Software sleuth solves engineering problems

MARK J. ANDERSON
PATRICK J. WHITCOMB
Application Engineers
Stat-Ease Inc.
Minneapolis, Minn.

Engineers at an aluminum-casting company were pulling their hair out trying to understand why a particular part came off the line filled with inclusions. Like most engineers faced with a challenge, they would change one manufacturing variable at a time, such as material temperature or the injection rate of the liquid aluminum into the die, but none seemed to reduce the defect count below an acceptable value. In desperation they turned to statistical software and a process called design for experiments.

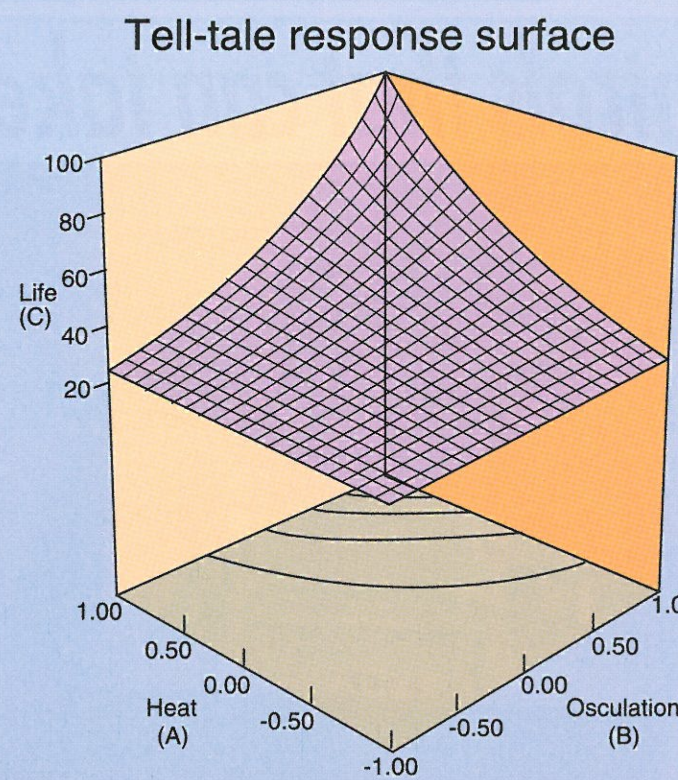
Having conducted so many trial-and-error experiments, the engineers had already collected sufficient data to feed into the software. Findings indicated an unexpected reaction between the molten aluminum's temperature and the final-phase intensification pressure. Optimizing on only these two factors let the engineers eventually reduce the defect rate to zero.

TWO-LEVEL SIMPLICITY

The technology hero in this case is not new; it has been around about 50 years. What's relatively new is design-of-experiments or DOE software that quickly grinds through necessary calculations to draw illuminating plots. Engineers can make best use of the technology in a simple form called a two-level factorial design. Working through a simple yet practical example introduces potential users to the terminology and general guidelines for using the technology.

The real-world example for the software's usefulness comes from bearing manufacturer SKF. Applying the technology to the interaction of components in a roller bearing let its engineers uncover a relationship that led to a many-fold increase in operating life. This and similar experiments saved the company millions of dollars and helped them fend off challenges in quality and price from off-shore manufacturers.

The two-level factorial method involves adjusting experimental factors to only high and low levels. For instance, if heat seems to be a factor in the design or process, running experiments at high and low temperatures collects information to evaluate its effect. The two-level design approach offers a parallel testing scheme that is more powerful than one-factor-at-a-time methods. By restricting tests to two levels, users minimize the number of experiments needed. The contrast between levels provides the driving



A test result called a response surface has been calculated by Design-Expert software from Stat-Ease Corp. after digesting the results from several structured bearing tests. The high point of the graph indicates that a particular dimension ratio in the bearing along with a heat treatment produces a rugged design capable of an unexpectedly long life. Without the scientific approach to testing that design of experiments provides, the bearing engineers probably would not have found the right conditions.

force to uncover the most dominant effects.

Engineers might construct two-level factorial designs with the help of a textbook. Of course they will have to run carefully constructed experiments. But with a little background, software can quickly grind through most of the drudge work. Users need only run experiments or tests with as little as one more than the number of factors they want to test. For example, one could test seven factors (heat, lubrication, dimensions, and so on) in eight runs or tests, or 15 factors in 16 runs.

Two-level factorial design

Experiments \ Number of Factors	2	3	4	5	6	7	8	9	10
4	Full	1/2 Fract.							
8		Full	1/2 Fract.	1/4 Fract.	1/8 Fract.	1/16 Fract.			
16			Full	1/2 Fract.	1/4 Fract.	1/8 Fract.	1/16 Fract.	1/32 Fract.	1/64 Fract.
32				Full	1/2 Fract.	1/4 Fract.	1/8 Fract.	1/16 Fract.	1/32 Fract.
64					Full	1/2 Fract.	1/4 Fract.	1/8 Fract.	1/16 Fract.

The two-factorial design guide assists setting up screening studies with a color code. For example, white squares indicate a full factorial problem. Green squares show a good design with high-resolution fractional factorials. These are the best choices because they are efficient. Yellow implies medium resolution; users should proceed with caution because two-factor interactions are likely to be confused with other two-factor interactions. Red suggests low-resolution fractions; main effects will be confused with two-factor interactions.

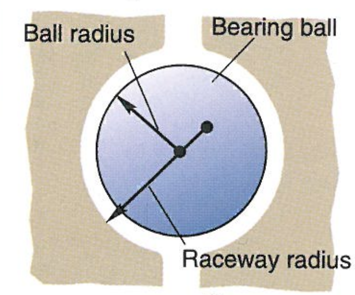
In the following example, the bearing manufacturer used two-level factorials to make a breakthrough improvement in a standard deep-groove bearing. The experimenters designed several variations of the bearing to study osculation, heat treatment of the inner ring, and cage material — conditions the engineers thought most affected bearing life.

Osculation is a ratio of the ball radius to the outer-ring raceway radius. Tests involved two levels of the ratio, one high and one low. Heat treating was also done in two levels, and the cages were made of two materials: steel and a less-expensive polymer. All possible combinations of these factors require eight experiments ($2^3 = 8$). From the eight runs one gets eight pieces of information — three main effects, three two-factor interactions, one three-factor interaction, and the overall average.

The accompanying cube plot shows results from tests in terms of bearing life. Simply performing the tests may uncover the right combination of conditions. But when results are not dramatic, significant effects can be hidden by variability in results. Or, users might think a result is significant when it's really just caused by normal variation. For example, statistical analysis reveals no significant difference between the 85-hr life and the 128-hr life in the tested bearings. In fact, the high values at the upper right edge indicate a previously unknown interaction between osculation and heat treatment. This breakthrough was not revealed by prior one-factor-at-a-time experimentation.

The specific design layout for the bearing case is shown in the table *Breakdown of a design matrix.* Columns A, B, and C represent control factors (osculation, heat treatment, and cage material). These are laid out according to a standard order that can be obtained from any textbook or software on design of experiments. Pluses indicate high levels and minuses low levels. Each column contains four pluses and four minuses. The matrix offers an important statistical property called or-

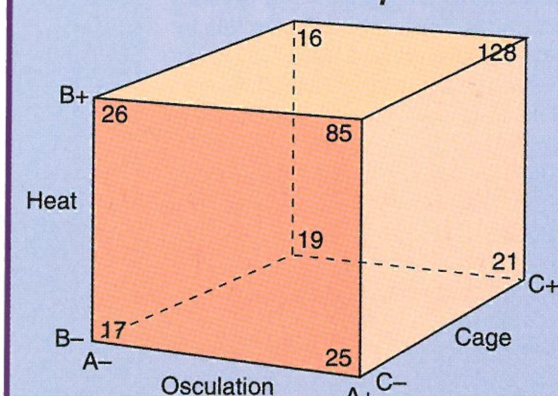
Bearing osculation



$$\text{Osculation} = \frac{R_{\text{bearing}}}{R_{\text{raceway}}}$$

Osculation is the ratio of the ball radius to race radius.

Cube plot shows two-level response



The cube plot for a 2^3 factorial test shows where the test factors have the most effect.

thogonality which means factors are not correlated. That's good.

If one just collected happenstance data from production records, it is highly unlikely one would get an array of factors shown in the table. It would probably show that factors such as temperature and pressure rise and fall together. As factors become more correlated, the error in estimation of their effects becomes larger. That's not good.

Breakdown of a design matrix

STANDARD ORDER	A	B	C	AB	AC	BC	ABC	LIFE (HR)	LOG_{10} LIFE
1	–1	–1	–1	+1	+1	+1	–1	17	1.23
2	+1	–1	–1	–1	–1	+1	+1	25	1.40
3	–1	+1	–1	–1	+1	–1	+1	26	1.41
4	+1	+1	–1	+1	–1	–1	–1	85	1.93
5	–1	–1	+1	+1	–1	–1	+1	19	1.28
6	+1	–1	+1	–1	+1	–1	–1	21	1.32
7	–1	+1	+1	–1	–1	+1	–1	16	1.20
8	+1	+1	+1	+1	+1	+1	+1	128	2.11
Effect (as is)	45.25	43.25	7.75	40.25	11.75	8.75	14.75		
Effect (log 10)	0.41	0.36	–0.015	0.30	0.66	–0.001	0.13		

An experiment with three main factors involves eight runs ($2^3 = 8$). The design matrix presents a method of organizing the data and results from a two-level bearing test with three factors. In the example, A stands for osculation, B for heat treatment, and C for cage material. The plus refers to a high level factor and minus for low level.

Orthogonal test matrices make effect estimation neat and easy. For example, the effect *E* of factor A is calculated by simply averaging the responses at the plus level and subtracting the average at the minus level.

$E = +A_{mean} - (-A_{mean})$

or

$$E = [(25 + 85 + 21 + 128) - (17 + 26 + 19 + 16)]/4 = 45.25$$

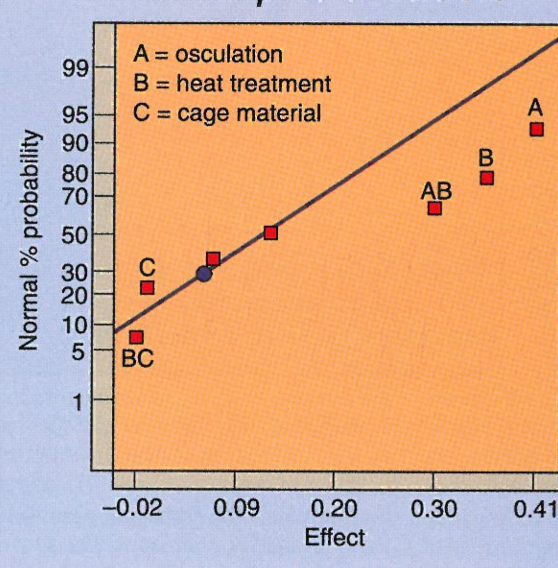
The table also shows the response (in this case the bearing life) varies by nearly an order of magnitude from 16 to 128 hr. In situations with such contrast, statisticians routinely perform a transformation of the response, most commonly with a logarithm. One can do the same on graph paper with log scales to plot data in a straight line. The log counteracts a common relationship: the true standard deviation increases as the true mean increases. In other words, error is a constant percentage of the average response. This violates an important statistical assumption, that the variation is a constant. If one cannot satisfy this assumption some statistics may come out wrong, so consider transforming the values with a common mathematical function.

Let engineering knowledge be a guide when selecting an appropriate transformation. For example, chemists might use the rule of thumb that says reaction rates double for every 10° of temperature increase.

When one cannot predict what a relationship should be, try log-log or square-function paper. A better statistical fit for the bearing-life data came from a log transformation. The transformed responses appear in the last column of the matrix. These figures were used to calculate the effects listed in the bottom row.

After studying the design matrix, it seems obvious that one should focus on the largest effects and dismiss the rest, right? Wrong. What if none of the effects are real and testers just measured results from random error? The vital few significant factors must be screened out of many trivial ones that occur by chance. One can do this easily with a graph called the normal plot. Textbooks provide details on how to construct these graphs, but DOE software can do it faster. Typically, users see a group of near-zero effects that form a line. After a noticeable gap one may find effects much smaller or larger than the others. Anything significant falls off to the bottom left or up-

Normal plot of effects



Plotting the effects calculated from the design matrix versus the normal percent probability indicate which effects are most influential. They appear away from the others to the right. In this case, drawing a line from the left-most (least influential) effects highlights that osculation, heat, and their combination are most critical.

per right of the line.

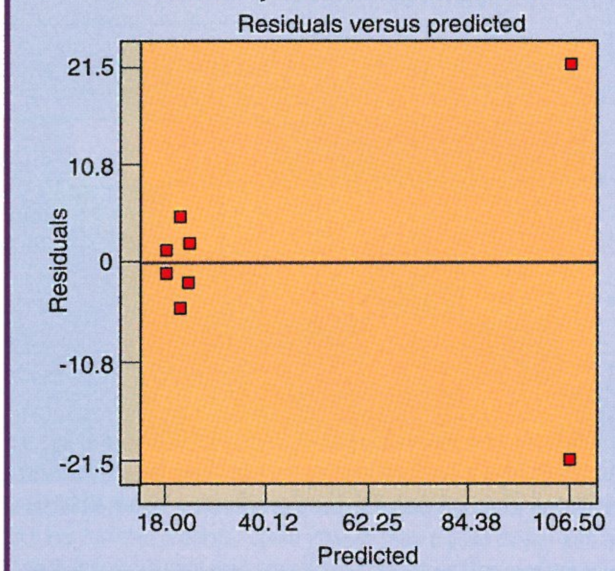
The figure *Normal plot of effects* shows those for the bearing tests. Significant effects are labeled. The near-zero effects fall on a straight line exhibiting normal scatter. These insignificant effects can be used to estimate experimental error. To be conservative, consider replicating the design to get estimates of pure error. This is an actual measure of error as opposed to assumed error from pooling insignificant effects.

Be sure to randomize the run order of the entire design. Do the same for replicate runs. For example, write the factors for each of eight sets of conditions and three repeat tests on separate sheets of paper and place them in a container. Pulling them out one by one would describe a random order that helps ensures valid figures. Without random testing, users leave themselves open to lurking or background factors such as gradual change in ambient temperature or machine wear. These could confound factor estimates.

With a valid estimate of experimental error, standard statistical analyses can validate the overall outcome and individual effects. A few texts provide hand-calculation methods for doing statistical analysis of two-level factorials, but it's much easier to let statistical software programs do the work.

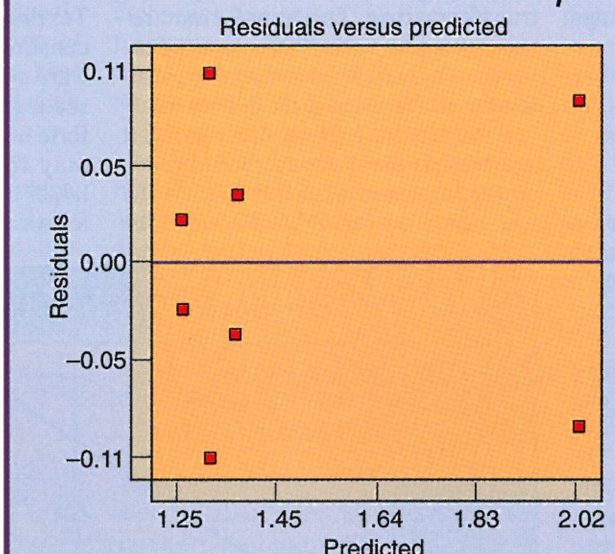
An indicator of sufficient capability in the software would be a feature that performs residual analysis. Residuals are the difference between actual and predicted response. They represent the error in predictions. Because of the variability of process and test, one cannot be right-on in each case. Just be sure that the residuals are about normal. This gives credibility to the statistics. Users can check the validity of their data by plotting residuals versus predicted level. Watch for a pattern of increasing variation sometimes called a megaphone shape. It indicates a violation of the statistical assumption that variance is a constant. The illustration *Normal plot of residuals* shows them for the bearing case with the model fitted to original data. It does not look good because residuals increase greatly when predicted values increase.

Normal plot of residuals



Plotting residuals versus predicted results clumps values and masks their possible influence.

How a transformation helps



Transforming the axes in the illustration *Normal plot of residuals* with a simple log function expands the values to reveal an even or normal scatter, and therefore, no hidden effects.

DOE textbooks can provide a bit more statistical advice on the subject. However, there is no substitute for knowing your process. This should guide in selecting a transformation. The illustration *How a transformation helps* shows the residual plot after transforming the *Y*-coordinate data by a log function. It looks good because there is no particular pattern other than a normal scatter.

Residual analysis also may reveal individual outliers. These are results that do not fit with the rest; they stand out or look different. But be careful. Don't delete points unless you can assign a special cause such as a broken fixture on a test machine. Quite often an outlier turns out to be simply an error in data entry. People easily transpose digits.

INTERPRETING RESULTS

With the foregoing information, one can generate a report. Start by making a plot of any significant main effects that are not part of a major interaction. For example, there are none in the bearing case. The effects form a hierarchical family: A, B, AB. This is a fairly typical outcome.

Next produce the interaction plots. The effects AB tell the entire story in this case. The lines on this plot are not parallel. This means the effect of one factor depends on the level of the other, so it would be inappropriate to display either of these main factors by themselves. On the interaction plot for the bearing case, for instance, factor A (osculation) has a greater impact when B (heat treatment) is at the higher level. The chart indicates that high A and high B most improve bearing life. This is dramatized by the illustration, *Tell-tale response surface*.

Before making a final recommendation on new factor levels, perform confirmation runs. One can predict the outcome with a simple equation that uses the overall average modified up or down depending on the level of each factor. The model coefficients are simply the effects divided by two (the difference between +1 and –1). Statisticians call this a coded equation because they plug in values of +1 for high levels and –1 for low. A midpoint setting is entered as 0. The predictive model for the bearing case is:

$$\mathrm{Log}(L) = 1.49 + 0.2A + 0.18B + 0.15AB$$

where L = bearing life, hr. and A, B, and AB = +1 or –1.

Plugging in the recommended settings in coded form gives the predicted outcome:

$$\mathrm{Log}(L) = 1.49 + 0.2(+1) + 0.18(+1) + 0.15(+1) = 2.02$$

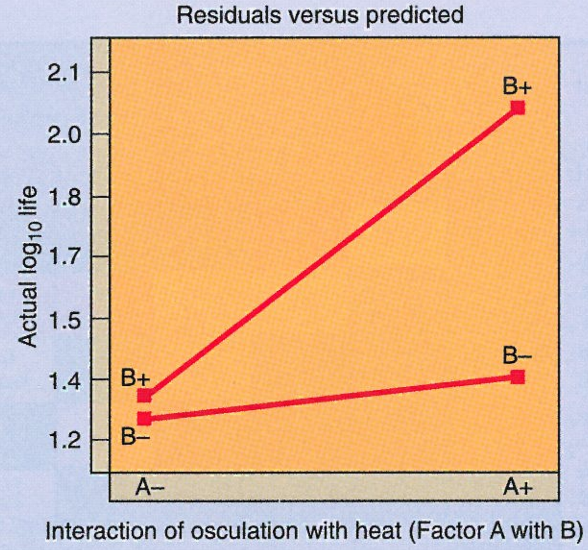
The transformation must be reversed to get the response back to the original units of measure. For example:

$$\text{Bearing life} = 10^{2.02}\ \text{hr} = 105\ \text{hr.}$$

This compares well with observed results. However, be prepared for some variation when confirming tests. Software should provide a confidence level on the expected values. Use this data to manage expectations. For instance, clients think engineers can predict events with certainty. Hedge on this by predicting an expected range. This way expectations will not be unrealistic.

The case study on bearing life illustrates how two-level factorials can be applied to a machine-design process with several variables. The bearing experiments uncovered a large interaction which led to a breakthrough production improvement. What's remarkable is that even nonresults provided useful information. The experimenters found the cage material has no effect. That means it could be set at its most economical level. Thus, significant savings also come from statistically insignificant factors. ■

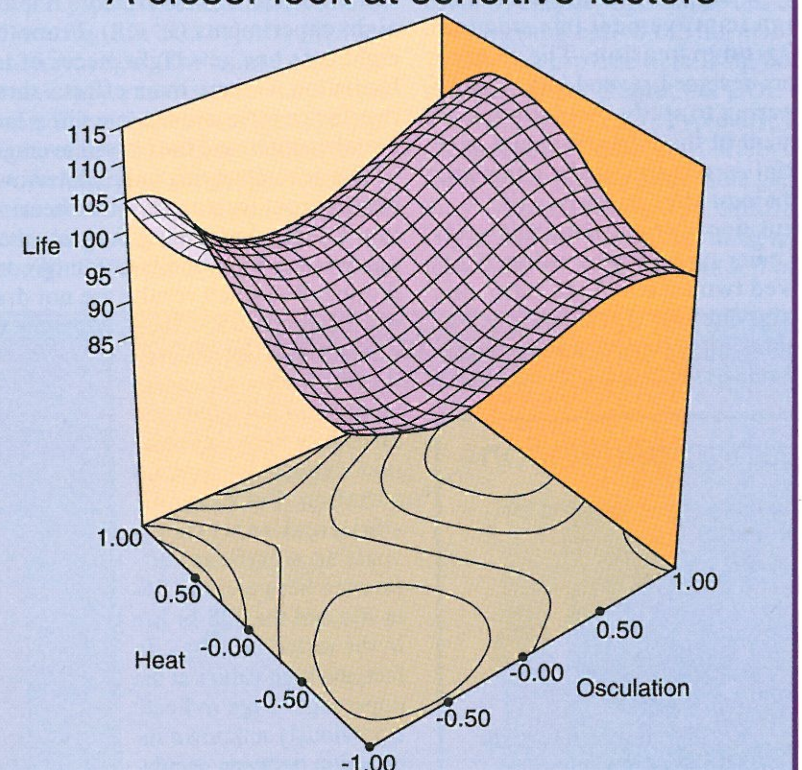
Graph of interactions



The interaction graph provides another method of pinpointing the most influential effects. The *Y* axis is a base 10 log of bearing life. The *X* axis is simply low and high osculation, factor A. Factor B is heat.

Further information is available from:
Stat-Ease Corporation
2021 East Hennepin Avenue
Suite 191
Minneapolis, MN 55413
Ph: 612-378-9449
Fax:612-378-2152
Web: www.statease.com
e-mail: info@statease.com

A closer look at sensitive factors



After engineers have digested the results from the first series of tests they may wish to fine-tune their results by pushing the osculation and heat treatments to higher levels. Such further studies might produce the more complex response surface in which maximums and minimums are less obvious but not less significant.