

# 9

## *Chapter*

# Design of Experiments

**Objective:** To study the basic design concepts for experiments and through which we can make comparisons of treatments with respect to the observed responses.

- 9.1 Introduction 466
- 9.2 Concepts from Experimental Design 467
- 9.3 Factorial Design 483
- 9.4 Optimal Design 487
- 9.5 The Taguchi Methods 489
- 9.6 Chapter Summary 493
- 9.7 Computer Examples 494
- Projects for Chapter 9 497



Genichi Taguchi

(Source: [http://www.amsup.com/BIOS/g\\_taguchi.html](http://www.amsup.com/BIOS/g_taguchi.html))

Genichi Taguchi (1924–) acquired his statistical skills under the guidance of Prof. Motosaburo Masuyama, one of the best statisticians of his time. After World War II, Japanese manufacturers were

struggling to survive with very limited resources. Taguchi revolutionized the manufacturing process in Japan through cost savings. He understood that all manufacturing processes are affected by outside influences—noise. However, Taguchi realized methods of identifying those noise sources that have the greatest effects on product variability. Isolating these factors to determine their individual effects can be a very costly and time-consuming process. Taguchi devised a way to use the so-called orthogonal arrays to isolate these noise factors from all others in a cost-effective manner. He introduced the loss function to quantify the decline of a customer's perceived value of a product as its quality declines. Taguchi referred to the ability of a process or product to work as intended regardless of uncontrollable outside influences as robustness. This was a novel concept in the design of experiments with profound influence in manufacturing. His ideas have been adopted by successful manufacturers around the globe because of their results in creating superior production processes at much lower costs.

### 9.1 INTRODUCTION

In statistics, we are concerned with the analysis of data generated from an experiment. It is desirable to take the necessary time and effort to organize the experiment appropriately so that we have the right type of data and sufficient amount of data to answer the questions of interest as clearly and efficiently as possible. This process is called *experimental design*. We can trace the roots of modern experimental design to the 1935 publication of the book *The Design of Experiments*, written by Sir Ronald A. Fisher. He showed how one could conduct credible experiments in the presence of many naturally fluctuating conditions such as the soil condition, temperature, and rainfall, in an agricultural experiment. Because then, the design principles that were developed for agricultural experiments were successfully adapted to industrial, military, and other applications. In modern industry it is essential to manufacture parts efficiently and with practically no defects. As a result, variation reduction in quality characteristics of these parts has become a major focus of quality and productivity improvement. Dr. Genichi Taguchi pioneered the use of design of experiments (DOE) in designing robust products—those relatively insensitive to changes in design parameters. Presently, DOE is used as an essential tool for improving the quality of goods and services. It is important to note that, unless a sound design is employed, it may be very difficult or even impossible to obtain valid conclusions from the resulting data. Also, properly designed experiments will generate more precise data while using substantially fewer experimental runs than ad hoc approaches. In industrial manufacturing, some of the major benefits of DOE are lower costs, simultaneous optimization of several factors, fast generation and organization of quantitative information, and overall quality improvement.

It is important to clearly identify the particular questions that an experiment is intended to answer (that is, the major objective of the experiment) before performing the experiment. These objectives may be to estimate or predict some unknown parameters, to explore relationships among various factors, to compare a collection of effects or parameters, or any combinations of these. When the intention is to compare parameters, the objective may be to corroborate a hypothesis, or to explore some simple relationships. In any design, it is necessary to identify the populations that are to be studied and the type of information about these populations that will be needed to answer the desired questions. While planning an experiment to investigate the primary objectives of the investigation,

we need to ensure that the measurement process is simple, the cost of the study is reasonable, the study can be concluded in a reasonable time frame, and the study produces reliable data. Because of the complex nature of real-world problems, planning an effective experiment is not an easy task. The important issues confronting one area, say engineering, will be different from those for another area such as biology or medicine. As a result, the design of experiments can take several forms. In this chapter, we will follow a general framework. Two of the major distinguishing elements of DOE are (1) simultaneous variation and evaluation of various factors, and (2) systematic removal of some of the possible test combinations to cut back experimental time and cost. Thus, a researcher should ensure that the statistical design is as simple as possible given the objectives of the experiment and within the practical constraints such as material, labor, and cost. Some other desirable criteria of a good design are that it provides unbiased estimates of treatment effects and the experimental error. In addition, it should be able to detect important small differences with sufficient precision, and it should provide an estimation of uncertainty in the conclusions and the confidence with which the result can be extended to other analogous situations. The experimental design determines the basic characteristics of the data collected. These data are then processed using statistical analysis techniques, with the goals of these analyses being determined by the experimental objectives. Conclusions are obtained by looking at the results of the statistical analyses.

## 9.2 CONCEPTS FROM EXPERIMENTAL DESIGN

In this section we introduce some of the basic definitions, methods, and procedures used in the experimental design. Many of the terms used have an agricultural basis, because the early development and applications of DOE were in the field of agriculture.

### 9.2.1 Basic Terminology

The first step in planning an experiment is to formulate a clear statement of objectives of the test program. The purpose of most statistical experiments is to determine the effect of one or more independent variables on the response variable. The main variable of interest in a study is the *response variable*, also called an *output variable*. These are the dependent variables (also referred to as criteria, effect, or predicted variable) in an experiment that describes the factors we are interested in predicting or comparing. The response variable is measured with different values of independent variables (representing those factors that are assumed to be the causes of the outcome) and analyzed to determine whether the independent variables have any effect. For example, in an agricultural experiment, the crop yield could be the response variable, whereas the type of soil, temperature, and rainfall could be the independent variables. We would like also to identify known or expected sources of variability in the experimental units, because one of the main aims of a designed experiment is to reduce the effect of these sources of variability on the answers to questions of interest. Hence, we must make a list of the factors that may affect the value of the response variable. We must also decide how many observations should be taken and what values should be chosen for each independent variable in each individual test run.

**Definition 9.2.1** *The variables that an experimenter is able to completely control in the DOE are called independent variables or treatment variables. These are also called input variables, explanatory variables, or factors.*

Basically, *factors* are independent variables whose effect on the response variable is a main objective of the study. These are control variables selected by the analyst for comparison. A factor is a general category or type of treatment. Factors can be either quantitative or qualitative based on whether the variable is measured on a numerical scale or not. For example, a rice field is divided into six parts, and each part is treated with a different fertilizer to see which produces the most rice. Here the response variable is the amount of rice output. The objective of the study is to compare the effects of different fertilizers on the rice output. Thus, the type of fertilizer is the factor.

**Definition 9.2.2** *Independent variables that are unknown or known but nonmanipulable are called nuisance variables.*

A factor can have different levels referred to as the *treatment* or *factor levels*. Different treatments constitute different levels of a factor. Levels are the values at which the factors are set in an experiment. The level of a variable or treatment means its amount or magnitude. For example, if the experimental units of a medication were given as 2.5 mg, 5 mg, and 10 mg, those amounts would be three levels of the treatment. *Level* is also used for categorical variables, such as drugs I, II, and III, where the three are different kinds of drugs, not different amounts of the same thing. Suppose four different groups of students are subjected to four different teaching methods. The students are the experimental units, the teaching methods are the treatments, and the four types of teaching methods constitute four levels of the factor "type of teaching." Note that this is a single-factor experiment, the factor being the method of teaching.

**Definition 9.2.3** *Noise is the effect of all the uncontrolled factors in an experiment.*

In some experiments, all the noise factors are known; however, in most cases only some of them are known. When an analyst controls the specification of the treatments and the method of allocating the experimental units to each of the treatments, the experiment is called *designed*. For example,  $n$  rats are randomly assigned to one of the five dose levels of an experimental drug under investigation. The analyst can also decide on the number  $n_i$  of rats for each dose level such that  $\sum_{i=1}^5 n_i = n$ .

Sometimes, conducting a designed experiment may not be practical or ethical. For example, if an analyst wants to know the relationship between fat content in a diet and the cholesterol level, it would be unethical and costly as well as time consuming to subject human volunteers to different fat-content diets. However, it is possible to observe the cholesterol levels of people who consume different diets. Care must be taken to record various other factors, such as exercise habits, age, and gender, before reporting any association between cholesterol levels and fat content of diets. The experiment is called *observational*, if the analyst is just an observer of the treatments on a sample of experimental units. Note that the *experimental units* are objects to which treatments are applied.

The crucial difference between an experiment and an observational study for comparing the effects of treatments is that, in an experiment, the researcher decides which experimental units receive which treatments, whereas in an observational study, the researcher simply compares experimental units that happen to be there that have received each of the treatments. Observational studies are often useful for identifying possible causes of treatment effects, and they are often cheaper. Their main disadvantage is that they are less conclusive. Only properly designed and executed experiments can

lead to reliable conclusions. Hence, in general designed experiments are preferred over observational experiments. In designing the experiment, there are almost always going to be constraints such as budget, time, and availability of experimental units.

The following example illustrates an *observational experiment*, where the analyst has control over the random sampling from the treatment populations as well as the size of each sample, but has no control over the assignment of the experimental units to the treatments.

---

**Example 9.2.1**

In order to compare the risk-taking tendency of the people that invest in mutual funds, samples are taken of individuals from three income groups—low income class, middle income class, and high income class. A score is given based on the percentage of their investment allocation on different types of mutual funds, such as large-cap, mid-cap, small-cap, hybrid, and specialty. The mean score for each income group is calculated. Identify each of the following elements: *response, factors and factor type(s), treatments, and experimental units*.

**Solution**

*The response is the variable of interest, which is the score given to each individual investor. The only factor investigated is the income class. This is a qualitative variable. The three income classes represent the levels of this factor. The treatment is the percentage investments in different types of mutual funds, such as large-cap, mid-cap, small-cap, hybrid, and specialty. The experimental unit is the individual investor.*

---

There are *single-factor experiments* and *multifactor experiments*. The previous example was a case of a single-factor experiment. Single-factor experiments have only one independent variable. Another example of a single-factor experiment is when we are interested in the effect of size of the screen of a computer monitor on the reading speed. In this case, the size of the screen is the single factor. If there are only two sizes, say 15-in. and 17-in. monitors, that we wish to compare, tests such as the two-sample *t*-test could be used to compare average reading speed. If there are more than two sizes of monitors, then the one-way ANOVA methods described in Chapter 10 could be used for analysis of the resulting data.

Even though the single-factor experiments are simple and elegant, they are costly and not very effective when there is more than one independent variable. Efficient use of resources is achieved through multifactor experiments in comparison to conducting many single-factor experiments. A multifactor experiment involves two or more independent variables and a dependent variable. Also, a greater range of questions could be answered using multifactor experiments. The resulting data are analyzed using ANOVA as described in Chapter 10. The following is an example of a multifactor experiment.

---

**Example 9.2.2**

In order to study the conditions under which a particular type of commercially raised fish reach maximum weight, an experiment is conducted at four water temperatures ( $60^{\circ}\text{F}$ ,  $70^{\circ}\text{F}$ ,  $80^{\circ}\text{F}$ ,  $90^{\circ}\text{F}$ ) and four water salinity levels (1%, 5%, 10%, 15%). Fish are raised in tanks with specific salinity levels and temperature

levels. There are 32 tanks and one of the four temperatures and one of the four salinity levels are assigned randomly to each tank. The weights are recorded at the beginning of the experiment and after 2 months. Identify each of the following elements: response, and factors and factor type(s). Write all the treatments from the factor-level combinations.

**Solution**

The response is the variable of interest, which is the weight gain of a fish. This experiment has two factors: water temperatures at four levels and water salinity at four levels. There are  $4 \times 4 = 16$  possible treatments:

(60°F, 1%) (60°F, 5%) (60°F, 10%) (60°F, 15%)  
 (70°F, 1%) (70°F, 5%) (70°F, 10%) (70°F, 15%)  
 (80°F, 1%) (80°F, 5%) (80°F, 10%) (80°F, 15%)  
 (90°F, 1%) (90°F, 5%) (90°F, 10%) (90°F, 15%)

It should be noted that there may be other factors, such as the density of the fish population, the initial size of the fish, and the type of feeding, that may affect weight gain of fish.

**Definition 9.2.4** *The experimental error explains the variation in the responses among experimental units that are assigned the same treatment and observed under identical experimental conditions.*

Experimental error can occur for many reasons, among them (1) the difference in the devices that record the measurements, (2) the natural dissimilarities in the experimental units prior to their receiving the treatment, (3) the variation in setting the treatment conditions, and (4) the effect on the response variable of all extraneous factors other than the treatment factors.

In order to construct confidence intervals on the treatment population means and to test hypotheses, it is necessary to obtain an estimate of the variance of experimental design. In a single-factor experiment with  $k$  levels, the estimate of the variance of experimental design could be taken as the pooled variance of responses from experimental units receiving the identical treatments. A large variance of experimental error will compromise the accuracy of inferences made from the experiments. Also, large amounts of experimental error make it difficult to determine whether the treatment has produced an effect or not, so one of the design goals is to reduce the experimental error. Bad execution of a design can lead to the whole experiment becoming a waste of time and resources. It is necessary to implement techniques to reduce experimental error in order to obtain more accurate inferences. One approach to reducing experimental error is to take extra care in conducting the experiment. The effect of experimental error can be reduced by using more homogeneous experimental materials (if available), and using the fundamental principles of replication, randomization, and blocking (see Section 9.2.2).

The *one-way analysis of variance* (in a single-factor experiment at several levels) enables one to compare several groups of observations, all of which are independent with the possibility of a different mean for each group. A test of significance is whether or not all the means are equal. *Two-way analysis of variance* is a method of studying the effects of two factors on the response variable.

There are other terms that are important in different applications. For example, in the medical field, the terms *blinding*, *double-blind*, and *placebo* are used. In a medical experiment, the comparison of treatments may be distorted if the patient, the person administering the treatment, and those evaluating it know which treatment is being allocated to which patient. It is therefore necessary to ensure that the patient, and/or the person administering the treatment, and/or the trial evaluators do not know (are blind to) which treatment is allocated to whom. If only the patient is unaware of the treatment, it is called *blinding*, and if both the patient and the person administering the treatment are blind to which treatment is being allocated, it is called *double-blinding*. In order to study the effect of a particular drug, experimenters divide the study population into two groups and treat one group with the drug and the other group with a so-called placebo, which could be just sugar pills. In order to clarify the objective of a design, it is necessary for an experimental designer to consult a wide range of people, especially those affected by the problem to be solved.

### 9.2.2 Fundamental Principles: Replication, Randomization, and Blocking

A good design of an experiment makes efficient use of resources to gather the data needed to meet the goals of the study. There are three fundamental principles that need to be considered in a good experimental design. They are *replication*, *randomization*, and *blocking*.

**Definition 9.2.5** *Replication means that the same treatment is applied (i) several times to the same experimental units, or (ii) one time to several similar experimental units, called replicate units.*

Replications are necessary for the estimation of the error variance in an experiment against which the differences among treatments are assessed. If an experiment is intended to test whether or not a number of treatments differ in their effects, these treatments must be applied to replicate units of the experiment. In order to show that two treatments have different mean effects, we need to measure several samples given the same treatment. For example, observing that one plant of a particular genotype is more resistant to a disease than another plant of a different genotype does not convey anything about the difference between the mean disease resistance of the two genotypes. This difference could have been caused by the environment or the inoculation procedure affecting the two plants differently. Hence, to make any inference about the mean difference between the genotypes, we have to test several plants of each type. Thus, increasing the number of replications increases the reliability of inferences drawn from the observed data. It is necessary to increase the number of replications to decrease the variance of the treatment effect estimates and also to provide more power for detecting differences in treatment effects. We should not confuse multiple observations of the same experimental unit with replication. Replication involves applying the treatment to a number of experimental units.

**Definition 9.2.6** *A block is a portion of the experimental unit that is more likely to be homogeneous within itself than with other units.*

*Blocking* refers to the distribution of the experimental units into blocks in such a way that the units within each block are more or less homogeneous. The experimenter uses information of the possible variability among units to group them in such a way that most of the unwanted experimental error can be removed through the block effect.

For blocking to be effective, the units should be arranged so that within-block variation is much smaller than between-block variation. As an example, suppose a researcher wishes to compare the yields of rice for four different kinds of fertilizers. In order to minimize the effect of environmental and soil conditions, the field may be divided into smaller blocks and each block is further parceled into four plots. Each variety of fertilizer is applied in each block with one in each parcel. This method ensures that the external conditions from plot to plot within a block will be relatively uniform. Then we can use the ANOVA methods to pool from block to block to obtain the within-block information about the treatment differences while avoiding between-block differences. The relevant analysis is given in Section 10.5. Time could also be a block factor, because the concentration or expertise could alter as one carries out a task, such as determining disease levels or scoring microscope slides.

**Definition 9.2.7 Randomization** is the process of assigning experimental units to treatment conditions in an entirely chance manner.

The main objective of randomization is to negate the effects of all uncontrolled extraneous variables. Usually, randomization is associated with design functions such as random sampling or selection, random assignment, and random order. Random assignment of experimental units to groups tends to spread out differences between subjects in unsymmetric or random ways so that there is no tendency to give an edge to any group. In any well-conducted experiment, randomization eliminates bias from the experiment, enables us to use statistical tests of significance, and creates valid estimates of experimental error. For instance, suppose we are measuring the time of flowering of plants in a glass house or a growth cabinet. If the pots are arranged so that all the plants of one variety are next to each other, and we observe that one variety flowers earlier than the rest, does this imply that this variety is inherently earlier-flowering, or does it suggest that the light and temperature conditions in that part of the cabinet or glass house cause plants to flower early? It is not possible to tell from an experiment designed in this manner. Randomizing the treatments in time or space is an insurance policy, to take account of variation that we may or may not know to exist under the conditions of our experiment. For instance, the levels of light in growth cabinets vary considerably, so randomizing the layout of the plants of different types is essential to make sure that no one type is consistently exposed to light and temperature levels that are particularly high or low. Another way of selecting experimental units is simply to use intact groups, such as all students in a particular statistics classroom. Results obtained this way may be highly biased and hence not desirable. It should be noted that random assignment does not completely eliminate the problem of correlated data values.

Now we study some steps that can be used for randomization. Suppose there are  $N$  homogeneous experimental units and  $k$  treatments. In order to randomly assign  $r_i$  experimental units to the  $i$ th treatment with  $\sum_{i=1}^k r_i = N$ , we could use the following steps.

#### PROCEDURE FOR RANDOM ASSIGNMENT

1. Number the experimental units from 1 to  $N$ .
2. Use a random number table or statistical software to get a list of numbers that are random permutations of the numbers 1 to  $N$ .

3. Give treatment 1 to the experimental units having the first  $r_1$  numbers in the list. Treatment 2 will be given to the next  $r_2$  numbers in the list, and so on; give treatment  $k$  to the last  $r_k$  units in the list.

The following example illustrates the random assignment procedure.

**Example 9.2.3**

In order to study the number of hours to relief provided by five different brands (A, B, C, D, E) of pain reliever, doses are administered to 25 subjects numbered 1 through 25 with each brand administered to five subjects. Develop a design using the random assignment procedure.

**Solution**

Using Minitab, we obtained the following random permutations of the numbers from 1 to 25.

1	8	7	12	10	25	23	4	6	3
9	21	5	24	18	16	22	14	17	15
20	13	2	11	19					

Using the randomized procedure, we obtain the design given in Table 9.1.

**Table 9.1**

Subject:	1	8	7	12	10	25	23	4	6	3	9	21	
Brand:	A	A	A	A	A	B	B	B	B	B	C	C	
Subject	5	24	18	16	22	14	17	15	20	13	2	11	19
Brand	C	C	C	D	D	D	D	E	E	E	E	E	

That is, subject number 8 will get brand A pain reliever, subject 23 will get brand B pain reliever, and so forth. We can rewrite Table 9.1 as shown in Table 9.2.

**Table 9.2**

Brand	Subject					
A	1	8	7	12	10	
B	25	23	4	6	3	
C	9	21	5	24	18	
D	16	22	14	17	15	
E	20	13	2	11	19	

*It should be noted that once we create the design, the actual data will contain the number of hours to relief for each individual.*

It is important to note that randomization may not be possible in some cases. Observational studies may be necessary whenever the researcher cannot use controlled randomized experiments. For example, if we want to study the effect of smoking on lung cancer, randomization will mean that we should be able to select a group of people and tell a randomly selected subgroup to smoke and the other subgroup not to smoke. This is not only practically impossible; it is also unethical to deliberately expose people to a potentially hazardous substance.

### 9.2.3 Some Specific Designs

In this subsection, we will introduce three specific designs: completely randomized design, randomized complete block design, and Latin square design. The structure of the experiment in a *completely randomized design* is presumed to be such that the treatments are assigned to the experimental units completely at random. Example 9.2.1 is one such a design. In order to create a completely randomized design, follow the procedure given in Section 9.2.2.

The *randomized complete block design* is a design in which the subjects are matched according to a variable that the experimenter wants to control. The subjects are put into groups (blocks) of the same size as the number of treatments. The elements of each block are then randomly assigned to different treatment groups so as to reduce the influence of unknown variables. For example, a researcher is carrying out a study of three different drugs for the treatment of high cholesterol. Suppose she has 45 patients and divides them into three treatment groups of 15 patients each. Using a randomized block design, the patients are rated and put in blocks of three based on the cholesterol level: The three patients with the highest cholesterol are put in the first block, those with the next highest levels are put in the second block, and so on to the 15th block. The three members of each block are then randomly assigned, one to each of the three treatment groups. If there is very little extraneous, systematic variation, complete randomization allows differences between the mean effects of the treatments to be estimated with higher precision than other designs. However, it does not allow for the possibility that there could be some unknown extraneous factors, so if in doubt, use a randomized complete block design.

Suppose we have  $k$  treatments and  $N$  experimental units. Further, assume that the experimental units can be grouped into  $b$  groups containing  $k$  experimental units, so that  $N = bk$ . We could use the following steps for a randomized complete block design.

#### PROCEDURE FOR RANDOMIZATION IN A RANDOMIZED COMPLETE BLOCK DESIGN

1. Group the experimental units into  $b$  groups (blocks) containing  $k$  homogeneous experimental units.
2. In group 1, number the experimental units from 1 to  $k$  and obtain a random permutation of numbers 1 to  $k$  using a random number generator.

3. In group 1, the experimental unit corresponding to the first number in the permutation receives treatment 1, the experimental unit corresponding to the second number in the permutation receives treatment 2, and so on.
4. Repeat steps 2 and 3 for each of the remaining blocks.

We illustrate the step-by-step procedure just given in the following example.

#### **Example 9.2.4**

In order to study the number of hours to relief provided by five different brands (A, B, C, D, E) of pain relievers for pain resulting from different causes [headache (H), muscle pain (M), pain due to cuts and bruises (CB)], doses are administered to five subjects each having similar types of pain. Create a randomized complete block design. Choose, as blocks, the different types of pain (H, M, or CB).

#### **Solution**

Using Minitab with  $k = 5$  we have generated the random permutations shown in Table 9.3 for each of the  $b = 3$  blocks of five numbers and assigned the treatments according to the procedure just explained. As the table indicates, among persons with headache, subject number 3 is treated with brand A pain killer, and so forth.

<b>Table 9.3</b>		
H	M	CB
3(A)	5(A)	1(A)
1(B)	4(B)	2(B)
2(C)	3(C)	4(C)
5(D)	1(D)	3(D)
4(E)	2(E)	5(E)

In the previous example, we had only one replication of each treatment per block. This idea can be generalized to have  $r$  replications of each treatment per block. Then the generalized randomized complete block design, with  $k$  treatments,  $b$  blocks, and  $r$  replications with  $N = kbr$  which has  $kr$  homogeneous experimental units, can be randomized as follows.

#### **PROCEDURE FOR A RANDOMIZED COMPLETE BLOCK DESIGN WITH $r$ REPLICATIONS**

1. Group the experimental units into  $b$  groups (called blocks), each containing  $rk$  homogeneous experimental units.
2. In group 1, number the experimental units from 1 to  $rk$  and generate a list of numbers that are random permutations of the numbers 1 to  $rk$ .

3. In group 1, assign treatment 1 to the experimental units having numbers given by the first  $r$  numbers in the list. Assign treatment 2 to the experiments having next  $r$  numbers in the list, and so on until treatment  $k$  receives  $r$  experimental units.
4. Repeat steps 2 and 3 for the remaining blocks of experimental units.

The following example illustrates this procedure.

#### **Example 9.2.5**

With the following modifications, consider Example 9.2.2. Three groups of subjects are considered, with each group having 15 subjects. Group I consists of subjects with only headache (H), group II of subjects only with muscle pain (M), and group III of subjects only pain due to cuts and bruises (CB). Of the 15 with headache (group I), three are treated with brand A pain killer, three with brand B, and so forth. Subjects with other types of pain are treated similarly. Here the number of replications is three for each type of drug and for each type of pain. Create a randomized complete block design with three replications.

#### **Solution**

Using Minitab, for the group with headache (H), we generate a random permutation of numbers 1 to 15. The first three are given pain killer A, the next three B, and so forth. The process is repeated for other types of pain killers. The design is given in Table 9.4 where "2(A)" means that patient 2 is given brand A pain killer.

**Table 9.4**

H	M	CB	H	M	CB
2(A)	8(A)	3(A)	15(C)	9(C)	11(C)
14(A)	13(A)	8(A)	7(D)	4(D)	2(D)
10(A)	5(A)	14(A)	5(D)	11(D)	13(D)
8(B)	2(B)	6(B)	6(D)	15(D)	5(D)
12(B)	1(B)	15(B)	3(E)	7(E)	1(E)
11(B)	10(B)	12(B)	9(E)	12(E)	4(E)
4(C)	3(C)	10(C)	13(E)	6(E)	9(E)
1(C)	14(C)	7(C)			

By increasing the number of replications, we can increase the accuracy of estimators of treatment means and the power of the tests of hypotheses regarding differences between treatment means. However, because of constraints such as cost, time needed to handle a large number of experimental units, and even availability of experimental units, it is not realistic to have a large number of

replications. It is then necessary to determine the minimum number of replications needed to meet reasonable specifications on the accuracy of estimators or on the power of tests of hypotheses. We give a simple procedure for determining the number of replications needed.

Let  $r$  be the number of replications that we need to determine. Let  $\sigma$  be the experimental standard deviation, and  $E$  be the desired accuracy of the estimator. Then the sample size required to be  $(1 - \alpha)100\%$  confident that the estimator is within  $E$  units of the true treatment mean,  $\mu$ , is

$$r = \frac{(z_{\alpha/2})^2 \hat{\sigma}^2}{E^2}.$$

The values of  $\hat{\sigma}$  could be obtained from past experiments, from a pilot study, or by using a rough estimator

$$\hat{\sigma} = (\text{largest observation} - \text{smallest observation})/4.$$

Following is an example for determining the appropriate number of replications.

#### Example 9.2.6

A researcher wants to know the effect of class sizes on the mean score in a standardized test. She wants to estimate the treatment means  $\mu_1, \mu_2, \mu_3$ , and  $\mu_4$  such that she will be 95% confident that the estimates are within 10 points of the true mean score. What is the necessary number of replications to achieve this goal? It is known from the previous experiments that scores have ranged from 46 to 98.

#### Solution

A rough estimator of  $\sigma$  is

$$\hat{\sigma} = \frac{\text{Range}}{4} = \frac{98 - 46}{4} = 13.$$

From the normal table,  $z_{0.025} = 1.96$ . The value of  $E = 10$ . Thus, the number of replications necessary is

$$r = \frac{(z_{\alpha/2})^2 \hat{\sigma}^2}{E^2} = \frac{(1.96)^2 (13)^2}{(10)^2} = 6.4923 \cong 7.$$

Thus, the researcher should use seven replications of each of the treatments to obtain the desired precision.

We have used the randomized complete block design when we wanted to control a single source of extraneous variation and there is only one factor of interest. When we need to compare  $k$  treatment means and there are two possible sources of extraneous variation, a *Latin square design* is the appropriate design of experiment.

**Definition 9.2.8** A  $k \times k$  Latin square design contains  $k$  rows and  $k$  columns. The  $k$  treatments are randomly assigned to the rows and columns so that each treatment appears in every row and column of the design.

It was the famous mathematician Leonhard Euler who introduced Latin squares in 1783 as a new kind of magic squares. Even though the idea is fairly elementary, a systematic use of Latin squares to

the design of experiments was advanced by Ronald A. Fisher only around 1921. Fisher realized that in a two-dimensional plot of land, the systematic error due to variation in soil and other factors could be minimized by a suitable Latin square partition of the plot.

The following example illustrates a case in which the experimental problems are affected by two sources of extraneous variation, the type of car and type of driver used.

### Example 9.2.7

A gasoline company is interested in comparing the effect of four gasoline additives (A, B, C, D) on the gas mileage achieved per gallon. Four cars (I, II, III, IV) and four drivers (1, 2, 3, 4) will be used in the experiment. Create a Latin square design.

#### Solution

We can filter out the variability due to type of cars used by ensuring that in each row only one of the additive types appears. Also, to filter the driver effect, use each additive only once for each driver. One such randomization results in the Latin square design given in Table 9.5.

**Table 9.5**

Cars	Drivers			
	1	2	3	4
I	D	B	A	C
II	C	A	D	B
III	B	D	C	A
IV	A	C	B	D

To construct a basic Latin square, one can use the following method, which we will present only for the  $4 \times 4$  Latin square of Example 9.2.7.

#### PROCEDURE FOR CONSTRUCTING A $4 \times 4$ LATIN SQUARE

1. Begin with the first row as A, B, C, D.
2. Generate each succeeding row by taking the first letter of the preceding row and placing it last, which has the effect of moving the other letters one position to the left.
3. Randomly assign one block factor to the rows and the other to the columns.
4. Randomly assign levels of the row factor, column factor, and treatment to row positions, column positions, and letters, respectively.

In step 2 of the foregoing procedure, instead of using the cyclic placement of rows, we can perform a cyclic placements for the columns. Accordingly, change the procedures in steps 3 and 4.

The following example illustrates a  $4 \times 4$  Latin square design.

**Example 9.2.8**

Using the previous procedure, construct a Latin square for the case of Example 9.2.7.

**Solution**

Following the procedure just given, the Latin square in Example 9.2.7, the basic Latin square is represented by Table 9.6.

Cars	Drivers			
	1	2	3	4
I	A	B	C	D
II	B	C	D	A
III	C	D	A	B
IV	D	A	B	C

Now one random assignment of cars, I, II, III, IV, is to the rows 4, 3, 2, 1 (this is a random order of numbers 1, 2, 3, 4) of Table 9.6. This gives Table 9.7.

Cars	Drivers			
	1	2	3	4
I	D	A	B	C
II	C	D	A	B
III	B	C	D	A
IV	A	B	C	D

Now one random assignment of the drivers 1, 2, 3, 4 is to the columns 1, 2, 4, 3 (this is a random order of numbers 1, 2, 3, 4) of Table 9.7, resulting in the Latin square shown in Table 9.8.

**Table 9.8**

Cars	Drivers			
	1	2	3	4
I	D	A	C	B
II	C	D	B	A
III	B	C	A	D
IV	A	B	D	C

Now along with this Latin square, we can represent the corresponding observations (numbers in parentheses are the gas mileage in miles per gallon) as shown in Table 9.9.

**Table 9.9**

Cars	Drivers			
	1	2	3	4
I	D(18)	A(22)	C(25)	B(19)
II	C(22)	D(24)	B(26)	A(24)
III	B(21)	C(20)	A(22)	D(23)
IV	A(17)	B(24)	D(23)	C(21)

Note that if we use the notation 1 for additive A, 2 for additive B, 3 for additive C, and 4 for additive D, the Latin square in the previous example can be rewritten as shown in Table 9.10.

**Table 9.10**

Cars	Drivers			
	1	2	3	4
I	4	1	3	2
II	3	4	2	1
III	2	3	1	4
IV	1	2	4	3

**Table 9.11**

A	B	C
C	A	B
B	C	A
3 × 3		

A	B	C	D	E
B	A	E	C	D
C	D	A	E	B
D	E	B	A	C
E	C	D	B	A
5 × 5				

This representation will be convenient if we need to write down a model. In order to test for the treatment effects, one could use the ANOVA methods discussed in Chapter 10.

For Latin square experiments involving  $k$  treatments, it is necessary to include  $k$  observations for each treatment resulting in a total of  $k^2$  observations. Table 9.11 shows two examples of Latin squares for  $n = 3$ , and  $n = 5$ .

We have used the Latin square design to eliminate two extraneous sources of variability. In order to eliminate three extraneous sources of variability, we can use a design called the *Greco-Latin square*. Greco-Latin squares are also called *orthogonal Latin squares*. This design consists of  $k$  Latin and  $k$  Greek letters. In this design, we take a Latin square and superimpose upon it a second square with treatments denoted by Greek letters. In this superimposed square, each Latin letter coincides with exactly one of each Greek letter. In our gasoline example, if we introduce the effect of, say, four different days, represented by Greek letters, then Table 9.12 shows the  $4 \times 4$  Greco-Latin square.

**Table 9.12**

$A\alpha$	$B\beta$	$C\gamma$	$D\delta$
$B\delta$	$A\gamma$	$D\beta$	$C\alpha$
$C\beta$	$D\alpha$	$A\delta$	$B\gamma$
$D\gamma$	$C\delta$	$B\alpha$	$A\beta$

We will not go into more detail on this design, or on the many other similar designs.

When developing an experimental design, it is important for the researcher to learn more about the terminology as well as the intricacies of the field in which the experiment will be performed. It is also important to observe that there are many other practical constraints affecting the design of experiments. For example, experiments are done by organizations and individuals that have limited resources of money and time. Appropriating these resources within the constraints is an integral part of planning an experiment. Also, many problems are approached sequentially in several stages. Planning for each stage is built on what has been learned before. Dealing with these types of issues is beyond the scope of this book.

**EXERCISES 9.2**

- 9.2.1.** In order to study the conditions under which hash-brown potatoes will absorb the least amount of fat, an experiment is conducted with four frying durations (2 min, 3 min, 4 min, 5 min) and using four different types of fats (animal fat I, animal fat II, vegetable fat I, vegetable fat II). The amount of fat absorbed is recorded. Identify each of the following elements: response, factors, and factor type(s). Write all the treatments from the factor-level combinations.
- 9.2.2.** A team of scientists is interested in the effects of vitamin A, vitamin C, and vitamin D on the number of offspring born for a specific species of mice. An experiment is set up using the same species of mice. The mice are randomly assigned to three groups. Each mouse in the study gets the same amount of food and daily exercise and is kept at the same temperature. One group of mice gets extra vitamin A, another group gets extra vitamin C, and the remaining group gets extra vitamin D. The supplements are added to their food. The number of offspring are counted and recorded for each group.
- What is the response variable?
  - What is the factor?
- 9.2.3.** Thirty rose bushes are numbered 1 to 30. Three different fertilizers are to be applied to 10 bushes each. Develop a design using the random assignment procedure.
- 9.2.4.** Three different fertilizers are to be applied to five bushes each for three varieties of flower plants: gardenia (G), rose (R), and jasmine (J). Create a randomized complete block design. Choose as blocks the different types of plants (G, R, or J).
- 9.2.5.** With the following modifications, consider Exercise 9.2.4. Three groups of flower plants are considered, with each group having nine plants. Group I consists of gardenia (G), group II consists of rose (R), and group III consists of jasmine (J). Of the nine gardenias (group I), three are treated with brand A fertilizer, three with brand B, and three with brand C fertilizer. Other plant types are treated similarly. Here the number of replications is three for each type of fertilizer and for each type of plants. Create a randomized complete block design with three replications.
- 9.2.6.** What are the reasons for using randomization in Exercises 9.2.3 to 9.2.5?
- 9.2.7.** Suppose a food processing company wants to package sliced pineapples in cans. They have four different processing plants, say, A, B, C, and D. Suppose they have 56 truckloads (numbered 1 to 56) of pineapples collected from different parts of the country. In order to get some uniformity in taste, it is better to randomly assign the trucks to the four plants. Develop a design using the random assignment procedure.
- 9.2.8.** In Exercise 9.2.1, suppose there are four pans and 24 packets of hash-brown potatoes. Randomly select six of the 24 packets to be fried with each of the fat types.
- Create a randomized complete block design.
  - Create a Latin square design.

- 9.2.9.** A chemist is interested in the effects of five different catalysts (A, B, C, D, E) on the reaction time of a chemical process. There are five batches of new material (1, 2, 3, 4, 5). She decides to study the effect of each catalyst on each material for 5 days (1, 2, 3, 4, 5). Construct a Latin square design for this experiment.
- 9.2.10.** Suppose a dating service wants to schedule dates for four women, Anna, Carol, Judy, and Nancy, with Ed, John, Marcus, and Richard on Thursday, Friday, Saturday, and Sunday in such a way that each man dates each woman in the 4 days. Create a Latin-square design displaying a schedule that the dating service could follow.
- 9.2.11.** In order to test the relative effectiveness of four different fertilizer mixtures on an orange crop, a Florida farmer applies the fertilizer and measures the yield per unit area when it harvests. The four experiments cannot be carried out on the same plot of land. Devise a Latin square arrangement of dividing a single plot into a  $4 \times 4$  grid of subplots for administering the fertilizers (labeled randomly A, B, C, D).
- 9.2.12.** A researcher wants to know the effect of four different types of fertilizers on the mean number of tomatoes produced. He wants to estimate the treatment means  $\mu_1$ ,  $\mu_2$ ,  $\mu_3$ , and  $\mu_4$  such that he will be 90% confident that the estimates are within five tomatoes of the true mean number of tomatoes. What is the necessary number of replications to achieve this goal? It is known from previous experiments that the numbers of tomatoes per plant have ranged from 20 to 60.

### 9.3 FACTORIAL DESIGN

In this section, we introduce a treatment design where the treatments are constructed from several factors rather than just being  $k$  levels of a single factor. The treatments are combinations of levels of the factors. A *factorial experiment* can be defined as an experiment in which the response variable is observed at all factor-level combinations of the independent variables. A *factorial design* is used to evaluate two or more factors simultaneously. In general, there are three ways to obtain experimental data: one-factor-at-a-time, full factorial, and fractional factorial. The most efficient design is the fractional factorials. A simple approach for examining the effect of multiple factors is the one-at-a-time approach. The advantages of factorial designs over one-factor-at-a-time experiments is that they allow interactions to be spotted. An interaction occurs when the effect of one factor varies with the level of another factor or with some combination of levels of other factors when there are multiple factors.

The *one-way analysis of variance*, discussed in the next chapter, enables us to compare several groups of observations, all of which are independent with the possibility of a different mean for each group. A test of significance is whether or not all the means are equal. *Two-way analysis of variance* is a way of studying the effects of two factors separately, such as their main effects, and together, with their interaction effect.

#### 9.3.1 One-Factor-at-a-Time Design

In one-factor-at-a-time design, one conducts the experiment with one factor at a time. Here we hold all factors constant except one and take measurements on the response variable for several levels

of this one factor, then choose another factor to vary, keeping all others constant, and so forth. We are familiar with this type of experiment from undergraduate chemistry or physics labs. One of the drawbacks of this method is that all factors are evaluated while the other factors are at a single setting. For example, in the case of Example 9.2.2, we would set a fixed temperature and study the effect of water salinity on fish weight gains, and then set a fixed water salinity and vary temperature. All this is time consuming.

### Example 9.3.1

Consider the following hypothetical data, in which two types of diet (fat, carbohydrates) in two levels (high, medium) were administered for a week for a sample of individuals. At the end of the week, each subject was put on a treadmill and time of exhaustion, in seconds, was measured. The objective was to determine the factor-level combination that will give maximum time of exhaustion. Table 9.13 gives average time to exhaustion for each combination of diet.

Discuss this as a one-factor-at-a-time experiment to predict average time of exhaustion.

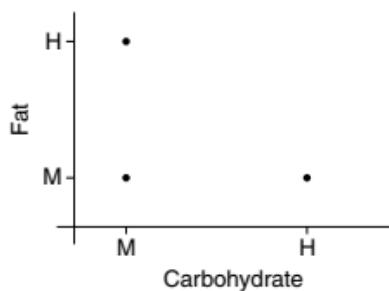
### Solution

We can see that the average time of exhaustion decreases when fat content is increased from medium to high while holding carbohydrate at medium. The average time of exhaustion also decreases when carbohydrate content is increased from medium to high while holding fat at medium. Thus, it is tempting to predict that increasing both fat and carbohydrate consumption will result in a lower average time of exhaustion. The problem with this reasoning is that the prediction is based on the assumption that the effect

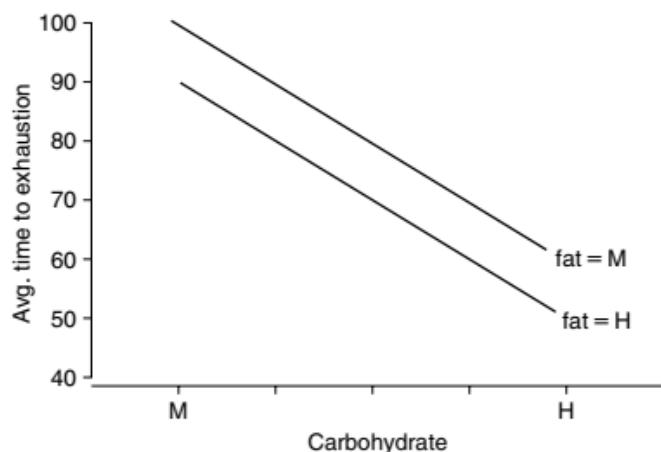
**Table 9.13**

Average time to exhaustion	Fat	Carbohydrate
88	High	Medium
98	Medium	Medium
77	Medium	High
74	High	High

of one factor is the same for both levels of the other factor. Changing the fat content from medium to high, keeping carbohydrate at medium, and the carbohydrate content from medium to high, keeping fat at medium, reduced the average time of exhaustion by approximately 10 seconds. The question then is, can we predict that increasing both fat and carbohydrate content to high will lower the average time of exhaustion to approximately 67 seconds? To answer this question, we need to administer high levels of both diets to a sample and observe the average time of exhaustion. If it is 67 seconds, then our observation is correct. However, what if the observation is 74 seconds? The average time of exhaustion has been lowered, but not as much. If this happens, we say that the two factors interact. When factors interact, the effect of one factor on the response is not the same for different levels of the other factor. Hence, the information obtained from the one-factor-at-a-time approach would lead to an invalid prediction.



■ FIGURE 9.1 One-factor-at-a-time approach.



■ FIGURE 9.2 No interaction.

The factor-level combination for a one-factor-at-a-time approach of Example 9.3.1 can be seen from Figure 9.1.

If there is no interaction, we get Figure 9.2, which shows average time to exhaustion with three given points and a possible point of around 68 seconds.

**Definition 9.3.1** *Two factors I and II are said to interact if the difference in mean responses for different levels of one factor is not constant across levels of the second factor.*

If there is interaction, the lines in Figure 9.2 might cross each other, in which case a one-factor-at-a-time approach may not be the appropriate design. In that case, the following alternative designs will give more accurate data.

### 9.3.2 Full Factorial Design

One way to get around the problem of interaction in one-factor-at-a-time design is to evaluate all possible combinations of factors in a single experiment. This is called a *full factorial experiment*. The main benefit of a full factorial design is that every possible data point is collected. The choice of

optimum condition becomes easy. For example, in an experiment such as the one in Example 9.2.2, one could conduct a full factorial design. The simplest form of factorial experiment involves two factors only and is called a *two-way layout*. A full factorial experiment with  $n$  factors and two levels for each factor is called a  $2^n$  *factorial experiment*. A full factorial experiment is practical if only a few factors (say, fewer than five) are being investigated. Beyond that, this design becomes time consuming and expensive.

### 9.3.3 Fractional Factorial Design

In a fractional factorial experiment, only a fraction of the possible treatments are actually used in the experiment. A full factorial design is the ideal design, through which we could obtain information on all main effects and interactions. But because of the prohibitive size of the experiments, such designs are not practical to run. For instance, consider Example 9.2.2. Now if we were to add say, two different densities, three sizes of fish, and three types of food, the number of factors becomes five, and total number of distinct treatments will be  $4 \times 4 \times 2 \times 3 \times 3 = 288$ . This method becomes very time consuming and expensive. The number of relatively significant effects in a factorial design is relatively small. In these types of situations, fractional factorial experiments are used in which trials are conducted on only a well-balanced subset of the possible combinations of levels of factors. This allows the experimenter to obtain information about all main effects and interactions while keeping the size of the experiment manageable. The experiment is carried out in a single systematic effort. However, care should be taken in selection of treatments in the experiment so as to be able to answer as many relevant questions as possible. The fractional factorial design is useful when the number of factors is large. Because we are reducing the number of factors, a fractional factorial design will not be able to evaluate the influence of some of the factors independently. Of course, the question is how to choose the factors and levels we should use in a fractional factorial design. The question of how fractional factorial designs are constructed is beyond the scope of this book.

## EXERCISES 9.3

- 9.3.1.** Suppose a large retail chain decides to introduce clothing in two types of materials' (ordinary, fine) qualities. Each store will have two different proportions (40%, 60%) displayed. At the end of the month, profits from each store for these two types of clothing are recorded. Table 9.3.1 represents the average profits for each of the quality–proportion combinations.

<b>Table 9.3.1</b>		
Average profit	Quality	Proportion
\$10,000	Fine	40%
\$25,000	Ordinary	40%
\$9500	Ordinary	60%
?	Fine	60%

Discuss this as a one-factor-at-a-time experiment to predict the average amount of profit.

- 9.3.2.** Draw graphs for the data to represent quality–proportion combinations (a) for the one-factor-at-a-time approach, and (b) for the case where there is no interaction.
- 9.3.3.** Discuss how a fractional factorial design can be performed for the problem in Exercise 9.3.1.
- 9.3.4.** Suppose a researcher wants to conduct a series of experiments to study the effect of fertilizer and temperature on plant growth. She uses four different brands of fertilizers in three different settings for the rose plants of the same age and of similar growth.
- How many factor-level combinations are possible in this experiment?
  - Each experiment makes use of one fertilizer–temperature combination (one-factor-at-a-time design). How should she implement randomization in this experiment?

## 9.4 OPTIMAL DESIGN

In 1959, J. Kiefer presented a paper to the Royal Statistical Society about his work on the theory of optimal design. He was trying to answer the major question, “How do we find the best design?” This work initiated a whole new field of optimal design. The methods of optimal experimental design provide the technical tools for building experimental designs to attain well-defined objectives with efficiency and with minimum cost. The cost can be the monetary cost, time, number of experimental runs, and so on. There are many methods of achieving optimal designs such as sequential (simplex) or simultaneous experiment designs. In sequential design, experiments are performed in succession in a direction of improvement until the optimum is reached. Simultaneous experiment designs such as response surface designs are used to build empirical models. A survey by Atkinson in 1988 contains many references on optimal design.

In this section, we focus only on one simple example to illustrate the ideas of optimal design in terms of choosing appropriate sample size. It is not possible to have a single design that is best for securing information concerning all types of population parameters. Indeed, it is beyond the scope of this section to present a general theory of optimal design.

### 9.4.1 Choice of Optimal Sample Size

The sample size estimation is an essential part of experimental design; otherwise, sample size may be very high or very low. If sample size is too low, the experiment will lack the accuracy to provide dependable answers to the questions we are investigating. If sample size is too large, time and resources will be wasted, often for insignificant gain. We now illustrate a simple case of optimal sample size determination.

Let  $X_{11}, \dots, X_{1n_1}$  be a random sample from population 1 with mean  $\mu_1$  and variance  $\sigma_1^2$  and  $X_{21}, \dots, X_{2n_2}$  be random samples from population 2 with mean  $\mu_2$  and variance  $\sigma_2^2$ . Assume that the two samples are independent. Then we know that  $\bar{X}_1 - \bar{X}_2$  is an unbiased estimator of  $\mu_1 - \mu_2$  with standard error

$$\begin{aligned}\sigma_{(\bar{X}_1 - \bar{X}_2)}^2 &= \text{Var}(\bar{X}_1 - \bar{X}_2) \\ &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.\end{aligned}$$

Suppose that there is a restriction that the total observations should be  $n$ , that is,  $n_1 + n_2 = n$ . Such a restriction may be due to cost factors or to a shortage of available subjects. An important design question is how to choose the sample sizes  $n_1$  and  $n_2$  so as to maximize the information in the data relevant to the parameter  $\mu_1 - \mu_2$ . We know that the samples contain maximum information when the standard error is minimum. Hence, the problem reduces to minimization of  $\text{Var}(\bar{X}_1 - \bar{X}_2)$ . Let  $a = \frac{n_1}{n}$  be the fraction on  $n$  observations that is assigned to sample 1. Then  $n_1 = na$  and  $n_2 = n(1-a)$ , and we have

$$\begin{aligned}\text{Var}(\bar{X}_1 - \bar{X}_2) &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \\ &= \frac{\sigma_1^2}{na} + \frac{\sigma_2^2}{n(1-a)}.\end{aligned}$$

The problem is now reduced to finding an  $a$  that minimizes the function  $g(a) = \frac{\sigma_1^2}{na} + \frac{\sigma_2^2}{n(1-a)}$ . This problem can be solved using calculus. By taking the derivative with respect to  $a$ ,  $\frac{d}{da}g(a)$  and equating it to zero, we have

$$-\frac{\sigma_1^2}{na^2} + \frac{\sigma_2^2}{n(1-a)^2} = 0.$$

Multiplying throughout by  $na^2(1-a)^2$ , we have

$$-\sigma_1^2(1-a^2) + \sigma_2^2a^2 = 0$$

which results in the quadratic equation

$$(\sigma_2^2 - \sigma_1^2)a^2 + 2\sigma_1^2a - \sigma_1^2 = 0.$$

Using the quadratic formula, we obtain the two roots as

$$a_1 = \frac{\sigma_1}{\sigma_1 + \sigma_2}$$

and

$$a_2 = \frac{\sigma_1}{\sigma_1 - \sigma_2}.$$

However,  $a_2$  cannot be the solution because, if  $\sigma_1 > \sigma_2$ , then  $a_2 > 1$ , otherwise  $a_2 < 0$ ; both are not admissible because  $a$  is a fraction. Hence,

$$a = \frac{\sigma_1}{\sigma_1 + \sigma_2} \quad \text{and} \quad 1 - a = \frac{\sigma_2}{\sigma_1 + \sigma_2}.$$

Using the second derivative test, we can verify that this indeed is a minimum for  $\text{var}(\bar{X}_1 - \bar{X}_2)$ . From this analysis we can see that the sample sizes that maximize the information in the data relevant to the parameter  $\mu_1 - \mu_2$  subject to the constraint  $n_1 + n_2 = n$  are

$$n_1 = \frac{\sigma_1}{\sigma_1 + \sigma_2} n \quad \text{and} \quad n_2 = \frac{\sigma_2}{\sigma_1 + \sigma_2} n.$$

As a special case, we can see that when  $\sigma_1^2 = \sigma_2^2$ , the optimal design is to take  $n_1 = n_2$ .

### EXERCISES 9.4

- 9.4.1. A total of 100 sample points were taken from two populations with variances  $\sigma_1^2 = 4$  and  $\sigma_2^2 = 9$ . Find  $n_1$  and  $n_2$  that will result in the maximum amount of information about  $(\mu_1 - \mu_2)$ .
- 9.4.2. Suppose in Exercise 9.4.1 we want to take  $n = n_1 = n_2$ . How large should  $n$  be to obtain the same information as that implied by the solution of Exercise 9.4.1?

## 9.5 THE TAGUCHI METHODS

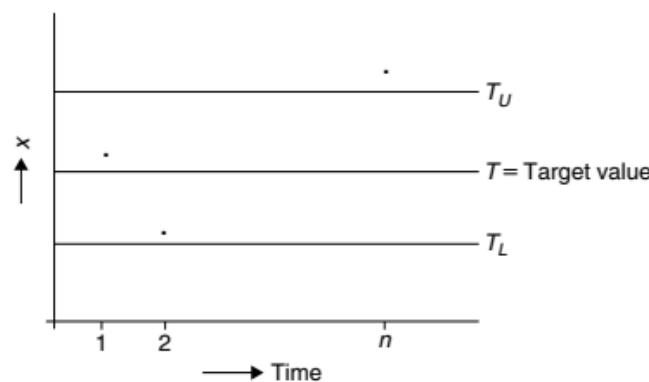
Taguchi methods were developed by Genichi Taguchi to improve the implementation of total quality control in Japan. These methods are claimed to have provided as much as 80% of Japanese quality gains. They are based on the design of experiments to provide near-optimal quality characteristics for a specific objective. A special feature of Taguchi methods is that they integrate the methods of statistical design of experiments into a powerful engineering process. The Taguchi methods are in general simpler to implement.

Taguchi methods are often applied on the Japanese manufacturing floor by technicians to improve their processes and their product. The goal is not just to optimize an arbitrary objective function, but also to reduce the sensitivity of engineering designs to uncontrollable factors or noise. The objective function used is the signal-to-noise ratio, which is then maximized. This moves design targets toward the middle of the design space so that external variation affects the behavior of the design as little as possible. This permits large reductions in both part and assembly tolerances, which are major drivers of manufacturing cost. Linking quality characteristics to cost through the Taguchi loss function (Taguchi and Yokoyama, 1994) was a major advance in quality engineering, as well as in the ability to design for cost. Taguchi methods are also called robust design. In 1982, the American Supplier Institute introduced Dr. Taguchi and his methods to the U.S. market.

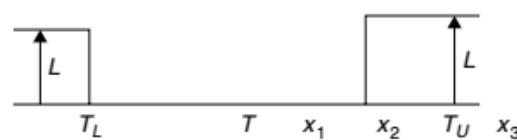
Using a well-planned experimental design, such as a fractional factorial design, it is possible to efficiently obtain information about the model and the underlying process. Clearly, the purpose of these methods is to control and ensure the quality of the end product. In the conventional approach, this is achieved by further testing a few end products that are randomly chosen or using control charts and making decisions based on certain preset criteria, such as acceptable or unacceptable. Thus, "quality" of the product is thought of as inside or outside of specifications. Instead, Taguchi suggested that we should specify a target value, and the quality should be thought of as the variation from the target.

As an example, suppose we make  $n$  observations of the output  $x_1, \dots, x_n$  of a process at times  $1, 2, \dots, n$ , as shown in Figure 9.3.

The control chart consists of a plot of observed output values ( $x_i$ 's) on the  $y$ -axis and the times of observation,  $1, 2, \dots, n$  on the  $x$ -axis, as shown in the figure. The letter  $T$  represents the target value. If



■ FIGURE 9.3 Control plot of processing times and outputs.



■ FIGURE 9.4 Loss function.

the output value is between  $T_L$  and  $T_U$ , the process is deemed to be operating satisfactorily; otherwise the process is said to be out of control and the output value is considered unsatisfactory.

Some other examples are (1) defining specification limits for acceptance, such as stating that the diameter of bolts must be between 9.8 mm and 10.2 mm with mean 10 mm, and (2) that the waiting time in a line should be less than 30 minutes for at least 90% of customers.

In all these situations, the specifications partition the state of the process as acceptable or unacceptable, that is, it classifies the state as a dichotomy. This is often called the “goal post mentality.”

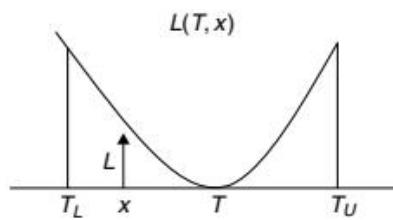
The basic idea of the Taguchi approach is a shift in mindset from demarking the quality as acceptable or unacceptable to a more flexible and realistic classification. The traditional approach to quality control does not take into account the size of departure from the target value. To accommodate the size of such departure as a significant factor in quality control, let us introduce the concept of loss function (see Chapter 11). If an output value  $x$  differs from the target value  $T$ , let  $L(T, x)$  denote the loss incurred, say in dollars. Other possible losses could also be reputation or customer satisfaction.

For the control chart example, we can assign the loss function

$$L(T, x) = \begin{cases} 0, & \text{if } T_U < x < T_L \\ L, & \text{if } x > T_L \text{ or } x < T_U \end{cases}$$

where  $L$  is a constant and  $x$  is the measured value. This is schematically shown in Figure 9.4.

From Figure 9.4, it is seen that we view outputs  $x_1$  and  $x_2$  as having equal quality, whereas  $x_2$  and  $x_3$  are considered to have vastly differing quality ( $x_2$  is acceptable and  $x_3$  is not acceptable). A more



■ FIGURE 9.5 Quadratic loss function.

reasonable conclusion would be that  $x_1$  has excellent quality, whereas  $x_2$  and  $x_3$  are similar, both being poor.

In Taguchi's approach, the loss function takes into account the size of departure from the target value. For example, a popular choice for the loss function is

$$L(T, X) = k(X - T)^2,$$

where

- $L$  = loss incurred,
- $k$  = constant,
- $X$  = actual value of the measured output, and
- $T$  = target value.

We can schematically represent the behavior as shown by Figure 9.5.

This form of loss function is called the *quadratic loss function*. The choice of  $k$  depends on the particular problem. For example, the scaling factor  $k$  can be used to convert loss into monetary units to accommodate comparisons of systems with different capital loss. Or, in product manufacturing, let  $D$  denote the allowed deviation from the target, and let  $A$  denote the loss due to a defective product. Then a choice of  $k$  can be  $k = (A/D)^2$ . As shown earlier, the average loss is  $E(L)$  and is given by

$$E(L) = k[(E(X) - T)^2 + \sigma^2] = k[(bias)^2 + variance]$$

where  $\sigma^2$  is the variance of  $X$  (measured quality, which is assumed to be random). In Taguchi, the variation from the target can be broken into components containing bias and product variation. Thus, if our aim is to minimize the expected loss,  $E(L)$ , we should not only require  $E(X) = \mu$  to be close to  $T$  but also should reduce the variance. It turns out that often these requirements are contradictory. The objective is to choose the design parameters (the factors that influence the quality) optimally to obtain the best quality product. In practice, the parameters  $\mu$  and  $\sigma^2$  are not known and are being estimated by  $\bar{X}$  and  $S^2$ , respectively. This results in the Taguchi loss function

$$\bar{L} = k[(\bar{X} - T)^2 + S^2].$$

This loss function penalizes small deviations from  $T$  only slightly, while assessing a larger penalty for responses far from the target. The expected loss is similar to a mean squared error loss, which we have seen in regression analysis in the form of least squares.

Why is controlling both bias and variance important? Suppose you want your community swimming pool temperature at 80°F, which is the  $T$  here. Suppose the temperature varies between 60°F and 100°F. Clearly the average (bias) is zero; however, it will be pretty uncomfortable to swim at 60°F or 100°F. Here the bias takes the ideal value of zero, but the variance is large. In another scenario, the variance may be small, but the average temperature may be farther away from the target value of 80°F (for example, the temperature is constant at 60°F). Hence, we want the pool temperature to be near to the target value of 80°F, with as small variance as possible (say, within 1°F to 2°F).

Taguchi coined the term *design parameters* as the generic description for factors that may influence the quality and whose levels we want to optimize. Taguchi's philosophy is to "design quality in" rather than to weed out the defective items after manufacturing. In order to obtain an optimal set of design parameters that affect the quality of the end product, the Taguchi method utilizes appropriately designed experiments. More specifically, orthogonal arrays are used for fractional factorial designs. Taguchi provides tables for these designs so that even a nonspecialist can use them. For two-level designs (high, low), we have a table for an  $L_4$  orthogonal array up to three factors; a table for an  $L_8$  orthogonal array up to seven factors; and so forth. Similar tables are available for three-level designs. We will not describe these design issues in this section. We refer the reader to specialized books on the subject for further details.

We can summarize the Taguchi approach to quality design as follows:

1. Taguchi's methods for experimental design are ready made and simple to use in the design of efficient experiments, even by nonexperts.
2. Taguchi's approach to total quality management is holistic and tries to design quality into a product rather than inspecting defects in the final product.
3. Taguchi's techniques can readily be applied to other fields such as management problems.

## EXERCISES 9.5

- 9.5.1.** Suppose the following data represent thickness between and within silicon wafers (in microns), with a target value of 14.5 microns.

13.688	13.788	14.173	14.557
13.925	14.545	13.797	14.778

Compute the Taguchi loss function.

- 9.5.2.** One of the commonly used performance measures in the Taguchi method is

$$\log \left( \frac{(\text{mean})^2}{s^2} \right),$$

where  $s^2$  is the sample variance. In general, the higher the performance measure, the better the design. This measure is called *robustness statistics*. For the problem of Exercise 9.5.1, suppose that we run the experiment by controlling various factors affecting the thickness. Table 9.5.1 shows the data obtained in four different runs.

**Table 9.5.1**

Run 1:	14.158	14.754	14.412	14.065	13.802	14.424	14.898	14.187
Run 2:	13.676	14.177	14.201	14.557	13.827	14.514	13.897	14.278
Run 3:	13.868	13.898	14.773	13.597	13.628	14.655	14.597	14.978
Run 4:	13.668	13.788	14.173	14.557	13.925	14.545	13.797	14.778

- (a) Using the robustness statistics given earlier, which of the processes gives us an improved performance?

- (b) Another commonly used performance statistic is

$$-\log(s^2).$$

Using this robustness statistic, which of the processes gives us an improved performance? Compare this with the results of part (a).

## 9.6 CHAPTER SUMMARY

In this chapter, we have learned some basic aspects of experimental design. Some fundamental definitions and tools for developing experimental designs such as randomization, replication, and blocking were introduced in Section 9.2. Basic concepts of factorial design were given in Section 9.3. In Section 9.4, we saw an example of optimal design. The Taguchi method was introduced in Section 9.5. In the next chapter, we introduce the analysis component. We have discussed only a very small collection of experimental designs in this chapter. There exist a wide variety of experimental designs to deal with a large number of treatments and to suit specific needs of research experiments in diverse fields. It is an exciting and growing area for the interested student to apply and explore.

We list some of the key definitions introduced in this chapter:

- Response variable (output variable)
- Independent variables (treatment variables or input variables or factors)
- Nuisance variables
- Noise
- Observational
- Experimental units
- Single-factor experiments
- Multifactor experiments
- Experimental error
- Blinding, double-blinding, and placebo
- Replication
- Block
- Randomization
- Completely randomized design
- Randomized complete block design
- $k \times k$  Latin square design

- Greco-Latin square
- design parameters

In this chapter, we have also learned the following important concepts and procedures.

- Procedure for random assignment
- Procedure for randomization in a randomized complete block design
- Procedure for a randomized complete block design with  $r$  replications
- Procedure for constructing a  $4 \times 4$  Latin square
- One-factor-at-a-time design
- Full factorial design
- Fractional factorial design
- Choice of optimal sample size
- The Taguchi methods

## 9.7 COMPUTER EXAMPLES

In this chapter, we present Minitab and SAS commands only. SPSS commands can be performed similarly to Minitab.

### 9.7.1 Minitab Examples

---

#### Example 9.7.1

Obtain a random permutation of numbers 1 to  $n$ .

**Solution**

Enter in **C1** the numbers 1 to  $n$ , say  $n = 10$ . Then

**Calc > random data > samples from column... >**  
enter sample **10** > rows from column(s) **C1** > Store samples in: **C2 > OK**

The result is a random permutation of numbers 1 to  $n (= 10)$ . One such permutation is given by

8 5 9 7 10 6 4 3 2 1

Now if we need to generate blocks of random permutations of numbers 1 to  $n (= 10)$ , in the foregoing steps, just store samples in **C3, C4, ...**

---

### 9.7.2 SAS Examples

---

#### Example 9.7.2

For the data of Example 9.2.4, conduct a randomized complete block design using SAS.

**Solution**

We represent blocks that are reasons for pain by  $H = 1$ ,  $M = 2$ , and  $CB = 3$ . Similarly five brands which are treatments by  $A = 1$ ,  $B = 2$ ,  $C = 3$ ,  $D = 4$ , and  $E = 5$ . Then we can use the following code to generate a randomized complete block design.

```
options nodate nonumber;
data a;
  do block = 1 to 3 ;
    do subject = 1 to 5;
      x = ranuni(0);
      output;
    end;
  end ;
proc sort; by block x;
data c; set a;
  trt = 1 + mod(N - 1, 5); /* mod = remainder of
                           N/5 */
proc sort; by block subject;
proc print;
  var block subject trt;
run;
```

We get the following output.

Completely randomized $2 \times 3$ design, 4 subjects per cell			
Obs	block	subject	trt
1	1	1	5
2	1	2	4
3	1	3	3
4	1	4	2
5	1	5	1
6	2	1	2
7	2	2	5
8	2	3	3
9	2	4	4
10	2	5	1
11	3	1	4
12	3	2	5
13	3	3	1
14	3	4	2
15	3	5	3

Note that the numbers in the column corresponding to a block identify the type of pain, the numbers in the subject column correspond to the subjects, and the numbers in the column corresponding to trt identify the brands. Using the corresponding letters, we can rewrite the foregoing table in the familiar form shown in Table 9.14.

<b>Table 9.14</b>		
<b>H</b>	<b>M</b>	<b>CB</b>
1(E)	1(B)	1(D)
2(D)	2(E)	2(E)
3(C)	3(C)	3(A)
4(B)	4(D)	4(B)
5(A)	4(A)	5(C)

The PLAN procedure constructs experimental designs. The PLAN procedure does not have a DATA= option in the PROC statement; in this procedure, both the input and output data sets are specified in the OUTPUT statement. We will use this to construct a Latin square design.

---

### Example 9.7.3

A gasoline company is interested in comparing the effect of four gasoline additives (A, B, C, D) on the gas mileage achieved per gallon. Four cars (1, 2, 3, 4) and four drivers (I, II, III, IV) will be used in the experiment. Create a Latin square design.

#### Solution

We can use the following program, where we represent the additives by 1 = A, 2 = B, 3 = C, and 4 = D.

```
Options nodate nonumber;
title 'Latin Square design for 4 additives';
proc plan seed=37432;
  factors rows=4 ordered cols=4 ordered/NOPRINT;
  treatments tmts=4 cyclic;
  output out=g
    rows cvals=('car 1' 'car 2' 'car 3' 'car 4')
      random
    cols cvals=('Driver 1' 'Driver 2' 'Driver 3'
      'Driver 4') random
    tmts nvals=(1 2 3 4) random;
run;
proc tabulate;
  class rows cols;
```

```

var tmst;
table rows, cols*(tmst*f=1.);
keylabel sum=' ';
run;

```

## PROJECTS FOR CHAPTER 9

### 9A. Sample Size and Power

Suppose that the experimenter is interested in comparing the true means of two independent populations. If two similar treatments are to be compared, the assumption of equality of variances is not unreasonable. Hence, assume that the common variance of the two populations is  $\sigma^2$ , and the experimenter has a prior estimate of the variance. We learned in Section 9.4 that in this case, the optimal design will be to take sample sizes  $n_1$  and  $n_2$  to be equal. Let  $n = n_1 = n_2$  be the size of the random sample that the experimenter should take from each population.

Now, suppose that the experimenter has decided to use the one-sided large sample test,  $H_0 : \mu_1 = \mu_2$  vs.  $H_a : \mu_1 > \mu_2$  with a fixed  $\alpha = P(\text{Type I error})$ . He wants to choose  $n$  to be so large that if  $\mu_1 = \mu_2 + k\sigma$ , he will get a fixed power  $(1 - \beta)$  of deciding  $\mu_1 > \mu_2$ . Recall that power of a test is the probability of (correctly) rejecting  $H_0$  when  $H_0$  is false. Find the approximate value of  $n$ . Note that, for a given  $\alpha$ , this will be an optimal sample size with a desired value of the power.

In particular, what should be the sample size in the hypothesis testing problem,  $H_0 : \mu_1 - \mu_2 = 0$  vs.  $H_a : \mu_1 - \mu_2 = 3$ , if  $\alpha = \beta = 0.05$ . Assume that  $\sigma = 7$ .

### 9B. Effect of Temperature on Spoilage of Milk

Suppose you have observed that milk in your refrigerator spoils very fast. You may be wondering whether it has anything to do with the temperature settings. Design an experiment to study the effect of temperature on spoiled milk, with at least three meaningful settings of the temperature. (i) Write a possible hypothesis for your experiment. (ii) What are the independent and dependent variables? (iii) Which variables are being controlled in this experiment? (iv) Discuss how you used the three basic principles of replication, blocking, and randomization. (v) What conclusions can you make? Think through any possible flaws in the design that may affect the integrity of your findings.