Research Strategies and the Control of Nuisance Variables

In: Experimental Design: Procedures for the Behavioral Sciences

By: Roger E. Kirk
Pub. Date: 2014
Access Date: September 6, 2016
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412974455
Online ISBN: 9781483384733
DOI: http://dx.doi.org/10.4135/9781483384733
Print pages: 1-29

Research Strategies and the Control of Nuisance Variables

1.1 Introduction

Sir Ronald Fisher, the statistician, eugenicist, evolutionary biologist, geneticist, and father of modern experimental design, observed that experiments are “only experience carefully planned in advance, and designed to form a secure basis of new knowledge” (Fisher, 1935a, p. 8). The design of experiments to investigate scientific or research hypotheses involves a number of interrelated activities:

1. Formulation of statistical hypotheses that are germane to the scientific hypothesis. A **statistical hypothesis** is a statement about (a) one or more parameters of a population or (b) the functional form of a population. Statistical hypotheses are rarely identical to scientific hypotheses; they are testable formulations of scientific hypotheses.
2. Determination of the experimental conditions (independent variable) to be used, the measurement (dependent variable) to be recorded, and the extraneous conditions (nuisance variables) that must be controlled.
3. Specification of the number of subjects (experimental units) required and the population from which they will be sampled.[1]
4. Specification of the procedure for assigning the subjects to the experimental conditions.
5. Determination of the statistical analysis that will be performed.

In short, an **experimental design** identifies the independent, dependent, and nuisance variables and indicates the way in which the randomization and statistical aspects of an experiment are to be carried out.

Subject Matter and General Organization of This Book

Experimental design, the subject of this book, refers to a plan for assigning subjects to experimental conditions and the statistical analysis associated with the plan. Selecting an appropriate plan and performing the correct statistical analysis are important facets of scientific research. However, the most important facet—identifying relevant research questions—is outside the scope of this book. The reader should remember that a carefully conceived and executed design is of no value if the scientific hypothesis that led to the experiment is without merit. Careful planning should always precede the data collection phase of an experiment. Data collection is usually the most costly and time-consuming aspect of an experiment. Advanced planning helps to ensure that the data can be used to maximum advantage. No amount of statistical wizardry can salvage a badly designed experiment.

Chapters 1 to 3 provide an overview of important design concepts and analysis tools that are used throughout the remainder of the book. Chapter 3 describes a procedure developed by Ronald A. Fisher called the **analysis of variance.** The procedure is used to decompose the total variation displayed by a set of observations into two or more identifiable sources of variation. Analysis of variance enables researchers to interpret the variability in designed experiments. Fisher showed that by comparing the variability among subjects treated differently to the variability among subjects treated alike, researchers can make informed choices between competing hypotheses in science and technology. A detailed examination of each analysis of variance design begins in Chapter 4. This examination includes a description of the design, conditions under which the design is appropriate, assumptions associated with the design, a computational example, and advantages and disadvantages of the design.

Two kinds of computational algorithms are provided for the designs. The first, referred to as the *classical sum-of-squares approach*, uses scalar algebra and is suitable for calculators. The second, called the *cell means model approach*, uses matrix algebra and is more suitable for computers.[2] In Chapters 7 and 13, I provide a brief description of a third computational algorithm: the *regression model approach*.

1.2 Formulation of Plans for the Collection and Analysis of Data

Acceptable Research Hypotheses

Some questions currently cannot be subjected to scientific investigation. For example, the questions “Can three or more angels dance on the head of a pin?” and “Does life exist in more than one galaxy in the universe?” cannot be answered because no procedures now exist for observing either angels or life on other galaxies. Scientists confine their research hypotheses to questions that can be answered by procedures that are available or that can be developed. Thus, the question concerning the existence of life on other galaxies currently cannot be investigated, but with continuing advances in space science, it is likely that eventually the question will be answered.

An experiment involves the manipulation of one or more variables by a researcher to determine the effect of this manipulation on another variable. Questions that provide the impetus for experimental research should be reducible to the form, *if A*, *then B*. For example, *if* albino rats are exposed to microwave radiation, *then* their food consumption will decrease. This research hypothesis can be investigated because procedures are available both for manipulating the radiation level and for measuring the food consumption of rats.

Much research departs from this pattern because nature rather than the researcher manipulates the independent variable. It would be unethical, for example, to study the effects of prenatal malnutrition on IQ by deliberately providing pregnant women with inadequate diets. Instead, the question is investigated by locating children whose mothers were malnourished during pregnancy and then comparing their IQs with those of children whose mothers were not malnourished. Research strategies in which the independent variable is not manipulated by the researcher include surveys, case studies, and naturalistic observation. These research strategies pose special problems for researchers who want to make causal inferences, as I discuss in Section 1.3.

Distinction Between Independent and Dependent Variables

In the radiation example cited earlier, the presence or absence of radiation is the independent variable—the variable that is manipulated by the researcher. More generally, an **independent variable** is any suspected causal event that is under investigation. The terms *independent variable* and *treatment* are interchangeable. A **dependent variable** is the measurement that is used to assess the effects, if any, of manipulating the independent variable. In the radiation example, the dependent variable is the amount of food consumed by the rats.

Selection of the Independent Variable

The independent variable in the radiation example is the presence or absence of radiation. The treatment has two levels. If the researcher is interested in the nature of the relationship between the radiation dose and food consumption, three or more levels of radiation must be used. The levels could be 0 microwatts, 20,000 microwatts, 40,000 microwatts, and 60,000 microwatts of radiation. This treatment is an example of a **quantitative independent variable** in which different treatment levels are different amounts of the independent variable.

When the independent variable is quantitative, the levels of the variable are generally chosen so that they are equally spaced. Usually there is little interest in the exact values of the treatment levels used in the experiment. In the radiation example, the research hypothesis also could be investigated using three other levels of radiation—say, 25,000, 50,000, and 75,000 microwatts in addition to the 0 microwatt control level. The treatment levels should cover a sufficiently wide range so that the effects of the independent variable can be detected if such effects exist. In addition, the number and spacing of the levels should be sufficient to define the shape of the function that relates the independent and dependent variables. Selection of the appropriate levels of the independent variable can be based on the results of previous experiments or on theoretical considerations. It may be beneficial to carry out a small pilot experiment to identify the most appropriate treatment levels. A pilot experiment also is useful for determining the number of experimental units required to test the statistical hypothesis.

Under the conditions described in Chapters 3 and 4, the levels of a quantitative independent variable can be selected randomly from a population of treatment levels. If this procedure is followed, a researcher can extrapolate from the results of the experiment to treatment levels that are not included in the experiment. If the treatment levels are not randomly sampled, the results of an experiment apply only to the specific levels included in the experiment.

Often a different type of independent variable is used. For example, if the treatment levels are unmodulated radiation, amplitude-modulated radiation, and pulse-modulated radiation, the treatment is called a **qualitative independent variable.** The different treatment levels represent different *kinds* rather than different *amounts* of the independent variable. The particular levels of a qualitative independent variable used in an experiment are generally of specific interest to a researcher. And the levels chosen are usually dictated by the research hypothesis.

Selection of the Dependent Variable

The choice of an appropriate dependent variable may be based on theoretical considerations, although in many investigations, the choice is determined by practical considerations. In the radiation example, other dependent variables that could be measured include the following:

1. Activity level of the rats in an activity cage
2. Body temperature of the rats
3. Emotionality of the rats as evidenced by their amount of defecation and urination
4. Problem-solving ability
5. Weight change of the rats
6. Speed of running in a straight-alley maze
7. Visual discrimination capacity

Several independent variables can be used in an experiment, but the designs described in this book are limited to the assessment of one dependent variable at a time. If it is necessary to evaluate two or more dependent variables simultaneously, a multivariate analysis of variance design can be used.[3] The selection of the most fruitful variables to measure should be determined by a consideration of the sensitivity, reliability, distribution, and practicality of the possible dependent variables. From previous experience, a researcher may know that one dependent variable is more sensitive than another to the effects of a treatment or that one variable is more reliable—that is, gives more consistent results—than another variable. Because behavioral research generally involves a sizable investment in time and material resources, the dependent variable should be reliable and maximally sensitive to the phenomenon under investigation. Choosing a dependent variable that possesses these two characteristics can minimize the amount of time and effort required to investigate a research hypothesis.

Other factors to consider in selecting a dependent variable are whether the population distributions for the various treatment levels are approximately normal and whether the populations have equal variances. I have more to say about these factors in Chapter 3 when I discuss the assumptions underlying the analysis of variance (**ANOVA**). If theoretical considerations do not dictate the selection of a dependent variable and if several alternative variables are equally sensitive and reliable, in addition to being normally distributed with equal variances, a researcher should select the variable that is most easily measured.

Nuisance Variables

In addition to independent and dependent variables, all experiments include one or more nuisance variables. **Nuisance variables** are undesired sources of variation in an experiment that affect the dependent variable. As the name implies, the effects of nuisance variables are of no interest per se. There are many potential sources of nuisance variables. For example, the calibration of equipment may change during the course of an experiment; the presentation of instructions may vary slightly from subject to subject; errors may occur in measuring or recording a subject's response; environmental factors such as room illumination, noise level, and room temperature may not be constant for all subjects; and subjects may experience lapses in attention, concentration, and interest.

In the radiation example, potential nuisance variables include the sex of the rats, differences in the weights of the rats prior to the experiment, presence of infectious diseases in one or more cages where the rats are housed, temperature variation among the cages, and differences in previous feeding experiences of the rats. If not controlled, nuisance variables can affect the outcome of an experiment. For example, if rats in the radiated groups suffer from some undetected disease, differences among the groups will reflect the effects of the disease in addition to radiation effects—if such effects exist.

The effect of a nuisance variable can take several forms. For example, a nuisance variable can systematically distort results in a particular direction, in which case the effect is called **bias.** Alternatively, a nuisance variable can increase the variability of the phenomenon being measured and thereby increase the error variance. **Error variance** is variability among observations that cannot be attributed to the effects of the independent variable. You also can think of error variance as differences in the performance of subjects who are treated alike. Sometimes a nuisance variable systematically distorts results in a particular direction and increases the error variance—the worst-case scenario.

Nuisance variables are undesired sources of variation and hence are threats to drawing valid inferences from research. Other threats to valid inference making are described in Sections 1.5 and 1.6.

1.3 Research Strategies

Research is performed for the following purposes: (1) to explore, (2) to describe or classify, (3) to establish relationships, or (4) to establish causality. Over the years, researchers have developed a variety of research strategies to accomplish these purposes. These strategies include the experiment, quasi-experiment, survey, case study, and naturalistic observation.

Experiments

An experiment enables a researcher to test a hypothesized relationship between an independent variable and a dependent variable by manipulating the independent variable. Experiments are usually performed in an environment that permits a high degree of control of nuisance variables. Such environments rarely duplicate real-life situations, but an experiment is still a useful way of obtaining knowledge. An **experiment** is characterized by the (1) manipulation by the researcher of one or more independent variables, (2) use of controls such as randomly assigning subjects or experimental units to the experimental conditions, and (3) careful observation or measurement of one or more dependent variables. The first and second characteristics—manipulation of an independent variable and the use of controls such as randomization—distinguish experiments from nonexperimental research strategies. The manipulation of one or more independent variables also is necessary for inferring causality. We infer that A causes Y if the following are true: A precedes Y (temporal precedence of A); whenever A is present, Y occurs (sufficiency of A); and A must be present for Y to occur (necessity of A).[4]

As noted earlier, this book is concerned with two aspects of experiments: the plan for assigning subjects to experimental conditions and the statistical analysis associated with the plan. Because the statistical analysis procedures for experiments also are applicable to other research strategies, I briefly describe some of these strategies next.

Quasi-Experiments

Quasi-experiments are similar to experiments except that the subjects are not randomly assigned to the independent variable. Quasi-experiments are used instead of experiments when random assignment is not possible or when, for practical or ethical reasons, it is necessary to use preexisting or naturally occurring groups such as subjects with a particular illness or subjects who have been sexually abused.[5]

An example of a well-designed quasi-experiment is the Newburgh-Kingston Caries-Fluorine Study (Hilleboe, 1956). This study was designed to determine the effect of adding fluoride to a community water supply. The cities studied, Newburgh and Kingston, New York, are located on the Hudson River about 35 miles apart. Beginning on May 2, 1945, sodium fluoride was added to the drinking water of Newburgh to bring the fluoride content from about 0.1 part per million to about 1.2 parts per million. The fluoride concentration of Kingston's water remained at about 0.1 part per million. In the year prior to adding fluoride to Newburgh's water supply, baseline data on the prevalence of tooth decay were obtained for school children aged 6 to 12. Baseline pediatric examinations also were performed on smaller samples. The baseline data in the two communities were similar for both tooth decay and general health. The effect of adding fluoride to Newburgh's water supply was evaluated 10 years later by examining more than 2000 children aged 6 to 16 in the two communities. For the 6- to 9-year-olds, the reduction in tooth decay in Newburgh relative to the rate in Kingston was 57%. For older children in Newburgh who had not used fluoridated water all their lives, the reduction was 41%. The tooth decay rate in Newburgh also was similar to that in Aurora, Illinois. Aurora has a naturally occurring fluoride level of about 1.2 parts per million—the same as that in Newburgh—and is known for its low level of tooth decay. The data from this quasi-experiment provide strong support for the efficacy of fluoridated water in preventing tooth decay.

The interpretation of the results of the Newburgh-Kingston study is relatively straightforward. The interpretation of the results of most quasi-experiments is often less straightforward because it is difficult to rule out all variables other than the independent variable as explanations for an observed difference. Researchers in the Newburgh-Kingston study attempted to rule out as many nuisance variables as possible. They chose two communities of comparable size on the Hudson River. And the communities had similar naturally occurring levels of fluoride in their water supplies. Because the communities are only 35 miles apart, they have similar climates and weather conditions. The potential variable of differences in the general health of children in the two communities was ruled out by a pediatric examination. Also, the tooth decay rate obtained with artificially fluoridated water in Newburgh was found to be similar to the rate in Aurora, which has about the same naturally occurring fluoride level.

There is always the possibility that some variable other than the higher fluoride level was responsible for the observed difference in tooth decay between Newburgh and Kingston. However, every effort, short of random assignment, was made to eliminate other variables as explanations for the observed difference. Random assignment is the best safeguard against undetected nuisance variables. As a general principle, the difficulty of unambiguously interpreting the outcome of research varies inversely with the degree of control that a researcher is able to exercise over randomization.

Surveys

Surveys rely on the technique of self-report to obtain information about such variables as people's attitudes, opinions, behaviors, and demographic characteristics. The data are collected by means of an interview or a questionnaire. Although surveys cannot establish causality, they can explore, describe, classify, and establish relationships among variables.

A survey enables a researcher to collect a considerable amount of information about a large number of people. If the survey respondents are representative of a population of interest, the results of the survey can be generalized to the population. Unfortunately, survey respondents are often not representative. For example, many people refuse to give phone interviews or respond to Internet questionnaires, and the return rate for questionnaires received in the mail is typically between 10% and 45%. In all likelihood, people who do not cooperate differ in significant ways from those who do. There are other problems with surveys. Some people tend to give socially acceptable answers or answers that they think the interviewer wants to hear. Also, people may have incomplete or inaccurate memories for past events. Despite these problems, surveys can be efficient, useful sources of information: witness the success of the U.S. census, Gallup Poll, Harris Poll, and Roper Poll.

Case Studies

In a **case study,** a researcher observes selected aspects of a subject's behavior over a period of time. The subject is usually a person, but it may be a setting such as a business, school, or neighborhood. Often the subject possesses an unusual or noteworthy condition. For example, Luria (1968) studied the Russian mnemonist Shereshevskii, who used mnemonic tricks and devices to remember phenomenal amounts of material. Significant discoveries also may result from studying less remarkable subjects. Jean Piaget's theory of intellectual development, for example, evolved from his intensive observation of his own three children. He presented tasks in a nonstandard manner to one child at a time in informal settings and observed the child's verbal and motor responses. Piaget did not attempt to systematically manipulate preselected independent variables, nor did he focus on just one or two dependent variables. Instead, his approach was quite flexible, which allowed him to alter his procedures and follow up on new hypotheses. His flexible case study approach uncovered knowledge about children's cognitive development that might not have been discovered by a more rigid experimental approach.

Case studies can lead to interesting insights that merit further investigation. However, case studies are particularly susceptible to the effects of nuisance variables. Furthermore, questions arise about the degree to which the findings generalize to other populations.

Naturalistic Observation

Naturalistic observation involves observing individuals or events in their natural setting without using manipulative interventions or measuring techniques that might intrude on the setting. Naturalistic observation is a passive form of research in the sense that the individual being observed determines the events that are available to be recorded. The researcher is an unobtrusive recorder of the ongoing events. Because a researcher can focus on only a finite number of events, decisions must be made concerning the events that will be observed. As in a case study, the researcher has the freedom to shift his or her focus to those events that seem most interesting. The data from naturalistic observations may be difficult to analyze, as when the researcher records a running description of a behavior, or easy to analyze, as when a frequency count of a behavior is made.

Naturalistic observation is one of the oldest methods for studying individuals and events. In some sciences, most notably astronomy, the strategy has led to extremely accurate predictions. Classic examples of naturalistic observation are Charles Darwin's voyages on the HMS *Beagle* as he compiled the data that led to the theory of evolution and Jane Goodall's (1971, 1986) study of chimpanzees in their natural habitat in Tanzania, which gave us a new appreciation for this highly social animal.

As a research strategy, naturalistic observation has two advantages over more controlled strategies such as the experiment. First, findings from naturalistic observations generalize readily to other real-life situations. Second, the strategy avoids the reactive arrangements problem that is described in Section 1.5. This problem is avoided because subjects are unaware that their behavior is being studied; hence, they do not react in an unnatural way as they might if they were aware that they were being studied. Unfortunately, there are some serious limitations associated with naturalistic observation. Although the strategy is useful for describing what happened, it does not yield much information about why something happened. To find out why something happened, it is necessary to tamper with the natural course of events. Also, the strategy is an inefficient way to answer “What if?” questions because the event of interest may occur infrequently or not at all in a natural setting.

In this section, I described five widely used research strategies. The strategies are presented in order of decreasing control of the independent and dependent variables. Research always involves a series of trade-offs—a theme I return to time and again. As our control of the independent and dependent variables decreases, our ability to unambiguously interpret the outcome of the research decreases, but our ability to generalize the results to the real world increases.

1.4 Other Research Strategies

The classification scheme for research strategies that I have described is widely used, but it is not exhaustive. There are numerous other ways of classifying research strategies. Each discipline tends to develop its own nomenclature and categories. This section describes some other ways of categorizing research strategies.

Ex Post Facto Studies

The term **ex post facto study** (after-the-fact study) refers to any nonexperimental research strategy in which subjects are singled out because they have already been exposed to a particular condition or because they exhibit a particular characteristic. In such studies, the researcher does not manipulate the independent variable or assign the experimental conditions to the subjects. The retrospective cohort study and the case-control study, described in the following section, are examples of ex post facto studies.

Retrospective and Prospective Studies

Retrospective and **prospective studies** are nonexperimental research strategies in which the independent and dependent variables occur before or after, respectively, the beginning of the study. Retrospective studies use historical records to look backward in time; prospective studies look forward in time. A retrospective study is particularly useful for studying the relationship between variables that occur infrequently or variables whose occurrence is difficult to predict. For example, much of our knowledge about the health effects of ionizing radiation came from studying persons exposed to the World War II bombings of Hiroshima and Nagasaki. A retrospective study also is useful when there is a long time interval between a presumed cause and effect. For example, a decade or more can pass between exposure to a carcinogen and the clinical detection of cancer.

There are two types of retrospective studies: retrospective cohort studies and case-control studies. In a **retrospective cohort study,** also called a **historical cohort study,** records are used to identify two groups of subjects: those who have and those who have not been exposed to the independent variable. Once the exposed and nonexposed groups have been identified, they are compared in terms of the frequency of occurrence of the dependent variable. Consider, for example, McMichael, Spirtas, and Kupper's (1974) study of workers in the rubber industry. Employment records were used to identify 6678 workers who were alive on January 1, 1964. The mortality experience of these workers over the following 9-year period was compared with the mortality experience of persons in the same age and gender categories in the U.S. population. The researchers found that the rubber workers had much higher death rates from cancers of the stomach, prostate, and hematopoietic tissues.

In a **case-control study,** also called a **case-referent study,** records are used to identify two groups of subjects: those who exhibit evidence of the dependent variable, called *cases*, and those who do not, called *controls*. The cases and controls are then compared in terms of their previous exposure to the independent variable. Consider, for example, the study by Clarke, Morgan, and Newman (1982), who investigated the relationship between cigarette smoking and cancer of the cervix. One hundred eighty-one women with cervical cancer (cases) and 905 women without cervical cancer (controls) were interviewed to determine their smoking histories. The researchers found that a much larger proportion of the cases than the controls had smoked cigarettes.

Neither the retrospective cohort study nor the case-control study can establish a causal relationship. However, the research strategies can suggest interesting relationships that warrant experimental investigation. In the retrospective cohort study, more than one dependent variable can be investigated, but only one independent variable can be studied at a time. In the case-control study, multiple independent variables can be investigated, but only one dependent variable can be studied at a time. Despite these and other limitations, both research strategies have been particularly useful in the health sciences.

As noted earlier, a **prospective study,** also called a **follow-up study, longitudinal study,** or **cohort study,** is a nonexperimental research strategy in which the independent and dependent variables are observed after the onset of the investigation. Subjects are classified as exposed or nonexposed based on whether they have been exposed to a naturally occurring independent variable. The exposed and nonexposed groups are then followed for a period of time, and the incidence of the dependent variable is recorded. A classic example is the Framingham Study (T. Gordon & Kannel, 1970), which attempted to identify factors related to the dependent variable of cardiovascular disease. In the study, more than 5000 persons living in Framingham, Massachusetts, who did not have clinical evidence of atherosclerotic heart disease were examined at 2-year intervals for more than 30 years. The study identified several factors, including hypertension, elevated serum cholesterol, and cigarette smoking, that were related to cardiovascular disease.

Prospective studies have advantages over retrospective studies. First, the purported cause (independent variable) clearly precedes the effect (dependent variable); second, the amount and quality of information are not limited by the availability of historical records or the recollections of subjects; and third, measures of the incidence of the dependent variable can be computed. But prospective studies have some serious limitations, too. If the dependent variable is a rare event, a prohibitively large sample may be required to find a sufficient number of subjects who develop the rare event. Also, the investigation of a chronic process using a prospective study may require years to complete. Unfortunately, lengthy studies often suffer from logistic problems such as keeping in touch with the subjects and turnover of the research staff. The distinguishing features of retrospective and prospective studies are summarized in Table 1.4-1.

Table 1.4-1 ▪ Distinguishing Features of Retrospective and Prospective Studies

	Time of Occurrence of Independent and Dependent Variables	
	Prior to Initiation of Study	After Initiation of Study
Subject Classified on Basis of Independent Variable	Retrospective cohort study (historical cohort study)	Prospective study (follow-up study, longitudinal study, cohort study)
Subject Classified on Basis of Dependent Variable	Case-control study (case-referent study)	

Longitudinal and Cross-Sectional Studies

The term **longitudinal study** refers to any research strategy in which the same individuals are observed at two or more times. Usually the time interval between observations is fairly long. For example, in the Framingham Study mentioned earlier, subjects were examined at 2-year intervals for more than 30 years in an attempt to identify factors related to cardiovascular disease.

A longitudinal study involves studying the same individuals over time. Identifying changes in individuals over time is not difficult, but identifying the cause of the changes can be a problem because it is difficult to control all nuisance variables over an extended period of time. As a result, a researcher is often faced with competing explanations for the observed changes. The longer the study, the more numerous the competing explanations. There are other problems with longitudinal studies. Over the course of a long study, subjects move, die, or decide to drop out of the study. Often the attrition rates for the groups being followed are different, which introduces another source of bias. Also, longitudinal studies tend to be more expensive and require a longer commitment of a researcher's time than cross-sectional studies, which are described next.

A **cross-sectional study** is any research strategy in which two or more cohorts are observed at the same time. As used here, a **cohort** denotes a person or group of people who have experienced a significant life event such as a birth, marriage, or illness during a given time interval—say, a calendar year or a decade. The Newburgh-Kingston Caries-Fluorine Study mentioned earlier involved several cohort comparisons: children living in Newburgh versus those living in Kingston and 6- to 9-year-olds versus older children.

Cross-sectional studies tend to be less expensive than longitudinal studies, and they provide more immediate results. Also, attrition of subjects is less likely to be a problem in cross-sectional studies. However, as mentioned earlier in discussing the Newburgh-Kingston Caries-Fluorine Study, there is always the possibility that even in a well-designed cross-sectional study, variables other than those under investigation are responsible for the observed difference in the dependent variable. As noted earlier, random assignment is the best safeguard against undetected nuisance variables.

Longitudinal-Overlapping and Time-Lag Studies

The two research strategies described in this section combine features of longitudinal and cross-sectional studies. A **longitudinal-overlapping study,** also called a **sequential study,** can be used to compress the time required to perform a longitudinal study. Suppose that a researcher wants to observe children at 2-year intervals from ages 5 through 13. A longitudinal study would require 8 years. This time can be compressed to 4 years by observing a group of 5-year-olds and a second group of 9-year-olds. The 5-year-old children are observed at ages 5, 7, and 9; the 9-year-old children are observed at ages 9, 11, and 13. Note the overlapping age: Both groups include 9-year-olds. The layout for this study is diagrammed in Figure 1.4-1, where O_1, O_2, and O_3 denote the first, second, and third observations of the children in each group, respectively. In addition to cutting the length of the study in half, this research strategy enables a researcher to compare 5- and 9-year-olds after completing the first set of observations. This comparison would be delayed for 4 years in a longitudinal study. The earlier discussion of the advantages and disadvantages of cross-sectional studies is applicable to a longitudinal-overlapping study.

Figure 1.4-1 ▪ Simplified layout for a longitudinal-overlapping study, where O_1, O_2, and O_3 denote, respectively, the first, second, and third observations (Obs.) on the children in $Group_1$ and $Group_2$.

	Subject's Age	1st Obs.	Subject's Age	2nd Obs.	Subject's Age	3rd Obs.
$Group_1$	5	O_1	7	O_2	9	O_3
$Group_2$	9	O_1	11	O_2	13	O_3

In a **time-lag study,** observations are made at two or more times but different subjects (cohorts) are measured at each time. Consider, for example, the annual administration of the Scholastic Aptitude Test to high school juniors and seniors. For a number of years, the test score means for seniors have been declining. This example of a time-lag study shares some of the characteristics of longitudinal and cross-sectional studies. The test scores are obtained at two or more times, as in a longitudinal study, but as in a cross-sectional study, different senior classes are observed at each testing period. The layout for this study is diagrammed in Figure 1.4-2, where the groups represent five senior classes that are each observed once and O_i denotes one of the i = 1, …, 5 observations.

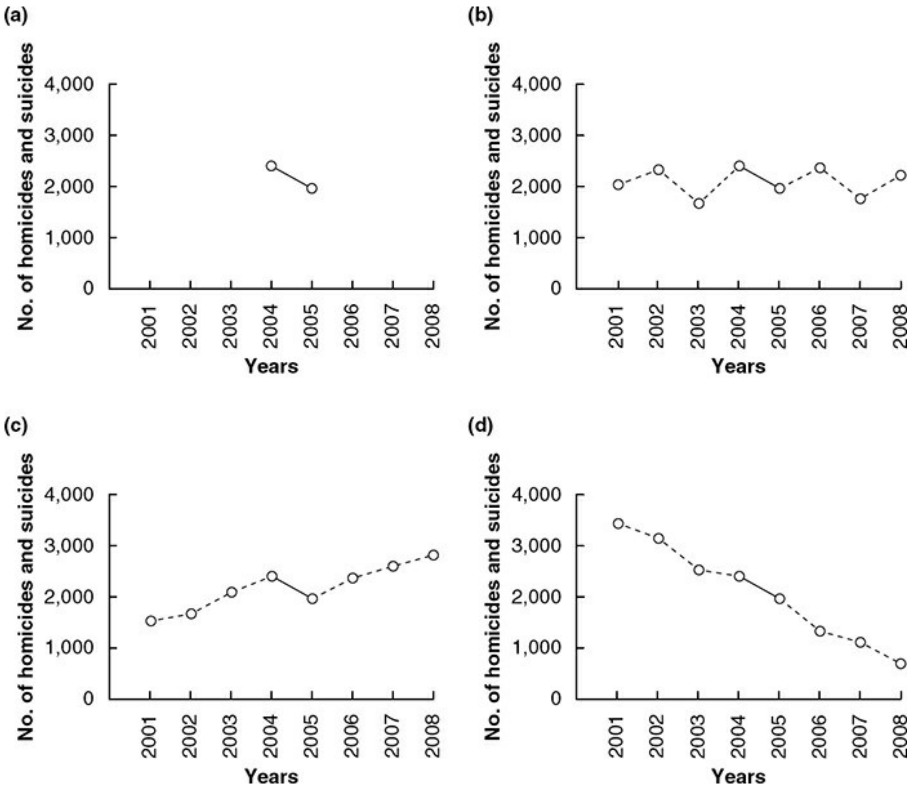
Figure 1.4-2 ▪ Simplified layout for a time-lag study, where O_1, …, O_5 denote observations (Obs.) on members of five senior classes denoted by $Group_1$ through $Group_5$.

	Year	1st Obs.	2nd Obs.	3rd Obs.	4th Obs.	5th Obs.
$Group_{1(Seniors)}$	2005	O_1				
$Group_{2(Seniors)}$	2006		O_2			
$Group_{3(Seniors)}$	2007			O_3		
$Group_{4(Seniors)}$	2008				O_4	
$Group_{5(Seniors)}$	2009					O_5

Time-Series and Single-Case Studies

A **time-series study** involves making multiple observations of one or more subjects or cohorts before and after the introduction of an independent variable. The independent variable may or may not be controlled by the researcher. Consider a study to determine the effect of banning the importation of assault rifles in 2005 on the incidence of homicides and suicides. One way to evaluate the effect of the ban is to compare the number of homicides and suicides in 2004 with the number in 2005. Suppose that the data in Figure 1.4-3(a) are obtained. Because of the reduction from 2004 to 2005, one might infer that the ban reduced the number of homicides and suicides. However, other nuisance variables such as an unusually cool summer could have been responsible for the reduction. A time-series study would provide stronger evidence for or against the effectiveness of banning the importation of assault rifles. Following this approach, a researcher would record the number of homicides and suicides for several years before and after the ban and note trends in the data. Consider the hypothetical data in Figures 1.4-3(b–d). Figure 1.4-3(b) suggests that the decrease in the number of homicides and suicides from 2004 to 2005 reflected nothing more than random year-to-year variation. Figure 1.4-3(c) suggests that the ban had only a temporary effect. Figure 1.4-3(d) suggests that the ban had no effect because similar reductions were observed during the years prior to and after the ban. These hypothetical examples illustrate the importance of obtaining multiple observations so that change can be viewed within a context.

Figure 1.4-3 ▪ Part (a) shows the decrease in the number of homicides and suicides following a ban on the importation of assault rifles in 2004. A time-series study can place the data in perspective. The hypothetical data in part (b) suggest that the decrease in the number of homicides and suicides from 2004 to 2005 reflected random year-to-year variation, part (c) suggests that the ban had a temporary effect, and part (d) suggests that the ban had no effect.



A **single-case study,** not to be confused with the case studies described in Section 1.3, has many of the characteristics of a time-series study. However, in a single-case study, multiple observations of a single subject are made before and after the introduction of an independent variable, and the researcher controls the independent variable.

Single-case studies were widely used in the behavioral sciences in the late 1880s and early 1900s. Examples include the pioneering work of Ebbinghaus (1850–1909) on forgetting, Wundt's (1832–1920) research on sensory and perceptual processes, and Titchener's (1867–1927) measurement of sensory thresholds. Researchers began to use large samples and random assignment in the 1920s and 1930s, primarily because of the influence of R. A. Fisher (1890–1962). B. F. Skinner's (1904–1990) research on schedules of reinforcement in the 1940s and 1950s rekindled an interest in single-case studies. This research strategy has proven to be particularly useful in assessing the effects of an intervention in clinical psychology research.[6]

The simplest single-case study uses an A-B design. The letter A denotes a baseline phase during which no intervention is in effect; the letter B denotes the intervention phase. The baseline phase serves three purposes: It provides data about a subject's performance prior to instituting an intervention, it provides a basis for predicting a subject's future performance in the absence of an intervention, and it indicates the normal variability in the subject's performance.

Consider an experiment to reduce the occurrence of thumb sucking of a 6-year-old named Bill. Bill usually sucked his thumb at bedtime while his mother read to him. During the baseline phase that lasted 3 days, Bill's mother read to him while an older sibling recorded the percent of story-reading time during which Bill sucked his thumb. During the treatment phase, when Bill began sucking his thumb, his mother would stop reading and remain quiet until Bill removed his thumb from his mouth. By the end of the seventh treatment day, Bill had stopped sucking his thumb when his mother read to him. The layout for this study is diagrammed in Figure 1.4-4, where O_i denotes one of the i = 1, …, n observations of the dependent variable.

Figure 1.4-4 ▪ Simplified layout for a single-case study, where O_1, O_2, …, O_i denote observations on a subject during the baseline period (A phase) and O_{i+1}, O_{i+2}, …, O_n denote observations during the treatment period (B phase). Any difference between the A and B phases in the mean of the observations or change in the trend of the observations is attributed to the intervention.

	Baseline (A Phase)	Treatment (B Phase)
Subject	$O_1, O_2, \ldots, O_i$	$O_{i+1}, O_{i+2}, \ldots, O_n$

In this example, the treatment appears to be related to the cessation of thumb sucking. But there is always the possibility that coincidental changes in a nuisance variable were completely or partly responsible for the cessation of thumb sucking. Statistical regression, which is described in Section 1.5, is a potential nuisance variable in this kind of research because the behavior that is to be altered is one that occurs frequently. Because of statistical regression, there is a tendency for the frequency of behaviors that have a high rate of occurrence to decrease in the absence of any intervention, as well as a tendency for the frequency of behaviors that have a low rate of occurrence to increase. In the thumb-sucking example, a stronger case for the efficacy of the treatment could be made if thumb sucking reappears when the treatment is withdrawn—that is, when Bill's mother continues reading even though Bill sucks his thumb. This modified design with the sequence of events

$$\text{baseline} \rightarrow \text{treatment} \rightarrow \text{baseline}$$

is diagrammed in Figure 1.4-5. Note that there are two opportunities to observe the effects of the treatment: the transition from the baseline to the treatment (A-B) and the transition from the treatment to the baseline (B-A). The presence of two transitions in the A-B-A design decreases the probability that changes in the dependent variable are the result of coincidental changes in a nuisance variable. A problem with this design is that the experiment ends on a baseline phase—a phase during which thumb sucking is expected to reappear. The solution to this problem is to reintroduce the B phase following the second A phase so that the experiment ends with the intervention phase. The design is called an A-B-A-B design and is shown in Figure 1.4-6. This design has the added advantage of providing three transitions: from A to B, from B to A, and from A to B. Hence there are three opportunities to evaluate the efficacy of the treatment. In a single-subject study, the use of one or more reversals in which a treatment is withdrawn to see whether the dependent variable returns to the baseline level can raise ethical questions. For example, if a treatment is successful in stopping an autistic child from repeatedly hitting his or her head against a wall, the withdrawal of the treatment and the subsequent return to head banging could result in physical injury to the child. In this example, the withdrawal of the treatment would be unacceptable.

Figure 1.4-5 ▪ Simplified layout for a single-case study, where O_1, O_2, …, O_n denote observations on a subject during a sequence of A-B-A phases.

	Baseline (A Phase)	Treatment (B Phase)	Baseline (A Phase)
Subject	$O_1, O_2, \ldots, O_i$	$O_{i+1}, O_{i+2}, \ldots, O_{i'}$	$O_{i'+1}, O_{i'+2}, \ldots, O_n$

Figure 1.4-6 ▪ Simplified layout for a single-case study, where O_1, O_2, …, O_n denote observations on a subject during a sequence of A-B-A-B phases.

	Baseline (A Phase)	Treatment (B Phase)	Baseline (A Phase)	Treatment (B Phase)
Subject	$O_1, O_2, \ldots, O_i$	$O_{i+1}, O_{i+2}, \ldots, O_{i'}$	$O_{i'+1}, O_{i'+2}, \ldots, O_{i''}$	$O_{i''+1}, O_{i''+2}, \ldots, O_n$

I have described a variety of research strategies in Section 1.3 and this section. In the next two sections, I briefly examine some threats to drawing valid inferences from research. In Section 1.7, I describe some general approaches to controlling nuisance variables and minimizing threats to valid inference making.

1.5 Threats to Valid Inference Making

Two goals of research are to draw valid conclusions about the effects of an independent variable and to make valid generalizations to populations and settings of interest. Shadish, Cook, and Campbell (2002), drawing on the earlier work of Campbell and Stanley (1966), have identified four categories of threats to these goals:[7]

1. **Statistical conclusion validity** is concerned with threats to valid inference making that result from random error and the ill-advised selection of statistical procedures.
2. **Internal validity** is concerned with correctly concluding that an independent variable is, in fact, responsible for variation in the dependent variable.
3. **Construct validity** of causes and effects is concerned with the possibility that operations that are meant to represent the manipulation of a particular independent variable can be construed in terms of other variables.
4. **External validity** is concerned with the generalizability of research findings to and across populations of subjects and settings.

This book is concerned with three of the threats to valid inference making: threats to statistical conclusion validity, internal validity, and external validity. In the discussion that follows, I focus on the three threats. The reader is encouraged to consult the original sources: Campbell and Stanley (1966) and Shadish et al. (2002). The latter book should be read by all researchers who, for one reason or another, are unable to randomly assign subjects to treatment conditions.

Threats to Statistical Conclusion Validity

1. **Low statistical power.** A researcher may fail to reject a false null hypothesis because the sample size is inadequate, irrelevant sources of variation are not controlled or isolated, or inefficient test statistics are used.
2. **Violated assumptions of statistical tests.** Test statistics require the tenability of certain assumptions. If these assumptions are not met, incorrect inferences may result. This threat is discussed in Section 3.5.
3. **Fishing for significant results and the error rate problem.** With certain test statistics, the probability of drawing one or more erroneous conclusions increases as a function of the number of tests performed. This threat to valid inference making is discussed in detail in Chapter 5.
4. **Reliability of measures.** The use of a dependent variable that has low reliability may inflate the estimate of the error variance and result in not rejecting a false null hypothesis.
5. **Reliability of treatment implementation.** Failure to standardize the administration of treatment levels may inflate the estimate of the error variance and result in not rejecting a false null hypothesis.
6. **Random irrelevancies in the experimental setting.** Variation in the environment (physical, social, etc.) in which a treatment level is administered may inflate the estimate of the error variance and result in not rejecting a false null hypothesis.

7. **Random heterogeneity of respondents.** Idiosyncratic characteristics of the subjects may inflate the estimate of the error variance and result in not rejecting a false null hypothesis.

Threats to Internal Validity

1. **History.** Events other than the administration of a treatment level that occur between the time a subject is assigned to the treatment level and the time the dependent variable is measured may affect the dependent variable.
2. **Maturation.** Processes not related to the administration of a treatment level that occur within subjects simply as a function of the passage of time (growing older, stronger, larger, more experienced, etc.) may affect the dependent variable.
3. **Testing.** Repeated testing of subjects may result in familiarity with the testing situation or acquisition of information that can affect the dependent variable.
4. **Instrumentation.** Changes in the calibration of a measuring instrument, shifts in the criteria used by observers and scorers, or unequal intervals in different ranges of a measuring instrument can affect the measurement of the dependent variable.
5. **Statistical regression.** When the measurement of the dependent variable is not perfectly reliable, there is a tendency for extreme scores to regress or move toward the mean. **Statistical regression** operates to (a) increase the scores of subjects originally found to score low on a test, (b) decrease the scores of subjects originally found to score high on a test, and (c) not affect the scores of subjects at the mean of the test. The amount of statistical regression is inversely related to the reliability of the test.
6. **Selection.** Differences among the dependent-variable means may reflect prior differences among the subjects assigned to the various levels of the independent variable.
7. **Mortality.** The loss of subjects in the various treatment conditions may alter the distribution of subject characteristics across the treatment groups.
8. **Interactions with selection.** Some of the foregoing threats to internal validity may interact with selection to produce effects that are confounded with or indistinguishable from treatment effects. Among these are selection history effects and selection maturation effects. For example, selection maturation effects occur when subjects with different maturation schedules are assigned to different treatment levels.
9. **Ambiguity about the direction of causal influence.** In some types of research—for example, correlational studies—it may be difficult to determine whether *X* is responsible for the change in *Y* or vice versa. This ambiguity is not present when *X* is known to occur before *Y*.
10. **Diffusion or imitation of treatments.** Sometimes the independent variable involves information that is selectively presented to subjects in the various treatment levels. If the

subjects in different levels can communicate with one another, differences among the treatment levels may be compromised.

11. **Compensatory rivalry by respondents receiving less desirable treatments.** When subjects in some treatment levels receive goods or services generally believed to be desirable and this becomes known to subjects in treatment levels that do not receive those goods and services, social competition may motivate the subjects in the latter group, the control subjects, to attempt to reverse or reduce the anticipated effects of the desirable treatment levels. Saretsky (1972) named this the "John Henry" effect in honor of the steel driver who, upon learning that his output was being compared with that of a steam drill, worked so hard that he outperformed the drill and died of overexertion.
12. **Resentful demoralization of respondents receiving less desirable treatments.** If subjects learn that the treatment level to which they have been assigned received less desirable goods or services, they may experience feelings of resentment and demoralization. Their response may be to perform at an abnormally low level, thereby increasing the magnitude of the difference between their performance and that of subjects assigned to the desirable treatment level.

Threats to External Validity

1. **Interaction of testing and treatment.** Results obtained under conditions of repeated testing may not generalize to situations that do not involve repeated testing. A pretest, for example, may sensitize subjects to a topic and, by focusing attention on the topic, enhance the effectiveness of a treatment. The opposite effect also can occur. A pretest may diminish subjects' sensitivity to a topic and thereby reduce the effectiveness of a treatment.
2. **Interaction of selection and treatment.** The constellation of factors that affect the availability of subjects to participate in an experiment may restrict the generalizability of results to populations that share the same constellation of factors. For example, if volunteers were used in an experiment, the results may generalize to only volunteer populations.
3. **Interaction of setting and treatment.** The unique characteristics of the setting in which results are obtained may restrict the generalizability of the results to settings that share the same characteristics. Results obtained in a classroom, for example, may not generalize to an assembly line.
4. **Interaction of history and treatment.** Occasionally results are obtained on the same day as a particularly noteworthy event. These results may be different from results that would have been obtained in the absence of the noteworthy event.
5. **Reactive arrangements.** Subjects who are aware that they are being observed may behave

differently than subjects who are not aware that they are being observed.

6. **Multiple-treatment interference.** When subjects are exposed to more than one treatment, the results may generalize to only populations that have been exposed to the same combination of treatments.

1.6 Other Threats to Valid Inference Making

Experimenter-Expectancy Effect

Controlling nuisance variables in research with human subjects is particularly challenging. Experiments with human subjects are social situations in which one person behaves under the scrutiny of another. The two people in this social situation have expectations about each other, communicate with each other, and form impressions about each other. The power of the subjects in the situation is always unequal: The researcher requests a behavior and the subject behaves. The researcher's overt request may be accompanied by other more subtle requests and messages. For example, body language, tone of voice, and facial expressions can communicate the researcher's expectations and desires concerning the outcome of an experiment. Such communications can affect a subject's performance. Rosenthal (1963) has reported that researchers tend to obtain from their subjects—whether human or animal—the data they want or expect to obtain. A researcher's expectations and desires also can influence the way he or she records, analyzes, and interprets data. According to Rosenthal (1969, 1978), observational or recording errors are usually small and unintentional. However, when such errors occur, more often than not they are in the direction of supporting the researcher's hypothesis. Sheridan (1976) has reported that researchers are much more likely to recompute and double-check results that conflict with their hypotheses than results that support their hypotheses. The effect of a researcher's expectations and desires on the outcome of an experiment is called the **experimenter-expectancy effect.**

Demand Characteristics

The experimenter-expectancy effect is one source of bias in an experiment; another source is what Orne (1962) has called demand characteristics. **Demand characteristics** refer to any aspect of the experimental environment or procedure that leads a subject to make inferences about the purpose of an experiment and to respond in accordance with—or in some cases, contrary to—the perceived purpose. Subjects are inveterate problem solvers. When they are told to perform a task, the majority will try to figure out what is expected of them and perform accordingly. Demand characteristics can result from rumors about an experiment, what subjects are told when they sign up for an experiment, the laboratory environment, or the communication that occurs during the course of an experiment. Demand characteristics influence a subject's perceptions of what is appropriate or expected.

Subject-Predisposition Effects

As I have discussed, an experimenter's expectations and motives can influence a subject's performance, and subjects often respond in ways that they think are appropriate or expected by the researcher. There is another source of bias in an experiment. Subjects, because of past experience, personality, and so on, come to experiments with a predisposition to respond in a particular way. I describe four kinds of subject-predisposition effects.

Cooperative-subject effect. The first predisposition is that of the cooperative subject whose main concern is to please the researcher and be a "good subject." Cooperative subjects are particularly susceptible to the experimenter-expectancy effect. They try, consciously or unconsciously, to provide data that support the researcher's hypothesis. This subject predisposition is called the **cooperative-subject effect.**

Screw you effect. A second group of subjects tends to be uncooperative and may even try to sabotage the experiment. Masling (1966) has called this predisposition the "**screw you effect.**" It can result from resentment over being required to participate in an experiment, from a bad experience in a previous experiment such as being deceived or made to feel inadequate, or from a dislike for the course or the professor associated with the experiment. Uncooperative subjects may try, consciously or unconsciously, to provide data that do not support the researcher's hypothesis.

Evaluation apprehension. A third group of subjects are apprehensive about being evaluated. Subjects with **evaluation apprehension** (Rosenberg, 1965) aren't interested in the experimenter's hypothesis, much less in sabotaging the experiment. Instead, their primary concern is to gain a positive evaluation from the researcher. The data they provide are colored by a desire to appear intelligent, well adjusted, and so on and to avoid revealing characteristics that they consider undesirable.

Faithful subjects. A fourth group of subjects have been labeled **faithful subjects** (Fillenbaum, 1966). Faithful subjects try to put aside their own hypotheses about the purpose of an experiment and to follow the researcher's instructions to the letter. Often they are motivated by a desire to advance scientific knowledge. The data produced by overly cooperative or uncooperative subjects or by subjects with evaluation apprehension can cause a researcher to draw a wrong conclusion. The data of faithful subjects, however, are not contaminated by such predispositions; faithful subjects simply try to do exactly what they are told to do.

Placebo Effect

The last source of bias that I describe is the placebo effect. A **placebo** is an inert substance or neutral stimulus that is administered to subjects as if it was the actual treatment condition. When subjects begin an experiment, they are not entirely naive. They have ideas, understandings, and perhaps a few misunderstandings about what will happen. If subjects expect that an experimental condition will have a particular effect, they are likely to behave in a manner consistent with their expectation. For example, subjects who believe that a medication will relieve a particular symptom may report feeling better even though they have received a chemically inert substance instead of the medication. Any change in the dependent variable attributable to receiving a placebo is called the **placebo effect.**

In the previous sections, I described a variety of threats to valid inference making: threats to statistical conclusion validity, internal validity, and external validity; the experimenter-expectancy effect; demand characteristics; subject-predisposition effects; and the placebo effect. This list of threats is far from complete. For a fuller discussion of threats to valid inference making, the reader should consult Shadish et al. (2002) and Rosenthal (1979). In the following section, I describe some procedures for controlling nuisance variables and minimizing threats to valid inference making.

1.7 Controlling Nuisance Variables and Minimizing Threats to Valid Inference Making

General Approaches to Control

Four general approaches are used to control nuisance variables. One approach is to hold the nuisance variable constant for all subjects. Examples are using only male rats of the same weight and presenting all instructions to subjects by means of an iPad, computer, or DVD player. Although a researcher may attempt to hold all nuisance variables constant, inevitably some variable will escape attention.

A second approach—one that is used in conjunction with the first—is to assign subjects randomly to the experimental conditions. Then known as well as unsuspected sources of variation are distributed over the entire experiment and thus do not selectively affect just one or a limited number of treatment levels. Random assignment has two other purposes. It permits the computation of an unbiased estimate of **error effects**—those effects not attributable to the manipulation of the independent variable—and it helps to ensure that the error effects are statistically independent. Through random assignment, a researcher creates two or more groups of subjects that at the time of assignment are probabilistically similar on the average. When random assignment is used, a researcher increases the magnitude of random variation among observations to minimize bias, which is the distortion of results in a particular direction. Random variation can be taken into account in evaluating the outcome of an experiment; it is more difficult to account for bias.

A third approach to controlling nuisance variables is to include the variable as one of the factors in the experimental design. This approach is illustrated in Section 2.2.

The three approaches for controlling nuisance variables illustrate the application of *experimental control* as opposed to the fourth approach, which is *statistical control*. In some experiments, it may be possible through the use of regression procedures (see Chapter 13) to statistically remove the effects of a nuisance variable. This use of statistical control is referred to as the *analysis of covariance*.

Some Specific Approaches to Control

In addition to the four general approaches just described, a variety of other procedures are used to control nuisance variables and minimize threats to valid inference making.

Single-blind procedure. In a **single-blind experiment,** subjects are not informed about the nature of their treatment or, when feasible, the purpose of the experiment. A single-blind procedure helps to minimize the effects of demand characteristics. Sometimes the purpose of an experiment cannot be withheld from subjects because of informed consent requirements that are imposed on the researcher. (Informed consent requirements are discussed in Section 1.8.)

Double-blind procedure. In a **double-blind experiment,** neither the subjects nor the researcher are informed about the nature of the treatment that the subjects receive. For example, in a drug study, the dose levels and placebo can be coded so that those administering the drug and those receiving the drug cannot identify the condition that is administered. A double-blind procedure helps to minimize experimenter-expectancy effects and demand characteristics.

Partial-blind procedure. Many treatments are of such a nature that they are easily identified by a researcher. In this case, a **partial-blind procedure** can be used in which the researcher does not know until just before administering the treatment level which level will be administered. In this way, experimenter-expectancy effects are minimized until the administration of the treatment level.

Deception. Deception occurs when subjects are not told the relevant details of an experiment or when they are told that the experiment has one purpose when in fact the purpose is really something else. Deception is used to direct a subject's attention away from the purpose of an experiment so as to minimize the effects of demand characteristics. Deception should never be used without a prior careful analysis of the ethical ramifications. (Ethical issues are discussed in Section 1.8.)

Disguised-experiment technique (unobtrusive experimentation). In the disguised-experiment technique, the subjects are not aware that they are participating in an experiment. The naturalistic-observation research strategy described in Section 1.3 is an example of this approach to minimizing the bias that might result from reactive arrangements and demand characteristics.

Multiple researchers. In some research areas, the characteristics of a researcher such as appearance, personality, inexperience, and so on can affect the results that are obtained. These researcher characteristics can seriously limit the generalizability of results. If several researchers are used, the researchers can be included as one of the variables in the experiment, and the significance of the variable can be evaluated.

Debriefing. It is a common practice to hold a postexperimental meeting with subjects at which time details of the experiment are shared. During this debriefing, subjects can be quizzed concerning their beliefs and expectations about the experiment. Information obtained at this time can be used to determine whether demand characteristics could have affected the results of the experiment.

Experimenter-expectancy control groups. The magnitude of the experimenter-expectancy effect can be determined by using several groups of researchers. One group of researchers is led to expect one experimental outcome, a second group is led to expect the opposite outcome, and a third group is led to believe that the treatment will have no effect on the dependent variable. Unfortunately, this procedure can be costly because it involves using numerous researchers and subjects.

Unrelated-experiment technique. The unrelated-experiment technique is designed to disguise the purpose of an experiment and minimize subject demand characteristics by separating the presentation of the independent variable from the measurement of the dependent variable. This technique requires subjects to participate in two experiments. In the first experiment, the subjects receive the independent variable. Later, the subjects are contacted and asked to participate in a second experiment at which time the dependent variable is measured. The researcher conveys the impression that the second experiment has no relationship to the first experiment.

Quasi-control group. This procedure uses a second control group, called a quasi-control group, to assess the effects of demand characteristics. The **quasi-control group** is exposed to all of the instructions and conditions that are given to the experimental group except that the treatment condition of interest is not administered. This group, unlike a regular control group, does not receive a placebo. Following the presentations of the instructions, the quasi-control subjects are asked to produce the data that they would have produced if they had actually received the treatment condition.

In a double-blind experiment, the quasi-control procedure can be carried one step further: Subjects can be asked to pretend that they have received the treatment condition and to behave accordingly—that is, to be simulators. At the conclusion of the experiment, the researcher is asked to identify the real subjects, control subjects, and simulators. Comparisons among the groups can be useful in detecting experimenter-expectancy effects and demand characteristics.

Yoked control procedure. Researchers would like the experiences of subjects in an experiment to be identical except for the independent variable. Unfortunately, it is difficult to keep the experiences of all subjects identical when a subject's behavior determines aspects of the experimental situation—for example, the number of shocks received in a learning experiment. A **yoked control procedure** allows a researcher to match two subjects on some important aspects of the experience they have in an experiment. In this procedure, two subjects—an active subject and a passive subject—are simultaneously exposed to the same experimental condition, but the behavior of only the active subject affects the outcome. Both members of the pair are subjected to the consequences of the active subject's behavior. For example, yoked active and passive subjects receive a shock each time the active subject makes an incorrect response, thus controlling the variable of number of shocks received.

1.8 Ethical Treatment of Subjects

In recent years, the research community has witnessed a renewed resolve to protect the rights and interests of humans and animals. Codes of ethics for research with human subjects have been adopted by a number of professional societies. Of particular interest are those of the American Educational Research Association (2011), American Evaluation Association (2008), American Psychological Association (2002), American Sociological Association (1999), and American Statistical Association (1999). These codes specify what is required and what is forbidden. In addition, they point out the ideal practices of the profession as well as ethical pitfalls. The 1970s saw the passage of laws to govern the conduct of research with human subjects. One law, which was originally enforced by the U.S. Department of Health, Education, and Welfare (HEW), now the Department of Health and Human Services (HHS), requires that all research funded by HHS involving human subjects be reviewed by an institutional review board (Weinberger, 1974, 1975). As a result, most institutions that conduct research have human subjects committees that screen all research proposals. These committees can disapprove research proposals or require additional safeguards for the welfare of subjects.

In addition to codes of ethics of professional societies, legal statutes, and peer review, perhaps the most important regulatory force within society is the individual researcher's ethical code. Researchers should be familiar with the codes of ethics and statutes relevant to their research areas and incorporate them into their personal codes of ethics.

Space does not permit an extensive examination of ethical issues here. For this the reader should consult the references above and the thorough and balanced treatment by Diener and Crandall (1978). However, I cannot leave the subject without listing some general guidelines.

1. A researcher should be knowledgeable about issues of ethics and values, take these into account in making research decisions, and accept responsibility for decisions and actions that have been taken. The researcher also is responsible for the ethical behavior of collaborators, assistants, and employees who have parallel obligations.
2. Subjects should be informed of aspects of research that might be expected to influence their willingness to participate. Failure to make full disclosure places an added responsibility on the researcher to protect the welfare and dignity of the subject. Subjects should understand that they have the right to decline to participate in an experiment and to withdraw at any time; pressure should not be used to gain cooperation.
3. Research subjects should be protected from physical and mental discomfort, harm, and danger. If risk of such consequences exists, a researcher must inform the subject of this. If harm does befall a subject, the researcher has an obligation to remove or correct the consequences.
4. Special care should be taken to protect the rights and interests of less powerful subjects such as children, minorities, patients, the poor, and prisoners.
5. Research deception should never be used without a prior careful ethical analysis. When the methodological requirements of a study demand concealment or deception, the researcher should take steps to ensure the subject's understanding of the reason for this action and afterward restore the quality of the relationship that existed. Where scientific or other compelling reasons require that this information be withheld, the researcher acquires a special responsibility to ensure that there are no damaging consequences for the subject.
6. Private information about subjects may be collected only with their consent. All such research information is confidential. Publication of research results should be in a form that

protects the subject's identity unless the subject agrees otherwise.

7. After data are collected, the researcher must provide the subjects with information regarding the nature of the study and relevant findings.
8. Results of research should be reported accurately and honestly, without omissions that might affect their interpretation.

A number of guides for research with animals have been published. Those engaged in such research should be familiar with the American Psychological Association's (1996) *Guidelines for Ethical Conduct in the Care and Use of Animals*.

1.9 Review Exercises[8]

1. Terms to remember:
   - a. statistical hypothesis (1.1)[9]
   - b. experimental unit (1.1)
   - c. experimental design (1.1)
   - d. analysis of variance (1.2)
   - e. independent variable (1.2)
   - f. dependent variable (1.2)
   - g. quantitative independent variable (1.2)
   - h. qualitative independent variable (1.2)
   - i. ANOVA (1.2)
   - j. nuisance variable (1.2)
   - k. bias (1.2)
   - l. error variance (1.2)
   - m. experiment (1.3)
   - n. quasi-experiment (1.3)
   - o. survey (1.3)
   - p. case study (1.3)
   - q. naturalistic observation (1.3)
   - r. ex post facto study (1.4)
   - s. retrospective and prospective study (1.4)
   - t. retrospective cohort studies (1.4)
   - u. case-control study (1.4)
   - v. longitudinal study (1.4)
   - w. cross-sectional study (1.4)
   - x. cohort (1.4)

y. longitudinal-overlapping study (1.4)
z. time-lag study (1.4)
aa. time-series study (1.4)
ab. single-case study (1.4)
ac. statistical conclusion validity (1.5)
ad. internal validity (1.5)
ae. construct validity (1.5)
af. external validity (1.5)
ag. statistical regression (1.5)
ah. demand characteristics (1.6)
ai. cooperative-subject effect (1.6)
aj. screw you effect (1.6)
ak. evaluation apprehension effect (1.6)
al. faithful subject (1.6)
am. placebo effect (1.6)
an. error effects (1.7)
ao. single-blind experiment (1.7)
ap. double-blind experiment (1.7)
aq. partial-blind procedure (1.7)
ar. quasi-control group (1.7)
as. yoked control procedure (1.7)

*2. [1.1] For each of the following, identify the experimental unit (EU) and the observational unit (OU).
*a. Fraternities at a large state university were randomly sampled and the members asked to complete several scales of the California Psychological Inventory.
*b. Cars at a roadblock were stopped at random and the occupants searched for illegal drugs.
c. Twenty students in an introductory psychology class were selected by random sampling and asked to participate in an experiment.
d. The time to run a straight-alley maze was recorded for each of five randomly sampled rats from 10 cages.
e. Telephone numbers obtained by random sampling from a directory were called and the respondents asked their political preference.

*3. [1.2] Which of the following are acceptable research hypotheses?
*a. Right-handed people tend to be taller than left-handed people.
*b. Behavior therapy is more effective than hypnosis in helping smokers kick the habit.

c. Most clairvoyant people are able to communicate with beings from outer space.
d. Rats are likely to fixate an incorrect response if it is followed by an intense noxious stimulus.

*4. [1.2] For each of the following studies, identify the (i) independent variable, (ii) dependent variable, and (iii) possible nuisance variables.

*a. Televised scenes portraying physical, cartoon, and verbal violence were shown to 20 preschool children. The facial expressions of the children were videotaped and then classified by judges.
*b. Power spectral analyses of changes in cortical electroencephalogram (EEG) were made during a 1- to 5-hour period of morphine administration in 10 female Sprague-Dawley rats.
c. The effects of four amounts of flurazepam on hypnotic suggestibility in men and women were investigated.
d. The keypecking rates of 20 female Silver King pigeons on fixed ratio reinforcement schedules of FR10, FR50, and FR100 were recorded.

*5. [1.2] For the independent variables in Exercise 4, indicate (i) which are quantitative and (ii) which are qualitative.

6. [1.3]

(a) List the ways in which experiments and quasi-experiments differ.
(b) Why wasn't the Newburgh-Kingston Caries-Fluorine Study an experiment?

7. [1.3] Describe how you would design the study described in Exercise 4a (a) as an experiment and (b) as a naturalistic observation study.

*8. [1.3–1.4] (i) Classify each of the following according to the most descriptive or definitive category in Sections 1.3 and 1.4: Use only one classification. (ii) What features of the studies prompted your classification?

*a. The effect of participation in the Boy Scouts, the independent variable, on the propensity for assuming leadership roles as an adult was investigated for a random sample of 400 men who were lifelong residents of Columbus, Ohio, and between the ages of 30 and 60. The subjects were classified as having held or not held a leadership role during the previous 5-year period. Records were then used to determine those men who had participated in the Boy Scouts.
*b. In a study of 86 lonely people, it was found that they display some of the characteristics of shy people: Lonely people disclose less personal information about themselves to opposite-sex friends than do nonlonely people, and they use inappropriate levels (too intimate or too impersonal) of self-disclosure in initial interactions.
*c. Two hundred thirty-two sixth graders took a test that measured arithmetic achievement. Two hundred of the students were matched on the basis of their achievement scores.

One member of each pair was randomly assigned to participate in a conventional arithmetic instruction program; the other member of the pair participated in an experimental arithmetic program. At the end of the semester, it was found that the arithmetic achievement scores of the students who participated in the experimental arithmetic program were higher than those of the other sample of students, $t(99) = 2.358$, $p = .020$, $g = .54$.

*d. Job performance ratings of graduates of a police academy were obtained for six classes from 2005 to 2010.

e. Cabdrivers in a large city were classified as expert drivers ($n = 14$), average drivers ($n = 33$), or poor drivers ($n = 11$), the independent variable, based on company records of their earnings for the previous 6-month period. All of the drivers were men between the ages of 26 and 45 and had driven a cab for at least 5 years. According to employment tests, expert drivers were superior to average and poor drivers in their ability to perceive large meaningful patterns and to do so with such speed that it appeared almost intuitive. Furthermore, expert drivers organized their knowledge of the city hierarchically, from large geographic areas down to smaller neighborhoods.

f. According to a national survey, the mean number of movies attended per month by 14-year-old boys in the United States is 4.8, with a standard deviation of 1.3. Juvenile court records in Houston, Texas, indicate that the corresponding statistics for a random sample of 31 boys who appeared in court are $\bar{Y} = 4.9$ and $\hat{\sigma} = 1.6$. The researcher concluded that the dispersion of the movie attendance distribution for boys who appeared in the Houston juvenile court, one of the dependent variables, was greater than that for the nation at large, $\chi^2(30) = 45.44$, $p < .05$.

g. A survey by the Centers for Disease Control and Prevention in Atlanta, Georgia, found that 27.6% of 15-year-old girls in 1999 had had premarital sex at least once. The comparable percentages for 2003 and 2008 were 53% and 77.2%, respectively.

h. The relationship between birth order and participation in dangerous sports such as hang gliding, auto racing, and boxing was investigated. Records at Florida State University were screened to obtain 50 men who were first-born, second-born, and so on and to identify their recreational activities while at the university.

i. Pediatricians in Oklahoma provided the names of 421 new mothers. The mothers' infant feeding practices were subsequently determined. Eight years later, elementary school records for 372 of these children indicated that the breast-fed babies had a higher level of performance in school than did those who had been bottle-fed.

j. Employment records were used to identify 86 men who had worked for a company in Cleveland, Ohio, that manufactured chemicals used as fire retardants. A second group of men, *n* = 89, was identified who worked for two other companies in Cleveland and had no exposure to the chemicals. Evidence of primary thyroid dysfunction was found in four of the exposed men; none of the unexposed men showed evidence of thyroid dysfunction.

*9. [1.5] Identify potential threats to internal validity for these studies.
   - *a. Exercise 8a
   - *b. Exercise 8b
   - c. Exercise 8d
   - d. Exercise 8g
   - e. Exercise 8h
   - f. Exercise 8i

*10. [1.5] Identify potential threats to external validity in these studies.
   - *a. Exercise 8c
   - *b. Exercise 8e
   - c. Exercise 8f
   - d. Exercise 8j

*11. [1.6] For the experiments in Exercise 8, indicate those for which the following are potential threats to valid inference making.
   - *a. Experimenter-expectancy effect
   - b. Demand characteristics
   - c. Subject-predisposition effects

12. [1.7] Two approaches to controlling nuisance variables and minimizing threats to valid inference making are holding the nuisance variable constant and using random assignment or random sampling. Indicate which experiments in Exercise 8 used these approaches and which approach was used.

13. [1.8] Section 1.8 lists eight general guidelines for the ethical treatment of subjects. Recognizing that all of the guidelines are important, select the five that you think are the most important and rank order them (assign 1 to the most important guideline). What do your selection and rankings reveal about your own ethical code?

[1] An **experimental unit** is that entity that is assigned to an experimental condition independently of other entities. An experimental unit may contain several observational units. For example, in an educational experiment, the experimental unit is often the classroom, but the individual students are the observational units. Administering an educational intervention to a classroom can result in nonindependence of the observational units. For a discussion of this problem, see Levin (1992).

[2]Readers who are interested only in the traditional approach to the analysis of variance can, without loss of continuity, omit the material on the cell means model.

[3]For a discussion of these designs, see R. J. Harris (2001); Lattin, Carroll, and Green (2003); Meyers, Gamst, and Guarono (2006); Stevens (2002); and Todman and Dugard (2007).

[4]Causality is a complex concept. For other definitions and views of causality, see Pearl (2000); Shadish (2010); Shadish, Cook, and Campbell (2002); Sobel (2008); and West and Thoemmes (2010).

[5]For an excellent treatment of quasi-experimental designs, see Shadish et al. (2002).

[6]Barlow, Nock, and Hersen (2009); Kazden (1982); and Morgan and Morgan (2009) provide in-depth discussions of single-case studies.

[7]The list of categories and threats to valid inference making are taken from Campbell and Stanley (1966) and Shadish et al. (2002). Responsibility for the interpretation of items in their lists is mine.

[8]Problems or portions thereof for which answers are given in Appendix F are denoted by *.

[9]The numbers in parentheses indicate the section in which the term is first described.

http://dx.doi.org/10.4135/9781483384733.n1