

---

## 2.1 Introduction

Although planning an experiment is an exciting process, it is extremely time-consuming. This creates a temptation to begin collecting data without giving the experimental design sufficient thought. Rarely will this approach yield data that have been collected in exactly the right way and in sufficient quantity to allow a good analysis with the required precision. This chapter gives a step by step guide to the experimental planning process. The steps are discussed in Sect. 2.2 and illustrated via real experiments in Sects. 2.3 and 2.5. Some standard experimental designs are described briefly in Sect. 2.4.

---

## 2.2 A Checklist for Planning Experiments

The steps in the following checklist summarize a very large number of decisions that need to be made at each stage of the experimental planning process. The steps are not independent, and at any stage, it may be necessary to go back and revise some of the decisions made at an earlier stage.

### Checklist

- (a) Define the objectives of the experiment.
- (b) Identify all sources of variation, including:
  - (i) treatment factors and their levels,
  - (ii) experimental units,
  - (iii) blocking factors, noise factors, and covariates.
- (c) Choose a rule for assigning the experimental units to the treatments.
- (d) Specify the measurements to be made, the experimental procedure, and the anticipated difficulties.
- (e) Run a pilot experiment.
- (f) Specify the model.
- (g) Outline the analysis.
- (h) Calculate the number of observations that need to be taken.
- (i) Review the above decisions. Revise, if necessary.

A short description of the decisions that need to be made at each stage of the checklist is given below. Only after all of these decisions have been made should the data be collected.

(a) **Define the objectives of the experiment.**

A list should be made of the precise questions that are to be addressed by the experiment. It is this list that helps to determine the decisions required at the subsequent stages of the checklist. It is advisable to list only the essential questions, since side issues will unnecessarily complicate the experiment, increasing both the cost and the likelihood of mistakes. On the other hand, questions that are inadvertently omitted may be unanswerable from the data. In compiling the list of objectives, it can often be helpful to outline the conclusions expected from the analysis of the data. The objectives may need to be refined as the remaining steps of the checklist are completed.

(b) **Identify all sources of variation.**

A source of variation is *anything* that could cause an observation to have a different numerical value from another observation. Some sources of variation are minor, producing only small differences in the data. Others are major and need to be planned for in the experiment. It is good practice to make a list of every conceivable source of variation and then label each as either major or minor. Major sources of variation can be divided into two types: those that are of particular interest to the experimenter, called “treatment factors,” and those that are not of interest, called “nuisance factors.”

(i) Treatment factors and their levels.

Although the term *treatment factor* might suggest a drug in a medical experiment, it is used to mean any substance or item whose effect on the data is to be studied. At this stage in the checklist, the treatment factors and their *levels* should be selected. The levels are the specific types or amounts of the treatment factor that will actually be used in the experiment. For example, a treatment factor might be a drug or a chemical additive or temperature or teaching method, etc. The levels of such treatment factors might be the different amounts of the drug to be studied, different types of chemical additives to be considered, selected temperature settings in the range of interest, different teaching methods to be compared, etc. Few experiments involve more than four levels per treatment factor.

If the levels of a treatment factor are quantitative (i.e., can be measured), then they are usually chosen to be equally spaced. Two levels are needed to model a linear trend, three levels for a quadratic trend, and so forth. If the response or log(response) should be well modeled by a rather simple function of the log of the factor level, then one may choose the factor levels to be equally spaced on a log scale. For convenience, treatment factor levels can be coded. For example, temperature levels 60, 70, 80°, . . . might be coded as 1, 2, 3, . . . in the plan of the experiment, or as 0, 1, 2, . . . With the latter coding, level 0 does not necessarily signify the absence of the treatment factor. It is merely a label. Provided that the experimenter keeps a clear record of the original choice of levels, no information is lost by working with the codes.

When an experiment involves more than one treatment factor, every observation is a measurement on some combination of levels of the various treatment factors. For example, if there are two treatment factors, temperature and pressure, whenever an observation is taken at a certain pressure, it must necessarily be taken at some temperature, and vice versa. Suppose there are four levels of temperature coded 1, 2, 3, 4 and three levels of pressure coded 1, 2, 3. Then there are twelve combinations of levels coded 11, 12, . . . , 43, where the first digit of each pair refers to the level

of temperature and the second digit to the level of pressure. Treatment factors are often labeled  $F_1, F_2, F_3, \dots$  or  $A, B, C, \dots$ . The combinations of their levels are called *treatment combinations* and an experiment involving two or more treatment factors is called a *factorial experiment*.

We will use the term *treatment* to mean a level of a treatment factor in a single factor experiment, or to mean a treatment combination in a factorial experiment.

(ii) Experimental units.

*Experimental units* are the “material” to which the levels of the treatment factor(s) are applied. For example, in agriculture these would be individual plots of land, in medicine they would be human or animal subjects, in industry they might be batches of raw material, factory workers, etc. If an experiment has to be run over a period of time, with the observations being collected sequentially, then the times of day can also be regarded as experimental units.

Experimental units should be representative of the material and conditions to which the conclusions of the experiment will be applied. For example, the conclusions of an experiment that uses university students as experimental units may not apply to all adults in the country. The results of a chemical experiment run in an 80° laboratory may not apply in a 60° factory. Thus it is important to consider carefully the scope of the experiment in listing the objectives in step (a).

It is important to distinguish experimental units from *observational units*—namely, what is measured to obtain observations. For example, in an experiment involving the feeding of animals in a pen to assess the effects of diet on weight gain, it may be that pens of animals fed together are the experimental units while the individual animals are the observational units. In most experiments, the experimental units and observational units are one and the same. However, when there is a distinction, it is important that the data analysis reflect it. Otherwise, mistakenly treating the observational units as experimental units would give the appearance that the experiment provides more data or replication than is indeed present.

(iii) Blocking factors, noise factors, and covariates.

An important part of designing an experiment is to enable the effects of the nuisance factors to be distinguished from those of the treatment factors. There are several ways of dealing with nuisance factors, depending on their nature.

It may be desirable to limit the scope of the experiment and to fix the level of the nuisance factor. This action may necessitate a revision of the objectives listed in step (a) since the conclusions of the experiment will not be so widely applicable. Alternatively, it may be possible to hold the level of a nuisance factor constant for one group of experimental units, change it to a different fixed value for a second group, change it again for a third, and so on. Such a nuisance factor is called a *blocking factor*, and experimental units measured under the same level of the blocking factor are said to be in the same *block* (see Chap. 10). For example, suppose that temperature was expected to have an effect on the observations in an experiment, but it was not itself a factor of interest. The entire experiment could be run at a single temperature, thus limiting the conclusions to that particular temperature. Alternatively, the experimental units could be divided into blocks with each block of units being measured at a different fixed temperature.

Even when the nuisance variation is not measured, it is still often possible to divide the experimental units into blocks of like units. For example, plots of land or times of day that are close together are more likely to be similar than those far apart. Subjects with similar characteristics are more likely to respond in similar ways to a drug than subjects with different characteristics. Observations made in the same factory are more likely to be similar than observations made in different factories.

Sometimes nuisance variation is a property of the experimental units and can be measured before the experiment takes place, (e.g., the blood pressure of a patient in a medical experiment, the I.Q. of a pupil in an educational experiment, the acidity of a plot of land in an agricultural experiment). Such a measurement is called a *covariate* and can play a major role in the analysis (see Chap. 9). Alternatively, the experimental units can be grouped into blocks, each block having a similar value of the covariate. The covariate would then be regarded as a blocking factor.

If the experimenter is interested in the variability of the response as the experimental conditions are varied, then nuisance factors are deliberately included in the experiment and not removed via blocking. Such nuisance factors are called *noise factors*, and experiments involving noise factors form the subject of *robust design*, discussed in Chap. 15.

(c) **Choose a rule by which to assign the experimental units to the levels of the treatment factors.**

The assignment rule, or the *experimental design*, specifies which experimental units are to be observed under which treatments. The choice of design, which may or may not involve blocking factors, depends upon all the decisions made so far in the checklist. There are several standard designs that are used often in practice, and these are introduced in Sect. 2.4. Further details and more complicated designs are discussed later in the book.

The actual assignment of experimental units to treatments should be done at random, subject to restrictions imposed by the chosen design. The importance of a random assignment was discussed in Sect. 1.1.4. Methods of randomization are given in Sect. 3.2.

There are some studies in which it appears to be impossible to assign the experimental units to the treatments either at random or indeed by any method. For example, if the study is to investigate the effects of smoking on cancer with human subjects as the experimental units, it is neither ethical nor possible to assign a person to smoke a given number of cigarettes per day. Such a study would therefore need to be done by observing people who have themselves chosen to be light, heavy, or nonsmokers throughout their lives. This type of study is an *observational study* and not an experiment. Although many of the analysis techniques discussed in this book could be used for observational studies, cause and effect conclusions are not valid, and such studies will not be discussed further.

(d) **Specify the measurements to be made, the experimental procedure, and the anticipated difficulties.**

The data (or observations) collected from an experiment are measurements of a response variable (e.g., the yield of a crop, the time taken for the occurrence of a chemical reaction, the output of a machine). The units in which the measurements are to be made should be specified, and these should reflect the objectives of the experiment. For example, if the experimenter is interested in detecting a difference of 0.5 gram in the response variable arising from two different treatments, it would not be sensible to take measurements to the nearest gram. On the other hand, it would be unnecessary to take measurements to the nearest 0.01 gram. Measurements to the nearest 0.1 gram would be sufficiently sensitive to detect the required difference, if it exists.

There are usually unforeseen difficulties in collecting data, but these can often be identified by taking a few practice measurements or by running a pilot experiment (see step (e)). Listing the anticipated difficulties helps to identify sources of variation required by step (b) of the checklist, and also gives the opportunity of simplifying the experimental procedure before the experiment begins.

Precise directions should be listed as to how the measurements are to be made. This might include details of the measuring instruments to be used, the time at which the measurements are to be made, the way in which the measurements are to be recorded. It is important that everyone involved in running the experiment follow these directions exactly. It is advisable to draw up a data collection sheet that shows the order in which the observations are to be made and also the units of measurement.

**(e) Run a pilot experiment.**

A pilot experiment is a mini experiment involving only a few observations. No conclusions are necessarily expected from such an experiment. It is run to aid in the completion of the checklist. It provides an opportunity to practice the experimental technique and to identify unsuspected problems in the data collection. If the pilot experiment is large enough, it can also help in the selection of a suitable model for the main experiment. The observed experimental error in the pilot experiment can help in the calculation of the number of observations required by the main experiment (step (h)).

At this stage, steps (a)–(d) of the checklist should be reevaluated and changes made as necessary.

**(f) Specify the model.**

The model must indicate explicitly the relationship that is believed to exist between the response variable and the major sources of variation that were identified at step (b). The techniques used in the analysis of the experimental data will depend upon the form of the model. It is important, therefore, that the model represent the true relationship reasonably accurately.

The most common type of model is the linear model, which shows the response variable set equal to a linear combination of terms representing the major sources of variation plus an error term representing all the minor sources of variation taken together. A pilot experiment (step (e)) can help to show whether or not the data are reasonably well described by the model.

There are two different types of treatment or block factors that need to be distinguished, since they lead to somewhat different analyses. The effect of a factor is said to be a *fixed effect* if the factor levels have been specifically selected by the experimenter and if the experimenter is interested in comparing the effects on the response variable of these specific levels. This is the most common type of factor and is the type considered in the early chapters. A model containing only fixed-effect factors (apart from the response and error random variables) is called a *fixed-effects model*.

Occasionally, however, a factor has an extremely large number of possible levels, and the levels included in the experiment are a random sample from the population of all possible levels. The effect of such a factor is said to be a *random effect*. Since the levels are not specifically chosen, the experimenter has little interest in comparing the effects on the response variable of the particular levels used in the experiment. Instead, it is the variability of the response due to the entire population of levels that is of interest. Models for which all factors are random effects are called *random-effects models*. Models for which some factors are random effects and others are fixed effects are called *mixed models*. Experiments involving random effects will be considered in Chaps. 17 and 18.

**(g) Outline the analysis.**

The type of analysis that will be performed on the experimental data depends on the objectives determined in step (a), the design selected in step (c), and its associated model specified in step (f).

The entire analysis should be outlined (including hypotheses to be tested and confidence intervals to be calculated). The analysis not only determines the calculations at step (h), but also verifies that the design is suitable for achieving the objectives of the experiment.

**(h) Calculate the number of observations needed.**

At this stage in the checklist, a calculation should be done for the number of observations that are needed in order to achieve the objectives of the experiment. If too few observations are taken, then the experiment may be inconclusive. If too many are taken, then time, energy, and money are needlessly expended.

Formulae for calculating the number of observations are discussed in Sects. 3.6 and 4.5 for the completely randomized design, and in later chapters for more complex designs. The formulae require a knowledge of the size of the experimental variability. This is the amount of variability in the data caused by the sources of variation designated as minor in step (b) (plus those sources that were forgotten!). Estimating the size of the experimental error prior to the experiment is not easy, and it is advisable to err on the large side. Methods of estimation include the calculation of the experimental error in a pilot experiment (step (e)) and previous experience of working with similar experiments.

**(i) Review the above decisions. Revise if necessary.**

Revision is necessary when the number of observations calculated at step (h) exceeds the number that can reasonably be taken within the time or budget available. Revision must begin at step (a), since the scope of the experiment usually has to be narrowed. If revisions are not necessary, then the data collection may commence.

It should now be obvious that a considerable amount of thought needs to precede the running of an experiment. The data collection is usually the most costly and the most time-consuming part of the experimental procedure. Spending a little extra time in planning helps to ensure that the data can be used to maximum advantage. No method of analysis can save a badly designed experiment.

Although an experimental scientist well trained in the principles of design and analysis of experiments may not need to consult a statistician, it usually helps to talk over the checklist with someone not connected with the experiment. Step (a) in the checklist is often the most difficult to complete. A consulting statistician's first question to a client is usually, "Tell me *exactly* why you are running the experiment. *Exactly* what do you want to show?" If these questions cannot be answered, it is not sensible for the experimenter to go away, collect some data, and worry about it later. Similarly, it is essential that a consulting statistician understand reasonably well not only the purpose of the experiment but also the experimental technique. It is not helpful to tell an experimenter to run a pilot experiment that eats up most of the budget.

The experimenter needs to give clear directions concerning the experimental procedure to all persons involved in running the experiment and in collecting the data. It is also necessary to check that these directions are being followed exactly as prescribed. An amusing anecdote told by Salvadori (1980) in his book *Why Buildings Stand Up* illustrates this point. The story concerns a quality control study of concrete. Concrete consists of cement, sand, pebbles, and water and is mixed in strictly controlled proportions in a concrete plant. It is then carried to a building site in a revolving drum on a large truck. A sample of concrete is taken from each truckload and, after seven days, is tested for compressive strength. Its strength depends partly upon the ratio of water to cement, and decreases as the proportion of water increases. The anecdote concerns a problem that occurred during construction of an airport

terminal in New York. Although the concrete reaching the site before noon showed good strength, some of the concrete arriving shortly after noon did not. The supervisor investigated the most plausible causes until he decided to follow the trucks as they went from the plant to the site. He spotted a truck driver regularly stopping for beer and a sandwich at noon, and to prevent the concrete hardening, he added extra water into the drums. Thus, Salvadori concludes “the prudent engineer must not only be cautious about material properties, but be aware, most of all, of human behavior.”

This applies to prudent experimenters, too! In the chapters that follow, most of the emphasis falls on the statistical analysis of well-designed experiments. It is crucial to keep in mind the ideas in these first sections while reading the rest of the book. Unfortunately, there are no nice formulae to summarize everything. Both the experimenter and the statistical consultant should use the checklist and lots of common sense!

---

### 2.3 A Real Experiment—Cotton-Spinning Experiment

The experiment to be described was reported in the November 1953 issue of the journal *Applied Statistics* by Robert Peake, of the British Cotton Industry Research Association. Although the experiment was run many years ago, the types of decisions involved in planning experiments have changed very little. The original report was not written in checklist form, but all of the relevant details were provided by the author in the article.

#### Checklist

(a) **Define the objectives of the experiment.**

At an intermediate stage of the cotton-spinning process, a strand of cotton (known as “roving”) thicker than the final thread is produced. Roving is twisted just before it is wound onto a bobbin. As the degree of twist increases, so does the strength of the cotton, but unfortunately, so does the production time and hence, the cost. The twist is introduced by means of a rotary guide called a “flyer.” The purpose of the experiment was twofold; first, to investigate the way in which different degrees of twist (measured in turns per inch) affected the breakage rate of the roving, and secondly, to compare the ordinary flyer with the newly devised special flyer.

(b) **Identify all sources of variation.**

(i) Treatment factors and their levels.

There are two treatment factors, namely “type of flyer” and “degree of twist.” The first treatment factor, flyer, has two levels, “ordinary” and “special.” We code these as 1 and 2, respectively. The levels of the second treatment factor, twist, had to be chosen within a feasible range. A pilot experiment was run to determine this range, and four non equally spaced levels were selected, 1.63, 1.69, 1.78, and 1.90 turns per inch. Coding these levels as 1, 2, 3, and 4, there are eight possible treatment combinations, as shown in Table 2.1.

The two treatment combinations 11 and 24 were omitted from the experiment, since the pilot experiment showed that these did not produce satisfactory roving. The experiment was run with the six treatment combinations 12, 13, 14, 21, 22, 23.

**Table 2.1** Treatment combinations for the cotton-spinning experiment

Flyer	Twist			
	1.63	1.69	1.78	1.90
Ordinary	(11)	12	13	14
Special	21	22	23	(24)

(ii) Experimental units.

An experimental unit consisted of the thread on the set of full bobbins in a machine on a given day. It was not possible to assign different bobbins in a machine to different treatment combinations. The bobbins needed to be fully wound, since the tension, and therefore the breakage rate, changed as the bobbin filled. It took nearly one day to wind each set of bobbins completely.

(iii) Blocking factors, noise factors, and covariates.

Apart from the treatment factors, the following sources of variation were identified: the different machines, the different operators, the experimental material (cotton), and the atmospheric conditions.

There was some discussion among the experimenters over the designation of the blocking factors. Although similar material was fed to the machines and the humidity in the factory was controlled as far as possible, it was still thought that the experimental conditions might change over time. A blocking factor representing the day of the experiment was contemplated. However, the experimenters finally decided to ignore the day-to-day variability and to include just one blocking factor, each of whose levels represented a machine with a single operator. The number of experimental units per block was limited to six to keep the experimental conditions fairly similar within a block.

(c) **Choose a rule by which to assign the experimental units to the treatments.**

A randomized complete block design, which is discussed in detail in Chap. 10, was selected. The six experimental units in each block were randomly assigned to the six treatment combinations. The design of the final experiment was similar to that shown in Table 2.2.

(d) **Specify the measurements to be made, the experimental procedure, and the anticipated difficulties.**

It was decided that a suitable measurement for comparing the effects of the treatment combinations was the number of breaks per hundred pounds of material. Since the job of machine operator included mending every break in the roving, it was easy for the operator to keep a record of every break that occurred.

The experiment was to take place in the factory during the normal routine. The major difficulties were the length of time involved for each observation, the loss of production time caused by changing the flyers, and the fact that it was not known in advance how many machines would be available for the experiment.

**Table 2.2** Part of the design for the cotton-spinning experiment

Block	Time order					
	1	2	3	4	5	6
I	22	12	14	21	13	23
II	21	14	12	13	22	23
III	23	21	14	12	13	22
IV	23	21	12	...	...	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮

**(e) Run a pilot experiment.**

The experimental procedure was already well known. However, a pilot experiment was run in order to identify suitable levels of the treatment factor “degree of twist” for each of the flyers; see step (b).

**(f) Specify the model.**

The model was of the form

$$\text{Breakage rate} = \text{constant} + \text{effect of treatment combination} + \text{effect of block} + \text{error} .$$

Models of this form and the associated analyses are discussed in Chap. 10.

**(g) Outline the analysis.**

The analysis was planned to compare differences in the breakage rates caused by the six flyer/twist combinations. Further, the trend in breakage rates as the degree of twist was increased was of interest for each flyer separately.

**(h) Calculate the number of observations that need to be taken.**

The experimental variability was estimated from a previous experiment of a somewhat different nature. This allowed a calculation of the required number of blocks to be done (see Sect. 10.5.2). The calculation was based on the fact that the experimenters wished to detect a true difference in breakage rates of at least 2 breaks per 100 pounds with high probability. The calculation suggested that 56 blocks should be observed (a total of 336 observations!).

**(i) Review the above decisions. Revise, if necessary.**

Since each block would take about a week to observe, it was decided that 56 blocks would not be possible. The experimenters decided to analyze the data after the first 13 blocks had been run. The effect of decreasing the number of observations from the number calculated is that the requirements stated in step (h) would not be met. The probability of detecting differences of 2 breaks per 100 lbs was substantially reduced.

**Table 2.3** Data from the cotton-spinning experiment

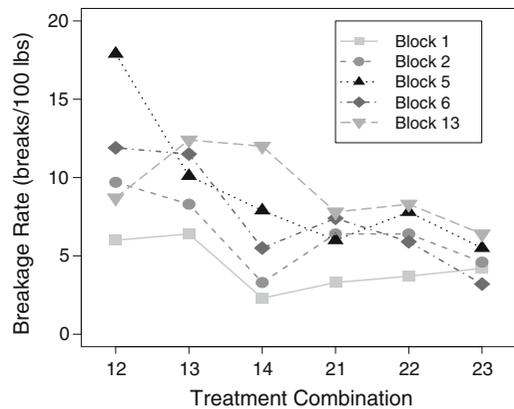
Treatment combination	Block number					
	1	2	3	4	5	6
12	6.0	9.7	7.4	11.5	17.9	11.9
13	6.4	8.3	7.9	8.8	10.1	11.5
14	2.3	3.3	7.3	10.6	7.9	5.5
21	3.3	6.4	4.1	6.9	6.0	7.4
22	3.7	6.4	8.3	3.3	7.8	5.9
23	4.2	4.6	5.0	4.1	5.5	3.2

Treatment combination	Block number						
	7	8	9	10	11	12	13
12	10.2	7.8	10.6	17.5	10.6	10.6	8.7
13	8.7	9.7	8.3	9.2	9.2	10.1	12.4
14	7.8	5.0	7.8	6.4	8.3	9.2	12.0
21	6.0	7.3	7.8	7.4	7.3	10.1	7.8
22	8.3	5.1	6.0	3.7	11.5	13.8	8.3
23	10.1	4.2	5.1	4.6	11.5	5.0	6.4

Source Peake (1953). Copyright © 1953 Royal Statistical Society. Reprinted with permission

**Fig. 2.1** A subset of the data for the cotton-spinning experiment



The results from the 13 blocks are shown in Table 2.3, and the data from five of these are plotted in Fig. 2.1. The data show that there are certainly differences in blocks. For example, results in block 5 are consistently above those for block 1. The breakage rate appears to be somewhat higher for treatment combinations 12 and 13 than for 23. However, the observed differences may not be any larger than the inherent variability in the data. Therefore, it is important to subject these data to a careful statistical analysis. This will be done in Sect. 10.5.

## 2.4 Some Standard Experimental Designs

An experimental design is a rule that determines the assignment of the experimental units to the treatments. Although experiments differ from each other greatly in most respects, there are some standard designs that are used frequently. These are described briefly in this section.

### 2.4.1 Completely Randomized Designs

A *completely randomized design* is the name given to a design in which the experimenter assigns the experimental units to the treatments completely at random, subject only to the number of observations to be taken on each treatment. Completely randomized designs are used for experiments that involve no blocking factors. They are discussed in depth in Chaps. 3–9 and again in some of the later chapters. The mechanics of the randomization procedure are illustrated in Sect. 3.2. The statistical properties of the design are completely determined by specification of  $r_1, r_2, \dots, r_v$ , where  $r_i$  denotes the number of observations on the  $i$ th treatment,  $i = 1, \dots, v$ .

The model is of the form

$$\text{Response} = \text{constant} + \text{effect of treatment} + \text{error} .$$

Factorial experiments often have a large number of treatments. This number can even exceed the number of available experimental units, so that only a subset of the treatment combinations can be observed. Special methods of design and analysis are needed for such experiments, and these are discussed in Chap. 15.

### 2.4.2 Block Designs

A *block design* is a design in which the experimenter partitions the experimental units into blocks, determines the allocation of treatments to blocks, and assigns the experimental units within each block to the treatments completely at random. Block designs are discussed in depth in Chaps. 10–14.

In the analysis of a block design, the blocks are treated as the levels of a single blocking factor even though they may be defined by a combination of levels of more than one nuisance factor. For example, the cotton-spinning experiment of Sect. 2.3 is a block design with each block corresponding to a combination of a machine and an operator. The model is of the form

$$\begin{aligned} \text{Response} = & \text{constant} + \text{effect of block} \\ & + \text{effect of treatment} + \text{error} . \end{aligned}$$

The simplest block design is the *complete block design*, in which each treatment is observed the same number of times in each block. Complete block designs are easy to analyze. A complete block design whose blocks contain a single observation on each treatment is called a *randomized complete block design* or, simply, a *randomized block design*.

When the block size is smaller than the number of treatments, so that it is not possible to observe every treatment in every block, a block design is called an *incomplete block design*. The precision with which treatment effects can be compared and the methods of analysis that are applicable depend on the choice of the design. Some standard design choices, and appropriate methods of randomization, are covered in Chap. 11. Incomplete block designs for factorial experiments are discussed in Chap. 13.

### 2.4.3 Designs with Two or More Blocking Factors

When an experiment involves two major sources of variation that have each been designated as blocking factors, these blocking factors are said to be either *crossed* or *nested*. The difference between these is illustrated in Table 2.4. Each experimental unit occurs at some combination of levels of the two blocking

**Table 2.4** Schematic plans of experiments with two blocking factors

(i) Crossed blocking factors				(ii) Nested blocking factors					
		Block					Block		
		Factor 1					Factor 1		
		1	2	3			1	2	3
Block	1	*	*	*	1		*		
Factor	2	*	*	*	2		*		
	3	*	*	*	3		*		
					Block	4		*	
					Factor	5		*	
					2	6		*	
						7			*
						8			*
						9			*

factors, and an asterisk denotes experimental units that are to be assigned to treatment factors. It can be seen that when the block factors are crossed, experimental units are used from all possible combinations of levels of the blocking factors. When the block factors are nested, a particular level of one of the blocking factors occurs at only one level of the other blocking factor.

**Crossed Blocking Factors**

A design involving two crossed blocking factors is sometimes called a “row–column” design. This is due to the pictorial representation of the design, in which the levels of one blocking factor are represented by rows and the levels of the second are represented by columns as in Table 2.4(i). An intersection of a row and a column is called a “cell.” Experimental units in the same cell should be similar. The model is of the form

$$\text{Response} = \text{constant} + \text{effect of row block} + \text{effect of column block} + \text{effect of treatment} + \text{error} .$$

Some standard choices of row–column designs with one experimental unit per cell are discussed in Chap. 12, and an example is given in Sect. 2.5.3 (p. 26) of a row–column design with six experimental units per cell.

The example shown in Table 2.5 is a basic design (prior to randomization) that was considered for the cotton-spinning experiment. The two blocking factors were “machine with operator” and “day.” Notice that if the column headings are ignored, the design looks like a randomized complete block design. Similarly, if the row headings are ignored, the design with columns as blocks looks like a randomized complete block design. Such designs are called Latin squares and are discussed in Chap. 12. For the cotton-spinning experiment, which was run in the factory itself, the experimenters could not guarantee that the same six machines would be available for the same six days, and this led them to select a randomized complete block design. Had the experiment been run in a laboratory, so that every machine was available on every day, the Latin square design would have been used, and the day-to-day variability could have been removed from the analysis of treatments.

**Nested (or Hierarchical) Blocking Factors.**

Two blocking factors are said to be nested when observations taken at two different levels of one blocking factor are automatically at two different levels of the second blocking factor as in Table 2.4(ii).

**Table 2.5** A Latin square for the cotton-spinning experiment

Machine with operator	Days					
	1	2	3	4	5	6
1	12	13	14	21	22	23
2	13	14	21	22	23	12
3	14	21	22	23	12	13
4	22	23	12	13	14	21
5	23	12	13	14	21	22
6	21	22	23	12	13	14

As an example, consider an experiment to compare the effects of a number of diets (the treatments) on the weight (the response variable) of piglets (the experimental units). Piglets vary in their metabolism, as do human beings. Therefore, the experimental units are extremely variable. However, some of this variability can be controlled by noting that piglets from the same litter are more likely to be similar than piglets from different litters. Also, litters from the same sow are more likely to be similar than litters from different sows. The different sows can be regarded as blocks, the litters regarded as subblocks, and the piglets as the experimental units within the subblocks. A piglet belongs only to one litter (piglets are nested within litters), and a litter belongs only to one sow (litters are nested within sows). The random assignment of piglets to diets would be done separately litter by litter in exactly the same way as for any block design.

In the industrial setting, the experimental units may be samples of some experimental material (e.g., cotton) taken from several different batches that have been obtained from several different suppliers. The samples, which are to be assigned to the treatments, are “nested within batches,” and the batches are “nested within suppliers.” The random assignment of samples to treatment factor levels is done separately batch by batch.

In an ordinary block design, the experimental units can be thought of as being nested within blocks. In the above two examples, an extra “layer” of nesting is apparent. Experimental units are nested within subblocks, subblocks are nested within blocks. The subblocks can be assigned at random to the levels of a further treatment factor. When this is done, the design is often known as a *split-plot design* (see Sect. 2.4.4).

#### 2.4.4 Split-Plot Designs

A *split-plot design* is a design with at least one blocking factor where the experimental units within each block are assigned to the treatment factor levels as usual, and *in addition*, the blocks are assigned at random to the levels of a further treatment factor. This type of design is used when the levels of one (or more) treatment factors are easy to change, while the alteration of levels of other treatment factors are costly, or time-consuming. For example, this type of situation occurred in the cotton-spinning experiment of Sect. 2.3. Setting the degree of twist involved little more than a turn of a dial, but changing the flyers involved stripping down the machines. The experiment was, in fact, run as a randomized complete block design, as shown in Table 2.2. However, it could have been run as a split-plot design, as shown in Table 2.6. The time slots have been grouped into blocks, which have been assigned at random to the two flyers. The three experimental units within each cell have been assigned at random to degrees of twist.

**Table 2.6** A split-plot design for the cotton-spinning experiment

	Time order					
	1	2	3	4	5	6
		Block I			Block II	
		Flyer 2			Flyer 1	
Machine I	Twist 2	Twist 1	Twist 3	Twist 2	Twist 4	Twist 3
		Flyer 2			Flyer 1	
Machine II	Twist 1	Twist 2	Twist 3	Twist 4	Twist 2	Twist 3
		Flyer 1			Flyer 2	
Machine III	Twist 4	Twist 2	Twist 3	Twist 3	Twist 1	Twist 2
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Split-plot designs also occur in medical and psychological experiments. For example, suppose that several subjects are assigned at random to the levels of a drug. In each time-slot each subject is asked to perform one of a number of tasks, and some response variable is measured. The subjects can be regarded as blocks, and the time-slots for each subject can be regarded as experimental units within the blocks. The blocks and the experimental units are each assigned to the levels of the treatment factors—the subject to drugs and the time-slots to tasks. Split-plot designs are discussed in detail in Chap. 19.

In a split-plot design, the effect of a treatment factor whose levels are assigned to the experimental units is generally estimated more precisely than a treatment factor whose levels are assigned to the blocks. It was this reason that led the experimenters of the cotton-spinning experiment to select the randomized complete block design in Table 2.2 rather than the split-plot design of Table 2.6. They preferred to take the extra time in running the experiment rather than risk losing precision in the comparison of the flyers.

## 2.5 More Real Experiments

Three experiments are described in this section. The first, called the “soap experiment,” was run as a class project by Suyapa Silvia in 1985. The second, called the “battery experiment,” was run by one of the authors. Both of these experiments are designed as completely randomized designs. The first has one treatment factor at three levels while the second has two treatment factors, each at two levels. The soap and battery experiments are included here to illustrate the large number of decisions that need to be made in running even the simplest investigations. Their data are used in Chaps. 3–5 to illustrate methods of analysis. The third experiment, called the “cake-baking experiment,” includes some of the more complicated features of the designs discussed in Sect. 2.4.

### 2.5.1 Soap Experiment

The checklist for this experiment has been obtained from the experimenter’s report. Our comments are in parentheses. The reader is invited to critically appraise the decisions made by this experimenter and to devise alternative ways of running her experiment.

**Checklist (Suyapa Silvia, 1985)****(a) Define the objectives of the experiment.**

The purpose of this experiment is to compare the extent to which three particular types of soap dissolve in water. It is expected that the experiment will answer the following questions: Are there any differences in weight loss due to dissolution among the three soaps when allowed to soak in water for the same length of time? What are these differences?

Generalizations to other soaps advertised to be of the same type as the three used for this experiment cannot be made, as each soap differs in terms of composition, i.e., has different mixtures of ingredients. Also, because of limited laboratory equipment, the experimental conditions imposed upon these soaps cannot be expected to mimic the usual treatment of soaps, i.e., use of friction, running water, etc. Conclusions drawn can only be discussed in terms of the conditions posed in this experiment, although they could give indications of what the results might be under more normal conditions.

(We have deleted the details of the actual soaps used).

**(b) Identify all sources of variation.****(i) Treatment factors and their levels**

The treatment factor, soap, has been chosen to have three levels: regular, deodorant, and moisturizing brands, all from the same manufacturer. The particular brands used in the experiment are of special interest to this experimenter.

The soap will be purchased at local stores and cut into cubes of similar weight and size—about 1'' cubes. The cubes will be cut out of each bar of soap using a sharp hacksaw so that all sides of the cube will be smooth. They will then be weighed on a digital laboratory scale showing a precision of 10 mg. The weight of each cube will be made approximately equal to the weight of the smallest cube by carefully shaving thin slices from it. A record of the preexperimental weight of each cube will be made.

(Note that the experimenter has no control over the age of the soap used in the experiment. She is assuming that the bars of soap purchased will be typical of the population of soap bars available in the stores. If this assumption is not true, then the results of the experiment will not be applicable in general. Each cube should be cut from a different bar of soap purchased from a random sample of stores in order for the experiment to be as representative as possible of the populations of soap bars.)

**(ii) Experimental units**

The experiment will be carried out using identical metal muffin pans. Water will be heated to 100°F (approximate hot bath temperature), and each section will be quickly filled with 1/4 cup of water. A pilot study indicated that this amount of water is enough to cover the tops of the soaps. The water-filled sections of the muffin pans are the experimental units, and these will be assigned to the different soaps as described in step (c).

**(iii) Blocking factors, noise factors, and covariates**

(Apart from the differences in the composition of the soaps themselves, the initial sizes of the cubes were not identical, and the sections of the muffin pan were not necessarily all exposed to the

same amount of heat. The initial sizes of the cubes were measured by weight. These could have been used as covariates, but the experimenter chose instead to measure the weight changes, that is, “final weight minus initial weight.” The sections of the muffin pan could have been grouped into blocks with levels such as “outside sections,” “inside sections,” or such as “center of heating vent” and “off-center of heating vent.” However, the experimenter did not feel that the experimental units would be sufficiently variable to warrant blocking. Other sources of variation include inaccuracies of measuring initial weights, final weights, amounts and temperature of water. All of these were designated as minor. No noise factors were incorporated into the experiment.)

(c) **Choose a rule by which to assign the experimental units to the levels of the treatment factors.**

An equal number of observations will be made on each of the three treatment factor levels. Therefore,  $r$  cubes of each type of soap will be prepared. These cubes will be randomly matched to the experimental units (muffin pan sections) using a random-number table.

(This assignment rule defines a completely randomized design with  $r$  observations on each treatment factor level, see Chap. 3).

(d) **Specify the measurements to be made, the experimental procedure, and the anticipated difficulties.**

The cubes will be carefully placed in the water according to the assignment rule described in paragraph (c). The pans will be immediately sealed with aluminum foil in order to prevent excessive moisture loss. The pans will be positioned over a heating vent to keep the water at room temperature. Since the sections will be assigned randomly to the cubes, it is hoped that if water temperature differences do exist, these will be randomly distributed among the three treatment factor levels. After 24 hours, the contents of the pans will be inverted onto a screen and left to drain and dry for a period of 4 days in order to ensure that the water that was absorbed by each cube has been removed thoroughly. The screen will be labeled with the appropriate soap numbers to keep track of the individual soap cubes.

After the cubes have dried, each will be carefully weighed. These weights will be recorded next to the corresponding preexperimental weights to study the changes, if any, that may have occurred. The analysis will be carried out on the differences between the post- and preexperimental weights.

### Expected Difficulties

- (i) The length of time required for a cube of soap to dissolve noticeably may be longer than is practical or assumed. Therefore, the data may not show any differences in weights.
- (ii) Measuring the partially dissolved cubes may be difficult with the softer soaps (e.g., moisturizing soap), since they are likely to lose their shape.
- (iii) The drying time required may be longer than assumed and may vary with the soaps, making it difficult to know when they are completely dry.
- (iv) The heating vent may cause the pan sections to dry out prematurely.

(After the experiment was run, Suyapa made a list of the actual difficulties encountered. They are reproduced below. Although she had run a pilot experiment, it failed to alert her to these difficulties ahead of time, since not all levels of the treatment factor had been observed.)

### Difficulties Encountered

- (i) When the cubes were placed in the warm water, it became apparent that some soaps absorbed water very quickly compared to others, causing the tops of these cubes to become exposed eventually. Since this had not been anticipated, no additional water was added to these chambers in order to keep the experiment as designed. This created a problem, since the cubes of soap were not all completely covered with water for the 24-hour period.
- (ii) The drying time required was also different for the regular soap compared with the other two. The regular soap was still moist, and even looked bigger, when the other two were beginning to crack and separate. This posed a real dilemma, since the loss of weight due to dissolution could not be judged unless all the water was removed from the cubes. The soaps were observed for two more days after the data was collected and the regular soap did lose part of the water it had retained.
- (iii) When the contents of the pans were deposited on the screen, it became apparent that the dissolved portion of the soap had become a semisolid gel, and a decision had to be made to regard this as “nonusable” and not allow it to solidify along with the cubes (which did not lose their shape).

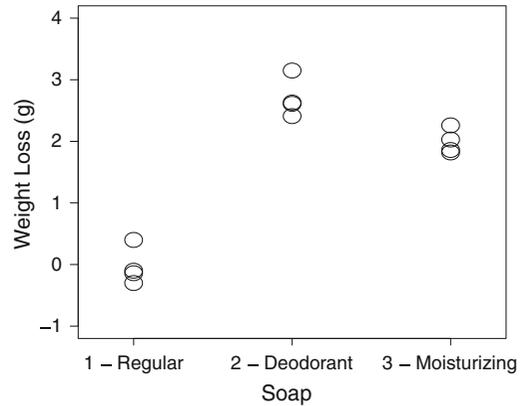
(The remainder of the checklist together with the analysis is given in Sect. 3.7. The calculations at step (h) showed that four observations should be taken on each soap type. The data were collected and are shown in Table 2.7. A plot of the data is shown in Fig. 2.2.)

The weightloss for each cube of soap measured in grams to the nearest 0.01 gm is the difference between the initial weight of the cube (pre-weight) and the weight of the same cube at the end of the experiment (post-weight). Negative values indicate a weight gain, while positive values indicate a weight loss (a large value being a greater loss). As can be seen, the regular soap cubes experienced the smallest changes in weight, and in fact, appear to have retained some of the water. Possible reasons for this will be examined in the discussion section (see Sect. 3.7.3). The data show a clear difference in the weight loss of the different soap types. This will be verified by a statistical hypothesis test (Sect. 3.7.2).

**Table 2.7** Weight loss for soaps in the soap experiment

Soap (Level)	Cube	Pre-weight (grams)	Post-weight (grams)	Weightloss (grams)
Regular (1)	1	13.14	13.44	-0.30
	2	13.17	13.27	-0.10
	3	13.17	13.31	-0.14
	4	13.17	12.77	0.40
Deodorant (2)	5	13.03	10.40	2.63
	6	13.18	10.57	2.61
	7	13.12	10.71	2.41
	8	13.19	10.04	3.15
Moisturizing (3)	9	13.14	11.28	1.86
	10	13.19	11.16	2.03
	11	13.06	10.80	2.26
	12	13.00	11.18	1.82

**Fig. 2.2** Weight loss for the soap experiment



## 2.5.2 Battery Experiment

### Checklist

(a) **Define the objectives of the experiment.**

Due to the frequency with which his family needed to purchase flashlight batteries, one of the authors (Dan Voss) was interested in finding out which type of nonrechargeable battery was the most economical. In particular, Dan was interested in comparing the lifetime per unit cost of the particular name brand that he most often purchased with the store brand where he usually shopped. He also wanted to know whether it was worthwhile paying the extra money for alkaline batteries over heavy duty batteries.

A further objective was to compare the lifetimes of the different types of battery regardless of cost. This was due to the fact that whenever there was a power cut, all the available flashlights appeared to have dead batteries! (Only the first objective will be discussed in Chaps. 3 and 4. The second objective will be addressed in Chap. 5.)

(b) **Identify all sources of variation.**

There are several sources of variation that are easy to identify in this experiment. Clearly, different duty batteries such as alkaline and heavy duty could well be an important factor in the lifetime per unit cost, as could the brand of the battery. These two sources of variation are the ones of most interest in the experiment and form the levels of the two treatment factors “duty” and “brand.” Dan decided not to include regular duty batteries in the experiment.

Other possible sources of variation include the date of manufacture of the purchased battery, and whether the lifetime was monitored under continuous running conditions or under the more usual setting with the flashlight being turned on and off, the temperature of the environment, the age and variability of the flashlight bulbs.

The first of these could not be controlled in the experiment. The batteries used in the experiment were purchased at different times and in different locations in order to give a wide representation of dates of manufacture. The variability caused by this factor would be measured as part of the natural variability (error variability) in the experiment along with measurement error. Had the dates been marked on the packets, they could have been included in the analysis of the experiment as covariates. However, the dates were not available.

The second of these possible sources of variation (running conditions) was fixed. All the measurements were to be made under constant running conditions. Although this did not mimic the usual operating conditions of flashlight batteries, Dan thought that the relative ordering of the different battery types in terms of life per unit cost would be the same. The continuous running setting was much easier to handle in an experiment since each observation was expected to take several hours and no sophisticated equipment was available.

The third source of variation (temperature) was also fixed. Since the family living quarters are kept at a temperature of about 68° in the winter, Dan decided to run his experiment at this usual temperature. Small fluctuations in temperature were not expected to be important.

The variability due to the age of the flashlight bulb was more difficult to handle. A decision had to be made whether to use a new bulb for each observation and risk muddling the effect of the battery with that of the bulb, or whether to use the same bulb throughout the experiment and risk an effect of the bulb age from biasing the data. A third possibility was to divide the observations into blocks and to use a single bulb throughout a block, but to change bulbs between blocks. Since the lifetime of a bulb is considerably longer than that of a battery, Dan decided to use the same bulb throughout the experiment.

#### (i) Treatment factors and their levels

There are two treatment factors each having two levels. These are battery “duty” (level 1 = alkaline, level 2 = heavy duty) and “brand” (level 1 = name brand, level 2 = store brand). This gives four treatment combinations coded 11, 12, 21, 22. In Chaps. 3–5, we will recode these treatment combinations as 1, 2, 3, 4, and we will often refer to them as the four different treatments or the four different levels of the factor “battery type.” Thus, the levels of battery type are:

Level	Treatment Combination
1	alkaline, name brand (11)
2	alkaline, store brand (12)
3	heavy duty, name brand (21)
4	heavy duty, store brand (22)

#### (ii) Experimental units

The experimental units in this experiment are the time slots. These were assigned at random to the battery types so as to determine the order in which the batteries were to be observed. Any fluctuations in temperature during the experiment form part of the variability between the time slots and are included in the error variability.

#### (iii) Blocking factors, noise factors, and covariates

As mentioned above, it was decided not to include a blocking factor representing different flashlight bulbs. Also, the date of manufacture of each battery was not available, and small fluctuations in room temperature were not thought to be important. Consequently, there were no covariates in the experiment, and no noise factors were incorporated.

(c) **Choose a rule by which to assign the experimental units to the levels of the treatment factor.**

Since there were to be no blocking factors, a completely randomized design was selected, and the time slots were assigned at random to the four different battery types.

(d) **Specify the measurements to be made, the experimental procedure, and the anticipated difficulties.**

The first difficulty was in deciding exactly how to measure lifetime of a flashlight battery. First, a flashlight requires two batteries. In order to keep the cost of the experiment low, Dan decided to wire a circuit linking just one battery to a flashlight bulb. Although this does not mimic the actual use of a flashlight, Dan thought that as with the constant running conditions, the relative lifetimes per unit cost of the four battery types would be preserved. Secondly, there was the difficulty in determining when the battery had run down. Each observation took several hours, and it was not possible to monitor the experiment constantly. Also, a bulb dims slowly as the battery runs down, and it is a judgment call as to when the battery is flat. Dan decided to deal with both of these problems by including a small clock in the circuit. The clock stopped before the bulb had completely dimmed, and the elapsed time on the clock was taken as a measurement of the battery life. The cost of a battery was computed as half of the cost of a two-pack, and the lifetime per unit cost was measured in minutes per dollar (min/\$).

(e) **Run a pilot experiment.**

A few observations were run as a pilot experiment. This ensured that the circuit did indeed work properly. It was discovered that the clock and the bulb had to be wired in parallel and not in series, as Dan had first thought! The pilot experiment also gave a rough idea of the length of time each observation would take (at least four hours), and provided a very rough estimate of the error variability that was used at step (h) to calculate that four observations were needed on each treatment combination.

### **Difficulties Encountered**

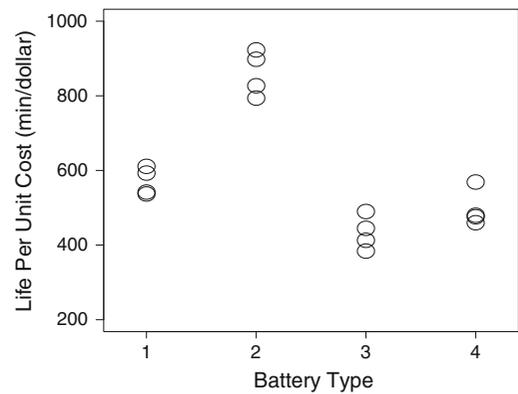
The only difficulty encountered in running the main experiment was that during the fourth observation, it was discovered that the clock was running but the bulb was out. This was due to a loose connection. The connection was repaired, a new battery inserted into the circuit, and the clock reset.

### **Data**

The data collected in the main experiment are shown in Table 2.8 and plotted in Fig. 2.3. The experiment was run in 1993.

**Table 2.8** Data for the battery experiment

Battery type	Life (min)	Unit cost (\$)	Life per unit cost	Time order
1	602	0.985	611	1
2	863	0.935	923	2
1	529	0.985	537	3
4	235	0.495	476	4
1	534	0.985	542	5
1	585	0.985	593	6
2	743	0.935	794	7
3	232	0.520	445	8
4	282	0.495	569	9
2	773	0.935	827	10
2	840	0.935	898	11
3	255	0.520	490	12
4	238	0.495	480	13
3	200	0.520	384	14
4	228	0.495	460	15
3	215	0.520	413	16

**Fig. 2.3** Battery life per unit cost versus battery

### 2.5.3 Cake-Baking Experiment

The following factorial experiment was run in 1979 by the baking company Spillers Ltd. (in the U.K.) and was reported in the *Bulletin in Applied Statistics* in 1980 by S.M. Lewis and A.M. Dean.

**Checklist**

**(a) Define the objectives of the experiment.**

The experimenters at Spillers, Ltd. wanted to know how “cake quality” was affected by adding different amounts of glycerine and tartaric acid to the cake mix.

**(b) Identify all sources of variation.**

(i) Treatment factors and their levels

The two treatment factors of interest were glycerine and tartaric acid. Glycerine was called the “first treatment factor” and labeled  $F_1$ , while tartaric acid was called the “second treatment factor” and labeled  $F_2$ . The experimenters were very familiar with the problems of cake baking and determinations of cake quality. They knew exactly which amounts of the two treatment factors they wanted to compare. They selected four equally spaced amounts of glycerine and three equally spaced amounts of tartaric acid. These were coded as 1, 2, 3, 4 for glycerine and 1, 2, 3 for tartaric acid. Therefore, the twelve coded treatment combinations were 11, 12, 13, 21, 22, 23, 31, 32, 33, 41, 42, 43.

(ii) Identify the experimental units

Before the experimental units can be identified, it is necessary to think about the experimental procedure. One batch of cake-mix was divided into portions. One of the twelve treatment combinations (i.e., a certain amount of glycerine and a certain amount of tartaric acid) was added to each portion. Each portion was then thoroughly mixed and put into a container for baking. The containers were placed on a tray in an oven at a given temperature for the required length of time. The experimenters required an entire tray of cakes to make one measurement of cake quality. Only one tray would fit on any one shelf of an oven. An experimental unit was, therefore, “an oven shelf with a tray of containers of cake-mix,” and these were assigned at random to the twelve treatment combinations.

(iii) Blocking factors, noise factors, and covariates

There were two crossed blocking factors. The first was time of day with two levels (morning and afternoon). The second was oven, which had three levels, one level for each of the three ovens that were available on the day of the experiment. Each cell (defined by oven and time of day) contained six experimental units, since an oven contained six shelves (see Table 2.9). Each set of six experimental units was assigned at random to six of the twelve treatment combinations, and it

**Table 2.9** Basic design for the baking experiment

Oven codes	Time of day codes											
	1						2					
1	11	13	22	24	32	34	12	14	21	23	31	33
2	12	14	21	23	32	34	11	13	22	24	31	33
3	12	14	22	24	31	33	11	13	21	23	32	34

was decided in advance which six treatment combinations should be observed together in a cell (see step (c) of the checklist).

Although the experimenters expected differences in the ovens and in different runs of the same oven, their experience showed that differences between the shelves of their industrial ovens were very minor. Otherwise, a third blocking factor representing oven shelf would have been needed.

It was possible to control carefully the amount of cake mix put into each container, and the experimenters did not think it was necessary to monitor the precooked weight of each cake. Small differences in these weights would not affect the measurement of the quality. Therefore, no covariates were used in the analysis.

**(c) Choose a rule by which to assign the experimental units to the levels of the treatment factors.**

Since there were two crossed blocking factors, a row–column design with six experimental units per cell was required. It was not possible to observe every treatment combination in every cell. However, it was thought advisable to observe all twelve treatment combinations in each oven, either in the morning or the afternoon. This precaution was taken so that if one of the ovens failed on the day of the experiment, the treatment combinations could still all be observed twice each. The basic design (before randomization) that was used by Spillers is shown in Table 2.9. The experimental units (the trays of containers on the six oven shelves) need to be assigned at random to the 6 treatment combinations cell by cell. The oven codes need to be assigned to the actual ovens at random, and the time of day codes 1 and 2 to morning and afternoon.

---

## Exercises

**Exercises 1–7 refer to the list of experiments in Table 2.10.**

1. Table 2.10 gives a list of experiments that can be run as class projects. Select a simple experiment of interest to you, but preferably not on the list. Complete steps (a)–(c) of the checklist with the intention of actually running the experiment when the checklist is complete.
2. For experiments 1 and 7 in Table 2.10, complete steps (a) and (b) of the checklist. There may be more than one treatment factor. Give precise definitions of their levels.
3. For experiment 2, complete steps (a)–(c) of the checklist.
4. For experiment 3, complete steps (a)–(c) of the checklist.
5. For experiment 4, list sources of variation. Decide which sources can be controlled by limiting the scope of the experiment or by specifying the exact experimental procedure to be followed. Of the

**Table 2.10** Some simple experiments

- 
- |    |   |
|----|---|
| 1. | Compare the growth rate of bean seeds under different watering and lighting schedules.  |
| 2. | Does the boiling point of water differ with different concentrations of salt?   |
| 3. | Compare the strengths of different brands of paper towel.   |
| 4. | Do different makes of popcorn give different proportions of unpopped kernels? What about cooking methods?   |
| 5. | Compare the effects of different locations of an observer on the speed at which subjects locate the occurrences of the letter “e” in a written passage. |
| 6. | Do different colored candles burn at different speeds?  |
| 7. | Compare the proportions of words remembered from lists of related or unrelated words, and under various conditions such as silence and distraction.     |
| 8. | Compare the effects of different colors of exam paper on students’ performance in an examination.   |
-

remaining sources of variation, decide which are minor and which are major. Are there any blocking factors in this experiment?

6. For experiment 6, specify what measurements should be made, how they should be made, and list any difficulties that might be expected.
7. For experiment 8, write down all the possible sources of variation. In your opinion, should this experiment be run as a completely randomized design, a block design, or a design with more than one blocking factor? Justify your answer.
8. Read critically through the checklists in Sect. 2.5. Would you suggest any changes? Would you have done anything differently? If you had to criticize these experiments, which points would you address?
9. The following description was given by Clifford Pugh in the 1953 volume of *Applied Statistics*.

“The widespread use of detergents for domestic dish washing makes it desirable for manufacturers to carry out tests to evaluate the performance of their products. . . . Since foaming is regarded as the main criterion of performance, the measure adopted is the number of plates washed before the foam is reduced to a thin surface layer. The five main factors which may affect the number of plates washed by a given product are (i) the concentration of detergent, (ii) the temperature of the water, (iii) the hardness of the water, (iv) the type of “soil” on the plates, and (v) the method of washing used by the operator. . . . The difficulty of standardizing the soil is overcome by using the plates from a works canteen (cafeteria) for the test and adopting a randomized complete block technique in which plates from any one course form a block . . . . One practical limitation is the number of plates available in any one block. This permits only four . . . tests to be completed (in a block).”

Draw up steps (a)–(d) of a checklist for an experiment of the above type and give an example of a design that fits the requirements of your checklist.