Software Sleuth Solves Engineering Problems

Mark J. Anderson* and Patrick J. Whitcomb
Stat-Ease, Inc.
*mark@statease.com

Design of experiments (DOE) is a powerful tool that you can use to achieve breakthrough product improvements. Here's just a few examples:

- An aluminum diecaster discovers the key factors needed to reduce defects from 15% to near zero.[1]
- After four years of unfruitful effort, John Deere Engine Works solves a paint adhesion problem, saving nearly $500,000 per year.[2]
- Engineers from Eastman-Kodak use DOE to redesign from a metal clip. They use additional DOEs to reduce time needed to retool the precision stamper. The end result is a ten-fold reduction in scrap, with faster setup and elimination of costly lubricants.[3]
- SKF, the world's largest bearing manufacturer, uncovers an interaction that leads to a many-fold increase in part life. This and similar experiments saved tens of millions of dollars and helped the company withstand the Japanese challenge in quality and price.[4,5]

All of these improvements came from a very simple form of DOE called two-level factorial design. A previous article outlined the broad procedures for doing this type of DOE.[6] In this article you will be given the primary details. We will do this from an engineering perspective, with an emphasis on the practical aspects. The bearing case noted above will form the basis for explanation.

Two-level factorials for maximum efficiency

Two-level factorial design is a statistically-based method that involves simultaneous adjustment of experimental factors at only two levels: high and low. The two-level design approach offers a parallel testing scheme that's much more efficient than one-factor-at-a-time. By restricting the tests to only two levels, you minimize the number of experiments. The contrast between levels gives you the necessary driving force for product improvement.

Two-level factorial designs can be constructed with the aid of a textbook[7], or better yet, with a statistical software package, most of which offer design of experiments capabilities[8]. You will find designs available with as little as one more than the number of factors you want to test. (See Figure 1). For example you could test 7 factors in 8 runs, or 15 factors in 16 runs. However, these "saturated" designs provide very poor resolution: main effects will be confused with two-factor interactions. We advise that you avoid running such low resolution designs because you may miss an important interaction.

DOE reveals incredible two-factor interaction

SKF, a Swedish manufacturer with plants in 14 countries, used two-level factorials to make a breakthrough improvement in a standard deep-groove bearing. The experimenters set aside one production line for a study of 3 factors:

A. Osculation: two levels of the ratio of the ball radius to the radius of the outer ring raceway. (See Figure 2.)
B. Heat treatment of inner ring: two levels.
C. Cage design: steel versus a cheaper polymer

All possible combinations of these factors required 8 experiments. This DOE is symbolized mathematically as 2^3. From the 8 runs you can get 8 pieces of information: 3 main effects, 3 two-factor interactions, 1 three-factor interaction, plus the overall average.

Figure 3 shows the results from the DOE on bearing life. The high values at the upper right edge indicate a previously unknown interaction between osculation and heat. This breakthrough was not revealed by prior one-factor-at-a-time experimentation.

The specific design layout for the bearing case is shown in Table 1. Columns A, B and C represent the control factors. These are laid out according to a standard order that can be obtained from any textbook on design of experiments (or provided by software). Note the balanced array of plus (high) and minus (low) levels in the test matrix. Each column contains 4 pluses and 4 minuses, which provides statistical power. The matrix offers a very important statistical property called “orthogonality” which means that factors are not correlated. If you just collected happenstance data from production records, it is highly unlikely you would get an array of factors like this. You would probably find that factors such as temperature and pressure go up and down together. As factors become more and more correlated, the error in estimation of their effects becomes larger and larger. That’s not good.

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

$$Effect = Mean\ A_{+} - Mean\ A_{-} \quad (1)$$

Applying a transformation to satisfy statistical assumptions

Notice in Table 1 that the response varies by nearly an order of magnitude, from 16 to 128 hours. In situations like this, statisticians routinely perform a *transformation* of the response, most commonly with a logarithm. Engineering students do the same thing when they use special graph paper, such as log scale, to get the data to come out in a straight line. The log counteracts a very common relationship: the true standard deviation increases as the true mean goes up. In other words, the error is a constant percentage of the average response. This violates an important statistical assumption - that the variation is a constant. If you cannot satisfy this assumption some statistics may come out wrong, so consider a transformation. Ideally, your engineering knowledge will guide you in the selection of an appropriate transformation. However, if you can’t predict what the relationship should be, try the log or square root. We got a better statistical fit from the log transformation of the bearing life data. The transformed responses can be seen in the last column of Table 1.

Using statistical principles to pick significant factors

You now know how to calculate effects. It seems obvious that you should pick the largest ones and run with those - right? Wrong! How do you know where to make the cut off? What if none of the effects are real, and you've just measured results due to random error? Somehow the vital few significant factors must be screened out of the trivial many that occur due to chance. You can do this easily with a graph called a "normal plot". Textbooks provide details on how to construct these graphs, but you will find it much easier to let DOE software do it for you. Typically you will see a group of near-zero effects that form a line. Then after a noticeable gap you may find effects that are much smaller or much larger than the others. Anything significant will fall off to the bottom left or upper right of the line. Figure 4 shows the normal plot of effects for the bearing case. 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.

If you want to be conservative, consider replicating the design to get estimates of "pure" error. Be sure to randomize the run order of your entire design, including replicates. Otherwise you leave yourself open to "lurking factors", such as gradual change in ambient temperature or machine wear, that could confound your factor estimates.

Given a valid estimate of experimental error, regardless of the source, standard statistical analyses can then be performed to validate the overall outcome and individual effects. Textbooks provide hand-calculation schemes for doing statistical analysis of two-level factorials, but it's much easier to let a statistical software program to do this work for you.

Validating the results by looking at residuals

Be sure you choose a program that provides "residual analysis" capabilities. Residuals are the difference between actual and predicted response. They represent the error in your predictions. Because of the variability of process and test, you won't be right on in every case. Just be sure that the residuals are approximately normal. Then you can rely on the statistics. You can check this by plotting residuals on normal or half-normal paper. You should also look at the plot of residuals versus predicted level. Beware of a pattern of increasing variation - a megaphone shape. This indicates a violation of the statistical assumption that variance is a constant. Figure 5a shows the plot of residuals for the bearing case with the model fitted to original data. It doesn't look good.

If you do see a bad patterns on the residual plots, consider the use of a response transformation. The log transformation often helps. You might also try a square root or one of many other functions. Refer to a DOE textbook for statistical advice on this subject. However, there is no substitute for your process knowledge. This should guide you in selection of a transformation. Figure 5b shows the residual plot after transforming the data by the log. It looks good. There's no longer any particular pattern other than a normal scatter.

Residual analysis also may reveal individual outliers. But be careful, don't delete points unless you can assign a special cause, such as a temporary breakdown in an agitator or the like. Quite often an outlier turns out to be simply an error in data entry. Digits can get transposed very easily.

Interpreting the results

Now you are ready to make your report. Start by making a plot of any significant main effects that are not part of a significant interaction. 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. In this case AB will tell the entire story (Figure 6). Notice that the lines on this plot are not parallel. This means that 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. For example, on the interaction plot for the bearing case, notice that factor A (osculation) has a much bigger impact when B (heat) is at the higher rate. Clearly it's best to go with high A and high B. This is dramatized by Figure 7, which shows a three-dimensional response surface graph for bearing life.

Before you make a final recommendation on the new factor levels, it would be wise to perform confirmation runs. You can predict the outcome with a simple equation that uses the overall average modified up or down depending on the level to which you set each factor. The model coefficients are simply the effects divided by 2. Statisticians call this a "coded" equation because you plug in values of plus 1 for high and minus 1 for low levels. (A midpoint setting is entered as 0.)

For the bearing case the predictive model is:

$$Log_{10}\,Bearing\ Life = 1.49 + 0.2A + 0.18B + 0.15AB \qquad (2)$$

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

$$Log_{10}\,Bearing\ Life = 1.49 + 0.2(+1) + 0.18(+1) + 0.15(+1)$$
$$= 2.02 \qquad (3)$$

Then to get the response back to the original units of measure, the transformation must be reversed.

$$Bearing\ life = 10^{2.02} = 105\ hours \qquad (4)$$

This compares well with the observed results. However, be prepared for some variation when you do confirmation tests. Your software should provide a confidence interval on the expected values. Use this data to "manage expectations".

What's in it for you

The case study on bearing life illustrates how two-level factorials can be applied to a machine design process with several variables. The design of experiments uncovered a large interaction which led to a breakthrough product improvement. The experimenters found that one of the factors, cage design, had no effect, so it could be set at its most economical level. Thus, significant savings can be achieved even from statistically insignificant factors.

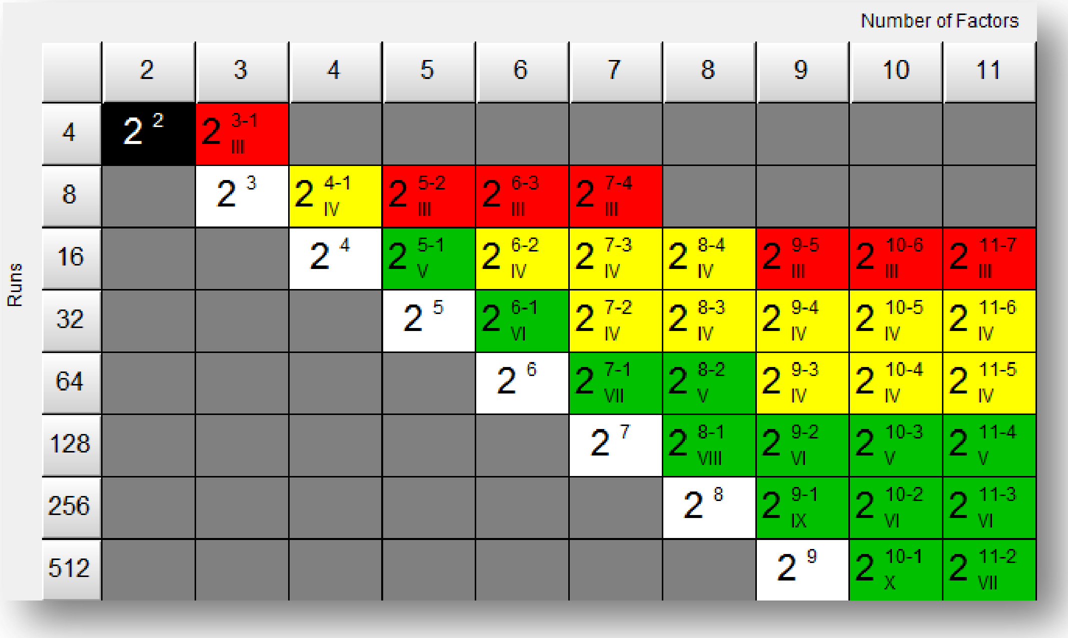
If you equip yourself with the basic tools of statistical DOE, you will be in a position to make the most of opportunities such as those presented at the outset of this article. Your reputation will be enhanced and the competitive position of your company advanced.

Literature Cited

1) "Diecaster achieves zero-defect parts," *Quality in Manufacturing*, (March/April, 1994)
2) Mills, W., "John Deere Saves $500K Annually with DOE," *Scitech Journal*, (July 1995).
3) Runke, P., "Design of Experiments Software Saves Kodak Thousands," *MetalForming*, (April, 1996).
4) Box, G. E. P., "George's Column," *Quality Engineering, Vol. 2 No. 3,* (1990).
5) Hellstrand, C., "The Necessity of Modern Quality Improvement and Some Experience with its Implementation in the Manufacture of Rolling Bearings," Report No. 35, Center for Quality and Productivity Improvement, University of Wisconsin, (March 1989)
6) Burnham, R. A., "A better way to design experiments," *Machine Design*, (April 1996)
7) Montgomery, D.C., *Design and Analysis of Experiments*, 3rd ed., John Wiley & Sons, Inc, New York, 1991.
8) Helseth, T.J., et al, *Design-Ease*, Windows or Macintosh, Stat-Ease, Inc, Minneapolis ($395).

Table 1. 2^3 Design Matrix, Data and Effects for Bearing Case

Standard Order	A	B	C	AB	AC	BC	ABC	Life (hours)	Log Base 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		



Runs \ Number of Factors	2	3	4	5	6	7	8	9	10	11
4	2^2	2^{3-1}_{III}								
8		2^3	2^{4-1}_{IV}	2^{5-2}_{III}	2^{6-3}_{III}	2^{7-4}_{III}				
16			2^4	2^{5-1}_{V}	2^{6-2}_{IV}	2^{7-3}_{IV}	2^{8-4}_{IV}	2^{9-5}_{III}	2^{10-6}_{III}	2^{11-7}_{III}
32				2^5	2^{6-1}_{VI}	2^{7-2}_{IV}	2^{8-3}_{IV}	2^{9-4}_{IV}	2^{10-5}_{IV}	2^{11-6}_{IV}
64					2^6	2^{7-1}_{VII}	2^{8-2}_{V}	2^{9-3}_{IV}	2^{10-4}_{IV}	2^{11-5}_{IV}
128						2^7	2^{8-1}_{VIII}	2^{9-2}_{VI}	2^{10-3}_{V}	2^{11-4}_{V}
256							2^8	2^{9-1}_{IX}	2^{10-2}_{VI}	2^{11-3}_{VI}
512								2^9	2^{10-1}_{X}	2^{11-2}_{VII}

Figure 1. Options for Standard 2-Level Factorials (Up to 10 factors in 64 runs)

Color Key (like a stoplight):

- White - full factorial: no problems but may require too many runs
- Green - high resolution (V) fractional factorial: Go ahead, it's a good design
- Yellow - medium resolution (IV) fraction: Proceed with caution, main effects OK, but two-factor interactions confused with other two-factor interactions
- Red - low resolution (III) fraction: Stop and think, main effects confused with two-factor interactions

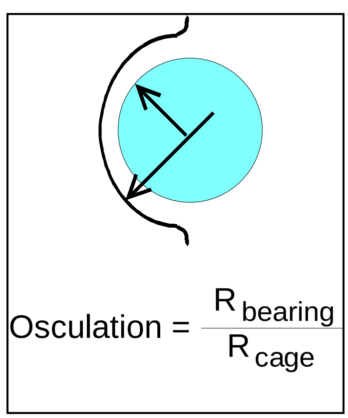


Figure 2. Osculation of bearing (R = Radius)

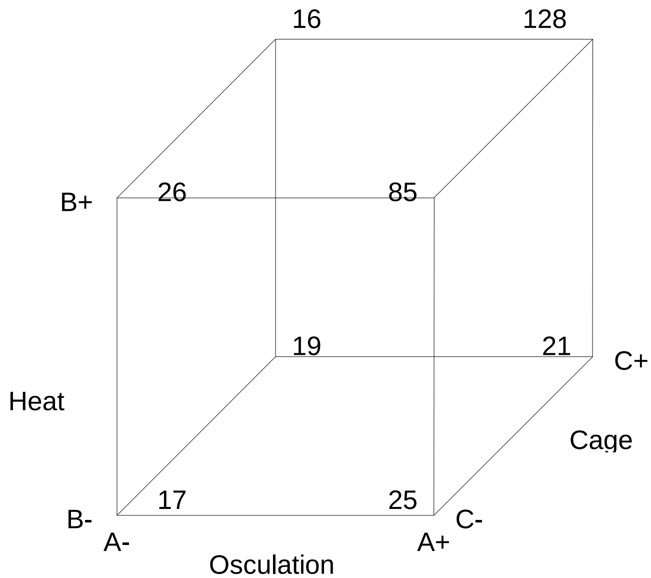


Figure 3. Cube Plot Showing Breakthrough Increase in Hours of Bearing Life

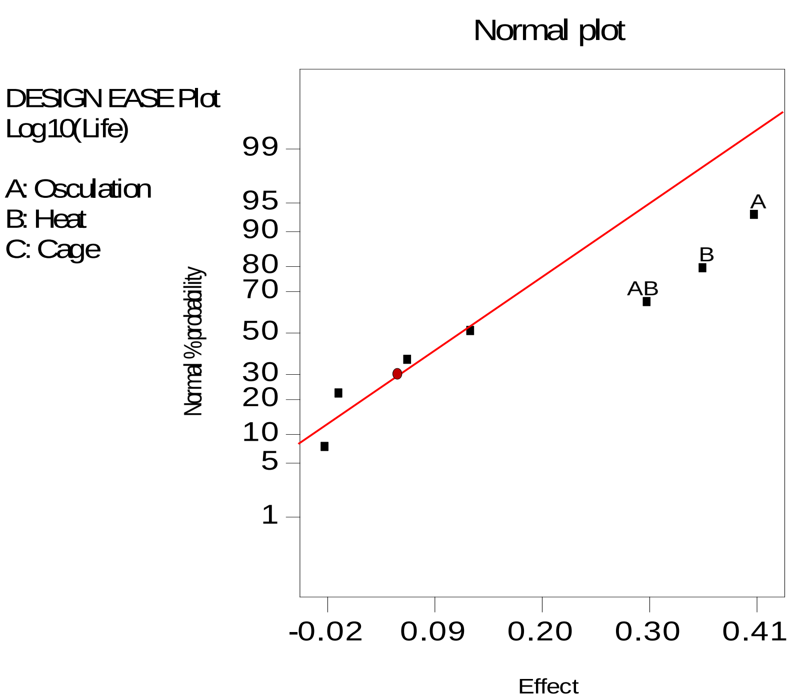


Figure 4. Normal Plot of Effects

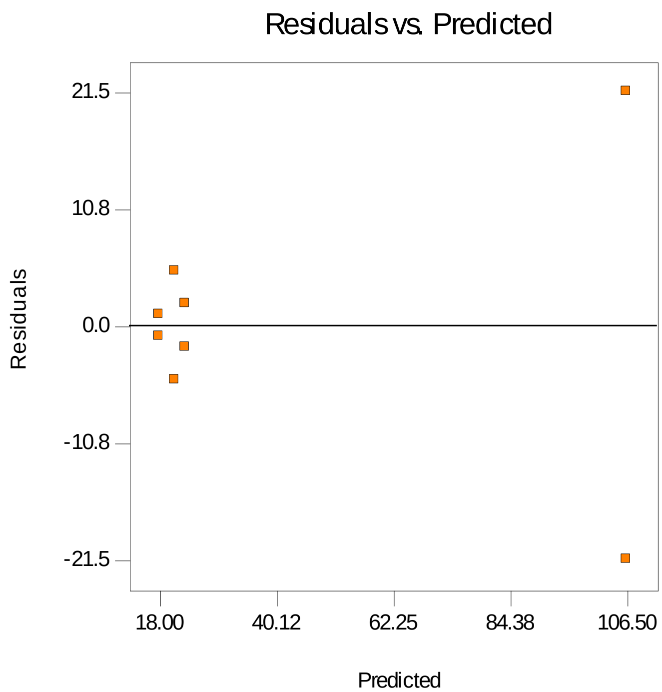


Figure 5a. Normal Plot of Residuals from Model Fitted to Raw Data

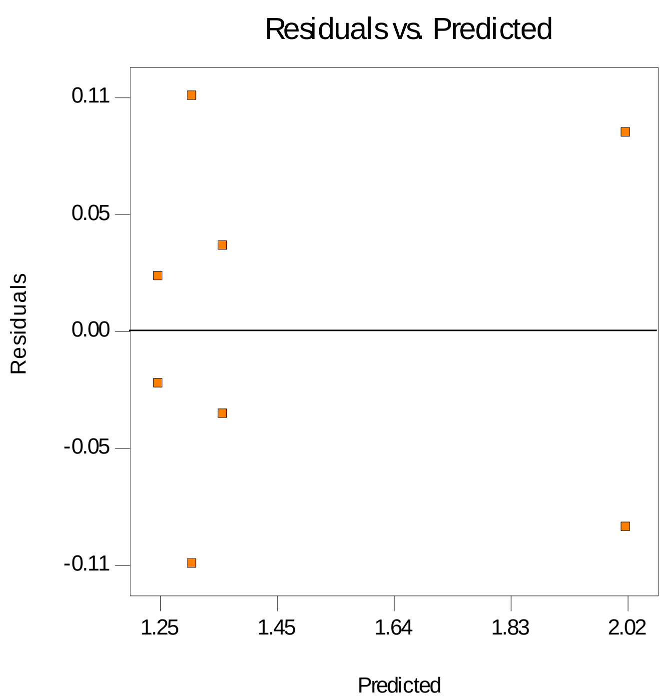


Figure 5b. Normal Plot of Residuals from Model Fitted to Transformed Data

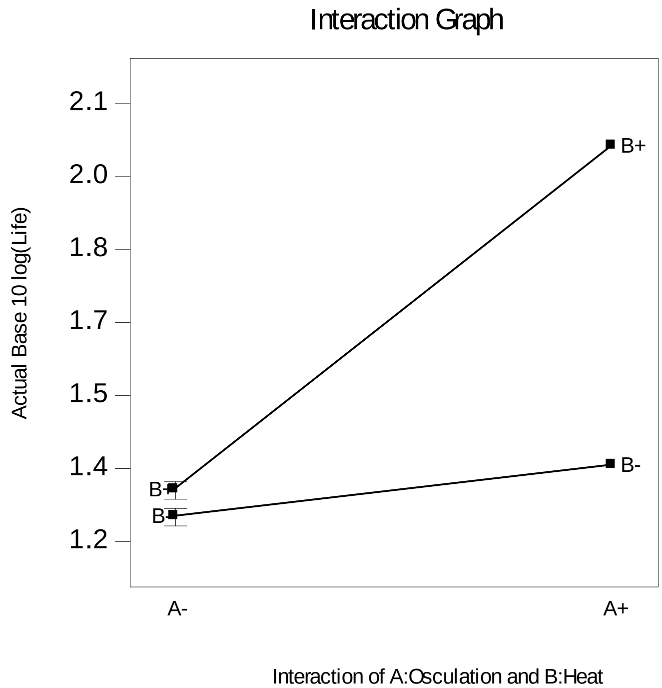


Figure 6. Interpretation Plot for Interaction AC

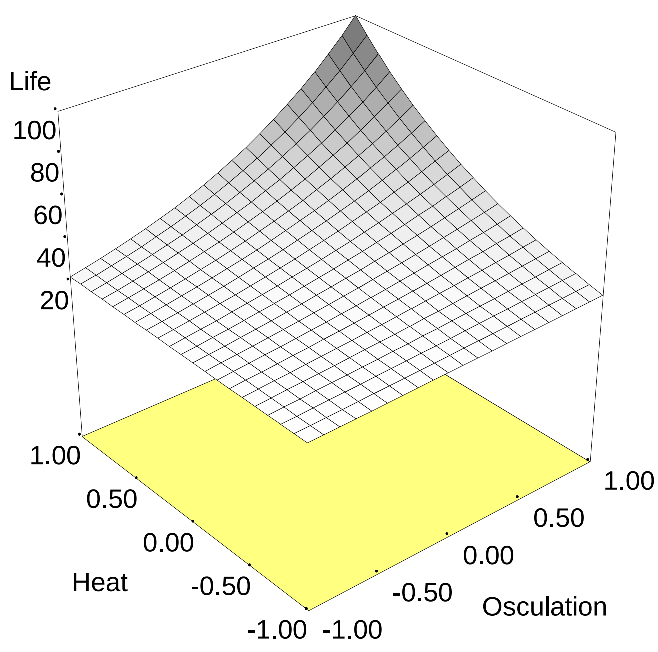


Figure 7. Response Surface Graph Illustrating Exponential Increase in Bearing Life Due to Interaction of Two Factors