

QUASI-EXPERIMENTS AND CORRELATIONAL STUDIES

Michael L. Raulin and Anthony M. Graziano
State University of New York at Buffalo, USA

Experimental research

Quasi-experimental procedures

Non-equivalent control-group designs

Differential research designs

Interrupted time-series designs

Single-subject designs

Reversal (ABA) design

Multiple baseline design

Single-subject, randomized, time-series design

Correlational approaches

Simple correlations

Advanced correlational techniques

Simple linear regression

Advanced regression techniques

Path analysis

Conclusion

Further reading

References

Scientific research is a process of inquiry – sequences of asking and answering questions about the nature of relationships among variables (e.g., How does *A* affect *B*? Do *A* and *B* vary together? Is *A* significantly different from *B*? and so on). Scientific research is carried out at many levels that differ in the types of questions asked and, therefore, in the procedures used to answer them. Thus, the choice of which methods to use in research is largely determined by the kinds of questions that are asked.

In research, there are four basic questions about the relationships among variables: causality, differences between groups, direction and strength of relationships, and contingencies. This chapter will focus primarily on three of them: methods applied to questions of causality (experiments and quasi-experiments), methods to address questions about group differences

(differential research), and methods used to answer questions about the direction and strength of relationships among variables (correlations). (For more detail on these topics see Graziano and Raulin, 1993.)

EXPERIMENTAL RESEARCH

Experiments are the most effective way to address questions of causality – does one variable have an effect upon another (e.g., Does a new drug *X* reduce anxiety?). A precise and well-controlled experiment eliminates alternative explanations of results. If a well-designed experiment indicates that anxiety decreases when drug *X* is given, we shall have high confidence that it was drug *X*, and not some other factor, that brought about the decrease.

In an experiment, the causal hypothesis (that *X*, the independent variable, will affect *Y*, the dependent variable) is tested by manipulating *X* and observing *Y*. In the simplest experiment, the independent variable is manipulated by presenting it to one group of subjects and not presenting it to another group. The two groups must be comparable prior to the manipulation so that any post-manipulation group difference is clearly due to the independent variable and not to some other, extraneous variable. If an uncontrolled, extraneous variable may have affected the outcome, then we have an alternative explanation of the results. Consequently, we cannot be sure which variables were actually responsible for the effects. Comparability of groups before the manipulation is assured by assigning subjects randomly to the groups.

Well-designed experiments have sufficient controls to eliminate alternative explanations, allowing us to draw causal conclusions. Uncontrolled variables threaten the validity of the experiment and our conclusions. For example, suppose you develop a headache while working for hours at your computer. You stop, go into another room, and take two aspirin. After about 15 minutes your headache is gone and you return to work. Like most people, you would probably conclude that the aspirin eliminated the headache, that is, you would infer a causal relationship between variable *X* (aspirin) and variable *Y* (headache). But other variables besides the aspirin might have been responsible for the improvement. You stopped working for awhile. You left the room and stretched your legs. You may have closed your eyes for a few minutes and rubbed your temples. You took a cool drink of water when you swallowed the aspirin. Mostly, you took a break from the intensive work and relaxed for a few minutes. All of those variables are potential contributors to the observed effect; they are confounding factors in the research. Therefore, it is not clear that it was the aspirin alone that reduced the headache. To be confident of a causal relationship between the aspirin and headache reduction, one would have to carry out an experiment designed

to control these confounding factors and to eliminate the alternative explanations.

The discussion above refers to one of the most important concepts in experimentation – validity. A research study has good validity when it controls confounding factors and thus eliminates rival explanations, allowing a causal conclusion. To state it another way, uncontrolled extraneous variables threaten the validity of experiments.

QUASI-EXPERIMENTAL PROCEDURES

Experiments allow us to draw causal inferences with the greatest confidence. However, there are conditions under which we cannot meet all demands of a true experiment but still want to address causal questions. In these situations we can use quasi-experimental designs. Using quasi-experiments in clinical and field situations to draw cautious causal inferences is preferable to not experimenting at all.

Quasi-experimental designs resemble experiments but are weak on some of the characteristics. Quasi-experiments include a comparison of at least two levels of an independent variable, but the manipulation is not always under the experimenter's control. For example, suppose we are interested in the health effects of a natural disaster such as a destructive tornado. We cannot manipulate the tornado but we can compare those who experienced the tornado with a group of people who did not. Likewise, in many field situations we cannot assign subjects to groups in an unbiased manner. Indeed, we often cannot assign subjects at all, but must accept the natural groups as they exist. Thus, in quasi-experimental designs:

- 1 We state a causal hypothesis.
- 2 We include at least two levels of the independent variable, although we may not manipulate it.
- 3 We usually cannot assign subjects to groups, but must accept existing groups.
- 4 We include specific procedures for testing hypotheses.
- 5 We include some controls for threats to validity.

Compare this list with the characteristics of a *true* experiment:

- 1 We state a causal hypothesis.
- 2 We manipulate the independent variable.
- 3 We assign subjects randomly to groups.
- 4 We use systematic procedures to test the hypothesized causal relationships.
- 5 We use specific controls to reduce threats to validity.

There are a number of basic quasi-experimental designs, but we shall focus on four of the most important: non-equivalent control-group designs,

differential research designs, interrupted time-series designs, and single-subject designs.

Non-equivalent control-group designs

We can test causal hypotheses with confidence if we randomly assign subjects to groups, because such groups are likely to be equivalent at the beginning of the study. However, sometimes subjects cannot be assigned randomly to groups, and the groups may not be equivalent on some variables at the beginning of the study. Campbell and Stanley (1966) popularized the non-equivalent control-group design by suggesting that already existing groups can be similar to one another on most relevant variables even though there is no systematic assignment of subjects to groups. The more similar natural groups are to one another, the closer the design approximates a true experiment. Furthermore, Cook and Campbell (1979) showed that it is sometimes possible to draw strong conclusions from non-equivalent control-group studies, even when the groups are different, provided the researcher carefully evaluates all potential threats to validity.

There are two problems with non-equivalent groups: groups may be different on the dependent measure(s) at the start of the study, and there may be other differences between groups. To address the first issue, we include a pre-test measure. The pre-test tells us how similar the groups are on the dependent variable(s) at the beginning of the study. The more similar the groups are, the greater control we have. To address the second issue – that groups may differ on variables other than the dependent variable – it is important to rule out each potential confounding variable. To do this, we must first identify potential confounding variables, measure them, and carefully rule them out. Figure 1 shows six possible outcomes of a non-equivalent control-group design. In Figures 1(a), 1(b), and 1(c) the pattern of scores for the experimental and control groups suggests no effect of the independent variable. In Figure 1(a), neither group changes; the pre-test scores suggest that the group differences existed prior to the independent variable manipulation. Both Figures 1(b) and 1(c) show an equivalent increase in the groups on the dependent measure from pre-test to post-test, suggesting that there is no effect of the independent variable. Again, the pre-test in 1(b) allows us to rule out the hypothesis that the group post-test differences are due to the independent variable. In Figure 1(d), groups equivalent at the beginning of the study diverge; there does appear to be an effect of the independent variable. In both Figures 1(e) and 1(f), the groups differ on the dependent measure at pre-test, and the experimental group changes more than the control group after the manipulation. Figure 1(f) shows a slight change in the control group but a marked change in the experimental group, suggesting an effect of the independent variable. However, there is still the potentially confounding factor of regression to the mean. Regression is a potential source

of confounding whenever we begin an experiment with extreme scores. The marked pre-test difference between groups in Figure 1(f) may represent extreme scores for the experimental group. In the course of the experiment the scores for the group may have returned to the mean level represented by the control group. Consequently, we cannot be confident in attributing the results to the causal effects of the independent variable. In Figure 1(e), the control group does not change but the experimental group changes markedly in the predicted direction, even going far beyond the level of the control group. This is called a crossover effect. The results give us considerable confidence in a causal inference. Maturation (normal changes in subjects over time) and history (changes in subjects during the study due to events other than the independent variable manipulation) are unlikely alternative hypotheses because the control group should also have been affected by these factors. Regression to the mean is also an unlikely alternative hypothesis because the experimental group increased not only to the mean of the control group but also beyond it. With these results a quasi-experimental design gives us fairly good confidence in a causal inference.

These examples are reasonably interpretable. Other situations described by Cook and Campbell (1979) are more difficult, or even impossible, to

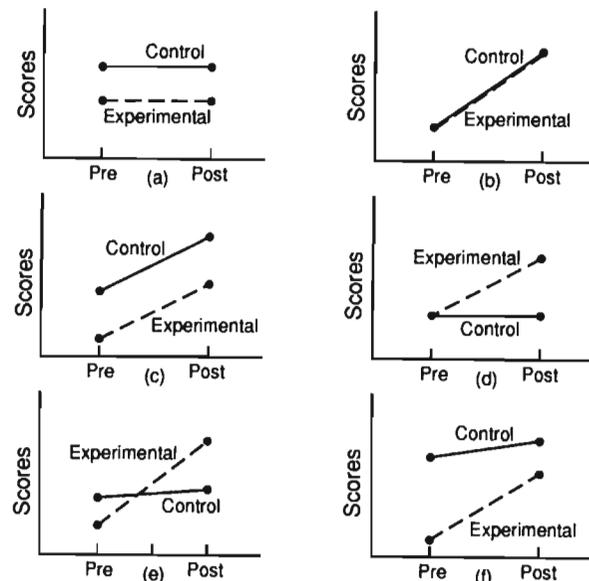


Figure 1 Possible outcomes from non-equivalent control group studies

interpret. Using non-equivalent control group designs appropriately requires considerable expertise.

Differential research designs

In differential research, pre-existing groups (e.g., diagnostic groups) are compared on one or more dependent measures. There is no random assignment to groups; subjects are classified into groups and measured on the dependent variable. In essence, the researcher measures two variables (the variable defining the group and the dependent variable). Consequently, many researchers classify differential research as a variation of correlational research. We believe that differential research designs can employ control procedures not available in straight correlational research and therefore should be conceptualized as somewhere between quasi-experimental and correlational designs. We cannot draw causal conclusions from differential research, but we can test for differences between groups.

A typical differential research study might compare depressed and non-depressed subjects. The dependent variable is selected for its theoretical significance. For example, we might measure people's judgment of the probability of succeeding on a test of skill, as Alloy and Abramson (1979) did. They hypothesized that depressed subjects would be more likely to expect failure. This hypothesis was based on a causal model of depression that suggested that one's attributional style would affect the risk for depression. This causal model could not be tested directly with a differential research design, but the data from this study could test the plausibility of the model. If there were attributional differences in depressed and non-depressed subjects, then it is plausible that these differences predated the depression, perhaps even contributing to the development of depression.

One could think of a differential research study as including an implicit manipulation that occurs prior to data collection – a manipulation that created the defining characteristic of the groups. This would be equivalent to comparing a group of people who lived through a tornado and a group who had never experienced one. The groups are defined by an event that predated the research study. However, the exposure to a tornado is likely a random event, so this comparison is conceptually close to an experiment. The subjects are assigned randomly, although not by the researcher; one group experiences the independent variable (i.e., the tornado), although again not controlled by the researcher; dependent measures are taken after the manipulation. We would classify such a study as a strong quasi-experimental design and would feel justified in drawing rather strong causal conclusions on the basis of our group comparisons. However, when the manipulation is something less likely to be random, such as becoming depressed or not, the possibility of confounding is greatly increased.

Whenever you start with pre-existing groups, the groups are likely to differ

on many variables other than the variable that defines the groups. For example, if we compare schizophrenic patients with a randomly selected non-patient control sample, the groups would differ not only on diagnosis, but probably also on social class, education level, average IQ, amount of hospitalization, history of medication use, and the social stigma associated with a psychiatric diagnosis. Any difference between our schizophrenic and control samples could be due to the disease or any of the other differences listed above. These variables are all confounded with the diagnosis of the subjects. It is impossible to draw a strong conclusion.

Confounding is the norm in differential research, even in cases where you might not expect it. For example, if we compare children of various ages in a cross-sectional developmental study, we might expect that the children made it into the various groups by the random factor of when they were born. That may be true, but there are likely to be differences between the groups that are a function of historical factors unique to a given age range of children. These might include major historical events occurring at critical ages, differences in economic conditions that affect what resources are available to the children at any given age, differences in school systems that may be the result of budget issues or political pressures, or even the impact of a single teacher. (Note that most samples of subjects for research come from accessible populations from a narrow geographic area. Therefore, it is possible that a single teacher could differentially affect the results.) Differences between age groups that are the result of a different set of historical experiences are known as cohort effects.

The ideal control group in differential research is identical to the experimental group on everything except the variable that defines the groups. This ideal often is impossible. Therefore, researchers attempt to equate the groups on critical variables, that is, variables that could confound the interpretation of the results. A variable can confound the results, first, if it has an effect on the dependent variable, and second, if there is a mean difference on the variable in the groups being compared. For example, IQ might confound results in a study of cognitive styles, but hair colour is unlikely to because hair colour is probably unrelated to cognitive styles. However, even though IQ is a potential confounding variable, it cannot confound the results *unless* there is a mean difference between the groups. In differential research, it is common to include control groups that are matched on one or more of these critical variables to avoid confounding. For more discussion of this strategy, see Chapman and Chapman (1973) and Graziano and Raulin (1993).

Interrupted time-series designs

In interrupted time-series designs, a single group of subjects is measured several times both before and after some event or manipulation. The multiple

measures over time strengthen the design considerably over a simple pre-post design, controlling many potential confounding factors. A major potential confounding factor in the simple pre-post study is regression to the mean. Behaviour fluctuates over time, displaying considerable variability. The intervention might be applied only at a high point in that natural variation, just before the behaviour decreased again. Thus, the observed changes in behaviour may not be due to the treatment at all but only to the natural variability of behaviour. The same reduction might have been observed even if we had not applied the treatment. The multiple measures of the interrupted time-series design give several points of comparison, allowing us to rule out the effects of regression to the mean. We can see the natural variability and can see if the post-treatment change exceeds the natural variability.

Figure 2 shows the results of an interrupted time-series study of disruption in autistic children (Graziano, 1974). Disruptive behaviour of four autistic children was monitored for a full year before the treatment (relaxation training) was introduced and a year following the treatment. Note that the variability during baseline disappears after treatment, with disruptive behaviour dropping to zero and remaining there for a full year. Such results are not likely due to normal fluctuation or regression to the mean. They also seem unlikely to be due to maturation of all subjects during the same period of time. With time-series designs, however, there are still two potentially confounding factors – history and instrumentation. History can confound results in any procedure that requires a fairly long period of time because other events might account for changes in the dependent variable. Thus, when using the interrupted time-series design, the experimenter must identify potential confounding due to history and carefully rule it out. Instrumentation is another potential threat to validity. When new programmes are started there may be accompanying changes in the way records are kept. The

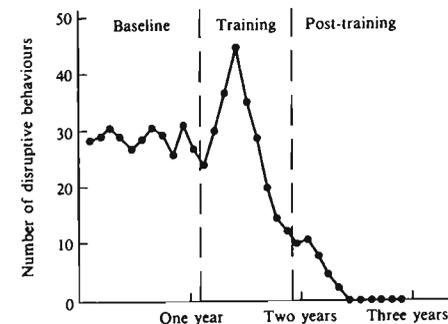


Figure 2 An interrupted time-series design showing the effects of relaxation treatment on disruptive behaviour in autistic children

researcher must be careful to determine that an apparent change is not due to changes in record-keeping.

The interrupted time-series design is useful in clinical or naturalistic settings where the effects of some event, naturally occurring or manipulated, can be assessed by taking multiple measurements both before and after the event. It can also be used in studies where the presumed causal event occurs to all members of a population. For example, the effects of a policy change (e.g., a change in the speed limit) could be evaluated with an interrupted time-series design using routinely gathered data (e.g., traffic fatality counts).

The interrupted time-series design can be improved by adding one or more comparison groups. In our hypothetical study of the effects of a change in speed limit on the number of fatalities, we could use comparable data from a neighbouring state that did not reduce the speed limit. Such a comparison would help to control for potentially confounding factors such as history and maturation.

Graphical presentations of data in the interrupted time-series design can provide considerable information. In a time-series study, the change in the time graph must be sharp to be interpreted as anything other than only a normal fluctuation. Slight or gradual changes are difficult to interpret. But in a time-series design, simply inspecting the graph is not enough. Testing the statistical significance of pre-post differences in time-series designs requires sophisticated procedures, which are beyond the scope of this chapter (see Glass, Willson, & Gottman, 1975; or Kazdin, 1992).

Single-subject designs

Single-subject designs were developed early in the history of experimental psychology and were used in both human and animal learning studies. Since the early 1960s they have become popular in clinical psychology. With single-subject designs we are able to manipulate independent variables, to observe their effects on dependent variables, to draw causal inferences, and to do so with a single subject. Modern clinical psychology is now heavily reliant on behaviour modification treatment methods, and behaviour modification research utilizes single-subject research designs refined from the work of B. F. Skinner. For more information on single-subject designs, consult Barlow and Hersen (1984), Kratochwill (1978), and Sidman (1960).

Single-subject designs are variations of time-series designs. The same subject is exposed to all manipulations, and we take dependent measurements of the same subject at different points in time. This allows us to compare measures taken before and after some naturally occurring event or an experimental manipulation. The basic comparison is between the same subject's own pre-treatment and post-treatment responses. Note that at its simplest level, this resembles the pre-test-post-test comparison – a relatively weak non-experimental design. A control group would strengthen a pre-test

– post-test design, but a control group is not possible when we have only one subject. Single-subject designs improve on the pre-post design, not by adding a control group, but by adding more conditions to the experiment. If the dependent variable changes in the predicted direction at each manipulation, we can have reasonable confidence that the manipulation is responsible for the observed change in the dependent variable. There are several single-subject designs, including reversal or ABA designs, multiple baseline designs, and single-subject, randomized, time-series designs.

Reversal (ABA) design

In reversal or ABA designs, the effects of an independent variable on a dependent variable are demonstrated by measuring the dependent variable at three or four points in time. There is a no-treatment baseline period during which the dependent behaviour is only observed, a treatment period in which the manipulation is carried out, and a return or reversal to the no-treatment condition. The effects of the independent variable (the treatment) on the dependent variable (the behaviour to be changed) is demonstrated if the behaviour changes in the predicted direction whenever the conditions are reversed. We often strengthen the design by measuring the dependent variable several times during each condition. A hypothetical study will help to describe the general format used. The study concerns self-stimulatory behaviour of a retarded child, Terry. After observing Terry in the classroom, a psychologist forms the tentative hypothesis that the teacher's attention is reinforcing the self-stimulatory behaviour. That is, whenever Terry begins her self-stimulatory activity, the teacher tries to soothe and comfort her. The teacher does not realize that it may be her efforts to help Terry control the behaviour that are actually helping to maintain it.

To test the hypothesis the psychologist sets up an ABA design, in which condition *A*, the baseline, involves the teacher's usual approach of attending to Terry whenever she displays the self-stimulatory behaviour. Condition *B* is the treatment – a differential reinforcement procedure in which the teacher provides attention and support for Terry whenever she refrains from the self-stimulatory behaviour, but withdraws attention when Terry engages in self-stimulatory behaviour. Precise observations are carried out for one hour at the same time each day. The graph in Figure 3 shows the behavioural changes as the *A* and *B* conditions are sequentially reversed. The graph suggests that there may be a causal relationship between teacher attention and Terry's self-stimulatory behaviour. Notice that the psychologist has not limited the approach to only three conditions, ABA, but has added another reversal at the end for an ABAB procedure. The ABA sequence is sufficient to suggest causality, but the demonstration of a causal link creates an ethical demand to return Terry to the optimal state.

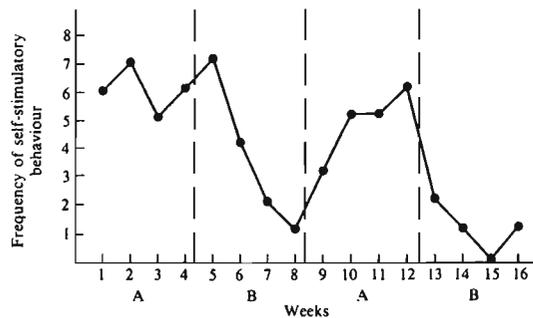


Figure 3 A reversal design showing the effects of contingent reinforcement on self-stimulatory behaviour of a single child

Multiple baseline design

Although the ABA design can provide a powerful demonstration of the effect of one variable on another, there are situations in which reversal procedures are not feasible or ethical. For example, suppose in our example of Terry above, the self-stimulatory behaviour was injurious, such as severe head-banging. We would be unwilling to reverse conditions once we achieved improved functioning because it could risk injury. Instead, we could use a multiple baseline design.

In the multiple baseline design, the effects of the treatment are demonstrated on different behaviours successively. To illustrate, we shall use an example similar to the previous example. Suppose that a fifth-grade boy (about 10 years old) is doing poorly in school, although he appears to have the ability to achieve at a high level. He also disrupts class frequently and often fights with other students. A psychologist spends several hours observing the class and notes some apparent contingencies regarding the boy's behaviour. The teacher, attempting to control the boy, pays more attention to him when he is disruptive – scolding, correcting, lecturing him, and making him stand in a corner whenever he is caught fighting. The psychologist notes that the boy seems to enjoy the attention. However, on those rare occasions when he does his academic work quietly and well, the teacher ignores him completely. "When he is working, I leave well enough alone," the teacher says. "I don't want to risk stirring him up." Based on these observed contingencies, the psychologist forms the tentative hypothesis that the contingent teacher attention to the boy's disruptive behaviour and fighting may be a major factor in maintaining these behaviours, whereas the teacher's failure to reward the boy's good academic work may account for its low occurrence. The psychologist sets up a multiple-baseline design to test the hypothesis about the importance of teacher attention on disruptive

behaviour, fighting, and academic performance. The independent variable here is teacher attention.

Figure 4 shows the sequence of phases of the hypothetical study. During Phase 1 all three dependent variables are measured while the teacher continues the usual procedure of trying to punish the disruption and fighting while ignoring the positive academic behaviour. As seen in Figure 4, disruptive behaviour and fighting are high and academic performance is low. In Phase 2, the teacher's attention to fighting is withdrawn and positive attention is made contingent on academic work. In Phase 3, these procedures continue and the teacher withdraws attention for disruption as well as for fighting. The measured changes in the dependent variables associated with the independent variable manipulations provide evidence for the hypothesis that contingent teacher attention is an important controlling factor in the child's behaviour.

Single-subject, randomized, time-series design

When a reversal design is not appropriate and a multiple baseline procedure is not feasible because we want to study only one behaviour, the single-subject, randomized, time-series design can be used. This design is an interrupted time-series design for a single subject with one additional element – the randomized assignment of the manipulation in the time-series.

The single-subject, randomized, time-series design could be applied in the example above, but let us take another example. Suppose that Joey, another child in the special class, does not complete his daily work. During lesson periods, when he should be responding to a workbook lesson, Joey looks around the room or just closes his eyes and does no work. Reminders from

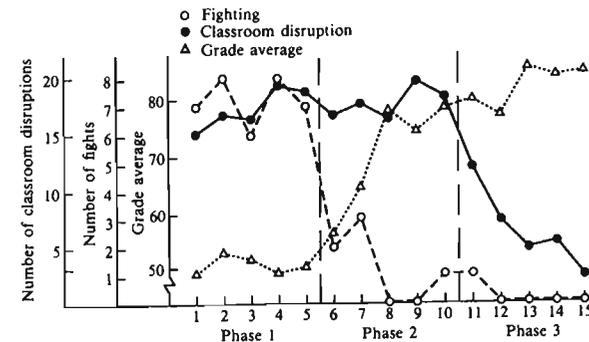


Figure 4 A multiple baseline design showing improvement in disruptive behaviour, fighting, and academic performance for a single child contingent upon teacher attention

the teacher have little effect. An effective motivational intervention with children is a token reinforcement system in which paper or plastic tokens are given to the child whenever he engages in the desired behaviour. The tokens serve as immediate secondary reinforcement for the desired behaviour. They are saved by the child and cashed in for items and privileges. If we were to employ a single-subject, randomized, time-series design, we might decide to measure the child's homework achievement for 6 weeks (30 school days). We might decide that we want at least 5 days before and after the implementation of the treatment. This ensures adequate pre-treatment and post-treatment measures. We then use a table of random numbers to select randomly one of the middle 20 days as our point for introducing the manipulation. The manipulation is the use of token reinforcement for homework achievement. Suppose we randomly select the ninth day as the point for introducing the token reinforcement programme. The beginning of the manipulation is preceded by 8 days of baseline measurement followed by 22 days of measurements of the dependent variable under the token reinforcement condition. If the time graph shows a marked improvement in homework achievement coincident with the ninth measurement, we have a convincing illustration of the effects of the token reinforcement. Note that it is unlikely that such marked improvement would occur by chance, or because of maturational or historical factors, at exactly the point at which we have randomly introduced the treatment.

CORRELATIONAL APPROACHES

Like quasi-experimental designs, correlation designs are used in situations in which the manipulation of an independent variable is either impossible or unethical. Because there is no experimental manipulation, one must be cautious in drawing causal conclusions. In fact, most correlational procedures are not powerful enough to justify causal interpretations.

Simple correlations

The correlation coefficient is probably the single most widely computed statistic in psychology. In many research studies, the purpose of the study is to produce measures of relationships between variables (i.e., correlations). Even in experimental designs or other designs, it is common to compute numerous correlation coefficients to help interpret the data. These correlation coefficients may or may not be reported in the final paper, but they are often routinely computed. Correlations between demographic variables and performance on the dependent measure often help us to identify potential confounding variables in a current study or in a future study that might be run.

The most commonly used correlation coefficient is the Pearson product-moment correlation. This coefficient is used when both variables are

measured on an interval or ratio scale. A Spearman rank-order correlation is preferred if at least one of the two variables is measured on an ordinal scale. Computational procedures for either of these coefficients are readily available in almost any undergraduate statistics textbook. There are other correlation coefficients available as well, which we shall discuss shortly. The range for both the Pearson and Spearman correlations is -1.00 to $+1.00$. A correlation of $+1.00$ indicates a perfect positive relationship (i.e., as one variable increases the other increases by a predictable amount). A correlation of -1.00 indicates a perfect negative relationship (i.e., as one variable increases the other variable decreases by a predictable amount). The sign indicates the direction of the relationship and the absolute size of the correlation indicates the strength of the relationship.

Correlations are most easily visualized in a scatter plot. Each person is plotted in a coordinate system in which their location is determined by the scores on variables X and Y . Figure 5 gives several examples of scatter plots, each indicating a particular degree of relationship. The actual product-moment correlation is indicated next to each scatter plot. Figures 5(a) and 5(b) illustrate scatter plots for strong positive and negative correlations, respectively. Figures 5(c) and 5(d) illustrate zero correlations. Note a zero correlation is often described as circular scatter plot. The scatter plot is circular, however, only if the variance on variables X and Y are equal, something that rarely occurs. Figure 5(d), for example, illustrates a zero correlation where variable X has a greater variance than variable Y . In this situation, the circular correlation is elongated horizontally. Figure 5(e) shows a perfect positive relationship with all the points clearly lining up on a straight line. Figure 5(f) shows the powerful effect of a single deviant score, especially when you have a small sample. In this case, 14 of the 15 data points clearly seem to show a zero correlation, but the correlation when you include the fifteenth point (at 10, 10) is $.77$. Finally, scatter plots shown in Figure 5(g) and 5(h) illustrate non-linear correlations. The product-moment correlation is sensitive only to the linear component. In Figure 5(g), the correlation is essentially zero, whereas in Figure 5(h), the correlation is somewhat positive. In both Figure 5(g) and 5(h), the product-moment correlation is an inappropriate measure of relationship. Simple product-moment correlation should be used only in situations where you anticipate a linear relationship between variables X and Y .

As mentioned earlier, the strength of the relationship between X and Y is illustrated by the size of the correlation regardless of the sign. A commonly used index is the square of the correlation, which can be interpreted as the proportion of variability in one variable that is predictable on the basis of knowing the scores on the second variable. This statement is usually shortened to "the proportion of variance accounted for".

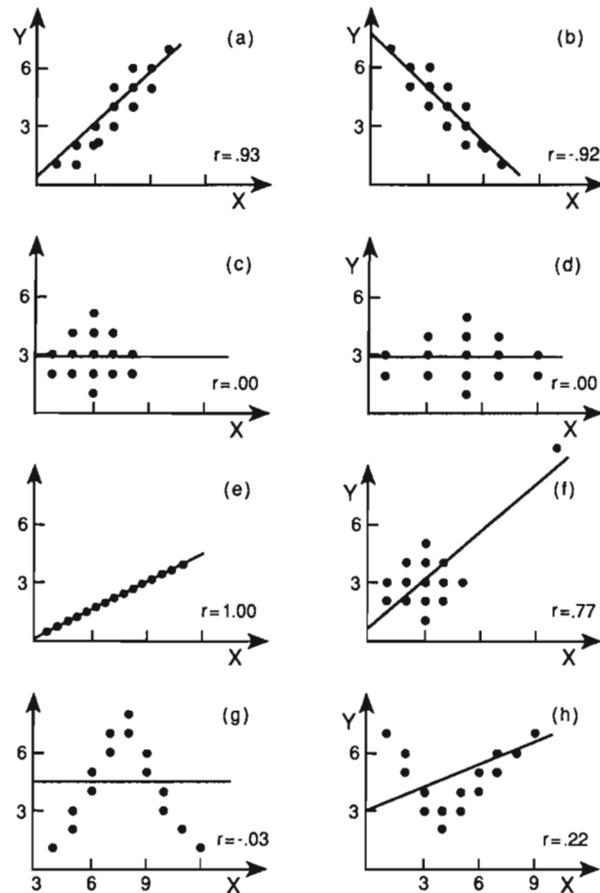


Figure 5 Several examples of scatter plots, regression lines, and the product-moment correlations that each plot represents

Advanced correlational techniques

More sophisticated correlational procedures are also available. You can correlate one variable with an entire set of variables (multiple correlation) or one set of variables with another set of variables (canonical correlation). It is also possible to correlate one variable with another after statistically removing the effects of a third variable (partial correlation). Discussion of these procedures is beyond the scope of this chapter, but the interested reader is referred to Nunnally (1976) for a more detailed discussion.

Simple linear regression

Regression techniques utilize the observed relationship between two or more variables to make predictions. The simplest regression technique is linear regression, with one variable being predicted on the basis of scores on a second variable. The equation below shows the general form for the prediction equation.

$$\text{Predicted } Y \text{ score} = (b \times X) + a$$

The values of the slope b and the intercept a in the above equation are a function of the observed correlation and the variances for both the X and Y variables. The computational detail can be found in virtually any undergraduate statistics text (e.g., Shavelson, 1988). In each of the eight scatter plots shown in Figure 5, the regression line for predicting Y from X has been drawn. When the correlation is zero the regression line is horizontal with an intercept at the mean for Y . In other words, if X and Y are unrelated to one another the mean of the Y distribution is the best prediction of Y , regardless of the value of X .

Advanced regression techniques

It is possible to use the relationships of several variables to the variable that you wish to predict in a procedure known as multiple regression. Multiple regression is also a linear regression technique, except that instead of working in the two-dimensional space indicated in Figure 5, we are now working in N -dimensional space, where N is equal to the number of predictor variables plus one. For example, if you have two predictor measures and one criterion measure, the prediction equation would be represented by a line in the three-dimensional space defined by these measures. If you have several predictor measures, visualizing multiple regression is difficult, even though the procedure is conceptually straightforward.

Although it is possible to put all of the predictor variables into the regression equation, it is often unnecessary to do so in order to get accurate predictions. The most commonly used procedure for a regression analysis is a procedure called stepwise regression, which uses a complex algorithm to enter variables into the equation one at a time. The algorithm starts by entering the variable that has the strongest relationship with the variable that you want to predict, and then selects additional variables on the basis of the incremental improvement in prediction that each variable provides. The computational procedures for stepwise regression are too complex to be done without the use of a statistical analysis program such as SPSS or SPSS/PC+ (Norusis, 1990b).

All of the regression models discussed above assume linear regression. In Figure 5(g) and 5(h), the scatter plot suggests that there is a non-linear

relationship between variables X and Y . When a non-linear relationship exists, there are non-linear statistical procedures for fitting a curve to the data. These procedures are well beyond the scope of this chapter: the interested reader is referred to Norusis (1990a).

Path analysis

A procedure that is rapidly becoming the standard for analysis of correlational data is path analysis. Path analysis is one of several regression procedures that fall under the general category of latent variable models. All latent variable models make the assumption that the observed data are due to a set of unobserved (i.e., latent) variables. Factor analysis is probably the most widely used of the latent variable models.

Path analysis seeks to test the viability of a specified causal model by factoring the matrix of correlations between variables within the constraints of the model. This process is best illustrated with an example. Suppose that we had three variables (A , B , and C) that we hypothesize are causally related to another variable (E). Further, we hypothesize that the causal effects of variables B and C on variable E are indirect – that is, variables B and C are causally related to a variable D which in turn is causally related to variable E . This model is illustrated in Figure 6. Straight lines in a path model represent hypothesized causal connections, while curved lines represent potential correlations that are not hypothesized to be causal. To make our example more intuitive, we shall present a scenario where such a model might make sense. We shall assume that we are studying high-risk sexual behaviour (i.e., sexual behaviour that increases the risk of HIV infection). Variable E in this model is *safe-sex behaviour in heterosexual college students*. We are hypothesizing that safe-sex behaviour is a function of *knowledge of safe-sex procedures* (variable D) and how *vulnerable* one feels (variable A). Whether one obtains knowledge of safe-sex procedures is hypothesized to be a function of *whether such information is available to people* (variable B) and *whether one believes that AIDS is not just a disease found in gay men or IV drug users* (variable C). In our model, we are suggesting that simply making information about safe-sex procedures available will not lead to the practice of safe sex unless (1) the information is learned, which will occur only if (2) the person believes that heterosexual sex represents a risk. Even if the person knows about safe sex procedures, the behaviour will not be practised unless (3) the person feels vulnerable to AIDS at the time of engaging in sex. We could expand this model by adding other variables if we wanted. For example, we might hypothesize that vulnerability is increased if the person knows people with HIV infection and is temporarily decreased if the person has been drinking (i.e., additional variables F and G are causally related to variable A). However, we shall restrict our model to the five variables shown in Figure 6 to illustrate the procedures.

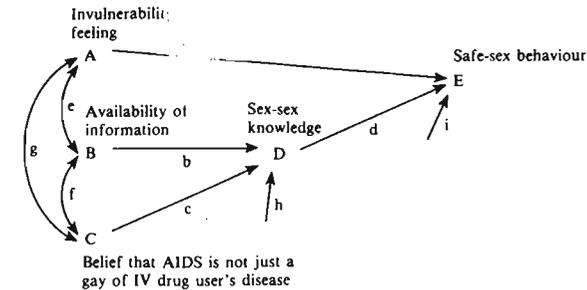


Figure 6 A hypothetical example of a path analysis

We have labelled the various paths in Figure 6 with lower-case letters. We shall solve for the strength of path coefficients for each of these paths. We have also included residual arrows for variables D and E . These residual arrows will have a strength that represents the unexplained variance in our model. Technically, all variables would have residual arrows, but it is customary not to include them with the initial variables (i.e., variables A , B , and C in our model). These variables represent our starting-point; our model does not address the question of what factors cause these initial states.

To evaluate the feasibility of our causal model, we need measures of the variables on a large sample of subjects. We need not have measures of all of the variables to do the computation; some variables may be unseen (latent). We start by computing a correlation matrix from which the path coefficients are computed. The actual path coefficient computations are beyond the scope of this chapter (the interested reader is referred to Loehlin, 1992). Path analysis tests the feasibility of a hypothesized causal model. If the absolute value of the path coefficients in the model are generally large and the residual coefficients are generally small, the model is feasible. You strive to create as parsimonious a path model as possible. Generally, as you include more paths, you will explain more of the variance, but you also run the risk of capitalizing on chance variance.

CONCLUSION

If given a choice, a researcher should use an experimental design. Manipulating variables and observing their effects on other variables, coupled with the other controls that are a part of experimental research, gives us the greatest confidence that the observed relationship is causal. However, there are many circumstances in which experimental research is impractical or unethical. This chapter has described some of the many quasi-experimental and correlational designs available for these situations. Data

from these designs should be interpreted with caution because the possibility of confounding is much larger than with experimental designs. However, these designs have proved their value in psychological research.

FURTHER READING

- Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behaviour change* (2nd edn). New York: Pergamon.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research on teaching*. Chicago, IL: Rand McNally.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi experimentation: Design and analysis issues for field studies*. Chicago, IL: Rand McNally.
- Graziano, A. M., & Raulin, M. L. (1993). *Research methods: A process of inquiry* (2nd edn). New York: Harper Collins.
- Kazdin, A. E. (1992). *Research design in clinical psychology* (2nd edn). New York: Macmillan.
- Loehlin, J. C. (1992). *Latent variable models: An introduction to factor, path, and structural analyses* (2nd edn). Hillsdale, NJ: Lawrence Erlbaum.

REFERENCES

- Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingencies in depressed and nondepressed subjects: Sadder but wiser? *Journal of Experimental Psychology: General*, 108, 447-485.
- Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behaviour change* (2nd edn). New York: Pergamon.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research on teaching*. Chicago, IL: Rand McNally.
- Chapman, L. J., & Chapman, J. P. (1973). *Disordered thought in schizophrenia*. Englewood Cliffs, NJ: Prentice-Hall.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi experimentation: Design and analysis issues for field studies*. Chicago, IL: Rand McNally.
- Glass, G. V., Willson, V. L., & Gottman, J. M. (1975). *Design and analysis of time series*. Boulder, CO: Laboratory of Educational Research Press.
- Graziano, A. M. (1974). *Child without tomorrow*. Elmsford, NY: Pergamon.
- Graziano, A. M., & Raulin, M. L. (1993). *Research methods: A process of inquiry* (2nd edn). New York: Harper Collins.
- Kazdin, A. E. (1992). *Research design in clinical psychology* (2nd edn). New York: Macmillan.
- Kratochwill, T. R. (Ed.) (1978). *Single-subject research: Strategies for evaluating change*. New York: Academic Press.
- Loehlin, J. C. (1992). *Latent variable models: An introduction to factor, path, and structural analyses* (2nd edn). Hillsdale, NJ: Lawrence Erlbaum.
- Norusis, M. J. (1990a). *SPSS-PC + Advanced Statistics 4.0*. Chicago, IL: Statistical Package for the Social Sciences.
- Norusis, M. J. (1990b). *SPSS-PC + Statistics 4.0*. Chicago, IL: Statistical Package for the Social Sciences.
- Nunnally, J. C. (1976). *Psychometric theory* (2nd edn). New York: McGraw-Hill.

- Shavelson, R. J. (1988). *Statistical reasoning for the behavioral sciences* (2nd edn). Boston: Allyn & Bacon.
- Sidman, M. (1960). *Tactics of scientific research: Evaluating scientific data in psychology*. New York: Basic Books.