data are met. Furthermore, the F test assumes that no interaction exists among the two blocking variables and the independent variable (also the treatment). To test this interaction assumption, a Latin square design must have two or more observations per cell (i.e., $n \geq 2$). Otherwise, this assumption of no interaction is assumed for data.

If treatments of the independent variable included in a study (e.g., contextualized and abstract) are regarded as fixed and the two blocking variables are regarded likewise, the design is a fixed-effects design. The statistical model prescribed for a fixed-effects design is Model I (Kirk, 1995). If, however, one of the blocking variables is regarded as a random factor, the corresponding design is a mixed-effects design and the model is Model III. If treatments included in a study represent only a sample of all treatments, and hence the corresponding independent variable is a random factor, the design is a random-effects design and Model II is an appropriate statistical model for the data. For additional readings on Latin square designs, readers are encouraged to consult the entry titled "Latin Square Design," Kirk (1995), or Maxwell and Delaney (1990).

For the three research designs and the corresponding statistical models presented above, it is assumed that data are collected and measured without errors. If measurement errors are present in the data, RELIABILITY needs to be taken into consideration when the F statistic is computed (Cleary & Linn, 1969; Levin & Subkoviak, 1977). Furthermore, the VALIDITY of any causal relationship established between independent variables and one or more dependent variables in an experiment can be threatened by a number of factors. For details on threats to external, internal, and statistical conclusion validities, readers are encouraged to consult Campbell and Stanley (1966) and Kirk (1995).

Within the field of probability theory, an EXPERIMENT is a process or procedure that leads to a single outcome whose probability of occurrence is to be determined. An experiment can be a laboratory study (e.g., engaging students in collaborative learning to improve their social skills) or a process of observing an aspect of behavior in a sample taken from a population (e.g., noting the birth months of two randomly chosen guests at a dinner party of 50). Both the experiment and the outcome (the improvement of social skills or two guests born in the same month) should be well defined such that a researcher could formulate rules for determining the probability (likelihood) of obtaining the specific outcome, among all possible outcomes. The approach to probability determination may be based on either (a) the subjective-personalistic framework, (b) the logical or classical framework, or (c) the empirical-relative or frequency framework. In this context, an experiment may also be called a simple experiment.

—Chao-Ying Joanne Peng

See also BLOCK DESIGN, FIELD EXPERIMENTATION, LATIN SQUARE, QUASI-EXPERIMENT, RELIABILITY, VALIDITY

REFERENCES

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally College Publishing.

Cleary, T. A., & Linn, R. L. (1969). Error of measurement and the power of a statistical test. *The British Journal of Mathematical and Statistical Psychology, 22*(1), 49–55.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally College Publishing.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715–730.

Fisher, R. A. (1971). *Design of experiments.* New York: Hafner Press. (Original work published 1935.)

Glass, G. V, Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the analysis of variance and covariance. *Review of Educational Research, 42,* 237–288.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole.

Levin, J. R., & Subkoviak, M. J. (1977). Planning an experiment in the company of measurement error. *Applied Psychological Measurement, 1*(3), 331–338.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective.* Belmont, CA: Wadsworth.

Vogt, W. P. (1999). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences* (2nd ed.). Thousand Oaks, CA: Sage.

EXPERIMENTAL DESIGN

CHARACTERISTIC FEATURES OF EXPERIMENTS

Two characteristics set experimentation apart from other methods of social inquiry. First, experimentation involves a planned intervention. Unlike passive observation, in which the researcher attempts to trace the effects of naturally occurring fluctuations in putative causes, experiments create a disturbance in the environment and track its consequences. For example, an observational study of welfare reform might examine whether rates of joblessness vary with over-time or cross-jurisdiction variations in program requirements. By contrast, an experimental investigation would assign program participants to different sets of requirements in order to examine whether they affected joblessness.

The second characteristic of experiments is RANDOM ASSIGNMENT. Although the term EXPERIMENT is used loosely in common parlance to refer any type of intervention, in social science it has come to refer to studies in which units of observation are assigned at random to treatment and control conditions. Because social scientists cannot create the equivalent of a physics lab, in which all extraneous causes are eliminated, they instead rely on random assignment, which, as R. A. Fisher pointed out in his classic work *The Design of Experiments* (1935), is the only *procedure* that guarantees the comparability of treatment and control groups. So long as randomization is carried out faithfully, we can be sure that treatment and control groups differ solely as a result of chance.

Randomization therefore enables researchers to make precise statistical statements about the likelihood that any postintervention differences between treatment and control groups resulted from fortuitous differences between the groups, as opposed to the intervention in question. Precise probability statements typically are unavailable in non-experimental social science, where the independent variables are generated by an unknown process. Beyond the usual statistical uncertainty that arises from limited sample size, non-experimental research involves considerable uncertainty about whether the observed correlation between two variables, X and Y, reflects the causal influence of X on Y, of Y on X, or of some third variable Z on both X and Y.

PROBLEMS CONFRONTING EXPERIMENTAL DESIGN

Although, in principle, experiments represent the most reliable means of drawing causal inferences, in practice experiments confront several problems. Problems of INTERNAL VALIDITY arise when an experiment fails to isolate the true effects of a particular causative agent. For example, the Lanarkshire milk experiment, an early study designed to test whether distributing milk in schools improved students' growth rates, was undone when teachers reassigned underweight students in the control group to receive milk supplements. Thus, the contrast between treatment and control groups reflected both the effects of milk and the vagaries of teacher reassignments.

Concerns about EXTERNAL VALIDITY arise when the environment within which an experiment takes place or the people who participate in the experiment differ in important respects from the places and populations about whom the researcher intends to generalize. This concern arises frequently when LABORATORY EXPERIMENTS attempt to simulate economic or political environments using college students as subjects. In response to concerns about external validity, social scientists have turned increasingly to FIELD EXPERIMENTS, or experiments conducted in real-world settings. Such studies have examined the effectiveness of voter mobilization campaigns, preschool education programs, methods for encouraging compliance with tax rules, and a wide array of other interventions. Nevertheless, practical constraints limit the range of field experiments. Social scientists lack the resources and authority to manipulate large-scale causative factors, such as legislative institutions or the religiosity of the population.

Related to concerns about external validity is the issue of homogeneous treatment effects. Are all subjects equally influenced by a given intervention? If the answer is yes, the range of research opportunities expands, because one may study situations in which only some of those assigned to the treatment group actually receive treatment. For example, if a voter mobilization campaign has the same effect on everyone it contacts, but it reaches only half of the people it seeks to contact, its effect is twice as strong as a naive comparison of treatment and control groups would suggest. The assumption of homogeneous treatment effects, in other words, enables us to extrapolate easily from those who were actually exposed to a treatment to those whom researchers sought to treat.

Whether treatment effects are in fact homogeneous is an empirical question, not unlike the question of whether college students are as susceptible to social influences as those outside the university. Concerns about homogeneity and external validity underscore the importance of REPLICATING experimental findings in different settings and populations.

Ethical limitations are a further constraint. Experimental interventions may adversely affect the subjects involved, as well as the broader society. Besides the practical constraints of devising feasible interventions, researchers much follow procedural safeguards to ensure that subjects are neither harmed nor coerced. These considerations generally remove from consideration far-reaching experiments involving social or economic policy, but as Donald T. Campbell (1969) pointed out, the social costs of *not* conducting experiments must also be taken into account. Governments, firms, and organizations continually intervene in the world, and the question is whether their interventions can be structured in such a way as to generate useful knowledge.

EXPERIMENTS AND SCIENTIFIC ADVANCEMENT

Despite these limitations, experiments remain the gold standard for adjudicating causal claims. A single well-crafted experiment—conducted in a real-world setting and of sufficient size to produce statistically precise conclusions—can overshadow a large body of research based on observational data.

Part of the allure of experiments is their elegant transparency. In contrast to non-experimental data analysis, the analysis of experimental results often requires little more than elementary statistical methods, and the choice of statistical techniques seldom has a material bearing on the results. The experimental design largely dictates the manner in which the data will be analyzed statistically; committing to a plan of analysis ex ante helps guard against post hoc decisions that may bias the results in a particular direction. Experimental procedures not only lead to clearer causal inferences but also free the analyst from the moral hazards of data mining.

—Donald P. Green

See also EXPERIMENT

REFERENCES

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24*, 409–429.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Boston: Houghton Mifflin.

Fisher, R. A. (1935). *The design of experiments*. New York: Hafner Publishing.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Wilmington, MA: Houghton Mifflin.

EXPERIMENTER EXPECTANCY EFFECT

The experimenter expectancy effect is one of the sources of artifact or error in scientific inquiry (see INVESTIGATOR EFFECT). Specifically, it refers to the unintended effect of experimenters' hypotheses or expectations on the results of their research.

Some expectation of how the research will turn out is virtually a constant in science. Social scientists, like other scientists generally, conduct research specifically to examine hypotheses or expectations about the nature of things. In the social and behavioral sciences, the HYPOTHESIS held by the investigator can lead him or her unintentionally to alter behavior toward the research participants in such a way as to increase the likelihood that participants will respond so as to confirm the investigator's hypothesis or expectations. We are speaking, then, of the investigator's hypothesis as a self-fulfilling prophecy: One prophesies an event, and the expectation of the event then changes the behavior of the prophet in such a way as to make the prophesied event more likely. The history of science documents the occurrence of this phenomenon with the case of the horse Clever Hans as a prime example (Pfungst, 1911/1965).

The first experiments designed specifically to investigate the effects of experimenters' expectations on the results of their research employed human research participants. Graduate students and advanced undergraduates in the field of psychology were employed to collect data from introductory psychology students. The experimenters showed a series of photographs of faces to research participants and asked participants to rate the degree of success or failure reflected in the photographs. Half the experimenters, chosen at random, were led to expect that their research participants would rate the photos as being of more successful people. The remaining half of the experimenters were given the opposite expectation—that their research participants would rate the photos as being of less successful people. Despite the fact that all experimenters were instructed to conduct a perfectly standard