relative proportions of the components present in the mixture. For mixture experiments, if we let x_i represent the proportion of the $i-$th component in the mixture where the number of components is q, then

$$0 \le x_i \le 1,\ \ i = 1,2, \dots ,q \ \ , \tag{1.1}$$

and

$$\sum_{i=1}^{q} x_i = 1 .$$

The canonical form of the quadratic mixture model is

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum \sum_{i<j=2}^{q} \beta_{ij} x_i x_j \quad ,$$

and the canonical form of the special cubic mixture model is

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum \sum_{i<j=2}^{q} \beta_{ij} x_i x_j + \sum \sum_{i<j<k=3}^{q} \sum \beta_{ijk} x_i x_j x_k .$$

Then mixture experiment model is represented by

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \tag{1.2}$$

where $\underline{y}$ is the $n \times 1$ vector of responses, X is the $n \times p\ (p < n)$ model matrix, $\underline{\beta}$ is the $p \times 1$ vector of unknown coefficients, and $\underline{\varepsilon}$ is the $n \times 1$ vector of random errors. With solving the normal equations, the least-squares estimator of $\underline{\beta}$ is $\underline{b} = (X'X)^{-1}X'\underline{y}$.

Outliers can distort many quantitative descriptive measures used for data analysis. Several numerical measures for the regression diagnostics (for examples, residuals, standardized residuals, externally studentized residuals, Cook' s distance (Cook, 1977 & 1979), DFFITS and DFBETAS (Belsley et al., 1980) are common in regression analysis. We can also use a variety of graphical summaries for regression diagnostics (including leverage plot, Cook's distance plot, spread-level plot (Emerson and Strenio, 1983), added variable plot (Velleman and Welsch, 1981), and regression influence plot (Fox, 2008).

As a simple graphical method for quantifying the impact of outliers, Jang and Anderson-Cook (2013) proposed the firework plot, which gives an intuitive visual summary of the impact of outliers on different descriptive summaries of the data. By examining a continuum of weights for different data points in a data set, we are able to gain understanding about how their contributions impact the global summaries. In regression, understanding the impact of high influence points on the estimation of model parameters can aid with interpretation and help reveal the robustness of estimates to individual observations. These plots provide a class of graphical methods for evaluating the effect of outliers in univariate and bivariate data, as well as in regression settings. In this paper, we extend these ideas to the mixture experiments model setting.

Outliers and influential points can distort the quantitative measures and parameter estimates in mixture experiments analysis. Depending on the form of the mixture experiments model, the impact of observations on the parameter estimates may not be entirely transparent. Usually we use the regression diagnostics tools as a method for visualizing the impact of these points on the model and to detect violations of the model assumptions. To complement these tools, we propose the 3−D firework plot and pairwise firework plot matrix to give an intuitive visual summary of the impact of outliers on the estimated parameters and the residual sum of squares, since mixture experiments model estimates are known to be sensitive to outliers.

The paper is organized as follows. In Chapter 2, review of literature about outlier and influential observation, firework plot. And in Chapter 3, the 3−D firework plot and the pairwise firework plot matrix for mixture experiments are illustrated with a mixture experiment example to give an intuitive visual summary of the impact of outliers on the estimated parameters and the residual sum of squares. And the 3−D firework plot and the pairwise firework plot matrix for mixture experiments under constraints on the component proportions are illustrated with a mixture experiment example to give an intuitive visual summary of the impact of outliers on the estimated parameters and the residual sum of squares. Conclusions and discussion are given in Chapter 4.

2. REVIEW OF LITERATURE

2.1 Regression Diagnostics

Several numerical measures for the regression diagnostics are common in regression analysis. There are some literatures about outlier and influential observations.

Outliers are measuerd by residuals, internally studentized residuals, and externally studentized residuals. Influential observations are measured by hat matrix, cook's distance (Cook, 1977&1979), COVRATIO, AP Statistic, DFFITS (Belsley et al., 1980), and DFBETAS (Belsley et al., 1980).

Table 2.1. Regression Diagnostics

Outliers	Influential observations
Rasiduals	Hat Matrix Cook's Distance
Internally Studentized Residuals	COVRATIO AP Statistic
Externally Studentized Residuals	DFFITS DFBETAS

Now we will review some regression diagnostics measures.

2.2 Measures of Outliers

In regression, we define an outlier as an observation which is large or small compared to other observations in the data set. The observation as an outlier is in determining whether remove or include is significant in regression.

2.2.1 Residuals Residual defined as the difference between y_i and $\hat{y}_i$, y_i is the $i-$th value of the variable to be predicted, and $\hat{y}_i$ is its predicted value.

The sum of squares of residuals, SSE, is a measure of the difference between the data and an estimated model.

A small SSE indicates a tight fit of the model to the data.

SSE is given by

$$\mathrm{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i{}^2 = \underline{e}'\underline{e}$$

where $\underline{e} = (e_1, \dots, e_n)^t$.

2.2.2 Internally Studentized Residuals (Standardized Residuals)

When the residuals $e_i$have substantially different variances $\mathrm{Var}(e_i)$, as given in $\mathrm{Var}(e_i) = \sigma^2(1-h_{ii})$, it is well to consider the magnitude of e_i relative to $\sigma(e_i)$ to give recognition to differences in their sampling errors. We know that an unbiased estimator of this variance is

$$s^2(e_i) = MSE(1-h_{ii}),$$

where MSE is mean squared error, i.e. SSE/n−k.

The ratio of e_i to $s(e_i)$ is called the internally studentized residual and will be denoted by e_i^*, i.e. ,

$$e_i^* = \frac{e_i}{s(e_i)} = \frac{e_i}{\sqrt{MSE(1-h_{ii})}}.$$

A large absolute value of e_i^* indicates that the observation is an outlier.

2.2.3 Externally Studentized Residuals When the fitted regression, there is based on the cases excluding the $i-$th one. The method is related to measure the $i-$th residual $e_i = y_i - \hat{y}_i$.

We need to delete the $i-$th case, fit the regression function to the remaining $n \times 1$ cases, and compare the point estimate of the expected value when the X levels are those of the $i-$th case, denoted by $\hat{Y}_{i(i)}$, with the actual observed value Y_i. The residual is called a deleted residual and is denoted by d_i, i.e. ,

$$d_i = Y_i - \hat{Y}_{i(i)}.$$

An algebraically equivalent expression for d_i which does not require a recomputation of the fitted regression function omitting the $i-$th case is

$$d_i = \frac{e_i}{1 - h_{ii}} \ ,$$

where e_i is the ordinary residual including the $i-$th case and h_{ii} is the leverage value for this case.

Then, we know that the estimated variance of d_i is

$$s^2(d_i) = \frac{MSE_{(i)}}{1 - h_{ii}},$$

where MSE_i is the MSE without the $i-$th case.

And it can be shown that

$$\frac{d_i}{s(d_i)} \sim t(n-p-1) \ .$$

The studentized deleted residual, denoted by d_i^*, is defined by

$$d_i^* = \frac{d_i}{s(d_i)} \ .$$

It follows that an algebraically equivalent expression for d_i^* is

$$d_i^* = \frac{e_i^*}{\sqrt{MSE_{(i)}(1 - h_{ii})}} \ .$$

The d_i, hence, follows the $\mathrm{t}-$distribution with $\mathrm{n-p-1}$ degree of freedom. Fortunately, the studentized deleted residuals d_i^* can be calculated without having to fit new regression functions each time a different case is omitted.

A simple relationship between MSE and MSE_i is

$$(\mathrm{n-p})\mathrm{MSE} = (\mathrm{n-p-1})MSE_{(i)} + \frac{{e_i}^2}{1 - h_{ii}} \ .$$

Using this relationship in $d_i^* = \frac{e_i}{\sqrt{MSE_{(i)}(1-h_{ii})}}$ yields the following equivalent expression for d_i^*,

$$d_i^* = e_i\left[\frac{n-p-1}{SSE(1-h_{ii})-e_i^2}\right]^{\frac{1}{2}}.$$

Thus, the studentized residuals d_i^* can be calculated from the residuals e_i, the error sum of squares SSE, and the leverage valuse h_{ii}, which are all for the fitted regression with all the n cases.

2.3 Measures of Influential Observations

Influential observations are those observations influence the fitted regression equation as compared to other observations in the data set. As a measure to identify the influential observations, the difference between $\underline{b}_{(i)}$ and $\underline{b}$ is used.

Here $\underline{b}_{(i)}$ is the vector of the estimated regression coefficients obtained without the i −th case and $\underline{b}$ is one with all n cases included.

2.3.1 Hat Matrix The i−th diagonal element of $\mathrm{H} = \mathrm{X}(\mathrm{X}'X)^{-1}X'$, which is called hat matrix, is defined as

$$h_{ii} = \underline{x}_i'(X'X)^{-1}\underline{x}_i$$

where $\underline{x}_i'$ is the i−th row of X, $i = 1,2,\ldots,n$.

The h_{ii} are called leverage values and the observations with large values of h_{ii} are called high−leverage points. Hoaglin and Welsch (1978) suggested that the observations with $h_{ii} > 2k/n$ high−leverage points.

In particular, h_{ii} has the following properties ;

$$0 \le h_{ii} \le 1\ ,\ \ \sum_{i=1}^{n} h_{ii} = p\ ,$$

where p is the number of regression parameters in the regression function.

2.3.2 Cook's Distance Cook's distance measure D_i has the structure for measuring the combined impact of the differences in the estimated regression coefficients without the i−th case and defined by

$$D_i = \frac{(\underline{b} - \underline{b}_{(i)})' X' X (\underline{b} - \underline{b}_{(i)})}{pMSE}\ .$$

An algebraically equivalent expression is

$$D_i = \frac{{e_i}^2}{pMSE} \cdot \frac{h_{ii}}{(1 - h_{ii})^2}\ .$$

D_i is a function of e_i and h_{ii} and is proportional to them. Therefore we know that the i−th case can be influential observations if the percentile value is large. Also where the difference between $\underline{b}_i$ and $\underline{b}$ is large, it implies that the i−th case has a substantial influence on the regression fitting.

2.3.3 Measure of the Influence on the Fitted Values : DFFITS A useful measure of the influence that case i has on the fitted value $\hat{Y}_i$, is given by

$$
\begin{aligned}
(DFFITS)_i &= \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{MSE_{(i)} h_{ii}}} \\
&= e_i \left[\frac{n-p-1}{SSE(1-h_{ii}) - e_i{}^2}\right]^{\frac{1}{2}} \left(\frac{h_{ii}}{1-h_{ii}}\right)^{\frac{1}{2}} \\
&= d_i{}^* \left(\frac{h_{ii}}{1-h_{ii})}\right)^{1/2}.
\end{aligned}
$$

If case i is an outlier with high leverage value, $\left(h_{ii}/(1-h_{ii})\right)^{1/2}$ will be greater than 1 and $(DFFITS)_i$ will tend to be large absolutely. As a guideline for identifying influential cases, Belsley et al. (1980) suggest considering a case influential if the absolute value of DFFITS exceeds 1 for small to medium−size data sets and $2\sqrt{p/n}$ for large data sets.

2.3.4 Measure of Influence on the Regression Coefficients : DFBETAS Another measure of the influence of the i−th case on each regression coefficient $b_k (k = 0,1,\ldots,p-1)$ is the difference between the estimated regression coefficient b_k based on all n cases and the regression coefficient obtained by an appropriate standardization. The measure is termed as DFBETAS, and is defined by

$$
(DFBETAS)_{k(i)} = \frac{b_k - b_{k(i)}}{\sqrt{MSE_{(i)} c_{kk}}}, \quad k = 0,1,\ldots,p-1,
$$

where c_{kk} is the k−th diagonal element of $(X'X)^{-1}$.

A large absolute value of $(DFBETAS)_{k(i)}$ is indicative of a large impact of the i−th case on the k−th regression coefficient. As a guideline for identifying influential cases Belsley, et al. (1980) suggest considering a case influential if the absolute value of DFBETAS exceeds 1 for small to medium size data sets and $2\sqrt{n}$ for large data sets.

2.4 Firework Plot

Jang and Anderson-Cook(2013) proposed a new set of graphical summaries, called firework plots, as simple tools for evaluating the impact of outliers in data exploration and regression assessment. Outliers can distort many measures for data analysis. Firework plot provides to easy and simple visual summary for the outliers. There were two examples for explaining firework plot, one is in data exploration and the other is in regression assessment by Jang and Anderson-Cook(2013).

First, to illustrate how information can be gleaned from the plot, consider the data set of 2012 MLB team values in data exploration.

Table 2.2. Data Set of 2012 MLB Team Values

Rank	Team	Current Value ($mil)	One Year Value Change (%)	Debt/Value (%)	Revenue ($mil)	Operation Income ($mil)
1	New York Yankees	1850	9	2	439	10
2	Los Angeles Dodgers	1400	75	41	230	1.2
3	Boston Red Sox	1000	10	24	310	25.4
4	Chicago Cubs	879	14	69	266	28.1
5	Philadelphia phillies	723	19	24	249	−11.6
6	New York Mets	719	−4	69	225	−40.8
7	Texas Rangers	674	20	55	233	15.3
8	Los Angeles Angels of Anaheim	656	18	3	226	−1.2
9	San francisco Giants	643	14	16	230	8.8
10	Chicago White Sox	600	14	7	214	10.7
11	St. Lous Cardinals	591	14	47	233	25
12	Seattle Mariners	585	30	0	210	2.2
13	Houston Astros	549	16	41	196	24.3
14	Minnesota Twins	510	4	20	213	16.6
15	Atlanta Braves	508	5	0	203	20.7
16	Washington Nationals	480	15	52	200	25.9
17	Detroit Tigers	478	24	39	217	8.2
18	Colorado Rockies	464	12	15	193	14.4
19	Baltimore Orioles	460	12	33	179	12.9
20	San Diego Padres	458	13	44	163	23.2
21	Miami Marlins	450	25	32	148	8.9
22	Milwaukee Brewers	448	19	27	195	19.2
23	Arizona Diamondbacks	447	13	39	186	27.2
24	Cincinnati Reds	424	13	10	185	17.1
25	Toronto Blue Jays	413	23	0	188	24.9
26	Cleveland Indians	410	16	27	178	30.1
27	Kansas City Royals	354	1	14	161	28.5
28	Pittsburgh Pirates	336	11	38	168	15.9
29	Tampa Bay Rays	323	−2	36	161	26.2
30	Oakland Athletics	321	5	28	160	14.6

From Figure 2.1, we see that overall mean for all observations is located where all the lines converge, which is 605.10. Here w_i means that the weighted value is i the values for the mean based on removing a single observation range between 562.17 and 614.90, depending on which observation is removed. Similarly the overall standard deviation is 318.44, and ranges between 222.75 and 340.08 when a single observation is removed. Removing large values lead to reductions in the overall mean, while removing extreme values lead to reductions in the standard deviation.

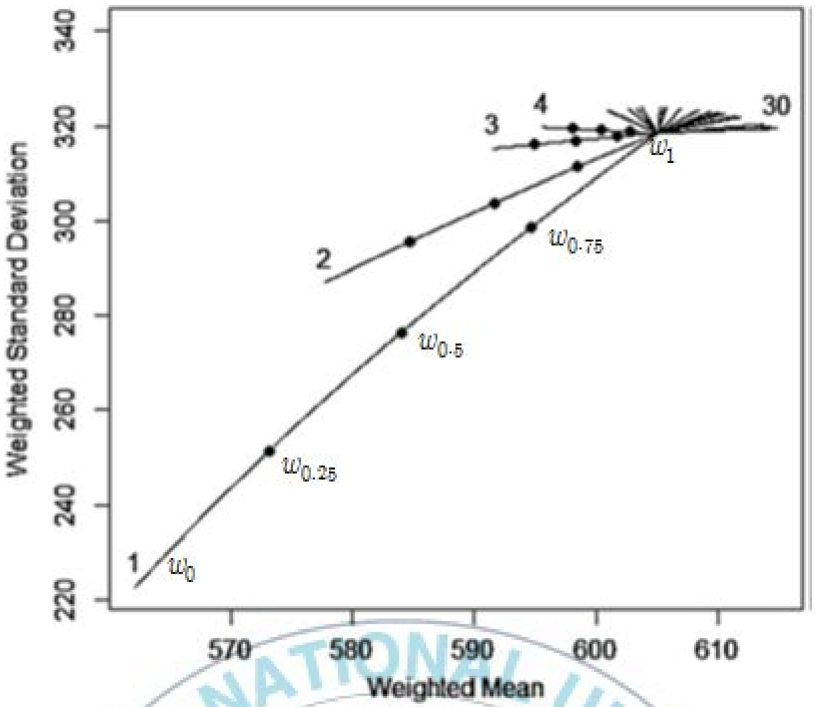


Figure 2.1. Mean-standard Deviation Firework Plot for 'Current Value'(The number is the MLB rank of each team)

The New York Yankees (labeled 1) has a substantial impact on both the mean and standard deviation and can be considered an outlier. This observation has the largest effect on the mean (605.10->562.17) and standard deviation (318.44->222.75).

From Figure 2.2, we see that New York Yankees (labeled 1) and Los Angeles Dodgers (labeled 2) are two outliers with each impact in different directions.

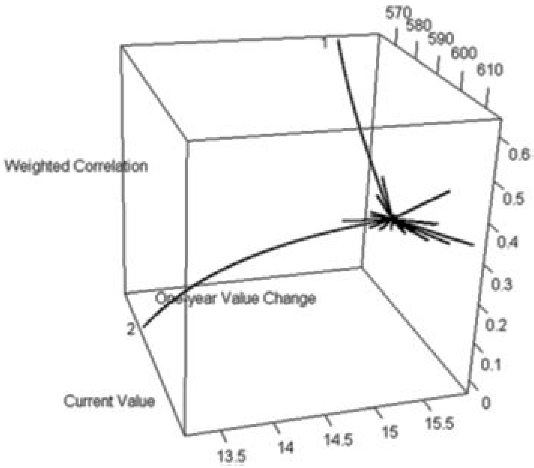


Figure 2.2. 3-D Mean-correlation Coefficient Firework Plot for 'Current Value' and 'One Year Value Change'.

Figure 2.3, also shows that New York Yankees (labeled 1) and Los Angeles Dodgers (labeled 2) are two outliers.

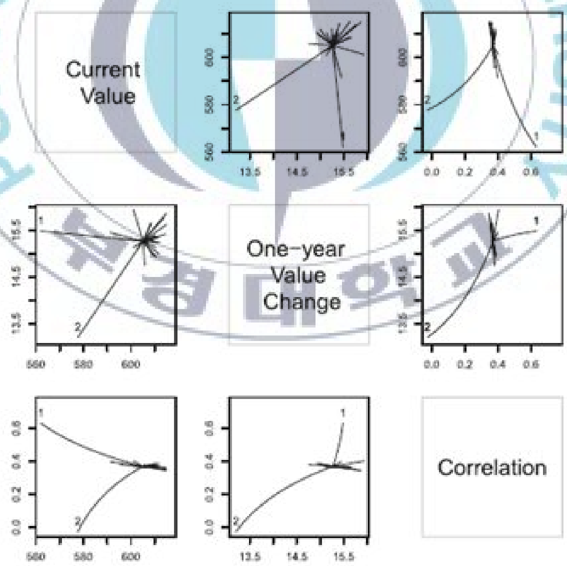


Figure 2.3. Pairwise Mean-correlation Firework Plot for 'Current Value' and 'One Year Value Change'.

We can obtain a synergy effect for evaluating the impact of outliers if we use this pairwise firework plot matrix with 3−D firework plot.

Next, consider scatter plot for Mickey data set in regression assessment to illustrate how information can be gleaned from the plot.

With Figure 2.4, note how observations 2 and 18 fall further away from most of the remaining observations, while observation 19 does not appear to follow the predominant relationship between the two variables.

In the 3−D firework plot (Figure 2.5) for the Mickey data set, the X−axis represents the weighted estimated intercept, the Y−axis the weighted estimated slope, and the Z−axis the weighted standard error.

By rotating the 3−D firework plot, we can find an orientation that highlights the outliers in the Mickey data set.

From Figure 2.5, we see that changing the weight of the 19^{th} observation produces a dramatic change in the weighted SSE, while changing the weight of the 18^{th} observation alters both the slope and intercept estimates substantially.

Hence, we find that the 19^{th} observation is an outlier and the 18^{th} observation is an influential observation.

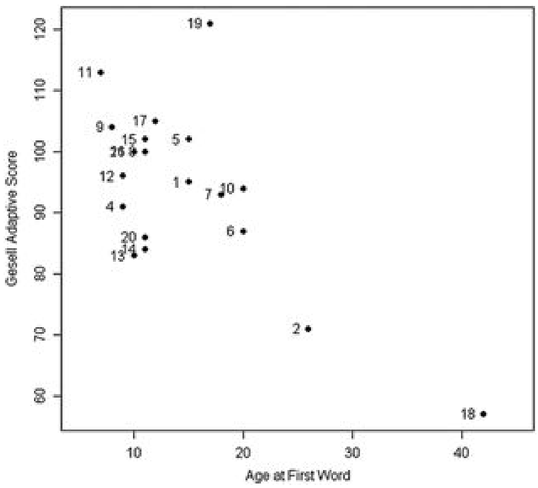


Figure 2.4. Scatterplot for Mickey Data Set.

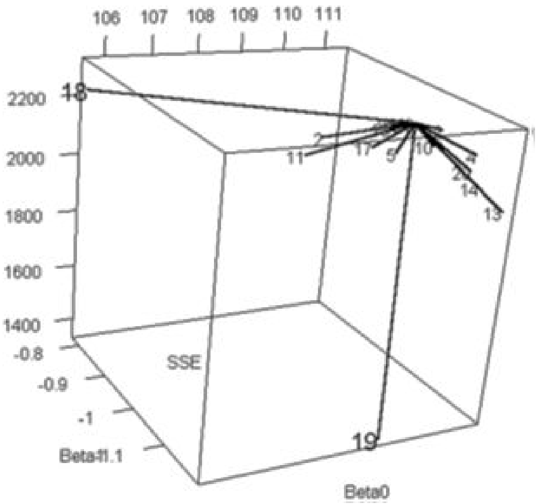


Figure 2.5. 3-D Firework Plot for Mickey Data Set.

Graphical tools can be helpful for evaluating the degree of sensitivity of individual observations in a data set. For exploratory data analysis, understanding the impact on the mean, standard deviation, and correlation coefficient can provide greater understanding of the relationship of the observations to each other. The firework plot provides useful information with a simple and easy to understand visual summary for evaluating the impact of outliers and influential sets of observation. Several variations of the plots provide flexibility about how the data can be examined and which characteristics are highlighted. The 2-D, 3-D, and pairwise matrix versions of the firework plots allow for different visualizations of key features of the data. In addition, the regression-based versions of the plots allow visualization of the impact on linear regression parameters based on individual observations. The plots complement other graphical summaries commonly used in regression analysis.

3. 3-D FIREWORK PLOT AND PAIRWISE FIREWORK PLOT MATRIX FOR EVALUATING THE EFFECT OF OUTLIERS

3.1 Firework Plot for Evaluating the Effect of Outliers in Mixture Experiments

Outliers can distort the quantitative measures and parameter estimates in mixture experiments analysis. Usually we use the regression diagnostics tools as methods for evaluating the impact of these points on the model and to detect violations of the model assumptions. To complement these tools, we propose the 3-D and pairwise firework plots to give an intuitive visual summary of the impact of outliers on the estimated regression coefficients and the residual sum of squares, where least squares estimators are known to be sensible to outliers. By examining a continuum of weights for different data points in a data set, we are able to see how their contributions impact the mixture experiments analysis summaries.

For a data set, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, we can use the weighted least-squares estimator and obtain the corresponding weighted residual sum of squares (SSE) as follows:

$$\underline{\mathrm{b}} = (\mathrm{X'WX})^{-1}\mathrm{X'W}\underline{\mathrm{y}} \qquad (3.1)$$

$$SSE = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2$$

where W is a diagonal matrix with weights $w_1, w_2, \ldots, w_n$ on its diagonal.

The 3-D firework plot uses curves to connect three values of the estimated regression coefficients and the residual sum of squares, SSE, calculated by changing the weights for each observation from 1 to 0, while all other observation weights are held fixed with a value of 1. With this 3-D firework plot, we can identify outliers and quantify their impact on the estimated model parameters in simple linear regression. For each (x_i, y_i), we calculate the weighted regression coefficients and the weighted SSE ineq.(3.1) with $w_j = 1\ (j \neq i\ ,\ j = 1,2, \ldots n)$ and the fractional deletion of x_i with sequential changing weights, w_i from 1 (equal contribution for all observations with no deletion) to 0 (complete deletion). Hence, we can obtain 3-D curves, for the range of the estimated quantities as the weights for each observation are adjusted. The 3-D firework plot shows the range of values for each characteristic of interest from down weighting or removing a single observation. By seeing the pattern of these curves, we can evaluate the relative impact of all of the observations in simple linear regression. We can consider the 3-D firework plot as a graphical representation of infinitesimal perturbation method among the in fluence measures in the regression diagnostics.

In the pairwise firework plot matrix, each panel uses curves to connect the values of the weighted least-squares estimators and weighted residual sum of squares, SSE, calculated by changing weights for each observation from 1 to 0. With this firework plot matrix, we can identify outliers and quantify their impact on the estimated parameters in multiple linear regression analysis.

To illustrate how information can be gleaned from the plot, consider the etch rate experiment data set (Myers et al., 2009, 571p). Wet chemical etching is often performed on the backs of silicon wafers prior to metallization in the semiconductor industry. The etching solution is a mixture of three different acids: A, B, and C. Table 3.1 shows augmented simplex−lattice design for the etch rate experiment. y is a response variable (etch rate in $\overset{0}{A}/\min$).

Table 3.1. Augmented Simplex−lattice Design for the Etch Rate Experiment

No	x_1 (Acid A)	x_2 (Acid B)	x_3 (Acid C)	Y
1	1	0	0	540
2	1	0	0	560
3	0	1	0	330
4	0	1	0	350
5	0	0	1	295
6	0	0	1	260
7	1/2	1/2	0	610
8	1/2	0	1/6	425
9	0	1/2	1/6	330
10	1/3	1/3	1/3	800
11	1/3	1/3	1/3	850
12	2/3	1/6	1/6	710
13	1/6	2/3	1/6	640
14	1/6	1/6	2/3	460

The fitted special cubic model using this data set is

$$\hat{y} = 550.200x_1 + 344.723x_2 + 268.295x_3 + 689.537x_1x_2$$
$$-9.035x_1x_3 + 58.108x_2x_3 + 9243.336x_1x_2x_3$$

Figure 3.1 shows the pairwise firework plot matrix for the etch rate experiment data set. In Figure 3.2, we can see the pairwise firework plot matrix corresponding to mixed quadratic and cubic terms and SSE for the etch rate experiment data set. And Figure 3(a) indicates 3−D firework plot corresponding to linear terms and Figure 3(b) indicates 3−D firework plot corresponding to mixed quadratic and cubic terms and SSE. From these Figures, we see that 7^{th}, 8^{th}, and 9^{th} observations are three influential observations with the relative largest change of the weighted estimated regression coefficients and that 14^{th} observation is an outlier with the relative largest change of the weighted SSE. It is interesting that 7^{th}, 8^{th}, and 9^{th} observations have the relative largest change of the weighted estimated regression coefficients corresponding to mixed quadratic and cubic terms β_{12}, β_{13}, β_{23}, β_{123}, , but that these observations have the small change of the weighted estimated regression coefficients corresponding to linear terms β_1, β_2, β_3.

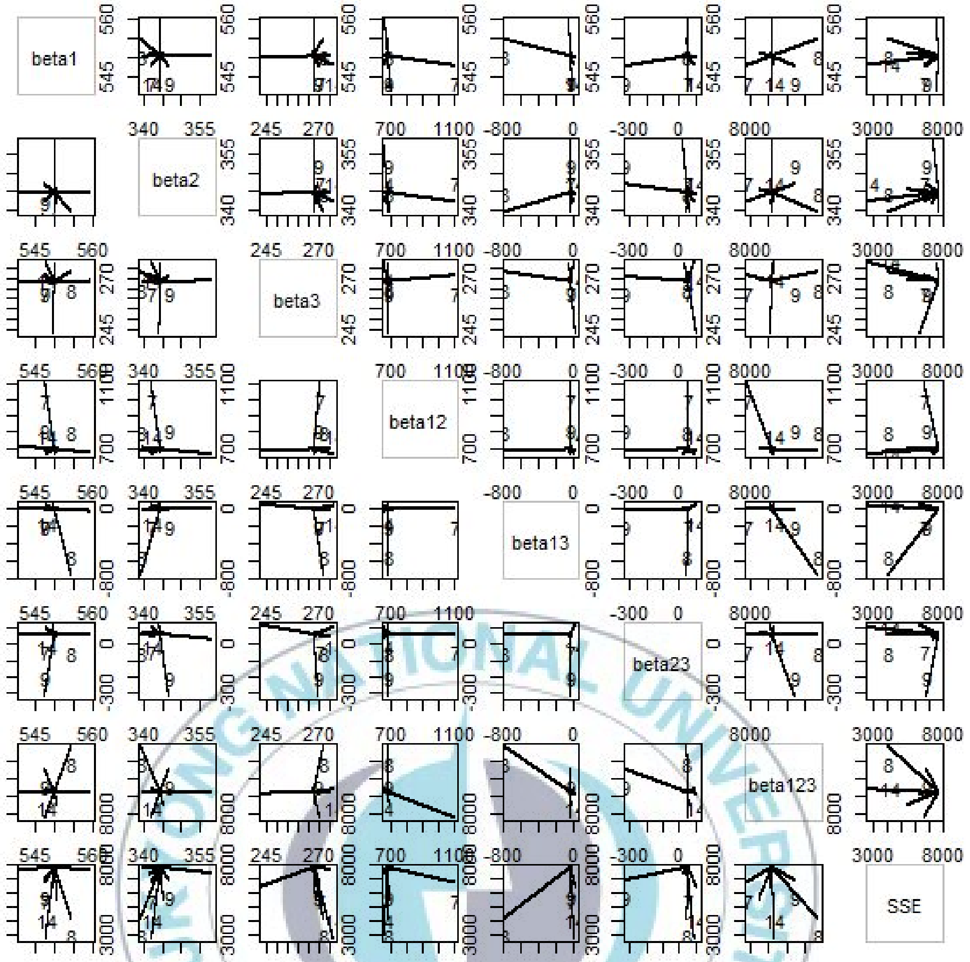


Figure 3.1. Pairwise Firework Plot Matrix for the Etch Rate Experiment Data Set.

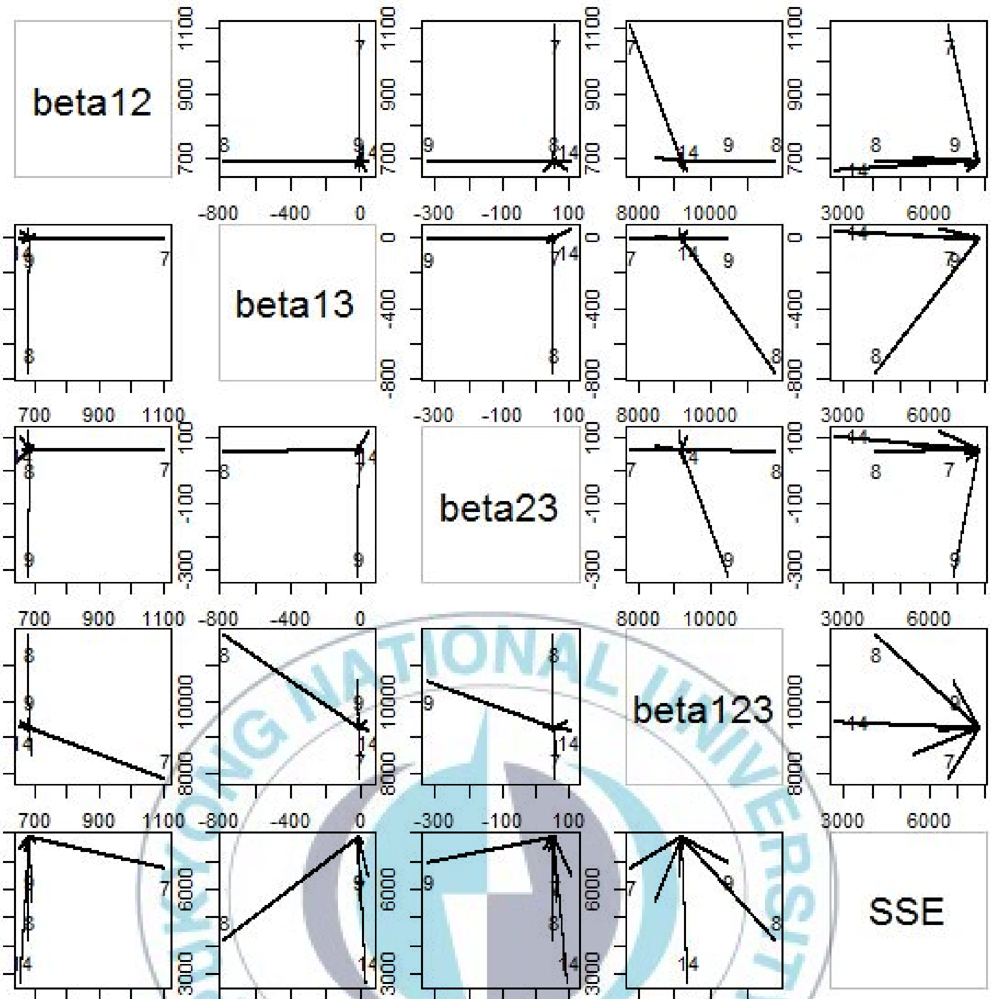
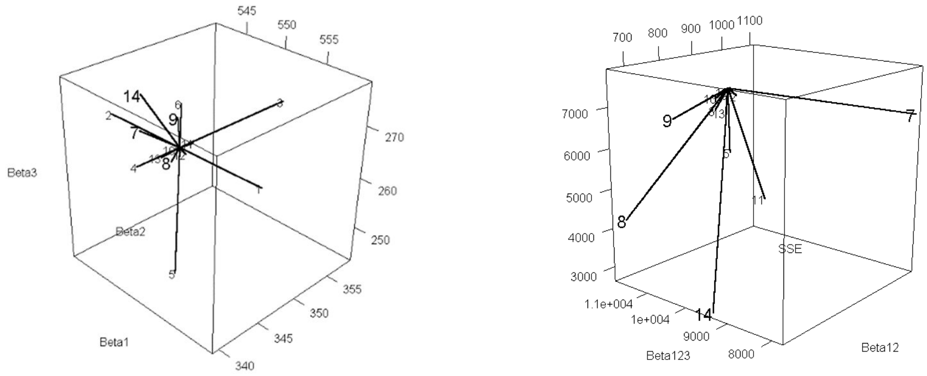


Figure 3.2.   Pairwise Firework Plot Matrix Corresponding to Mixed Quadratic and Cubic Terms and SSE for the Etch Rate Experiment Data Set.



(a) (b)

Figure 3.3. 3-D Firework Plot Corresponding to (a) Linear Terms (b) Mixed Quadratic and Cubic Terms and SSE for the Etch Rate Experiment Data Set.

For comparison with firework plots, Table 3.2 shows the regression diagnostics numerical results for the etch rate experiment. Using 3-D firework plot and pairwise firework plot matrix, we can obtain the same results which match up with the results from the regression diagnostics numerical results.

Table 3.2. The Regression Diagnostics Numerical Results for the Etch Rate Experiment

NO	Actual Value	Predicted Value	R−Student	Hat_ii	Cook_Distance
1	540.0	550.2	−0.40	0.483	0.024
2	560.0	550.2	0.38	0.483	0.022
3	330.0	344.7	−0.58	0.483	0.050
4	350.0	344.7	0.20	0.483	0.006
5	295.0	268.3	1.13	0.483	0.164
6	260.0	268.3	−0.32	0.483	0.016
7	610.0	619.8	−0.99	0.912	1.453
8	425.0	407.0	2.31	0.912	4.862
9	330.0	321.0	0.89	0.912	1.204
10	800.0	812.2	−0.43	0.375	0.018
11	850.0	812.2	1.58	0.375	0.176
12	710.0	717.4	−0.23	0.206	0.002
13	640.0	620.2	0.64	0.206	0.016
14	460.0	523.8	−3.38	0.206	0.170

3.2 Firework Plot for Evaluating the Effect of Outliers in Mixture Experiments under Constraints on the Component Proportions

In some mixture problems, constraints on the individual components arise as following:

$$L_i \leq x_i \leq U_i, i = 1,2,\dots,q. \tag{3.2}$$

3−D Firework plot and pairwise firework plot matrix can be used for evaluating the effect of outliers in mixture experiments under constraints on the component proportions.

To illustrate how information can be gleaned from the plot, consider the railroad flare experiment data set (McLean and Anderson, 1966; Myers et al., 2009, 604p) about the illumination level. This data set consists of a response variable and 4 components (y: illumination level(1000 candles), x_1: magnesium, x_2: sodium nitrate, x_3: strontium nitrate, x_4: binder) with 15 observations under the following constraints,

$$0.40 \le x_1 \le 0.60, 0.10 \le x_2 \le 0.50, 0.10 \le x_3 \le 0.50, 0.03 \le x_4 \le 0.08$$

Table 3.3 shows an extreme vertices design for the railroad flare experiment.

The fitted special cubic model using this data set is

$$\hat{y} = -734.307x_1 + 1029.821x_2 + 849.011x_3 - 9874.473x_4 + 19162.127x_1x_4$$
$$-26120.150x_2x_3 + 52817.778x_1x_2x_3 + 116303.493x_2x_3x_4$$

Table 3.3. Extreme Vertices Design for the Railroad Flare Experiment

No	x1	x2	x3	x4	Y
1	0.4	0.1000	0.4700	0.030	75
2	0.4	0.1000	0.4200	0.080	180
3	0.6	0.1000	0.2700	0.030	195
4	0.6	0.1000	0.2200	0.080	300
5	0.4	0.4700	0.1000	0.030	145
6	0.4	0.4200	0.1000	0.080	230
7	0.6	0.2700	0.1000	0.030	220
8	0.6	0.2200	0.1000	0.080	350
9	0.5	0.1000	0.3450	0.055	220
10	0.5	0.3450	0.1000	0.055	260
11	0.4	0.2725	0.2725	0.055	190
12	0.6	0.1725	0.1725	0.055	310
13	0.5	0.2350	0.2350	0.030	260
14	0.5	0.2100	0.2100	0.080	410
15	0.5	0.2225	0.2225	0.055	425

Figure 3.4 shows the pairwise firework plot matrix for the railroad flare experiment data set. Figure 3.5 shows the pairwise firework plot matrix corresponding to linear terms and SSE for the railroad flare experiment data set. Figure 3.6 shows the pairwise firework plot matrix correspond−ing to mixed quadratic and cubic terms and SSE for the railroad flare experiment data set. Figure 3.7 (a)−(b) shows 3−D firework plot corresp−onding to linear terms and SSE for the railroad flare experiment data set. Figure 3.7 (c)−(d) shows 3−D firework plot corresponding to mixed quad−ratic and cubic terms and SSE for the railroad flare experiment data set. From these Figures, we see that 11^{th} and 13^{th} observations are two influential observations with the relative largest change of the weighted estimated regression coefficients and that 15^{th} observation is an outlier with the relative largest change of the weighted SSE.

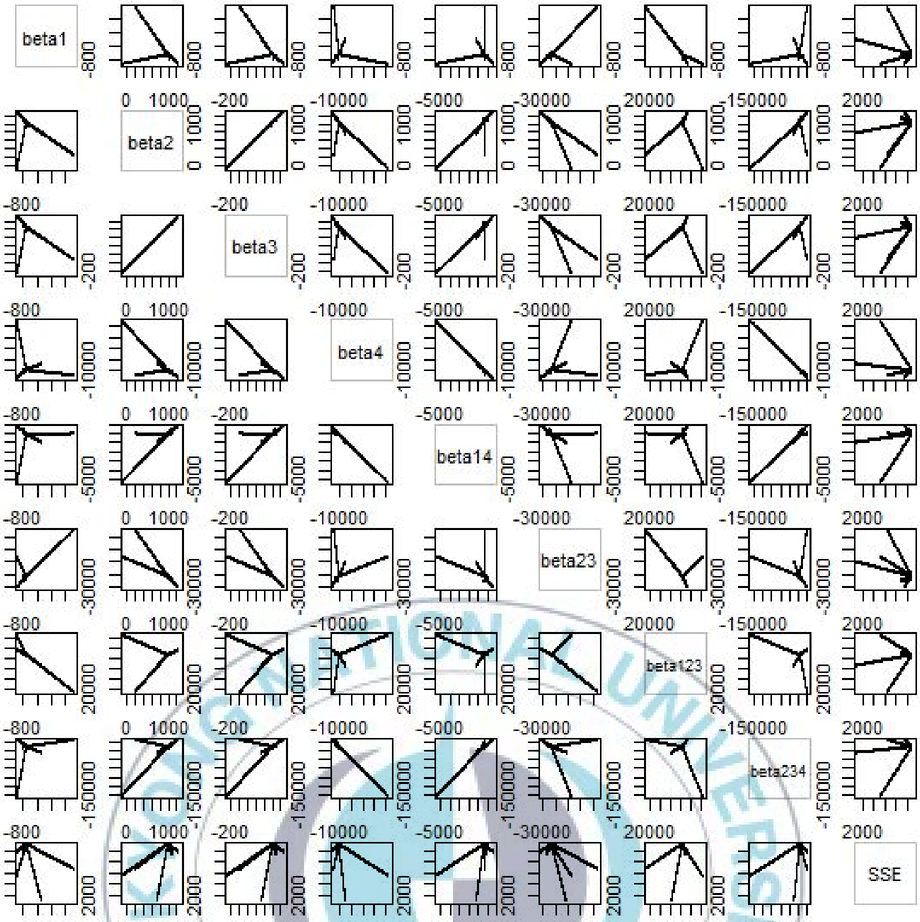


Figure 3.4. The Pairwise Firework Plot Matrix for the Railroad Flare Experiment Data Set.

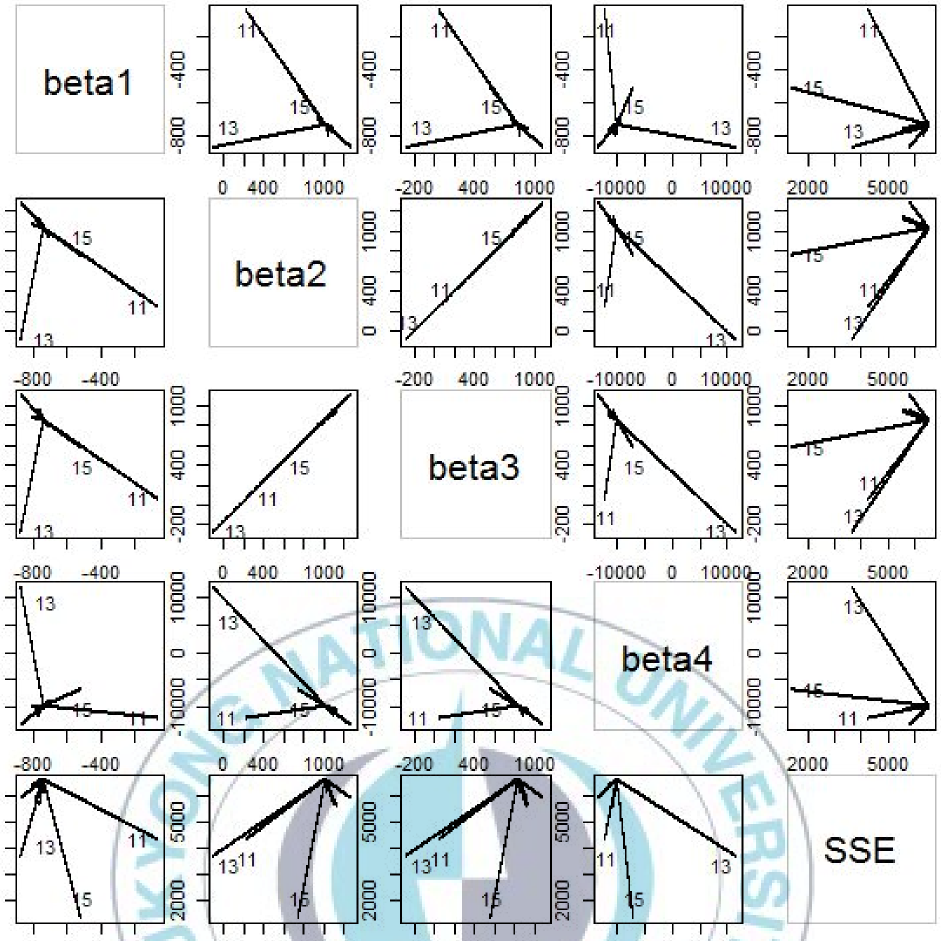


Figure 3.5. The Pairwise Firework Plot Matrix Corresponding to Linear Terms and SSE for the Railroad Flare Experiment Data Set.

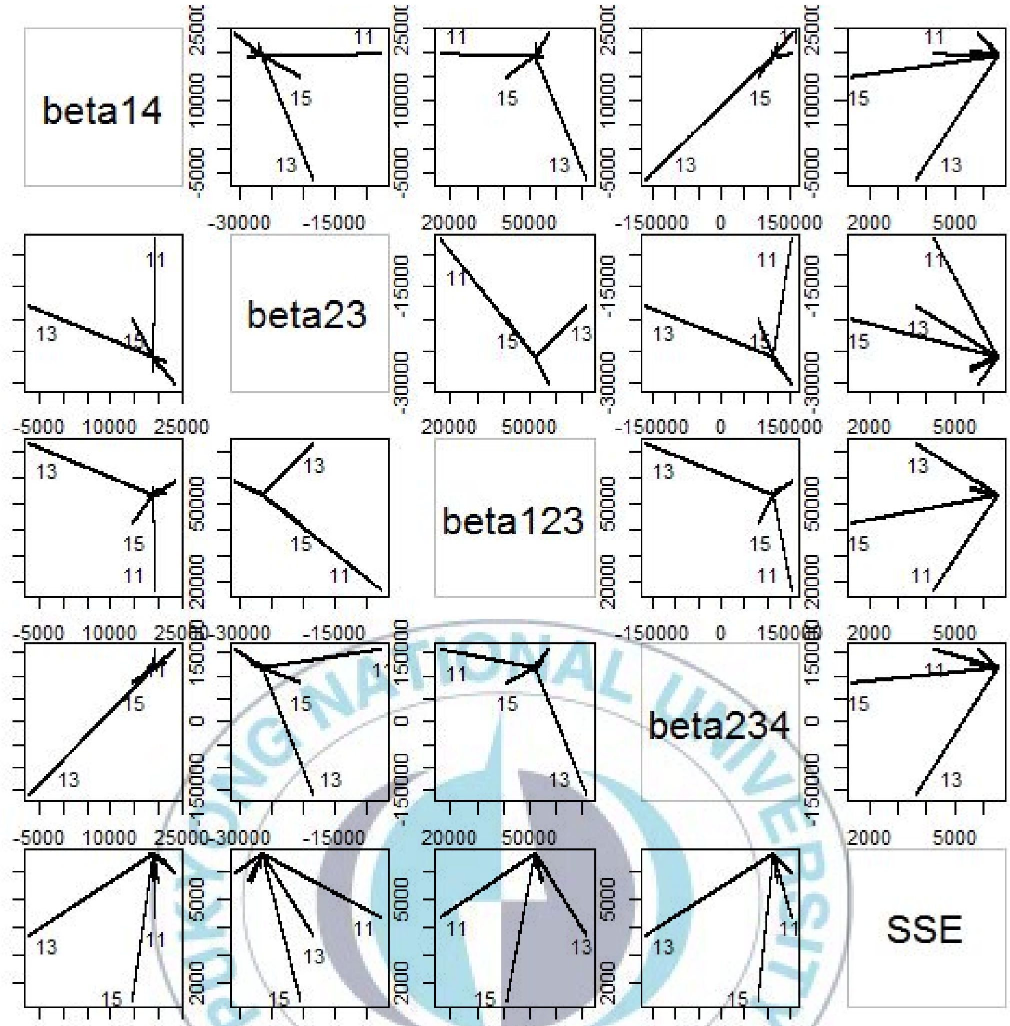


Figure 3.6. The Pairwise Firework Plot Matrix Corresponding to Mixed Quadratic and Cubic Terms and SSE for the Railroad Flare Experiment Data Set.

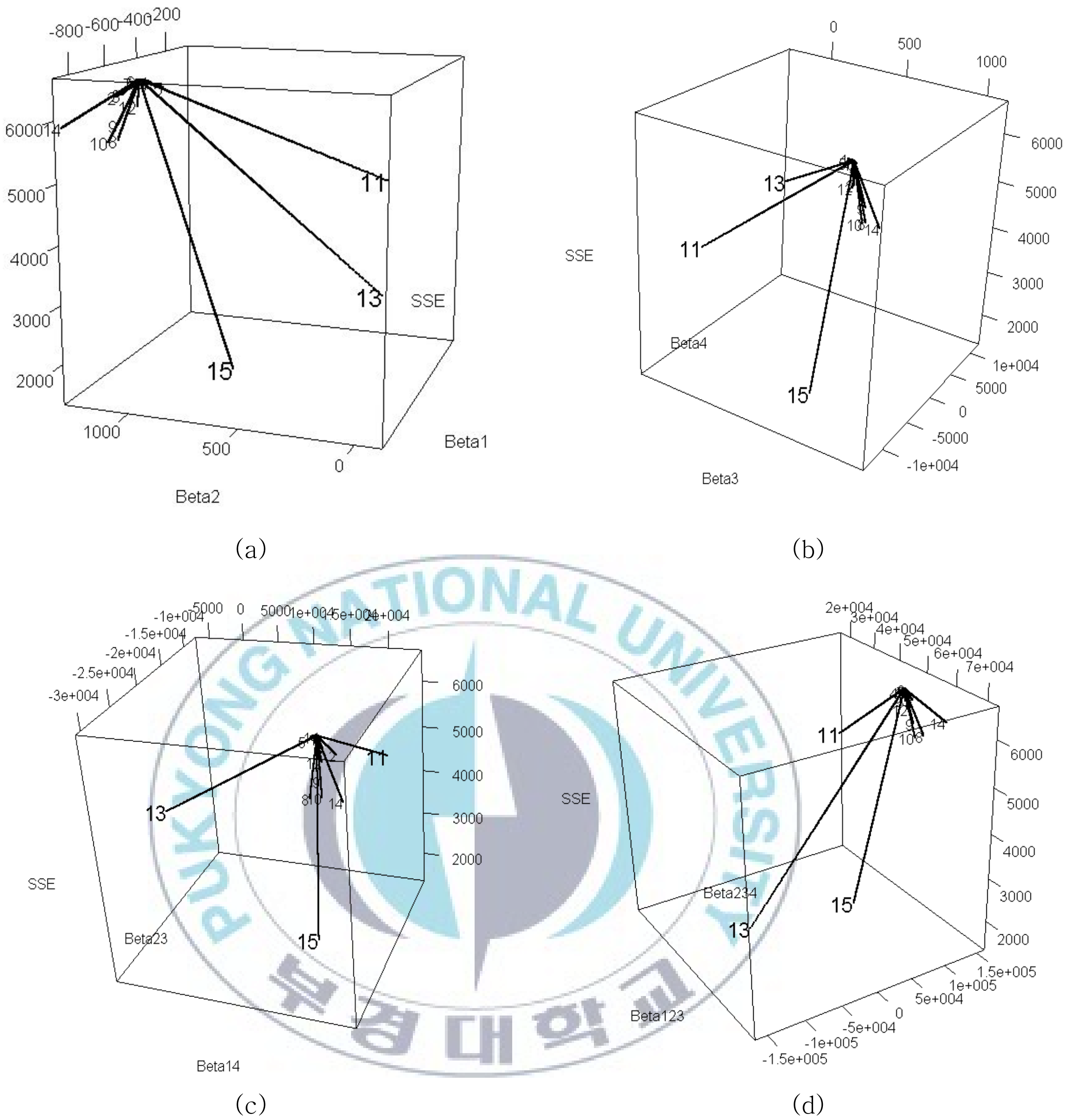


Figure 3.7. 3-D Firework Plot Corresponding to (a)-(b) Linear Terms (c)-(d) Mixed Quadratic and Cubic Terms and SSE for the Railroad Flare Experiment Data Set

For comparison with firework plots, Table 3.4 shows the regression diagnostics numerical results for the railroad flare experiment. Using 3−D firework plot and pairwise firework plot matrix, we can obtain the same results which match up with the results from the regression diagnostics numerical results.

Table 3.4. The Regression Diagnostics Numerical Results for the Railroad Flare Experiment

NO	Actual Value	Predicted Value	R−Student	Hat_ii	Cook_Distance
1	75	71.321	0.19275	0.666	0.01074
2	180	170.146	0.49825	0.629	0.05892
3	195	184.925	0.43434	0.494	0.02608
4	300	306.250	−0.26707	0.495	0.01008
5	145	138.221	0.35783	0.666	0.03646
6	230	228.005	0.09891	0.629	0.00241
7	220	215.663	0.18464	0.494	0.00483
8	350	327.947	1.01383	0.495	0.12549
9	220	234.247	−0.84445	0.227	0.02737
10	260	287.546	−1.02551	0.227	0.03842
11	190	200.960	−1.77669	0.947	5.39995
12	310	328.868	−0.65325	0.186	0.01327
13	260	274.201	−2.17475	0.931	5.16620
14	410	426.981	−0.84426	0.588	0.13242
15	425	365.720	4.75960	0.325	0.33319

4. CONCLUSION

Graphical tools can be helpful for evaluating the degree of sensitivity of individual observations in a data set. The firework plot provides useful information with a simple and easy to understand visual summary for evaluating the impact of outliers and influential sets of observation. Variations of the plots provide flexibility about how the data can be examined and which characteristics are highlighted. 3−D and pairwise matrix versions of the firework plots allow for different visualizations of key features of the data. The regression−based versions of the plots allow visualization of the impact on mixture experiments model parameters based on individual observations. 3−D firework plot and pairwise firework plot matrix can be used as a graphical exploratory data analysis tool for executing regression diagnostics in mixture experiments. These plots are closely related with several numerical measures for the regression diagnostics and complement other graphical summaries commonly used in mixture experiments model analysis.

References

Belsley, D. A., Kuh, E., Welch, R. E. (1980), *Regression Diagnostics: Identifying Influential Data and Source of Collinearity*, New York: Wiley.

Cook, R. D. (1977), "Detection of Influential Observation in Linear Regression," *Technometrics*, 19, 15−18.

Cook, R. D.(1979), "Influential Observation in Linear Regression," *Journal of the American Statistical Association*, 74, 169−174.

Emerson, J. D. and Strenio, J. (1983), The Spread−versus−Level plot in Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (Eds.) *Understanding Robust and Exploratory Data Analysis*, New York: Wiley.

Fox, J. (2008), *Applied Regression Analysis and Generalized Linear Models,* 2nd ed., New York: Sage.

Jang, D. H. and Anderson−Cook C. M. (2013), "Firework Plot as a Graphical Exploratory Data Analysis Tool for Evaluating the Impact of Outliers in Data Exploration and Regression," *Quality and Reliability Engineering International,* in press.

McLean, R. A. and Anderson, V. L. (1966), "Extreme Vertices Design of Mixture Experiments," *Technometrics*, 8, 447−454.

Myers, R. H., Montgomery, D. C., Anderson−Cook, C. M. (2009), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 3rd ed., New York: Wiley.

Velleman, P. F. and Welsch, R. E. (1981), "Efficient Computing of Regression Diagnostics," *The American Statistician*, 35, 234−242.

Appendix

```
#######   3.1 R source   #######

install.packages('rgl')
install.packages('scatterplot3d')
install.packages('leaps')
install.packages('car')
install.packages('MASS')

library(rgl)
library(scatterplot3d)
library(leaps)
library(car)
library(MASS)

#######   자료 읽기   #######

EtchRate=read.table("E:\\EtchRate.txt",header=T)
EtchRate
attach(EtchRate)

x1
x2

x12=EtchRate$x1*EtchRate$x2
x12

x1
x3

x13=EtchRate$x1*EtchRate$x3
x13

x2
x3

x23=EtchRate$x2*EtchRate$x3
x23

x1
x2
x3

x123=EtchRate$x1*EtchRate$x2*EtchRate$x3
x123
X=cbind(x1,x2,x3,x12,x13,x23,x123)
X
XX=t(X)%*%X
XX
혼합물모형=lm(EtchRate$y~x1+x2+x3+x12+x13
+x23+x123-1)
summary(혼합물모형)
anova(혼합물모형)

coef(혼합물모형)

n=length(EtchRate$y);m=2001
n

mat1=c(rep(0,n*m))
wbeta1=matrix(mat1,nrow=n)
wbeta2=matrix(mat1,nrow=n)
wbeta3=matrix(mat1,nrow=n)
wbeta12=matrix(mat1,nrow=n)
wbeta13=matrix(mat1,nrow=n)
wbeta23=matrix(mat1,nrow=n)
wbeta123=matrix(mat1,nrow=n)
wSSE=matrix(mat1,nrow=n)

######   Calculation of Weighted Least-Squares
Estimators   ######

k=1
for(i in 1:n)
{
for(j in seq(1,0,by=-0.0005))
{
w=c(rep(1,n))
w[i]=j
wlsm=lm(EtchRate$y~x1+x2+x3+x12+x13+x23
+x123-1,weights=w)
wbeta1[i,k]=coef(wlsm)[1]
wbeta2[i,k]=coef(wlsm)[2]
wbeta3[i,k]=coef(wlsm)[3]
wbeta12[i,k]=coef(wlsm)[4]
wbeta13[i,k]=coef(wlsm)[5]
wbeta23[i,k]=coef(wlsm)[6]
wbeta123[i,k]=coef(wlsm)[7]
```

```
wSSE[i,k]=sum(w*(EtchRate$y-wbeta1[i,k]*x1-wbeta2[i,k]*x2-wbeta3[i,k]*x3-wbeta12[i,k]*x12-wbeta13[i,k]*x13-wbeta23[i,k]*x23-wbeta123[i,k]*x123)^2)
k=k+1
}
k=1
}

###### Firework Matrix ######
par(mfrow=c(8,8))
par(mar=c(1,1,1,1))

plot.new()
box(col="grey")
text(.5,.5,"beta1",cex=1)

plot(wbeta2,wbeta1,xlab="",ylab="",cex=0.1)
text(wbeta2[7,m],wbeta1[7,m],7,pos=1,col="black")
text(wbeta2[8,m],wbeta1[8,m],8,pos=3,col="black")
text(wbeta2[9,m],wbeta1[9,m],9,pos=1,col="black")
text(wbeta2[14,m],wbeta1[14,m],14,pos=4,col="black")
par(new=F)

plot(wbeta3,wbeta1,xlab="",ylab="",cex=0.1)
text(wbeta3[7,m],wbeta1[7,m],7,pos=1,col="black")
text(wbeta3[8,m],wbeta1[8,m],8,pos=3,col="black")
text(wbeta3[9,m],wbeta1[9,m],9,pos=2,col="black")
text(wbeta3[14,m],wbeta1[14,m],14,pos=2,col="black")
par(new=F)
plot(wbeta12,wbeta1,xlab="",ylab="",cex=0.1)
text(wbeta12[7,m],wbeta1[7,m],7,pos=1,col="black")
text(wbeta12[8,m],wbeta1[8,m],8,pos=3,col="black")
text(wbeta12[9,m],wbeta1[9,m],9,pos=2,col="black")
text(wbeta12[14,m],wbeta1[14,m],14,pos=4,col="black")
par(new=F)
plot(wbeta13,wbeta1,xlab="",ylab="",cex=0.1)
text(wbeta13[7,m],wbeta1[7,m],7,pos=1,col="black")
text(wbeta13[8,m],wbeta1[8,m],8,pos=3,col="black")
text(wbeta13[9,m],wbeta1[9,m],9,pos=3,col="black")
text(wbeta13[14,m],wbeta1[14,m],14,pos=2,col="black")
par(new=F)

plot(wbeta23,wbeta1,xlab="",ylab="",cex=0.1)
text(wbeta23[7,m],wbeta1[7,m],7,pos=1,col="black")
text(wbeta23[8,m],wbeta1[8,m],8,pos=3,col="black")
text(wbeta23[9,m],wbeta1[9,m],9,pos=4,col="black")
text(wbeta23[14,m],wbeta1[14,m],14,pos=1,col="black")
par(new=F)

plot(wbeta123,wbeta1,xlab="",ylab="",cex=0.1)
text(wbeta123[7,m],wbeta1[7,m],7,pos=1,col="black")
text(wbeta123[8,m],wbeta1[8,m],8,pos=1,col="black")
text(wbeta123[9,m],wbeta1[9,m],9,pos=1,col="black")
text(wbeta123[14,m],wbeta1[14,m],14,pos=1,col="black")
par(new=F)

plot(wSSE,wbeta1,xlab="",ylab="",cex=0.1)
text(wSSE[7,m],wbeta1[7,m],7,pos=2,col="black")
text(wSSE[8,m],wbeta1[8,m],8,pos=3,col="black")
text(wSSE[9,m],wbeta1[9,m],9,pos=1,col="black")
text(wSSE[14,m],wbeta1[14,m],14,pos=4,col="black")
par(new=F)
plot(wbeta1,wbeta2,xlab="",ylab="",cex=0.1)
text(wbeta1[7,m],wbeta2[7,m],7,pos=3,col="black")
text(wbeta1[8,m],wbeta2[8,m],8,pos=3,col="black")
text(wbeta1[9,m],wbeta2[9,m],9,pos=2,col="black")
text(wbeta1[14,m],wbeta2[14,m],14,pos=4,col="black")
par(new=F)
```

```
plot.new()
box(col="grey")
text(.5, .5,"beta2",cex=1)

plot(wbeta3,wbeta2,xlab="",ylab="",cex=0.1)
text(wbeta3[7,m],wbeta2[7,m],7,pos=3,col="black"
)
text(wbeta3[8,m],wbeta2[8,m],8,pos=3,col="black"
)
text(wbeta3[9,m],wbeta2[9,m],9,pos=3,col="black"
)
text(wbeta3[14,m],wbeta2[14,m],14,pos=3,col="bla
ck")
par(new=F)

plot(wbeta12,wbeta2,xlab="",ylab="",cex=0.1)
text(wbeta12[7,m],wbeta2[7,m],7,pos=3,col="black
")
text(wbeta12[8,m],wbeta2[8,m],8,pos=3,col="black
")
text(wbeta12[9,m],wbeta2[9,m],9,pos=3,col="black
")
text(wbeta12[14,m],wbeta2[14,m],14,pos=4,col="b
lack")
par(new=F)

plot(wbeta13,wbeta2,xlab="",ylab="",cex=0.1)
text(wbeta13[7,m],wbeta2[7,m],7,pos=3,col="black
")
text(wbeta13[8,m],wbeta2[8,m],8,pos=3,col="black
")
text(wbeta13[9,m],wbeta2[9,m],9,pos=3,col="black
")
text(wbeta13[14,m],wbeta2[14,m],14,pos=2,col="b
lack")
par(new=F)
plot(wbeta23,wbeta2,xlab="",ylab="",cex=0.1)
text(wbeta23[7,m],wbeta2[7,m],7,pos=3,col="black
")
text(wbeta23[8,m],wbeta2[8,m],8,pos=2,col="black
")
text(wbeta23[9,m],wbeta2[9,m],9,pos=4,col="black
")
text(wbeta23[14,m],wbeta2[14,m],14,pos=3,col="b
lack")
par(new=F)

plot(wbeta123,wbeta2,xlab="",ylab="",cex=0.1)
text(wbeta123[7,m],wbeta2[7,m],7,pos=3,col="blac
k")
text(wbeta123[8,m],wbeta2[8,m],8,pos=3,col="blac
k")
text(wbeta123[9,m],wbeta2[9,m],9,pos=3,col="blac
k")
text(wbeta123[14,m],wbeta2[14,m],14,pos=3,col="
black")
par(new=F)

plot(wSSE,wbeta2,xlab="",ylab="",cex=0.1)
text(wSSE[7,m],wbeta2[7,m],7,pos=3,col="black")
text(wSSE[8,m],wbeta2[8,m],8,pos=4,col="black")
text(wSSE[9,m],wbeta2[9,m],9,pos=3,col="black")
text(wSSE[14,m],wbeta2[14,m],14,pos=4,col="blac
k")
par(new=F)

plot(wbeta1,wbeta3,xlab="",ylab="",cex=0.1)
text(wbeta1[7,m],wbeta3[7,m],7,pos=1,col="black"
)
text(wbeta1[8,m],wbeta3[8,m],8,pos=1,col="black"
)
text(wbeta1[9,m],wbeta3[9,m],9,pos=2,col="black"
)
text(wbeta1[14,m],wbeta3[14,m],14,pos=1,col="bla
ck")
par(new=F)

plot(wbeta2,wbeta3,xlab="",ylab="",cex=0.1)
text(wbeta2[7,m],wbeta3[7,m],7,pos=1,col="black"
)
text(wbeta2[8,m],wbeta3[8,m],8,pos=1,col="black"
)
text(wbeta2[9,m],wbeta3[9,m],9,pos=1,col="black"
)
text(wbeta2[14,m],wbeta3[14,m],14,pos=1,col="bla
ck")
par(new=F)

plot.new()
box(col="grey")
text(.5, .5,"beta3",cex=1)

plot(wbeta12,wbeta3,xlab="",ylab="",cex=0.1)
text(wbeta12[7,m],wbeta3[7,m],7,pos=1,col="black
")
text(wbeta12[8,m],wbeta3[8,m],8,pos=1,col="black
")
```

```
text(wbeta12[9,m],wbeta3[9,m],9,pos=1,col="black
")
text(wbeta12[14,m],wbeta3[14,m],14,pos=1,col="b
lack")
par(new=F)

plot(wbeta13,wbeta3,xlab="",ylab="",cex=0.1)
text(wbeta13[7,m],wbeta3[7,m],7,pos=3,col="black
")
text(wbeta13[8,m],wbeta3[8,m],8,pos=1,col="black
")
text(wbeta13[9,m],wbeta3[9,m],9,pos=1,col="black
")
text(wbeta13[14,m],wbeta3[14,m],14,pos=1,col="b
lack")
par(new=F)

plot(wbeta23,wbeta3,xlab="",ylab="",cex=0.1)
text(wbeta23[7,m],wbeta3[7,m],7,pos=2,col="black
")
text(wbeta23[8,m],wbeta3[8,m],8,pos=1,col="black
")
text(wbeta23[9,m],wbeta3[9,m],9,pos=4,col="black
")
text(wbeta23[14,m],wbeta3[14,m],14,pos=1,col="b
lack")
par(new=F)

plot(wbeta123,wbeta3,xlab="",ylab="",cex=0.1)
text(wbeta123[7,m],wbeta3[7,m],7,pos=1,col="blac
k")
text(wbeta123[8,m],wbeta3[8,m],8,pos=1,col="blac
k")
text(wbeta123[9,m],wbeta3[9,m],9,pos=1,col="blac
k")
text(wbeta123[14,m],wbeta3[14,m],14,pos=1,col="
black")
par(new=F)

plot(wSSE,wbeta3,xlab="",ylab="",cex=0.1)
text(wSSE[7,m],wbeta3[7,m],7,pos=1,col="black")
text(wSSE[8,m],wbeta3[8,m],8,pos=1,col="black")
text(wSSE[9,m],wbeta3[9,m],9,pos=3,col="black")
text(wSSE[14,m],wbeta3[14,m],14,pos=4,col="blac
k")
par(new=F)

plot(wbeta1,wbeta12,xlab="",ylab="",cex=0.1)
text(wbeta1[7,m],wbeta12[7,m],7,pos=1,col="black
")
text(wbeta1[8,m],wbeta12[8,m],8,pos=3,col="black
")
text(wbeta1[9,m],wbeta12[9,m],9,pos=3,col="black
")
text(wbeta1[14,m],wbeta12[14,m],14,pos=3,col="b
lack")
par(new=F)

plot(wbeta2,wbeta12,xlab="",ylab="",cex=0.1)
text(wbeta2[7,m],wbeta12[7,m],7,pos=1,col="black
")
text(wbeta2[8,m],wbeta12[8,m],8,pos=3,col="black
")
text(wbeta2[9,m],wbeta12[9,m],9,pos=3,col="black
")
text(wbeta2[14,m],wbeta12[14,m],14,pos=3,col="b
lack")
par(new=F)

plot(wbeta3,wbeta12,xlab="",ylab="",cex=0.1)
text(wbeta3[7,m],wbeta12[7,m],7,pos=1,col="black
")
text(wbeta3[8,m],wbeta12[8,m],8,pos=3,col="black
")
text(wbeta3[9,m],wbeta12[9,m],9,pos=3,col="black
")
text(wbeta3[14,m],wbeta12[14,m],14,pos=3,col="b
lack")
par(new=F)

plot.new()
box(col="grey")
text(.5,.5,"beta12",cex=1)

plot(wbeta13,wbeta12,xlab="",ylab="",cex=0.1)
text(wbeta13[7,m],wbeta12[7,m],7,pos=1,col="blac
k")
text(wbeta13[8,m],wbeta12[8,m],8,pos=3,col="blac
k")
text(wbeta13[9,m],wbeta12[9,m],9,pos=4,col="blac
k")
text(wbeta13[14,m],wbeta12[14,m],14,pos=3,col="
black")
par(new=F)

plot(wbeta23,wbeta12,xlab="",ylab="",cex=0.1)
```

```
text(wbeta23[7,m],wbeta12[7,m],7,pos=1,col="black")
text(wbeta23[8,m],wbeta12[8,m],8,pos=3,col="black")
text(wbeta23[9,m],wbeta12[9,m],9,pos=3,col="black")
text(wbeta23[14,m],wbeta12[14,m],14,pos=3,col="black")
par(new=F)

plot(wbeta123,wbeta12,xlab="",ylab="",cex=0.1)
text(wbeta123[7,m],wbeta12[7,m],7,pos=1,col="black")
text(wbeta123[8,m],wbeta12[8,m],8,pos=3,col="black")
text(wbeta123[9,m],wbeta12[9,m],9,pos=3,col="black")
text(wbeta123[14,m],wbeta12[14,m],14,pos=3,col="black")
par(new=F)

plot(wSSE,wbeta12,xlab="",ylab="",cex=0.1)
text(wSSE[7,m],wbeta12[7,m],7,pos=1,col="black")
text(wSSE[8,m],wbeta12[8,m],8,pos=3,col="black")
text(wSSE[9,m],wbeta12[9,m],9,pos=3,col="black")
text(wSSE[14,m],wbeta12[14,m],14,pos=4,col="black")
par(new=F)

plot(wbeta1,wbeta13,xlab="",ylab="",cex=0.1)
text(wbeta1[7,m],wbeta13[7,m],7,pos=1,col="black")
text(wbeta1[8,m],wbeta13[8,m],8,pos=3,col="black")
text(wbeta1[9,m],wbeta13[9,m],9,pos=1,col="black")
text(wbeta1[14,m],wbeta13[14,m],14,pos=1,col="black")
par(new=F)

plot(wbeta2,wbeta13,xlab="",ylab="",cex=0.1)
text(wbeta2[7,m],wbeta13[7,m],7,pos=1,col="black")
text(wbeta2[8,m],wbeta13[8,m],8,pos=3,col="black")
text(wbeta2[9,m],wbeta13[9,m],9,pos=1,col="black")
text(wbeta2[14,m],wbeta13[14,m],14,pos=1,col="black")
par(new=F)

plot(wbeta3,wbeta13,xlab="",ylab="",cex=0.1)
text(wbeta3[7,m],wbeta13[7,m],7,pos=2,col="black")
text(wbeta3[8,m],wbeta13[8,m],8,pos=3,col="black")
text(wbeta3[9,m],wbeta13[9,m],9,pos=1,col="black")
text(wbeta3[14,m],wbeta13[14,m],14,pos=1,col="black")
par(new=F)

plot(wbeta12,wbeta13,xlab="",ylab="",cex=0.1)
text(wbeta12[7,m],wbeta13[7,m],7,pos=1,col="black")
text(wbeta12[8,m],wbeta13[8,m],8,pos=3,col="black")
text(wbeta12[9,m],wbeta13[9,m],9,pos=1,col="black")
text(wbeta12[14,m],wbeta13[14,m],14,pos=1,col="black")
par(new=F)

plot.new()
box(col="grey")
text(.5, .5,"beta13",cex=1)

plot(wbeta23,wbeta13,xlab="",ylab="",cex=0.1)
text(wbeta23[7,m],wbeta13[7,m],7,pos=1,col="black")
text(wbeta23[8,m],wbeta13[8,m],8,pos=3,col="black")
text(wbeta23[9,m],wbeta13[9,m],9,pos=4,col="black")
text(wbeta23[14,m],wbeta13[14,m],14,pos=1,col="black")
par(new=F)

plot(wbeta123,wbeta13,xlab="",ylab="",cex=0.1)
text(wbeta123[7,m],wbeta13[7,m],7,pos=1,col="black")
text(wbeta123[8,m],wbeta13[8,m],8,pos=3,col="black")
```

```
text(wbeta123[9,m],wbeta13[9,m],9,pos=1,col="bla
ck")
text(wbeta123[14,m],wbeta13[14,m],14,pos=1,col
="black")
par(new=F)

plot(wSSE,wbeta13,xlab="",ylab="",cex=0.1)
text(wSSE[7,m],wbeta13[7,m],7,pos=2,col="black"
)
text(wSSE[8,m],wbeta13[8,m],8,pos=3,col="black"
)
text(wSSE[9,m],wbeta13[9,m],9,pos=1,col="black"
)
text(wSSE[14,m],wbeta13[14,m],14,pos=4,col="bl
ack")
par(new=F)

plot(wbeta1,wbeta23,xlab="",ylab="",cex=0.1)
text(wbeta1[7,m],wbeta23[7,m],7,pos=1,col="black
")
text(wbeta1[8,m],wbeta23[8,m],8,pos=1,col="black
")
text(wbeta1[9,m],wbeta23[9,m],9,pos=3,col="black
")
text(wbeta1[14,m],wbeta23[14,m],14,pos=1,col="b
lack")
par(new=F)

plot(wbeta2,wbeta23,xlab="",ylab="",cex=0.1)
text(wbeta2[7,m],wbeta23[7,m],7,pos=1,col="black
")
text(wbeta2[8,m],wbeta23[8,m],8,pos=1,col="black
")
text(wbeta2[9,m],wbeta23[9,m],9,pos=3,col="black
")
text(wbeta2[14,m],wbeta23[14,m],14,pos=1,col="b
lack")
par(new=F)

plot(wbeta3,wbeta23,xlab="",ylab="",cex=0.1)
text(wbeta3[7,m],wbeta23[7,m],7,pos=1,col="black
")
text(wbeta3[8,m],wbeta23[8,m],8,pos=1,col="black
")
text(wbeta3[9,m],wbeta23[9,m],9,pos=3,col="black
")
text(wbeta3[14,m],wbeta23[14,m],14,pos=1,col="b
lack")
par(new=F)

plot(wbeta12,wbeta23,xlab="",ylab="",cex=0.1)
text(wbeta12[7,m],wbeta23[7,m],7,pos=1,col="blac
k")
text(wbeta12[8,m],wbeta23[8,m],8,pos=1,col="blac
k")
text(wbeta12[9,m],wbeta23[9,m],9,pos=3,col="blac
k")
text(wbeta12[14,m],wbeta23[14,m],14,pos=1,col="
black")
par(new=F)

plot(wbeta13,wbeta23,xlab="",ylab="",cex=0.1)
text(wbeta13[7,m],wbeta23[7,m],7,pos=1,col="blac
k")
text(wbeta13[8,m],wbeta23[8,m],8,pos=1,col="blac
k")
text(wbeta13[9,m],wbeta23[9,m],9,pos=3,col="blac
k")
text(wbeta13[14,m],wbeta23[14,m],14,pos=1,col="
black")
par(new=F)

plot.new()
box(col="grey")
text(.5,.5,"beta23",cex=1)

plot(wbeta123,wbeta23,xlab="",ylab="",cex=0.1)
text(wbeta123[7,m],wbeta23[7,m],7,pos=1,col="bla
ck")
text(wbeta123[8,m],wbeta23[8,m],8,pos=1,col="bla
ck")
text(wbeta123[9,m],wbeta23[9,m],9,pos=3,col="bla
ck")
text(wbeta123[14,m],wbeta23[14,m],14,pos=1,col
="black")
par(new=F)

plot(wSSE,wbeta23,xlab="",ylab="",cex=0.1)
text(wSSE[7,m],wbeta23[7,m],7,pos=1,col="black"
)
text(wSSE[8,m],wbeta23[8,m],8,pos=1,col="black"
)
text(wSSE[9,m],wbeta23[9,m],9,pos=3,col="black"
)
text(wSSE[14,m],wbeta23[14,m],14,pos=4,col="bl
ack")
par(new=F)

plot(wbeta1,wbeta123,xlab="",ylab="",cex=0.1)
```

```
text(wbeta1[7,m],wbeta123[7,m],7,pos=4,col="blac
k")
text(wbeta1[8,m],wbeta123[8,m],8,pos=1,col="blac
k")
text(wbeta1[9,m],wbeta123[9,m],9,pos=1,col="blac
k")
text(wbeta1[14,m],wbeta123[14,m],14,pos=1,col="
black")
par(new=F)

plot(wbeta2,wbeta123,xlab="",ylab="",cex=0.1)
text(wbeta2[7,m],wbeta123[7,m],7,pos=4,col="blac
k")
text(wbeta2[8,m],wbeta123[8,m],8,pos=1,col="blac
k")
text(wbeta2[9,m],wbeta123[9,m],9,pos=1,col="blac
k")
text(wbeta2[14,m],wbeta123[14,m],14,pos=1,col="
black")
par(new=F)

plot(wbeta3,wbeta123,xlab="",ylab="",cex=0.1)
text(wbeta3[7,m],wbeta123[7,m],7,pos=2,col="blac
k")
text(wbeta3[8,m],wbeta123[8,m],8,pos=1,col="blac
k")
text(wbeta3[9,m],wbeta123[9,m],9,pos=1,col="blac
k")
text(wbeta3[14,m],wbeta123[14,m],14,pos=1,col="
black")
par(new=F)

plot(wbeta12,wbeta123,xlab="",ylab="",cex=0.1)
text(wbeta12[7,m],wbeta123[7,m],7,pos=2,col="bla
ck")
text(wbeta12[8,m],wbeta123[8,m],8,pos=1,col="bla
ck")
text(wbeta12[9,m],wbeta123[9,m],9,pos=1,col="bla
ck")
text(wbeta12[14,m],wbeta123[14,m],14,pos=1,col
="black")
par(new=F)

plot(wbeta13,wbeta123,xlab="",ylab="",cex=0.1)
text(wbeta13[7,m],wbeta123[7,m],7,pos=2,col="bla
ck")
text(wbeta13[8,m],wbeta123[8,m],8,pos=1,col="bla
ck")
text(wbeta13[9,m],wbeta123[9,m],9,pos=1,col="bla
ck")
text(wbeta13[14,m],wbeta123[14,m],14,pos=1,col
="black")
par(new=F)

plot(wbeta23,wbeta123,xlab="",ylab="",cex=0.1)
text(wbeta23[7,m],wbeta123[7,m],7,pos=1,col="bla
ck")
text(wbeta23[8,m],wbeta123[8,m],8,pos=1,col="bla
ck")
text(wbeta23[9,m],wbeta123[9,m],9,pos=4,col="bla
ck")
text(wbeta23[14,m],wbeta123[14,m],14,pos=1,col
="black")
par(new=F)

plot.new()
box(col="grey")
text(.5,.5,"beta123",cex=1)

plot(wSSE,wbeta123,xlab="",ylab="",cex=0.1)
text(wSSE[7,m],wbeta123[7,m],7,pos=2,col="black
")
text(wSSE[8,m],wbeta123[8,m],8,pos=1,col="black
")
text(wSSE[9,m],wbeta123[9,m],9,pos=1,col="black
")
text(wSSE[14,m],wbeta123[14,m],14,pos=4,col="b
lack")
par(new=F)

plot(wbeta1,wSSE,xlab="",ylab="",cex=0.1)
text(wbeta1[7,m],wSSE[7,m],7,pos=2,col="black")
text(wbeta1[8,m],wSSE[8,m],8,pos=1,col="black")
text(wbeta1[9,m],wSSE[9,m],9,pos=1,col="black")
text(wbeta1[14,m],wSSE[14,m],14,pos=3,col="blac
k")
par(new=F)

plot(wbeta2,wSSE,xlab="",ylab="",cex=0.1)
text(wbeta2[7,m],wSSE[7,m],7,pos=1,col="black")
text(wbeta2[8,m],wSSE[8,m],8,pos=1,col="black")
text(wbeta2[9,m],wSSE[9,m],9,pos=1,col="black")
text(wbeta2[14,m],wSSE[14,m],14,pos=3,col="blac
k")
par(new=F)

plot(wbeta3,wSSE,xlab="",ylab="",cex=0.1)
```

```
text(wbeta3[7,m],wSSE[7,m],7,pos=2,col="black")
text(wbeta3[8,m],wSSE[8,m],8,pos=1,col="black")
text(wbeta3[9,m],wSSE[9,m],9,pos=4,col="black")
text(wbeta3[14,m],wSSE[14,m],14,pos=3,col="black")
par(new=F)

plot(wbeta12,wSSE,xlab="",ylab="",cex=0.1)
text(wbeta12[7,m],wSSE[7,m],7,pos=1,col="black")
text(wbeta12[8,m],wSSE[8,m],8,pos=1,col="black")
text(wbeta12[9,m],wSSE[9,m],9,pos=4,col="black")
text(wbeta12[14,m],wSSE[14,m],14,pos=3,col="black")
par(new=F)

plot(wbeta13,wSSE,xlab="",ylab="",cex=0.1)
text(wbeta13[7,m],wSSE[7,m],7,pos=1,col="black")
text(wbeta13[8,m],wSSE[8,m],8,pos=1,col="black")
text(wbeta13[9,m],wSSE[9,m],9,pos=2,col="black")
text(wbeta13[14,m],wSSE[14,m],14,pos=3,col="black")
par(new=F)

plot(wbeta23,wSSE,xlab="",ylab="",cex=0.1)
text(wbeta23[7,m],wSSE[7,m],7,pos=2,col="black")
text(wbeta23[8,m],wSSE[8,m],8,pos=1,col="black")
text(wbeta23[9,m],wSSE[9,m],9,pos=4,col="black")
text(wbeta23[14,m],wSSE[14,m],14,pos=3,col="black")
par(new=F)

plot(wbeta123,wSSE,xlab="",ylab="",cex=0.1)
text(wbeta123[7,m],wSSE[7,m],7,pos=1,col="black")
text(wbeta123[8,m],wSSE[8,m],8,pos=1,col="black")
text(wbeta123[9,m],wSSE[9,m],9,pos=1,col="black")
text(wbeta123[14,m],wSSE[14,m],14,pos=3,col="black")
par(new=F)

plot.new()
box(col="grey")
text(.5,.5,"SSE",cex=1)

###### 3D firework plot ######

plot3d(wbeta1[1:n,],wbeta2[1:n,],wbeta3[1:n,],xlab="Beta1",ylab="Beta2",zlab="Beta3",main="",size=2)
text3d(wbeta1[-c(7,8,9,14),m],wbeta2[-c(7,8,9,14),m],wbeta3[-c(7,8,9,14),m],seq(n)[-c(7,8,9,14)],adj=1)
text3d(wbeta1[c(7,8,9,14),m],wbeta2[c(7,8,9,14),m],wbeta3[c(7,8,9,14),m],c(7,8,9,14),adj=1,cex=1.5)

plot3d(wbeta12[1:n,],wbeta123[1:n,],wSSE[1:n,],xlab="Beta12",ylab="Beta123",zlab="SSE",main="",size=2)
text3d(wbeta12[-c(7,8,9,14),m],wbeta123[-c(7,8,9,14),m],wSSE[-c(7,8,9,14),m],seq(n)[-c(7,8,9,14)],adj=1)
text3d(wbeta12[c(7,8,9,14),m],wbeta123[c(7,8,9,14),m],wSSE[c(7,8,9,14),m],c(7,8,9,14),adj=1,cex=1.5)

######## 회귀진단 #########

추정반응값=fitted(혼합물모형)
스튜던트잔차=rstudent(혼합물모형)
반응값_잔차=cbind(y,추정반응값,스튜던트잔차)
반응값_잔차
#hat: h_ii, coefficients: y_i 제거 후 구한 추정회귀계수, sigma: y_i 제거 후 구한 root(MSE), wt.res: 가중잔차(weighted(deviance) residual)
influence(혼합물모형)
# 영향점 검출 통계량들
영향력=influence.measures(혼합물모형)
영향력
# 영향점만 출력
which(apply(영향력$is.inf,1,any))
summary(영향력)
# 특이점[이상점] 검출 통계량
ls.diag(혼합물모형)
# residual plot
residualPlots(혼합물모형)
# 특이점검정: Bonferronni p-value
outlierTest(혼합물모형)
```