Principles and Techniques

<div style="text-align:right">**1**</div>

1.1 Design: Basic Principles and Techniques

1.1.1 The Art of Experimentation

One of the first questions facing an experimenter is, "How many observations do I need to take?" or alternatively, "Given my limited budget, how can I gain as much information as possible?" These are not questions that can be answered in a couple of sentences. They are, however, questions that are central to the material in this book. As a first step towards obtaining an answer, the experimenter must ask further questions, such as, "What is the main purpose of running this experiment?" and "What do I hope to be able to show?"

Typically, an experiment may be run for one or more of the following reasons:

  (i) to determine the principal causes of variation in a measured response,
 (ii) to find the conditions that give rise to a maximum or minimum response,
(iii) to compare the responses achieved at different settings of controllable variables,
(iv) to obtain a mathematical model in order to predict future responses.

Observations can be collected from *observational studies* as well as from *experiments*, but only an experiment allows conclusions to be drawn about cause and effect. For example, consider the following situation:

The output from each machine on a factory floor is constantly monitored by any successful manufacturing company. Suppose that in a particular factory, the output from a particular machine is consistently of low quality. What should the managers do? They could conclude that the machine needs replacing and pay out a large sum of money for a new one. They could decide that the machine operator is at fault and dismiss him or her. They could conclude that the humidity in that part of the factory is too high and install a new air conditioning system. In other words, the machine output has been observed under the current operating conditions (an observational study), and although it has been very effective in showing the management that a problem exists, it has given them very little idea about the *cause* of the poor quality.

It would actually be a simple matter to determine or rule out some of the potential causes. For example, the question about the operator could be answered by moving all the operators from machine to machine over several days. If the poor output follows the operator, then it is safe to conclude that the operator is the cause. If the poor output remains with the original machine, then the operator is

blameless, and the machine itself or the factory humidity is the most likely cause of the poor quality. This is an "experiment." The experimenter has control over a possible cause in the difference in output quality between machines. If this particular cause is ruled out, then the experimenter can begin to vary other factors such as humidity or machine settings.

It is more efficient to examine all possible causes of variation simultaneously rather than one at a time. Fewer observations are usually needed, and one gains more information about the system being investigated. This simultaneous study is known as a "factorial experiment." In the early stages of a project, a list of all factors that conceivably could have an important effect on the response of interest is drawn up. This may yield a large number of factors to be studied, in which case special techniques are needed to gain as much information as possible from examining only a subset of possible factor settings.

The art of designing an experiment and the art of analyzing an experiment are closely intertwined and need to be studied side by side. In designing an experiment, one must take into account the analysis that will be performed. In turn, the efficiency of the analysis will depend upon the particular experimental design that is used to collect the data. Without these considerations, it is possible to invest much time, effort, and expense in the collection of data which seem relevant to the purpose at hand but which, in fact, contribute little to the research questions being asked. A guiding principle of experimental design is to "keep it simple." Interpretation and presentation of the results of experiments are generally clearer for simpler experiments.

Three basic techniques fundamental to experimental design are replication, blocking, and randomization. The first two help to increase precision in the experiment; the last is used to decrease bias. These techniques are discussed briefly below and in more detail throughout the book.

1.1.2 Replication

Replication is the repetition of experimental conditions so that the effects of interest can be estimated with greater precision and the associated variability can be estimated.

There is a difference between "replication" and "repeated measurements." For example, suppose four subjects are each assigned to a drug and a measurement is taken on each subject. The result is four independent observations on the drug. This is "replication." On the other hand, if one subject is assigned to a drug and then measured four times, the measurements are not independent. We call them "repeated measurements." The variation recorded in repeated measurements taken at the same time reflects the variation in the measurement process, while the variation recorded in repeated measurements taken over a time interval reflects the variation in the single subject's response to the drug over time. Neither reflects the variation in independent subjects' responses to the drug. We need to know about the latter variation in order to generalize any conclusion about the drug so that it is relevant to all similar subjects.

1.1.3 Blocking

A designed experiment involves the application of treatments to experimental units to assess the effects of the treatments on some response. The "experimental units," which may be subjects, materials, conditions, points in time, or some combination of these, will be variable and induce variation in the response. Such variation in experimental units may be intentional, as the experimental conditions under which an experiment is run should be representative of those to which the conclusions of the experiment are to be applied. For inferences to be broad in scope, the experimental conditions should

be appropriately varied. Blocking is a technique that can often be used to control and adjust for some of the variation in experimental units.

To block an experiment is to divide, or partition, the experimental units into groups called blocks in such a way that the experimental units in each block are intended to be relatively similar, so that treatments assigned to experimental units in the same block can be compared under relatively similar experimental conditions. If blocking is done well, then comparisons of two or more treatments are made more precisely in the experiment than similar comparisons from an unblocked design. For example, in an experiment to compare the effects of two skin ointments for rash, the two treatments can be compared more precisely on two arms of the same person than on the arms of two different people. Either circumstance can be replicated, ideally using subjects randomly sampled from or at least representative of the population of interest.

1.1.4   Randomization

The purpose of randomization is to prevent systematic and personal biases from being introduced into the experiment by the experimenter. A random assignment of subjects or experimental material to treatments prior to the start of the experiment ensures that observations that are favored or adversely affected by unknown sources of variation are observations "selected in the luck of the draw" and not systematically selected.

Lack of a random assignment of experimental material or subjects leaves the experimental procedure open to *experimenter bias*. For example, a horticulturist may assign his or her favorite variety of experimental crop to the parts of the field that look the most fertile, or a medical practitioner may assign his or her preferred drug to the patients most likely to respond well. The preferred variety or drug may then appear to give better results no matter how good or bad it actually is.

Lack of random assignment can also leave the procedure open to *systematic bias*. Consider, for example, an experiment involving drying time of three paints applied to sections of a wooden board, where each paint is to be observed four times. If no random assignment of order of observation is made, many experimenters would take the four observations on paint 1, followed by those on paint 2, followed by those on paint 3. This order might be perfectly satisfactory, but it could equally well prove to be disastrous. Observations taken over time could be affected by differences in atmospheric conditions, fatigue of the experimenter, systematic differences in the wooden board sections, etc. These could all conspire to ensure that any measurements taken during the last part of the experiment are, say, underrecorded, with the result that paint 3 appears to dry faster than the other paints when, in fact, it may be less good. The order 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3 helps to solve the problem, but it does not remove it completely (especially if the experimenter takes a break after every three observations).

There are also analytical reasons to support the use of a random assignment. It will be seen in Chaps. 3 and 4 that common forms of analysis of the data depend on the F and t distributions. It can be shown that a random assignment ensures that these distributions are the correct ones to use. The interested reader is referred to Kempthorne (1977).

To understand the meaning of randomization, consider an experiment to compare the effects on blood pressure of three exercise programs, where each program is observed four times, giving a total of 12 observations. Now, given 12 subjects, imagine making a list of all possible assignments of the 12 subjects to the three exercise programs so that 4 subjects are assigned to each program. (There are 12!/(4!4!4!), or 34,650 ways to do this.) If the assignment of subjects to programs is done in such a way that every possible assignment has the same chance of occurring, then the assignment is said to be a *completely random assignment*. Completely randomized designs, discussed in Chaps. 3–7 of this book, are randomized in this way. It is, of course, possible that a random assignment itself could

lead to the order 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3. If the experimenter expressly wishes to avoid certain assignments, then a different type of design should be used. An experimenter should not look at the resulting assignment, decide that it does not look very random, and change it.

Without the aid of an objective randomizing device, it is not possible for an experimenter to make a random assignment. In fact, it is not even possible to select a single number at random. This is borne out by a study run at the University of Delaware and reported by Professor Hoerl in the *Royal Statistical Society News and Notes* (January 1988). The study, which was run over several years, asked students to pick a number at random between 0 and 9. The numbers 3 and 7 were selected by about 40% of the students. This is twice as many as would be expected if the numbers were truly selected at random.

The most frequently used objective mechanism for achieving a random assignment in experimental design is a random number generator. A random number generator is a computer program that gives as output a very long string of digits that are integers between 0 and 9 inclusive and that have the following properties. All integers between 0 and 9 occur approximately the same number of times, as do all pairs of integers, all triples, and so on. Furthermore, there is no discernible pattern in the string of digits, and hence the name "random" numbers.

The random numbers in Appendix Table A.1 are part of a string of digits produced by a random number generator (in SAS$^{®}$ version 6.09 on a DEC Model 4000 MODEL 610 computer at Wright State University). Many experimenters and statistical consultants will have direct access to their own random number generator on a computer or calculator and will not need to use the table. The table is divided into six sections (pages), each section containing six groups of six rows and six groups of six columns. The grouping is merely a device to aid in reading the table. To use the table, a random starting place must be found. An experimenter who always starts reading the table at the same place always has the same set of digits, and these could not be regarded as random. The grouping of the digits by six rows and columns allows a random starting place to be obtained using five rolls of a fair die. For example, the five rolls giving 3, 1, 3, 5, 2 tells the experimenter to find the digit that is in Sect. 3 of the table, row group 1, column group 3, row 5, column 2. Then the digits can be read singly, or in pairs, or triples, etc. from the starting point across the rows.

The most common random number generators on computers or calculators generate n-digit real numbers between zero and one. Single digit random numbers can be obtained from an n-digit real number by reading the first digit after the decimal point. Pairs of digits can be obtained by reading the first two digits after the decimal point, and so on. The use of random numbers for randomization is shown in Sects. 3.2, 3.8.1 and 3.9.1.

1.2   Analysis: Basic Principles and Techniques

In the analysis of data, it is desirable to provide both graphical and statistical analyses. Plots that illustrate the relative responses of the factor settings under study allow the experimenter to gain a feel for the practical implications of the statistical results and to communicate effectively the results of the experiment to others. In addition, data plots allow the proposed model to be checked and aid in the identification of unusual observations, as discussed in Chap. 5. Statistical analysis quantifies the relative responses of the factors, thus clarifying conclusions that might be misleading or not at all apparent in plots of the data.

The purpose of an experiment can range from exploratory (discovering new important sources of variability) to confirmatory (confirming that previously discovered sources of variability are sufficiently major to warrant further study), and the philosophy of the analysis depends on the purpose of the experiment. In the early stages of experimentation the analysis may be exploratory, and one would plot and analyze the data in any way that assists in the identification of important sources of variation. In

later stages of experimentation, analysis is usually confirmatory in nature. A mathematical model of the response is postulated and hypotheses are tested and confidence intervals are calculated.

In this book, we use *linear models* to model our response and the *method of least squares* for obtaining estimates of the parameters in the model. These are described in Chap. 3. We also use restricted maximum likelihood estimation of parameters in Chap. 19. Our models include random "error variables" that encompass all the sources of variability not explicitly present in the model. We operate under the assumption that the error terms are normally distributed. However, most of the procedures in this book are generally fairly robust to nonnormality, provided that there are no extreme observations among the data.

It is rare nowadays for experimental data to be analyzed by hand. Most experimenters and statisticians have access to a computer package that is capable of producing, at the very least, a basic analysis of data for the simplest experiments. To the extent possible, for each design discussed, we shall present useful plots and methods of analysis that can be obtained from most statistical software packages. We will also develop many of the mathematical formulas that lie behind the computer analysis. This will enable the reader more easily to appreciate and interpret statistical computer package output and the associated manuals. Computer packages vary in sophistication, flexibility, and the statistical knowledge required of the user. The SAS software (see SAS Institute Inc., 2004) is one of the better commercial statistical packages for analyzing experimental data. The R software (see R Core Team, 2017) is a command line software for statistical computing and graphics which is freely available on the web. Both packages can handle every model discussed in this book, and although they require some knowledge of experimental design on the part of the user, neither is difficult to learn. We provide some basic SAS and R statements and resulting output at the end of most chapters to illustrate data analysis. A reader who wishes to use a different computer package can run the equivalent analyses on his or her own package and compare the output with those shown. It is important that every user know exactly the capabilities of his or her own package and also the likely size of rounding errors.

It is not our intent to teach the best use of SAS and R software, and readers may find better ways of achieving the same analyses. SAS software, being a commercial package, requires purchase of a license, but many universities and companies already have site licenses. R is a free software environment for statistical computing and graphics. Links for downloading and installing R are provided in Sect. 3.9.