Experimental and observational studies:

A common goal for a statistical research project is to investigate causality, and in particular to draw a conclusion on the effect of changes in the values of predictors or independent variables on dependent variables. There are two major types of causal statistical studies: experimental studies and observational studies. In both types of studies, the effect of differences of an independent variable (or variables) on the behavior of the dependent variable are observed. The difference between the two types lies in how the study is actually conducted. Each can be very effective. An experimental study involves taking measurements of the system under study, manipulating the system, and then taking additional measurements using the same procedure to determine if the manipulation has modified the values of the measurements. In contrast, an observational study does not involve experimental manipulation. Instead, data are gathered and correlations between predictors and response are investigated. While the tools of data analysis work best on data from randomized studies, they are also applied to other kinds of data—like natural experiments and observational studies—for which a statistician would use a modified, more structured estimation method (e.g., Difference in differences estimation and instrumental variables, among many others) that produce consistent estimators.

The basic steps of a statistical experiment are:

1.Planning the research, including finding the number of replicates of the study, using the following information: preliminary estimates regarding the size of treatment effects, alternative hypotheses, and the estimated experimental variability. Consideration of the selection of experimental subjects and the ethics of research is necessary. Statisticians recommend that experiments compare (at least) one new treatment with a standard treatment or control, to allow an unbiased estimate of the difference in treatment effects.

2.Design of experiments, using blocking to reduce the influence of confounding variables, and randomized assignment of treatments to subjects to allow unbiased estimates of treatment effects and experimental error. At this stage, the experimenters and statisticians write the experimental protocol that will guide the performance of the experiment and which specifies the primary analysis of the experimental data.

3. Performing the experiment following the experimental protocol and analyzing the data following the experimental protocol.

4. Further examining the data set in secondary analyses, to suggest new hypotheses for future study.

5. Documenting and presenting the results of the study.

Scientific control

A scientific control is an experiment or observation designed to minimize the effects of variables other than the independent variable. This increases the reliability of the results, often through a comparison between control measurements and the other measurements. Scientific controls are a part of the scientific method.

In controlled experiments, the same experiment is done in at least two parallel experiments that differ in only one way, with one experiment being the "control arm" and the other being the "experimental arm".

Controls eliminate alternate explanations of experimental results, especially experimental errors and experimenter bias. Many controls are specific to the type of experiment being performed, as in the molecular markers used in SDS-PAGE experiments, and may simply have the purpose of ensuring that the equipment is working properly. The selection and use of proper controls to ensure that experimental results are valid (for example, absence of confounding variables) can be very difficult. Control measurements may also be used for other purposes: for example, a measurement of a microphone's background noise in the absence of a signal allows the noise to be subtracted from later measurements of the signal, thus producing a processed signal of higher quality.

For example, if a researcher feeds an experimental artificial sweetener to sixty laboratory rats and observes that ten of them subsequently become sick, the underlying cause could be the sweetener itself or something unrelated. Other variables, which may not be readily obvious, may interfere with the experimental design. For instance, the artificial sweetener might be mixed with a dilutant and it might be the dilutant which causes the effect. To control for the effect of the dilutant, the same test is run twice; once with the artificial sweetener in the dilutant, and another done exactly the same way, but using the dilutant alone. Now the experiment is controlled for the dilutant and the experimenter can distinguish between sweetener, dilutant and non-treatment.

Controls are most often necessary where a confounding factor cannot easily be separated from the primary treatments. For example, it may be necessary to use a tractor to spread fertilizer where there is no other practicable way to spread fertilizer. The simplest solution is to have a treatment where a tractor is driven over plots without spreading fertilizer and in that way the effects of tractor traffic are controlled.

The simplest types of control are negative and positive controls, and both are found in many different types of experiments. These two controls, when both are successful, are usually sufficient to eliminate most potential confounding variables: it means that the experiment produces a negative result when a negative result is expected, and a positive result when a positive result is expected.

Randomization

In randomization, the groups that receive different experimental treatments are determined randomly. While this does not ensure that there are no differences between the groups, it ensures that the differences are distributed equally, thus correcting for systematic errors.

For example, in experiments where crop yield is affected (e.g. soil fertility), the experiment can be controlled by assigning the treatments to randomly selected plots of land. This mitigates the effect of variations in soil composition on the yield.

Average treatment effect

The average treatment effect (ATE) is a measure used to compare treatments (or interventions) in randomized experiments, evaluation of policy interventions, and medical trials. The ATE measures the difference in mean (average) outcomes between units assigned to the treatment and units assigned to the control. In a randomized trial (i.e., an experimental study), the average treatment effect can be estimated from a sample using a comparison in mean outcomes for treated and untreated units. However, the ATE is generally understood as a causal parameter (i.e., an estimate or property of a population) that a researcher desires to know, defined without reference to the study design or estimation procedure. Both observational studies and experimental study designs with random assignment may enable one to estimate an ATE in a variety of ways.

Design of experiments

The design of experiments (DOE, DOX, or experimental design) is the design of any task that aims to describe or explain the variation of information under conditions that are hypothesized to reflect the variation. The term is generally associated with experiments in which the design introduces conditions that directly affect the variation, but may also refer to the design of quasi-experiments, in which natural conditions that influence the variation are selected for observation.

In its simplest form, an experiment aims at predicting the outcome by introducing a change of the preconditions, which is represented by one or more independent variables, also referred to as "input variables" or "predictor variables." The change in one or more independent variables is generally hypothesized

to result in a change in one or more dependent variables, also referred to as "output variables" or "response variables." The experimental design may also identify control variables that must be held constant to prevent external factors from affecting the results.

Experimental design involves not only the selection of suitable independent, dependent, and control variables, but planning the delivery of the experiment under statistically optimal conditions given the constraints of available resources. There are multiple approaches for determining the set of design points (unique combinations of the settings of the independent variables) to be used in the experiment.

Fisher's principles

A methodology for designing experiments was proposed by Ronald Fisher, in his innovative books: The Arrangement of Field Experiments (1926) and The Design of Experiments (1935). Much of his pioneering work dealt with agricultural applications of statistical methods. As a mundane example, he described how to test the lady tasting tea hypothesis, that a certain lady could distinguish by flavour alone whether the milk or the tea was first placed in the cup. These methods have been broadly adapted in the physical and social sciences, are still used in agricultural engineering and differ from the design and analysis of computer experiments.

1. Comparison

In some fields of study it is not possible to have independent measurements to a traceable metrology standard. Comparisons between treatments are much more valuable and are usually preferable, and often compared against a scientific control or traditional treatment that acts as baseline.

2. Randomization

Random assignment is the process of assigning individuals at random to groups or to different groups in an experiment, so that each individual of the population has the same chance of becoming a participant in the study. The random assignment of individuals to groups (or conditions within a group) distinguishes a rigorous, "true" experiment from an observational study or "quasi-experiment". There is an extensive body of mathematical theory that explores the consequences of making the allocation of units to treatments by means of some random mechanism (such as tables of random numbers, or the use of randomization devices such as playing cards or dice). Assigning units to treatments at random tends to mitigate confounding, which makes effects due to factors other than the treatment to appear to result from the treatment.

The risks associated with random allocation (such as having a serious imbalance in a key characteristic between a treatment group and a control group) are calculable and hence can be managed down to an acceptable level by using enough experimental

units. However, if the population is divided into several subpopulations that somehow differ, and the research requires each subpopulation to be equal in size, stratified sampling can be used. In that way, the units in each subpopulation are randomized, but not the whole sample. The results of an experiment can be generalized reliably from the experimental units to a larger statistical population of units only if the experimental units are a random sample from the larger population; the probable error of such an extrapolation depends on the sample size, among other things.

3.Statistical replication

Measurements are usually subject to variation and measurement uncertainty; thus they are repeated and full experiments are replicated to help identify the sources of variation, to better estimate the true effects of treatments, to further strengthen the experiment's reliability and validity, and to add to the existing knowledge of the topic. However, certain conditions must be met before the replication of the experiment is commenced: the original research question has been published in a peer-reviewed journal or widely cited, the researcher is independent of the original experiment, the researcher must first try to replicate the original findings using the original data, and the write-up should state that the study conducted is a replication study that tried to follow the original study as strictly as possible.

4.Blocking

Blocking is the non-random arrangement of experimental units into groups (blocks/lots) consisting of units that are similar to one another. Blocking reduces known but irrelevant sources of variation between units and thus allows greater precision in the estimation of the source of variation under study.

5.Orthogonality

Orthogonality concerns the forms of comparison (contrasts) that can be legitimately and efficiently carried out. Contrasts can be represented by vectors and sets of orthogonal contrasts are uncorrelated and independently distributed if the data are normal. Because of this independence, each orthogonal treatment provides different information to the others. If there are T treatments and T – 1 orthogonal contrasts, all the information that can be captured from the experiment is obtainable from the set of contrasts.