

Vocabulary

Term	Definition
Bias	Any systematic failure of a sampling method to represent its population is bias. It is almost impossible to recover from bias, so efforts to avoid it are well spent. Common errors include relying on voluntary response, undercoverage of the population, nonresponse bias, or response bias.
Blinding	Any individual associated with an experiment who is not aware of how subjects have been allocated to treatment groups is said to be blind.
Block	When groups of experimental units are similar, it is often a good idea to gather them together into blocks. By blocking we isolate the variability attributable to the differences between the blocks so that we can see the differences caused by the treatments more clearly.
Census	A sample that consists of the entire population is called a census.
Cluster sample	A sampling design in which entire groups, or clusters , are chosen at random. Cluster sampling is usually selected as a matter of convenience, practicality, or cost. Each cluster should be heterogeneous (and representative of the population), so all the clusters should be similar to each other.
Confounding	When the levels of one factor are associated with the levels of another factor so their effects cannot be separated, we say that these two factors are confounded.
Control group	The experimental units assigned to a baseline treatment level, typically either the default treatment, which is well understood, or a null, placebo treatment. Their responses provide a basis for comparison.
Convenience sample	A convenience sample consists of the individuals who are conveniently available. Convenience samples often fail to be representative because every individual in the population is not equally convenient to sample.
Designs	In a randomized block design , the randomization occurs only within blocks. In a completely randomized design , all experimental units have an equal chance of receiving any treatment.
Experiment	An experiment <i>manipulates</i> factor levels to create treatments, <i>randomly assigns</i> subjects to these treatment levels, and then <i>compares</i> the responses of the subject groups across treatment levels.
Experimental units	Individual on whom an experiment is performed. Usually called subjects or participants when they are human.
Factor	A variable whose levels are controlled by the experimenter. Experiments attempt to discover the effects that differences in factor levels may have on the responses of the experimental units.
Level	The specific values that the experimenter chooses for a factor are called the levels of the factor.

Vocabulary

Term	Definition
Matching	Any attempt to force a sample to resemble specified attributes of the population is a form of matching. Matching may help make better samples, but it is no substitute for randomizing.
Matching	In a retrospective or prospective study, subjects who are similar in ways not under study may be matched and then compared with each other on the variables of interest. Matching, like blocking, reduces unwanted variation.
Multistage sample	Sampling schemes that combine several sampling methods are called multistage samples. For example, a national polling service may stratify the country by geographical regions, select a random sample of cities from each region, and then interview a cluster of residents in each city.
Nonresponse bias	Bias introduced to a sample when a large fraction of those sampled fails to respond. Those who do respond are likely to not represent the entire sample. Voluntary response bias is a form of nonresponse bias, but nonresponse may occur for other reasons. For example, those who are at work during the day won't respond to a telephone survey conducted only during working hours.
Observational study	A study based on data in which no manipulation of factors has been employed.
Outcome	An individual result of a component of a simulation is its outcome
Placebo	A treatment known to have no effect, administered so that all groups experience the same conditions. Many subjects respond to such a treatment (a response known as a placebo effect). Only by comparing with a placebo can we be sure that the observed effect of a treatment is not due simply to the placebo effect.
Placebo effect	The tendency of many human subjects (often 20% or more of the experiment subjects) to show a response even when administered a placebo.
Population	The entire group of individuals or instances about whom we hope to learn.
Population parameter	A numerically valued attribute of a model for a population. We rarely expect to know the true value of a population parameter, but we do hope to estimate it from sampled data. For example, the mean income of all employed people in the country is a population parameter.
Principles of experimental design	Control aspects of the experiment that we know may have an effect on the response, but that are not the factors being studied. Randomize subjects to treatments to even out effects that we cannot control. Replicate over as many subjects as possible. Results for a single subject are just anecdotes. If, as often happens, the subjects of the experiment are not a representative sample from the population of interest, replicate the entire study with a different group of subjects, preferably from a different part of the population. Block to reduce the effects of identifiable attributes of the subjects that cannot be controlled.

Vocabulary

Term	Definition
Prospective study	An observational study in which subjects are followed to observe future outcomes. Because no treatments are deliberately applied, a prospective study is not an experiment. Nevertheless, prospective studies typically focus on estimating differences among groups that might appear as the groups are followed during the course of the study.
Random	An event is random if we know what outcomes could happen, but not which particular values will happen.
Random assignment	To be valid, an experiment must assign experimental units to treatment groups at random. This is called random assignment.
Random numbers	Random numbers are hard to generate. Nevertheless, several internet sites offer an unlimited supply of equally likely random values.
Randomization	The best defense against bias is randomization, in which each individual is given a fair, random chance of selection.
Representative	A sample is said to be representative if the statistics computed from it accurately reflect the corresponding population parameters.
Response	A variable whose values are compared across different treatments. In a randomized experiment, large response differences can be attributed to the effect of differences in treatment level.
Response bias	Anything in a survey design that influences responses falls under the heading of response bias. One typical response bias arises from the wording of questions, which may suggest a favored response. Voters, for example, are more likely to express support of "the president" than support of the particular person holding that office at the moment.
Response variables	Values of the response variable record the results of each trial with respect to what we were interested in.
Retrospective study	An observational study in which subjects are selected and then their previous conditions or behaviors are determined. Because retrospective studies are not based on random samples, they usually focus on estimating differences between groups or associations between variables.
Sample	A (representative) subset of a population, examined in hope of learning about the population.
Sample size	The number of individuals in a sample. The sample size determines how well the sample represents the population, not the fraction of the population sampled.
Sample survey	A study that asks questions of a sample drawn from some population in the hope of learning something about the entire population. Polls taken to assess voter preferences are common sample surveys.

Vocabulary

Term	Definition
Sampling frame	A list of individuals from whom the sample is drawn is called the sampling frame. Individuals who may be in the population of interest but who are not in the sampling frame cannot be included in any sample.
Sampling variability	The natural tendency of randomly drawn samples to differ, one from another. Sometimes, unfortunately, called <i>sampling error</i> , sampling variability is no error at all, but just the natural result of random sampling.
Simple random sample (SRS)	A simple random sample of sample size n is one in which each set of n elements has an equal chance of selection.
Simulation	A simulation models random events by using random numbers to specify event outcomes with relative frequencies that correspond to the true real-world relative frequencies we are trying to model.
Simulation component	The most basic situation in a simulation in which something happens at random.
Single-blind, double-blind	There are two main classes of individuals who can affect the outcome of an experiment: * those who could <i>influence the results</i> (the subjects, treatment administrators, or technicians). * Those who <i>evaluate the results</i> (judges, treatment physicians, etc.). When every individual in <i>either</i> of these classes is blinded, an experiment is said to be single-blind. When everyone in <i>both</i> classes is blinded, we call the experiment double-blind.
Statistic, sample statistic	Statistics are values calculated for sampled data. Those that correspond to, and thus estimate, a population parameter, are of particular interest. For example, the mean income of all employed people in a representative sample can provide a good estimate of the corresponding population parameter. The term "sample statistic" is sometimes used, usually to parallel the corresponding term, "population parameter."
Statistically significant	When an observed difference is too large for us to believe that it is likely to have occurred naturally, we consider the difference to be statistically significant. Subsequent chapters will show specific calculations and give rules, but the principle remains the same.
Stratified random sampling	A sampling design in which the population is divided into several subpopulations, or strata , and random samples are then drawn from each stratum. If the strata are homogeneous but are different from each other, a stratified sample may yield more consistent results.
Systematic sample	A sample drawn by selecting individuals systematically from a sampling frame. When there is no relationship between the order of the sampling frame and the variables of interest, a systematic sample can be representative.
Treatment	The process, intervention, or other controlled circumstance applied to randomly assigned experimental units. Treatments are the different levels of a single factors or are made up of combinations of levels of two or more factors.

Vocabulary

Term	Definition
Trial	The sequence of several components representing events that we are pretending will take place.
Undercoverage	A sampling scheme that biases the sample in a way that gives a part of the population less representation than it has in the population, suffers from undercoverage.
Voluntary response bias	Bias introduced to a sample when individuals can choose on their own whether to participate in the sample. Samples based on voluntary response are always invalid and cannot be recovered, no matter how large the sample size.