Analysis of Variance and Design of Experiments-II

MODULE VIII

LECTURE - 34

RESPONSE SURFACE DESIGNS

Dr. Shalabh
Department of Mathematics & Statistics
Indian Institute of Technology Kanpur

Least squares analysis

Using the least squares principle, the normal equations for the β_i^* of the model

$$y(x_1, x_2, ..., x_k)_\ell = \beta_0^* + \sum_{i=1}^{k} \beta_i^* x_i + \varepsilon(x_1, x_2, ..., x_k)_\ell,$$

are obtained as

$$(L,\ D)'(L,\ D)\hat{\beta}^* = (L,\ D)'y$$

or

$$N\hat{\beta}^* = \begin{pmatrix} L'y \\ D'y \end{pmatrix}.$$

Thus

$$\hat{\beta}_0^* = \frac{1}{N} \sum_{x_1, x_2, ..., x_k} \sum_{\ell} y(x_1, x_2, ..., x_k)_\ell = \bar{y}$$

and

$$\hat{\beta}_i^* = \frac{1}{N} d_i' y$$

$$= \frac{1}{N}[(\text{sum of all observations with } x_i = 1) \text{ - } (\text{sum of all observations with } x_i = 1)].$$

Then

$$Var(\hat{\beta}_i^*) = \frac{1}{N}\sigma_\varepsilon^2$$

and

$$Cov(\hat{\beta}_i^*, \hat{\beta}_{i'}^*) = 0.$$

So, for any given point $z=(z_1,z_2,...,z_k)'$ in the ER given by $\{-1\le z_i\le 1; i=1,2,...,k\}$ we obtain the predicted response

$$\hat{y}(z)=\hat{\beta}_0^*+\sum_{i=1}^{k}\hat{\beta}_i^* z_i$$

with variance

$$Var[\hat{y}(z)]=\frac{1}{N}\left(1+\sum_{i=1}^{k} z_i^2\right)\sigma_\varepsilon^2.$$

In order to know that which of the factors are influential and also to know the response surface given by $\hat{y}(z)$, we need to obtain an estimate of σ_ε^2. This is achieved, as usual, through the ANOVA as given in the following table with notations

$$y(x)_\ell = y(x_1,x_2,...,x_k)_\ell$$

$$\sum_x = \sum_{x_1,x_2,...,x_k}$$

$$D_1 = SS(Total) - N\sum_{i=1}^{k}\left(\hat{\beta}_i^*\right)^2$$

$$D_2 = D_1 - SS(PE)$$

ANOVA for first-order response surface design

Source	Degrees of freedom	Sum of squares
Regression	k	
β_1^*	1	$N(\hat{\beta}_1^*)^2$
β_2^*	1	$N(\hat{\beta}_2^*)^2$
$\vdots$	$\vdots$	$\vdots$
β_k^*	1	$N(\hat{\beta}_k^*)^2$
Error	$r2^k - k - 1$	$D_1 = SSE$
- Lack-of-fit error	$2^k - k - 1$	$D_2 = \mathrm{SS}(LOF)$
- Pure error	$2^k(r-1)$	$\sum_x \sum_\ell \left[y(x)_\ell - \overline{y}(x)_o \right]^2 = \mathrm{SS}(PE)$
Total	$N - 1$	$\sum_{xo\ell} \left[y(x)_\ell - \overline{y}(o)_o \right]^2$

Note that the *SSE* consists of following two parts:

i. The usual error sum of squares for a CRD, denoted here by SS(*PE*) and

ii. The lack of fit (LOF) sum of the sums of squares for all the interactions for the 2^k factorial denoted here by SS(*LOF*). This sum of squares can be used to test whether the postulated model

$$y(x_1, x_2, ..., x_k)_\ell = \beta_0^* + \sum_{i=1}^{k} \beta_i^* x_i + \varepsilon(x_1, x_2, ..., x_k)_\ell$$

provides a sufficiently good enough fit to the data.

To test whether the i^{th} factor contributes in explaining the response, we use the following *F* - statistic

$$F_i = \frac{\text{SS}(\beta_i^*)}{\text{MS}(E)} \quad (i = 1, 2, ..., k)$$

which follows the *F*-distribution with 1 and $(N - k - 1)$ degrees of freedom. Suppose without loss of generality, only the first k_1 factors are important. Then instead of using the model $y(x_1, x_2, ..., x_k)_\ell = \beta_0^* + \sum_{i=1}^{k} \beta_i^* x_i + \varepsilon(x_1, x_2, ..., x_k)_\ell$, the following model based on k_1 factors is used:

$$y(x_1, x_2, ..., x_k)_\ell = \beta_0^* + \sum_{i=1}^{k_1} \beta_i^* x_i + \varepsilon(x_1, x_2, ..., x_k)_\ell$$

and the predicted response then becomes

$$\hat{y}(z) = \hat{\beta}_0^* + \sum_{i=1}^{k_1} \hat{\beta}_i^* z_i$$

with the estimated variance as

$$\widehat{Var}\left[\hat{y}(z)\right] = \frac{1}{N}\left(1 + \sum_{i=1}^{k_1} z_i^2\right) MSE$$

Then the responses for two different sets of input variables, $z = \left(z_1, z_2, ..., z_{k_1}\right)'$ and $w = \left(w_1, w_2, .., w_{k_1}\right)'$ are compared by considering the difference in the predicted values based on these input variable as

$$\hat{y}(z) - \hat{y}(w) = \sum_{i=1}^{k_1} \hat{\beta}_i^* (z_i - w_i)$$

and its estimated variance is given by

$$\widehat{Var}\left[\hat{y}(z) - \hat{y}(w)\right] = \frac{1}{N} \sum_{i=1}^{k_1} (z_i - w_i)^2 \, MSE.$$

Similarly, the experimenter can also consider the differences in responses if some of the input variables are kept constant at a desired level and the remaining input variables are varied to achieve optimum response in ER.

Since the true response surface is being approximated and due to experimental error, there may not exist a single level combination which achieves the optimum response. Instead of this, there may exist a neighbourhood in which the optimum may lie and this optimum may not be significantly different from each other.

Alternative Design

It may not be a good idea to use the full 2^k factorial to estimate the parameters of a first-order response surface as this may involve large number of observations to handle. There are basically two ways to reduce the number of experimental points. One way is to replicate each design point $(x_1, x_2, ..., x_k)$ only once and in such case

SS(*PE*) = 0 and *SS*(*E*) = *SS*(*LOF*).

Another alternative is to use only a fraction of a 2^k factorial either as a single replicate or as a CRD with more than one replications. In either case, the experimenter has to choose a fraction such that all the *k* main effects are estimable with sufficient degrees of freedom for error so that comparisons like $\hat{y}(z) - \hat{y}(w)$ can be made with satisfactory statistical power as measured by its variance. This means that if we need to choose a very small fraction, then this can be achieved by fractional factorials, with several replications for each design point.

An important aspect in a 2^k factorial is that the blocking can be introduced easily without sacrificing the estimation of the main effects. This will help in reducing the experimental material as well as the cost and provide simplicity in the experimentation. We discussed this aspect in fractional factorial module.

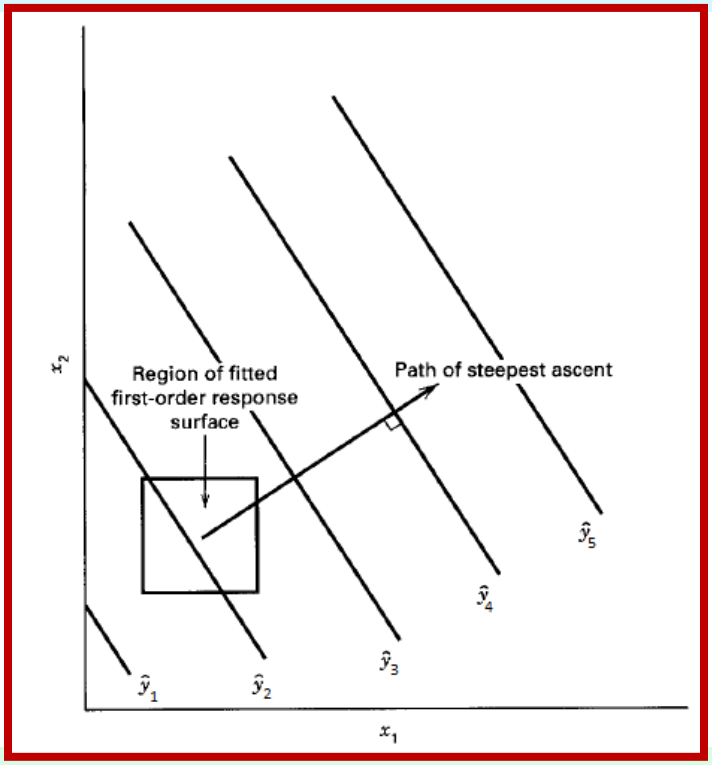
The method of steepest ascent

In many experimental conditions, the initial estimate of the optimum operating conditions for the system may be away from the actual optimum. In such conditions, one would like to move rapidly to the general vicinity of the optimum. It is expected to have a procedure which is simple to use and economically efficient. When the experimenter is remote away from the optimum, then usually it is assumed that a first-order model is an adequate approximation to the true surface in a small region of the *x*'s.

The **method of steepest ascent** is a procedure for moving sequentially along the path of steepest ascent, i.e., in the direction of the maximum increase in the response. If minimization is desired, then this technique is called as the **method of steepest descent**. The fitted first-order model is

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i$$

and the first-order response surface can be represented as the contours of $\hat{y}$. The contours are a series of parallel lines such as shown in following figure:



First order response surface and path of steepest ascent

The direction of steepest ascent is the direction in which $\hat{y}$ increases most rapidly. Such direction is parallel to the normal to the fitted response surface.

The experimenter usually takes as the **path of steepest ascent** the line through the center of the region of interest and normal to the fitted surface. Thus, the steps along the path are proportional to the regression coefficients $\hat{\beta}_i's$. The actual step size is determined by the experimenter based on process knowledge or other practical considerations.

The experiments are continued to be conducted along the path of steepest ascent until no further increase in response is observed.

Then a new first-order model which may be a fit, a new path of steepest ascent determined, and the procedure continued. Finally, the experimenter will arrive in the vicinity of the optimum.

This is judged by the lack of fit test of a first-order model. Some additional experiments are conducted to obtain a more precise estimate of the optimum at this point.