Responses are aggregated and summarized, and participant identity should be confidential or anonymous. Statistical procedures for reporting survey data include frequencies, percent, cross-tabulations (cross-tabs), chi-square statistic, phi coefficient, Kendall coefficient, and the gamma statistic.

Correlation

Correlation research is used to explore relationships between or among two or more variables. Correlation studies are useful for establishing predictive validity, establishing test reliability, and describing relationships. Simple correlation procedures involve ascertaining the relationship between two variables, whereas partial correlation procedures are used to control for a variable that may influence the correlation between two other variables. A multiple correlation coefficient (multiple regression) indicates the relationship between the best combination of independent variables and a single dependent variable. Canonical correlation indicates the relationship between a set of independent variables and a set of dependent variables. The kind of correlation coefficient computed depends on the type of measurement scale used and the number of variables.

Marie Kraska

See also Data Cleaning; Experimental Design; Hypothesis; Nonexperimental Design; Quasi-Experimental Design; Sample; Survey

Further Readings

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Myers, R. H. (1990). *Classical and modern regression with applications* (2nd ed.). Pacific Grove, CA: Duxbury.

Ramsey, F. L., & Schafer, D. W. (2002). *The statistical sleuth* (2nd ed.). Pacific Grove, CA: Duxbury.

Salkind, N. J. (2000). *Statistics for people who think they hate statistics*. Thousand Oaks, CA: Sage.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Wiersma, W., & Jurs, S. G. (2005). *Research methods in education: An introduction* (8th ed.). Needham Heights, MA: Allyn & Bacon.

Quasi-Experimental Design

A scientific experiment is a controlled set of observations aimed at testing whether two or more variables are causally related. William Shadish, Thomas Cook, and Donald Campbell describe two broad types of experiments: (a) randomized experiments, in which study units are randomly assigned to observational conditions; and (b) quasi-experiments, in which study units are not randomly assigned to observational conditions because of ethical or practical constraints. Although it is more difficult to draw causal inferences from quasi-experiments than from randomized experiments, careful planning of quasi-experiments can lead to designs that allow for strong causal inferences.

In order to infer a relationship between cause and effect, three requirements must be met: Cause must precede effect; cause must be related to effect; and, aside from the cause, no alternative explanation for the effect must be plausible. Randomized and quasi-experiments do not differ with respect to the first two requirements. However, with respect to the third requirement, randomized experiments have an advantage over quasi-experiments. Because study units are randomly assigned to conditions in randomized experiments, alternative explanations (e.g., confounding variables) are equally likely across these conditions and can be ruled out. But because quasi-experiments lack random assignment between conditions, alternative explanations are difficult to rule out. This entry focuses on the validity of, common designs of, and inferences drawn from quasi-experiments.

Validity

Inferences based on an experiment are only as good as the evidence that supports them. The term *validity* is used to refer to the relation between the conclusion of an inference and its supporting evidence. In experimentation, inferences (i.e., conclusions) are valid if they are plausible.

A number of conditions must be met in order to draw a valid inference based on an experiment.

These conditions fall into four categories. First, the *internal validity* of an inference refers to whether the covariation between the experimental manipulation and the experimental outcome does indeed reflect a causal relationship between the manipulation and outcome. Second, *external validity* refers to the generalizability of an inference (i.e., do the results of the experiment apply outside of the experimental setting?). Third, *statistical conclusion validity* refers to the validity of inferences about the covariation between manipulation and outcome. Fourth, *construct validity* refers to the validity of inferences about the higher order construct(s) that the experimental manipulation operationalizes.

Threats to Internal Validity

Factors that influence the nature and strength of inferences are referred to as threats to validity. Of particular relevance to quasi-experimental designs are threats to internal validity as they increase the likelihood that a plausible alternative explanation for the experimental outcome exists. Shadish and colleagues identify the following threats to internal validity:

Ambiguous temporal precedence: Lack of clarity about which variable occurred first may yield confusion about which variable is the cause and which is the effect.

Selection: Systematic differences over conditions in respondent characteristics that could also cause the observed effect.

History: Events occurring concurrently with treatment could cause the observed effect.

Maturation: Naturally occurring changes over time could be confused with a treatment effect.

Regression: When units are selected for their extreme scores, they will often have less extreme scores on other variables, an occurrence that can be confused with a treatment effect.

Attrition: Loss of respondents to treatment or to measurement can produce artificial effects if that loss is systematically correlated with conditions.

Testing: Exposure to a test can affect scores on subsequent exposures to that test, an occurrence that can be confused with a treatment effect.

Instrumentation: The nature of a measure may change over time or conditions in a way that could be confused with a treatment effect.

Additive and interactive effects of threats to internal validity: The impact of a threat can be added to that of another threat or may depend on the level of another threat.

Common Designs

Because threats to internal validity are prominent in quasi-experiments, care must be taken to ensure that the influence of these threats to validity is minimized. Shadish and colleagues discuss three principles useful in this respect: identification and study of plausible threats to internal validity; design controls that limit threats to internal validity (e.g., control groups, pretest/posttest designs); and specific hypotheses that limit the number of viable alternative explanations (e.g., predicted interactions or inclusion of *nonequivalent dependent variables*, that is, a dependent variable that is predicted not to change because of the manipulation but is expected to respond to threats to internal validity the same way as the dependent variable being studied). Four types of common quasi-experimental designs are discussed, each of which has its own advantages and disadvantages concerning threats to internal validity. This discussion and examples of quasi-experimental designs draw on the work of Shadish and colleagues.

Designs Without a Control Group

One-Group Posttest-Only Design

This is a simple design involving a posttest on participants (O_1) following a manipulation (X).

$$X \quad O_1$$

For example, suppose knowledge of the causes of sudden infant death syndrome (SIDS) is low and stable within a community. Public health officials create a media campaign to raise awareness regarding factors affecting SIDS. Following the campaign, a sample of citizens is surveyed concerning their knowledge of these factors. If the citizens are aware of factors affecting SIDS (O_1), one may infer that this is due to the media campaign (X).

Nonetheless, the one-group posttest-only design is a very weak design. It is impossible to guarantee temporal precedence, and furthermore, nearly all other threats to internal validity may apply. One way to improve this design, although it will remain weak, is to use multiple, unique posttests (O_{1A} through O_{1N}).

$$X \quad \{O_{1A} \; O_{1B} \; \ldots \; O_{1N}\}$$

Results of the several posttests can be assessed individually and compared to the hypothesized outcomes for each individual posttest and the manipulation. This decreases the likelihood of an invalid inference based on just a single prediction.

One-Group Pretest–Posttest Design

Instead of having a single observation, as in the previous design, the one-group pretest–posttest design has a pretest measure (O_1) before manipulation (X) as well as a posttest measure (O_2) following treatment.

$$O_1 \quad X \quad O_2$$

For example, Jonathan Duckhart studied the effects of a program to reduce environmental lead in low-income urban housing units in Baltimore. Lead levels in each home were measured at pretest and following the intervention. Lead levels decreased between pretest and posttest, supporting the conclusion that the program was effective.

Because of the pre- and posttest, temporal precedence is more easily established, although the effect could still have been caused by history or maturation. Possible improvements involve using a double pretest,

$$O_1 \quad O_2 \quad X \quad O_3$$

or a nonequivalent dependent variable.

$$\{O_{1A} \quad O_{1B}\} \quad X \quad \{O_{2A} \quad O_{2B}\}$$

A double pretest can give an estimate of biases that may exist in the observed effect of the manipulation at posttest. Differences between the posttest and the second pretest that are similar to differences between the two pretests are likely due to factors other than the manipulation. Using a nonequivalent dependent variable reduces many threats to internal validity because any changes from pretest to posttest not due to the manipulation should affect the original dependent variable as well as the nonequivalent dependent variable.

Control Group Designs

Nonequivalent Groups Posttest-Only Design

The nonequivalent groups posttest-only design is similar to the one-group posttest-only design with a control group added (where NR indicates nonrandom assignment).

$$\begin{array}{ccc} NR & X & O_1 \\ \hline NR & & O_2 \end{array}$$

For example, suppose psychotherapy researchers obtained two samples of individuals with depression: those who have attended psychotherapy and those who have not. If those who attend psychotherapy have fewer depressive symptoms than those who did not, the researchers may conclude that psychotherapy reduced the symptoms.

The added control group is a definite improvement over the design lacking a control group as it indicates which effects occur without the manipulation. However, the design is still rather weak because the experimental and control groups may have differed on many nonmanipulated variables related to outcome. This situation is often referred as *selection bias*. To estimate selection bias, this design can include an independent pretest sample that does not receive the experimental manipulation.

$$\begin{array}{cccc} NR & O_1 \mid & X & O_2 \\ \hline NR & O_1 \mid & & O_2 \end{array}$$

Here, the pretest and posttest samples are independent across time (i.e., they may be sampled at the same moment in time, and thus on different study units).

Further improvements include matching (where study units with similar scores are matched across experimental and control groups), using internal controls (where the control group consists of a sample drawn from a population similar to that of the experimental group), using multiple nonequivalent control groups, and using a predicted interaction (a highly differentiated causal

hypothesis that predicts one particular interaction but excludes others).

Nonequivalent Groups Pretest–Posttest Design

The nonequivalent groups pretest–posttest design is similar to the one-group pretest–posttest design with a control group added.

$$\begin{array}{cccc} \text{NR} & \text{O}_1 & \text{X} & \text{O}_2 \\ \hline \text{NR} & \text{O}_1 & & \text{O}_2 \end{array}$$

Note that although similar to the nonequivalent groups posttest-only design with a pretest, the current design includes a dependent pretest, that is, pretest and posttest data are gathered on the same study units. For example, Grace Carter, John Winkler, and Andrea Biddle compared scientists who had received a Research Career Development Award from the National Institutes of Health in order to improve their research careers to those who did not. They found that at posttest, scientists who had received such an award did better than those who did not. However, the former also exceeded the latter at pretest, thus calling into the question the effectiveness of the research awards.

Because pretest and posttest data are gathered on both experimental and control groups, only one of which receives the experimental manipulation, the existence of a possible selection bias may be estimated. Insofar as selection is present, it may magnify the effects of other threats to internal validity (e.g., maturation).

Improvements of this design include a double pretest that allows for the assessment of the selection-maturation threat to internal validity,

$$\begin{array}{ccccc} \text{NR} & \text{O}_1 & \text{O}_2 & \text{X} & \text{O}_3 \\ \hline \text{NR} & \text{O}_1 & \text{O}_2 & & \text{O}_3 \end{array};$$

switching replications, which entails delivering the manipulation to control group at a later date,

$$\begin{array}{ccccc} \text{NR} & \text{O}_1 & \text{X} & \text{O}_2 & \text{O}_3 \\ \hline \text{NR} & \text{O}_1 & & \text{O}_2 & \text{X} \quad \text{O}_3 \end{array};$$

using a reversed treatment control group, in which two manipulations are delivered that are opposite in direction of effect,

$$\begin{array}{cccc} \text{NR} & \text{O}_1 & \text{X}_+ & \text{O}_2 \\ \hline \text{NR} & \text{O}_1 & \text{X}_- & \text{O}_2 \end{array};$$

or direct measurement of threats to validity and incorporating these estimates into statistical analysis of the outcomes.

In addition to these modifications, cohort (successive, comparable groups) controls may also be used to improve the nonequivalent groups pretest-posttest design:

$$\begin{array}{cccc} \text{NR} & \text{O}_1 & & \\ \hline \text{NR} & & \text{X} & \text{O}_2 \end{array}.$$

The first group, which is similar to the second group in relevant aspects, does not receive the experimental manipulation, whereas the second group does. Because of the similarity of the two groups, any differences between them are assumed to be related to the manipulation.

This simple cohort control design may be further improved by adding pretests.

$$\begin{array}{ccccc} \text{NR} & \text{O}_1 & \text{O}_2 & & \\ \hline \text{NR} & & & \text{O}_3 \quad \text{X} & \text{O}_4 \end{array}$$

Interrupted Time-Series Designs

In an interrupted time series design, the same variable is measured repeatedly over time.

$$\text{O}_1 \ \text{O}_2 \ \text{O}_3 \ \text{O}_4 \ \text{O}_5 \ \text{X} \ \text{O}_6 \ \text{O}_7 \ \text{O}_8 \ \text{O}_9 \ \text{O}_{10}$$

A change in intercept or slope of the time series is expected at the point in time where the manipulation was delivered. For example, A. John McSweeny studied the effects of the Cincinnati Bell phone company instituting a charge of 20 cents per call for local directory assistance. At the point in time when this charge was added, a significant and immediate drop in the number of local directory assistance calls is visible in the data. This illustrates how a change in intercept of a time series can be used to judge the effect of a manipulation. A change in slope can also be used to judge the effect of a manipulation. For example, Julian Roberts and Robert Geboyts studied the impact of the reform of Canadian sexual assault law instituted in order to increase reporting of sexual assault crimes. They found a relatively flat slope in the years before the reform (that is, the reported number of

assaults remained fairly constant), whereas in the years following the reform, a steady increase in the reported number of assaults was observed.

A major threat to this type of design is history—that is, other factors occurring at the same time as the manipulation that may cause the outcome under investigation. Several improvements can be made to this design. A nonequivalent control group may be added:

$$\begin{array}{l} O_1\ O_2\ O_3\ O_4\ O_5\ X\ O_6\ O_7\ O_8\ O_9\ O_{10} \\ \hline O_1\ O_2\ O_3\ O_4\ O_5 \qquad O_6\ O_7\ O_8\ O_9\ O_{10} \end{array}$$

A nonequivalent dependent variable may be added:

$$\begin{array}{l} O_{A1}\ O_{A2}\ O_{A3}\ O_{A4}\ O_{A5}\ X\ O_{A6}\ O_{A7}\ O_{A8}\ O_{A9}\ O_{A10} \\ \hline O_{B1}\ O_{B2}\ O_{A3}\ O_{B4}\ O_{B5} \qquad O_{B6}\ O_{B7}\ O_{B8}\ O_{A9}\ O_{A10} \end{array}$$

The manipulation may be implemented and subsequently removed at a known time:

$$O_1\ O_2\ O_3\ O_4\ O_5\ X\ O_6\ O_7\ O_8\ O_9\ \not{X}\ O_{10}\ O_{11}\ O_{12}\ O_{13}$$

This may also be done multiple times, resulting in a multiple replications design:

$$O_1\ O_2\ X\ O_3\ O_4\ \not{X}\ O_5\ O_6\ X\ O_7\ O_8\ \not{X}\ O_9\ O_{10}\ X\ O_{11}\ O_{12}\ \not{X}\ O_{13}\ O_{14}$$

Delivering the manipulation to two nonequivalent groups at different times results in a switching replications design:

$$\begin{array}{l} O_1\ O_2\ O_3\ X\ O_4\ O_5\ O_6\ O_7\ O_8 \qquad O_9\ O_{10}\ O_{11} \\ \hline O_1\ O_2\ O_3 \qquad O_4\ O_5\ O_6\ O_7\ O_8\ X\ O_9\ O_{10}\ O_{11} \end{array}$$

Regression Discontinuity Design

In regression discontinuity designs, experimental manipulation is based on a cutoff score (C) on an assignment variable (O_A) measured before manipulation.

$$\begin{array}{llll} O_A & C & X & O_2 \\ O_A & C & & O_2 \end{array}$$

If the manipulation has an effect, then regression analysis of the data obtained from this design should reveal a discontinuity at the cutoff score corresponding to the size of the manipulation effect. If the manipulation has no effect, the regression line should be continuous. For example, Charles Wilder studied how the institution of the Medicaid program affected medical visits and found, perhaps unsurprisingly, that household income was positively correlated with medical visits. More importantly, however, the data also showed a dramatic increase in medical visits at the cutoff score for Medicaid eligibility, which supports the inference that the Medicaid program does indeed stimulate medical visits.

Differential attrition may cause a discontinuity in the regression line that resembles an effect of the manipulation and thus is a threat to internal validity. History is a plausible threat to validity if factors affecting the outcome occur only for study units on one side of the cutoff.

Inferences

Randomized experiments allow for solid inferences about a proposed causal relation of two variables. However, randomization is often not practically or ethically possible, making a quasi-experimental design necessary. Although threats to internal validity are often highly plausible in quasi-experimental designs, researchers can still draw valid causal inferences if they identify plausible threats to internal validity and select quasi-experimental designs that address those threats. Often, the design adjustments require elaborate changes, such as administering and removing the intervention multiple times. However, a small, simple change often can make a large difference, such as adding a nonequivalent dependent variable. In either case, causal inferences are strengthened, which is the primary purpose of experimentation.

Scott Baldwin and Arjan Berkeljon

See also Cause and Effect; Experimental Design; Internal Validity; Research Design Principles; Threats to Validity; Validity of Research Conclusions

Further Readings

Carter, G. M., Winkler, J. D., & Biddle, A. K. (1987). *An evaluation of the NIH research career development award*. Santa Monica, CA: RAND.

Duckhart, J. P. (1998). An evaluation of the Baltimore Community Lead Education and Reduction Corps

(CLEARCorps) program. *Evaluation Review, 22,* 373–402.
McSweeney, A. J. (1978). The effects of response cost on the behavior of a million persons: Charging for directory assistance in Cincinnati. *Journal of Applied Behavioral Analysis, 11,* 47–51.
Roberts, J. V., & Gebotys, R. J. (1992). Reforming rape laws: Effects of legislative change in Canada. *Law and Human Behavior, 16,* 555–573.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.
Wilder, C. S. (1972). *Physician visits, volume, and interval since last visit, U.S. 1969* (No. Series 10, No. 75; DHEW Pub. No. HSM 72–1064). Rockville, MD: National Center for Health Statistics.

Quetelet's Index

Quetelet's Index, more commonly known as the body mass index (BMI), is a measure of weight relative to height. Originally developed by 19th-century mathematician Lambert Adolphe Jacques Quetelet, it is a standard measurement that is widely used by health professionals and researchers as an index of body fat. The formula used to calculate this index is weight (in kilograms) divided by height (in meters squared). Height in meters can be calculated by dividing height in centimeters by 100. For example, a person who weighs 71 kilograms and is 165 centimeters tall has a BMI of 26 [$(71\ \text{kg})/(1.65\ \text{m})^2 = 26$]. Alternatively, BMI can be calculated by dividing weight (in pounds) by height (in inches squared) multiplied by a conversion factor of 703. For example, a person who weighs 145 pounds and is 65 inches (or 5′ 5″) tall has a BMI of 24 {$[(145\ \text{lbs.})/(65\ \text{in.})^2] \times 703 = 24$}.

Although the body mass index is not a direct measure of adiposity (body fat), it is correlated with direct measures of body fat, such as dual energy x-ray absorptiometry and underwater (hydrostatic) weighing. The advantages of using BMI as a proxy for adiposity are that it is inexpensive, it is easy to obtain, it does not require extensive training, and it is simple to calculate relative to other methods.

Interpretation for Adults

BMI is often used to classify overweight and obesity, but can be interpreted for the entire weight spectrum. Typically, a person is considered underweight if he or she has a BMI of 18.5 or less and normal weight between 18.5 and 24.9. A BMI of 25.0 to 29.9 is classified as overweight, and 30 or above is considered obese. The obese category can be further subdivided into Class I obesity (30.0 to 34.9), Class II obesity (35.0 to 39.9), and Class III or extreme obesity (40 or above). Recent estimates suggest that two thirds of the U.S. adult population are classified as overweight according to these categories, and nearly one third are obese.

Interpretation for Children and Adolescents

Although the calculation for BMI is the same for adults and children, the interpretation differs. For adults (aged 20 years and older), BMI is interpreted the same for both men and women using the categories listed above. However, for children and adolescents (aged 2 through 19 years), the interpretation is based on age- and sex-specific percentiles, which reflects the fact that adiposity changes with age and differs among boys and girls. Therefore, it is not appropriate to use the BMI categories for adults to interpret BMI and determine the weight category for children and adolescents. BMI can be used to identify children and adolescents who are either currently overweight or at risk of becoming overweight, based on BMI-for-age growth charts provided by the Centers for Disease Control and Prevention. After obtaining the BMI using the method described above, this number is plotted on the BMI-for-age growth charts for the appropriate sex to yield a percentile ranking, which indicates how the child's BMI compares to children of the same age and sex. Although children are usually not classified as obese, the current recommendation is that BMI values that meet or exceed the 95th percentile of the BMI growth charts for their age and sex should be categorized as overweight, and those who are in the 85th percentile to the 95th percentile are classified as at risk of overweight. These growth curves are J-shaped and were constructed to identify BMI scores that show a trajectory toward overweight ($\text{BMI} \geq 25$) or obesity ($\text{BMI} \geq 30$) in adulthood.