First Steps in Experimental Design—
The Screening Experiment

John A. Wass

“Statistical Viewpoint” addresses principles of statistics useful to practitioners in compliance and validation. We intend to present these concepts in a meaningful way so as to enable their application in daily work situations.

Reader comments, questions, and suggestions are needed to help us fulfill our objective for this column. Please send your comments to coordinating editor Susan Haigney at shaigney@advanstar.com.

KEY POINTS
The following key points are discussed in this article:

• Design of experiments (DOE) consists of three basic stages: screening (to identify important factors), response surface methodology (to define the optimal space), and model validation (to confirm predictions)
• A critical preliminary step in the screening stage is for subject matter experts to identify the key list of factors that may influence the process
• A DOE design consists of a table with rows that represent experimental trials and columns (vectors) that give the corresponding factor levels. In a DOE analysis, the factor level columns are used to estimate the corresponding factor main effects
• Interaction columns in a design are formed as the “dot” product of two other columns. In a DOE analysis, the interaction columns are used to estimate the corresponding interaction effects
• When two design columns are identical, the corresponding factors or interactions are aliased and their corresponding effects cannot be distinguished
• A desirable feature of a screening design is orthogonality in which the vector products of any two main effect or interaction columns sum to zero. Orthogonality means that all estimates can be obtained independently of one another
• DOE software provides efficient screening designs with columns that are not aliased and from which orthogonal estimates can be obtained
• Full-factorial designs include all combinations of factor levels and provide a predictive model that includes main effects and all possible interactions
• Fractional factorial (screening) designs include fewer trials and may be more efficient than the corresponding full factorial design
• The concept of aliasing is one of the tools that can be used to construct efficient, orthogonal, screening designs
• Center points are often included in screening designs to raise the efficiency and to assess lack of model fit due to curvature
• The order of running and testing experimental trials is often randomized to protect against the presence of unknown lurking variables
• Blocking variables (such as day or run or session) may be included in a design to raise the design efficiency

ABOUT THE AUTHOR
John A. Wass, Ph.D., is a consulting statistician with Quantum Cats Consulting in the Chicago area, as well as a contributing editor at Scientific Computing and administrator of a regional statistical software group. He may be reached by e-mail at john.wass@tds.net.
• Factor effects in screening designs may be missed because they were not included in the screening experiment, because they were not given sufficiently wide factor ranges, because the design was underpowered for those factors, because trial order was not properly randomized or blocked, or because of an inadequate model.

INTRODUCTION
In days of old (i.e., the author’s undergraduate years), we were introduced to the joys of manual calculations and analysis of variance (ANOVA). Experimenters would change one factor at a time and identify what they felt were “optimal” processing conditions. With the advent of personal computers and the dissemination of more efficient techniques by statisticians, problems of increasing complexity were solved. This not only enlightened the basic researcher, but permitted scientists and engineers to design more robust products and processes. In fact, the statistical design of experiments (DOE) has been called the most cost-effective quality and productivity optimization method known. In this brief introduction, we will concentrate on practical aspects and keep mathematical theory to a minimum.

The advent of DOE brought a modicum of order to the wild west of one-factor-at-a-time changes. The technique has many variations but consists of the following three basic stages:

• Screening—to exclude extraneous effects considered as noise
• Response surface methodology—to finely define the optimal result space
• Model validation—to confirm predictions.

Each is quite important. In this paper we will concentrate on the first stage, screening design and analysis.

There are many commercial software packages, either standalone or modules, within general statistics programs that will support DOE. Some of these are listed later in this discussion. Each has its unique strengths and weaknesses. For this paper, JMP® has been used. The principles will be the same for most programs; although, the user interfaces, algorithms, and output will vary.

THEORY
The literature of DOE is replete with names such as full factorial, fractional factorial, runs, power, levels, and interactions. In addition, we have categorical and continuous factors and a variety of design names. Fortunately, in screening we usually confine ourselves to the fractional factorial designs. Unfortunately, as with everything in real-life, there is a price to pay for every extra bit of information required.

We can show an experimental design as a table. An example is presented in Figure 1. Each row in the table corresponds to an experimental trial. The table columns indicate the levels of the experimental factors. For screening designs we usually consider only two levels, usually coded +/-1. In this notation, + represents “high” and – represents “low.” We may also include other columns that indicate interactions among the factors. The columns giving the experimental factor levels permit us to estimate “main effects” and the interaction columns permit us to estimate “interaction effects.” We will say more about main effects and interaction effects below. In addition to factor level and interaction columns, we may record one or more columns of measured variable values that result from each trial. We refer to these dependent variables as the experimental responses.

Screening designs are useful as they are a practical compromise between cost and information. Their main contribution is in suggesting which of many factors that may impact a result are actually the most important. Because screening designs require fewer runs, they are far less costly than the more informative full-factorial designs where the practitioner uses all combinations of factor levels. It has been suggested that no more than 25% of the total budget for DOE be spent on the screening runs. Screening runs are usually a prelude to further experimentation, namely the response surface and confirmatory runs, where specific information is gained around target (desired) outcomes.

Key Assumption For Screening Studies
In screening designs, we make the assumption that our real-world processes are driven by only a few factors, the others being relatively unimportant. This usually works quite well but it is a crucial assumption that requires careful consideration by subject matter experts. Also keep in mind that the fractional factorial designs may be upgraded to full factorial designs (main effects plus all interactions) if there are only a few main effects. This allows us to observe interactions at a reasonable cost.

Number Of Runs
With screening designs, responses are taken only for a small fraction of the total possible combinations to reduce the number of runs and thus cost. The total number of runs is calculated by raising the number of levels to the power of the number of factors (e.g., for three factors at two levels each we have runs = 2^3 = 2x2x2 = 8). This is actually a full factorial design as we are testing...
all combinations of factor levels. Full factorial designs allow us to build predictive models that include the main
effects of each factor as well as interactions. This brings
us to three important concepts of these models: interac-
tion (the effects of one factor on another), orthogonality
(all factors are independent of one another), and alias-
ing (when the effects due to multiple factors cannot be
distinguished).

**Interactions**

One of the more important things that practitioners need
to know about is that main factors may affect each other
in ways known and unknown (i.e., interaction among
effects). For example, the interaction of two reagents in a
chemical process may be a significant driver of the overall
process (think enzyme and substrate). In deciding which
are important, statistically and physically, it is necessary
to consult with the bench scientists and technicians to
to get a handle on what is already known and suspected to
be important to the process. Too few factors risk missing
something important. Including too many factors
will render the screening more costly and lead to a lack
of orthogonality due to aliasing. Aliasing occurs when
two columns in the design (referred to by statisticians
as a vector of the input space) are identical or when one
column is identical to another formed from the interac-
tion of two columns (i.e., the “vector” or “dot” product
of two columns).

Figure 1 presents a design table for a full factorial design
in three factors. This design requires eight trials (rows) and
has three factors (A, B, C) for which we can estimate main
effects. It is also possible to estimate all possible two-way
interactions with this design (AB, AC, BC) as well as the
single three-way interaction ABC that is shown in the
design table. If later, the design is augmented with a fourth
factor D (all runs not shown below), we have a problem.
Now the contribution to any measured outcome (effect)
from ABC is indistinguishable from D and, therefore, we
do not know if the driver was D or ABC.

The A, B, and C columns give the levels (coded +/-) of
eight experimental design factors. The ABC interaction
column is formed as the “dot” product of the A, B, and
C columns. Notice how each row in the ABC column is
formed as a product of the corresponding levels of A, B,
and C. In the analysis of such a design, the ABC interac-
tion column would be used to estimate the corresponding
ABC interaction effect.

The sums of the levels of the A, B, C, and ABC columns
all equal zero. This means that the design is balanced.
Balanced designs give more precise estimates of main and
interaction effects than unbalanced designs.

**Figure 1: Aliasing in a simple design.**

<table>
<thead>
<tr>
<th>Run</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>ABC</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Sum</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Orthogonality**

Further, the “dot” product of any two of the columns A,
B, C, or ABC will also sum to zero (try it and see). This
more subtle design characteristic is called “orthogonality”
and is critical to good experimental design. To under-
stand why orthogonality is so important, we return to
our concept of aliasing. Aliasing is the extreme absence
of orthogonality. It is impossible to separately estimate
the effects corresponding to two design columns that are
aliased. In a sense, such estimates are 100% correlated
(the statistical term is “confounded”). In contrast, when
two design columns are orthogonal, the corresponding
effect estimates have zero correlation. This means that
errors in estimating one effect do not, on average, bias
our estimate of the other effect. Orthogonal designs
prevent us from accidentally confusing the effects of
two different factors.

DOE software provides us with orthogonal (or nearly
orthogonal) screening designs in which the main effect
and interaction columns are not aliased. These allow us
to estimate the corresponding effects without worrying
about possible correlations. This lack of correlation usu-
ally implies that the estimates are independent.

Figure 2 gives a good mental image of the value of
orthogonality and the independence it provides in our
estimates. Let the three mutually perpendicular (orthog-
onal) axes X, Y, and Z represent dimensions along which
three estimates from some experimental design may lie.
The results of an experiment may then be indicated as a
point in that three dimensional space. If we repeat the
experiment many times, the resulting estimates will form
a cluster of points in space, centered about the “true”
effects being estimated. With orthogonal designs, the
cluster of points will be spherical in shape indicating a
lack of correlation (or independence) among the esti-
mates. In designs with aliasing, the points will fall along
a one-dimensional line or two-dimensional plane that
indicate a complete correlation of three or two effects,

---

**Orthogonality**

Further, the “dot” product of any two of the columns A,
B, C, or ABC will also sum to zero (try it and see). This
more subtle design characteristic is called “orthogonality”
and is critical to good experimental design. To under-
stand why orthogonality is so important, we return to
our concept of aliasing. Aliasing is the extreme absence
of orthogonality. It is impossible to separately estimate
the effects corresponding to two design columns that are
aliased. In a sense, such estimates are 100% correlated
(the statistical term is “confounded”). In contrast, when
two design columns are orthogonal, the corresponding
effect estimates have zero correlation. This means that
errors in estimating one effect do not, on average, bias
our estimate of the other effect. Orthogonal designs
prevent us from accidentally confusing the effects of
two different factors.

DOE software provides us with orthogonal (or nearly
orthogonal) screening designs in which the main effect
and interaction columns are not aliased. These allow us
to estimate the corresponding effects without worrying
about possible correlations. This lack of correlation usu-
ally implies that the estimates are independent.

Figure 2 gives a good mental image of the value of
orthogonality and the independence it provides in our
estimates. Let the three mutually perpendicular (orthog-
onal) axes X, Y, and Z represent dimensions along which
three estimates from some experimental design may lie.
The results of an experiment may then be indicated as a
point in that three dimensional space. If we repeat the
experiment many times, the resulting estimates will form
a cluster of points in space, centered about the “true”
effects being estimated. With orthogonal designs, the
cluster of points will be spherical in shape indicating a
lack of correlation (or independence) among the esti-
mates. In designs with aliasing, the points will fall along
a one-dimensional line or two-dimensional plane that
indicate a complete correlation of three or two effects,
respectively. In between these extremes will be designs that will produce ellipsoid-shaped clusters, whose axes are not parallel to X, Y, and Z. Such designs are less efficient than fully orthogonal designs and may result in misinterpretation of screening results.

Similarly, when the columns in our experimental design are unaliased and when the vector products of any two columns sum to zero, the corresponding effect estimates are mutually independent. This means that random errors in estimating one effect will not (on average) bias the estimate of another effect. Designing for orthogonality is excellent protection against aliasing.

**Aliasing**
The concept of aliasing is also useful in constructing efficient fractional factorial designs. The design in Figure 1 includes an ABC interaction column. In the real world, two-way interaction effects may well be present, but three-way interactions such as ABC are often considered unlikely to be important. This is referred to as the "sparcity of effects" principle. Based on our knowledge of the process, we may be willing to assume that the ABC interaction effect is not present. In that case we might consider including factor D in our experiment and setting its levels to those indicated by ABC. In this way we can estimate the main effects of all four factors, although we sacrifice our ability to learn about the ABC interaction. This is an example of using aliasing as a design tool. We indicate this aliasing mathematically by writing "ABC=D". This is not meant to imply that the effects of D are the same as the interaction of ABC, but only that our experiment cannot distinguish the main effect of D from the ABC interaction. We have now created a design with four factors in only eight trials (a full factorial design with four factors would have required $2^4=16$ trials).

As this is an introduction we will not delve into the more advanced concepts of saturation, efficiency, foldover, and the more complex screening designs. Any of the textbooks cited in the reference section may be consulted for these topics.

**GENERAL TECHNIQUES**
In most modern software it is a straightforward matter to design an experiment and analyze the results (with a little practice and much consulting of the manual). It’s the decision as to what factors to include that is critical. It is strongly advised that you not throw in “everything and the kitchen sink” for fear of missing something. This is where consultation with subject matter experts like the process engineers and bench scientists familiar with the process or product is crucial.

Figure 3 is an example of an experimental design produced by the JMP software. The software allows fast generation of an experimental sheet showing the factors and their levels, plus a column for the results.

This design has three input factors (X1-X3) and a single output (Y). The levels indicate low (-1), median (0), and high (1) levels of the factor. The trials for which all factors are at their median level are called “center points” and may represent target or control factor settings. The 000 points are the center points and are useful for testing linearity in the process. It is a simple and inexpensive way to check for curvature. We now have nine design points, each replicated twice to yield 18 runs. This design is a full factorial design in which each of the nine trials in the full factorial is replicated twice. Replication is sometimes needed to provide sufficient statistical power. These factor levels are obviously coded but the actual numbers could be entered and more factors screened against multiple outputs, if desired.

**An Example**
The medical diagnostics industry makes great use of DOE for myriad products. In this example, we will look at the design and analysis of a clinical laboratory kit pack for an unspecified analyte. We wish to determine and minimize the product variability given certain known conditions of reagent concentration and physical/environmental factors for the chemical reactions occurring in the kit pack. The input factors are as follows:

- Reagent 1
- Reagent 2
- Enzyme
- Temperature
- Mixing speed.
In this example, the two reagents are combined with the enzyme in a reaction vessel at a specified temperature. The mixture is agitated at several mixing speeds and the researcher wants to know how important each of the factors (i.e., reagents, enzyme, temp, and mix speed) is to the concentration of the final product. All are held at either low, median, or high levels. The single output (response) variable is the concentration of the product. We will minimize the variability of the concentration measurement (i.e., for the given inputs we wish to hold the measured concentration to be within a stated reference range with minimal variability). Also runs are randomized within a block, and block order may be randomized if necessary. Note that although this screen is designed for maximal efficiency and precision, only main effects order effects may be estimated. After gaining management approval for the cost, we obtain the design using JMP software. Note that the block column here indicates the session in which to perform each trial (see Figure 4). Notice that we have five factors and that a full factorial design would require $2^5$ or 32 runs yet our design only has 18. The reduction comes in an aspect of DOE called design efficiency. This is a measure of the efficiency of any given design in covering the design space, compared to a 100% efficient design that would cover all of the points needed to extract maximal information about the dependency of the results upon the input factors. Modern DOE software employs highly efficient algorithms based upon some very complex mathematics that will reduce the number of points needed while minimally impacting efficiency.

Having designed the experiment, we go into the lab and collect the data then return to the JMP design table and fill in the Y column of the table (see Figure 5). The data can be analyzed in JMP. The first step is to specify the model. The model can then be automatically generated as an effect screen using JMP’s standard least squares platform. The model contains a “list” of all the factor effects that we wish (and are able) to estimate. Our estimates will include the five main factor effects, the block effect and the two-way interaction between reagents 1 and 2 (see Figure 6).

Running this model shows us an actual vs. predicted plot for the product (see Figure 7). As the p-value from the analysis of variance shows that the model is very significant ($p<<0.01$), the $R^2$ is 92% and the estimate of the standard deviation of the process noise (RMSE) is quite low, we are initially happy with the model fit. However, Figure 8 conditions may vary from session to session. We would like to design the experiment in some way so that any session differences are not misconstrued as factor effects. So our statistician suggests the use of blocking in our design. Blocking is the arranging of experimental units (the runs) in groups, or blocks, that are similar to one another.

We block our designs to reduce random noise between sessions (in this case) as the experiment was carried out over several time periods and we expect the data within any one block to be more homogeneous than that between blocks. Thus, greater precision is obtained by doing within-block comparisons, as inter-block differences are eliminated. Also runs are randomized within a block, and block order may be randomized if necessary.
shows the parameter estimates that test the significance of the individual factors to the overall model.

Based upon the usual p-value cutoff of 0.05, only the effects of the two reagents were statistically significant. The intercept is usually not of interest in the physical sense of the model as it only gives information about the “0” point of the x-axis.

As the chemists suspected, the reagents are quite important to the product and the mix speed was not. However, the lack of an enzyme and temperature effects is surprising and may indicate a flaw in the design. Either the wrong values of those factors were chosen (the range was not sufficiently wide to see the effect), the design was under-powered for those factors (Power in this case is taken as: not enough runs to discern a difference given that one truly exists), or both. The one interaction and the block tests cannot truly be discerned (as to the true effect on product) as we have not the resolution to do so with this design. We can test the power, however, for the surprises in the main effects (see Figure 9).

These are both low and we would need more data to determine true lack of significance. For guidance on what constitutes adequate power, see the “Tips, Tricks, and Pitfalls” section. The Pareto Plot (see Figure 10) shows the results graphically.

The prediction profiler (see Figure 11) shows us how to set factors to maximize our product and the mean and 95% confidence interval for the mean response.

Our task is now to determine the best settings for the temperature and enzyme factors and to augment the design with enough data to ensure adequate power. Once this is achieved, we can zero in on maximal setting and minimal variation using response surface methodology.

Other Designs
Fractional and full factorials are not the only screening designs available. The following are among the most popular:

- **Plackett-Burman.** Resolution is higher than simple factorial and this is a good technique for large numbers of input factors. Also more flexible as the runs need not be a power of 2. All columns are balanced and pairwise orthogonal.

- **Box-Behnken.** Efficient technique for modeling quantitative three-level factors. Hold some of the factors at their centerpoints and are easily blocked.

**Figure 4:** Example problem screening design.
**Figure 5:** Test data.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Block</th>
<th>Reagent 1</th>
<th>Reagent 2</th>
<th>Enzyme</th>
<th>Temp</th>
<th>Mix Speed</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 00000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.4</td>
</tr>
<tr>
<td>2 ++++</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3 +++++</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>2.3</td>
</tr>
<tr>
<td>4 +++++</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1.8</td>
</tr>
<tr>
<td>5 +++++</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1.5</td>
</tr>
<tr>
<td>6 +++++</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1.9</td>
</tr>
<tr>
<td>7 +++++</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>8 +++++</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>9 +++++</td>
<td>2</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1.6</td>
</tr>
<tr>
<td>10 ++++</td>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1.9</td>
</tr>
<tr>
<td>11 ++++</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>2.6</td>
</tr>
<tr>
<td>12 0000</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.6</td>
</tr>
<tr>
<td>13 ++++</td>
<td>3</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>2.2</td>
</tr>
<tr>
<td>14 ++++</td>
<td>3</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1.4</td>
</tr>
<tr>
<td>15 ++++</td>
<td>4</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>16 ++++</td>
<td>4</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1.8</td>
</tr>
<tr>
<td>17 ++++</td>
<td>4</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>2.1</td>
</tr>
<tr>
<td>18 ++++</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3.1</td>
</tr>
</tbody>
</table>

**Figure 6:** Data analysis.

**Figure 7:** Actual by predicted plot.
Statistical Viewpoint.

- **Box-Wilson.** These are central composite designs that may be used when significant interaction is suspected. They have the desirable property of rotatability where the predicted response may be estimated with equal variance from any direction from the center of the design space.

**TIPS, TRICKS, AND PITFALLS**

Experimental design is an art. There are many detailed techniques and lessons learned from years of experience that novices may have to gradually learn on their own—the hard way. The following comments are meant to minimize the pain for experimenters who wish to take advantage of these modern experimental tools.

**Curvature**

First and foremost is the fact that screening is a linear process and as such does not take curvature and higher order interactions that generate curvature into account. If the model renders a poor fit, and the graph of the data appears to suggest curvature (points clustered above and below the line), the experimenter is well advised to go to a more complex model (e.g., add a center point or quadratic). Also note that screening may underestimate the true number of main factors, depending upon choices and basic knowledge of the process. Lack-of-fit should also be carefully assessed, as by its very nature screening will evidence poor lack of fit for any curved process despite the fact that the selections of main factors were complete. Bottom line: there are problems with only looking at first order effects.

**Randomization**

One of the more important concepts in all of DOE is to randomize the order of runs. This averages over noise that may be due to a specific time of run, type of chemical prep, temperature, pressure, or other condition that may be concentrated in space and time. When runs may be dependent upon long physical startup times, the experimenter may choose to block certain factors.

**Power**

Here we refer to statistical power (i.e., the probability of detecting a difference if one truly exists). Most software will perform this test and it should never be ignored. In certain cases, if the power is found to be unacceptably low, it might dictate rerunning the entire experiment with a larger number of runs. Acceptable power is determined by the experimenter or may be dictated by company or governmental rules. As a rule of thumb: for purely research purposes, 70-80% may be adequate. For process

---

**Figure 8:** Parameter estimates.

| Term            | Estimate | Std Error | t Ratio | Prob>|t| |
|-----------------|----------|-----------|---------|------|
| Intercept       | 2.00375  | 0.06      | 33.40   | <.0001* |
| Reagent 1       | 0.45     | 0.063246  | 7.12    | 0.0001* |
| Reagent 2       | 0.3125   | 0.063246  | 4.94    | 0.0011* |
| Enzyme          | 0.1375   | 0.063246  | 2.17    | 0.0614  |
| Temp            | 0.125    | 0.063246  | 1.98    | 0.0835  |
| Mix Speed       | 0.05     | 0.063246  | 0.79    | 0.4520  |
| Block[1]        | -0.00375 | 0.1       | -0.44   | 0.6710  |
| Block[2]        | -0.15375 | 0.107703  | -1.43   | 0.1513  |
| Block[3]        | 0.13625  | 0.1       | 1.36    | 0.2102  |
| Reagent 1*Reagent 2 | 0.0375   | 0.063246  | 0.59    | 0.5696  |

**Figure 9:** Statistical power.

<table>
<thead>
<tr>
<th>Enzyme Power Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
</tr>
<tr>
<td>( \alpha )</td>
</tr>
<tr>
<td>( \sigma )</td>
</tr>
<tr>
<td>( \delta )</td>
</tr>
<tr>
<td>Number</td>
</tr>
<tr>
<td>Power</td>
</tr>
</tbody>
</table>

**Temp Power Details**

<table>
<thead>
<tr>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
</tr>
<tr>
<td>( \sigma )</td>
</tr>
<tr>
<td>( \delta )</td>
</tr>
<tr>
<td>Number</td>
</tr>
<tr>
<td>Power</td>
</tr>
</tbody>
</table>

**Figure 10:** Pareto plot of transformed estimates.
development 90+% is desirable and in certain critical cases (e.g., HIV diagnostics) 98% may be the least that is acceptable.

**Aliasing**
As was mentioned, due to low number of runs and resolution in most screening designs, we encounter situations where it is difficult to assign effects unambiguously to single factors. To get around this problem the experimenter may consider the following: select different design generators, increase the number of runs, or hold one or more factors constant.

**Replication**
One of the most important concepts in statistics is replication. The use of replication provides an estimate of the magnitude of noise. An adequate number of replicates will ensure a more precise estimate of the variation about a process mean, as well as facilitating the detection of true differences. For screening, we can usually afford but a single replicate, and in some cases it may not be necessary to replicate all of the points in a design. The heavy replication of points is usually left to the response surface methodologies where we really need to get a sharper picture of the reduced design space. Replication is one of the most underappreciated necessities of statistical analyses and the one that will generate the most headaches if ignored.

**SOFTWARE**
There are numerous software products available to assist the practitioner in design and analysis of their experiments. The author has had experience with the following commercial packages:
- JMP (www.jmp.com)
- Design Expert (www.statease.com)
- MODDE (www.umetrics.com)
- Unscrambler (www.camo.no)
- Minitab (www.minitab.com)
- SYSTAT (www.systat.com)
- STATISTICA (www.statsoft.com)
- GenStat (www.vsni.co.uk).

**CONCLUSIONS**
Modern experimental design is sometimes art as well as science. It is the objective of this column to acquaint the reader with the rudiments of the screening design, introduce them to the nomenclature and supplement the learning experience with a real-world example. Furthermore, as the process is not exact, several rules-of-thumb are given to provide guidance in those situations where hard and fast rules may not be available.

**GENERAL REFERENCES**