

Experimental design

Food Formulation,
Dr. Ali Nasirpour

What is the objective of the study?

- A. Screen out insignificant factors and identify significant factors. Get some idea about the existence of interaction effects. Use a Factorial Design that is color-coded White, Green, or Yellow.
- B. Understand main effects and get complete information about 2 factor interactions. Use a Factorial Design that is color-coded White or Green.
- C. Quickly trouble-shoot a process and get immediate results. You will have to go back later and do follow-up experimentation to understand why the conditions work. Use a Factorial Design that is color-coded Yellow or Red.
- D. Test a product, process, or method for ruggedness. Perhaps verify that a test method works for all operators or verify that small variations in design don't adversely affect the product. The idea for all of these is to run a minimal number of runs, testing all variables at once, and you hope to find nothing significant. If no factors are significant, then the changes you made don't significantly affect the product, process, or method. If something IS significant, then you can't be certain which factor it is, but you know that follow-up experimentation will be necessary. Use a Factorial Design color-coded Yellow or Red. Other options are Plackett-Burman or Taguchi designs.

E. Characterize how the significant factors affect your responses (for optimization purposes.) Use a Response Surface Design. Use Central Composite to study each factor at 5 levels, use Box Behnken to study each factor at 3 levels. D-Optimal will create a custom design based on the type of polynomial that you want to fit.

F. Formulation work. If your product is actually a mixture, or a formulation, such as a food product, a drug formulation, or a chemical composition, then you should use a Mixture Design. These designs allow you to set a total amount for the mixture. Then each component range is constrained by the fact that as you increase the amount of one component, the other components must decrease in order to keep the total amount constant. Mixture D-optimal designs are most commonly used because they allow the most flexibility in your component ranges.

Identify Factors and Levels

During the brainstorming session, make a list of all the factors that could possibly influence any of the responses you want to measure. Don't try to evaluate factors as they are listed, just write them all down. It is likely that your list may end up with 30 or more factors!

Next, group the factors into categories such as – most likely, somewhat likely and least likely to influence the responses. Also note if the factor is easily controllable or difficult to control. From this list, decide how many and which factors you want to include in the experiment. For the other factors, decide if they will be held constant, monitored, or ignored.

Remember, it is not necessary to include absolutely everything in one giant experiment. Smaller, sequential experiments teach you about the process quickly and generally cost less.

Once you have determined the factors to study, you must decide on a range for each one. The range should be large enough so that you expect to see a difference in the response(s), but not so large that you "fall off the cliff" and don't get any measurable results. Ideally, the experiment will produce some good and some bad product so you can learn what causes the difference.

Determine Number of Runs

How many runs are necessary to get significant results? This depends on how many factors you have, how much information you want (how many effects you want to estimate) and how much noise or process variation is in the system you'll be experimenting on. The more noise in the system, the more runs you'll need to do.

Factorial Design Selection

Factorial designs are used primarily for screening significant factors, but can also be used sequentially to model and refine a process.

- 2 Level Factorial designs (2-15 factors) – Full and fractional designs are available to explore many factors, setting each factor to only two levels. Fractional factorials are efficient designs used to screen many factors to find the few that are significant. Full factorials can be used to estimate the effects of all interactions.
- Irregular Fraction designs (4-9 factors) – These are a special set of Resolution V designs, usually with fewer runs than a normal fractional factorial, that allow clean estimation of the main effects and two-factor interactions.
- Taguchi designs (up to 63 factors) – A set of classic designs from Taguchi teachings. These may be used as a base to build a particular design. Note that all analyses will be completed using standardized ANOVA reports and interaction graphs.

Uses of RSM

- To determine the factor levels that will simultaneously satisfy a set of desired specifications
- To determine the optimum combination of factors that yield a desired response and describes the response near the optimum
- To determine how a specific response is affected by changes in the level of the factors over the specified levels of interest

Limitations to RSM

- Large variations in the factors can be misleading
- Critical factors may not be correctly defined or specified
- Range of levels of factors too narrow or too wide --optimum can not be defined
- Lack of use of good statistical principles

Introduction to Response Surface Methodology

- Response Surface Methodology (RSM) is useful for the modeling and analysis of programs in which a response of interest is influenced by several variables and the objective is to **optimize** this response.
- For example: Find the levels of temperature (x_1) and pressure (x_2) to maximize the yield (y) of a process.

$$y = f(x_1, x_2) + \varepsilon$$

- Response surface:

$$\eta = E(y) = f(x_1, x_2)$$

- The function f is unknown
- Approximate the true relationship between y and the independent variables by the lower-order polynomial model.

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$$

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j + \varepsilon$$

Response Surface Design Selection

Response surface method (RSM) designs help you quantify the relationships between one or more measured responses and the vital input factors. If you have 5 or more factors, consider doing a two-level factorial screening design first. At least some of the factors for RSM must be quantitative, continuous variables. The objective is to find a desirable location in the design space. This could be a maximum, a minimum or an area where the response is stable over a range of the factors. Goals might include meeting a set of specifications for several responses simultaneously.

RSM Design Points Needed

You may want to determine for yourself how many design points are needed for a good response surface design. The response surface method produces a mathematical model that you can use to predict a response. You should provide the following design features to build a good model:

1. Enough unique design points to estimate all the terms in the postulated model: linear, 2FI, quadratic or cubic. The number of model terms increases in proportion to the number of factors studied.
2. Extra unique design points, above what's needed for estimating the model and pure error to test how the model fits the data. These points must be at locations in the design space that are different from the model points. They are used in a "Lack of Fit" test for the model. You should specify at least four of these extra points to give an adequate statistical test.

3. Replicates of some design points to estimate the experimental, or pure, error. This is the error to be expected in the response if the experiment is repeated starting from scratch. Typically, the center point of the design is repeated, often four or more times. This gives an adequate estimate of the variation of the response and provides the number of degrees of freedom needed for an adequate statistical test of the model. You may choose to duplicate other points in the design if you desire better estimates of the response at those areas in the experimental space.

As a guide, here is a table of the number of coefficients in linear, quadratic, and cubic equations for the given number of factors.

# FACTORS	LINEAR	QUADRATIC	CUBIC
2	3	6	10
3	4	10	20
4	5	15	35
5	6	21	56
6	7	28	84
7	8	36	120

In summary, to ensure success in RSM modeling, you should allow for sufficient model points, plus at least 4 replicate points, plus at least 4 other points to determine the fit of the model.

Factor Coding in Response Surface Designs

Design descriptions and analyses are best done with coded factors. Coding reduces the range of each factor to a common scale, -1 to +1, regardless of its relative magnitude. Scaling establishes factor levels that can be orthogonal (or nearly so). Also, it is easier to think in terms of changes from low to high for the factors than to think about their actual values. For example, one factor may vary from 100 to 200 while another varies from 0.1 to 0.5. Typical coding has -1 as the lower level of a factor, +1 as the upper level, and 0 as the middle level.

RSM Model Terms

The degree of the model indicates the presence of certain terms.

Degree	Description
1	Linear - A, B, C, ...
2	Quadratic - AB, AC, BC, A ² , B ² , C ² ...
3	Cubic - ABC, A ² B, A ² C, AB ² , A ³ , B ³ , C ³ ,...

RSM D-Optimal Design

The D-optimal criteria, one of several "alphabetic" optimalities, was developed to select design points in a way that minimizes the variance associated with the estimates of specified model coefficients.

Here's how you set up a D-optimal design in Design-Expert.

1. Select the number of factors and levels (add constraints if needed).
2. Select the model that you want to fit with the Edit Model button (Quadratic is default).
3. DESIGN-EXPERT creates overall candidate point set depending on model chosen (Create Candidate Points button).
4. DESIGN-EXPERT chooses design points, including replicates.

You may want to check the quality of the design by using the Design Evaluation.

Candidate Points

Mixture Design Selection

Choose a mixture design if your experiment meets these two criteria:

1. The components add to a fixed total. For instance, A is 10% of the mixture, B is 30%, and C is the remaining 60%. If the percentage of one component is increased, then the percentage of one or more of the other components must be decreased. If your component amounts do NOT depend on each other, then you can experiment using response surface designs rather than mixture designs.
2. The response must be a function of the proportions of the components. For example, the flavor of ice cream depends on the relative proportions of the ingredients, not on the total amount of ice cream. If your response is not related to the ingredient proportions, then you should use Response Surface designs.

There are several mixture design choices. The simplex designs can be used whenever the components form a simplex region. For non-simplex spaces choose a d-optimal design.

On the first screen in the mixture design process, enter the names of the components, their allowed low and high levels, the total amount of the mixture, and the units the total is measured in. The low and high entries are constraints on a single component (For example, there are legal requirements on the level of fat in ice cream). To enter multi-component linear constraints (example $A+B<10$), click on the Edit Constraints button.

If you choose low and high levels such that the component ranges are not the same size, or you add any linear constraints, your design space will not form a simplex; therefore, you will not be permitted to build a Simplex Lattice or Simplex Centroid design. There are several other designs that can handle a non-simplex space, but d-optimal is the best choice.

Components will be assigned an alphabetical character. The limits on any component will range from 0% - not in the mixture, to 100% - only this ingredient is present, a pure component. The total of all of the ingredients must be 100% or less. The total may be less than 100% if the components form a constant portion in a larger mix.

The scale used by default is based on fractions, from 0 to 1. If you want to enter percent, then enter the limits and the total in percent, from 0 to 100.

Mixture Model Terms

The degree of a mixture model indicates the presence of certain terms. The following are in the Scheffe model form.

Degree	Description
1	Linear - A, B, C, ...
2	Quadratic - AB, AC, BC ...
3	Special Cubic - ABC, ...
3	Cubic - AB(A - B), AC(A - C), ...

Mix D-Optimal Design

The points selected are model dependent and are selected with only this criterion in mind. Therefore, for an adequate design the D-Optimal points should be augmented to provide for estimates of pure error by replication, and for determining the lack of fit using excess design points. After you build the design, you can check its adequacy by going to the Design Evaluation screen.