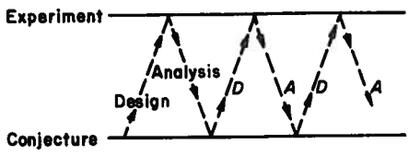

FIGURE 1.3. The iterative nature of experimentation.

refer only to *statistical* design and analysis. By "design," we mean the synthesis of a suitable experiment to test, estimate, and develop a current conjectured model. By "analysis" we mean the treatment of the experimental results, leading either to verification of a postulated model and the working out of its consequences, or to the forming of a new or modified conjecture. These complementary processes are used many times during an investigation and, by their intelligent alternation, the experimenter's knowledge of the system becomes steadily greater.

The indeterminancy of the learning process, but the ultimate possibility of its convergence, is familiarly exemplified by the playing of the game of 20 questions. Ten players engaged independently in playing 20 questions all might succeed eventually in discovering that the object in question was the left ear of the Statue of Liberty but they would almost certainly reach that conclusion by different routes. Let us consider how they do it.

A player may be supplied initially with the information that the object is "mineral." This limits consideration to a certain class of objects which he has to reduce to one as rapidly as possible. When the game is played so that each question concerns a choice between two specific alternatives, the best design of experiment (each "experiment" here is a question asked) is one which will classify the possible objects into two approximately equi-probable groups. A good (but not uniquely good) experiment would thus consist of asking "Is it metal or stone?;" a poor experiment would consist of asking "Is it the stone in my ring?." The experiment having been performed and the data (in the form of an answer) having become available, the player would now analyze the reply in the light of any relevant prior knowledge he possessed and would then form some new or modified conjecture as to the nature of the object in question. To the question, "Is it metal or stone?" the answer "stone" might conjure up in his mind, on the one hand, buildings, monuments, mountains, and so on, and, on the other hand, small stones, both precious (such as diamonds, rubies) and nonprecious (such as pebbles). The player might feel that a good way to discriminate between these two types of objects in the light of information then available would be the question "Is it larger or smaller than this room?." The answer to this question would now result in new analysis which would raise new conjectures and give rise to a new question and so the game would continue.

As another example of iterative investigation, consider a detective employed in solving a mystery. Some data in the form of a body, and the facts that the door was locked on the inside and that Mr. X would benefit from the will, are already known, or have been discovered. These data lead to conjectures on the part of the investigator which in turn lead to certain data-producing actions, or "experiments." As readers of detective novels will realize, the sequence of events which finally leads to the detection of the culprit is by no means unique. Alternative good (but by no means unique) experiments might be a visit to Baker Street subway station to question the ticket collector about a one-armed man with a scar under his left eye, or a visit to the Manor House to find out if the flower bed beneath Lady Cynthia's window shows a tell-tale footprint. The skill of the detective is partially measured by his ability to conceive, at each stage, appropriate conjectures and to select those experiments which, in the light of these conjectures, will best illuminate current aspects of the case. He pursues a strategy which (he hopes) will cause him to follow *one* of the paths leading to the unmasking of the assassin (preferably before the latter has had time to flee the country).

It is through the iterative process of "learning as we go" that the problem of indeterminancy which we mention may be resolved. We shall find that, in the application of the investigative processes described in this book, a multidimensional iteration occurs in which modification in the *location* of the experimental runs, in *scaling*, in *transformation*, and in the *complexity* of the contemplated model, all can occur. While we cannot, and should not, attempt to ensure uniqueness for the investigation, we can ensure that the crucial components of good design and analysis on which convergence depends are organized so as to illuminate and stimulate the ideas of the experimenter as effectively as possible, and so lead him quickly along one (nonunique) path to the truth. Figure 1.4 illustrates this general idea.

The truly iterative nature of some investigations may sometimes be obscured by the length of time taken by each iterative cycle. In this case it may be possible to see the wider iteration only by "stepping back" and examining what occurs over months or years. In this wider context, iteration may skip from one investigator to another, even from one country to another, and its phases may be very long. Even in this situation, however, it is important to bear in mind that the important consideration is the *overall* acquisition of knowledge.

What is being said is not, of course, peculiar to response surface methodology. "Traditional" designs such as randomized blocks, latin squares, and factorial designs have, since their inception, been used by statisticians as building blocks in iterative learning sequences. The possibility of rapid convergence of such a sequence depends, to an important

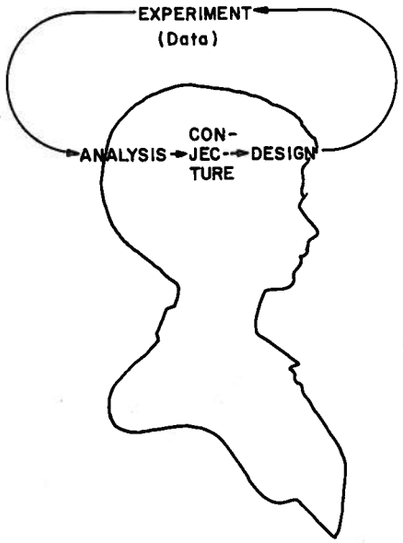

FIGURE 1.4. The iterative experimental process in relation to the experimenter.

extent, on the efficiency of these designs and their associated analyses, and their ability to illuminate and stimulate the ideas of the investigator. This notion was inherent in R. A. Fisher's* general attitude toward the use of these designs in scientific investigation.

1.4. SOME CLASSES OF PROBLEMS (WHICH, HOW, WHY)

Within the foregoing iterative context, we now consider some specific classes of scientific problems and the characteristics of models used in solving these problems.

Mechanistic and Empirical Models

It is helpful in this discussion to distinguish between empirical models and mechanistic models. We consider first what we might mean by a purely mechanistic model. Suppose that, in the study of some physical phenomenon, we know enough of its physical mechanism to *deduce* the form of the

*Sir Ronald Fisher (1890–1962), the famous British statistician, was responsible for many of the basic ideas of experimental design. His books *Statistical Methods for Research Workers* and *The Design of Experiments* are classic.

functional relationship linking the mean value η of the output to the levels ξ of the inputs via an expression

$$E(y) = \eta = f(\xi, \theta), \tag{1.4.1}$$

where $\xi = (\xi_1, \xi_2, \ldots, \xi_k)'$ is a set of input variables, measuring, for example, initial concentrations of reactants, temperatures, and pressures, and where $\theta = (\theta_1, \theta_2, \ldots, \theta_p)'$ represents a set of physical parameters measuring such things as activation energies, diffusion coefficients, and thermal conductivities. Then we should say that Eq. (1.4.1) represented a mechanistic model.

In practice some or all of the θ's would need to be estimated from the data. Also, mechanistic knowledge might most naturally be expressed by a set of differential equations or integral equations for which (1.4.1) was the solution.

Often, however, the necessary physical knowledge of the system is absent or incomplete and consequently no mechanistic model is available. In these circumstances, it could often be realistically assumed that the relationship between η and ξ would be smooth and, consequently, that $f(\xi, \theta)$ could be locally *approximated* (over limited ranges of the experimental variables ξ) by an interpolation function $g(\xi, \beta)$, such as a polynomial. In this latter expression, the β's, which are the elements of β, would be coefficients in the interpolation function. They would be related to, but be distinct from, the parameters (the θ's) of the physical system. The interpolation function $g(\xi, \beta)$ could provide a local empirical model for the system and, as we have said, would act simply as a mathematical French curve.

Now the theoretical mechanistic model $\eta = f(\xi, \theta)$, and the purely empirical model $\eta = g(\xi, \beta)$, as defined above, represent extremes. The former would be appropriate in the extreme case where a great deal was accurately known about the system, and the latter would be appropriate in the other extreme case, where nothing could be assumed except that the response surface was locally smooth. The situation existing in most real investigations is somewhere in between and, as experimentation proceeds, and we gain information, the situation can change. Because real problems occur at almost all points between the extremes mentioned above, a variety of statistical tools is needed to cope with them.

Both the state of ignorance in which we begin our experimental work, and the state of comparative knowledge to which we wish to be brought, will determine our approach. It must be realized, of course, that no real problem ever *quite* fits any prearranged category. With this proviso in mind, it is nevertheless helpful to distinguish the basic *types* of problems shown in Table 1.1. These are categorized in terms of what is unknown about the

extract the maximum amount of caffeine η when hot water is added to a particular kind of tea leaves. Suppose, further, that we have a pressure cooker at our disposal. The operability region $O(\xi)$ would define the temperatures at which the cooker could be operated; these might extend from 30 to 120°C (from room temperature to the temperature attainable when maximum pressure was applied). However, the region of interest $R(\xi)$ typically would be much more limited; initially, at least, $R(\xi)$ would likely be in the immediate vicinity of the boiling point of water—from 95 to 100°C, say. Depending on what the experiments conducted in this smaller region show, the region of interest might then be changed to cover some other more promising region $R'(\xi)$ within $O(\xi)$, and so on.

It should be carefully noted in this formulation that, in order best to explore some region $R(\xi)$, of current interest, we do not need to perform all the experimental runs within $R(\xi)$. Some runs might be inside and some outside, or all might possibly be outside.

Now, writing $f(\xi)$ for the polynomial approximation, we have

$$E(y) = \eta(\xi) \tag{13.1.2}$$

for all ξ, and

$$\eta(\xi) \doteq f(\xi) \tag{13.1.3}$$

over some limited region of interest $R(\xi)$. The fact that the polynomial is an approximation does not necessarily detract from its usefulness because all models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind. For, if $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ is a vector of random errors having zero vector mean, and if $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ and $\mathbf{f}(\boldsymbol{\xi}) = \{f(\xi_1), f(\xi_2), \ldots, f(\xi_n)\}$ where $\xi_1, \xi_2, \ldots, \xi_n$ are n observations on ξ, the true model is not

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\xi}) + \boldsymbol{\varepsilon} \tag{13.1.4}$$

but

$$\mathbf{y} = \boldsymbol{\eta}(\boldsymbol{\xi}) + \boldsymbol{\varepsilon}, \tag{13.1.5}$$

that is

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\xi}) + \boldsymbol{\delta}(\boldsymbol{\xi}) + \boldsymbol{\varepsilon}, \tag{13.1.6}$$

where

$$\boldsymbol{\delta}(\boldsymbol{\xi}) = \boldsymbol{\eta}(\boldsymbol{\xi}) - \mathbf{f}(\boldsymbol{\xi}) \tag{13.1.7}$$

is the vector discrepancy, which we would like to be small over $R(\xi)$, between actual and approximate models. In other words, there are *two* types of errors which must be taken into account:

1. Systematic, or bias, errors $\boldsymbol{\delta}(\boldsymbol{\xi}) = \boldsymbol{\eta}(\boldsymbol{\xi}) - \mathbf{f}(\boldsymbol{\xi})$, the difference between the expected value of the response, $E(\mathbf{y}) = \boldsymbol{\eta}(\boldsymbol{\xi})$ and the approximating function $\mathbf{f}(\boldsymbol{\xi})$.
2. Random errors $\boldsymbol{\varepsilon}$.

Although the above implies that systematic errors $\boldsymbol{\delta}(\boldsymbol{\xi})$ are always to be expected, there has been, since the time of Gauss (1777–1855), an unfortunate tendency to ignore them, and to concentrate only on the random errors $\boldsymbol{\varepsilon}$. This seems to have been because nice mathematical results are possible when this is done. In choosing an experimental design, the ignoring of systematic error is *not* an innocuous approximation and misleading results may well be obtained, as we shall see in due course.

13.2. THE COMPETING EFFECTS OF BIAS AND VARIANCE

Consider first the elementary case where there is only one important predictor variable ξ. Figure 13.1 represents a typical situation in which the

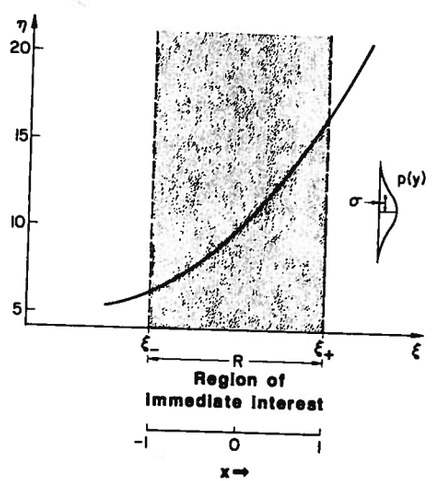

FIGURE 13.1. A hypothetical unknown relationship existing between $E(y) = \eta$ and ξ. At the right is shown the error distribution $p(y)$ which has standard deviation σ.

TABLE 12A.2. Comparison of extra sums of squares taken up by various models

Model	Source	Extra SS	df	MS
Second degree	Model	3004.2	9	333.8
	Residual	67.8	14	4.8
Canonical form	Model	2965.4	6	494.2
	Residual	106.6	17	6.3
Mechanism	Model	2986.7	5	597.3
	Residual	85.3	18	4.7

Comparison of Various Models

Some measure of the comparative representational ability of the various models we have fitted to the conversion data can be had by comparing the sums of squares explained by the models, and their corresponding degrees of freedom. It is convenient to examine the *extra* sums of squares beyond that accounted for by fitting only a mean, namely, $y = \beta_0 + \varepsilon$. (It must be remembered that the mechanistic model does not have a β_0 term in it, so that the "extra" sum of squares in this case is not one that can be attributed to specific "extra" parameters.) These are shown in Table 12A.2, the degrees of freedom in each case being the total number of parameters minus one for b_0. The mechanistic model appears to possess remarkably good representational properties, for comparatively few parameters.

CHAPTER 13

Design Aspects of Variance, Bias, and Lack of Fit

13.1. THE USE OF APPROXIMATING FUNCTIONS

Typically, in a scientific investigation, some response y is measured whose mean value $E(y) = \eta$ is believed to depend on a set of variables $\xi = (\xi_1, \xi_2, \ldots, \xi_k)'$. The exact functional relationship between them,

$$E(y) = \eta = \eta(\xi) \tag{13.1.1}$$

is usually unknown and possibly unknowable. We often represent the function $\eta(\xi)$ as the solution to some set of time- and space-dependent ordinary or partial differential equations. However, we have only to think of the flight of a bird, the fall of a leaf, or the flow of water through a valve, to realize that, even using such equations, we are likely to be able to approximate only the main features of such a relationship. In this book, we employ even cruder approximations using polynomials that exploit local smoothness properties. Approximations of this type are, however, often perfectly adequate locally. Over a short distance, the flight of the bird might be approximated by a straight line function of time; over somewhat longer distances, a quadratic function might be used. Over narrow ranges, the flow of water through the valve might be similarly approximated by a straight line function of the valve opening.

Mathematically, this idea may be expressed by saying that, over a limited range of interest $R(\xi)$, the main characteristics of a smooth function may be represented by the low-order terms of a Taylor series approximation. It should be understood that the *region of interest* $R(\xi)$ will lie within a larger *region of operability* $O(\xi)$, defined as the region over which experiments *could* be conducted if desired, using the available apparatus. For example, suppose we are anxious to discover the water temperature ξ which will

extract the maximum amount of caffeine η when hot water is added to a particular kind of tea leaves. Suppose, further, that we have a pressure cooker at our disposal. The operability region $O(\xi)$ would define the temperatures at which the cooker could be operated; these might extend from 30 to 120°C (from room temperature to the temperature attainable when maximum pressure was applied). However, the region of interest $R(\xi)$ typically would be much more limited; initially, at least, $R(\xi)$ would likely be in the immediate vicinity of the boiling point of water—from 95 to 100°C, say. Depending on what the experiments conducted in this smaller region show, the region of interest might then be changed to cover some other more promising region $R'(\xi)$ within $O(\xi)$, and so on.

It should be carefully noted in this formulation that, in order best to explore some region $R(\xi)$, of current interest, we do not need to perform all the experimental runs within $R(\xi)$. Some runs might be inside and some outside, or all might possibly be outside.

Now, writing $f(\xi)$ for the polynomial approximation, we have

$$E(y) = \eta(\xi) \tag{13.1.2}$$

for all ξ, and

$$\eta(\xi) \doteq f(\xi) \tag{13.1.3}$$

over some limited region of interest $R(\xi)$: The fact that the polynomial is an approximation does not necessarily detract from its usefulness because all models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind. For, if $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ is a vector of random errors having zero vector mean, and if $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ and $\mathbf{f}(\xi) = \{f(\xi_1), f(\xi_2), \ldots, f(\xi_n)\}$ where $\xi_1, \xi_2, \ldots, \xi_n$ are n observations on ξ, the true model is not

$$\mathbf{y} = \mathbf{f}(\xi) + \boldsymbol{\varepsilon} \tag{13.1.4}$$

but

$$\mathbf{y} = \boldsymbol{\eta}(\xi) + \boldsymbol{\varepsilon}, \tag{13.1.5}$$

that is

$$\mathbf{y} = \mathbf{f}(\xi) + \boldsymbol{\delta}(\xi) + \boldsymbol{\varepsilon}, \tag{13.1.6}$$

where

$$\boldsymbol{\delta}(\xi) = \boldsymbol{\eta}(\xi) - \mathbf{f}(\xi) \tag{13.1.7}$$

is the vector discrepancy, which we would like to be small over $R(\xi)$, between actual and approximate models. In other words, there are *two* types of errors which must be taken into account:

1. Systematic, or bias, errors $\boldsymbol{\delta}(\xi) = \boldsymbol{\eta}(\xi) - \mathbf{f}(\xi)$, the difference between the expected value of the response, $E(\mathbf{y}) = \boldsymbol{\eta}(\xi)$ and the approximating function $\mathbf{f}(\xi)$.
2. Random errors $\boldsymbol{\varepsilon}$.

Although the above implies that systematic errors $\boldsymbol{\delta}(\xi)$ are always to be expected, there has been, since the time of Gauss (1777–1855), an unfortunate tendency to ignore them, and to concentrate only on the random errors $\boldsymbol{\varepsilon}$. This seems to have been because nice mathematical results are possible when this is done. In choosing an experimental design, the ignoring of systematic error is *not* an innocuous approximation and misleading results may well be obtained, as we shall see in due course.

13.2. THE COMPETING EFFECTS OF BIAS AND VARIANCE

Consider first the elementary case where there is only one important predictor variable ξ. Figure 13.1 represents a typical situation in which the

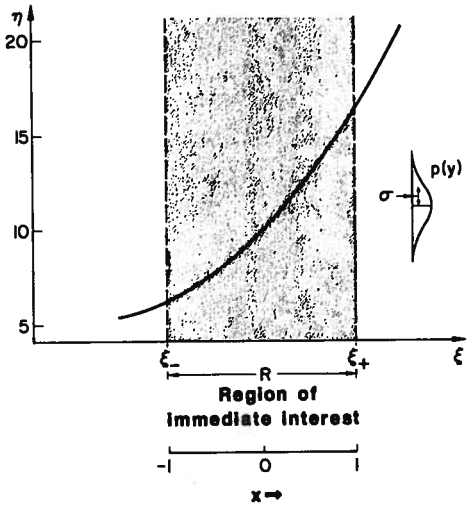

FIGURE 13.1. A hypothetical unknown relationship existing between $E(y) = \eta$ and ξ. At the right is shown the error distribution $p(y)$ which has standard deviation σ.