# 22

# Experimental Design

**KEY WORDS** *blocking, Box-Behnken, composite design, direct comparison, empirical models, emulsion breaking, experimental design, factorial design, field studies, interaction, iterative design, mechanistic models, one-factor-at-a-time experiment, OFAT, oil removal, precision, Plakett-Burman, randomization, repeats, replication, response surface, screening experiments, standard error,* t-*test.*

> "It is widely held by nonstatisticians that if you do good experiments statistics are not necessary. They are quite right.…The snag, of course, is that doing good experiments is difficult. Most people need all the help they can get to prevent them making fools of themselves by claiming that their favorite theory is substantiated by observations that do nothing of the sort.…" (Coloquhon, 1971).

We can all cite a few definitive experiments in which the results were intuitively clear without statistical analysis. This can only happen when there is an excellent experimental design, usually one that involves direct comparisons and replication. Direct comparison means that nuisance factors have been removed. Replication means that credibility has been increased by showing that the favorable result was not just luck. (If you do not believe me, I will do it again.) On the other hand, we have seen experiments where the results were unclear even after laborious statistical analysis was applied to the data. Some of these are the result of an inefficient experimental design.

*Statistical experimental design* refers to the work plan for manipulating the settings of the independent variables that are to be studied. Another kind of experimental design deals with building and operating the experimental apparatus. The more difficult and expensive the operational manipulations, the more statistical design offers gains in efficiency.

This chapter is a descriptive introduction to experimental design. There are many kinds of experimental designs. Some of these are one-factor-at-a-time, paired comparison, two-level factorials, fractional factorials, Latin squares, Graeco-Latin squares, Box-Behnken, Plackett-Burman, and Taguchi designs. An efficient design gives a lot of information for a little work. A "botched" design gives very little information for a lot of work. This chapter has the goal of convincing you that one-factor-at-a-time designs are poor (so poor they often may be considered botched designs) and that it is possible to get a lot of information with very few experimental runs. Of special interest are two-level factorial and fractional factorial experimental designs. Data interpretation follows in Chapters 23 through 48.

## What Needs to be Learned?

Start your experimental design with a clear statement of the question to be investigated and what you know about it. Here are three pairs of questions that lead to different experimental designs:

1.a. If I observe the system without interference, what function best predicts the output $y$?
b. What happens to $y$ when I change the inputs to the process?
2.a. What is the value of $\theta$ in the mechanistic model $y = x^{\theta}$?
b. What smooth polynomial will describe the process over the range $[x_1, x_2]$?



**TABLE 22.1**
Five Classes of Experimental Problems Defined in Terms of What is Unknown in the Model, $\eta = f(\mathbf{X}, \theta)$, Which is a Function of One or More Independent Variables **X** and One or More Parameters $\theta$

| Unknown | Class of Problem | Design Approach | Chapter |
|---|---|---|---|
| $f, \mathbf{X}, \theta$ | Determine a subset of important variables from a given larger set of potentially important variables | Screening variables | 23, 29 |
| $f, \theta$ | Determine empirical "effects" of known input variables **X** | Empirical model building | 27, 38 |
| $f, \theta$ | Determine a local interpolation or approximation function, $f(\mathbf{X}, \theta)$ | Empirical model building | 36, 37, 38, 40, 43 |
| $f, \theta$ | Determine a function based on mechanistic understanding of the system | Mechanistic model building | 46, 47 |
| $\theta$ | Determine values for the parameters | Model fitting | 35, 44 |

*Source:* Box, G. E. P. (1965). Experimemtal Strategy, Madison WI, Department of Statistics, Wisconsin Tech. Report #111, University of Wisconsin-Madison.

3.a. Which of seven potentially active factors are important?
b. What is the magnitude of the effect caused by changing two factors that have been shown important in preliminary tests?

A clear statement of the experimental objectives will answer questions such as the following:

1. What factors (variables) do you think are important? Are there other factors that might be important, or that need to be controlled? Is the experiment intended to show which variables are important or to estimate the effect of variables that are known to be important?
2. Can the experimental factors be set precisely at levels and times of your choice? Are there important factors that are beyond your control but which can be measured?
3. What kind of a model will be fitted to the data? Is an empirical model (a smoothing polynomial) sufficient, or is a mechanistic model to be used? How many parameters must be estimated to fit the model? Will there be interactions between some variables?
4. How large is the expected random experimental error compared with the expected size of the effects? Does my experimental design provide a good estimate of the random experimental error? Have I done all that is possible to eliminate bias in measurements, and to improve precision?
5. How many experiments does my budget allow? Shall I make an initial commitment of the full budget, or shall I do some preliminary experiments and use what I learn to refine the work plan?

Table 22.1 lists five general classes of experimental problems that have been defined by Box (1965). The model $\eta = f(\mathbf{X}, \theta)$ describes a response $\eta$ that is a function of one or more independent variables **X** and one or more parameters $\theta$. When an experiment is planned, the functional form of the model may be known or unknown; the active independent variables may be known or unknown. Usually, the parameters are unknown. The experimental strategy depends on what is unknown. A well-designed experiment will make the unknown known with a minimum of work.

## Principles of Experimental Design

Four basic principles of good experimental design are direct comparison, replication, randomization, and blocking.

### Comparative Designs

If we add substance $X$ to a process and the output improves, it is tempting to attribute the improvement to the addition of $X$. But this observation may be entirely wrong. $X$ may have no importance in the process.

Its addition may have been coincidental with a change in some other factor. The way to avoid a false conclusion about $X$ is to do a comparative experiment. Run parallel trials, one with $X$ added and one with $X$ not added. All other things being equal, a change in output can be attributed to the presence of $X$. Paired $t$-tests (Chapter 17) and factorial experiments (Chapter 27) are good examples of comparative experiments.

Likewise, if we passively observe a process and we see that the air temperature drops and output quality decreases, we are not entitled to conclude that we can cause the output to improve if we raise the temperature. Passive observation or the equivalent, dredging through historical records, is less reliable than direct comparison. If we want to know what happens to the process when we change something, we must observe the process when the factor is actively being changed (Box, 1966; Joiner, 1981).

Unfortunately, there are situations when we need to understand a system that cannot be manipulated at will. Except in rare cases (TVA, 1962), we cannot control the flow and temperature in a river. Nevertheless, a fundamental principle is that we should, whenever possible, do designed and controlled experiments. By this we mean that we would like to be able to establish specified experimental conditions (temperature, amount of $X$ added, flow rate, etc.). Furthermore, we would like to be able to run the several combinations of factors in an order that we decide and control.

## Replication

Replication provides an internal estimate of random experimental error. The influence of error in the effect of a factor is estimated by calculating the standard error. All other things being equal, the standard error will decrease as the number of observations and replicates increases. This means that the precision of a comparison (e.g., difference in two means) can be increased by increasing the number of experimental runs. Increased precision leads to a greater likelihood of correctly detecting small differences between treatments. It is sometimes better to increase the number of runs by replicating observations instead of adding observations at new settings.

Genuine repeat runs are needed to estimate the random experimental error. "Repeats" means that the settings of the $x$'s are the same in two or more runs. "Genuine repeats" means that the runs with identical settings of the $x$'s capture all the variation that affects each measurement (Chapter 9). Such replication will enable us to estimate the standard error against which differences among treatments are judged. If the difference is large relative to the standard error, confidence increases that the observed difference did not arise merely by chance.

## Randomization

To assure validity of the estimate of experimental error, we rely on the principle of randomization. It leads to an unbiased estimate of variance as well as an unbiased estimate of treatment differences. Unbiased means free of systemic influences from otherwise uncontrolled variation.

Suppose that an industrial experiment will compare two slightly different manufacturing processes, A and B, on the same machinery, in which A is always used in the morning and B is always used in the afternoon. No matter how many manufacturing lots are processed, there is no way to separate the difference between the machinery or the operators from morning or afternoon operation. A good experiment does not assume that such systematic changes are absent. When they affect the experimental results, the bias cannot be removed by statistical manipulation of the data. Random assignment of treatments to experimental units will prevent systematic error from biasing the conclusions.

Randomization also helps to eliminate the corrupting effect of serially correlated errors (i.e., process or instrument drift), nuisance correlations due to lurking variables, and inconsistent data (i.e., different operators, samplers, instruments).

Figure 22.1 shows some possibilities for arranging the observations in an experiment to fit a straight line. Both replication and randomization (run order) can be used to improve the experiment.

Must we randomize? In some experiments, a great deal of expense and inconvenience must be tolerated in order to randomize; in other experiments, it is impossible. Here is some good advice from Box (1990).



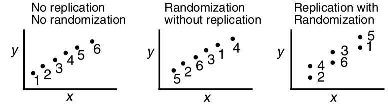



**FIGURE 22.1** The experimental designs for fitting a straight line improve from left to right as replication and randomization are used. Numbers indicate order of observation.

1. In those cases where randomization only slightly complicates the experiment, always randomize.
2. In those cases where randomization would make the experiment impossible or extremely difficult to do, but you can make an honest judgment about existence of nuisance factors, run the experiment without randomization. Keep in mind that wishful thinking is not the same as good judgment.
3. If you believe the process is so unstable that without randomization the results would be useless and misleading, and randomization will make the experiment impossible or extremely difficult to do, then do not run the experiment. Work instead on stabilizing the process or getting the information some other way.

## Blocking

The paired $t$-test (Chapter 17) introduced the concept of blocking. Blocking is a means of reducing experimental error. The basic idea is to partition the total set of experimental units into subsets (blocks) that are as homogeneous as possible. In this way the effects of nuisance factors that contribute systematic variation to the difference can be eliminated. This will lead to a more sensitive analysis because, loosely speaking, the experimental error will be evaluated in each block and then pooled over the entire experiment.

Figure 22.2 illustrates blocking in three situations. In (a), three treatments are to be compared but they cannot be observed simultaneously. Running A, followed by B, followed by C would introduce possible bias due to changes over time. Doing the experiment in three blocks, each containing treatment A, B, and C, in random order, eliminates this possibility. In (b), four treatments are to be compared using four cars. Because the cars will not be identical, the preferred design is to treat each car as a block and balance the four treatments among the four blocks, with randomization. Part (c) shows a field study area with contour lines to indicate variations in soil type (or concentration). Assigning treatment A to only the top of the field would bias the results with respect to treatments B and C. The better design is to create three blocks, each containing treatment A, B, and C, with random assignments.

## Attributes of a Good Experimental Design

A good design is simple. A simple experimental design leads to simple methods of data analysis. The simplest designs provide estimates of the main differences between treatments with calculations that amount to little more than simple averaging. Table 22.2 lists some additional attributes of a good experimental design.

If an experiment is done by unskilled people, it may be difficult to guarantee adherence to a complicated schedule of changes in experimental conditions. If an industrial experiment is performed under production conditions, it is important to disturb production as little as possible.

In scientific work, especially in the preliminary stages of an investigation, it may be important to retain flexibility. The initial part of the experiment may suggest a much more promising line of investigation, so that it would be a bad thing if a large experiment has to be completed before any worthwhile results are obtained. Start with a simple design that can be augmented as additional information becomes available.

**TABLE 22.2**

Attributes of a Good Experiment

A good experimental design should:

1. Adhere to the basic principles of randomization, replication, and blocking.
2. Be simple:
   a. Require a minimum number of experimental points
   b. Require a minimum number of predictor variable levels
   c. Provide data patterns that allow visual interpretation
   d. Ensure simplicity of calculation
3. Be flexible:
   a. Allow experiments to be performed in blocks
   b. Allow designs of increasing order to be built up sequentially
4. Be robust:
   a. Behave well when errors occur in the settings of the $x$'s
   b. Be insensitive to wild observations
   c. Be tolerant to violation of the usual normal theory assumptions
5. Provide checks on goodness of fit of model:
   a. Produce balanced information over the experimental region
   b. Ensure that the fitted value will be as close as possible to the true value
   c. Provide an internal estimate of the random experimental error
   d. Provide a check on the assumption of constant variance

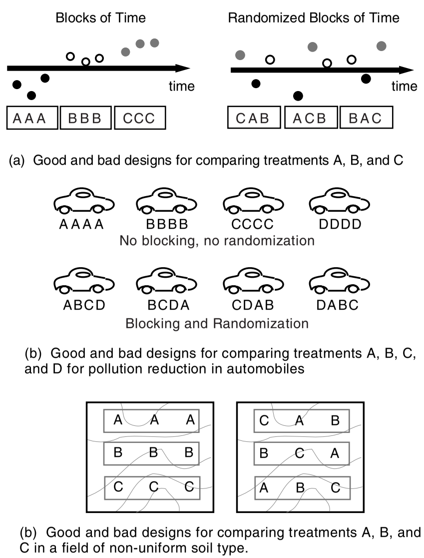



**FIGURE 22.2** Successful strategies for blocking and randomization in three experimental situations.



## One-Factor-At-a-Time (OFAT) Experiments

Most experimental problems investigate two or more factors (independent variables). The most inefficient approach to experimental design is, "Let's just vary one factor at a time so we don't get confused." If this approach does find the best operating level for all factors, it will require more work than experimental designs that simultaneously vary two or more factors at once.

These are some advantages of a good multifactor experimental design compared to a one-factor-at-a-time (OFAT) design:

- It requires less resources (time, material, experimental runs, etc.) for the amount of information obtained. This is important because experiments are usually expensive.
- The estimates of the effects of each experimental factor are more precise. This happens because a good design multiplies the contribution of each observation.
- The interaction between factors can be estimated systematically. Interactions cannot be estimated from OFAT experiments.
- There is more information in a larger region of the factor space. This improves the prediction of the response in the factor space by reducing the variability of the estimates of the response. It also makes the process optimization more efficient because the optimal solution is searched for over the entire factor space.

Suppose that jar tests are done to find the best operating conditions for breaking an oil–water emulsion with a combination of ferric chloride and sulfuric acid so that free oil can be removed by flotation. The initial oil concentration is 5000 mg/L. The first set of experiments was done at five levels of ferric chloride with the sulfuric acid dose fixed at 0.1 g/L. The test conditions and residual oil concentration (oil remaining after chemical coagulation and gravity flotation) are given below.

| | | | | | |
|---|---|---|---|---|---|
| **$FeCl_3$ (g/L)** | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 |
| **$H_2SO_4$ (g/L)** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **Residual oil (mg/L)** | 4200 | 2400 | 1700 | 175 | 650 |

The dose of 1.3 g/L of $FeCl_3$ is much better than the other doses that were tested. A second series of jar tests was run with the $FeCl_3$ level fixed at the apparent optimum of 1.3 g/L to obtain:

| | | | |
|---|---|---|---|
| **$FeCl_3$ (g/L)** | 1.3 | 1.3 | 1.3 |
| **$H_2SO_4$ (g/L)** | 0 | 0.1 | 0.2 |
| **Oil (mg/L)** | 1600 | 175 | 500 |

This test seems to confirm that the best combination is 1.3 g/L of $FeCl_3$ and 0.1 g/L of $H_2SO_4$.

Unfortunately, this experiment, involving eight runs, leads to a wrong conclusion. The response of oil removal efficiency as a function of acid and iron dose is a valley, as shown in Figure 22.3. The first one-at-a-time experiment cut across the valley in one direction, and the second cut it in the perpendicular direction. What appeared to be an optimum condition is false. A valley (or a ridge) describes the response surface of many real processes. The consequence is that one-factor-at-a-time experiments may find a false optimum. Another weakness is that they fail to discover that a region of higher removal efficiency lies in the direction of higher acid dose and lower ferric chloride dose.

We need an experimental strategy that (1) will not terminate at a false optimum, and (2) will point the way toward regions of improved efficiency. Factorial experimental designs have these advantages. They are simple and tremendously productive and every engineer who does experiments of any kind should learn their basic properties.

We will illustrate two-level, two-factor designs using data from the emulsion breaking example. A two-factor design has two independent variables. If each variable is investigated at two levels (high and

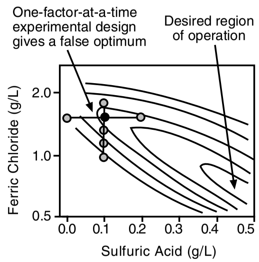



**FIGURE 22.3** Response surface of residual oil as a function of ferric chloride and sulfuric acid dose, showing a valley-shaped region of effective conditions. Changing one factor at a time fails to locate the best operating conditions for emulsion breaking and oil removal.

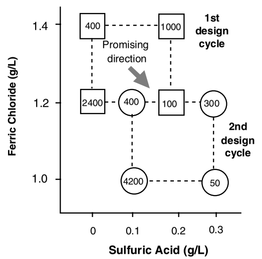



**FIGURE 22.4** Two cycles (a total of eight runs) of two-level, two-factor experimental design efficiently locate an optimal region for emulsion breaking and oil removal.

low, in general terms), the experiment is a two-level design. The total number of experimental runs needed to investigate two levels of two factors is $n = 2^2 = 4$. The $2^2$ experimental design for jar tests on breaking the oil emulsion is:

| Acid (g/L) | $FeCl_3$ (g/L) | Oil (mg/L) |
|---|---|---|
| 0 | 1.2 | 2400 |
| 0 | 1.4 | 400 |
| 0.2 | 1.2 | 100 |
| 0.2 | 1.4 | 1000 |

These four experimental runs define a small section of the response surface and it is convenient to arrange the data in a graphical display like Figure 22.4, where the residual oil concentrations are shown in the squares. It is immediately clear that the best of the tested conditions is high acid dose and low $FeCl_3$ dose. It is also clear that there might be a payoff from doing more tests at even higher acid doses and even lower iron doses, as indicated by the arrow. The follow-up experiment is shown by the circles in Figure 22.4.

The eight observations used in the two-level, two-factor designs come from the 28 actual observations made by Pushkarev et al. (1983) that are given in Table 22.3. The factorial design provides information



**TABLE 22.3**

Residual Oil (mg/L) after Treatment by Chemical Emulsion Breaking and Flotation

| $FeCl_3$ Dose (g/L) | Sulfuric Acid Dose (g/L $H_2SO_4$) 0 | 0.1 | 0.2 | 0.3 | 0.4 |
|---|---|---|---|---|---|
| 0.6 | — | — | — | — | 600 |
| 0.7 | — | — | — | — | 50 |
| 0.8 | — | — | — | 4200 | 50 |
| 0.9 | — | — | 2500 | 50 | 150 |
| 1.0 | — | 4200 | 150 | 50 | 200 |
| 1.1 | — | 2400 | 50 | 100 | 400 |
| 1.2 | 2400 | 1700 | 100 | 300 | 700 |
| 1.3 | 1600 | 175 | 500 | — | — |
| 1.4 | 400 | 650 | 1000 | — | — |
| 1.5 | 350 | — | — | — | — |
| 1.6 | 1600 | — | — | — | — |

*Source:* Pushkarev et al. 1983. *Treatment of Oil-Containing Wastewater*, New York, Allerton Press.

that allows the experimenter to iteratively and quickly move toward better operating conditions if they exist, and provides information about the interaction of acid and iron on oil removal.

## More about Interactions

Figure 22.5 shows two experiments that could be used to investigate the effect of pressure and temperature. The one-factor-at-a-time experiment (shown on the left) has experimental runs at these conditions:

| Test Condition | Yield |
|---|---|
| (1) Standard pressure and standard temperature | 10 |
| (2) Standard pressure and new temperature | 7 |
| (3) New pressure and standard temperature | 11 |

Imagine a total of $n = 12$ runs, 4 at each condition. Because we had four replicates at each test condition, we are highly confident that changing the temperature at standard pressure decreased the yield by 3 units. Also, we are highly confidence that raising the temperature at standard pressure increased the yield by 1 unit.

Will changing the temperature at the new pressure also decrease the yield by 3 units? The data provide no answer. The effect of temperature on the response at the new temperature cannot be estimated.

Suppose that the 12 experimental runs are divided equally to investigate four conditions as in the two-level, two-factor experiment shown on the right side of Figure 22.5.

| Test Condition | Yield |
|---|---|
| (1) Standard pressure and standard temperature | 10 |
| (2) Standard pressure and new temperature | 7 |
| (3) New pressure and standard temperature | 11 |
| (4) New pressure and new temperature | 12 |

At the standard pressure, the effect of change in the temperature is a decrease of 3 units. At the new pressure, the effect of change in temperature is an increase of 1 unit. The effect of a change in temperature depends on the pressure. There is an *interaction* between temperature and pressure. The experimental effort was the same (12 runs) but this experimental design has produced new and useful information (Czitrom, 1999).

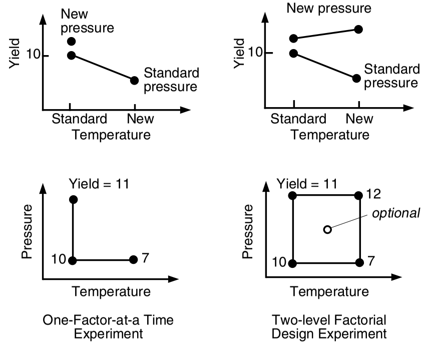



**FIGURE 22.5** Graphical demonstration of why one-factor-at-a-time (OFAT) experiments cannot estimate the two-factor interaction between temperature and pressure that is revealed by the two-level, two-factor design.

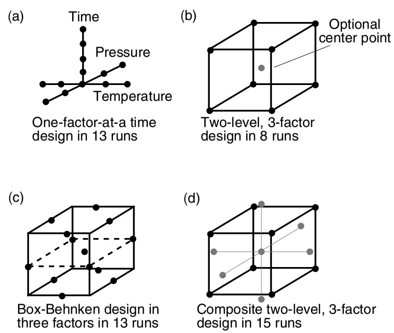



**FIGURE 22.6** Four possible experimental designs for studying three factors. The worst is (a), the one-factor-at-a-time design (top left). (b) is a two-level, three-factor design in eight runs and can describe a smooth nonplanar surface. The Box-Behnken design (c) and the composite two-level, three-factor design (d) can describe quadratic effects (maxima and minima). The Box-Behnken design uses 12 observations located on the face of the cube plus a center point. The composite design has eight runs located at the corner of the cube, plus six "star" points, plus a center point. The corner and star points are equidistant from the center (i.e., located on a sphere having a diameter equal to the distance from the center to a corner).

It is generally true that (1) the factorial design gives better precision than the OFAT design if the factors *do* act additively; and (2) if the factors *do not* act additively, the factorial design can detect and estimate interactions that measure the nonadditivity.

As the number of factors increases, the benefits of investigating several factors simultaneously increases. Figure 22.6 illustrates some designs that could be used to investigate three factors. The one-factor-at-a time design (Figure 22.6a) in 13 runs is the worst. It provides no information about interactions and no information about curvature of the response surface. Designs (b), (c), and (d) do provide estimates



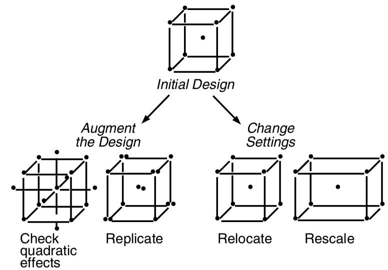



**FIGURE 22.7** Some of the modifications that are possible with a two-level factorial experimental design. It can be stretched (rescaled), replicated, relocated, or augmented.

of interactions as well as the effects of changing the three factors. Figure 22.6b is a two-level, three-factor design in eight runs that can describe a smooth nonplanar surface. The *Box-Behnken design* (c) and the composite two-level, three-factor design (d) can describe quadratic effects (maxima and minima). The Box-Behnken design uses 12 observations located on the face of the cube plus a center point. The *composite design* has eight runs located at the corner of the cube, plus six "star" points, plus a center point. There are advantages to setting the corner and star points equidistant from the center (i.e., on a sphere having a diameter equal to the distance from the center to a corner).

Designs (b), (c), and (d) can be replicated, stretched, moved to new experimental regions, and expanded to include more factors. They are ideal for iterative experimentation (Chapters 43 and 44).

## Iterative Design

Whatever our experimental budget may be, we never want to commit everything at the beginning. Some preliminary experiments will lead to new ideas, better settings of the factor levels, and to adding or dropping factors from the experiment. The oil emulsion-breaking example showed this. The importance of iterative experimentation is discussed again in Chapters 43 and 44. Figure 22.7 suggests some of the iterative modifications that might be used with two-level factorial experiments.

## Comments

A good experimental design is simple to execute, requires no complicated calculations to analyze the data, and will allow several variables to be investigated simultaneously in few experimental runs.

Factorial designs are efficient because they are balanced and the settings of the independent variables are completely uncorrelated with each other (orthogonal designs). Orthogonal designs allow each effect to be estimated independently of other effects.

We like factorial experimental designs, especially for treatment process research, but they do not solve all problems. They are not helpful in most field investigations because the factors cannot be set as we wish. A professional statistician will know other designs that are better. Whatever the final design, it should include replication, randomization, and blocking.

Chapter 23 deals with selecting the sample size in some selected experimental situations. Chapters 24 to 26 explain the analysis of data from factorial experiments. Chapters 27 to 30 are about two-level factorial and fractional factorial experiments. They deal mainly with identifying the important subset of experimental factors. Chapters 33 to 48 deal with fitting linear and nonlinear models.

## References


Berthouex, P. M. and D. R. Gan (1991). "Fate of PCBs in Soil Treated with Contaminated Municipal Sludge," *J. Envir. Engr. Div.*, ASCE, 116(1), 1–18.

Box, G. E. P. (1965). Experimental Strategy, Madison, WI, Department of Statistics, Wisconsin Tech. Report #111, University of Wisconsin–Madison.

Box, G. E. P. (1966). "The Use and Abuse of Regression," *Technometrics,* 8, 625–629.

Box, G. E. P. (1982). "Choice of Response Surface Design and Alphabetic Optimiality," *Utilitas Mathematica,* 21B, 11–55.

Box, G. E. P. (1990). "Must We Randomize?," *Qual. Eng.,* 2, 497–502.

Box, G. E. P., W. G. Hunter, and J. S. Hunter (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building,* New York, Wiley Interscience.

Colquhoun, D. (1971). *Lectures in Biostatistics,* Oxford, England, Clarendon Press.

Czitrom, Veronica (1999). "One-Factor-at-a Time Versus Designed Experiments," *Am. Stat.,* 53(2), 126–131.

Joiner, B. L. (1981). "Lurking Variables: Some Examples," *Am. Stat.,* 35, 227–233.

Pushkarev et al. (1983). *Treatment of Oil-Containing Wastewater,* New York, Allerton Press.

Tennessee Valley Authority (1962). *The Prediction of Stream Reaeration Rates,* Chattanooga, TN.

Tiao, George, S. Bisgarrd, W. J. Hill, D. Pena, and S. M. Stigler, Eds. (2000). *Box on Quality and Discovery with Design, Control, and Robustness,* New York, John Wiley & Sons.


## Exercises

**22.1** Straight Line. You expect that the data from an experiment will describe a straight line. The range of $x$ is from 5 to 50. If your budget will allow 12 runs, how will you allocate the runs over the range of $x$? In what order will you execute the runs?

**22.2** OFAT. The instructions to high school science fair contestants states that experiments should only vary one factor at a time. Write a letter to the contest officials explaining why this is bad advice.

**22.3** Planning. Select one of the following experimental problems and (a) list the experimental factors, (b) list the responses, and (c) explain how you would arrange an experiment. Consider this a brainstorming activity, which means there are no wrong answers. Note that in 3, 4, and 5 some experimental factors and responses have been suggested, but these should not limit your investigation.

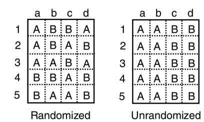
1. Set up a bicycle for long-distance riding.
2. Set up a bicycle for mountain biking.
3. Investigate how clarification of water by filtration will be affected by such factors as pH, which will be controlled by addition of hydrated lime, and the rate of flow through the filter.
4. Investigate how the dewatering of paper mill sludge would be affected by such factors as temperature, solids concentration, solids composition (fibrous vs. granular material), and the addition of polymer.
5. Investigate how the rate of disappearance of oil from soil depends on such factors as soil moisture, soil temperature, wind velocity, and land use (tilled for crops vs. pasture, for example).
6. Do this for an experiment that you have done, or one that you would like to do.

**22.4** Soil Sampling. The budget of a project to explore the extent of soil contamination in a storage area will cover the collection and analysis of 20 soil specimens, or the collection of 12 specimens with duplicate analyses of each, or the collection of 15 specimens with duplicate analyses of 6 of these specimens selected at random. Discuss the merits of each plan.



**22.5** Personal Work. Consider an experiment that you have performed. It may be a series of analytical measurements, an instrument calibration, or a process experiment. Describe how the principles of direct comparison, replication, randomization, and blocking were incorporated into the experiment. If they were not practiced, explain why they were not needed, or why they were not used. Or, suggest how the experiment could have been improved by using them.

**22.6** Trees. It is proposed to study the growth of two species of trees on land that is irrigated with treated industrial wastewater effluent. Ten trees of each species will be planted and their growth will be monitored over a number of years. The figure shows two possible schemes. In one (left panel) the two kinds of trees are allocated randomly to 20 test plots of land. In the other (right panel) the species A is restricted to half the available land and species B is planted on the other. The investigator who favors the randomized design plans to analyze the data using an independent $t$-test. The investigator who favors the unrandomized design plans to analyze the data using a paired $t$-test, with the average of 1a and 1b being paired with 1c and 1d. Evaluate these two plans. Suggest other possible arrangements. Optional: Design the experiment if there are four species of tress and 20 experimental plots.




**22.7** Solar Energy. The production of hot water is studied by installing ten units of solar collector A and ten units of solar collector B on homes in a Wisconsin town. Propose some experimental designs and discuss their advantages and disadvantages.

**22.8** River Sampling. A river and one of its tributary streams were monitored for pollution and the following data were obtained:

| | | | | | |
|---|---|---|---|---|---|
| **River** | 16 | 12 | 14 | 11 | |
| **Tributary** | 9 | 10 | 8 | 6 | 5 |

It was claimed that this proves the tributary is cleaner than the river. The statistician who was asked to confirm this impression asked a series of questions. *When were the data taken? All in one day? On different days? Were the data taken during the same time period for the two streams? Were the temperatures of the two streams the same? Where in the streams were the data taken? Why were these points chosen? Are they representative?*

Why do you think the statistician asked these questions? Are there other questions that should have been asked? Is there any set of answers to these questions that would justify the use of a $t$-test to draw conclusions about pollution levels?