

## 3.2 Designing Experiments

**Definition.** An *observational study* observes individuals and measures variables of interest but does not attempt to influence the responses. An *experiment*, on the other hand, deliberately imposes some treatment on individuals in order to observe their responses.

**Definition.** The individuals on which the experiment is done are the *experimental units*. When the units are human beings, they are called *subjects*. A specific experimental condition applied to the units is called a *treatment*.

**Definition.** The explanatory variables in an experiment are often called *factors*. Many experiments study the joint effects of several factors. In such an experiment, each treatment is formed by combining a specific value (often called a *level*) of each of the factors.

**Example 3.8.** Researchers studying the absorption of a drug into the bloodstream inject the drug (the treatment) into 25 people (the subjects). The response variable is the concentration of the drug in a subject's blood, measured 30 minutes after the injection. This experiment has a single factor with only one level. If three different doses of the drug are injected, there is still a single factor (the dosage of the drug), now with three levels. The three levels of the single factor are the treatments that the experiment compares.

**Note.** Experimentation allows us to study the effects of the specific treatments we are interested in. Moreover, we can control the environment of the experimental units to hold constant factors that are of no interest to us. Another advantage of experiments is that we can study the combined effects of several factors simultaneously. The interaction of several factors can produce effects that could not be predicted from looking at the effect of each factor alone.

### Comparative Experiments

**Example 3.10.** Ulcers in the upper intestine are unfortunately common in modern society. “Gastric freezing” is a clever treatment for ulcers. The patient swallows a deflated balloon with tubes attached, then a refrigerated solution is pumped through the balloon for an hour. The idea is that cooling the stomach will reduce its production of acid and so relieve ulcers. An experiment reported in the *Journal of the American Medical Association* showed that gastric freezing did reduce acid production and relieve ulcer pain. The treatment was safe and easy and was widely used for several years. The gastric freezing experiment was poorly designed. The patients’ response may have been due to the *placebo effect*. A placebo is a dummy treatment that can have no physical effect. Many patients respond favorably to *any* treatment, even a placebo, presumably because of trust in the doctor and expectations of a cure. This response to a dummy treatment is the placebo effect. A second experiment, done several years later, divided ulcer patients into two groups. One group was treated by gastric freezing as before.

The other group received a placebo treatment in which the solution in the balloon was at body temperature rather than freezing. The results: 34% of the 82 patients in the treatment group improved, but so did 38% of the 78 patients in the placebo group. This and other designed experiments showed that gastric freezing was no better than a placebo, and its use was abandoned.

**Note.** Experiments should compare treatments rather than attempt to assess a single treatment in isolation. The group of patients who received a sham treatment is called a *control group*, because it enables us to control the effects of lurking variables on the outcome. Control of the effects of lurking variables is the first principle of statistical design of experiments.

### Completely Randomized Experiments

**Definition.** The use of chance to divide experimental units into groups is called *randomization*. Randomization is the second major principle of statistical design of experiments.

**Example 3.12.** Many utility companies have introduced programs to encourage energy conservation among their customers. An electric company considers placing electronic indicators in households to show what the cost would be if the electricity use at that moment continued for a month. Will indicators reduce electricity use? Would cheaper methods work almost as well? The company decides to design an experiment.

One cheaper approach is to give customers a chart and information about monitoring their electricity use. The experiment compares these two approaches (indicator, chart) and also a control. The control group of customers receives information about energy conservation but no help in monitoring electricity use. The response variable is total electricity used in a year. The company finds 60 single-family residences in the same city willing to participate, so it assigns 20 residences at random to each of the 3 treatments. (See page 205 for an outline of the design). To carry out the random assignment, label the 60 households 01 to 60. Enter Table B and select an SRS of 20 to receive the indicators. Continue in Table B, selecting 20 more to receive charts. The remaining 20 form the control group.

**Definition.** In a *completely randomized* experimental design, all the experimental units are allocated at random among all the treatments.

### The Logic of Experimental Design

**Note.** The logic behind a randomized comparative design is as follows:

- Randomization produces groups of experimental units that should be similar in all respects before the treatments are applied.
- Comparative design insures that influences other than the experimental treatments operate equally on all groups.
- Therefore, differences in the response variable must be due to the effects of the treatments. That is, the treatments not only are

associated with the observed differences in the response but must actually cause them.

The great advantage of randomized comparative experiments is that they can produce data that give good evidence for a cause-and-effect relationship between the explanatory and response variables.

**Definition.** An observed effect too large to attribute plausibly to chance is called *statistically significant*.

**Note.** If we observe statistically significant differences among the groups after a comparative randomized experiment, we have good evidence that the treatments actually caused these differences.

**Definition.** A third principle of statistical design of experiments is *replication*: repeat each treatment on a large enough number of experimental units or subjects to allow the systematic effects of the treatments to be seen.

**Note.** The basic principles of statistical design of experiments are:

1. *Control* the effects of lurking variables on the response, most simply by comparing several treatments.
2. *Randomization*, the use of impersonal chance to assign subjects to treatments.
3. *Replication* of the experiment on many subjects to reduce chance variation in the results.

## Cautions about Experimentation

**Definition.** In a *double-blind* experiment, neither the subjects nor the people who have contact with them know which treatment a subject received.

**Note.** The most serious potential weakness of experiments is *lack of realism*. The subjects or treatments or setting of an experiment may not realistically duplicate the conditions we really want to study.

## Other Experimental Designs

**Definition.** A *block* is a group of experimental units or subjects that are similar in ways that are expected to affect the response to the treatments. In a *block design*, the random assignment of units to treatments is carried out separately within each block.

**Note.** A simple and common special type of block design is the *matched pairs design*. Matched pairs designs compare just two treatments. Each block consists of just two units, as closely matched as possible. These units are assigned at random to the treatments.

**Example 3.16.** Pepsi once wanted to demonstrate that Coke drinkers in fact prefer Pepsi when they taste both colas blind. The subjects, all people who said they were Coke drinkers, tasted both colas from glasses without brand markings and said which they liked better. This

is a matched pairs design in which each subject compares the two colas. Because responses may depend on which cola is tasted first, the order of tasting should be chosen at random for each subject. When more than half the Coker drinkers chose Pepsi, Coke claimed that the experiment was biased. The Pepsi glasses were marked  $M$  and Coke glasses were marked  $Q$ . Aha, said Coke, this just shows that people like the letter  $M$  better than the letter  $Q$ . A careful experiment would in fact take care to avoid any distinction other than the actual treatments.