current flow is obtained through the ammeter (usually $X = 0$), then the unknown resistance is calculated as

$$Y = \frac{BD}{C} - \frac{X^2}{C^2E}[A(D + C) + D(B + C)][B(C + D) + F(B + C)] \qquad (1\text{–}1)$$

The engineer wants to design the circuit so that overall gage capability is good; that is, he would like for the standard deviation of the measurement error to be small. He has decided that $A = 20\ \Omega$, $C = 2\ \Omega$, $D = 50\ \Omega$, $E = 1.5$ V, and $F = 2\ \Omega$ are the best choices for the design parameters so far as gage capability is concerned, but the overall measurement error is still too high. This is likely due to the tolerances that have been specified on the circuit components. These tolerances are ±1 percent for each resistor (A, B, C, D, and F) and ±5 percent for the power supply E. These tolerance bands can be used to define high and low factor levels, and an experiment can be performed to determine which circuit components have the most critical tolerances and by how much these critical tolerances must be tightened to produce adequate gage capability. This will result in a design specification that tightens only the most critical tolerances the minimum amount possible consistent with desired measurement capability; consequently, a lower-cost design that is easier to manufacture will result.

Notice that in this experiment it is unnecessary to actually build the hardware since the response from the circuit can be calculated via Equation 1–1. The actual response variable for the experiment should be the standard deviation of Y. However, an equation for the standard deviation of Y for the circuit can be found by using a Taylor series expansion of Equation 1–1, as explained in many engineering statistics textbooks [for example, see Hines and Montgomery (1990), Chapter 5]. Therefore, the entire experiment can be performed using a computer model of the Wheatstone bridge. ■

from MONTGOMERY, D.C. (1997)
Design and Analysis of Experiments, 4th Edn
Wiley

1–3 BASIC PRINCIPLES

If an experiment such as the ones described in Examples 1–1 through 1–4 is to be performed most efficiently, then a scientific approach to planning the experiment must be employed. By the **statistical design of experiments,** we refer to the process of planning the experiment so that appropriate data that can be analyzed by statistical methods will be collected, resulting in valid and objective conclusions. The statistical approach to experimental design is necessary if we wish to draw meaningful conclusions from the data. When the problem involves data that are subject to experimental errors, statistical methodology is the only objective approach to analysis. Thus, there are two aspects to any experimental problem: the design of the experiment and the statistical analysis of the data. These two subjects are closely related since the method of analysis depends directly on the design employed. Both topics will be addressed in this book.

The three basic principles of experimental design are **replication, randomization,** and **blocking.** By replication we mean a repetition of the basic experiment.

In the metallurgical experiment discussed in Section 1–1, replication would consist of treating a specimen by oil quenching and treating a specimen by saltwater quenching. Thus, if five specimens are treated in each quenching medium, we say that five **replicates** have been obtained. Replication has two important properties. First, it allows the experimenter to obtain an estimate of the experimental error. This estimate of error becomes a basic unit of measurement for determining whether observed differences in the data are really *statistically* different. Second, if the sample mean (e.g., $\bar{y}$) 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. For example; if σ^2 is the variance of an individual observation and there are n replicates, then the variance of the sample mean is

$$\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n}$$

The practical implication of this is that if we had $n = 1$ replicates and observed $y_1 = 145$ (oil quench) and $y_2 = 147$ (saltwater quench), we would probably be unable to make satisfactory inferences about the effect of the quenching medium—that is, the observed difference could be the result of experimental error. On the other hand, if n was reasonably large, and the experimental error was sufficiently small, then if we observed $\bar{y}_1 < \bar{y}_2$, we would be reasonably safe in concluding that saltwater quenching produces a higher hardness in this particular aluminum alloy than does oil quenching.

Randomization is the cornerstone underlying the use of statistical methods in experimental design. By randomization we mean that both the allocation of the experimental material and 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, we also assist in "averaging out" the effects of extraneous factors that may be present. For example, suppose that the specimens in the above experiment are of slightly different thicknesses and that the effectiveness of the quenching medium may be affected by specimen thickness. If all the specimens subjected to the oil quench are thicker than those subjected to the saltwater quench, then we may be continually handicapping one quenching medium over the other. Randomly assigning the specimens to the quenching media alleviates this problem.

Blocking is a technique used to increase the precision of an experiment. 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. A simple example of the blocking principle is given in Section 2–5.1 of Chapter 2.

These basic principles of experimental design are an important part of every experiment. We will illustrate and emphasize them repeatedly throughout this book.

1–4 GUIDELINES FOR DESIGNING EXPERIMENTS

To use the statistical approach in designing and analyzing an experiment, it is necessary that everyone involved in the experiment 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.

1. Recognition of and statement of the problem. This may seem to be a rather obvious point, but in practice it is often not 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, the customer, and operating personnel (who usually have much insight and who are too often ignored). A clear statement of the problem often contributes substantially to a better understanding of the phenomena and the final solution of the problem. For this reason, a **team approach** to designing experiments is recommended.

2. Choice of factors, levels, and ranges. (As noted in Table 1–1, steps 2 and 3 are often done simultaneously, or in the reverse order.) 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. Thought must also be given to how these factors are to be controlled at the desired values and how they are to be measured. For instance, in the flow solder experiment, the engineer has defined 12 variables that may affect the occurrence of solder defects. The engineer will also have to decide on a region of interest for each variable (that is, the range over which each factor will be varied) and on how many levels of each variable to use. Process knowledge is required to do this.

Table 1–1 Guidelines for Designing an Experiment

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

[a] In practice, steps 2 and 3 are often done simultaneously, or in reverse order.

This process knowledge is usually a combination of practical experience and theoretical understanding. It is important to investigate all factors that may be of importance and to not be overly influenced by past experience, particularly when we are in the early stages of experimentation or when the process is not very mature.

When the objective of the experiment is factor screening or process characterization, it is usually best to keep the number of factor levels low. Generally, two levels work very well in factor screening studies. Choosing the region of interest is also important. In factor screening, the region of interest should be relatively large—that is, the range over which the factors are varied should be broad. As we learn more about which variables are important and which levels produce the best results, the region of interest will usually become more narrow.

3. 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. Gauge capability (or measurement error) is also an important factor. If gauge capability is inadequate, then 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.

We reiterate how crucial it is to bring out all points of view and process information in steps 1 through 3 above. We refer to this as **pre-experimental planning.** It is unlikely that one person has all the knowledge required to do this adequately in many situations. Therefore, we argue strongly 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 pre-experimental planning activities above are done correctly, this step is relatively easy. Choice of design involves the consideration of sample size (number of replicates), the selection of a suitable run order for the experimental trials, and the 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 catalog 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 prefer to see several alternatives instead of relying on a computer recommendation in most cases.) These programs will usually also provide [illegible]orksheet (with the order of the runs randomized) for use in conducting the experiment.

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. 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 or research and development environment.

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 if it has been performed according to the design, then 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. 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. 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 hypothe[illegible] 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.

(1958) Statistical Methods for Research Workers 13th Edn

1–5 HISTORICAL PERSPECTIVE

The late Sir Ronald A. Fisher was the innovator in the use of statistical methods in experimental design. For several years he was responsible for statistics and data analysis at the Rothamsted Agricultural Experiment Station in London, England. Fisher developed and first used the analysis of variance as the primary method of statistical analysis in experimental design. In 1933, Fisher took a professorship at the University of London. He later was on the faculty of Cambridge University and held visiting professorships at several universities throughout the world. For an excellent biography of Fisher, see J. F. Box (1978). While Fisher was clearly the pioneer, there have been many other significant contributors to the literature of experimental design, including F. Yates, R. C. Bose, O. Kempthorne, W. G. Cochran, R. H. Myers, J. S. Hunter, W. G. Hunter, and G. E. P. Box. The bibliography at the end of the book contains several works by these authors.

Many of the early applications of experimental design methods were in the agricultural and biological sciences, and as a result, much of the terminology of the field is derived from this heritage. However, the first industrial applications of experimental design began to appear in the 1930s, initially in the British textile and woolen industry. After World War II, experimental design methods were introduced to the chemical and process industries in the United States and Western Europe. These industry groups are still very fertile areas for using experimental design in product and process development. The semiconductor and electronics industry has also used experimental design methods for many years with considerable success.

In recent years, there has been a revival of interest in experimental design in the United States because many industries discovered that their off-shore competitors have been using designed experiments for many years and that this has been an important factor in their competitive success. The day is approaching (hopefully rapidly) when all engineers will receive formal training in experimental design as part of their undergraduate education. The successful integration of experimental design into the engineering profession is a key factor in the future competitiveness of the industrial base of the United States.

1–6 SUMMARY: USING STATISTICAL TECHNIQUES IN EXPERIMENTATION

Much of the research in engineering, science, and industry is empirical and makes extensive use of experimentation. Statistical methods can greatly increase the efficiency of these experiments and often strengthens the conclusions so obtained. The proper use of statistical techniques in experimentation requires that the experimenter keep the following points in mind:

1. Use your nonstatistical knowledge of the problem. Experimenters are usually highly knowledgeable in their fields. For example, a civil engineer working on a problem in hydrology typically has considerable practical experience and formal academic training in this area. In some fields there is a large body of physical theory on which to draw in explaining relationships between factors and responses. This type of nonstatistical knowledge is invaluable in choosing factors, determining factor levels, deciding how many replicates to run, interpreting the results of the analysis, and so forth. Using statistics is no substitute for thinking about the problem.

2. Keep the design and analysis as simple as possible. Don't be overzealous in the use of complex, sophisticated statistical techniques. Relatively simple design and analysis methods are almost always best. This is a good place to reemphasize step 4 of the procedure recommended in Section 1–4. If you do the design carefully and correctly, the analysis will almost always be relatively straightforward. However, if you botch the design badly, it is unlikely that even the most complex and elegant statistics can save the situation.

3. Recognize the difference between practical and statistical significance. Just because two experimental conditions produce mean responses that are statistically different, there is no assurance that this difference is large enough to have any practical value. For example, an engineer may determine that a modification to an automobile fuel injection system may produce a true mean improvement in gasoline mileage of 0.1 mi/gal. This is a statistically significant result. However, if the cost of the modification is $1000, then the 0.1 mi/gal difference is probably too small to be of any practical value.

4. Experiments are usually iterative. Remember that, in most situations, it is unwise to design too comprehensive an experiment at the start of a study. Successful design requires knowledge of the important factors, the ranges over which these factors are varied, the appropriate number of levels for each factor, and the proper units of measurement for each factor and response. Generally, we are not well-equipped to answer these questions at the beginning of the experiment, but we learn the answers as we go along. This argues in favor of the *iterative* or *sequential* approach discussed previously. Of course, there are situations where comprehensive experiments are entirely appropriate, but as a general rule, most experiments should be iterative. Consequently, we usually should not invest more than about 25 percent of the resources of experimentation (runs, budget, time, etc.) in the initial design. Often these first efforts are just learning experiences, and some resources must be available to accomplish the final objectives of the experiment.