CHAPTER SEVEN

Experimental Design

CHAPTER OBJECTIVES

After completing this chapter, the reader should be able to:

- understand the basic concepts and definitions of experimental design;
- discuss the characteristics of experimental design;
- understand the use of full and fractional factorial experimental designs;
- understand the use of analysis of variance (ANOVA) in analyzing experimental results; and
- discuss Taguchi Methods for experimental design.

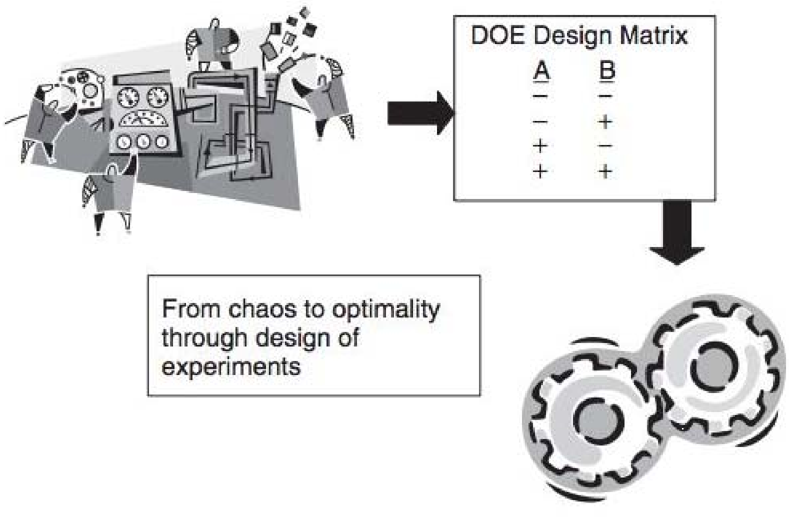
Even relatively simple systems are affected by a number of variables. For example, some of the factors that may affect the quality of a hole drilled into a block of wood include the type of drill bit, the sharpness of the bit, the rotational speed of the bit, the feed rate of the bit into the work, the type of wood, and the moisture content of the wood. To really optimize the performance of this process, the quality manager or engineer would need to determine the best settings for these six variables. Many industrial processes involve hundreds of variables, any or all of which can affect the quality of the output. Determining which of these variables are most important, how they may interact with each other, and how to adjust them for optimal performance is the domain of experimental design. In our example, designed experiments can assist the quality professional in determining which type of drill bit to use, how long to use it before resharpening, how many revolutions per minute (rpm) to run the bit, how fast to feed the bit into the work, adjustments necessary for different types of wood being drilled, and how dry the wood should be before processing in order for the drilling process to produce consistent, top-quality holes.

An experiment where one variable is studied while the other variables are held constant can be inefficient and suffers from the inability to assess interactions among the variables. Factorial designs allow for the efficient testing of the main effects of the variables as well as the interaction effects among the variables. Interpretation of the data obtained from factorial designs is often accomplished using analysis of variance (ANOVA). Factorial designs coupled with analysis using ANOVA enable the experimenter to determine all main effects (i.e., the effect on the process of changing one variable) and all interaction effects (the effect on the process of the interaction of several variables).

Genichi Taguchi made a significant contribution by adapting fractional factorial orthogonal arrays (balanced both ways) to experimental design so that the time and cost of experimentation is reduced while validity and reproducibility are maintained. Taguchi's approach is disciplined and structured to make it easy for quality managers and engineers to apply. A full factorial design with seven factors at two levels would require $2^7 = 128$ experiments while Taguchi's L_8 orthogonal array requires only eight experiments.

- Is this process operating in a way that quality is maximized and cost is minimized?
- What can we do to improve the performance of this process?
- Will this change to the process create undesired effects on other aspects of the process?

Questions such as these are often asked but frequently dealt with using just intuition, assumptions, and experience. They may be addressed best by use of carefully designed experiments in conjunction with technical knowledge about the processes and experience in working with the processes. The result of a program of process improvement using experimental design is a process that performs optimally with a minimum of variation. Process owners also will better understand the relationships among the many variables in the process and the performance that they desire. Ultimately this translates into higher quality, improved customer satisfaction, and increased profits.



BASIC CONCEPTS AND DEFINITIONS

Experimental design is defined as "a formal plan that details the specifics for conducting an experiment, such as which responses, factors, levels, blocks, treatments, and tools are to be used" (Quality Glossary, 2002, 49). Designed experiments are important tools for optimizing processes, identifying interactions among process variables, and reducing variation in processes.

In design of experiments (DOE) the experimenter changes the inputs to a process in a systematic way and evaluates resulting changes in the outputs. The inputs are called the independent variables (X-variables) and the outputs are called the dependent variables (Y-variables). A *dependent variable* is the variable of primary interest in an experiment. *Independent variables* are those we believe that affect the measurements obtained on the dependent variable for a given experiment. It is possible for several independent variables to be under investigation simultaneously. *Environmental variables* are those which may or may not affect experimental results but which are not easily controlled by the experimenter. Examples of environmental variables would be ambient temperature and humidity.

In an experimental investigation, each independent variable that is assumed to influence the dependent variable of interest is called a *factor*. The various logical categories or intensities of the factors being investigated are referred to as *levels* or *treatments*. In a given experiment, several independent variables, or factors, might be under investigation simultaneously. A specific combination of factor levels imposed on a single unit of experimental material is called a treatment combination. The experiment in Figure 7.1 consists of three levels (0.2%, 0.4%, and 0.65%) of one factor, chlorination. The commonly used notation is to show the number of factors as the exponent for the number of levels. For example, an experiment that examines two factors at two levels would be referred to as a 2^2 full-factorial design. One with three factors at two levels would be referred to as a 2^3 full-factorial design (see Figure 7.4).

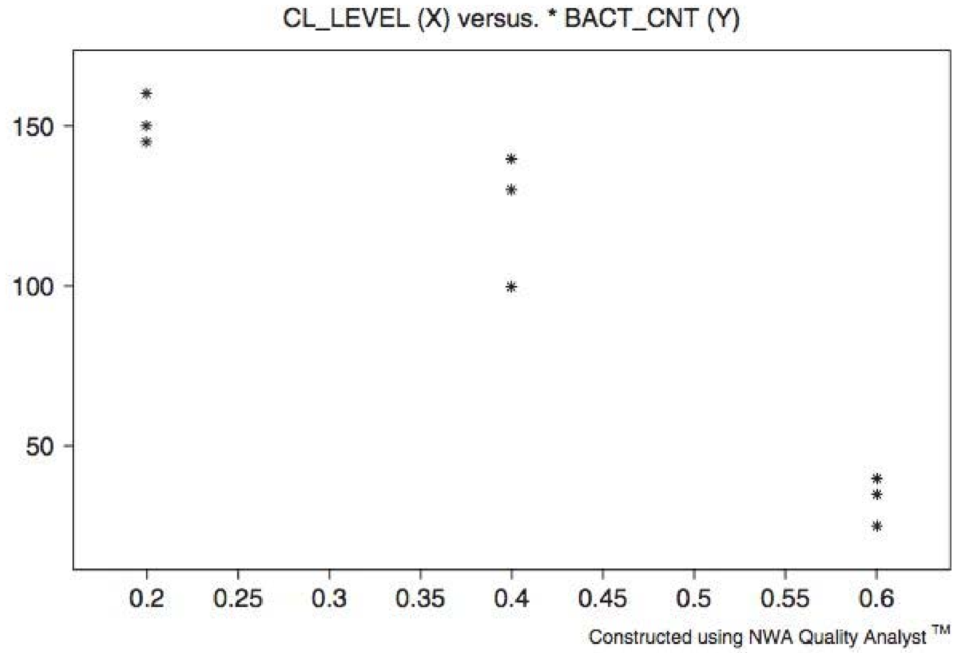
Three basic concepts of experimental design are replication, randomization, and local control (Fisher, 1958). When two or more identical experimental units are subjected to the same treatment, the experiment is said to be replicated. *Replication* is necessary in an experiment in order to provide a measure of experimental error and thus "to increase the precision of the estimates of the effects" (Montgomery, 2005). When two identical experimental units receive the same treatment or treatment combination (replication) and yield different responses or measurements, the difference between what statistical analysis of a sample says a value should be and the actual value based on the population as a whole is defined as *error*. Error arises due to random forces and due to other variables that contribute to variation (such as environmental variables), but which are not singled out for investigation (i.e., not included as factors in the experiment). The experiment in Figure 7.1 consists of three replications.

Randomization is a requirement for the statistical methods (ANOVA and regression) used to analyze the results of designed experiments and also minimizes the effect of procedural bias. Randomization is achieved through the use of random

Trial	Chlorination Level	Bacterial Count
1	1	150
2	2	100
3	1	145
4	1	160
5	2	140
6	3	25
7	2	130
8	3	35
9	3	40

Chlorination Levels: 1 - 0.2%
2 - 0.4%
3 - 0.6%

A simple analysis of this experiment can be done using a scatter diagram.



It appears that the relationship between chlorination levels and bacterial counts is fairly linear, and that as chlorine levels are increased, bacterial counts decrease.

Figure 7.1. Single Factor Experiment

numbers to determine the order of experimental trials and to assign the order for specific inputs. The experiments in Figure 7.4 do not show randomization. Before running the experiment, the order of the four trials for the 2^2 design and the eight trials for the 2^3 design should be randomly assigned.

Local control refers to the use of *blocks*. Blocks are a "subdivision of the experiment space into relatively homogeneous experimental units between which the experimental error can be expected to be smaller than would be expected should a similar number of units be randomly located within the entire experimental space" (ASQC Statistics Division, 1983, 62). An example of the use of blocking would be the testing of fertilizers by applying the different treatments to test plots within a small block of the field. The external variables (e.g., amount of sunlight, basic soil fertility, drainage) within the block would be expected to vary less than over the entire field. This provides for "a more homogeneous experiment subspace" (ASQC Statistics Division, 1983, 62). Blocking may be used as an alternative to a completely randomized experiment.

Consider a case of testing the hardness of material used in a metal stamping process. The experimenter has been provided several different stamping tools—that is, different designs, materials, weights, and so forth. The experimenter wants to determine which of various stamping tools will provide the best impact on the metal plates being stamped. In a completely randomized design, each tool would be used to stamp randomized pieces of stamping material and the results compared. Although this sounds okay, a closer examination would reveal that random variations in the stamping material could introduce a bias to the experiment that may skew the test results. In order to minimize this bias, a randomized block design is used. In this case, a number of pieces of stamping material are randomly selected. All tools are tested at random points on each piece of material. By using this method, we have minimized the potential bias from variations in the stamping material and thus have a more robust data set for comparing the various stamping tools.

Experiments can be performed by anyone wishing to answer a question. We all perform experiments of one kind or another. For example, we may conduct an experiment to determine the quickest way to get from our house in the suburbs to a point in the central business district. We may consider three alternative modes of travel (independent variable): personal vehicle (car), city bus, and light rail. We may collect our data by randomly selecting a mode on each of the nine days so that each mode is taken three times. Our decision could be made based on the average travel time (dependent variable) for each mode.

We may find that the decision made based on this simple experiment does not provide the expected outcome. The process may be more complicated than we thought when we designed this simple experiment. This might be due to one or more uncontrolled variables in this process. Among the possible uncontrolled variables would be time of day and weather. Weather would be a difficult, but not impossible, variable to incorporate into your experiment. Time of day is much more easily included, and based upon your prior knowledge of the process, is very likely to influence travel time. Time of day might interact with the mode selected. During rush hour, travel time for a car might be the highest, while during non-rush hour it might be the lowest. Time of day might interact less with bus travel time since buses may be able to use high occupancy vehicle lanes. Time of day might not interact at all with rail travel times since rail should not be affected by highway congestion. In order to understand this interaction effect, a new experiment must be designed that incorporates time of day as an independent variable.

EXPERIMENTAL DESIGN CHARACTERISTICS

In addition to replication, randomization, and local control, it is important to consider balance, efficiency, and fit when designing experiments. *Balance* refers to having equal numbers of observations, n, in each cell. Balance is important because it facilitates the use of more multiple comparison procedures. It can also prevent certain forms of bias from affecting the experimental results. An *orthogonal design* provides both horizontal and vertical balancing—that is, all pairs of factors at particular levels appear together an equal number of times.

Efficiency has to do with how focused or precise the experimental design is. For example, because a higher order full factorial design requires the running of a large number of trials, we may use a fractional factorial design (i.e., look at a specific set of variables, while ignoring others). The selection of variable for a fractional factorial design would be based on technical knowledge and experience with the process. The fractional factorial design would be more efficient than a full k-factorial design, but it would not take into account all factors and their possible interactions. So we may miss a relationship that we did not expect, as only those factors previously identified as pertinent are being evaluated.

Fit refers to how well the experimental design fits the actual population. Fit is usually calculated using error, which measures the difference between what our experimental analysis tells us the outcome should be and what the actual outcome is using the entire population.

The process of designing and conducting an experiment consists of seven steps (Montgomery, 2005, 14):

1. Recognize and define the problem to be addressed by the experiment.
2. Select the response variable(s).
3. Select the factors, levels, and ranges.
4. Select and appropriate experimental design.

5. Perform the experiment.
6. Statistically analyze the data.
7. Form conclusions and recommendations.

TYPES OF DESIGN

There are various types of experimental designs. When there are several factors of interest in an experiment and all main effects and interaction effects are to be examined, a factorial design should be used. In a *full-factorial design* or experiment, all possible combinations of the levels of the factors are investigated in each complete trial of the experiment. Thus if there are two factors, A and B, with j levels of factor A and k levels of factor B, then each replicate contains all jk possible combinations. In a *fractional factorial design*, a subset of the factors is evaluated.

Single-Factor Design

When we desire information on the relationship between a single input factor and the output factor in the experiment, a *one-factor or single-factor design* is appropriate. A single-factor design saves time and effort because it focuses on a single item rather than all possible combinations of many factors, yet it may result in a suboptimal solution for the same reason. A full-factorial design is best used when we are concerned about the effects of the interactions of the multiple factors because single factor experiments cannot assess these interaction effects. A full-factorial design is best used when the need for a complete data set outweighs the potential cost savings of a smaller, more focused experiment.

An example of a single-factor experiment would be the analysis of the effect of input concentration of a chlorinator on bacterial counts in a stream of water. Three levels of chlorination are to be evaluated, and the trials are to be replicated three times. Using the random number table in Appendix A results in the sequence of trials shown in Figure 7.1.

One-Factor-at-a-Time Design

When multiple factors are to be considered, a simplistic approach is to vary one factor at a time and hold the other factors constant. This is a very inefficient experimental design, and, unless all possible combinations are run, does not allow for the estimation of interaction effects among the variables. Therefore this design is only useful if no factor interactions exist.

Consider a chemical process with three independent variables (factors)—temperature, pressure, and time—and one dependent variable, yield. A possible one-factor-at-a-time experiment is shown in Figure 7.2. The one-factor-at-a-time experimental design provides incomplete and possibly misleading information. With the addition of two more runs (four more with replication), complete information can be obtained by using the 2^3 full-factorial design.

2a. Hold everything constant except time. Run two replications of two different levels of time and select the time with the highest yield.

Run	Time	Temp.	Pressure
1	−	+	+
2	+	+	+
1	−	+	+
2	+	+	+

Assume that the results show that the longer time (+) is better.

2b. Then fix time at the "optimum level", hold pressure constant, and run two replications of two different levels of temperature.

Run	Time	Temp.	Pressure
1	+	−	+
2	+	+	+
1	+	−	+
2	+	+	+

Assume that the results show that the lower temperature (−) is better.

2c. Then fix time and temperature at the "optimum levels," and run two replications of two different levels of pressure.

Run	Time	Temp.	Pressure
1	+	−	−
2	+	−	+
1	+	−	−
2	+	−	+

Assume that the results show that the higher pressure (+) is better.

The experimenter now believes that the optimum conditions for this process are long time (+), low temperature (−), and high pressure (+). However, since all possible combinations of these variables have not been assessed, this conclusion may very well be erroneous.

To arrive at this conclusion, the experimenter executed six runs—with replication, a total of 12 trials. Had the experimenter used a 2^3 full-factorial design, she would need to execute eight runs—with replication, a total of 16 trials. But, with the full-factorial design, she would be able to assess all main effects and all interaction effects. This provides a much greater likelihood of identifying the true optimal settings for the three factors.

2d. Full-factorial design.

Run	Time	Temp.	Pressure
1	−	−	−
2	+	−	−
3	−	+	−
4	+	+	−
5	−	−	+
6	+	−	+
7	−	+	+
8	+	+	+

Runs 9–16 would be a replication of the above.

Figure 7.2. Comparison of One-Factor-at-a-Time with Full-Factorial Design

An engineer is interested in finding the values of temperature and time that maximize yield. Suppose we fix the temperature at 160°F (the current operating level) and perform five runs at different levels of time. The results of this series of runs are shown in the table below.

Run Time (Hour)	Yield (%)
0.5	42
1.0	50
1.5	62
2.0	70
2.5	66

This experiment indicates that maximum yield is achieved at approximately 2.0 hours of reaction time. To optimize the temperature, the engineer fixes the time at 2.0 hours (the apparent optimum) and performs five runs at different temperatures.

Temperature (°F)	Yield (%)
140	36
150	44
160	76
170	62
180	31

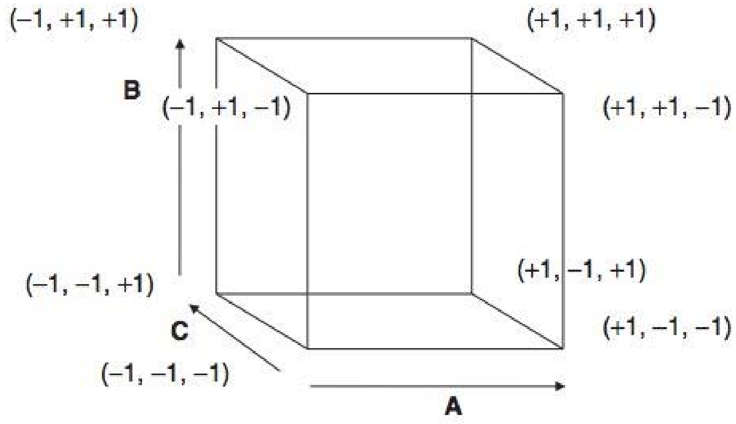
Based on the data from the second experiment, maximum yield occurs at 160°F.

Therefore, we would conclude that running the process at 160°F and 2.0 hours is the best set of operating conditions, resulting in a yield of about 73 percent [(70 + 76)/2].

The downside of one-factor design is that it fails to detect possible interactions between the two factors (in this case time and temperature). For the example above, a full factorial experimental design indicates that optimal yield is really around 95 percent at a temperature of between 180°F and 190°F and a time of around 0.5 hours.

Full-Factorial Design

A full-factorial design involves testing all possible combinations of the factors in an experiment at a number of levels. A full-factorial design allows for the estimation of all main effects and all interaction effects. The difficulty is that this design requires a considerable investment of time and resources. This becomes of increasing importance as the number of factors increases. A full-factorial design for two levels of three factors (2^3) is shown in Figure 7.2d and in Figure 7.3.



Run	A	B	C	A×B	A×C	B×C	A×B×C
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

These four columns are part of the analysis matrix and show the interactions that can be assessed. The three columns following the Run column constitute the design matrix.

Note that this design examines only the corners of the cube. If there is an indication that the relationships may be nonlinear, it would be advantageous to examine the center points as well as the corners.

Figure 7.3. 2^3 Full-Factorial Design

2^2

Run	Factor A	Factor B
1	−	−
2	−	+
3	+	−
4	+	+

2^3

Run	Factor A	Factor B	Factor C
1	−	−	−
2	+	+	−
3	−	+	−
4	+	+	−
5	−	−	+
6	+	−	+
7	−	+	+
8	+	+	+

Figure 7.4. Some Full-Factorial Designs

The design in Figure 7.3 assumes a linear relationship among the factors. The corners of the cube represent the two levels of the factors, that is, −1 (minimum setting) and +1 (maximum setting). If a nonlinear relationship is suspected, it would be necessary to run the center points of the cube shown. A 3^3 factorial design that examines three levels of the three factors will provide more complete evidence of the linearity of the relationships. But this comes at the cost of the need to make 27 runs rather than eight runs.

Students running a 2^3 full-factorial-designed experiment to optimize the settings on the catapult to maximize the distance that the ball is thrown. [Exercise 1]

FRACTIONAL FACTORIAL DESIGN

When there are many factors, the number of runs required for a full-factorial design becomes unmanageable. A fractional factorial design involves testing an orthogonal (balanced) subset of all possible combinations of the factors. A fractional factorial design allows for the estimation of all main effects and selected interaction effects. The sacrifice in information obtained is balanced by the reduced resource requirements compared with a full-factorial design. A 2^{3-1} fractional factorial design for the example in Figure 7.1 is shown in Figure 7.5. The run number is the number these runs would be in a full-factorial design.

Often fractional factorial designs are used for *screening experiments* to identify critical factors, to identify the direction and magnitude of the main effects, and to estimate factor interactions. This could lead to additional experiments examining only the critical factors. The results of a screening experiment might also be helpful in identifying critical factors to monitor more closely in order to better control a process.