

Design and Analysis of Experiments

8th Edition

Douglas C. Montgomery

Wiley, 2012

pg. 14-21

1.4 Guidelines for Designing Experiments

To use the statistical approach in designing and analyzing an experiment, it is necessary for everyone involved in the experiment to have a clear idea in advance of exactly what is to be studied, how the data are to be collected, and at least a qualitative understanding of how these data are to be analyzed. An outline of the recommended procedure is shown in Table 1.1. We now give a brief discussion of this outline and elaborate on some of the key points. For more details, see Coleman and Montgomery (1993), and the references therein. The **supplemental text material** for this chapter is also useful.

1. Recognition of and statement of the problem. This may seem to be a rather obvious point, but in practice often neither it is simple to realize that a problem requiring experimentation exists, nor is it simple to develop a clear and generally accepted statement of this problem. It is necessary to develop all ideas about the objectives of the experiment. Usually, it is important to solicit input from all concerned parties: engineering, quality assurance, manufacturing, marketing, management, customer, and operating personnel (who usually have much insight and who are too often ignored). For this reason, a **team approach** to designing experiments is recommended.

It is usually helpful to prepare a list of specific problems or questions that are to be addressed by the experiment. A clear statement of the problem often contributes substantially to better understanding of the phenomenon being studied and the final solution of the problem.

It is also important to keep the overall objectives of the experiment in mind. There are several broad reasons for running experiments and each type of experiment will generate its own list of specific questions that need to be addressed. Some (but by no means all) of the reasons for running experiments include:

- a. **Factor screening or characterization.** When a system or process is new, it is usually important to learn which factors have the most influence on the response(s) of interest. Often there are a lot of factors. This usually indicates that the experimenters do not know much about the system so screening is essential if we are to efficiently get the desired performance from the system. Screening experiments are extremely important when working with new systems or technologies so that valuable resources will not be wasted using best guess and OFAT approaches.
- b. **Optimization.** After the system has been characterized and we are reasonably certain that the important factors have been identified, the next objective is usually optimization, that is, find the settings or levels of

■ **TABLE 1.1**
Guidelines for Designing an Experiment

1. Recognition of and statement of the problem]	Pre-experimental planning
2. Selection of the response variable ^a		
3. Choice of factors, levels, and ranges ^a		
4. Choice of experimental design		
5. Performing the experiment		
6. Statistical analysis of the data		
7. Conclusions and recommendations		

^aIn practice, steps 2 and 3 are often done simultaneously or in reverse order.

the important factors that result in desirable values of the response. For example, if a screening experiment on a chemical process results in the identification of time and temperature as the two most important factors, the optimization experiment may have as its objective finding the levels of time and temperature that maximize yield, or perhaps maximize yield while keeping some product property that is critical to the customer within specifications. An optimization experiment is usually a follow-up to a screening experiment. It would be very unusual for a screening experiment to produce the optimal settings of the important factors.

- c. **Confirmation.** In a confirmation experiment, the experimenter is usually trying to verify that the system operates or behaves in a manner that is consistent with some theory or past experience. For example, if theory or experience indicates that a particular new material is equivalent to the one currently in use and the new material is desirable (perhaps less expensive, or easier to work with in some way), then a confirmation experiment would be conducted to verify that substituting the new material results in no change in product characteristics that impact its use. Moving a new manufacturing process to full-scale production based on results found during experimentation at a pilot plant or development site is another situation that often results in confirmation experiments—that is, are the same factors and settings that were determined during development work appropriate for the full-scale process?
- d. **Discovery.** In discovery experiments, the experimenters are usually trying to determine what happens when we explore new materials, or new factors, or new ranges for factors. In the pharmaceutical industry, scientists are constantly conducting discovery experiments to find new materials or combinations of materials that will be effective in treating disease.
- e. **Robustness.** These experiments often address questions such as under what conditions do the response variables of interest seriously degrade? Or what conditions would lead to unacceptable variability in the response variables? A variation of this is determining how we can set the factors in the system that we can control to minimize the variability transmitted into the response from factors that we cannot control very well. We will discuss some experiments of this type in Chapter 12.

Obviously, the specific questions to be addressed in the experiment relate directly to the overall objectives. An important aspect of problem formulation is the recognition that one large comprehensive experiment is unlikely to answer the key questions satisfactorily. A single comprehensive experiment requires the experimenters to know the answers to a lot of questions, and if they are wrong, the results will be disappointing. This leads to wasting time, materials, and other resources and may result in never answering the original research questions satisfactorily. A **sequential** approach employing a series of smaller experiments, each with a specific objective, such as factor screening, is a better strategy.

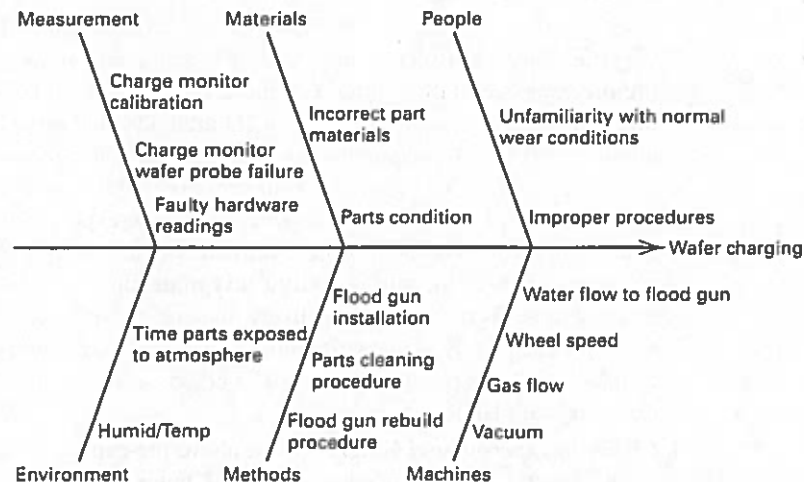
2. **Selection of the response variable.** In selecting the response variable, the experimenter should be certain that this variable really provides useful information about the process under study. Most often, the average or standard deviation (or both) of the measured characteristic will be the response variable. Multiple responses are not unusual. The experimenters must decide how each response will be measured, and address issues such as how will any measurement system be calibrated and

how this calibration will be maintained during the experiment. The gauge or measurement system capability (or measurement error) is also an important factor. If gauge capability is inadequate, only relatively large factor effects will be detected by the experiment or perhaps additional replication will be required. In some situations where gauge capability is poor, the experimenter may decide to measure each experimental unit several times and use the average of the repeated measurements as the observed response. It is usually critically important to identify issues related to defining the responses of interest and how they are to be measured *before* conducting the experiment. Sometimes designed experiments are employed to study and improve the performance of measurement systems. For an example, see Chapter 13.

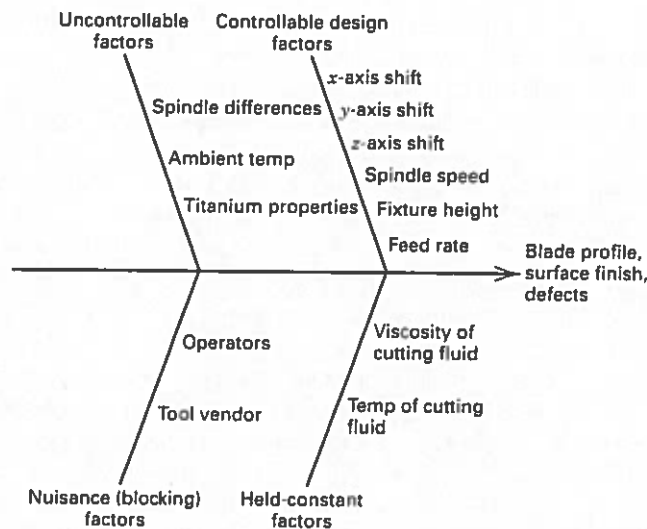
3. *Choice of factors, levels, and range.* (As noted in Table 1.1, steps 2 and 3 are often done simultaneously or in the reverse order.) When considering the factors that may influence the performance of a process or system, the experimenter usually discovers that these factors can be classified as either **potential design factors** or nuisance factors. The potential design factors are those factors that the experimenter may wish to vary in the experiment. Often we find that there are a lot of potential design factors, and some further classification of them is helpful. Some useful classifications are **design factors**, **held-constant factors**, and **allowed-to-vary factors**. The design factors are the factors actually selected for study in the experiment. Held-constant factors are variables that may exert some effect on the response, but for purposes of the present experiment these factors are not of interest, so they will be held at a specific level. For example, in an etching experiment in the semiconductor industry, there may be an effect that is unique to the specific plasma etch tool used in the experiment. However, this factor would be very difficult to vary in an experiment, so the experimenter may decide to perform all experimental runs on one particular (ideally “typical”) etcher. Thus, this factor has been held constant. As an example of allowed-to-vary factors, the experimental units or the “materials” to which the design factors are applied are usually nonhomogeneous, yet we often ignore this unit-to-unit variability and rely on randomization to balance out any material or experimental unit effect. We often assume that the effects of held-constant factors and allowed-to-vary factors are relatively small.

Nuisance factors, on the other hand, may have large effects that must be accounted for, yet we may not be interested in them in the context of the present experiment. Nuisance factors are often classified as **controllable**, **uncontrollable**, or **noise factors**. A controllable nuisance factor is one whose levels may be set by the experimenter. For example, the experimenter can select different batches of raw material or different days of the week when conducting the experiment. The blocking principle, discussed in the previous section, is often useful in dealing with controllable nuisance factors. If a nuisance factor is uncontrollable in the experiment, but it can be measured, an analysis procedure called the **analysis of covariance** can often be used to compensate for its effect. For example, the relative humidity in the process environment may affect process performance, and if the humidity cannot be controlled, it probably can be measured and treated as a covariate. When a factor that varies naturally and uncontrollably in the process can be controlled for purposes of an experiment, we often call it a noise factor. In such situations, our objective is usually to find the settings of the controllable design factors that minimize the variability transmitted from the noise factors. This is sometimes called a process robustness study or a robust design problem. Blocking, analysis of covariance, and process robustness studies are discussed later in the text.

The **cause-and-effect diagram** can be a useful technique for organizing some of the information generated in pre-experimental planning. Figure 1.10 is the cause-and-effect diagram constructed while planning an experiment to resolve problems with wafer charging (a charge accumulation on the wafers) encountered in an etching tool used in semiconductor manufacturing. The cause-and-effect diagram is also known as a **fishbone diagram** because the “effect” of interest or the response variable is drawn along the spine of the diagram and the potential causes or design factors are organized in a series of ribs. The cause-and-effect diagram uses the traditional causes of measurement, materials, people, environment, methods, and machines to organize the information and potential design factors. Notice that some of the individual causes will probably lead directly to a design factor that



■ **FIGURE 1.10** A cause-and-effect diagram for the etching process experiment



■ **FIGURE 1.11** A cause-and-effect diagram for the CNC machine experiment

will be included in the experiment (such as wheel speed, gas flow, and vacuum), while others represent potential areas that will need further study to turn them into design factors (such as operators following improper procedures), and still others will probably lead to either factors that will be held constant during the experiment or blocked (such as temperature and relative humidity). Figure 1.11 is a cause-and-effect diagram for an experiment to study the effect of several factors on the turbine blades produced on a computer-numerical-controlled (CNC) machine. This experiment has three response variables: blade profile, blade surface finish, and surface finish defects in the finished blade. The causes are organized into groups of controllable factors from which the design factors for the experiment may be selected, uncontrollable factors whose effects will probably be balanced out by randomization, nuisance factors that may be blocked, and factors that may be held constant when the experiment is conducted. It is not unusual for experimenters to construct several different cause-and-effect diagrams to assist and guide them during preexperimental planning. For more information on the CNC machine experiment and further discussion of graphical methods that are useful in preexperimental planning, see the supplemental text material for this chapter.

We reiterate how crucial it is to bring out all points of view and process information in steps 1 through 3. We refer to this as **pre-experimental planning**. Coleman and Montgomery (1993) provide worksheets that can be useful in pre-experimental planning. Also see the **supplemental text material** for more details and an example of using these worksheets. It is unlikely that one person has all the knowledge required to do this adequately in many situations. Therefore, we strongly argue for a team effort in planning the experiment. Most of your success will hinge on how well the pre-experimental planning is done.

4. **Choice of experimental design.** If the above pre-experimental planning activities are done correctly, this step is relatively easy. Choice of design involves consideration of sample size (number of replicates), selection of a suitable run order for the experimental trials, and determination of whether or not blocking or other randomization restrictions are involved. This book discusses some of the more important types of

experimental designs, and it can ultimately be used as a guide for selecting an appropriate experimental design for a wide variety of problems.

There are also several interactive statistical software packages that support this phase of experimental design. The experimenter can enter information about the number of factors, levels, and ranges, and these programs will either present a selection of designs for consideration or recommend a particular design. (We usually prefer to see several alternatives instead of relying entirely on a computer recommendation in most cases.) Most software packages also provide some diagnostic information about how each design will perform. This is useful in evaluation of different design alternatives for the experiment. These programs will usually also provide a worksheet (with the order of the runs randomized) for use in conducting the experiment.

Design selection also involves thinking about and selecting a tentative **empirical model** to describe the results. The model is just a quantitative relationship (equation) between the response and the important design factors. In many cases, a low-order polynomial model will be appropriate. A **first-order** model in two variables is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

where y is the response, the x 's are the design factors, the β 's are unknown parameters that will be estimated from the data in the experiment, and ε is a random error term that accounts for the experimental error in the system that is being studied. The first-order model is also sometimes called a **main effects** model. First-order models are used extensively in screening or characterization experiments. A common extension of the first-order model is to add an **interaction** term, say

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

where the cross-product term $x_1 x_2$ represents the two-factor interaction between the design factors. Because interactions between factors is relatively common, the first-order model with interaction is widely used. Higher-order interactions can also be included in experiments with more than two factors if necessary. Another widely used model is the **second-order** model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

Second-order models are often used in optimization experiments.

In selecting the design, it is important to keep the experimental objectives in mind. In many engineering experiments, we already know at the outset that some of the factor levels will result in different values for the response. Consequently, we are interested in identifying *which* factors cause this difference and in estimating the *magnitude* of the response change. In other situations, we may be more interested in verifying uniformity. For example, two production conditions A and B may be compared, A being the standard and B being a more cost-effective alternative. The experimenter will then be interested in demonstrating that, say, there is no difference in yield between the two conditions.

5. **Performing the experiment.** When running the experiment, it is vital to monitor the process carefully to ensure that everything is being done according to plan. Errors in experimental procedure at this stage will usually destroy experimental validity. One of the most common mistakes that I have encountered is that the people conducting the experiment failed to set the variables to the proper levels on some runs. Someone should be assigned to check factor settings before each run. Up-front planning to prevent mistakes like this is crucial to success. It is easy to

underestimate the logistical and planning aspects of running a designed experiment in a complex manufacturing or research and development environment.

Coleman and Montgomery (1993) suggest that prior to conducting the experiment a few trial runs or pilot runs are often helpful. These runs provide information about consistency of experimental material, a check on the measurement system, a rough idea of experimental error, and a chance to practice the overall experimental technique. This also provides an opportunity to revisit the decisions made in steps 1–4, if necessary.

6. **Statistical analysis of the data.** Statistical methods should be used to analyze the data so that results and conclusions are **objective** rather than judgmental in nature. If the experiment has been designed correctly and performed according to the design, the statistical methods required are not elaborate. There are many excellent software packages designed to assist in data analysis, and many of the programs used in step 4 to select the design provide a seamless, direct interface to the statistical analysis. Often we find that simple **graphical methods** play an important role in data analysis and interpretation. Because many of the questions that the experimenter wants to answer can be cast into an hypothesis-testing framework, hypothesis testing and confidence interval estimation procedures are very useful in analyzing data from a designed experiment. It is also usually very helpful to present the results of many experiments in terms of an **empirical model**, that is, an equation derived from the data that express the relationship between the response and the important design factors. Residual analysis and model adequacy checking are also important analysis techniques. We will discuss these issues in detail later.

Remember that statistical methods cannot prove that a factor (or factors) has a particular effect. They only provide guidelines as to the reliability and validity of results. When properly applied, statistical methods do not allow anything to be proved experimentally, but they do allow us to measure the likely error in a conclusion or to attach a level of confidence to a statement. The primary advantage of statistical methods is that they add objectivity to the decision-making process. Statistical techniques coupled with good engineering or process knowledge and common sense will usually lead to sound conclusions.

7. **Conclusions and recommendations.** Once the data have been analyzed, the experimenter 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.

Throughout this entire process, it is important to keep in mind that experimentation is an important part of the learning process, where we tentatively formulate hypotheses about a system, perform experiments to investigate these hypotheses, and on the basis of the results formulate new hypotheses, and so on. This suggests that experimentation is **iterative**. It is usually a major mistake to design a single, large, comprehensive experiment at the start of a study. A successful experiment requires knowledge of the important factors, the ranges over which these factors should be varied, the appropriate number of levels to use, and the proper units of measurement for these variables. Generally, we do not perfectly know the answers to these questions, but we learn about them as we go along. As an experimental program progresses, we often drop some input variables, add others, change the region of exploration for some factors, or add new response variables. Consequently, we usually experiment **sequentially**, and as a general rule, no more than about 25 percent of the available resources should be invested in the first experiment. This will ensure

that sufficient resources are available to perform confirmation runs and ultimately accomplish the final objective of the experiment.

Finally, it is important to recognize that all experiments are designed experiments. The important issue is whether they are well designed or not. Good pre-experimental planning will usually lead to a good, successful experiment. Failure to do such planning usually leads to wasted time, money, and other resources and often poor or disappointing results.

1.5 A Brief History of Statistical Design

There have been four eras in the modern development of statistical experimental design. The agricultural era was led by the pioneering work of Sir Ronald A. Fisher in the 1920s and early 1930s. During that time, Fisher was responsible for statistics and data analysis at the Rothamsted Agricultural Experimental Station near London, England. Fisher recognized that flaws in the way the experiment that generated the data had been performed often hampered the analysis of data from systems (in this case, agricultural systems). By interacting with scientists and researchers in many fields, he developed the insights that led to the three basic principles of experimental design that we discussed in Section 1.3: randomization, replication, and blocking. Fisher systematically introduced statistical thinking and principles into designing experimental investigations, including the factorial design concept and the analysis of variance. His two books [the most recent editions are Fisher (1958, 1966)] had profound influence on the use of statistics, particularly in agricultural and related life sciences. For an excellent biography of Fisher, see Box (1978).

Although applications of statistical design in industrial settings certainly began in the 1930s, the second, or industrial, era was catalyzed by the development of response surface methodology (RSM) by Box and Wilson (1951). They recognized and exploited the fact that many industrial experiments are fundamentally different from their agricultural counterparts in two ways: (1) the response variable can usually be observed (nearly) immediately, and (2) the experimenter can quickly learn crucial information from a small group of runs that can be used to plan the next experiment. Box (1999) calls these two features of industrial experiments **immediacy** and **sequentiality**. Over the next 30 years, RSM and other design techniques spread throughout the chemical and the process industries, mostly in research and development work. George Box was the intellectual leader of this movement. However, the application of statistical design at the plant or manufacturing process level was still not extremely widespread. Some of the reasons for this include an inadequate training in basic statistical concepts and methods for engineers and other process specialists and the lack of computing resources and user-friendly statistical software to support the application of statistically designed experiments.

It was during this second or industrial era that work on **optimal** design of experiments began. Kiefer (1959, 1961) and Kiefer and Wolfowitz (1959) proposed a formal approach to selecting a design based on specific objective optimality criteria. Their initial approach was to select a design that would result in the model parameters being estimated with the best possible precision. This approach did not find much application because of the lack of computer tools for its implementation. However, there have been great advances in both algorithms for generating optimal designs and computing capability over the last 25 years. Optimal designs have great application and are discussed at several places in the book.

The increasing interest of Western industry in quality improvement that began in the late 1970s ushered in the third era of statistical design. The work of Genichi Taguchi [Taguchi