KEYS TO SUCCESSFUL DESIGNED EXPERIMENTS

Mark J. Anderson and Shari L. Kraber

Consultants, Stat-Ease, Inc., Minneapolis, MN (e-mail: Mark@StatEase.com)

ABSTRACT

This paper identifies eight keys to success in applying statistical tools for design of experiments (DOE). Quality managers who grasp these keys will be better able to support use of DOE in their organization. Ultimately this will lead to breakthrough improvements in product quality and process efficiency.

1.0 Introduction

Design of experiments (DOE) provides a powerful means to achieve breakthrough improvements in product quality and process efficiency. This leads to increased market share, decreased costs, and big gains in profit. So why don't more manufacturers use design of experiments (DOE)? In some cases it's simple ignorance, but even when companies provide proper training, experimenters resist DOE because it requires planning, discipline and the use of statistics. Fear of statistics is widespread, even among highly educated scientists and managers. Quality professionals can play a big role in helping their colleagues overcome these barriers.

2.0 Discussion

Using DOE successfully depends on understanding eight fundamental concepts. To illustrate these keys to success, we'll follow two case studies:

- Increasing the life of a bearing
- Reducing shrinkage of plastic parts from an injection molding process

The focus will be on 2-level factorial design, where each input variable is varied at high (+) and low (–) levels. These designs are very simple, yet extremely powerful. For example, Figure 1 shows the results for a 2-level design on 3 factors affecting bearing life. Note the large increase at the upper right corner of the cube. In this case two factors interact to produce an unexpected breakthrough in product quality. One-factor-at-a-time (OFAT) experimentation will never reveal interactions like this.

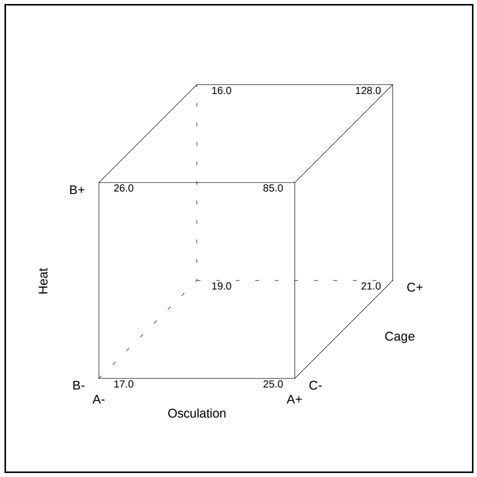

Figure 1: Two-Level Design for 3 Factors (Response: Bearing Life)

Full factorials are much more efficient than OFAT, because they make use of multivariate design. It's simply of matter of parallel processing (with factorial design) versus serial processing (with OFAT). Furthermore, you don't need to run the full number of 2-level combinations, particularly when you get to 5 or more factors. By making use of fractional designs, the 2-level approach can be extended to many factors. Therefore, these DOEs are ideal for screening many factors to identify the vital few that significantly effect your response. The injection molding case will demonstrate the use of fractional 2-level design.

2.1 Key #1: Set Good Objectives

The first decision before designing an experiment is "what is the objective, or purpose, of this study?" The focus of the study may be to screen out the factors that are not critical to the process, or it may be to optimize a few critical factors. A well-defined objective leads the experimenter to the correct DOE.

For example, in the initial stage of process development or troubleshooting, the appropriate design choice is a fractional two-level factorial. This DOE screens a large number of factors in a minimal number of runs. However, if the process is already close to optimum conditions, then a response surface design may be most appropriate. It will explore a few factors over many levels.

If you do not identify the objectives of a study, you may pay the consequences:

- trying to study too many or too few factors
- not measuring the correct responses
- arriving at conclusions that are already known

In essence, vague objectives lead to lost time and money, as well as feelings of frustration for all involved. Identifying the objective up-front builds a common understanding of the project and expectations for the outcome.

In our case study of the injection molder, management wants to reduce variation in the shrinkage of their parts. If the shrinkage can be stabilized, then mold dimensions can be adjusted so the parts can be made consistently.

The factors and levels to be studied are:

FACTOR NAME	UNITS	LOW LEVEL	HIGH LEVEL
A: Mold Temp	Degrees F	130	180
B: Holding Pressure	PSIG	1200	1500
C: Booster Pressure	PSIG	1500	1800
D: Moisture	Percent	.05	.15
E: Screw Speed	Inches/sec	1.5	4.0
F: Cycle Time	Seconds	25	30
G: Gate Size	Mils	30	50

Table 1: Factors for a DOE Case Study on a Molding Process

The experimenters have chosen a 2-level factorial design with 32 runs. A full set of combinations would require 128 runs (2^7) so this represents a 1/4th fraction.

2.2 Key #2: Measure Responses Quantitatively

Many DOE's fail because the response cannot be measured quantitatively. A classic example is found with visual inspections for quality. Traditionally, process operators or inspectors have developed a qualitative system that they use to determine if a product passes or fails. At best, they may have boundary samples of minimally acceptable product. Although this system may be OK for production, it does not have enough precision for a good DOE. Pass/fail data can be used in DOE, but it's very crude. For example, if your process typically produces a 0.1% defective rate you would expect to find 5 out of 5000 parts defective. In order to execute a simple designed experiment that investigated 3 factors in 8 experimental runs on such a process, you would need to utilize a minimum of 40,000 parts (8 x 5000). This would assure getting enough defects to judge improvement, but can you afford the cost?

For the purposes of experimentation, a rating scale works well. Even crude scaling from 1 to 5 will be far better than the simple pass/fail method. Define the scale by providing benchmarks in the form of defective units or pictures. Train three to five people to use the scale. During the experiment, each trained inspector should rate each unit. Some inspectors may tend to rate somewhat high or low, but this bias can be removed in the analysis via blocking (see key #5 below). For a good DOE, the testing method must consistently produce reliable results.

In the case study on injection molding, the experimenters will measure percent shrinkage at a critical dimension on the part. In addition, they could rate the parts for imperfections in surface quality.

2.3 Key #3: Replicate to Dampen Uncontrollable Variation (Noise)

Common sense tells you that the more times you replicate (repeat) a given set of conditions, the more precisely you can estimate the response. Replication improves the chance of detecting a statistically significant effect (the signal) in the midst of natural process variation (the noise). In some processes, the noise drowns out the signal. Before you do a DOE, it helps to assess the ratio of signal to noise. Then you can determine how many runs will be required for the DOE. You first must decide how much of a signal you want to be able to detect. Then you must estimate the noise. This can be determined from:

- control charts
- process capability studies
- analysis of variance (ANOVA) from prior DOEs
- best guess based on experience

The statisticians who developed two-level factorial designs incorporated 'hidden' replication within the test matrices. The level of replication is a direct function of the size of the DOE. You can use the following table to determine how many two-level factorial runs you need to provide a 90% probability of detecting the desired signal (see reference by Watts). If you can't afford to do the necessary runs, then you must see what can be done to decrease noise. For example, the table shows a minimum of 64 runs for a signal to noise ratio of 1. However, if you could cut the noise in half, the signal to noise ratio would double (to 2), thus reducing your runs by 4-fold (from 64 to 16). If you cannot reduce noise, then you must accept an increase in the detectable signal (the minimum effect that will be revealed by the DOE).

Signal to Noise Ratio (Δ/σ)	Minimum Number of Runs
1.0	64
1.4	32
2.0	16
2.8	8

Table 2: Number of Runs for 2-Level Factorial as a Function of Signal to Noise

You can improve the power of the DOE by adding actual replicates where conditions are duplicated. You can't just get by with repeat samples and/or measurements. The entire process must be repeated from start to finish. If you do submit several samples from a given experimental run, enter the response as an average.

For our case study on injection molding, control charts reveal a standard deviation of 0.60. Management would like to detect an effect of magnitude 0.85. Therefore the signal to noise ratio is approximately 1.4. The appropriate number of runs for this two-level factorial experiment is 32 runs. They decide not to add further replicates due to time constraints, but several parts will be made from each run. The response then becomes the average shrinkage per part, thus dampening out variability in parts and the measurement itself.

2.4 Key #4: Randomize the Run Order

The order in which you run the experiments should be randomized to avoid influence by uncontrolled variables such as tool wear, ambient temperature, and changes is raw material. These changes, which often are time-related, can significantly influence the response. If you don't randomize the run order, the DOE may indicate factor effects that are really due to an uncontrolled variables that just happened to change at the same time. For example, let's assume that you run an experiment to keep your copier from jamming so often during summer months. During the day-long DOE, you first run all the low levels of a setting (factor "A"), and then run the high levels. Meanwhile the humidity increases by 50%, creating a significant effect on the response. (Physical properties of paper are very dependent on humidity.) In the analysis stage, factor A then appears to be significant, but it really is the change in the humidity level that caused the effect. Randomization would have prevented this confusion. Always do it!

2.5 Key #5: Block Out Known Sources of Variation.

Blocking screens out noise caused by known sources of variation, such as raw material batch, shift changes, or machine differences. By dividing your experimental runs into homogeneous blocks, and then arithmetically removing the difference, you increase the sensitivity of your DOE.

Do not block on anything that you want to study. For example, if you want to measure the difference between two raw material suppliers include them as a factor to study in your DOE.

In the injection molding case study, management would like the experimenters to include all the machines in the DOE. There are 4 lines in the factory, which may differ slightly. The experimenters divide the DOE into 4 blocks of 8 runs per production line. By running all lines simultaneously, the experiment will get done 4 times faster. However, in this case, where the DOE already is fractionated, there is a cost associated with breaking it up into blocks: the interaction of A: Mold Temperature and B: Holding Pressure cannot be estimated due to aliasing. Aliasing is an unfortunate side-effect of fractional and/or blocked factorials. We will discuss this in the next section.

2-6 Key #6: Know Which Effects (if any) will be Aliased

An alias indicates that you've changed two or more things at the same time in the same way. Even unsophisticated experimenters know better. Aliasing is a critical and often over-looked feature of Plackett-Burman, Taguchi designs, or standard fractional factorials.

For example, if you try to study 3 factors in only 4 runs, a half-fraction, the main effects become aliased with the two-factor interactions. If you're lucky, only main effects will be active, but more likely there will be at least one interaction. The bearings case (depicted in Figure 1) can be manipulated to show how dangerous it can be to run such a low-resolution fraction. Table 3 shows the full-factorial test matrix. In this case, the interaction AB is very significant, so it's included in the matrix. (Note that this column is the product of columns A and B.) The half-fraction is represented by the enlarged and emboldened rows - the responses for the other runs have been struck out. Observe that in the highlighted area the pattern of the minuses (lows) and pluses (highs) for AB is identical to that of factor C. Therefore, by going to the half-fraction, we've now totally confused the real effect of AB with factor C.

Factor A: Osculation	Factor B: Heat	Interaction AB	Factor C: Cage	Response: Life (Hours)
–1	–1	+1	–1	~~17~~
+1	**–1**	**–1**	**–1**	**25**
–1	**+1**	**–1**	**–1**	**26**
+1	+1	+1	–1	~~85~~
–1	**–1**	**+1**	**+1**	**19**
+1	–1	–1	+1	~~21~~
–1	+1	–1	+1	~~16~~
+1	**+1**	**+1**	**+1**	**128**

Table 3: Test Matrix for Bearings Case (Shaded Area is Half-Fraction)

Aliasing is a problem which can be avoided by doing only full 2-level factorials or high resolution fractionals, but this would not be practical. The reference by Box provides details on aliasing for the standard fractional designs. However, it's very difficult to find any published information on the alias structures for Plackett-Burman or Taguchi designs. These designs are often very low in resolution and therefore give very misleading results on specific effects. If you must deal with these non-standard designs, always do a design evaluation to see what's aliased. Good DOE software will give you the necessary details, even if runs are deleted or levels changed. Then if any effects are significant you will know whether to rely on the results or do further verification.

The injection molding study is a fractional factorial design with mediocre resolution: several two-factor interactions are aliased. A design evaluation provides the specifics: CE=CE+FG, CF=CF+EG, CG=CG+EF, where the equal sign indicates aliasing. For example, if you evaluate the effects matrix for CE versus FG, you will see a perfect correlation. The plus symbol in the alias relationship tells you that the calculated effect could be due to CE plus FG. If these or any of the other aliased interactions are significant, further work will be needed.

2-7 Key#7: Do a Sequential Series of Experiments

Designed experiments should be executed in an iterative manner so that information that is learned in one experiment can be applied to the next. For example, rather than running a very large experiment with many factors and using up the majority of your resources, consider starting with a smaller experiment and then building upon the results. A typical series of experiments consists of a screening design (fractional factorial) to identify the significant factors, a full factorial or response surface design to fully characterize or model the effects, followed up with confirmation runs to verify your results. If you make a mistake in the selection of your factor ranges or responses in a very large experiment, it can be very costly. Plan for a series of sequential experiments so you can remain flexible. A good rule of thumb is not to invest more than 25 percent of your budget in the first DOE. For example, in the injection molding case, follow-up studies or confirmation runs should be included in the overall plan.

2-8 Key #8: Always Confirm Critical Findings

After all the effort that goes into planning, running, and analyzing a designed experiment, it is very exciting to get the results of your work. There is a tendency to eagerly grab the results and rush out to production and say, “We have the answer! This will solve the problem!” BEFORE doing that, you need to take the time to do a confirmation run and verify the outcome. Good software packages will provide you with a prediction interval to compare the results within some degree of confidence. Remember that in statistics you never deal with absolutes - there is always uncertainty in your recommendations. Be sure to double-check your results.

In the injection molding case, the results of the experiment revealed a significant interaction between booster pressure and moisture (CD). The interaction is not aliased with any other 2-factor interactions, so it’s a clear result. As shown by the flat line on Figure 2, shrinkage will be stabilized by keeping moisture low (D−). This is known as a robust operating condition.

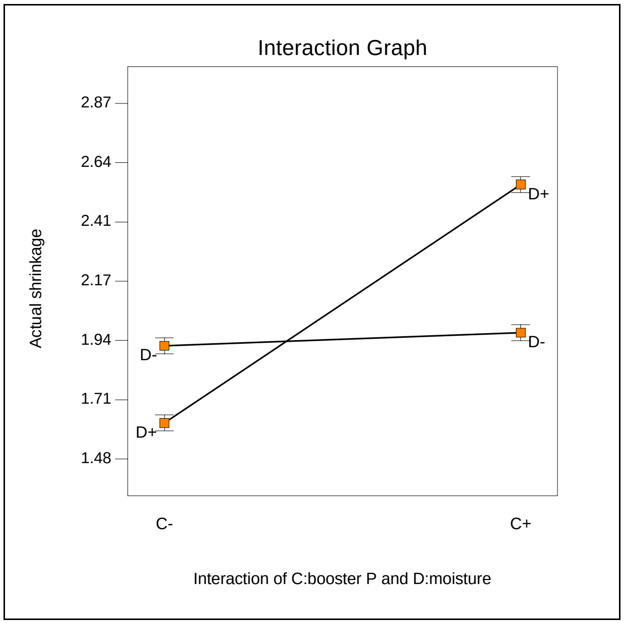

Figure 2: Interaction Plot for CD Effect on Shrinkage

Contour graphs and 3D projections help you visualize your response surface. To achieve robust operating conditions, look for flat areas in the response surface. Figure 3 shows a 3D graph for the factors C and D in the injection molding system. Notice how it flattens as the moisture is decreased. This is a region of stability for the shrinkage response.

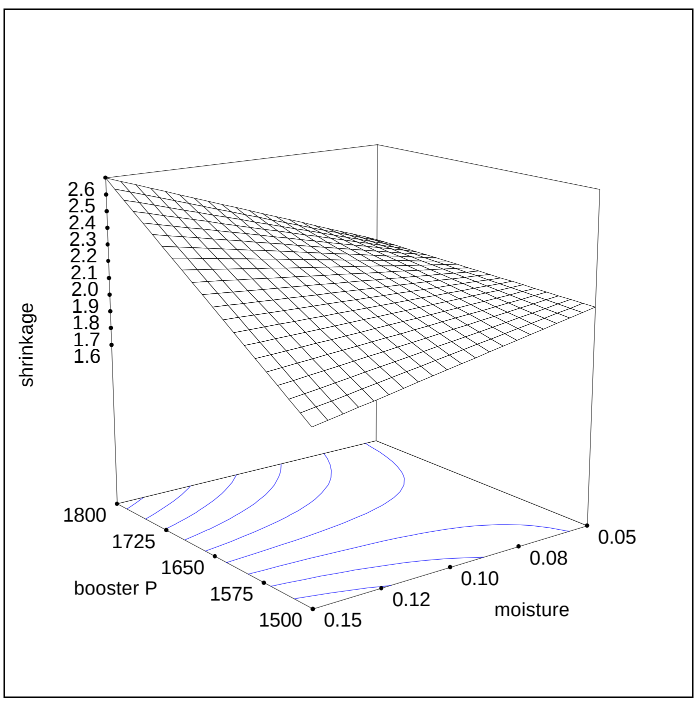

Figure 3: 3D Response Surface for Shrinkage

The 3D surfaces are very impressive, but they're only as good as the data generated to create the predictive model. The results still must be confirmed. If you want to generate more sophisticated surfaces, you should follow-up with response surface methods (RSM) for optimization. These designs require at least 3 levels of each factor, so you should restrict your study to the vital few factors that survive the screening phase.

3.0 CONCLUSION

Design of experiments is a very powerful tool that can be utilized in all manufacturing industries. The keys to success are:

1. Set good objectives.
2. Measure responses quantitatively.
3. Replicate to dampen uncontrollable variation.
4. Randomize the run order.
5. Block out known sources of variation.
6. Know which effects (if any) will be aliased.
7. Do a sequential series of experiments.
8. Always confirm critical findings.

Quality managers who encourage use of DOE, and promote these keys, will greatly increase the chances for making breakthrough improvements in product quality and process efficiency.

REFERENCES

Anderson, M. J., Whitcomb, P. J. [1997], "Software Sleuth Solves Engineering Problems", *Machine Design*, 6/5/97.

Anderson, M. J., Whitcomb, P. J. [1997], "Breakthrough Improvements with Experiment Design", *Rubber and Plastics News*, 6/16/97.

Box, Hunter and Hunter [1978], *Statistics for Experimenters*, John Wiley & Sons.

Watts, E. G. [1997], "Explaining Power Using Two-Level Factorial Designs", *Journal of Quality Technology*, Vol. 29, No. 3, (July).