Chapter 4

Anonymity When the names of individuals participating in a study are not known even to the director of the study.

Bias The design of a statistical study shows bias if it systematically favors certain outcomes.

Block A group of experimental units that are known before the experiment to be similar in some way that is expected to affect the response to the treatments.

Census A study that attempts to collect data from every individual in the population.

Cluster sample To take a cluster sample, first divide the population into smaller groups. Ideally, these clusters should mirror the characteristics of the population. Then choose an SRS of the clusters. All individuals in the chosen clusters are included in the sample.

Completely randomized design When the treatments are assigned to all the experimental units completely by chance.

Confidentiality A basic principle of data ethics that requires individual data to be kept private.

Confounding When two variables are associated in such a way that their effects on a response variable cannot be distinguished from each other.

Control An important experimental design principle. Researchers should control for lurking variables that might affect the response by using a comparative design and ensuring that the only systematic difference between the groups is the treatment administered.

Control group An experimental group whose primary purpose is to provide a baseline for comparing the effects of the other treatments. Depending on the purpose of the experiment, a control group may be given a placebo or an active treatment.

Convenience sample A sample selected by taking the members of the population that are easiest to reach; particularly prone to large bias.

Double-blind An experiment in which neither the subjects nor those who interact with them and measure the response variable know which treatment a subject received.

Experiment Deliberately imposes some treatment on individuals to measure their responses.

Experimental units The smallest collection of individuals to which treatments are applied.

Explanatory variable A variable that helps explain or influences changes in a response variable.

Factor The explanatory variables in an experiment are often called factors.

Inference about cause and effect Using the results of an experiment to conclude that the treatments caused the difference in responses. Requires a well-designed experiment in which the treatments are randomly assigned to the experimental units.

Inference about the population Using information from a sample to draw conclusions about the larger population. Requires that the individuals taking part in a study be randomly selected from the population of interest.

Informed consent A basic principle of data ethics. Individuals must be *informed* in advance about the nature of a study and any risk of harm it may bring. Participating individuals must then *consent* in writing.

Institutional review board A basic principle of data ethics. All planned studies must be approved in advance and monitored by an institutional review board charged with protecting the safety and well-being of the participants.

Lack of realism When the treatments, the subjects, or the environment of an experiment are not realistic. Lack of realism can limit researchers' ability to apply the conclusions of an experiment to the settings of greatest interest.

Level A specific value of an explanatory variable (factor) in an experiment.

Lurking variable A variable that is not among the explanatory or response variables in a study but that may influence the response variable.

Matched pair **A** common form of blocking for comparing just two treatments. In some matched pairs designs, each subject receives both treatments in a random order. In others, the subjects are matched in pairs as closely as possible, and each subject in a pair is randomly assigned to receive one of the treatments.

Margin of error A numerical estimate of how far the sample result is likely to be from the truth about the population due to sampling variability.

Nonresponse Occurs when a selected individual cannot be contacted or refuses to cooperate; an example of a nonsampling error.

Nonsampling error The most serious errors in most careful surveys are nonsampling errors. These have nothing to do with choosing a sample—they are present even in a census. Some common examples of nonsampling errors are nonresponse, response bias, and errors due to question wording.

Observational study Observes individuals and measures variables of interest but does not attempt to influence the responses.

Placebo An inactive (fake) treatment.

Placebo effect Describes the fact that some subjects respond favorably to any treatment, even an inactive one (placebo).

Population In a statistical study, the population is the entire group of individuals about which we want information.

Random assignment An important experimental design principle. Use some chance process to assign experimental units to treatments. This helps create roughly equivalent groups of experimental units by balancing the effects of lurking variables that aren't controlled on the treatment groups.

Random sampling The use of chance to select a sample; is the central principle of statistical sampling.

Randomized block design Start by forming blocks consisting of individuals that are similar in some way that is important to the response. Random assignment of treatments is then carried out separately within each block.

Replication An important experimental design principle. Use enough experimental units in each group so that any differences in the effects of the treatments can be distinguished from chance differences between the groups.

Response bias A systemic pattern of incorrect responses.

Response variable A variable that measures an outcome of a study.

Sample The part of the population from which we actually collect information. We use information from a sample to draw conclusions about the entire population.

Sampling error Mistakes made in the process of taking a sample that could lead to inaccurate information about the population**.** Bad sampling methods and undercoverage are common types of sampling error.

Sample survey A study that uses an organized plan to choose a sample that represents some specific population. We base conclusions about the population on data from the sample.

Sampling frame The list from which a sample is actually chosen.

Simple random sample (SRS) The basic random sampling method. An SRS gives every possible sample of a given size the same chance to be chosen. We often choose an SRS by labeling the members of the population and using random digits to select the sample.

Single-blind An experiment in which either the subjects or those who interact with them and measure the response variable, but not both, know which treatment a subject received.

Statistically significant An observed effect so large that it would rarely occur by chance.

Strata Groups of individuals in a population that are similar in some way that might affect their responses.

Stratified random sample To select a stratified random sample, first classify the population into groups of similar individuals, called strata. Then choose a separate SRS from each stratum to form the full sample.

Subjects Experimental units that are human beings.

Table of random digits A long string of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 with these properties:
• Each entry in the table is equally likely to be any of the 10 digits 0 through 9.
• The entries are independent of each other. That is, knowledge of one part of the table gives no information about any other part.

Treatment A specific condition applied to the individuals in an experiment. If an experiment has several explanatory variables, a treatment is a combination of specific values of these variables.

Undercoverage Occurs when some members of the population are left out of the sampling frame; a type of sampling error.

Voluntary response samples People decide whether to join a sample based on an open invitation; particularly prone to large bias.

Wording of questions The most important influence on the answers given to a survey. Confusing or leading questions can introduce strong bias, and changes in wording can greatly change a survey's outcome. Even the order in which questions are asked matters.