Response Surface Designs

With response surface designs, the goal is to optimize a response with respect to a set of quantitative variables, often by using sequential experimentation. We use polynomial models to approximate the response surface. These methods are applicable to experiments with several experimental factors, but we shall follow the text's discussion and focus on the simpler case of using two factors. The process often begins by using a 2^n factorial design with additional center points, using a **first-order design**. If curvature is detected (**lack of fit** to the first-order model), then we conclude that we are near the optimum and add points to obtain a **second-order design**. If curvature is not detected, we are far from the optimum and use the method of **steepest ascent** to get points headed (hopefully) toward to optimum. As we recognize that we are near the optimum, we use a second-order design.

First-order designs and analysis

In the vinylation example from the text, there are two factors, temperature and pressure. To analyze linear effects of these factors, a 2^2 factorial design is used at predesignated low and high values of temperature and pressure. We also include four points in the center of the low and high values. Note that the values are typically coded so that -1 is the low value, 1 is the high value, and the center points have 0 values. From this data, we fit the model

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_{ij}$$

where i indexes the observations and j indexes replication at a set of (x_{1i}, x_{2i}) values. An example of analysis of this kind of model is shown on page 428 of our text. Individual tests are conducted for the x_i coefficients, and also tests are discussed for interaction and lack of fit. The lack-of-fit test uses two sources of error: the first source is the variation among values sharing x levels (the center points), called pure error, and the second source is the rest of the residual variation, called lack-of-fit. In a 2^n design with m center points, there are $2^n + m - 1$ degrees of freedom with n for the x terms, $m - 1$ for pure error, and the rest for lack of fit.

If there is no lack-of-fit:

Then the linear model seems to approximate the response surface in this region, which means we are far from an optimum. In this case we want to obtain additional responses (design points) that go closer to the optimum. To do this we use the method of steepest ascent, which just means that future design points should increase the values of x_1 and x_2 for the design center in proportion to $(\widehat{\beta}_1, \widehat{\beta}_2)$ until our observed responses no longer increase, suggesting that we are near the optimum response. In the vinylation example, the first order prediction equation is

$$\widehat{y}_{ij} = 20 + 8x_{1i} + 4x_{2i},$$

so we increase 4 units in the x_2 direction for every 8 units in the x_1 direction. Once we are near the optimum, we then use a second-order design in this new region.

If there is lack-of-fit:

If the lack-of-fit test is significant, it indicates that the current region has curvature and should be near the optimum. We then augment our first-order design with additional points to yield a second-order design.

Second-order designs and analysis

A design for curved surfaces needs to have more distinct x values to detect and check lack-of-fit to a quadratic surface. Designs based on 3^n or higher levels per factor can require many observations, so instead the **central composite design (CCD)** can be used. It includes three parts: a 2^n factorial part, an axial part that consists of x values that are 0 in all but one coordinate, and a part that is a set of center points.

Choosing the axial values (α) **and number of center points** (m)

The axial value α and the number of center points m are usually chosen to achieve a rotatable (the variance of $\widehat{y}_{ij}$ is constant at points equally distant from the center) design with uniform precision for the estimated surface within one unit of the design center. A table of these values is shown in the text.

Second-order analysis

The second order model (with two x factors) is

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_{ij}.$$

As with a first-order analysis, we conduct tests for the design variables (which now involves more terms) and perform a lack-of-fit test. We can also perform a canonical analysis to understand the type of second order surface, the location of the optimum value, and canonical directions. The canonical analysis transforms the second order model into

$$\widehat{y}_{ij} = \widehat{y}_s + \lambda_1 Z_{1i}^2 + \lambda_2 Z_{2i}^2,$$

where the Z variables are rotated axes and the λ values are used to describe the fitted quadratic model. If all λ values are positive, then the surface is a minimum; if all λ values are negative the surface is a maximum. If some are positive and some negative, the surface is a saddle point. If some λ values are close to zero, then the surface has one or more ridges. The value $\widehat{y}_s$ is the optimum, and the Z variables express the rotation that is performed to eliminate cross-product (x_1x_2) terms from the original model. As shown in the text and on the SAS code on the webpage, we can use standard statistical software to calculate the λ values and Z variables for the canonical analysis.