CHAPTER 2

THEORY AND LITERATURE SURVEYS

2.1 What is Experimental Design ?

A designed experiment is a test or series of tests in which purposeful changes are made to the input variables of a process so that we may observe and identify corresponding changes in the output response. In Figure 2.1, the process can be visualized as some combination of machines, methods and people that transforms an input material into an output product, This output product has one or more observable quality characteristics or responses. Some of the process variables $x_1, x_2, \ldots, x_p$ are controllable, while others $z_1, z_2, \ldots, z_p$ are uncontrollable.

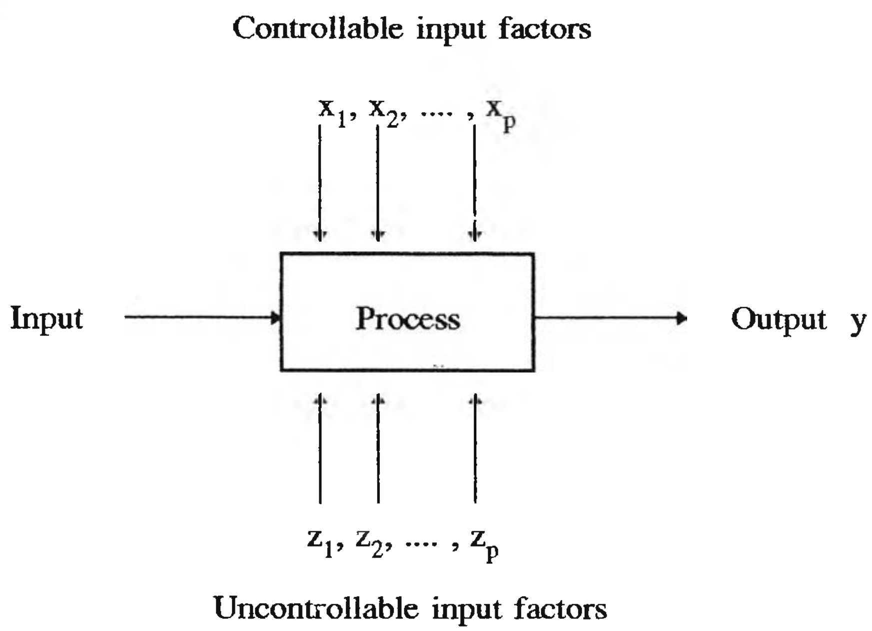

Figure 2.1 : General Model of a Process.

The objectives of designed experiment may include

1. Determining which variables are most influential on the response, y.
2. Determining where to set the influential x's so that y is near the nominal requirement.
3. Determining where to set the influential x's so that variability in y is small.
4. Determining where to set the influential x's so that the effects of the uncontrollable variables z are minimized.

Experimental design is a critically important engineering tool for improving a manufacturing process. It also has extensive application in the development of new processes. Application of these techniques early in process development can result in

1. Improved yield.
2. Reduced variability and closer conformance to nominal.
3. Reduced development time.
4. Reduced overall costs.

2.2 Basic Principles

The three basic principles of experimental design are replication, randomization, and blocking.

Replication means a repetition of the experiment. Replication has two important benefits. First, it allows the experimenter to obtain an estimate of the experimental error. This estimate of error becomes a basic unit of measurement for detemining whether observed differences in the data are really statistically different. Second, if the sample mean is used to estimate the effect of a factor in the experiment, then replication permits the experimenter to obtain a more precise estimate of this effect.

Randomization means that the order in which the individual runs or trials of the experiment are to be performed are randomly determined. Statistical methods require that the observations or errors be independently distributed random variables. Randomization usually makes this assumption valid. By properly randomizing the experiment, the effects of nuisance variable is balanced out.

Blocking is a technique used to increase the precision of an experiment. This technique is used in order to control or remove variability arising from nuisance variables. A block is a portion of the experimental material that should be more homogeneous than the entire set of material. Blocking involves making comparisons among the conditions of interest in the experiment within each block.

2.3 Type of Designed Experiment

By the number of factor, the designed experiment can be classified as single- factor experiment, factorial experiment, and 2^k factorial experiment.

Single-factor experiment is the designed experiment for testing effect of a factor, which has more than two levels, on responses.

Factorial experiment is used to study the effects of two or more factors. The effects of factors include a main effect and an interaction effect on the interesting responses. This experiment is suitable for more than two levels of each factor.

Only two levels of each factor, for two or more factors, especially in several factors (k factors), 2^k factorial experiment is widely used to study the joint effect of the factors on a response. The 2^k design is particularly useful in the early stages of experimental work to screen factor that does not affect an response variable out, so called the factor screening experiment.

In Figure 2.2, the main effect of factor A and B have on an response when factor levels (A_1, A_2, B_1,and B_2) change, but there is no interaction effect.

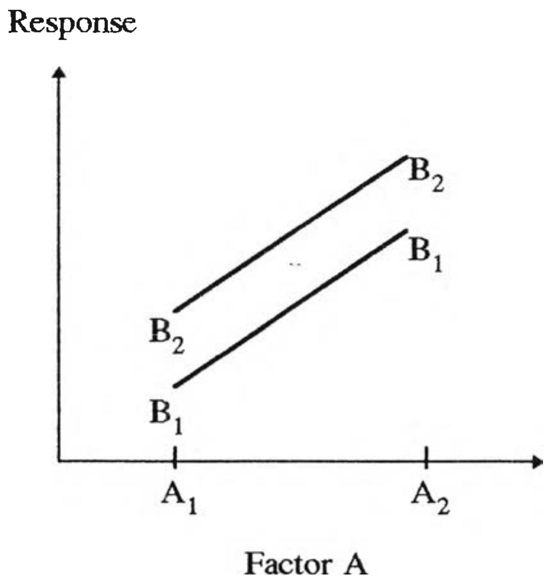

Figure 2.2 : A Factorial Experiment without Interaction.

In Figure 2.3, the main effect of factor A and B have on an response when factor levels (A_1, A_2, B_1,and B_2) change. Also there is the interaction effect between factors A and B, the two lines are not parallel.

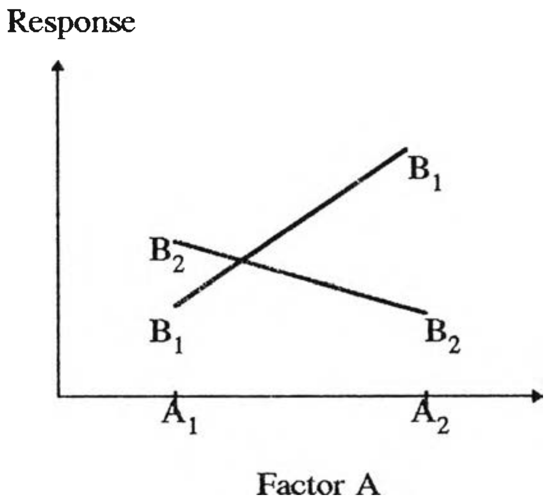

Figure 2.3 : A Factorial Experiment with Interaction.

2.4 Analysis of Variance

The method of Analysis of Variance (ANOVA) is applied to the designed experiment to draw conclusion about the effect of factors on an response.

For example, experiment with two factors, the hypotheses about the model of observations, which will be tested by ANOVA, are as follows.

The observations resulting from the experiment are shown in Figure 2.4 may be described by the model

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \text{ , by } \begin{array}{l} i = 1, 2, \ldots\ldots, a \\ j = 1, 2, \ldots\ldots, b \\ k = 1, 2, \ldots\ldots, n \end{array}$$

where μ is the overall mean effect, τ_i is the effect of the i th level of the row factor A, β_j is the effect of the j th level of column factor B, $(\tau\beta)_{ij}$ is the effect of the interaction between τ_i and β_j, and ε_{ijk} is a random error component. Both factors are assumed to be fixed, and the treatment effects are defined as deviations from the overall mean, so $\sum_{i=1}^{a}\tau_i = 0$ and $\sum_{j=1}^{b}\beta_j = 0$. Similarly, the interaction effects are fixed and are defined such that $\sum_{i=1}^{a}(\tau\beta)_{ij} = \sum_{j=1}^{b}(\tau\beta)_{ij} = 0$. Since there are n replicates of the experiment, there are abn total observations.

Factor B

Factor A	1	2	.....	b
1	$y_{111}, y_{112}, \ldots\ldots, y_{11n}$	$y_{121}, y_{122}, \ldots\ldots, y_{12n}$		$y_{1b1}, y_{1b2}, \ldots\ldots, y_{1bn}$
2	$y_{211}, y_{212}, \ldots\ldots, y_{21n}$	$y_{221}, y_{222}, \ldots\ldots, y_{22n}$		$y_{2b1}, y_{2b2}, \ldots\ldots, y_{2bn}$
. .				
a	$y_{a11}, y_{a12}, \ldots\ldots, y_{a1n}$	$y_{a21}, y_{a22}, \ldots\ldots, y_{a2n}$		$y_{ab1}, y_{ab2}, \ldots\ldots, y_{abn}$

Figure 2.4 : General Arrangement for a Two-factor Factorial Design.

In the factorial experiment, both row and column factors or treatment, A and B, are of equal interest. So the hypotheses about the equality of row treatment effects will be tested.

$$H_0 : \tau_1 = \tau_2 = \ . \ . \ . \ = \tau_a = 0$$
$$H_1 : \text{ at least one } \tau_i \neq 0$$

And the hypotheses about the equality of column treatment effects will be tested.

$$H_0 : \beta_1 = \beta_2 = \ . \ . \ . \ = \beta_b = 0$$
$$H_1 : \text{ at least one } \beta_j \neq 0$$

Finally, the hypotheses about interaction effect will also be tested.

$$H_0 : (\tau\beta)_{ij} = 0 \qquad\qquad \text{for all i, j}$$
$$H_1 : \text{ at least one } (\tau\beta)_{ij} \neq 0$$

These hypotheses are tested using ANOVA of the fixed effects model by computing sum of squares , mean squares, and ratio of mean squares (F_o) as follows.

The total sum of squares is computed as usual by

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn} \quad ,$$
$$y_{...} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}$$

The sums of squares for the main effects are

$$SS_A = \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - \frac{y_{...}^2}{abn}$$

$$SS_B = \frac{1}{an}\sum_{j=1}^{b} y_{.j.}^2 - \frac{y_{...}^2}{abn}$$

The sum of squares for the interaction effect is

$$SS_{AB} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{...}^2}{abn} - SS_A - SS_B$$

The sum of squares of error is

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_o
A treatments	SS_A	a-1	$MS_A = \frac{SS_A}{a-1}$	$F_o = \frac{MS_A}{MS_E}$
B treatments	SS_B	b-1	$MS_B = \frac{SS_B}{b-1}$	$F_o = \frac{MS_B}{MS_E}$
Interaction	SS_{AB}	(a-1)(b-1)	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_o = \frac{MS_{AB}}{MS_E}$
Error	SS_E	ab(n-1)	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	SS_T	abn-1		

Table 2.1 : The Analysis of Variance Table for the Two Factors.

And $F_{\alpha, v1, v2}$ can be obtained from the table of percentage points of the F distribution, α is the significant level, and V_1 and V_2 are the degrees of freedom.

In Table 2.1, we would reject H_o if F_o of A treatments is more than $F_{\alpha, a-1, ab(n-1)}$, we conclude that factor A significantly affects an response. In the same way, we would reject H_o if F_o of B treatments is more than $F_{\alpha, b-1, ab(n-1)}$, we conclude that factor B significantly affects an response.

And we would reject H_o if F_o of interaction is more than $F_{\alpha, (a-1)(b-1), ab(n-1)}$, we conclude that there is an interaction effect between the two factors on an response.

2.5 Model Adequacy Checking

As the analysis of variance assumes that the model errors are normally and independently distributed with the same variance in each factor level, abbreviated NID(0, σ^2), these assumptions can be checked by examining the residuals. A residual is defined as the difference between the actual observation and the value that would be obtained from a least-squares fit of the underlying analysis of variance model to the sample data. For example, the residuals for the two-factor factorial model are

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk} \quad \text{or}$$
$$e_{ijk} = y_{ijk} - \bar{y}_{ijk}$$

The normality assumption can be checked by constructing a normal probability plot of the residuals, plotting residuals ranked in ascending order (k) versus their cumulative probability points P_k = (k-0.5)/ n , n is number of all observations in the experiment.

To check the assumption of equal variance at each factor level, plot the residuals against the factor levels and the fitted values, and then compare the spread in the residuals.

2.6 Duncan's Multiple Range Test

A procedure is widely used for comparing individual means of either factor, either the row averages or the column averages, when using the fixed effects models.

For example, two-factor factorial experiment, R_p in the equation below are used to compare with difference between two means.

$$R_p = r_\alpha(p,f)S_{\bar{y}_{ij.}} \quad \text{for p= 2, 3, ..., a or b}$$

From Duncan's table of significant ranges, obtain the value $r_\alpha(p,f)$, for p= 2, 3, ..., a or b, where α is the significance level and f is the number of degrees of freedom for error, $S_{\bar{y}_{ij.}} = \sqrt{\frac{MS_E}{n}}$, and n replicates .

2.7 Choice of number of replicates

Operating characteristic curve can be used to find the number of replicates for the designed experiment, for the two-factor factorial experiment, using the following formula.

$$\Phi^2 = \frac{naD^2}{2b\sigma^2}$$

where n is the number of replicates, a levels of factor A, b levels of factor B, D is the difference in mean, σ is standard deviation, V_1 = b-1, and V_2 = ab(n-1). Using Φ resulting from trials of n , α, V_1, and V_2 in the operating characteristic curve leads to β risk that could be acceptable to select the number of replicates.

2.8 Guidelines for Designing Experiments

Montgomery(1991) gives an outline of the recommended procedure as follows.

2.8.1 Recognition of and Statement of the problem.

In practice, it is often difficult to realize that a problem requiring formal designed experiments exists, so it may not be easy to develop a clear and generally accepted statement of the problem. However, it is absolutely essential to fully develop all ideas about the problem and about the specific objectives of the experiment.

A clear statement of the problem and the objectives of the experiment often contribute substantially to better process understanding and eventual solution of the problem.

2.8.2. Choice of Factor and Levels.

The experimenter must choose the factors to be varied in the experiment, the ranges over which these factors will be varied, and the specific levels at which runs will be made. Process knowledge including practical experience and theoretical understanding is required to do this. This step determines type of experiment whether single-factor experiment or factorial experiment or 2^k factorial experiment.

2.8.3 Selection of the Response Variable.

In selecting the response variable, the experimenter should be certain that the variable really provides useful information about the process under study. Most often the average or standard deviation (or both) of the measured characteristic will be the response variable.

2.8.4 Choice of Experimental Design.

Choice of design involves consideration of number of replicates, selection of a suitable run order for the experimental trials, and whether or not blocking or other randomization restrictions are involved.

2.8.5 Performing the Experiment.

When running the experiment, it is vital to carefully monitor the process to ensure that everything is being done according to plan. Errors in experimental procedure at this stage will usually destroy experimental validity. Up-front planning is crucial to success. It is easy to underestimate the logistical and planning aspects of running a designed experiment in a complex manufacturing environment.

2.8.6 Data Analysis.

Statistical methods should be used to analyze the data so that results and conclusions are objective rather than judgmental. If the experiment has been designed correctly and if it has been performed according to the design, then the type of statistical methods required is not elaborate. Many excellent software packages are available to assist in the data analysis, and simple graphical methods play an important role in data interpretation. Residual analysis and model validity checking are also important.

2.8.7 Conclusions and Recommendations.

Once the data have been analyzed, the experiment must draw practical conclusions about the results and recommend a course of action. Graphical methods are often useful in this stage, particularly in presenting the results to others. Follow-up runs and confirmation testing should also be performed to validate the conclusions from the experiment.

2.9 Literature Surveys

Brydson (1995) presented chemical data of melamine-formaldehyde resins, chemical structure of the resins, the method of melamine production, and resinification. The resinification, which is the method of the melamine-formaldehyde resin production, can be applied to this research at the reactor laboratory by controlling temperature level of 80-100 degrees of celsius and pH of the resin. Another important point of this book, the pH of melamine-formaldehyde resins can be adjusted using sodium carbonate or sodium hydroxide.

Hines and Montgomery (1980) presented the fundamental and application of statistics. The contents of tests of hypotheses, analysis of variance, and design of experiments are focused for doing this thesis. The hypotheses testing about the differences of two means and two variances of two groups of data can be applied to this research to draw conclusion about the data resulting from the laboratory and the process. Also the hypotheses testing is a part of the analysis of variance. The analysis of variance (ANOVA) is an important statistical tool for analyzing the difference of means of many groups of data, more than two group. The ANOVA can be classified as the three ways, which are fixed effects model, random effects model, and mixed model, depending on the objectives and the selected factor that are fixed or randomized . The ANOVA can also be used as an important part of the design of experiment since ANOVA is used to find effects of factors in the experiment. The fundamental concept of design of experiment is introduced as an application of ANOVA , how to analyze data of the designed experiment.

Juran and Gryna (1993) presented the fundamental concepts of quality control, quality improvement, and quality planning, special and common cause of variation, statistical tools for analyzing data, statistical process control (SPC) , control chart, and the process capability. The reduction in variation is emphasized and is done using the concept of statistical process control (SPC). A way of using SPC concept is the control chart to monitoring process. When special causes of variation are eliminated, the process capability is more than 1.00, by implementing the control chart, the next step is the design of experiment for reducing the common causes of variation.

Kume (1992) presented the development, interpretation, and application of control chart, also other techniques for quality control. The out-of-control patterns of control chart are classified such as outside of limits, hugging, run, trend, periodicity.

Montgomery and Runger (1994) presented techniques of statistics and probability to solve technical problems, including hypotheses testing, estimation, and the experimental design comprising a single factor design, factorial experimental design, 2^k factorial experimental design. Also the analysis of variance for the fixed effects model, the random effects model, the mixed model is explained how to analyze the results of these model. Moreover, The mean estimation using confidence interval can be used to draw conclusion about the results of experiments in this research. The hypotheses testing about difference of means and variances can be employed to ensure the results of experiments by a level of significance.

Montgomery (1996), (1997) explained application of design of experiment for quality control, also analysis of variance (ANOVA) which is statistical method for analyzing the factors of the process. The experimental design technique is introduced for the purpose of the reduction in variation. Many applications of experimental design are a screening experiment, process optimization, and so on. The designed experiment depends on the number of factors. The single-factor experiment is performed by running only a factor. An experiment that is done by two factors or more is suitable for factorial designed experiment, and a 2^k factorial experiment is a particular form of factorial experiment, there are two levels of each factor. Also he gives the good guildline for designing experiments. In another book, the particular details for designing experiment and the methods of statistical analysis for the experiment are presented, starting from giving the definition of design of experiment, basic principles of experimental design comprising replication, randomization, and blocking. And the designed experiment can be classified as single-factor experiment, factorial experiment, and 2^k factorial experiment by number of factor. Also the effect of factor consists of the main effect and interaction effect. Consequently, the collected data from designed experiment is analyzed by the analysis of variance (ANOVA) to examine the factor's effects, based on the hypotheses testing about difference of means of response variable. And assumptions of ANOVA that the model error are normally and independently distributed with the same variance in each treatment combination must be checked by residual analysis, which comprises the use the normal probability plot of residuals and the plots of residuals against the factor levels and the fitted values. Other techniques in this book used in this research for analysing the data are Duncan's multiple range test to examine the difference of means in each treatment and the use of operating characteristic curve to find appropriate number of replicates for the two-factor factorial experiment.

Vale and Taylor (1964) explained how to produce melamine compound, which type of process of melamine compound, how to test and measure the curing time of melamine compound.

Krisada Asawarungsaengkul (1999) presented the quality improvement for process of harddisk by applying the design of experiment technique, 2^k factorial design. He uses the method of design of experiment to investigate 5 factors that might affect the number of chips and cracks of harddisk, a response or a quality characteristic. As the result of the experiment , two factors which are cutting speed and direction of cut significantly affect the chips and cracks. Finally, the suitable condition of the two factors is found by his design of experiment, leading to the reduction in the number of chips and cracks.

Songphon Phichatwattana (1998) applied the design of experiment to the process of hard disc drive by analyzing 4 factors, resulting in 3 factors that signficantly affect the pull strength of read/ write head of hard disc drive, which is a response. Consequently, he can meet a appropriate level of the three factors by the method of design of experiment, bringing about maximum pull strength as required.

Tossapol Kiatcharoenpol (1995) presented how to design the experiment for 4 factors affecting 6 responses of lacquering process and how to evaluate and conclude the result of the experiment that is the suitable condition of the four controllable factors for the optimum level of the six responses of the process.