LECTURE 9

Mixture Designs

1. Design and Analysis of Mixture Experiments

There are many times when the product we are interested in is a mixture. In other words, we are more interested in the proportions than the total amounts of the components. In product formulations, examples would be gasoline, soaps or detergents, beverages, cake mixes, soups, and so on. There are examples in process engineering, as well. For example, in the production of semiconductor wafers we might be interested in the proportions of various acids for the acid wash.

The fact that the proportions must add up to one is the key attribute of mixture designs. Specifically, the settings for various factors must satisfy

$$x_i \geq 0, \qquad \text{for all } i$$
$$\sum_i x_i = 1$$

The design region for mixture proportions is a simplex, a regularly sided figure of dimenstion $k-1$ with k vertices (and usually embedded into a k dimensional space. For example, with two factors, the simplex is the line segment from (0,1) to (1,0). With three factors, the simplex would have vertices at (1,0,0), (0,1,0), and (0,0,1). There is a corresponding simplex coordinate system.
Draw pictures.

We can now consider models for mixture experiments. The usual first order model is

$$E(y) = \beta_0 + \sum \beta_i x_i.$$

However, since $\sum x_i = 1$ for a mixture model, the β_i's will not be uniquely determined. We could choose to eliminate one of the x_i's, but a better approach was suggested by Scheffé. In the equation above multiply β_0 by $1 = \sum x_i$ to get

$$E(y) = \sum (\beta_0 + \beta_i) x_i.$$

Relabeling the β_i's, we get the following canonical forms.
Linear:

$$E(y) = \sum \beta_i x_i,$$

Quadratic:

$$E(y) = \sum \beta_i x_i + \sum \sum_{i<j} x_i x_j,$$

©Steven Buyske and Richard Trout.

Special Cubic:

$$E(y) = \sum \beta_i x_i + \sum \sum\nolimits_{i<j} x_i x_j + \sum \sum \sum\nolimits_{i<j<k} x_i x_j x_k,$$

Full Cubic:

$$E(y) = \sum \beta_i x_i + \sum \sum\nolimits_{i<j} \beta_{ij} x_i x_j + \sum \sum\nolimits_{i<j} \delta_{ij} x_i x_j (x_i - x_j) + \sum \sum \sum\nolimits_{i<j<k} x_i x_j x_k.$$

This last is rarely used. There are many other possible models. We mention just one, the Draper and St. John model, which can be useful when some of the components work well in small amounts (spices, for example, in food products).

$$E(y) = \sum \beta_i x_i + \sum \nu_i x_i^{-1}$$
$$E(y) = \sum \beta_i x_i + \sum \sum\nolimits_{i<j} \beta_{ij} x_i x_j + \sum \nu_i x_i^{-1}.$$

The terms in the canonical mixture models have simple interpretations.
Draw pictures.
Geometrically, the parameter β_j represents the expected response from a pure mixture with $x_j = 1$ (and all other components zero). The $\sum \beta_i x_i$ term is called the linear blending term. The quadratic terms should not be thought of as interaction but instead are called nonlinear blending terms. If β_{ij} is positive, the term is synergistic, while if it is negative it is called antagonistic.

For constructing an ANOVA table, the usual formulas apply. That is

$$SS_{Total} = \sum (y_i - \bar{y})^2$$
$$SS_{Reg} = \sum (\hat{y}_i - \bar{y})^2$$
$$SS_{Error} = (y_i - \hat{y}_i)^2.$$

Of course, the first has $n - 1$ degrees of freedom, the second $p - 1$, and the last $n - p$. The F statistic is, as usual,

$$F = \frac{SS_{Reg}/(p-1)}{SS_{Error}/(n-p)},$$

, while

$$R^2 = \frac{SS_{Reg}}{SS_{Total}},$$

and

$$R^2_{Adjust} = 1 - \frac{SS_{Error}/(n-p)}{SS_{Total}/(n-1)},$$

Let's look at a simple, though real, example courtesy of Lynne Hare. First, the data.

Stearine	Oil	SFI-50F
1	0	14.7
2/3	1/3	17.5
1/3	2/3	24.0
0	1	35.5

First, let's fit a linear model $E(y) = \beta_1 x_1 + \beta_2 x_2$. The results are

ANOVA

Source	df	SS	MS	F
Total	3	256.37	85.46	
Model	1	237.36	237.36	24.96
Residual	2	19.01	9.51	

$$R^2 = .926$$
$$R_A^2 = .889$$

Now we will try a quadratic model $E(y) = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$. The results are

ANOVA

Source	df	SS	MS	F
Total	3	256.37	85.46	
Model	2	256.27	128.14	1281.4
Residual	1	0.10	0.10	

$$R^2 = .9996$$
$$R_A^2 = .9988$$
$$\text{Additional SS due to curvature} = 18.91$$
$$F = 189.1$$
$$F_{.05}(1, 1) = 161.4$$

There are a number of standard mixture designs. The first set are known as the Simplex-Lattice $\{q, m\}$ designs, due to Scheffé. Here we have q components with $m+1$ equally spaced values from 0 to 1. The Simplex-Lattice Design includes every possible combination of these (remembering that $\sum x_i = 1$). The total number of points is then

$$\binom{m+q-1}{m}.$$

For example, here are three 3-component examples: $\{3, 2\}$, $\{3, 3\}$, and $\{3, 4\}$. *Draw picture*.

Another popular design is the Simplex-Centroid, also due to Scheffé. This will consist of $2^q - 1$ points. There are

- q vertices of the form $(1, 0, \ldots, 0)$
- $\binom{q}{2}$ points of the form $(1/2, 1/2, 0, \ldots, 0)$
-
-
- $\binom{q}{r}$ points of the form $(1/r, 1/r, \ldots, 0)$
- 1 point of the form $(1/q, \ldots, 1/q)$.

Here's the picture for the 4-component example.

For testing lack-of-fit, a popular choice of design is a simplex-centroid with q added interior points of the form $((q+1)/2q, 1/2q, 1/2q, \ldots, 1/2q)$. These points are sometimes called "axial check points" Here's a picture of a 3-component, 10 run design.

As an example, consider measurements of SFI-50 resulting from blends of stearine, vegetable oil solids.

Run	Stearine	Veg. Oil	Veg. Oil Solids	SFI-50 F
1	1	0	0	4.6
2	0	1	0	35.5
3	0	0	1	55.5
4	1/2	1/2	0	14.4
5	1/2	0	1/2	25.7
6	0	1/2	1/2	46.1
7	1/3	1/3	1/3	27.4
8	2/3	1/6	1/6	14.5
9	1/6	2/3	1/6	32.0
10	1/6	1/6	2/3	42.5

With this data, the x_2x_3 term is insignificant, giving a model of

$$y = 4.6x_1 + 35.9x_2 + 56.0x_3 - 21.5x_1x_2 - 16.6x_1x_3.$$

With this model, we get $R_A^2 = .9981$, and the ANOVA table looks like

ANOVA

Source	df	SS	MS
Total	9	2250.58	
Model	4	2248.02	562.00
Residual	5	2.57	.51

The additional SS due to x_1x_2 and x_1x_3 is 37.06, giving an F-statistic of 36.33 and a p-value of 0.0011.

Although a program like Design-Expert will analyze mixture designs with aplomb, analyzing the results with SAS is a little tricky. See the code below; the first proc glm will calculate the estimates correctly, but not the ANOVA, F-test, R^2, and so on (because the no-intercept option does not adjust the SS for the overall mean); the second proc glm will calculate the ANOVA correctly, but not the estimates of the linear terms, but the additional code will calculate the correct linear estimates. The data concerns an experiment with fruit juice. One final comment about data analysis: any residual checking should be done with Studentized residuals, because points in mixture designs can have substantial differences in their leverage values. (Recall that the Studentized residual r_i is given by

$$r_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}},$$

where h_{ii} comes from the hat matrix $H = X(X'X)^{-1}X'$.

```
options ls=76 ps=62 ;
data;
input dpoint x1 x2 x3 y1 y2 y3;
array ys{3} y1-y3;
do i=1 to 3;
y = ys{i};
output;
end;
keep x1 x2 x3 y;
cards;
1 1 0 0 4.3 4.7 4.8
2 .5 .5 0 6.3 5.8 6.1
3 0 1 0 6.5 6.2 6.1
5 0 0 1 6.9 7 7.4
6 .5 0 .5 6.1 6.5 5.9
7 .34 .33 .33 6 5.8 6.4
8 .72 .14 .14 5.4 5.8 6.6
9 .14 .57 .29 5.7 5 5.6
10 .14 .29 .57 5.2 6.4 6.4
run;

proc glm;
model y=x1 x2 x3 x1*x2 x1*x3 x2*x3 / noint;
run;

proc glm;
model y=x1 x2 x1*x2 x1*x3 x2*x3;
estimate 'beta1' intercept 1 x1 1;
estimate 'beta2' intercept 1 x2 1;
estimate 'beta3' intercept 1;
run;
```

One graphical approach to understanding a mixture experiment analysis, besides contour plotting, is to look at response trace plots. These plot the estimated response along the line from a vertex, through the centroid, to the opposite edge. One can look at the response trace lines for many variables at once. Any that are nearly flat indicate inactive components.

For screening designs, Snee and Marquardt (1976) recommend the following design. For q components, take

- q pure components,
- q interior points, half-way between the vertices and the centroid,
- 1 centroid, and
- q endpoints—all permutations of $(0, 1/(q-1), \ldots, 1/(q-1))$,

giving $3q + 1$ points altogether. The results can be analyzed by looking at the response trace plots.

In many mixture situations, there will be constraints. In this case, the entire simplex cannot be used, and the feasible region will be some polytope. One approach is called the extreme vertex design. All of the vertices of the polytope are used, as well as the centroid of the region, and possibly centroids of the various edges, faces, and so on.

One generally better approach is to use a D-optimal approach. D-optimality means minimizing the determinant of the $X'X$ matrix. We will discuss it in depth next class. The XVERT algorithm of Snee and Marquardt (1974) is the basis for many computer implementations of this approach.

Another good approach is the so-called Distance-Based Design. In this case the algorithm picks points that are spread out uniformly in the feasible region.

A final important aspect of constrained components is the pseudocomponent approach. We define new components by

$$x_i' = \frac{x_i - a_i}{1 - \sum a_i},$$

where a_i is the lower bound of x_i. If the feasible region has only lower bounds, the result is a new, full-size, simplex. If upper bounds are present as well, the new region will at least be simpler than the old one.

If there are process variables (e.g., temperature, cooking time) involved besides mixture variables, there are two usual approaches. The first is to transform the q mixture variables into $q - 1$ independent variables and then proceed in the usual way. One typically does this by using ratios of components. That is $r_1 = x_1/x_3$, $r_2 = x_2/x_3$. The second approach is to directly model the mixture components. The idea is to do a mixture experiment at each point of a factorial design. In any case, the matter is tricky: there are many terms in the model, the variances and covariances of the coefficients will be large, and the interpretation of significant terms can be unclear.

Future research in mixture designs will need to include

- Blocking
- Process Variable Problem
- “Best Design Criteria”

- Other Model Forms
- Education
- Graphical Display

Reference: J. A. Cornell (1990), *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data,* 2nd edition, John Wiley & Sons, New York.