Interpreting Power in Mixture DOE – Simplified

By Patrick J. Whitcomb, Stat-Ease, Inc., Minneapolis

(abridged by Mark Anderson mark@statease.com)

A Frequently Asked Question (FAQ):

"I evaluated my planned mixture experiment and found that it had very low power.

Could you tell me what is wrong with the design?"

We hear this often from statistically-savvy formulators who set up their experiments with software that calculates power. In most cases, there really is no problem with the mixture design. The lack of power stems from the nature of the beast – constraints in the ingredients imposed by the formulator to make feasible and reasonable mixtures.

To keep this answer short and sweet, we will assume that:

1. You are knowledgeable about mixture design. If not, take a look at a primer we posted at www.statease.com/pubs/mixdoe.pdf. It details a fun experiment that Pat and I did on home made play-putty.
2. You have some understanding of the concepts behind statistical power. For background on this subject, read the introduction to Oehlert and Whitcombs' "Sizing Fixed Effects for Computing Power in Experimental Designs" at www.statease.com/pubs/power.pdf.

A Quick Answer by Example

Table 1 lays out a prototypical mixture problem that leads to the FAQ noted above. It involves three components (labeled A, B and C) that sum to a total of 1 (or 100 percent). Let's assume that C catalyzes a reaction between A and B. Notice that it (component C) is constrained to 0.1, presumably because it is very potent, but the other two ingredients (A and B) are allowed to range completely from 0 to 1. (A very similar situation came up while making the play putty mentioned earlier – a small amount of borax catalyzes a reaction in white glue diluted by water.)

Table 1: A mixture problem with one component highly-constrained

0.0	≤	A	≤	1.0
0.0	≤	B	≤	1.0
0.0	≤	C	≤	0.1
Total = 1.0				

Figure 1 shows the design space within the triangular mixture coordinates. By constraining component C to such a narrow range, only a sliver of the possible mixture space remains valid.

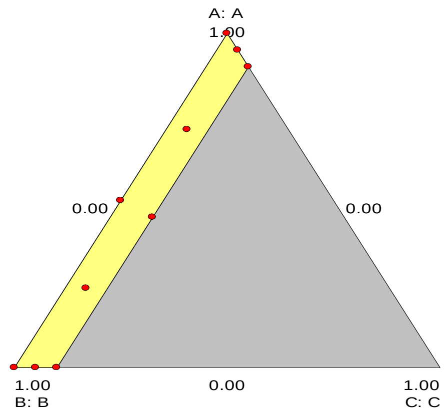

Figure 1: The feasible region

The dots in Figure 1 represent blends for a proposed design of experiments that provides a reasonable coverage of the feasible region. Table 2 shows the power statistics for this 10-point mixture design.

Table 2: Power statistics (simplified)

Term	R_i^2	Power (2σ effect)
A	0.672	5.0 %
B	0.672	5.0 %
C	0.999	5.0 %
AB	0.393	47.7 %
AC	0.999	5.0 %
BC	0.999	5.0 %

For all the model terms except AB, the power for an anticipated effect that's twice as large as the experimental error (σ - sigma) is at its minimum level – the specified significance threshold (α - alpha) of 5 percent. The lack of power stems from the tremendous collinearity (R_i^2 near 1) caused by component C being so severely constrained, which squeezes all the design points onto a relatively narrow track. The one exception, AB, exhibits an appreciable power because it does not depend much on what happens with C, thus its collinearity (R_i^2 = 0.393), although not perfect (R_i^2 of 0), remains reasonable.

The solution to the low power observed in Table 2 seems obvious: Simply replicate the mixture design.

Sidebar: A Two-Sigma (2σ) Effect Desired in Play Putty

Consider making your own play putty, for which you wish to generate a delightful 'bounce.' For the sake of discussion, let's assume that when comparing alternate putty formulations a difference less than 3 inches will not generate any delight – in other words, this is the minimum threshold for a practical effect. Prior experimentation reveals a standard deviation of 1.5 inches. Thus, the effect for which power must be evaluated will be two (3/1.5) standard deviations (2 σ).

Let's put this to the test with our example by duplicating each of the ten points three times to produce 30 proposed blends. Table 3 shows the resulting evaluation for power, which may surprise you.

Table 3: Power after replicating the mixture design three times

Term	Power (2σ effect)
A	5.0 %
B	5.0 %
C	5.0 %
AB	98.6 %
AC	5.0 %
BC	5.0 %

As you can see, this does no good other than increasing the power for the AB term.

Why does replication fail to help much in a case like this with one component severely constrained relative to the remaining formulation? To shed light on this perplexing question, we generated the 3D surfaces in Figures 2a and b. The actual responses came from equation (1).

$$\hat{y} = 4X_1 + 4X_2 + 8X_3 + 16X_1X_3 \quad (1)$$

This equation, called a "Scheffé polynomial", is laid out in terms of where X_1, X_2 and X_3, which are the mathematical symbols for mixture components A, B and C; respectively. The second-order term X_1X_3 makes this a "quadratic" equation. The surface in Figure 2a falls within the feasible region – component C restricted to a maximum of 0.1; whereas the one shown in 2b uses equation 1 for extrapolation of component C to its maximum value of 1 (or 100 percent). The spheres in Figure 2b represent design points in this "what-if" scenario of the constraint being removed from C.

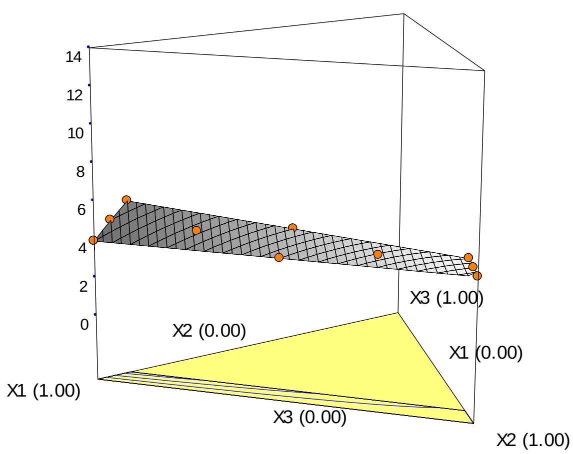

Figure 2a: 3D surface – feasible region

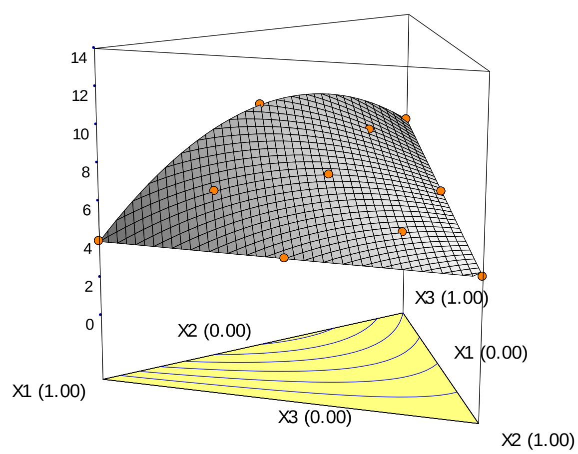

Figure 2b: Extrapolated surface

To illustrate why it's so hard to achieve power in this case, we restricted the view to only the X_1-X_3 (A-C) edge of these 3D surfaces in Figure 3. The bottom axis is labeled "C", but keep in mind that component A makes up the difference to 100 percent on this graph of response.

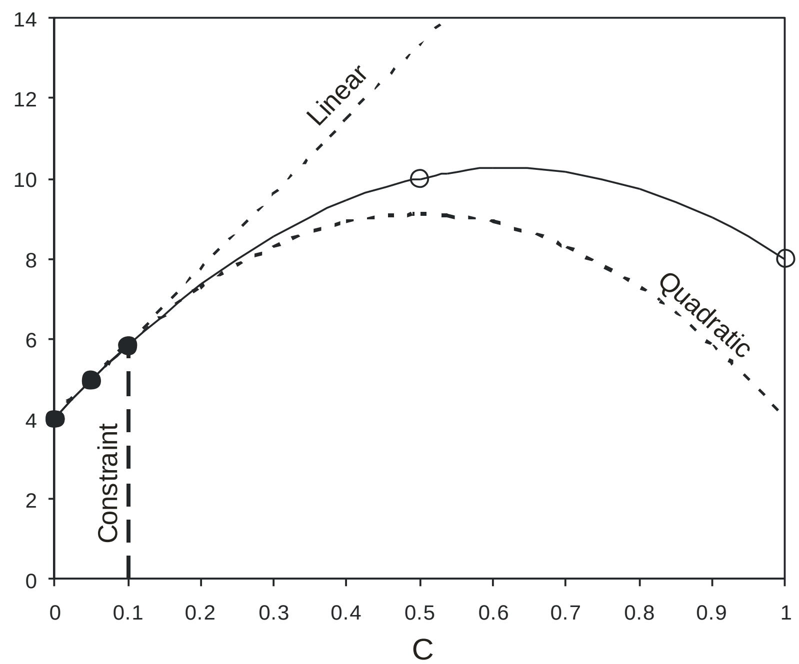

Figure 3: Viewing only the A-C edge

The solid line on Figure 3 comes from equation 1 – our specified response model, with two open circles representing the design points shown in Figure 3b (the extrapolation). The top dotted line illustrates the pure linear effect. The bottom dotted line represents a pure quadratic effect. Note how this curve forms a perfect ‘rainbow’: If you connect the end-points, there will be no slope.

You can now see that these two extremes exhibit very different behavior over the entire span of component C from 0 to 1, but they cannot be distinguished over the constrained region ($C \leq 0.1$).

Herein lies the problem; either a pure linear model, pure quadratic or infinite combinations of the two will adequately represent the response behavior within the constrained region. Therefore the ability to associate response behavior with individual terms is almost nonexistent, so power is at its minimum no matter how many runs or replicates are made.

Don’t Despair! Despite Low Power, Mixture Model Still Useful

Despite the demonstrated lack of power on individual terms, models of responses from mixture may still prove to be quite useful, more so as the number of blends in the experiment increase. In Figures 4a and b, notice how much replication lowers the standard error (SE) for prediction.

> *“Even the most stupid of men, by some instinct of nature, by himself and without any instruction, is convinced that the more observations have been made, the less danger there is of wandering from one’s goal.”*
>
> *Jacob Bernoulli (1654-1705), The ‘Father of Uncertainty’*

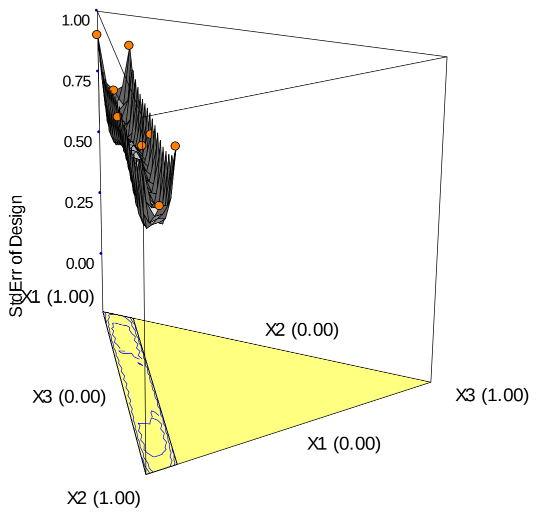

Figure 4a: Unreplicated Mixture design

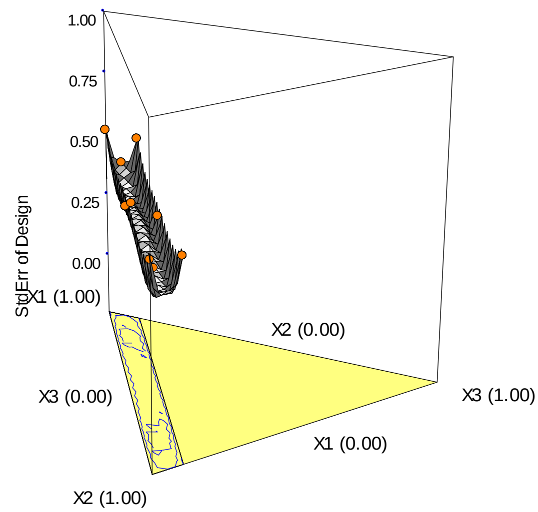

Figure 4b: Three replicates

The maximum SE falls from 0.9 to 0.52 from left (Figure 4a) to right (4b). This translates to a 42 percent improvement in precision of prediction as a result of the replication. Common sense prevails – the more you invest in a design, the more precisely you can predict response!

Sidebar: How Would Replication Help For An Experiment On Play Putty?

Recall from the earlier sidebar that the standard deviation for formulating play putty and measuring its bounce is 1.5 inches. This provides a context for the reduction in standard error (SE) seen in Figure 4 due to replication, which are evaluated with an assumed standard deviation of 1. For the play-putty example, the actual reduction in maximum SE for prediction becomes 1.35 inches (0.9 * 1.5) to 0.78 inches (0.52 * 1.5) – still a 42 percent improvement in precision.

But is really necessary to replicate a mixture experiment like this? For play putty, the answer was "No"! The design, run unreplicated, produced significant effects for the bounce response, which you can in a report posted at www.statease.com/news/news0212.pdf. More details can be found at www.statease.com/playputty.html.

Summary

The goal of mixture design is to generate a map of the response over a specified region of formulation. As the design space becomes more constrained, thus producing increased collinearity among the model terms, power will decrease. However, if statistical analysis shows that the actual experiment produced a significant model overall, then the map will be useful and improvements are sure to follow.