Copyright © 2011 IEEE. Reprinted, with permission, from Huairui Guo and Adamantios Mettas, “Design of Experiments and Data Analysis,” *2011 Reliability and Maintainability Symposium*, January, 2011.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of ReliaSoft Corporation's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

# Design of Experiments and Data Analysis

## Huairui Guo, Ph. D. & Adamantios Mettas

Huairui Guo, Ph.D., CPR.
ReliaSoft Corporation
1450 S. Eastside Loop
Tucson, AZ 85710 USA
*e-mail*: Harry.Guo@ReliaSoft.com

Adamantios Mettas, CPR
ReliaSoft Corporation
1450 S. Eastside Loop
Tucson, AZ 85710 USA
*e-mail*: Adam.Mettas@ReliaSoft.com

Tutorial Notes © 2011 AR&MS

# SUMMARY & PURPOSE

Design of Experiments (DOE) is one of the most useful statistical tools in product design and testing. While many organizations benefit from designed experiments, others are getting data with little useful information and wasting resources because of experiments that have not been carefully designed. Design of Experiments can be applied in many areas including but not limited to: design comparisons, variable identification, design optimization, process control and product performance prediction. Different design types in DOE have been developed for different purposes. Many engineers are confused or even intimidated by so many options.

This tutorial will focus on how to plan experiments effectively and how to analyze data correctly. Practical and correct methods for analyzing data from life testing will also be provided.

# Huairui Guo, Ph.D., CRP

Huairui Guo is the Director of Theoretical Development at ReliaSoft Corporation. He received his Ph.D. in Systems and Industrial Engineering from the University of Arizona. He has published numerous papers in the areas of quality engineering including SPC, ANOVA and DOE and reliability engineering. His current research interests include repairable system modeling, accelerated life/degradation Testing, warranty data analysis and robust optimization. Dr. Guo is a member of SRE, IIE and IEEE. He is a Certified Reliability Professional (CRP).

# Adamantios Mettas, CRP

Adamantios Mettas is the VP of Product Development at ReliaSoft Corporation, spearheading ReliaSoft's innovative product development group. In his previous position as ReliaSoft's Senior Scientist, he played a key role in the development of ReliaSoft's software including Weibull++, ALTA, RGA and BlockSim by deriving statistical formulations and models and developing numerical methods for solving them. Mr. Mettas has trained more than 2,000 engineers throughout the world on the subjects of Life Data Analysis, FMEA, Warranty Analysis, System Reliability, RCM, DOE and Design for Reliability, as well as advanced topics such as Accelerated Testing, Repairable Systems and Reliability Growth. He has published numerous papers on these topics. In addition to training, Mr. Mettas is part of ReliaSoft's consulting arm and has been involved in various projects across different industries, including Oil & Gas, Power Generation, Automotive, Semiconductor, Defense and Aerospace. Mr. Mettas holds an M.S. in Reliability Engineering from the University of Arizona and he is a Certified Reliability Professional (CRP).

# Table of Contents

1. Introduction ..... 1
2. Statistical Background ..... 2
3. Two Level Factorial Design ..... 3
4. Response Surface Methods (RSM) ..... 6
5. DOE for Life Testing ..... 9
6. Conclusions ..... 10
7. References ..... 11
8. Tutorial Visuals ..... 12

## 1. INTRODUCTION

The most effective way to improve product quality and reliability is to integrate them in the design and manufacturing process. Design of Experiments (DOE) is a useful tool that can be integrated into the early stages of the development cycle. It has been successfully adopted by many industries, including automotive, semiconductor, medical devices, chemical products, etc. The application of DOE is not limited to engineering. Many successful stories can be found in other areas. For example, it has been used to reduce administration costs, improve the efficiency of surgery processes, and establish better advertisement strategies.

### *1.1 Why DOE*

DOE will make your life easier. For many engineers, applying DOE knowledge in their daily work will reduce lots of trouble. Here are two examples of bad experiments that will cause trouble.

*Example 1*: Assume the reliability of a product is affected by voltage. The usage level voltage is 10. In order to predict the reliability at the usage level, fifty units are available for accelerated life testing. An engineer tested all fifty units at a voltage of 25. Is this a good test?

*Example 2*: Assume the reliability of a product is affected by temperature and humidity. The usage level is 40 degrees Celsius and 50% relative humidity. In order to predict the reliability at the usage level, fifty units are available for accelerated life testing. The design is conducted in the following way:

| Number of Units | Temperature (Celsius) | Humidity (%) |
|---|---|---|
| 25 | 120 | 95 |
| 25 | 85 | 85 |

*Table 1 – Two Stress Accelerated Life Test*

Will the engineer be able to predict the reliability at the usage level with the failure data from this test?

### *1.2 What DOE Can Do*

DOE can help you design better tests than the above two examples. Based on the objectives of the experiments, DOE can be used for the following purposes [1, 2]:

1. *Comparisons*. When you have multiple design options, several materials or suppliers are available, you can design an experiment to choose the best one. For example, in the comparison of six different suppliers that provide connectors, will the components have the same expected life? If they are different, how are they different and which is the best?
2. *Variable Screening.* If there are a large number of variables that can affect the performance of a product or a system, but only a relatively small number of them are important, a screening experiment can be conducted to identify the important variables. For example, the warranty return is abnormally high after a new product is launched. Variables that may affect the life are temperature, voltage, duty cycle, humidity and several other factors. DOE can be used to quickly identify the troublemakers and a follow-up experiment can provide the guidelines for design modification to improve the reliability.
3. *Transfer Function Exploration.* Once a small number of variables have been identified as important, their effects on the system performance or response can be further explored. The relationship between the input variables and output response is called the *transfer function*. DOE can be applied to design efficient experiments to study the linear and quadratic effects of the variables and some of the interactions between the variables.
4. *System Optimization.* The goal of system design is to improve the system performance, such as to improve the efficiency, quality, and reliability. If the transfer function between variables and responses has been identified, the transfer function can be used for design optimization. DOE provides an intelligent sequential strategy to quickly move the experiment to a region containing the optimum settings of the variables.
5. *System Robustness.* In addition to optimizing the response, it is important to make the system robust against "noise," such as environmental factors and uncontrolled factors. Robust design, one of the DOE techniques, can be used to achieve this goal.

### *1.3 Common Design Types*

Different designs have been used for different experiment purposes. The following list gives the commonly used design types.

1. For comparison
   - One factor design
2. For variable screening
   - 2 level factorial design
   - Taguchi orthogonal array
   - Plackett-Burman design
3. For transfer function identification and optimization
   - Central composite design
   - Box-Behnken design
4. For system robustness
   - Taguchi robust design

The designs used for transfer function identification and optimization are called *Response Surface Method* designs. In this tutorial, we will focus on *2 level factorial design* and *response surface method designs*. They are the two most popular and basic designs.

### *1.4 General Guidelines for Conducting DOE*

DOE is not only a collection of statistical techniques that enable an engineer to conduct better experiments and analyze data efficiently; it is also a philosophy. In this section, general guidelines for planning efficient experiments will be given. The following seven-step procedure should be followed [1, 2].

*1. Clarify and State Objective*. The objective of the experiment should be clearly stated. It is helpful to prepare a list of specific problems that are to be addressed by the experiment.

*2. Choose Responses*. Responses are the experimental outcomes. An experiment may have multiple responses based on the stated objectives. The responses that have been chosen should be measurable.

*3. Choose Factors and Levels.* A factor is a variable that is going to be studied through the experiment in order to understand its effect on the responses. Once a factor has been selected, the value range of the factor that will be used in the experiment should be determined. Two or more values within the range need to be used. These values are referred to as *levels* or *settings*. Practical constraints of treatments must be considered, especially when safety is involved. A cause-and-effect diagram or a fishbone diagram can be utilized to help identify factors and determine factor levels.

*4. Choose Experimental design.* According to the objective of the experiments, the analysts will need to select the number of factors, the number of level of factors, and an appropriate design type. For example, if the objective is to identify important factors from many potential factors, a screening design should be used. If the objective is to optimize the response, designs used to establish the factor-response function should be planned. In selecting design types, the available number of test samples should also be considered.

*5. Perform the Experiment.* A design matrix should be used as a guide for the experiment. This matrix describes the experiment in terms of the actual values of factors and the test sequence of factor combinations. For a hard-to-set factor, its value should be set first. Within each of this factor's settings, the combinations of other factors should be tested.

*6. Analyze the Data.* Statistical methods such as regression analysis and ANOVA (Analysis of Variance) are the tools for data analysis. Engineering knowledge should be integrated into the analysis process. Statistical methods cannot prove that a factor has a particular effect. They only provide guidelines for making decisions. Statistical techniques together with good engineering knowledge and common sense will usually lead to sound conclusions. Without common sense, pure statistical models may be misleading. For example, models created by smart Wall Street scientists did not avoid, and probably contributed to, the economic crisis in 2008.

*7. Draw Conclusions and Make Recommendations.* Once the data have been analyzed, practical conclusions and recommendations should be made. Graphical methods are often useful, particularly in presenting the results to others. Confirmation testing must be performed to validate the conclusion and recommendations.

The above seven steps are the general guidelines for performing an experiment. A successful experiment requires knowledge of the factors, the ranges of these factors and the appropriate number of levels to use. Generally, this information is not perfectly known before the experiment. Therefore, it is suggested to perform experiments iteratively and sequentially. It is usually a major mistake to design a single, large, comprehensive experiment at the start of a study. As a general rule, no more than 25 percent of the available resources should be invested in the first experiment.

## 2. STATISTICAL BACKGROUND

Linear regression and ANOVA are the statistical methods used in DOE data analysis. Knowing them will help you have a better understand of DOE.

### 2.1 Linear Regression[2]

A general linear model or a multiple regression model is:

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + \varepsilon \qquad (1)$$

Where: $Y$ is the response also called output or dependent variable. $X_i$ is the predictor also called input or independent variable. $\varepsilon$ is the random error or noise, which is assumed to be normally distributed with mean 0 and variance $\sigma^2$, usually noted as $\varepsilon \sim N(0, \sigma^2)$. Because $\varepsilon$ is normally distributed, then for a given value of *X*, *Y* is also normally distributed and $Var(Y) = \sigma^2$.

From the model, it can be seen that the variation or the difference of *Y* consists of two parts. One is the random part of $\varepsilon$. The other is the difference caused by the difference of the *X* values. For example, consider the data in Table 2:

| Observation | $X_1$ | $X_2$ | $Y$ | $Y$ Mean |
|---|---|---|---|---|
| 1 | 120 | 90 | 300 | 325 |
| 2 | 120 | 90 | 350 | |
| 3 | 85 | 95 | 150 | 170 |
| 4 | 85 | 95 | 190 | |
| 5 | 120 | 95 | 400 | 415 |
| 6 | 120 | 95 | 430 | |

*Table 2 – Example Data for Linear Regression*

Table 2 has three different combinations of $X_1$ and $X_2$, showing at different colors. For each combination, there are two observations. Because of the randomness caused by $\varepsilon$, these two observations are different although they have the same *X* values. This difference usually is called *within-run variation*. The mean values of *Y* at the three combinations are different too. This difference is caused by the difference of $X_1$ and $X_2$ and usually is called *between-run variation*.

If the between-run variation is significantly larger than the within-run variation, it means most of the variation of *Y* is caused by the difference of *X* settings. In other words, *X*s significantly affect *Y*. The difference of *Y*s caused by the *X*s are much more significant than the difference caused by the noise.

From Table 2, we have the feeling that the between-run variation is larger than the within-run variation. To confirm this, statistical methods should be applied. The amount of the total variation of *Y* is defined by the *sum of squares*:

$$SS_T = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 \qquad (2)$$

Where $Y_i$ is the *i*th observed value and $\bar{Y}$ is the mean of all the observations. However, since $SS_T$ is affected by the number of observations, to eliminate this effect, another metric called *mean squares* is used to measure the normalized

variability of *Y*.

$$MS_T = \frac{SS_T}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 \quad (3)$$

Equation (3) is also the unbiased estimator of $Var(Y)$.

As mentioned before, the total sum of squares can be partitioned into two parts: within-run variation caused by random noise (called sum of squares of error $SS_E$) and the between-run variation caused by different values of *X*s (called sum of squares of regression $SS_R$).

$$SS_T = SS_R + SS_E = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad (4)$$

Where: $\hat{Y}_i$ is the predicted value for the *i*th test. For tests with the same *X* values, the predicted values are the same.

Similar to equation (3), the mean squares of regression and the mean squares of error are calculated by:

$$MS_R = \frac{SS_R}{p} = \frac{1}{p}\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 \quad (5)$$

$$MS_E = \frac{SS_E}{n-1-p} = \frac{1}{n-1-p}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \quad (6)$$

Where: *p* is the number of *Xs*.

When there is more than one input variable, $SS_R$ can be further divided into the variation caused by each variable, such as:

$$SS_R = SS_{X_1} + SS_{X_2} + ... + SS_{X_p} \quad (7)$$

The mean squares of each input variable $MS_{X_i}$ is compared with $MS_E$ to test if the effect of $X_i$ is significantly greater than the effect of noise.

The mean squares of regression $MS_R$ is used to measure the between-run variance that is caused by predictor *X*s. The mean squares of error $MS_E$ represents the within-run variance caused by noise. By comparing these two values, we can find out if the variance contributed by *X*s is significantly greater than the variance caused by noise. ANOVA is the method used for the comparison in a statistical way.

### *2.2 ANOVA (Analysis of Variance)*

The following ratio

$$F_0 = \frac{MS_R}{MS_E} \quad (8)$$

is used to test the following two hypotheses:

*H0*: *There is no difference between the variance caused by Xs and the variance caused by noise*.

*H1*: *The variance caused by Xs is larger than the variance caused by noise.*

Under the null hypothesis, the ratio follows the *F* distribution with degree of freedoms of $p$ and $n-1-p$. By applying ANOVA to the data for this example, we get the following ANOVA table.

The third column shows the values for the sum of squares. We can easily verify that:

$$SS_T = SS_R + SS_E = SS_{X_1} + SS_{X_2} + SS_E \quad (9)$$

The fifth column shows the *F* ratio of each source. All the values are much bigger than 1. The last column is the *P* value. The smaller the *P* value is, the larger the difference between the variance caused by the corresponding source and the variance caused by noise. Usually, a significance level $\alpha$, such as 0.05 or 0.1 is used to compare with the *P* values. If a *P* value is less than $\alpha$, the corresponding source is said to be *significant* at the significance level of $\alpha$. From Table 3, we can see that both variables $X_1$ and $X_2$ are significant to the response *Y* at the significance level of 0.1.

| Source of Variation | Degrees of Freedom | Sum of Squares | Mean Squares | F Ratio | P Value |
|---|---|---|---|---|---|
| Model | 2 | 6.14E+04 | 3.07E+04 | 36.86 | 0.0077 |
| X1 | 1 | 6.00E+04 | 6.00E+04 | 72.03 | 0.0034 |
| X2 | 1 | 8100 | 8100 | 9.72 | 0.0526 |
| Residual | 3 | 2500 | 833.3333 | | |
| Total | 5 | 6.39E+04 | | | |

*Table 3 – ANOVA Table for the Linear Regression Example*

Another way to test whether or not a variable is significant is to test whether or not its coefficient is 0 in the regression model. For this example, the linear regression model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_1 X_2 + \varepsilon \quad (10)$$

If we want to test whether or not $\beta_1 = 0$, we can use the following hypothesis:

$$H0: \ \beta_1 = 0$$

Under this null hypothesis, the statistic is a *t* distribution:

$$T_0 = \frac{\beta_1}{se(\beta_1)} \quad (11)$$

$Se(\beta_1)$ is the standard error of $\beta_1$ that is estimated from the data. The *t* test results are given in Table 4.

| Term | Coefficient | Standard Error | T Value | P Value |
|---|---|---|---|---|
| Intercept | -2135 | 588.7624 | -3.6263 | 0.0361 |
| $X_1$ | 7 | 0.8248 | 8.487 | 0.0034 |
| $X_2$ | 18 | 5.7735 | 3.1177 | 0.0526 |

*Table 4 – Coefficients for the Linear Regression Example*

Table 3 and Table 4 give the same *P* values.

With linear regression and ANOVA in mind, we can start discussing DOE now.

## *3. TWO LEVEL FACTORIAL DESIGNS*

Two level factorial designs are used for factor screening. In order to study the effect of a factor, at least two different settings for this factor are required. This also can be explained from the viewpoint of linear regression. To fit a line, two points are the minimal requirement. Therefore, the engineer who tested all the units at a voltage of 25 will not be able to predict the life at the usage level of 10 volts. With only one voltage value, the effect of voltage cannot be evaluated.

### 3.1 Two Level Full Factorial Design

When the number of factors is small and you have the resources, a full factorial design should be conducted. Here we will use a simple example to introduce some basic concepts in DOE.

For an experiment with two factors, the factors usually are called A and B. Uppercase letters are used for factors. The first level or the low level is represented by -1, while the second level or the high level is represented by 1. There are four combinations of a 2 level 2 factorial design. Each combination is called a *treatment*. Treatments are represented by lowercase letters. The number of test units for each treatment is called the number of *replicates*. For example, if you test two samples at each treatment, the number of replicates is 2. Since the number of replicates for each factor combination is the same, this design is also *balanced*. A two level factorial design with *k* factors usually is written as $2^k$ design and read as "2 to the power of 3 design" or "2 to the 3 design." For a $2^2$ design, the design matrix is:

| Treatment | Factors | | Response |
|---|---|---|---|
| | A | B | |
| *-1* | -1 | -1 | 20 |
| *a* | 1 | -1 | 30 |
| *b* | -1 | 1 | 25 |
| *ab* | 1 | 1 | 35 |

*Table 5 – Treatments for 2 Level Factorial Design*

This design is *orthogonal*. This is because the sum of the product of A and B is zero, which is $(-1\times-1)+(1\times-1)+(-1\times1)+(1\times1)=0$. An orthogonal design will reduce the estimation uncertainty of the model coefficients.

The following linear regression model is used for the analysis:

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_{12}X_1X_2+\varepsilon \qquad (12)$$

Where: $X_1$ is for factor A; $X_2$ is for factor B; and their interaction is represented by $X_1X_2$. The effects of A and B are called *main effects*. The effects of their interaction are called *two-way interaction effects*. These three effects are the three sources for the variation of *Y*. Since equation (12) is a linear regression model, the ANOVA method and the *t*-test given in Section 2 can be used to test whether or not one effect is significant.

For a balanced design, a simple way to calculate the effect of a factor is to calculate the difference of the mean values of the response at its high and low setting. For example, the effect of A can be calculated by:

$$\text{Effet of A} = \text{Avg. at A}_{\text{high}} - \text{Avg. at A}_{\text{low}}$$
$$=\frac{30+35}{2}-\frac{20+25}{2}=10 \qquad (13)$$

### 3.2 Two Level Fractional Factorial Design

When you increase the number of factors, the number of test units will increase quickly. For example, to study 7 factors, 128 units are needed. In reality, responses are affected by a small number of main effects and lower order interactions. Higher order interactions are relatively unimportant. This statement is called the *sparsity of effects principle*. According to this principle, fractional factorial designs are developed. These designs use fewer samples to estimate main effects and lower order interactions, while the higher order interactions are considered to have negligible effects.

Consider a $2^3$ design. 8 test units are required for a full factorial design. Assume only 4 test units are available because of the cost of the test. Which 4 of the 8 treatments should you choose? A full design matrix with all the effects for a $2^3$ design is:

| Order | A | B | AB | C | AC | BC | ABC |
|---|---|---|---|---|---|---|---|
| 1 | -1 | -1 | 1 | -1 | 1 | 1 | -1 |
| 2 | 1 | -1 | -1 | -1 | -1 | 1 | 1 |
| 3 | -1 | 1 | -1 | -1 | 1 | -1 | 1 |
| 4 | 1 | 1 | 1 | -1 | -1 | -1 | -1 |
| 5 | -1 | -1 | 1 | 1 | -1 | -1 | 1 |
| 6 | 1 | -1 | -1 | 1 | 1 | -1 | -1 |
| 7 | -1 | 1 | -1 | 1 | -1 | 1 | -1 |
| 8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*Table 6 – Design Matrix for a $2^3$ Design*

If the effect of ABC can be ignored, the following 4 treatments can be used in the experiment.

| Order | A | B | AB | C | AC | BC | ABC |
|---|---|---|---|---|---|---|---|
| 2 | 1 | -1 | -1 | -1 | -1 | 1 | 1 |
| 3 | -1 | 1 | -1 | -1 | 1 | -1 | 1 |
| 5 | -1 | -1 | 1 | 1 | -1 | -1 | 1 |
| 8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*Table 7 –Fraction of the Design Matrix for a $2^3$ Design*

In Table 7, the effect of ABC cannot be estimated from the experiment because it is always at the same level of 1. Since Table 7 uses only half of the treatments from the full factorial design in Table 6, it is represented by $2^{3-1}$ and read as "2 to the power of 3 minus 1 design" or "2 to the 3 minus 1 design."

From Table 7, you will also notice that some columns have the same values. For example, column AB and C are the same. Using equation (13) to calculate the effect of AB and C, we will end up with the same procedure and result. Therefore, from this experiment, the effect of AB and C cannot be distinguished because they change with the same pattern. In DOE, effects that cannot be separated from an experiment are called *confounded effects* or *aliased effects*. A list of aliased effects is called the *alias structure*. For the design of Table 7, the alias structure is:

*[I]=I+ABC; [A]=A+AC; [B]=B+BC; [C]=C+AB*

Where: *I* is the effect of the intercept in the regression model, which represents the mean value of the response. The alias for

I is called the *defining relation*. For this example, the defining relation is written as $I$ = ABC. In a design, $I$ may be aliased with several effects. The order of the shortest effects that aliased with $I$ is the "*resolution*" of this design.

From the alias structure, we can see that main effects are confounded with 2-way interactions. For example, the estimated effect for C in fact is the combination effect of C and AB.

Checking Table 7, we can see it is a full factorial design if we have only factor A and B. Therefore, A and B usually are called *basic factors* and the full factorial design for them is called the *basic design*. A fractional factorial design is generated from its basic design and basic factor. For this example, the values for factor C are generated from the values of the basic factors A and B using the relation C=AB. Usually AB is called the *factor generator* for C.

By now, it should be clear that the design given in Table 1 at the beginning of this tutorial is not a good design. If you check the design in terms of coded value, the answer is obvious. Table 8 shows the design again.

| Number of Units | Temperature (Celsius) | Humidity (%) |
|---|---|---|
| 25 | 120 (1) | 95 (1) |
| 25 | 85 (-1) | 85 (-1) |

*Table 8 – Two Stress Accelerated Life Test (Coded Value)*

In this design, temperature and humidity are confounded. In fact, to study the effect of two factors, at least three different settings are required. From the linear regression point of view, at least three unique settings are needed to solve three parameters: the intercept, the effect of factor A and the effect of factor B. If their interaction is also to be estimated, four different settings should be used. Many DOE software packages can generate a design matrix for you according to the number of factors and the level of factors. It will help you avoid bad designs such as the one given in Table 8.

### *3.3 An Example of a Fractional Factorial Design*

Assume an engineer wants to identify the factors that affect the yield of a manufacturing process for integrated circuits. By following the DOE guidelines, five factors are brought up and a two level fractional factorial design is decided to be used [1]. The five factors and their levels are given in Table 9.

| Factor | Name | Unit | Low (-1) | High (1) |
|---|---|---|---|---|
| A | Aperture Setting | | small | large |
| B | Exposure Time | minutes | 20 | 40 |
| C | Develop Time | seconds | 30 | 45 |
| D | Mask Dimension | | small | large |
| E | Etch Time | minutes | 14.5 | 15.5 |

*Table 9 –Factor Settings for a Five Factor Experiment*

With five factors the total number of runs required for a full factorial is $2^5 = 32$. Running all of the 32 combinations is too expensive for the manufacturer. At the initial investigation, only main effects and two factor interactions are of interest, while higher order interactions are considered to be unimportant. It is decided to carry out the investigation using the $2^{5-1}$ design, which requires 16 runs. The defining relation is I=ABCDE, or in other words, the generator for factor E is E=ABCD. Table 10 gives the experiment data.

| Run Order | A | B | C | D | E | Yield |
|---|---|---|---|---|---|---|
| 1 | Large | 20 | 30 | Large | 15.5 | 10 |
| 2 | Large | 20 | 45 | Large | 14.5 | 21 |
| 3 | Small | 40 | 45 | Small | 15.5 | 45 |
| 4 | Small | 20 | 45 | Small | 14.5 | 16 |
| 5 | Large | 40 | 30 | Small | 15.5 | 52 |
| 6 | Large | 40 | 45 | Small | 14.5 | 60 |
| 7 | Small | 40 | 30 | Large | 15.5 | 30 |
| 8 | Small | 20 | 45 | Large | 15.5 | 15 |
| 9 | Large | 20 | 30 | Small | 14.5 | 9 |
| 10 | Small | 40 | 30 | Small | 14.5 | 34 |
| 11 | Small | 20 | 30 | Small | 15.5 | 8 |
| 12 | Large | 40 | 30 | Large | 14.5 | 50 |
| 13 | Small | 40 | 45 | Large | 14.5 | 44 |
| 14 | Small | 20 | 30 | Large | 14.5 | 6 |
| 15 | Large | 20 | 45 | Small | 15.5 | 22 |
| 16 | Large | 40 | 45 | Large | 15.5 | 63 |

*Table 10 –Design Matrix and Results*

Since the design has only 16 unique factor combinations, it can be used to estimate only 16 parameters in the linear regression model. If we include all main and 2-way interactions in the model, we get the following ANOVA table.

| Source of Variation | DF | Sum of Squares | Mean Squares | F | P Value |
|---|---|---|---|---|---|
| Model | 15 | 5775.4375 | 385.0292 | - | - |
| A | 1 | 495.0625 | 495.0625 | - | - |
| B | 1 | 4590.0625 | 4590.0625 | - | - |
| C | 1 | 473.0625 | 473.0625 | - | - |
| D | 1 | 3.0625 | 3.0625 | - | - |
| E | 1 | 1.5625 | 1.5625 | - | - |
| AB | 1 | 189.0625 | 189.0625 | - | - |
| AC | 1 | 0.5625 | 0.5625 | - | - |
| AD | 1 | 5.0625 | 5.0625 | - | - |
| AE | 1 | 5.0625 | 5.0625 | - | - |
| BC | 1 | 1.5625 | 1.5625 | - | - |
| BD | 1 | 0.0625 | 0.0625 | - | - |
| BE | 1 | 0.0625 | 0.0625 | - | - |
| CD | 1 | 3.0625 | 3.0625 | - | - |
| CE | 1 | 0.5625 | 0.5625 | - | - |
| DE | 1 | 7.5625 | 7.5625 | - | - |
| Residual | 0 | | | | |
| Total | 15 | 5775.4375 | | | |

*Table 11 –ANOVA Table with All Effects*

There are no F ratio and P values in the above table. This is because there are no replicates in this experiment when all the effects are considered. Therefore, there is no way to estimate the error term in the regression model. This is why the SSE (Sum of Squares of Error), labeled as *Residual* in Table 11, is 0. Without SSE, the estimation of the random error, how can we test whether or not an effect is significant compared to random error? Don't panic. Statisticians have already developed methods to deal with this situation. When there is no error in a screening experiment, Lenth's method can be used to identify significant effects. Lenth's method assumes that all the effects should be normally distributed with a mean of 0, given the hypothesis that they are not significant. If any effects are significantly different from 0, they should be considered significant. So we can check the normal probability plot for effects.

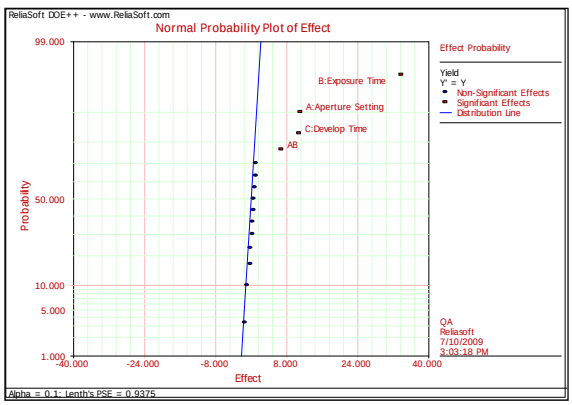


*Figure 1-Effect Probability Plot Using Lenth's Method*

From Figure 1, the main effect A, B, C and the 2-way interaction AB are identified as significant at a significance level of 0.1. Since the rest of the effects are not significant, they can be treated as noise and used to estimate the sum of squares of error. In DOE, it is a common practice to pool non-significant effects into error. With only A, B, C and AB in the model, we get the following ANOVA table.

| Source of Variation | DF | Sum of Squares | Mean Squares | F Ratio | P Value |
|---|---|---|---|---|---|
| **Model** | 4 | 5747.25 | 1436.8125 | 560.7073 | 1.25E-12 |
| A | 1 | 495.0625 | 495.0625 | 193.1951 | 2.53E-08 |
| B | 1 | 4590.0625 | 4590.0625 | 1791.244 | 1.56E-13 |
| C | 1 | 473.0625 | 473.0625 | 184.6098 | 3.21E-08 |
| AB | 1 | 189.0625 | 189.0625 | 73.7805 | 3.30E-06 |
| Residual | 11 | 28.1875 | 2.5625 | | |
| Lack of Fit | 3 | 9.6875 | 3.2292 | 1.3964 | 0.3128 |
| Pure Error | 8 | 18.5 | 2.3125 | | |
| Total | 15 | 5775.4375 | | | |

*Table 12 –ANOVA Table with Significant Effects*

From Table 12, we can see that effects A, B, C and AB are indeed significant because their P values are close to 0.

Once the important factors have been identified, follow-up experiments can be conducted to optimize the process. *Response Surface Methods* are developed for this purpose.

## 4. *RESPONSE SURFACE METHODS (RSM)*

Response surface methods (RSM) are used to estimate the transfer functions at the optimal region. The estimated function is then used to optimize the responses. The quadratic model is the model used in RSM. Similar to the factorial design, linear regression and ANOVA are the tools for data analysis in RSM. Let's use a simple example to illustrate this type of design.

### *4.1 Initial Investigation*

Assume the yield from a chemical process has been found to be affected by two factors [1]:

- Reaction Temperature
- Reaction Time

The current operating conditions of 230 Fahrenheit and 65 minutes give a yield of about 35%. The engineers decide to explore the current conditions in the range [L=225, H=235] Fahrenheit and [L=55, H=75] minutes to see how the temperature and time affect the yield. The design matrix is:

| Std. Order | Point Type | A: Temperature | B: Reaction Time | Yield (%) |
|---|---|---|---|---|
| 1 | 1 | -1 | -1 | 33.95 |
| 2 | 1 | 1 | -1 | 36.35 |
| 3 | 1 | -1 | 1 | 35 |
| 4 | 1 | 1 | 1 | 37.25 |
| 5 | 0 | 0 | 0 | 35.45 |
| 6 | 0 | 0 | 0 | 35.75 |
| 7 | 0 | 0 | 0 | 36.05 |
| 8 | 0 | 0 | 0 | 35.3 |
| 9 | 0 | 0 | 0 | 35.9 |

*Table 13-Design Matrix for the Initial Experiment*

Table 13 is the design matrix in coded values where -1 is the lower level and 1 is the higher level. For a given actual value for a numerical factor, its corresponding coded value can be calculated by:

$$\text{Coded value} = \frac{\text{Actual Value - Middle Value of High and Low}}{\text{Half of the Range Between High and Low}} \quad (13)$$

In Table 13, there are several replicated runs at the setting of (0, 0). These runs are called *center points*. Center points have the following two uses.

- To estimate random error.
- To check whether or not the curvature is significant.

At least five center points are suggested in an experiment.

When we analyze the data in Table 13, we get the ANOVA table below.

From Table 14, we know curvature is not significant. Therefore, the linear model is sufficient in the current experiment space. The linear model that includes only significant effects is:

$$Y = 35.6375 + 1.1625X_1 + 0.4875X_2 \quad (14)$$

Equation (14) is in terms of coded value. We can see both factors have a positive effect to the yield since their coefficients are positive. To improve the yield, we should explore the region with factor values larger than the current operation condition. There are many directions we can move to increase the yield, but which one is the fastest lane to approach the optimal region? By checking the coefficient of each factor in equation (14), we know the direction should be (1.1625, 0.4875). This also can be seen from the contour plot in Figure 2.

| Source of Variation | DF | Sum of Squares | Mean Squares | F Ratio | P Value |
|---|---|---|---|---|---|
| Model | 4 | 6.368 | 1.592 | 16.4548 | 0.0095 |
| A | 1 | 5.4056 | 5.4056 | 55.8721 | 0.0017 |
| B | 1 | 0.9506 | 0.9506 | 9.8256 | 0.035 |
| AB | 1 | 0.0056 | 0.0056 | 0.0581 | 0.8213 |
| Curvature | 1 | 0.0061 | 0.0061 | 0.0633 | 0.8137 |
| Residual | 4 | 0.387 | 0.0967 | | |
| Total | 8 | 6.755 | | | |

*Table 14-ANOVA Table for the Initial Experiment*

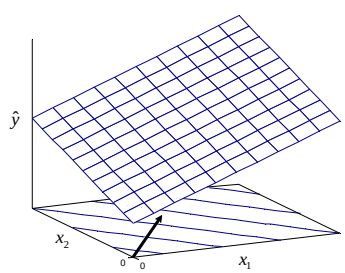


*Figure 2--Contour Plot of the Initial Experiment*

| Step | Factor Levels | | | | Yield (%) |
|---|---|---|---|---|---|
| | Coded | | Actual | | |
| | **A** | **B** | **A** | **B** | |
| Current Operation | 0 | 0 | 230 | 65 | 35 |
| 1 | 2.4 | 1 | 242 | 75 | 36.5 |
| 2 | 4.8 | 2 | 254 | 85 | 39.35 |
| 3 | 7.2 | 3 | 266 | 95 | 45.65 |
| 4 | 9.6 | 4 | 278 | 105 | 49.55 |
| 5 | 12 | 5 | 290 | 115 | 55.7 |
| 6 | 14.4 | 6 | 302 | 125 | 64.25 |
| 7 | 16.8 | 7 | 314 | 135 | 72.5 |
| 8 | 19.2 | 8 | 326 | 145 | 80.6 |
| 9 | 21.6 | 9 | 338 | 155 | 91.4 |
| **10** | **24** | **10** | **350** | **165** | **95.45** |
| 11 | 26.4 | 11 | 362 | 175 | 89.3 |
| 12 | 28.8 | 12 | 374 | 185 | 87.65 |

*Table 15-Path of Steepest Ascent*

From Figure 2, we know the fastest lane to increase yield is to move along the direction that is perpendicular to the contour lines. This direction is (1.1625, 0.4875), or about (2.4, 1) in terms of normalized scale. Therefore, if 1 unit of $X_2$ is increased, 2.39 units of $X_1$ should be used in order to keep moving on the *steepest ascent direction*. To convert the code values to the actual values, we should use the step size of (12 degree, 10 mins) for factor A and B. The table above gives the results for the experiments conducted along the *path of steepest ascent*.

From Table 15, it can be seen that at step 10, the factor setting is close to the optimal region. This is because the yield decreases on either side of this step. The region around setting of (350, 165) requires further investigation. Therefore, the analysts will conduct a $2^2$ factorial design with the center point of (350, 165) and the range of [L=345, H=355] for factor A (temperature) and [L=155, H=175] for factor B (reaction time). The design matrix is given in Table 16.

| Std. Order | Point Type | A:Temperature (F) | B:Reaction Time (min) | Yield (%) |
|---|---|---|---|---|
| 1 | 1 | 345 | 155 | 89.75 |
| 2 | 1 | 355 | 155 | 90.2 |
| 3 | 1 | 345 | 175 | 92 |
| 4 | 1 | 355 | 175 | 94.25 |
| 5 | 0 | 350 | 165 | 94.85 |
| 6 | 0 | 350 | 165 | 95.45 |
| 7 | 0 | 350 | 165 | 95 |
| 8 | 0 | 350 | 165 | 94.55 |
| 9 | 0 | 350 | 165 | 94.7 |

*Table 16-Factorial Design around the Optimal Region*

The ANOVA table for this data is:

| Source of Variation | DF | Sum of Squares | Mean Squares | F Ratio | P Value |
|---|---|---|---|---|---|
| Model | 4 | 37.643 | 9.4107 | 78.916 | 5E-04 |
| A | 1 | 1.8225 | 1.8225 | 15.283 | 0.017 |
| B | 1 | 9.9225 | 9.9225 | 83.208 | 8E-04 |
| AB | 1 | 0.81 | 0.81 | 6.7925 | 0.06 |
| Curvature | 1 | 25.088 | 25.088 | 210.38 | 1E-04 |
| Residual | 4 | 0.477 | 0.1193 | | |
| Total | 8 | 38.12 | | | |

*Table 17-ANOVA for the Experiment at the Optimal Region*

Table 17 shows curvature is significant at this experiment region. Therefore, the linear model is not enough for the relationship between factors and response. A quadratic model should be used instead. An experiment design that is good for the quadratic model should be used for further investigation. Central Composite Design (CCD) is one of these designs.

### *4.2 Optimization Using RSM*

Table 17 is the 2-level factorial design at the optimal region. CCD is build based on this factorial design and used to estimate the parameters for a quadratic model such as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon \quad (15)$$

In fact, a CCD can be directly augmented from a regular factorial design. The augmentation process is illustrated in Figure 3.

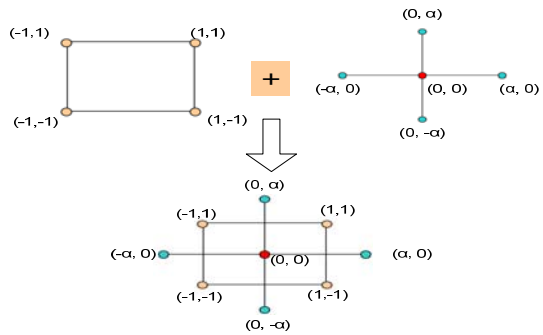


*Figure 3-Augmnent a Factorial Design to CCD*

Points outside the rectangle in Figure 3 are called *axial points* or *start points*. By adding several center points and axial points, a regular factorial design is augmented to a CCD. In Figure 3, we can see there are five different values for each factor. So CCD can be used to estimate the quadratic model in equation (15).

Several methods have been developed to calculate $\alpha$ to make the CCD have special properties such that the designed experiment can better estimate model parameters or can better explore the optimal region. The commonly used method to calculate $\alpha$ is:

$$\alpha = \left[ \frac{2^{k-f}(n_f)}{n_s} \right]^{1/4} \quad (16)$$

Where: $n_f$ is the number of replicates of the runs in the original factorial design. $n_s$ is the number of replicates of the runs at the axial points. $2^{k-f}$ represents the original factorial or fractional factorial design.

| Std. Order | Point Type | A:Temperature (F) | B:Reaction Time (min) | Yield (%) |
|---|---|---|---|---|
| 1 | 1 | 345 | 155 | 89.75 |
| 2 | 1 | 355 | 155 | 90.2 |
| 3 | 1 | 345 | 175 | 92 |
| 4 | 1 | 355 | 175 | 94.25 |
| 5 | 0 | 350 | 165 | 94.85 |
| 6 | 0 | 350 | 165 | 95.45 |
| 7 | 0 | 350 | 165 | 95 |
| 8 | 0 | 350 | 165 | 94.55 |
| 9 | 0 | 350 | 165 | 94.7 |
| 10 | -1 | 342.93 | 165 | 90.5 |
| 11 | -1 | 357.07 | 165 | 92.75 |
| 12 | -1 | 350 | 150.86 | 88.4 |
| 13 | -1 | 350 | 179.14 | 92.6 |

*Table 18-CCD around the Optimal Region*

We use equation (15) to calculate $\alpha$ for our example. So $\alpha = 1.414$. Since we already have five center points in the factorial design at the optimal region, we only need to add start points to have a CCD. The complete design matrix for the CCD is:shown in Table 18. The last 5 runs in the above table are added to the previous factorial design. The fitted quadratic model is:

$$Y = 94.91 + 0.74X_1 + 1.53X_2 - 1.52X_1^2 - 2.08X_2^2 + 0.45X_1X_2 \quad (16)$$

The ANOVA table for the quadratic model is:

| Source of Variation | DF | Sum of Squares | Mean Squares | F Ratio | P Value |
|---|---|---|---|---|---|
| **Model** | 5 | 65.0867 | 13.0173 | 91.3426 | 3.22E-06 |
| A | 1 | 4.3247 | 4.3247 | 30.3465 | 0.0009 |
| B | 1 | 18.7263 | 18.7263 | 131.4022 | 8.64E-06 |
| AB | 1 | 0.81 | 0.81 | 5.6838 | 0.0486 |
| AA | 1 | 16.0856 | 16.0856 | 112.8724 | 1.43E-05 |
| BB | 1 | 30.1872 | 30.1872 | 211.8234 | 1.73E-06 |
| Residual | 7 | 0.9976 | 0.1425 | | |
| Lack of Fit | 3 | 0.5206 | 0.1735 | 1.4551 | 0.3527 |
| Pure Error | 4 | 0.477 | 0.1193 | | |
| Total | 12 | 66.0842 | | | |

*Table 19-ANOVA Table for CCD at the Optimal Region*

As mentioned before, the model for CCD is used to optimize the process. Therefore, the accuracy of the model is very important. From Table 19, the Lack of Fit test, we see the P value is relatively large. It means the model can fit the data well. The Lack of Fit residual is the estimation of the variations of the terms that are not included in the model. If its amount is close to the pure error, which is the within-run variation, it can be treated as part of the noise. Another way to check the model accuracy is to check the residual plots.

Through the above diagnostic, we found the model is adequate and we can use it to identify the optimal factor settings. The optimal settings can be found easily by taking the derivative of each factor from equation (16) and setting them to 0. Many software packages can do optimization. The following results are from DOE++ from ReliaSoft [3].

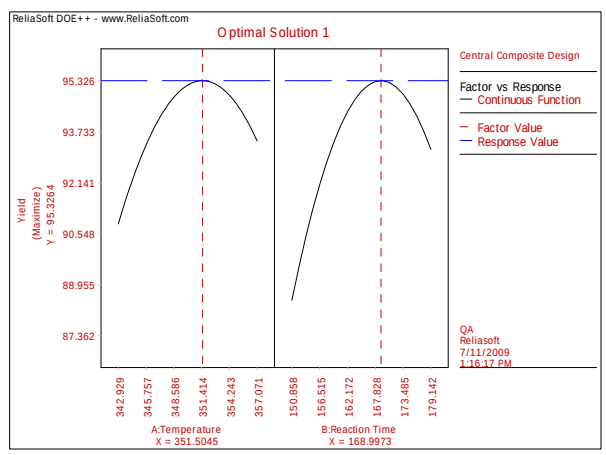


*Figure 4-Optimal Solution for the Chemical Process*

Figure 4 shows how the response changes with each

factor. The red dashed line points to the optimal value of each factor. we can see that the optimal settings are 351.5 degrees Fahrenheit for temperature and 169 minutes for the reaction time. At this setting, the predicted yield is 95.3%, which is much better than the yield at the current condition (about 35%).

## 5. DOE FOR LIFE TESTING

When DOE is used for life testing, the response is the life or failure time. However, because of the cost, time or other constraints, you may not have observed values of life for some test units. They are still functioning at the time when the test ends. The end time of the test is called the *suspension time* for the units that are not failed. Obviously, this time is not their "life." Should the suspension time be treated as the life time in order to analyze the data? In this section, we will discuss the correct statistical method for analyzing the data for life testing. First, let's explain some basic concepts in life data analysis.

### *5.1 Data Type for Life Test*

When the response is life, there are two types of data

- Complete Data
- Censored Data
  - Right Censored (Suspended)
  - Interval Censored

If a test unit is failed during the test and the exact failure time is known, the failure time is called *complete data*.

If a test unit is failed and you don't know the exact failure time -- instead, you know the failure occurs within a time range -- this time range is called *interval data*.

If a unit does not fail in the test, the end time of the test of the unit is called *right censored data* or *suspension data*.

Obviously, ignoring the censored data or treating them as failure times will underestimate the system reliability. However, in the use of the linear regression and ANOVA, an exact value for each observation is required. Therefore, engineers have to tweak the censored data in order to use linear regression and ANOVA. A simple way to tweak the data is to use the center point of the interval data as the failure time, and treat the suspension units as failed.

Even with the modification of the original data, another issue may still exist. In the using of linear regression and ANOVA, the response is assumed to be normally distributed. The $F$ and $T$ tests are established based on this normal distribution assumption. However, life time usually is not normally distributed.

Given the above reasons, correct analysis methods for data from life testing are needed.

### *5.2 Maximum Likelihood Estimation and Likelihood Ratio Test [2]*

Maximum Likelihood Estimation (MLE) can estimate model parameters to maximize the probability of the occurrence of an observed data set. It has been used successfully to handle different data types, such as complete data, suspensions and interval data. Therefore, we will use MLE to estimate the model parameters for life data from DOE.

Many distributions are used to describe lifetimes. The three most commonly used are [4]:

- Weibull distribution with probability density function (*pdf*):

$$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\mu}\right)^{\beta}} \tag{17}$$

- Lognormal distribution with *pdf*:

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^{\beta}} \tag{18}$$

- Exponential distribution with *pdf*:

$$f(t) = \frac{1}{m} e^{-\left(\frac{t}{m}\right)} \tag{19}$$

Assume there is only one factor (in the language of DOE), or stress (in the language of accelerated life testing) that affects the lifetime of the product. The life distribution and factor relationship can be described using the following graph.

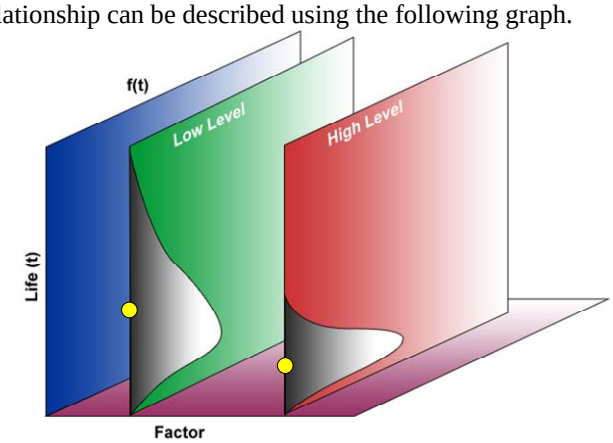


*Figure 5-pdf at Different Stress/Factor Levels*

Figure 5 shows that life decreases when a factor is changed from the low level to the high level. The *pdf* curves have the same shape while only the scale of the curve changes. The scale of the *pdf* is compressed at the high level. It means the failure mode remains the same, only the time of occurrence decreases at the high level. Instead of considering the entire scale of the *pdf*, a *life characteristic* can be chosen to represent the curve and used to investigate the effect of potential factors on life. The life characteristics for the three commonly used distributions are:

- Weibull distribution: $\eta$
- Lognormal distribution: $\mu$
- Exponential distribution: $m$

The life-factor relationship is studied to see how factors affect life characteristic. For example, a linear model can be used as the initial investigation for the relationship:

$$\mu' = \beta_0 + \beta_1 X_1 + \beta_1 X_2 + \ldots + \beta_{12} X_1 X_2 + \ldots \tag{20}$$

Where: $\mu' = \ln(\eta)$ for Weibull; $\mu' = \mu$ for lognormal and $\mu' = \ln(m)$ for exponential.

Please note that in equation (20) a logarithmic

transformation is applied to the life characteristics of the Weibull and exponential distributions. One of the reasons is because $\eta$ and $m$ can take only positive values.

To test whether or not one effect in equation (20) is significant, the likelihood ratio test is used:

$$LR(\text{effect } k) = -2\ln\frac{L(\text{effect k removed})}{L(\text{full model})} \tag{21}$$

Where $L()$ is the likelihood value. $LR$ follows a chi-squared distribution if effect $k$ is not significant.

If effect $k$ is not significant, whether or not it is removed from the full model of equation (20) will not affect the likelihood value much. It means the value of $LR$ will be close to 0. Otherwise, if the $LR$ value is very large, it means effect $k$ is significant.

### *5.3 Life Test Example*

Consider an experiment to improve the reliability of fluorescent lights [2]. Five factors A-E are investigated in the experiment. A $2^{5-2}$ design with factor generators D=AC and E=BC is conducted. The objective is to identify the significant effects that affect the reliability. Two replicates are used for each treatment. The test ends at the 20th day. Inspections are conducted every two days. The experiment results are given in Table 20.

| Run | A | B | C | D | E | Failure Time | |
|---|---|---|---|---|---|---|---|
| 1 | -1 | -1 | -1 | -1 | -1 | (14,16) | 20+ |
| 2 | -1 | -1 | 1 | 1 | 1 | (18,20) | 20+ |
| 3 | -1 | 1 | -1 | -1 | 1 | (8,10) | (10, 12) |
| 4 | -1 | 1 | 1 | 1 | -1 | (18,20) | 20+ |
| 5 | 1 | -1 | -1 | 1 | -1 | 20+ | 20+ |
| 6 | 1 | -1 | 1 | -1 | 1 | (12,14) | 20+ |
| 7 | 1 | 1 | -1 | 1 | 1 | (16,18) | 20+ |
| 8 | 1 | 1 | 1 | -1 | -1 | (12,14) | (14, 16) |

*Table 20- Data for the Life Test Example*

20+ means that the test unit was still working at the end of the test. So this experiment has suspension data. (14, 16) means that failure occurred at a time between the 14th and the 16th day. So this experiment also has interval data. The Weibull model is used as the distribution for the life of the fluorescent lights. The likelihood ratio test table is given below.

| Model | Effect | DF | Ln(LKV) | LR | P Value |
|---|---|---|---|---|---|
| Reduced | A | 1 | -20.7181 | 3.1411 | 0.0763 |
| | B | 1 | -24.6436 | 10.9922 | 0.0009 |
| | C | 1 | -19.2794 | 0.2638 | 0.6076 |
| | D | 1 | -25.7594 | 13.2237 | 0.0003 |
| | E | 1 | -21.0727 | 3.8504 | 0.0497 |
| Full | | 7 | -19.1475 | | |

*Table 21- LR Test Table for the Life Test Example*

Table 21 has a layout that is similar to the ANOVA table. This makes it easy to read for engineers who are familiar with ANOVA. From the P value column, we can see factor A, B, D and E are important to the product life. The estimated factor-life relationship is:

$$\ln(\eta) = 2.9959 + 0.1052A - 0.2256B - 0.0294C - 0.2477D + 0.1166E \tag{22}$$

The estimated shape parameter $\beta$ for the Weibull distribution is 7.27.

For comparison, the data was also analyzed using traditional linear regression and the ANOVA method. To apply linear regression and ANOVA, the data set was modified by using the center points as the failure time for interval data and treating the suspensions as failures. Results are given below.

| Source of Variation | DF | Sum of Squares | Mean Squares | F Ratio | P Value |
|---|---|---|---|---|---|
| Model | 5 | 143.3125 | 28.6625 | 4.2384 | 0.025 |
| A | 1 | 1.5625 | 1.5625 | 0.2311 | 0.6411 |
| B | 1 | 33.0625 | 33.0625 | 4.8891 | 0.0515 |
| C | 1 | 3.0625 | 3.0625 | 0.4529 | 0.5162 |
| D | 1 | 95.0625 | 95.0625 | 14.0573 | 0.0038 |
| E | 1 | 10.5625 | 10.5625 | 1.5619 | 0.2398 |
| Residual | 10 | 67.625 | 6.7625 | | |
| Total | 15 | 210.9375 | | | |

*Table 23- ANOVA Table for the Life Test Example*

In Table 23, only effects B and D are showing to be significant. The estimated linear regression model is:

$$Y = 16.9375 + 0.3125A - 1.4375B + 0.4375C - 2.4375D + 0.8127E \tag{23}$$

Comparing the results in Tables 22 and 23, we can see that they are quite different. The linear regression and ANOVA method failed to identify A and E as important factors at the significance level of 0.1.

## *6. CONCLUSION*

In this tutorial, simple examples were used to illustrate the basic concepts in DOE. Guidelines for conducting DOE were given. Three major topics were discussed in detail: 2-level factorial design, RSM and DOE for life tests.

Linear regression and ANOVA are the important tools in DOE data analysis. So, they are emphasized. For DOE involving censored data, the better method of MLE and likelihood ratio test should be used.

DOE involves many different statistical methods. Many useful techniques, such as blocking and randomization, random and mixed effect model, model diagnostic, power and sample size, measurement system study, RSM with multiple responses, D-optimal designs, Taguchi orthogonal array, Taguchi robust designs, mixture designs and so on are not covered in this tutorial [1, 2, 5, 6]. However, with the basic knowledge of this tutorial, readers should be able to learn most of them easily.

*REFERENCES*


1. D. C. Montgomery, *Design and Analysis of Experiments*, 5th edition, John Wiley and Sons, Inc. New York, 2001.
2. C. F. Wu and M. Hamad, *Experiments: Planning, Analysis, and Parameter Design Optimization*, John Wiley and Sons, Inc. New York, 2000.
3. ReliaSoft, http://www.ReliaSoft.com/doe/index.htm
4. ReliaSoft, *Life Data Analysis Reference*, ReliaSoft Corporation, Tucson, 2007.
5. R. H. Myers and D. C. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd edition, John Wiley and Sons, Inc. New York, 2002.
6. ReliaSoft, *Experiment Design and Analysis Reference*, ReliaSoft Corporation, Tucson, 2008.