SOCIOLOGY REFERENCE GUIDE

RESEARCH & EVALUATION METHODS

SOCIOLOGY REFERENCE GUIDE

RESEARCH & EVALUATION METHODS

The Editors of Salem Press

SALEM PRESS
A Division of EBSCO Information Services
Ipswich, Massachusetts

ISBN: 978-0-8242-1342-8

ebook: 978-1-42983-491-9

Contents

Introduction

In order for sociologists to prove the validity of their theories, it is important that they have adequate and valid statistical support. While the gathering and analysis of this data can often be tedious, the results can be extremely beneficial to a sociologist's cause.

The *Sociology Reference Guide* series is designed to provide a solid foundation for the research of various sociological topics. This volume ties together a valuable set of essays that overview the tools and methods used in research and evaluation. The essays ease into the matters of research by explaining the methods of data collection, the tools and theories used in the analysis of data, and the ethical implications involved in the process of empirical research.

This volume begins with the intricacies involved in the designing of a research project. Ruth A. Wienclaw explains what constitutes "good" research design and how to limit extraneous variables. An important stage in this process is hypothesis construction, which is the first step in testing the validity of a theory. The reference guide then details the various methods sociologists employ in their gathering of data. Collecting field data, surveying individuals, sampling populations, and conducting experiments are just a few examples of popular collection techniques. However, for data to be useable, all collection instruments and measures must be reliable and valid. An analysis of choosing proper means for data collection is examined in the volume's essay on reliability. Alexandra Howson then expounds upon the use of quantitative and qualitative research and evaluation methods, which provide important paradigms in the research of social and behavioral science.

In addition to the components involved in the gathering of data, the tools used in the analysis and understanding of research must also be taken into account. Probability theory can determine the likelihood that a hypothesized relationship between variables exists, but it does not necessarily prove that this hypothesis is correct. It does, however, help for expressing the level of confidence in the variable relationship. The types of variables that can play significant roles in hypothesis testing include independent variables, extraneous variables, intervening variables, and dependent variables. As statistical analysis with regard to sociological studies become more in-depth, studies on confidence intervals and correlation become more relevant. This volume covers several of these statistical topics, including inferential statistics, which allow sociologists to draw conclusions from data that are worth testing, and descriptive statistics, which allow individuals to better understand masses of data. This reference guide concludes with Wienclaw's examinations of the unethical use of statistics and the importance and necessity of maintaining proper ethics when conducting research in sociology.

Conducting research and analyzing data will likely remain extremely important to the more empirical elements involved in sociological study. This volume will provide readers with an overview of these issues and the diverse range of theories in the study of research and evaluation methods. Complete bibliographic entries, a list of suggested readings, and relevant terms and concepts finish the essay.

Designing a Research Project

Table of Contents

Abstract

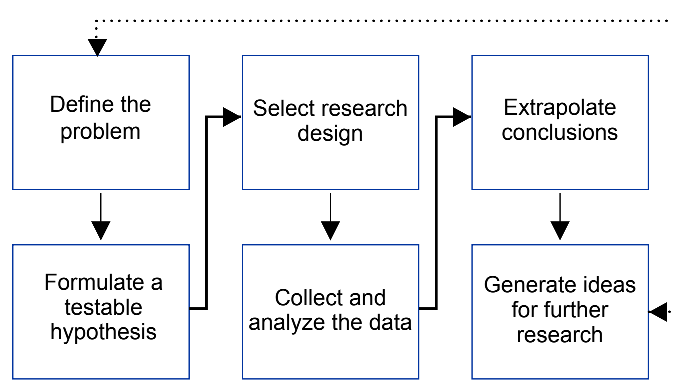
All science advances through the rigorous application of the scientific method. Part of this process involves the development of an empirical research design that can help researchers determine whether or not the hypothesis being tested is likely to be true. Good research design is based on a researcher's empirical observations and a review of the scientific literature. The information garnered from these sources is then formulated into a testable hypothesis that can be analyzed using inferential statistics. The research design used to test this hypothesis needs to not only consider the effect of various levels of the independent variables on the dependent variables, but also control as much as possible any extraneous variables that are not related to the research question but which might affect the results. The experimental data are analyzed to determine their statistical significance and the likelihood the null hypothesis is true using statistical tests.

Overview

Progress in the physical, behavioral, and social sciences is made through the systematic and rigorous application of the scientific method to observed real-world phenomena. The scientific method comprises the general procedures, guidelines, assumptions, and attitudes required for the organized and systematic collection, analysis, interpretation, and verification of data that can then be reproduced. The goal of the scientific method is to articulate or modify the laws and principles of a science. As shown in Figure 1, steps in the scientific method include

- problem definition based on observation and review of the literature;
- formulation of a testable hypothesis;
- selection of a research design;
- data collection and analysis;
- extrapolation of conclusions; and
- development of ideas for further research.

Figure 1: The Scientific Method



Observing & Researching Phenomena

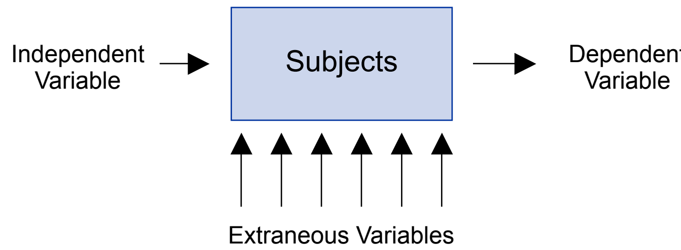
Typically, scientific research begins with the scientist's empirical observations. For example, I might observe that when I wear a business suit to a meeting even when other people are wearing more casual clothes, I tend to be afforded more respect than when I wear less formal attire. If curious, I might next look at social science literature to see if anyone else has observed such incidents and hypothesized an underlying cause. I might find that there is a large body of research on how to "dress for success." My literature review might reveal that other scientists have not only observed these behaviors but also theorized about their causes and conducted research to test their theories. Problem definition relies on both of these sources of information: the researcher's observations of real-world phenomena and the research results and theories that are described in the scientific literature.

If the literature review has not answered all my questions and I am still curious about the nature of this phenomenon, my next step would be to formulate a testable hypothesis. This is not necessarily as easy as it sounds; although it might be relatively easy to articulate a naïve theory concerning the relationship between attire and success in the workplace, such as that people who wear business attire are more likely to be successful at work, such statements are vague and not testable. To be able to test this tentative hypothesis using the scientific method, one must determine what factors are important in this theory and then operationally define the associated terms.

Identifying & Defining Variables

In the simplest research design, a stimulus, such as a person wearing either business attire or casual attire, is presented to the research subjects—in this case, potential customers, supervisors, or other people who might be encountered in a business setting. The responses of the subjects are then observed and recorded. From a research design point of view, both the stimulus and the response are called variables. The variables of most concern in the design of a research study are the independent variable, which is the stimulus or experimental condition that is hypothesized to affect the outcome (e.g., how one dresses in the workplace), and the dependent variable, which is the observed effect on outcome caused by the independent variable (e.g., the reactions of research subjects to people wearing business attire). As shown in Figure 2, researchers must also consider extraneous variables, or variables that can affect the outcome of the experiment but have nothing to do with the independent variable itself. For example, if the person wearing either business or casual attire is rude when interacting with a research subject, this rudeness will probably have a much stronger effect on the subject's response than how the person is dressed. Similarly, if the person has visible tattoos, is poorly groomed, or looks like the subject's ex-spouse, the subject's response may be a reaction to these extraneous variables rather than to the independent variable. Any number of extraneous variables may affect an experiment. The more of these variables that are accounted for and controlled in the experimental design, the more meaningful the results of the research study will be.

Figure 2: Research Variables



In addition to determining which variables are important in the research study, it is also essential to operationally define them. An operational definition is a definition that is stated in terms that can be observed and measured. In this example, the researcher might operationally define "business attire" as a dark suit with a white shirt or blouse. "Casual attire" might be operationally defined as shorts and a t-shirt. However, by operationally defining these terms in this manner, the researcher is by necessity limiting the generalizability of the research results. With these definitions, the researcher will only be able to draw a conclusion about subjects' reactions to people wearing dark suits and white shirts or blouses versus their reactions to people wearing shorts and t-shirts. A whole range of other work-appropriate attire exists: sports coats and blazers, colored shirts or blouses, Bermuda shorts, polo shirts, and any other type of clothing that could be worn in the workplace. The researcher must decide how many of these options are important to the theory and should be tested in the experiment. The researcher, believing that it is important to look at a range of clothing options, might decide that several conditions of the independent variable are needed—formal business attire, informal business attire, and business casual clothing, for example—and design an experiment that examines subjects' reactions to all three levels of formality. Or, based on the research literature, the experimenter might conclude that formal business attire has already been demonstrated to result in better treatment in the workplace and decide to examine the limits of this conclusion. Accordingly, he or she might design an experiment in which subjects are exposed to people variously wearing black suits, charcoal gray suits, and navy suits with white shirts or blouses to see if there is any difference in the way that the subjects react. At some point, however, the researcher will have to limit the definitions of the variables to a manageable number, which is done in part by determining which inferential statistical techniques are available to analyze the data.

Constructing a Hypothesis

After the variables are identified and defined, the researcher will develop a formal hypothesis for the experiment that can be analyzed with inferential statistics. For this purpose, hypotheses are stated in two ways. The first of these is called the null hypothesis (H0), which is a statement that asserts that there is no statistical difference between the status quo and the experimental condition. In other words, the null hypothesis states that the manipulation of the independent variable being studied made no difference on the dependent variable, or the subjects' responses. For example, a null hypothesis might state that there is no difference between the way people in the workplace react to people who wear dark suits and the way they react to people who wear business casual clothing. In addition, the researcher will develop an alternative hypothesis (H1) that states that there *is* a relationship between the two variables—for example, that people tend to be more respectful in the workplace of people who are wearing dark suits.

Designing an Experiment

Once the null hypothesis has been formulated, an experimental design is developed that allows the researcher to empirically test the hypothesis. Typically, the experimental design includes a control group that that does not receive the experimental condition and an experimental group that does receive the experimental condition. The presence of a control group helps minimize the influence of the extraneous variables and determine how accurately the data collected from the experimental group describes the relationship between the independent and dependent variables.

For example, if one wanted to determine whether or not hearing a political candidate's speech changed people's minds about that candidate, the researcher could divide a sample of people into two groups and collect data about their initial opinions regarding the candidate. One of the groups would then hear the speech, and afterwards, all of the subjects would be asked for their opinions again. If the opinions of the control group, or the group that did not hear the speech, did not change and the opinions of the experimental group, or the group that did hear the speech, did change, then the researcher may be able to conclude that the speech was influential. However, if both groups show a similar change in opinion, then the change is more than likely due to something other than the speech, since the control group was not exposed to the speech.

After running the experiment, the researcher then collects data from the people in the study to determine whether or not the experimental condition had any effect on the outcome. Once the data have been collected, they are statistically analyzed to determine whether the null hypothesis—that there is no difference between the control and experimental groups—should be accepted or rejected. By accepting the null hypothesis, the researcher is concluding that the independent variable had no effect on the dependent variable (e.g., that a political speech had no effect on the people who listened to it, or that the way people dress in the workplace does not affect the way they are treated). If, on the other hand, it is found that the results of the data analysis are statistically significant, the researcher will conclude that it is probable that the difference observed between the experimental and control groups is due not to chance but to a real underlying relationship between the independent variable and the dependent variable.

Applications

Part of research design and hypothesis development is determining how the research data will be statistically analyzed. It is important to note that the design of the experiment limits one's choices of how to analyze the data, and the researcher must determine which statistical tools will be used to analyze the data before data collection begins so that he or she can be assured that all the necessary information will be collected for analysis. In most cases, it is impossible to go back and collect additional data, meaning that the research study would need to be performed again from the beginning.

Inferential Statistics

In research studies, inferential statistics are used to test hypotheses to determine if the results of a study occur at a rate that is statistically significant, meaning that they are unlikely to be due to chance. There are a number of statistical methods for testing hypotheses, each of which is appropriate to a different type of experimental design. One commonly used class of statistical tests is the various *t*-tests. These tests are used to analyze the mean of a population or compare the means of two different populations. Another frequently used technique for analyzing data in applied settings is analysis of variance (ANOVA). This family of techniques is used to analyze the joint and separate effects of multiple independent variables on a single dependent variable and determine the statistical significances of the effects. For example, analysis of variance might be used if one wished to examine the differences between research subjects' reactions to people wearing black suits, navy suits, and grey suits. For more complicated situations, multivariate analysis of variance (MANOVA), an extension of this set of analysis of variance, allows researchers to test hypotheses involving the simultaneous effects of multiple independent variables on multiple dependent variables. Other inferential statistical tests include correlation, which determines the degree to which two variables are related, and regression analysis, which is used to build models of complex real-world data. Again, which statistical tool is used depends on the kinds of data available and the hypothesis the researcher is testing.

Because not every experimental situation in the behavioral and social sciences yields neat or ideal data, inferential statistical tools such as *t*-tests and analysis of variance are called parametric statistical tools, meaning that they make certain assumptions about the underlying distribution of the data they analyze. For instance, these tools assume that the measurement scale used to articulate the data has a meaningful zero point and intervals of equal size.

Fortunately, researchers do not need to misuse parametric statistics or forgo statistical analysis completely in situations where data do not meet the assumptions underlying parametric statistics. A number of nonparametric procedures that correspond to common parametric tests and do not make assumptions about the underlying distribution can be used when the shape and parameters of a distribution are known. These statistical tools are not as powerful as standard parametric statistics, but for situations in which the data set is less than perfect, they do allow the analyst to derive meaningful information.

Conclusion

In order for any branch of science to advance, it is necessary to conduct rigorous empirical research that adheres to the principles of the scientific method: observation and review of the literature, formulation of a testable hypothesis, selection of a research design, data collection and analysis, extrapolation of conclusions, and development of ideas for further research in the area.

In the simplest research design, a stimulus is presented to the research subjects. The responses of the subjects are then observed and recorded. The researcher needs to determine which independent and dependent variables are important to the research question and then operationally define the variables so that they can be statistically analyzed and used to draw meaningful conclusions. This information is then turned into a formal hypothesis that is stated two ways: a null hypothesis, which states that the manipulation of the independent variable being studied has no effect on the dependent variable, and an alternate hypothesis, which states that the value of the independent variable does have an effect on the dependent variable. Based on this information, an experimental design is developed that allows the researcher to control any extraneous variables as much as possible, thus allowing him or her to observe changes in the dependent variable that are concurrent with changes in the independent variable. Part of the process of designing an experiment involves determining how the research data will be analyzed so that the researcher can establish the statistical significance of the results and either accept or reject the null hypothesis.

Terms & Concepts

Analysis of Variance (ANOVA): A family of statistical techniques that analyze the joint and separate effects of multiple independent variables on a single dependent variable and determine the statistical significances of the effects.

Correlation: The degree to which two events or variables are consistently related. Correlation may be positive (as the value of one variable increases, the value of the other variable increases), negative (as the value of one variable increases, the value of the other variable decreases), or zero (the values of the two variables are unrelated). Correlation does not imply causation.

Data: In statistics, quantifiable observations or measurements that are used as the basis of scientific research.

Dependent Variable: The outcome variable or resulting behavior that changes depending on whether the subject receives the control or experimental condition.

Empirical Evidence: Evidence that is derived from or based on observation or experiment.

Independent Variable: The variable in an experiment or research study that is intentionally manipulated in order to determine its effect on the dependent variable.

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data. Inferential statistics are used to make inferences, such as drawing conclusions about a population from a sample, as well as in decision making.

Mean: An arithmetically derived measure of central tendency in which the sum of the values of all the data points is divided by the number of data points.

Null Hypothesis (H0): The statement that the findings of an experiment will show no statistical difference between the control condition and the experimental condition.

Operational Definition: A definition that is stated in terms that can be observed and measured.

Sample: A subset of a population. A random sample is a sample that is chosen at random from the larger population with the assumption that it will reflect the characteristics of the larger population.

Scientific Method: The general procedures, guidelines, assumptions, and attitudes required for the organized and systematic collection, analysis, and interpretation of data that can then be verified and reproduced. The goal of the scientific method is to either articulate or modify the laws and principles of a science. Steps in the scientific method include problem definition based on observation and review of the literature, formulation of a testable hypothesis, selection of a research design, data collection and analysis, extrapolation of conclusions, and development of ideas for further research in the area.

Statistical Significance: The degree to which an observed outcome is unlikely to have occurred due to chance.

Subject: A participant in a research study or experiment whose responses are observed, recorded, and analyzed.

Bibliography

Armore, S. J. (1966.) *Introduction to statistical analysis and inferences for psychology and education*. New York: John Wiley & sons.

Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate data analysis* . New York: John Wiley and Sons.

Dallal, G. E. (2007). Nonparametric statistics. Retrieved August 20, 2007, from http://www.jerrydallal.com/LHSP/npar.htm

Hanson-Hart, Z. (n.d.). *Statistical reasoning.* Retrieved July 24, 2007, from http://www.math.temple.edu/~zachhh/ch5.pdf.

Hollander, M. & Wolfe, D. A. (1973). *Nonparametric statistical methods* . New York: John Wiley and Sons.

Huff, D. (1954). *How to lie with statistics*. New York: W. W. Norton & Company.

MacKinnon, D. P. (2011). Integrating mediators and moderators in research design. *Research on Social Work Practice, 21*(6), 675–681. Retrieved November 5, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=66817231&site=ehost-live

Mitchell, Lada. (2003, Sep). Book review: Applied multivariate statistics for the social sciences. *Journal of the Royal Statistical Society: Series D (The Statistician), 52* (3), 418-20. Retrieved August 20, 2007 from EBSCO Online Database Business Source Complete. http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=10637883&site=ehostlive

Schaefer, R. T. (2002). *Sociology: A brief introduction* (4th ed.). Boston: McGraw-Hill.

Welsh, B., Peel, M., Farrington, D., Elffers, H., & Braga, A. (2011). Research design influence on study outcomes in crime and justice: A partial replication with public area surveillance. *Journal of Experimental Criminology, 7*(2), 183–198. Retrieved November 5, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=60874161&site=ehost-live

Wendt, O., & Miller, B. (2012). Quality appraisal of single-subject experimental designs: An overview and comparison of different appraisal tools. *Education & Treatment Of Children, 35*(2), 235–268. Retrieved November 5, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=74231752&site=ehost-live

Witte, R. S. (1980). *Statistics* . New York: Holt, Rinehart and Winston.

Suggested Reading

Calfee, R. C. (1975). *Human experimental psychology* . New York: Holt, Rinehart and Winston.

Gravetter, F. J. & Wallnau, L B. (2006). *Statistics for the behavioral sciences* . Belmont, CA: Wadsworth/Thomson Learning.

Seidman, E. (2012). An emerging action science of social settings. *American Journal of Community Psychology, 50*(1/2), 1–16. Retrieved November 5, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=78190954&site=ehost-live

Webster, M. & Sell, J. (2007). *Laboratory experiments in the social sciences* . New York: Academic Press.

Young, R. K. & Veldman, D. J. (1977). *Introductory statistics for the behavioral sciences* (3rd ed.). New York: Holt, Rinehart and Winston.

Essay by Ruth A. Wienclaw, PhD

Dr. Ruth A. Wienclaw holds a doctorate in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human-systems integration.

Hypothesis Construction

Table of Contents

Abstract

Based on their observations of real world phenomena, sociologists develop theories to explain and predict the behavior of humans within society. One of the first steps in testing the validity of a theory is to develop a hypothesis. A hypothesis is an empirically verifiable declaration that certain variables and their corresponding measures are related in a specific way proposed by a theory. To be of use in testing the validity of a theory, hypotheses must be stated so that they can be tested with the tools of inferential statistics. To this end, hypotheses express the relationship between the independent and dependent variables proposed by a theory in a way that permits them to be tested to determine the statistical likelihood of the observed results being due to chance or an underlying factor.

Overview

No matter where we are or what we are doing, we are constantly bombarded with all sorts of information. Sometimes this information is relevant to our current or future activities, and sometimes it is not. For example, as I sit in my office dictating this article to my computer, I am primarily aware of the process of trying to transform my thoughts into words. However, if I pay attention, I am also aware of other experiences, too. Certainly I hear my voice as I dictate, but I can also hear my headset amplifying my voice, telling me that my voice recognition software is receiving the *data* it needs to transcribe my words. In addition, I receive other sense experiences that I am choosing to ignore at this time: the heat of the halogen lamp sitting on my desk, the sunlight streaming through my office windows, the noise of the printer as it spits out a draft copy of the article, and the warm air softly blowing from a heating vent. If I am quiet and listen carefully, I can also hear my heart beating as well as noises coming from outside my office.

Obviously, I do not care about all this information, nor can I process it all at the same time. Unless I am in danger of touching my lamp, the heat it puts off is irrelevant. My heartbeat is not important either, unless it develops an arrhythmia or other aberration. Even the sound of my voice is irrelevant as long as I hear the words in my head and they get correctly transcribed onto the computer screen. I simply cannot maintain a high degree of attention to all these sense experiences at the same time, so I ignore most of them and focus merely on the ones that are important to the task at hand.

Just as I need to pay attention to or ignore the various inputs I receive as I sit in my office, so, too we must pay attention to or ignore the various inputs we receive as we interact with others. For example, if I am having difficulty downloading an article from a database, there are many potential reasons for the problem: I may have entered an incorrect access code; I may no longer have access authorization; my computer hardware or software may be malfunctioning; the database may be experiencing a technical problem; or the host server or my Internet service provider may be experiencing a technical problem. If I am unable to troubleshoot the problem on my own, I may contact technical support to gather additional data so that I can narrow down the source of the problem. Technical support may be able to give me additional data, or point out data that I am ignoring so that, between us, we can solve the problem.

As we work together, we develop and test hypotheses. For example, our initial hypothesis might be that I have entered an incorrect access code. The technical support person could then look up my account, confirm my access code, and ask me to enter it again. If this does not work, we might formulate a new hypothesis: that I no longer have access authorization. The technical support person could then contact the department that sets up authorizations and see if I have lost mine.

As complex as this process may be, however, troubleshooting a computer problem is a relatively simple task compared to interpreting human behavior. Sociologists task themselves with this work as they constantly formulate and reformulate hypotheses based on their observations in order to describe and predict the behavior of people within society.

Applications

What is a Hypothesis?

Hypotheses are developed from the observations of a researcher or research team. For example, based on my observation that I am much more likely to receive prompt and courteous service when I go to a department store while I am wearing my business clothes than when I am wearing my old gardening clothes, I may develop the hypothesis that clerks in retail stores give differential service depending on the perceived *socioeconomic status* and *social capital* of the person they are serving.

In scientific terms, a hypothesis is more than a question. For example, I may wonder aloud whether there is any relationship between the way that I dress and the way that I am treated by a sales clerk. To be useful from a scientific point of view, however, I need to operationally define my terms so that I can get a testable answer to my question. An *operational definition* is a definition that is stated in terms that can be observed and measured. For example, "the way that I dress" is open to many interpretations. To turn my question into a hypothesis, I need to operationally define all the terms in my question. So, I might operationally define "well-dressed" to mean being clean, well groomed, and wearing business attire, and "poorly dressed" to mean being dirty, poorly groomed, and wearing old, dirty clothes. Notice that by defining the terms in this manner I have left out a number of other possible scenarios, such as wearing old but clean and mended clothes, wearing formal wear, and wearing business clothes but not being well groomed. Similarly, I need to operationally define the meaning of "good service." To do this, I might develop a series of rating scales or criteria that measure the various components of service (e.g., the number of minutes it takes for the sales clerk responds to the customer standing at the counter, how much eye contact the sales clerk makes with the customer, how long the sales clerk listens before making a suggestion). Although these operational definitions (and their concomitant simplification of the original question) may not answer all the nuances of the original question, they do allow me to develop a hypothesis that I can actually test in the field.

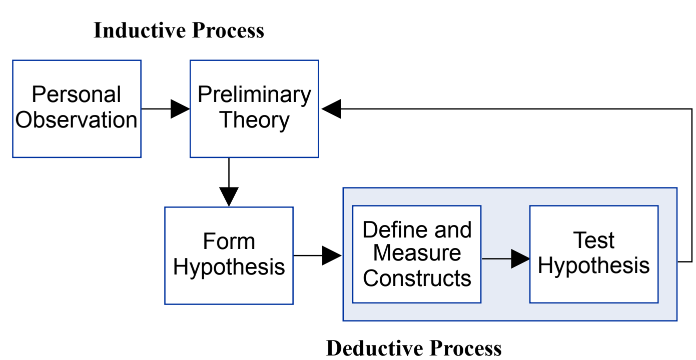
Scientifically speaking, a hypothesis is an *empirically verifiable* declaration describing the relationship between and corresponding measures of the independent and dependent variables as proposed by a theory. The *independent variable* is the variable that is manipulated by the researcher. In the example above, the independent variable is the manner in which a person is dressed (e.g., in business attire or in dirty old clothes). The *dependent variable,* so called because its value depends upon the degree of the independent variable to which the subject is exposed, is the subject's response to the independent variable (e.g., the level of service the sales clerk offers).

Null & Alternative Hypotheses

For the purposes of empirical research, a hypothesis is stated in two ways. The *null hypothesis (H0)* is the statement that there is no statistical difference between the *status quo* and the experimental condition (i.e., the treatment being studied made no difference on the end result). For example, a null hypothesis about sales clerks' responses to the way customers dress would state that there is no difference in the way sales clerks treats customers dressed in business attire and the way they treat customers dressed in dirty, old clothes. In effect, this null hypothesis states that there is no relationship between the independent variable of how people dress and the dependent variable of the level of service offered. The *alternative hypothesis (H1),* on the other hand, states that there *is* a relationship between the two variables (e.g., that sales clerks give better service to customers wearing business attire).

As shown in Figure 1, hypothesis construction and research design start with a theory that is based on real-world observation. To find out if this hypothesis is true, the researcher next needs to operationally define the various terms (i.e., *constructs*) in the hypothesis. The researcher would then run an experiment to test the hypothesis.

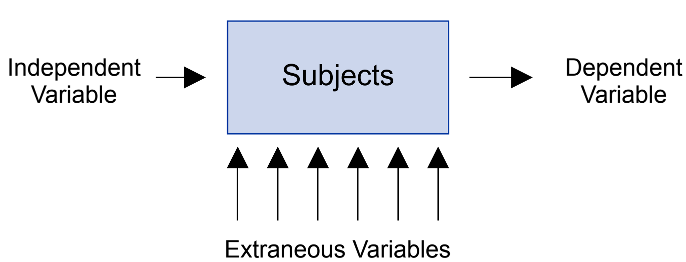
Figure 1: The Theory Building Process



In the simplest research design, a *stimulus* (e.g., a "customer" wearing business attire or dirty old clothes) is presented to the research subjects (e.g., sales clerks). The response of the subjects to the stimulus is observed and recorded (e.g., what level of service they gave the "customer"). There are three types of variables that are important in research. As discussed above, the variables of most concern in the design of a research study are the independent and the dependent variables. However, as shown in Figure 2, these are not necessarily the only variables that need to be controlled during a study. *Extraneous variables,* or variables that have nothing to do with the independent variable itself, can also affect the outcome of the experiment (e.g., the level of service given to the customer). To make my experiment valid, I need to define and control these variables.

For example, if a sales clerk has just dealt with a difficult customer or had a negative interaction with her boss, it is likely that the negative attitude created by that previous interaction will carry over to the next interaction. This transfer would be particularly likely if the person with whom the clerk just had a negative encounter was wearing clothes similar to those worn by the *research confederate.* Any number of extraneous can variables can affect the outcome of the research and lead to an erroneous interpretation of the results. Therefore, as much as possible, these variables need to be controlled. For example, the experiment could be set up so that the confederate would be the clerk's first customer of the day, or could only approach the clerk after he or she had been free for 10 minutes. Although it is impossible to control every possible extraneous variable—for instance, being the clerk's first customer does not rule out negative interactions the clerk may have had at home or while driving to work—the more of these variables that are accounted for and controlled in the experimental design, the more meaningful the experiment's results will be.

Figure 2: Research Variables



One of the reasons that researchers use hypotheses with operationally-defined variables rather than just asking general questions is so that they can statistically determine if the results they observe are due to some underlying factor or just chance. Used correctly, statistical analysis can help researchers determine if there is a relationship between the independent and dependent variables not only within the relatively restricted *sample* on which the research was based, but also, and more importantly, within the larger *population* of which the sample is assumed to be representative.

Analyzing Data

Statistical tools make certain assumptions about the nature of the data and their underlying *distribution.* As a result, not every statistical technique is appropriate for use with every set of data. Further, as discussed at the beginning of this article, the world is a complex place and the relationship between an observed result (behavior) and the stimulus or stimuli that caused it can be complex. Although multivariate statistical tools can be used in some complex situations, they, too, are limited in what they can do. Therefore, designing a good research study depends in part on two factors: controlling the situation so that the research is only measuring what it is supposed to measure, and including as many of the relevant factors as possible so that the research scenario accurately emulates the real world experience.

Conclusion

A hypothesis is an empirically verifiable declaration describing the relationship between and corresponding measures of the independent and dependent variables as proposed by a theory. In sociology, hypotheses are used to transform questions about the behavior of people in groups or societies into testable research designs that can be statistically analyzed to determine the probability of the observed results being due to an underlying factor or to chance. Hypotheses employ the use of operational definitions that are stated in terms that can be observed and measured. For purposes of scientific research, hypotheses are stated two ways. The null hypothesis is the formal statement that the findings of an experiment will show no statistical difference between the current condition, or control condition, and the experimental condition. The alternative hypothesis is the formal statement that there is a statistical difference between the two conditions. The development of a good research hypothesis must take into consideration not only the independent and dependent variables that are of interest, but also any extraneous variables that may affect the resultant behavior but are not directly related to the research question. In addition, a hypothesis must be stated in such a way that it can be mathematically analyzed to determine the statistical significance of the observed results.

Terms & Concepts

Confederate: A person who assists a researcher by pretending to be part of the experimental situation while actually only playing a rehearsed part meant to stimulate a response from the research subject.

Data: (*sing.* datum) In statistics, data are quantifiable observations or measurements that are used as the basis of scientific research.

Dependent Variable: The outcome variable or resulting behavior that changes depending on whether the subject receives the control or experimental condition (e.g., a consumer's reaction to a new cereal).

Distribution: A set of numbers collected from data and their associated frequencies.

Empirical: Theories or evidence that are derived from or based on observation or experiment.

Hypothesis: An empirically verifiable declaration describing the relationship between and corresponding measures of the independent and dependent variables as proposed by a theory.

Independent Variable: The variable in an experiment or research study that is intentionally manipulated in order to determine its effect on the dependent variable (e.g., the independent variable of type of cereal might affect the dependent variable of the consumer's reaction to it).

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data. Inferential statistics are used to make inferences, such as drawing conclusions about a population from a sample. Inferential statistics can also be used in decision-making.

Null Hypothesis (H0): The statement that the findings of the an experiment will show no statistical difference between the current condition, or control condition, and the experimental condition.

Operational Definition: A definition that is stated in terms that can be observed and measured.

Population: The entire group of subjects belonging to a certain category (e.g., all women between the ages of 18 and 27; all dry cleaning businesses; all college students).

Sample: A subset of a population. A random sample is a sample that is chosen at random from the larger population with the assumption that such samples tend to reflect the characteristics of the larger population.

Social Capital: The resources or benefits that people gain from the connections within and between their social networks.

Socioeconomic Status (SES): The position of an individual or group on the two vectors of social and economic status and their combination. Factors contributing to socioeconomic status include (but are not limited to) income, type and prestige of occupation, place of residence, and educational attainment.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables, or response variables. Extraneous variables are variables that affect the dependent variables but that are not related to the question under investigation in the study.

Bibliography

Black, K. (2006). *Business statistics for contemporary decision making* (4th ed.). New York: John Wiley & Sons.

Calderwood, K. A. (2012). Teaching inferential statistics to social work students: A decision-making flow chart. *Journal of Teaching in Social Work, 32* (2), 133–147. Retrieved November 4, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=74639228&site=ehost-live

Cohen, E. H., & Tresser, C. (2011a). Matrix assisted structural hypothesis construction. *BMS: Bulletin de Methodologie Sociologique (Sage Publications Ltd.),* (109), 5–19. Retrieved November 4, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=61825831&site=ehost-live

Cohen, E. H., & Tresser, C. (2011b). Matrix assisted structural hypothesis construction: Further explorations. *BMS: Bulletin de Methodologie Sociologique (Sage Publications Ltd.),* (112), 63–70. Retrieved November 4 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=67690797&site=ehost-live

Knuttila, K. M., & Magnan, A. *Introducing sociology: A critical approach*.5th ed. Oxford, UK: Oxford University Press.

Witte, R. S. (1980). *Statistics*. New York: Holt, Rinehart and Winston.

Suggested Reading

Anderson, M. L. & Taylor, H. F. (2002). *Sociology: Understanding a diverse society* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Feld, S. L. (1997, Mar). Mathematics in thinking about sociology. *Sociological Forum, 12* (1), 3–9. Retrieved March 13, 2008 from EBSCO Online Database Academic Search Complete http://web.ebscohost.com/ehost/pdf?vid=11&hid=17&sid=a3878266-c212-4e78-95f9-556697cc9da2%40sessionmgr102

Gravetter, F. J. & Wallnau, L B. (2006). *Statistics for the behavioral sciences.* Belmont, CA: Wadsworth/Thomson Learning.

Saetnan, A. R., Lomell, H. M., & Hamer, S. (Eds.) (2011). *The mutual construction of statistics and society.* New York, NY: Routledge.

Schaefer, R. T. (2002). *Sociology: A brief introduction* (4th ed.). Boston: McGraw -Hill.

Stockard, J. (2000). *Sociology: Discovering society* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Welkowitz, J., Cohen, B. H., & Lea, R. B. (2012). *Introductory statistics for the behavioral sciences.* 7th ed. Hoboken, NJ: John Wiley & Sons.

Young, R. K. & Veldman, D. J. (1977). *Introductory statistics for the behavioral sciences* (3rd ed.). New York: Holt, Rinehart and Winston.

Essay by Ruth A. Wienclaw, Ph.D.

Dr. Ruth A. Wienclaw holds a doctorate in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human/systems integration.

Field Data Collection

Table of Contents

Abstract

Data for behavioral research can be gathered in a number of ways. Although in experimental paradigms researchers have a great deal of control, the results of such studies often have limited generalizability and cannot account for the great complexity of variables experienced in the real world. On the other hand, field data collection techniques including field observation, field research, and unobtrusive measures offer the researcher little or no control but are rich sources of information about the way people actually act in the real world. The data gathered using these methods enable researchers to apply inductive reasoning to real world data so that the variables contributing to behavior can be better understood and testable hypotheses can be formulated. Field data collection tools are an important part of the behavioral scientist's toolbox and can add to our understanding of human behavior in significant and important ways.

Overview

Research in the behavioral and social sciences can be both fascinating and challenging. Not only does social science research require the application of the scientific method, it also requires the use of a great deal of creativity in order to obtain data concerning intangible constructs, phenomena that change when they are directly observed, or complex situations affected by numerous extraneous variables not directly related to the hypothesis being tested. The complexity of social science research was vividly and hilariously illustrated by a recently published cartoon. Captioned simply, "If Einstein had been a social scientist…," the cartoon showed the back of a wild-haired individual scribbling furiously on a blackboard. Rather than the expected E=MC2 equation, however, the concept of "relativity" was expressed in terms of the square root of parents multiplied by sibling rivalry, the square of in-laws divided by the first marriage, and numerous other nonsensical terms. The resultant equation filled the blackboard while the researcher continued to articulate variables and relationships that needed to be considered.

In the behavioral and social sciences, there often seems to be a plethora of variables that need to be taken into consideration in the quest to understand and predict behavior. As a result, the findings of research studies frequently bring up more questions than they answer. For example, a researcher might want to determine the relationship between the time it takes to readjust after the death of a spouse and the time that the couple had been married. This is a simple enough relationship at first glance, at least until one takes into account other considerations: How dependent had the spouses been on each other? Does the surviving spouse have a strong support network of family and friends that can help in the readjustment period? Does the surviving spouse have a strong religious faith? Did the couple have children who also survived? Was this a first marriage? If not, what caused the end of the first marriage? The list of possible variables other than length of the marriage that might also have an effect on the outcome of the relationship is seemingly endless. Even if a researcher could articulate all the major factors to be considered and design a research paradigm that could be analyzed using inferential statistics, he or she would still face an ethical problem in collecting the research. It would simply be unethical to randomly assign married people to the various experimental conditions and then manipulate whether or not their spouses lived.

Survey Research

Because of the complexity of social science issues and the ethical considerations in the treatment of experimental subjects, social and behavioral scientists are often required to be very creative in the operational definition of their variables and concomitant data collection methods. One way to collect data is through survey research, in which data about the opinions, attitudes, or reactions of the members of a sample are gathered using a survey instrument that is administered in a paper-and-pencil (or electronic) form or by an interviewer. Surveys have the advantage of being able to collect information on non-tangible constructs (e.g., feelings, attitudes, and opinions of subjects) that are difficult to collect directly. Further, surveys can be relatively inexpensive to administer, as they require no manipulation of variables and have relatively low costs associated with data collection. However, even well-written surveys cannot provide the researcher with the answers to all questions of interest. What one says and what one does are often two different things. Research subjects may lie on a survey instrument in order to look good to the researcher (or even to themselves) or give responses that are not well thought out because they are not motivated to participate honestly in the survey. On the other hand, subjects' actions in response to the manipulation of an independent variable tend to reflect their real reaction. However, in addition to situations where it is unethical to manipulate variables to measure subjects' reactions, there are also instances where the mere fact that a subject knows that a researcher is watching may change his or her behavior.

Further, survey instruments do not yield the same kind of neat interval or ratio data that are gathered in most experimental designs. This makes the data problematic to analyze. Many commonly used inferential statistical tools assume that the data being analyzed have been randomly selected from a population that has a normal distribution and require data that are interval or ratio in nature. This means that not only do the rank orders of the data have meaning (e.g., a value of 6 is greater than a value of 5) but so do the intervals between the values. In the physical sciences, such assumptions are typically easy to meet: it is clear that the difference between 1 gram of a chemical compound and 2 grams of a chemical compound is the same as the difference between 100 grams of the compound and 101 grams of the compound. Physical measurements have meaning because the weight scale has a true zero (i.e., we know what it means to have 0 grams of the compound) and the intervals between values are equal. However, it may not be quite as clear that the difference between 0 and 1 on a 100 point attitude scale is the same as the difference between 50 and 51 or between 98 and 99. These are value judgments, and the scale may not have a true zero. For example, the scale may go from 1 to 100 and not include a 0. Similarly, even if the scale does start at 0, it may be difficult to define what this value means. It can be difficult to articulate how a score of 0 on this scale differs significantly from a score of 1 or even what a score of zero means (e.g., does a score of zero mean that the person has no opinion?). In addition, one cannot tell from the scores why the subject assigned the values, a question often of interest to social scientists. Even if the various points on the scale were well-defined, different people may give vastly different responses to indicate the same attitude. Ratings are also subjective, and although numerical values may be assigned to them, they do not necessarily meet the requirement of parametric statistics that the data can be at the interval or ratio level.

Field Research and Observations

One way to collect better data in some situations is through the observation of subjects in a real world setting. This can include field research, the collection of unobtrusive measures, and field observations. Like experiments and surveys, these methods also have advantages and disadvantages. However, they provide the researcher with additional tools for data collection that can aid in the quest to understand and predict behavior. Like all research tools, these methods should be selected only after careful consideration of what data are needed, what the practical and ethical limitations are in collecting the data, and what the statistical limitations are for analyzing the data. Although field research tools offer the researcher less control than laboratory experiments and simulations, they have the advantage of allowing subjects to be observed in a natural setting where the intrusion of the experimenter is unlikely to be noticed.

The Uses and Drawback of Field Research

Field observation and research typically allow the researcher no control over the experimental situation. Therefore, this approach to data collection is often considered inferior to other methods. However, it must be remembered that it is through the application of inductive reasoning to individual observations in the real world that testable hypotheses are generated. This is not only one of the first steps in the scientific method, it is also an essential step without which more controlled data collection could not be conducted. Further, because of the complexity of human behavior in real world situations, it is often beneficial to observe people in field settings in order to better understand the interaction of variables causing their behavior. On the one hand, field observation and research frequently do not yield high quality, quantitative data that can be statistically analyzed. On the other hand, carefully controlled experimental research often restricts the operational definitions of variables used in the study to the extent that the results are far removed from the real world. It is important to note that both these approaches are useful tools when appropriately used.

Participant and Non-Participant Observation

Field research and observation constitute a set of data collection tools that allows researchers to directly observe behavior in natural, real-world situations. Perhaps the simplest of all methods of data collection is to merely observe subjects acting naturally in a real-world setting. This can be done with the researcher acting either as a participant or as a non-participant in the situation. For example, a researcher who is interested in how police officers treat suspects from arrest through arraignment could gather data in several ways. The researcher might develop a questionnaire that could be given to suspects asking them to rate how they were treated by the arresting and booking officers. Although this approach has the advantage of being able to potentially gather information about the subjects' attitudes, emotions, and opinions, it also presents subjects with a tempting venue to lie or exaggerate in order to appear more sympathetic. The researcher could also gather similar data through field observation. In a participatory observation paradigm, the researcher might train at the police academy to learn how to arrest and book a suspect, or the researcher might allow himself or herself to be arrested and observe what happens during the process. These scenarios, however, are not without difficulty. First, it is unlikely that a police department would allow a behavioral researcher to train at the academy and actually participate in arrest and booking procedures if he or she were not a police officer. Similarly, the researcher who poses as a suspect is unlikely to be able to be objective during the arrest and booking process. Further, in such a scenario, it would be virtually impossible to take notes on the behaviors of others or the reactions of oneself without alerting the subjects in the experiment to the fact that they were being observed. In this instance, non-participatory observation might produce more realistic data. The researcher could do a ride-along with police officers, taking notes on their interactions with the suspects while observing their behavior. The problem with this approach, however, is that the police officers—the subjects of the observation—would know that they were being observed and, therefore, might not act naturally.

Unobtrusive Research

For such situations, another approach to real-world data collection is available. Unobtrusive research is an approach in which the researcher collects data without directly interfacing with or talking to the subjects. The purpose of collecting unobtrusive measures is to create a situation in which valid data can be collected in a non-reactive way so that subjects behave naturally. There are a number of approaches to collecting unobtrusive measures. First, one might look at the various physical traces resulting from human behavior, including both erosion and accretion. For example, a study of the relative popularity of museum exhibits might compare the wear patterns in the linoleum tile surrounding the various exhibits. Floors around exhibit areas that show the most wear could be interpreted to be the most popular.

Another example of using physical traces to gather data is the practice of looking for information by going through subjects' trash. This approach might be taken in a situation where the researcher wants to gather information on how much milk is being given to children in a family but is dubious about the likelihood of the parents answering the question truthfully. Trash cans filled with beer bottles but no milk cartons over a series of days would be a good indication that no matter what the parents say, they are spending their money on beer rather than milk. Another way that data can be unobtrusively gathered is through an analysis of existing records. For example, researchers could review and analyze data found in actuarial records, political or judicial records, government records, the mass media, or any other hard copy or electronic document. Like the collection of data through the examination of physical traces, the analysis of data found in archives does not require the researcher to interact with the subject. However, archival information is limited: it is not always possible for researcher to find the archival data necessary to investigate the research question.

Another way to collect unobtrusive data on behavior is through the use of hidden hardware and controls. For example, a research situation could be set up to observe the behavior of a subject behind a one-way mirror. The subject's reaction could be recorded by the researcher on the other side of the mirror, or a microphone or video camera could be unobtrusively placed in the room in such a way that the subject does not notice. This would allow researchers not only to collect data unobtrusively but also to record it objectively for further analysis. Unobtrusive research techniques allow the observation of sensitive situations and events or ones in which the introduction of the researcher might change the situation. However, unobtrusive research can be far removed from normal situations and does not necessarily allow the researcher to collect all the data needed.

Applications

Case Study: Observing and Analyzing Graffiti

In a real world example of field research using unobtrusive methods, Kiofas and Cutshall (1985) looked at institutional cultures in a closed juvenile facility. Based on review of the literature, the authors determined that graffiti can often be a useful unobtrusive measure in the investigation of social and cultural phenomena. The data collected in this study consisted of 2,765 discrete pieces of transcribed graffiti from the walls of an institution for juvenile delinquents. The graffiti analyzed were found in 95 general population rooms. Two teams of research assistants transcribed the graffiti along with identifying data, including the location of the room and the wall on which each graffito was found. The researchers defined discrete units of graffiti using criteria such as handwriting, subject matter, and the writing tool used. Chained responses to the graffiti left by another were considered to be discrete units of graffiti. After transcription, the graffiti were grouped into 13 categories with a high degree of inter-rater reliability. The quality of the graffiti data was limited by the fact that in some of the rooms, the writings were so old that they could not be read and, therefore, were not included in the data analysis. In addition to analyzing the graffiti, the authors also looked at other sources of data, including official reports from the last years that the institution was open, related newspaper articles from the largest local newspaper, and interviews with former staff members and inmates of the facility. This supplementary information helped researchers better interpret the graffiti. The combination of analysis of the graffiti and other sources of information helped the authors better understand the kinds of various influences on the juvenile legal system. Newspaper reports written from a liberal point of view, for example, frequently focused on the impact of brutal incarceration conditions on naïve juveniles. On the other hand, interviews with conservative ex-staff who had worked at the detention center tended to portray the former inmates as both sophisticated and dangerous criminals.

The authors found that the analysis of the graffiti, supplemented by interviews and newspaper accounts, was useful in reconstructing the lives of inmates at that particular juvenile detention center. Many of the items of graffiti showed the significance of having peer support and belonging to an identifiable group within the facility. An analysis of the contents of the graffiti in different corridors in the institution also demonstrated the importance of time served by the inmates and the severity of deprivations that they endured. For example, early in an inmate's incarceration period, graffiti tended to be concerned with individuality and identity. As time served progressed, these concerns gave way to other concerns, in particular great antagonism towards authority figures.

Analysis of the graffiti allowed the researchers to gain insight into the juvenile institution in a way that interviews, surveys, or short-term observation alone could not. However, as with any unobtrusive measure, the authors also admit the limitations of this type of data in assessing institutional culture. Although the graffiti were a rich source of information that would have otherwise been unavailable to the researchers, they did not provide in-depth information in the same way that the interviews did. In addition, the authors pointed out that both observations and interview studies of prison inmates have significant problems of reliability and validity not encountered in the analysis of graffiti. Although graffiti does not provide a sufficient stand-alone measure for understanding the culture within a juvenile detention institution, it provides invaluable insights into this cultural situation.

Conclusion

Because of the complexity of social science issues and the ethical considerations in the treatment of human subjects, social and behavioral scientists often need to be quite creative in developing methods for collecting research data. Although research experiments allow researchers to control the variables in the study, the need to operationally define and restrict the experimental paradigm so that it can be statistically analyzed can lead to conclusions with very limited applications. Surveys can gather more in-depth information, but they are susceptible to problems related to scaling and lack of objectivity. Field observation, field research, and unobtrusive measures give the researcher little control but enable the application of inductive reasoning to real world data so that the variables contributing to behavior can be better understood and testable hypotheses can eventually be formulated. Like more controlled experimental tools, field tools have their place in the behavioral scientist's toolbox and can add to our understanding of human behavior in significant and important ways.

Terms and Concepts

Data: (*sing.* datum) In statistics, data are quantifiable observations or measurements that are used as the basis of scientific research.

Ethics: In scientific research, a code of moral conduct regarding the treatment of research subjects that is subscribed to by the members of a professional community. Many professional groups have a specific written code of ethics that sets standards and principles for professional conduct and the treatment of research subjects.

Field Observation: An approach to data collection in which the researcher directly observes behavior, experiences, and phenomena in the settings in which they naturally occur. Although field observation can provide in-depth insight the researcher might not otherwise be able to obtain, it often involves only a limited number of cases, making findings difficult to generalize.

Hypothesis: An empirically testable theory that certain variables and their corresponding measure are related in a specific way.

Inductive Reasoning: A type of logical reasoning in which inferences and general principles are drawn from specific observations or cases. Inductive reasoning is a foundation of the scientific method and enables the development of testable hypotheses from particular facts and observations.

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data. Inferential statistics are used to make inferences, such as drawing conclusions about a population from a sample, as well as in decision making.

Operational Definition: A definition that is stated in terms that can be observed and measured.

Scientific Method: General procedures, guidelines, assumptions, and attitudes required for the organized and systematic collection, analysis, interpretation, and verification of data that can be verified and reproduced. The goal of the scientific method is to articulate or modify the laws and principles of a science. Steps in the scientific method include problem definition based on observation and review of the literature, formulation of a testable hypothesis, selection of a research design, data collection and analysis, extrapolation of conclusions, and development of ideas for further research in the area.

Subject: A participant in a research study or experiment whose responses are observed, recorded, and analyzed.

Survey: (a) A data collection instrument used to acquire information on the opinions, attitudes, or reactions of people; (b) a research study in which members of a selected sample are asked questions concerning their opinions, attitudes, or reactions and their responses are recorded for purposes of scientific analysis, the results of which are typically used to extrapolate the findings from the sample to the underlying population; (c) to conduct a survey on a sample.

Survey Research: A type of research in which data about the opinions, attitudes, or reactions of the members of a sample are gathered using a survey instrument. The phases of survey research are goal setting, planning, implementation, evaluation, and feedback. Unlike experimental research, survey research does not allow for the manipulation of an independent variable.

Unobtrusive Research: An approach to data collection in which the researcher collects data without directly interfacing with the subjects. Unobtrusive research techniques allow the observation of sensitive situations and events or situations in which the presence of the researcher changes the situation. However, unobtrusive research is often far removed from normal situations.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables (response). Extraneous variables are variables that affect the response but are not related to the question under investigation in the study.

Bibliography

Kiofas, J. M. & Cutshall, C. R. (1985, Win). The social archeology of a juvenile facility: Unobtrusive methods in the study of institutional cultures. *Qualitative Sociology, 8* (4), 368-387. Retrieved April 15, 2008, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=10871953&site=ehost-live

Savage, M., & Silva, E. B. (2013). Field Analysis in Cultural Sociology. Cultural Sociology, 7(2), 111–126. Retrieved October 25, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=87909126&site=ehost-live

Stockard, J. (2000). *Sociology: Discovering society* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Sutton, J. (2011). An ethnographic account of doing survey research in prison: Descriptions, reflections, and suggestions from the field. Qualitative Sociology Review, 7(2), 45–63. Retrieved October 25, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=65288230&site=ehost-live

Waller, L. (2013). Interviewing the surveyors: Factors which contribute to questionnaire falsification (curbstoning) among Jamaican field surveyors. International Journal of Social Research Methodology, 16(2), 155–164. Retrieved October 25, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=85750329&site=ehost-live

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences.* Chicago: Rand McNally College Publishing Company.

Suggested Reading

Cochran, P. A. L., et al. (2008, Jan). Indigenous ways of knowing: Implications for participatory research and community. *American Journal of Public Health, 98* (1), 22-27. Retrieved April 15, 2008, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=28804932&site=ehost-live

Haer, R., & Becher, I. (2012). A methodological note on quantitative field research in conflict zones: Get your hands dirty. International Journal of Social Research Methodology, 15(1), 1–13. Retrieved October 25, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=67698514&site=ehost-live

Taylor, B. W. K. (2006, Jul). A feminist critique of Japanization: Employment and work in consumer electronics. *Gender, Work and Organization, 13* (4), 317-337. Retrieved April 15, 2008, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=20857183&site=ehost-live

Wood, E. J. (2006). The ethical challenges of field research in conflict zones. *Qualitative Sociology, 29* (3), 373-386. Retrieved April 15, 2008, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=22172175&site=ehost-live

Essay by Ruth A. Wienclaw, Ph.D.

Dr. Ruth A. Wienclaw holds a doctorate in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human/systems integration.

Surveys in Sociology Research

Table of Contents

Abstract

Ethical and practical considerations in applied research with human beings often mean that researchers are unable to experimentally manipulate independent variables to determine their effects. In such situations, survey research methodology allows researchers to gather and analyze data about phenomena of interest in order to help them better understand and explain the world around them. In survey research, participants are asked questions concerning their opinions, attitudes, or reactions through a structured data collection instrument for purposes of scientific analysis. These results are used to extrapolate the findings from the sample to the underlying population. Although there are a number of advantages to using survey research for data collection from human beings, there are also many disadvantages. Typically, survey research should be used only in those situations where data cannot be collected in other ways.

Overview

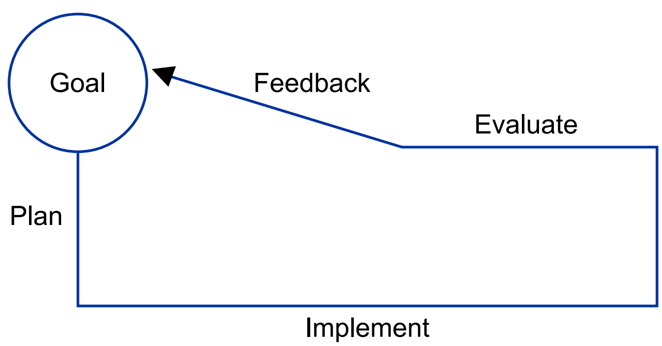
Applied research with human beings such as the kind that is done in many sociology studies precludes the manipulation of variables or random assignments to experimental groups for ethical reasons. For example, if one wanted to know the comparative effects of short-term and long-term unemployment on people, it would be unethical to randomly assign people to groups, fire one group from their jobs, and preclude from acquiring employment for a given period of time. Not only would such a study completely or partially stop the income for the persons in the experimental group, put their health and safety at risk due to potential inability to purchase food and shelter, and inversely impact their families, it would also result in various levels of psychological stress that could negatively impact them for the foreseeable future even once they were employed again. For this reason, survey research is often used to collect data from individuals already in whatever situation is of interest to the researcher. Survey research does not require the artificial external manipulation of variables (i.e., the experimenter has no control over who loses their job or how long they stay unemployed), but collects data from individuals who are already in the population of interest due to other factors (i.e., have already lost their jobs outside the scope of the research).

In general, a survey is a data collection instrument used to acquire information on the opinions, attitudes, or reactions of people. Examples of survey research are all around us and at some point in most of our lives we will participate in a study that uses surveys. The market research done at the mall where people are asked to participate in a blind taste test of two kinds of cola and answer a questionnaire specifying which they preferred and why, is a simple type of survey research. The survey run by the United States Census Bureau every 10 years to collect data about the homes and lifestyles of people across the country represents a more sophisticated type of survey research. Even the questionnaire "Is Your Spouse a Louse?" in a popular women's magazine is a type of survey (although it is not used as part of the survey research methodology and the research analysis and extrapolation is done just by the person responding to the instrument).

Survey research is a type of research in which data about the opinions, attitudes, or reactions of the members of a sample are gathered using a survey instrument. As opposed to experimental research, survey research does not allow for the manipulation of an independent variable. Survey research is a research study in which members of a selected sample are asked questions concerning their opinions, attitudes, or reactions are gathered using a survey instrument or questionnaire for purposes of scientific analysis; typically the results of this analysis are used to extrapolate the findings from the sample to the underlying population.

As shown in Figure 1, when used in scientific research, survey research follows the same general paradigm as any hypothesis testing or theory building process. Using the example above concerning the effects of unemployment on individuals and families, the researchers need to first determine what the goals of their study are. There are, for example, a wide range of consequences that someone may experience as a result of a period of unemployment and the concomitant lack of income, including inability to pay bills, loss of retirement savings and investments, change in diet (due to inability to buy the same types of food), loss of the esteem of others, feelings of inadequacy due to the inability to provide for one's family, and embarrassment from having to borrow money or sell possessions. Some of these results compound each other. The loss of self-esteem resulting from loss of employment, for example, may mean that the person has less self-confidence and does not present him/herself well in an interview, resulting in greater difficulty finding a job. The researchers need to determine which of these or other possible consequences of job loss they wish to investigate.

Figure 1: Survey Research Methodology



For example, the researchers may decide that they want to investigate both the physical and psychological consequences of unemployment. These terms, however, are rather nebulous and open-ended. The next step in the survey research methodology is to plan how the desired information will be collected. Specifically, in order to develop a good data collect instrument, they need to operationally define what they mean by "physical and psychological effects" of unemployment. They may start with their own knowledge and observations and add to these insights by interviewing people who have been unemployed to see what other effects might result from long-term unemployment. Based on this information, they would develop questions that would elicit the desired information from the survey participants. This means that they various factors in which they are interested need to be operationally defined and turned into unambiguous questions or items for inclusion on the survey. For example, survey items might include questions such as:

- "How often do you feel 'blue'?"
- "Do you have difficulty sleeping at night?"
- "Do you have difficulty getting up in the morning?"
- "Have you experienced any noticeable changes in your appetite since you lost your job?"

Using the principles of good psychometric question design, the researchers would then develop a survey instrument to be given to the sample that they have selected.

In addition to designing and developing the survey and selecting a sample, one must also determine how the survey will be delivered. Surveys can either be administered in written form through hard copy questionnaires that are mailed or given to prospective participants or administered by a trained interviewer in person or over the phone. Current survey methodology may also take advantage of newer technologies by administering the questionnaire over a cell phone, e-mail, or the Internet or by using a recording or using synthesized speech to administer the questions with respondents inputting their answers by punching in numbers on a telephone keypad or speaking the numbers which are then interpreted using voice recognition technology. Once survey data are collected, they are statistically analyzed to evaluate the responses and how they affect the researchers' theory. These results are then used to refine or expand the theory as necessary and as input for further research on the topic.

There are a number of advantages to using survey research to collect data. First, survey research methodology offers researchers a good way to collect data from a large number of participants. A survey can typically be comparatively easily and inexpensively administered to hundreds of participants as opposed to research in which variables are experimentally manipulated. Further, survey research can speed collection of data; rather than waiting for time to pass between the administration of the independent variable and the measurement of the dependent variable that is required by many experimental designs, survey research tends to ask about things that have happened in the past or that are currently occurring in the respondents' lives. In survey research there is no long waiting period to get the results. In addition, in some situations survey research is the only method available for data collection. It is impossible for ethical or practical considerations to manipulate variables for various topics (e.g., unemployment, death of a spouse). Survey research, however, can be used to gather information from individuals already experiencing the independent variable so that data otherwise unavailable can be gathered and analyzed. Finally, survey research is relatively cheap and easy, particularly when delivered through newer technologies. Costs associated with travel or experimental manipulation or variables are not required.

However, survey research also has a number of drawbacks. First, the researcher has no control over the experimental condition in survey research. Neither the value of the independent variable nor extraneous or intervening variables can be controlled. The most the researcher can do is to eliminate respondents from the final subject pool. In addition, survey research yields qualitative rather than quantitative data. Although nonparametric statistical methods can be used, these are more limited in scope than their parametric counterparts. Although nonparametric analysis can work with qualitative data that have no true zero or that are not based on a meaningful interval scale (e.g., what one person means by "dissatisfied" about an item in question and what another person may mean can be two completely different things). Further, it is easy for rating errors and other types of bias to taint the data. Poorly worded or ambiguous questions, response biases (e.g., never rating something as "excellent" because the person believes that everything can be improved), or differences in the way that multiple interviewers ask questions can all lead to inconsistent data. Although this latter problem can be overcome to some extent by training of interviewers and use of structured (rather than open-ended) survey instruments, human error will always taint the interview process. In addition, in most cases those responding to the survey instrument will care less about the results than will those who are asking the questions. Although longer surveys can ask series of questions to make sure that the data of interest is gathered, survey respondents tend to lose interest and not respond as carefully and honestly in such situations as the designers would like. Finally, there are very few situations in which survey participation can be forced. As a result, most surveys have a low response rate. This often means that the resultant sample — no matter how carefully it was designed by the researchers — is self-selected and prone to bias.

Applications

Sample selection is also part of the planning process for the survey research methodology. Certainly, one could just hang around the local unemployment office and ask everyone who comes in to answer the survey. However, this approach is also a way of selecting a sample, and is not necessarily the best way to get a representative sample of unemployed persons. Most people look for jobs online or the newspaper rather than going to an unemployment office. As a result, by using only people who go to an unemployment office the researchers would be restricting the kinds of people who will be chosen for the study to those who are in the office. This sample may not adequately represent the characteristics of the underlying population and may unintentionally bias the results. For example, if people who have been out of work for an extended period of time have decided that the unemployment office is not helpful in giving the viable jobs leads, they will stop coming to the unemployment office and, as a result, will not be included in the survey. Although this may be an acceptable situation if the researchers are interested only in the reactions of the newly unemployed, it is not an acceptable situation if the researchers are interested in the reactions of the long-term unemployed. One way to help alleviate this problem would be to collect demographic data on the survey such as the length of time the person has been unemployed. By collecting this information, the researchers could eliminate any participants who participated in the survey but who did not meet the requirement for having been unemployed for a pre-specified length of time. Another way to select a sample would be to randomly select from a population of people who are likely to be unemployed such as those who are at a shopping mall in the middle of the day or from a list of people who signed up for unemployment benefits. Random selection has the advantage that it will more than likely (based on the laws of probability) be representative of the underlying population. Such random selection helps alleviate the potential problem of asking people to participate in the survey who are not actually unemployed, although it does not completely eliminate it. People at the mall, for example, may work part-time, work from home, or be having a day off; people on the unemployment office list may have already found a job while others may not be on the list because they have been unemployed too long to qualify for benefits.

Another way to select samples is through systematic sampling where the researcher selects every nth person who walks through the door of the unemployment office or every nth name on the list of those who signed up for unemployment benefits. Sometimes, however, systematic sampling results in a sample that is self-selected on certain characteristics (e.g., there is a time limit for unemployment benefits; this means that people who are unemployed for a long period of time will not be included in the sample). To help ensure that the correct proportions of different demographics are included in the sample, one could define a stratified random sample. In this approach, the characteristics of participants is determined a priori (e.g., people unemployed 1-6 weeks, 7-12 weeks, 13 or more weeks). Within each of these subgroups (i.e., strata) a sample is randomly chosen in proportion to the proportion of that strata in the underlying population. This approach helps ensure that the subgroups of interest are included in the sample, but has the potential drawback of introducing bias in some instances.

When selecting a sample, it is important not to introduce bias into the sample so that the results truly represent the characteristics of the underlying population. In statistical terms, bias is the tendency for a given experimental design or implementation to unintentionally skew the results of the experiment. Selection bias occurs when the sample is selected in a way that is not representative of the underlying population. For example, as discussed above, recruiting participants in the study only from people who are present in an unemployment office may unfairly eliminate people who have been unemployed for a longer period of time. However, even with the best of intentions and the most rigorous sampling methods on the part of the researchers, bias can still be introduced into a sample. In very few situations is it possible to actually compel people to participate in a survey. In such situations, therefore, bias can be introduced into a sample through self-selection. This is a condition when members of the sample refuse to participate in the survey. For example, when the survey is delivered by mail, participants are free to complete the survey or not (in the great majority of the cases they do not). Similarly, relatively few people agree to participate in a survey over the phone. In many cases, those who self-select out of participating in the survey may have different characteristics than those who self-select to participate (e.g., desire for privacy, embarrassment at being unemployed. As a result, the self-selected sample chosen is often biased.

Conclusion

Applied research with human beings frequently precludes the manipulation of variables or random assignment for ethical or practical reasons. However, survey research allows researchers to collect data about human behavior and opinions in most such situations. In survey research participants are asked questions concerning their opinions, attitudes, or reactions through a structured data collection instrument for purposes of scientific analysis. These results are used to extrapolate the findings from the sample to the underlying population. Although there are a number of advantages to using survey research for data collection from human beings, there are also many disadvantages. Survey research should be used only in those situations where data cannot be collected in other ways.

Terms & Concepts

Bias: The tendency for a given experimental design or implementation to unintentionally skew the results of the experiment due to a nonrandom selection of participants.

Data: (sing. datum) In statistics, data are quantifiable observations or measurements that are used as the basis of scientific research.

Demographic Data: Statistical information about a given subset of the human population such as persons living in a particular area, shopping at an area mall, or subscribing to a local newspaper. Demographic data might include such information as age, gender, or income distribution.

Ethics: In scientific research, a code of moral conduct regarding the treatment of research subjects that is subscribed to by the members of a professional community. Many professional groups had a specific written code of ethics that sets standards and principles for professional conduct and the treatment of research subjects.

Hypothesis: An empirically-testable declaration that certain variables and their corresponding measure are related in a specific way proposed by a theory.

Independent Variable: The variable in an experiment or research study that is intentionally manipulated in order to determine its effect on the dependent variable (e.g., the independent variable of type of cereal might affect the dependent variable of the consumer's reaction to it).

Operational Definition: A definition that is stated in terms that can be observed and measured.

Population: The entire group of subjects belonging to a certain category (e.g., all women between the ages of 18 and 27; all dry cleaning businesses; all college students).

Probability: A branch of mathematics that deals with estimating the likelihood of an event occurring. Probability is expressed as a value between 0 and 1.0, which is the mathematical expression of the number of actual occurrences to the number of possible occurrences of the event. A probability of 0 signifies that there is no chance that the event will occur and 1.0 signifies that the event is certain to occur.

Sample: A subset of a population. A random sample is a sample that is chosen at random from the larger population with the assumption that such samples tend to reflect the characteristics of the larger population.

Skewed: A distribution that is not symmetrical around the mean (i.e., there are more data points on one side of the mean than there are on the other).

Survey: (a) A data collection instrument used to acquire information on the opinions, attitudes, or reactions of people; (b) A research study in which members of a selected sample are asked questions concerning their opinions, attitudes, or reactions are gathered using a survey instrument or questionnaire for purposes of scientific analysis; typically the results of this analysis are used to extrapolate the findings from the sample to the underlying population; (c) to conduct a survey on a sample.

Survey Research: A type of research in which data about the opinions, attitudes, or reactions of the members of a sample are gathered using a survey instrument. The phases of survey research are goal setting, planning, implementation, evaluation, and feedback. As opposed to experimental research, survey research does not allow for the manipulation of an independent variable.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables (response). Extraneous variables are variables that affect the response but that are not related to the question under investigation in the study.

Bibliography

Arsham, H. Questionnaire design and surveys sampling. University of Baltimore Website Retrieved September 11, 2007 from: http://home.ubalt.edu/ntsbarsh/stat-data/Surveys.htm

Axinn, W., Link, C., & Groves, R. (2011). Responsive survey design, demographic data collection, and models of demographic behavior. *Demography, 48*(3), 1127-1149. Retrieved November 4, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=62909910

Lavrakas, P. J., Shuttles, C. D., Steeh, C., & Fienberg, H. (2007). The state of surveying cell phone numbers in the United States. *Public Opinion Quarterly*, 71 (5), 840-854. Retrieved March 31, 2008 from EBSCO online database Academic Search Premier: http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=28452979&site=ehost-live

Porter, J., & Ecklund, E. (2012). Missing data in sociological research: An overview of recent trends and an illustration for controversial questions, active nonrespondents and targeted samples. *American Sociologist, 43*(4), 448-468. Retrieved November 4, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=84368865

Porter, S. R. & Whitcomb, M. E. (2007, Win). Mixed-mode contacts in web surveys. *Public Opinion Quarterly*, 71 (4), 635-648. Retrieved March 31, 2008 from EBSCO online database Academic Search Premier: http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=27772652&site=ehost-live

Smith, S. N., Fisher, S. D., & Heath, A. (2011). Opportunities and challenges in the expansion of cross-national survey research. *International Journal of Social Research Methodology, 14*(6), 485-502. Retrieved November 4, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=66808571

Suggested Reading

Brick, J. M., Edwards, W. S., & Lee, S. (2007). Sampling telephone numbers and adults, interview length, and weighting in the California health interview survey cell phone pilot study. *Public Opinion Quarterly*, 71 (5), p793-813. Retrieved March 31, 2008 from EBSCO online database Academic Search Premier: http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=28452977&site=ehost-live

Brown, G., Weber, D., Zanon, D., & de Bie, K. (2012). Evaluation of an online (opt-in) panel for public participation geographic information systems surveys. *International Journal Of Public Opinion Research, 24*(4), 534-545. Retrieved November 4, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=83932619

Naidoo, K. (2008, Jan). Researching reproduction: Reflections on qualitative methodology in a transforming society. *Forum: Qualitative Social Research*, 9 (1), 1-16. Retrieved March 31, 2008 from EBSCO online database SocINDEX with Full Text: http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=29973417&site=ehost-live

Olson, K., Smyth, J. D., & Wood, H. M. (2012). Does giving people their preferred survey mode actually increase survey participation rates? An experimental examination. *Public Opinion Quarterly, 76*(4), 611-635. Retrieved November 4, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=83746284

Saliba, D. et.al. (2001, Dec). The vulnerable elders survey: A tool for identifying vulnerable older people in the community. *Journal of the American Geriatrics Society*, 49 (12), 1691-1699. Retrieved March 19, 2008 from EBSCO online database Academic Search Premier: http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=5929191&site=ehost-live

Essay by Ruth A. Wienclaw, PhD

Ruth A. Wienclaw holds a doctorate in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human/systems integration.

Sampling

Table of Contents

Abstract

It is frequently impossible to gather data from every member of a population of interest. Therefore, sociologists and other researchers typically base their studies on samples of individuals that are drawn from the population of interest. In order to be useful for research purposes, these samples need to be drawn in such a way as to minimize the probability of introducing bias into the selection process so that the resulting sample truly represents the underlying population. There are a number of ways to do this, however, the comparative efficacy of various sampling approaches remains a matter of debate.

Overview

Why Do Researchers Take Samples?

To better understand the behavior of people within society, sociologists develop theories and collect and analyze data to test the validity of those theories. In some situations, it is relatively easy to gather data about opinions, behavior, or other characteristics of interest from every member of a population. For example, if I wish to know whether or not my class prefers to write one 10-page paper or two 5-page papers, I can simply ask my students for their preferences and make the assignment based on the majority opinion. This is easy to do because the class size is relatively small. I can easily collect the data and, since their motivation to respond is relatively high, the students' are likely to participate in my survey, giving me the data I need to make my decision.

If, on the other hand, I want to determine the same preference for all students in the university, or all students in all universities across the country or across the globe, the activity becomes more complicated. First, the sheer number of students in those larger populations makes the task of collecting information from all of them both costly and time-consuming. Second, although the students in my class may be motivated to answer my question because they are directly affected by the outcome, the students in these greater populations have no such motivation. As a result, the probability of collecting data from them all is rather low. However, I cannot in good conscience extrapolate the answer of my class to university students in general, students taking all my courses, or even university students taking this particular course from other professors at my university. There are too many differences between my class and those other, larger groups to make an accurate extrapolation. Although the members of my class have characteristics in common with the members of the other, larger groups (e.g., general age range, education level), it cannot necessarily be reasonably assumed that there are not other variables (e.g., workload, expectations) that may affect the responses of members of the other groups, making their answers different from those of this particular class.

To develop theories and build our knowledge of human behavior in society, however, it is often necessary to collect data about groups of people too large to poll individually. Rather than collecting data from a manageable group that has a low likelihood of representing the population that we wish to test, we instead take a sample of individuals from the larger group using a methodology that we believe will allow us to draw a sample that reflects the characteristics of the larger population. For example, although it may be impossible to collect data from every university student across the country, it may be possible to gather data from a representative sample of the population (e.g., university students taking an introductory sociology courses in several universities that have the characteristics in which we are interested). The sample that is selected can then be used in research based on the assumption that the sample has the same characteristics as the population as a whole.

Selection Bias

It is very important that the method used to draw the sample gives us a sample that is representative of the characteristics of the population in which we are interested. Otherwise, our sample will be biased and the results of our study will not represent the results that would have been obtained from the population in general.

Selection bias occurs when the sample asked to participate in a study is selected in a way that is not representative of the underlying population. One of the classic examples of a biased sample that led to an erroneous conclusion is the Gallup Poll result following the 1948 presidential election which predicted that Thomas Dewey would beat Harry Truman. Results obtained from biased samples cannot be meaningfully extrapolated to the population at large.

Applications

Defining Sample Characteristics

Before determining the best way to draw a sample, we must first operationally define which characteristics are important in the target population. Although in some cases it is of value to just randomly interview every 15th person who walks into a shopping mall, in most cases the target population needs to be better defined. For example, if we are interested in the opinions of shoppers on whether or not they would play the latest video game, it would be better to draw our sample from those shoppers who are more likely to play the game than from shoppers in general.

Sampling Methods

Once the population in which one is interested has been operationally defined, a sample needs to be drawn from the population. There are two general approaches used to select a representative sample. The first is random sampling, in which a subset of the population is randomly chosen for the sample. Choosing names out of a hat or using a random number generator or a list of names are examples of this approach. However, although this is a widely used technique and may in many cases accurately represent the larger population because it is based on random probability, it also may be skewed to unfairly represent some characteristic. As a result, in some situations it is important to use a stratified random sample. This technique takes into account the known characteristics of the population. For example, if it is known that half the sociology students in the country are women, a researcher might randomly select 100 women and 100 men for a research study so that both genders are represented in the sample in the same proportion that they appear in the population.

Representative samples can be drawn in a number of different ways. The simplest approach to sampling is to merely randomly select people from the population through such methods as having a computer pick names at random from a list or by selecting names from a hat. These randomly chosen individuals are then assigned to the sample. Based on the laws of probability, this approach will more than likely be representative of the underlying population. However, in practice, achieving a truly random sample can be more difficult than it sounds. Written surveys, for example, tend to have notoriously low return rates, and people are frequently loath to give out information over the phone. As a result, many of the people from whom one would like to collect data take themselves out of the sample. This self-selection means that the resultant sample is not truly random. Further, the characteristics that are common to the individuals who opt out of participating in the research may be less frequently observed in the rest of the sample. This means that the sample may not represent a significant segment of the underlying population.

Another way to select samples is through systematic sampling, which determines who will be included or excluded from the sample on the basis of an *a priori* rule. For example, the researcher could select every *nth* person who walks in the door of a mall to participate in the survey. Although it is easier to select the participants using this approach, it still may not be a truly random sample depending on the self-selection that occurs through factors like what door or time of day one chooses. Another approach would be to choose a convenience sample by asking whomever looks approachable, appears to be interested in the survey, or in some other way is most convenient to survey if he or she is willing to participate in the survey. Although this approach has the advantage of making the sample easy to choose, it is also very unlikely that a convenience sample will be truly representative of the underlying population. All the participants from whom it is convenient to collect data may share one or more characteristics such as attractiveness to the person who is collecting the data, extroversion, or not being employed full time.

One approach to trying to ensure that the correct proportions of different demographics are included in the sample is through the selection of a stratified random sample. In this approach, one *a priori* determines what general characteristics one wants to include in the sample (e.g., an equal number of women and men; equal numbers of children, young adults, and adults). Within each of these subgroups (also called "strata") a sample is randomly chosen in proportion to the proportion of that stratum within the population of interest. Stratified random sampling helps one gather information about specific subgroups in the population. This approach is also more likely to yield an accurate representation of each group than are some other sampling techniques. However, stratified random sampling may also introduce bias into the selection process.

Cluster sampling is another approach to sampling that is often used in sociological research. In cluster sampling, the population is divided into non-overlapping areas (i.e., clusters) and participants are randomly selected from each area. In cluster sampling as opposed to stratified random sampling, the clusters are heterogeneous rather than homogeneous. There are several advantages to cluster sampling. First, clustering makes data more convenient to obtain by restricting the areas from which the data are collected. In addition, it also tends to make the data more economical obtain by reducing expenses like travel costs related to data collection. However, if the elements of the clusters are similar, cluster sampling may be statistically less efficient than random sampling. In addition, if the elements in the clusters are the same, cluster sampling is no better than sampling a single unit from the cluster.

As noted above, it is important to use appropriate sampling methods to avoid introducing bias and obtain a truly representative sample from which one can extrapolate conclusions to a larger population. Statistically, bias is defined as the tendency for a given experimental design or implementation to unintentionally skew the results of the research. Selection bias occurs when the sample is selected in a way that introduces error and causes the resultant sample to not be representative of the underlying population. For example, if it was known that school-aged children were most likely to play a video game, trying to draw a sample from shoppers at a mall during school hours in the middle of the week would be unlikely to result in a representative sample.

Viewpoints

Which Sampling Methods Are the Best?

Although all of these approaches to sampling attempt to increase the probability that a random sample will be drawn, no method of sampling is perfect or without its drawbacks. Sometimes the choice of sampling method is limited by practical factors. In other cases, however, multiple methods may be feasible. In order to maximize the probability that the results of research done with a sample will be, in fact, representative of the underlying population, the choice of sampling method needs to be based on careful consideration rather than expedience. However, which method is best is a topic that has been heatedly debated for decades.

Area Sampling vs. Quota Sampling

Two approaches to sampling that are frequently used are area sampling and quota sampling. Area sampling is a type of multistage sampling that uses maps. Quota sampling is a type of stratified sampling in which the selection of the strata within the sample is not random, but is rather typically left to the discretion of the interviewer. Although some theorists and practitioners believe that quota sample is so innately prone to bias as to be completely worthless, others believe that this technique can be appropriate in some situations or even - with the implementation of adequate safeguards - be made highly reliable.

One of the main reasons that quota sampling continues to be used is that it is significantly less expensive than other methods, often costing only a third or half as much as random sampling techniques (Crawford, 1997). In addition, quota sampling is administratively much simpler to use than other methods because there is no need to randomly select sample members or continue to attempt to contact specific sample members who were unavailable during the first sampling attempt. Further, in some cases quota sampling is the most practical approach to sampling, such as with cases in which one needs to obtain immediate public reaction to an event. In these cases, the delay associated with determining a more random sample would taint the data through the introduction of memory errors. On the other hand, there are a number of drawbacks to the use of quota sampling techniques. Quota samples do not allow the researcher to estimate sampling errors because of their lack of randomness. This also means that potential sampling errors cannot be controlled. In addition, the interviewer using quota sampling may not obtain a representative sample because of experimenter error or bias, lack of opportunity for a more representative sample, or some other cause. Further, quota sampling is heavily dependent on the judgment of the interviewer and, therefore, more open to bias than random techniques.

Hockstim and Smith (1948) conducted a series of experiments to compare the relative efficacy of these two methods of sampling. The first experiment was designed to answer the question of how the composition of quota-control samples and stratified block samples differ. Two interviewers in each of 11 cities with populations over 50,000 were given the same number of ballots. One interviewer was given a simple quota assignment (i.e., gender, age, socioeconomic status) and the other interviewer was given a block assignment stratified by census tract and average monthly rent with superimposed quotas for gender and age. Blocks within each stratum were systematically selected. Interviewers were to select participants from lower, middle, and upper income households in the same proportions as they occurred within the blocks. The experiment was then repeated with a reversal of the quota and block assignments for the two interviewers within each pair. In both surveys, the block sample showed less bias on the education variable, although both sampling techniques resulted in bias on this variable. The block sample was somewhat superior to the quota sample for the variable of average rent, although both samples were comparable on the other three variables examined.

The second experiment examined the question of the effects of restriction on the interviewer's freedom in sampling with a block. In this experiment, a block sample was compared with a sample in which both blocks and dwelling units within the blocks were predetermined systematically in order to minimize bias resulting from the interviewers' choices. In this study, it was found that the less freedom an interviewer had in selecting dwelling units, the more representative the resultant sample was.

The third experiment examined the results of controlling the selection of respondents within households and requiring call-backs if the respondents did not answer. Samples were drawn from an area that contained a city of over 200,000 people, numerous small towns with populations between 2,500 and 25,000 people, and rural areas with villages, farms, and open country. In this experiment, there was very close agreement between the domal and area samples.

The authors drew three major conclusions from their research. First, they found that area samples yielded more representative cross sections of the population than did quota samples. Second, the use of mechanical or automatic selection tools to choose sample participants tends to make sample selection more representative and less subject to bias than does selection based on human opinion. Third, it was found that the requirement for callbacks for households that did not respond the first time was not always necessary. The researchers concluded that in certain circumstances carefully selected quota samples yield cross sections equivalent to those of area samples. The authors concluded that choice of sample method should be made with full consideration of the demands of the survey requirements (Hockstim & Smith, 1948).

Conclusion

Sampling is a group of techniques that are used to select a sample from a larger population so that research can be done with a manageable group and extrapolated to the larger population. There are many approaches to sampling. However, it is important that the sampling technique used results in a sample that represents the larger population and does not systematically introduce bias. Carefully chosen, a sample can help a sociologist draw meaningful results from data and better understand the behavior of people within society.

Terms & Concepts

Bias: The tendency for a given experimental design or implementation to unintentionally skew the results of the experiment due to a nonrandom selection of participants.

Data: (*sing.* datum) In statistics, data are quantifiable observations or measurements that are used as the basis of scientific research.

Population: The entire group of subjects belonging to a certain category (e.g., all women between the ages of 18 and 27; all dry cleaning businesses; all college students).

Probability: A branch of mathematics that deals with estimating the likelihood of an event occurring. Probability is expressed as a value between 0 and 1.0, which is the mathematical expression of the number of actual occurrences to the number of possible occurrences of the event. A probability of 0 signifies that there is no chance that the event will occur and a probability of 1.0 signifies that the event is certain to occur.

Sample: A subset of a population. A random sample is a sample that is chosen at random from the larger population with the assumption that such samples tend to reflect the characteristics of the larger population.

Sampling: A group of techniques that are used to select a sample from a larger population so that research can be done with a manageable group and extrapolated to the larger population.

Sampling Error: An error that occurs in statistical analysis when the sample does not represent the population.

Skewed: A distribution that is not symmetrical around the mean (i.e., there are more data points on one side of the mean than there are on the other).

Validity: The degree to which a survey or other data collection instrument measures what it purports to measure. A data collection instrument cannot be valid unless it is reliable.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables, also known as response variables. Extraneous variables are variables that affect the response but that are not related to the question under investigation in the study.

Bibliography

Arsham, H. (2008). "Questionnaire design and surveys sampling." University of Baltimore. Retrieved September 11, 2007, from http://home.ubalt.edu/ntsbarsh/stat-data/Surveys.htm

Black, K. (2006). *Business statistics for contemporary decision making* (4th ed.). New York: John Wiley & Sons.

Crawford, I. M. (1997). *Marketing research and information systems.* Rome: Food and Agriculture Organization of the United Nations. Retrieved March 14, 2008 from http://www.fao.org/docrep/W3241E/w3241e08.htm

Hochstim, J. R. & Smith, D. M. K. (1948, Spr). Area sampling or quota control? - Three sampling experiments. *Public Opinion Quarterly, 12* (1), 73-80. Retrieved March 14, 2008 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=11922609&site=ehost-live

Mouw, T., & Verdery, A.M. (2012). Network sampling with memory: A proposal for more efficient sampling from social networks. *Sociological Methodology, 42*(1), 206–256. Retrieved date from EBSCO online database, SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=83576852&site=ehost-live

Vearey, J. (2013). Sampling in an urban environment: Overcoming complexities and capturing differences. *Journal of Refugee Studies, 26*(1), 155–162. Retrieved date from EBSCO online database, SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=85919279&site=ehost-live

Suggested Reading

Anderson, M. L. & Taylor, H. F. (2002). *Sociology: Understanding a diverse society* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Chambers, R.L., & Clark, R.G. (2012). *An introduction to model-based survey sampling with applications*. New York: Oxford University Press.

Levy, P.S., & Lemeshow, S. (2011). *Sampling of populations: Methods and applications*. 4th ed. Hoboken, NJ: John Wiley & Sons.

McCormick, T. C. (1938, Oct). The role of statistics in social research: An elementary interpretation. *Social Forces, 17* (1), 47-51. Retrieved March 14, 2008 from EBSCO Online Database SocINDEX with Full Text. http://web.ebscohost.com/ehost/pdf?vid=7&hid=103&sid=2a8235e9-bed9-47a2-9173-27f10b92eff4%40sessionmgr102

Schaefer, R. T. (2002). *Sociology: A brief introduction* (4th ed.). Boston: McGraw -Hill.

Stockard, J. (2000). *Sociology: Discovering society* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Essay by Ruth A. Wienclaw, PhD

Dr. Ruth A. Wienclaw holds a PhD in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human/systems integration.

Experiments

Table of Contents

Abstract

Experimental research is one of the primary ways by which science advances. In experimental paradigms, independent variables (stimuli) are manipulated and the effect of the manipulation on the value of the dependent variable (response) is measured and analyzed. In addition, good experimental research is designed to control as much as possible the effects of extraneous and intervening variables, often by the use of control groups. Experimental research can be performed in the laboratory, through simulations, or through the manipulation of variables in field experiments. These approaches to experimental research allow the design of increasingly more realistic experimental paradigms but give the researcher decreasing control over the experimental situation. In addition to being concerned over designing an experiment that will collect uncontaminated data that can be meaningfully analyzed using inferential statistics, researchers using human subjects also need to be concerned about ethical considerations of the effects of their experiment on its subjects.

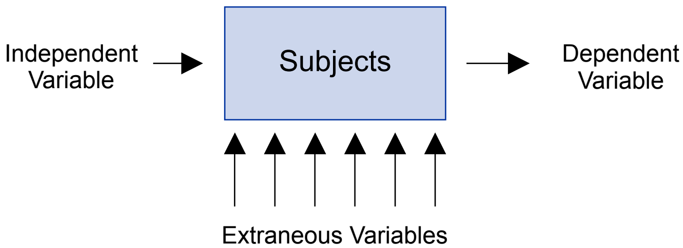
Overview

In order to be considered a true behavioral scientist, it is necessary for sociologists not only to observe and describe human behavior, but to perform research so that they can better understand and predict behavior. In some situations, it is only possible to collect data using surveys—data collection instruments used to acquire information on the opinions, attitudes, or reactions of people. In other situations, however, it is actually possible to control the research situation and manipulate variables in order to obtain better data and a clearer picture of the processes that underlie human behavior. These experiments are situations that are under the control of a researcher in which an experimental condition (independent variable) is manipulated and the effect on the experimental subject (dependent variable) is measured. Most experiments are designed using the principles of the scientific method and are analyzed using inferential statistics to determine whether or not the results are statistically significant.

Experimental Variable

In the simplest experimental design, a stimulus is presented to the research subjects and a response is observed and recorded. For example, one might present subjects with pictures of various situations and ask them to describe what they think they would do if they were in that situation. However, what one thinks one would do and what one would actually do can be very different and may depend on a number of factors. Behavior in the real world tends to be complicated, and several types of variables need to be considered when designing an experiment. Of most concern in most research paradigms are the independent variable (i.e., the stimulus or experimental condition that is hypothesized to affect behavior) and the dependent variable (i.e., the observed effect on behavior caused by the independent variable). However, there are typically other variables that need to be considered and controlled as much as possible, particularly in real-world research. As shown in Figure 1, extraneous variables that affect the outcome of the experiment but that have nothing to do with the independent variable may also need to be considered. For example, a person who responded to a picture of an automobile accident by saying that he or she would not stop to help might respond that way because he or she was feeling tired or ill that day and wanted nothing more than to get home. On another day, he or she might give a different response. In most real-world situations, there are innumerable variables that

Figure 1: Research Variables



are extraneous to the research question being asked but that still affect the outcome of the research. However, a well-designed experiment is created so that it controls for as many of the extraneous variables as possible. It is, of course, impossible in most cases to anticipate and control for every possible extraneous variable. However, the more of these that are accounted for and controlled in the experimental design, the more meaningful the results will be.

Another type of variable not directly related to the experiment but that might affect the results is the intervening variable. These are things that occur between the manipulation of the independent variable and the measurement of the dependent variable. For example, if a person in the experimental situation responded to a picture of an automobile accident responded that he or she would call 911 but would not stop to help later took a course in CPR, he or she might actually stop to help if encountered with the situation in the real world because of the intervening training. Like extraneous variables, intervening variables need to be controlled as much as possible in the experimental situation so that the effect of manipulation of the independent variable on the dependent variable can be determined and statistically analyzed. Extraneous and intervening variables are often controlled in an experiment by the inclusion of a control group comprising subjects that do not receive the experimental condition in order to level the effects of these variables.

Types of Experiments

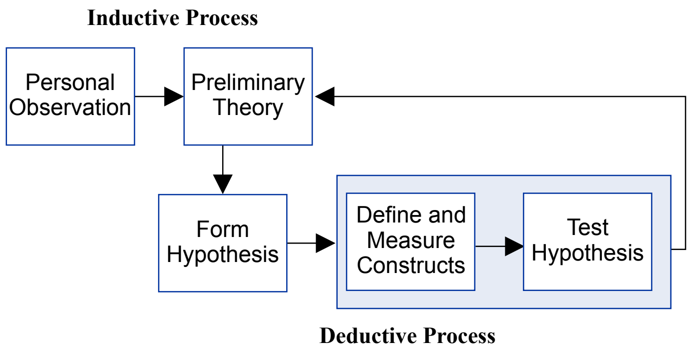
There are a number of ways to collect data that can be used by behavioral scientists. These range from laboratory experiments that allow the researcher great control over the conditions and variables in the study to secondary analysis methods that allow researchers to examine the results of studies done in the past but give them no control over the research design or data collection whatsoever. Laboratory research allows the most control over variables. However, it often is far-removed from real life. Laboratory methods tend to be more appropriate to basic research questions where the influences of the real world are not as important as in applied studies. However, as the research situation becomes more realistic, the research loses a greater degree of control over the situation. To be able to extrapolate research results to the real world and have them be meaningful, it is important to design an experiment that not only controls extraneous variables, but also is as realistic as possible.

There are several general types of experimental studies that can be used to explore the behavior of people in the real world. As discussed above, laboratory experiments allow researchers the most control over extraneous variables. However, laboratories tend to be far removed from the reality of how most people live their lives. Simulation is an approach to experimental research that allows a more realistic setting for the experiment while still allowing researchers a great degree of control. Using the example above, research subjects could be placed in a situation where they encounter a simulated automobile accident so that the researcher could determine how they actually would respond in such a condition. Or, the researcher could set up a field experiment in which the experimental condition is introduced into the real world and the researchers observe how people react. For example, the researcher could set up a situation where it looks like someone has had an automobile accident and then determine what percentage of people actually stop. Both simulations and field research have the advantages of being increasingly more realistic and, therefore, more likely to elicit real responses from people than is possible in the artificial setting of the laboratory. However, these approaches to experimental design also decrease the researcher's control over extraneous variables.

Designing an Experiment

The specifics of how one designs a study depend on the goals of the research and the practical constraints placed on the design by the statistical tools needed to analyze it. In general, however, experiments are designed to test hypotheses as part of the theory building process. As shown in Figure 2, research design starts with the development of a tentative theory that is based on real-world observation. For example, from personal experience, the researcher may have observed that people with certain personality traits are more prone to help others than are people without those traits. Based on these observations, the researcher might form an empirically testable hypothesis concerning the relationship between personality traits and willingness to help others. To find out if this hypothesis is true, the researcher would then operationally define the various terms in the hypothesis (i.e., personality traits, helping others). These definitions would include ways to measure the personality traits (e.g., through an existing personality test, the development of a new test, or ratings of friends) and specific criteria for what constitutes "helping others." The researcher would then conduct the experiment, statistically analyze the resulting data using inferential statistics, and—based on the statistical significance of the answer—determine whether it was likely for those possessing the studied personality trait to help others.

Figure 2: The Theory Building Process



In many cases, behavioral research can be more complicated than research in the physical sciences because of the complex nature of the real-world situation and the difficulty in operationally defining and measuring mental constructs. In addition, behavioral research needs to take into account ethical considerations for the treatment of research subjects. Although a compound in a chemistry experiment does not care if it is immersed in boiling water or flash frozen, a human subject would certainly object. In scientific research, ethics refers to a code of moral conduct regarding the treatment of research subjects that is subscribed to by the members of a professional community. Care must be taken in the design and conduct of experiments so that research subjects are not harmed.

Applications

The Stanford Prison Experiment

The ethical treatment of prisoners has been of concern for as long as there have been prisoners. In the aftermath of World War II, for example, the Fourth Geneva Convention specified standards for the ethical treatment of prisoners of war. During the Iraq and Afghan Wars, the ethical treatment of prisoners made the headlines with controversies over the treatment of prisoners at Abu Ghraib and the use of torture to gain information from prisoners in the War on Terror. One of the landmark studies on the treatment of prisoners was performed at Stanford University by a team lead by Philip Zimbardo in 1971. This study is an outstanding example of how behavioral research can be conducted in the real world through simulation, as well as of some of the practical and ethical problems that can be encountered when performing such experiments.

Zimbardo advertised in a local newspaper for male volunteers to participate in a study of the psychological effects of prison life. The applicants were given diagnostic interviews and personality tests to eliminate candidates with psychological problems, medical problems, or a history of drug or alcohol abuse. The application process reduced the number of applicants from 70 to 24. The resultant group was considered to be healthy, intelligent, middle-class individuals who tested within normal parameters on all measures administered. These individuals were randomly assigned to one of two groups: prison guards and prisoners. The experiment began with nine guards and nine prisoners. Guards worked in teams of three for eight-hour shifts, providing around the clock coverage of the "jail." Nine subjects were assigned to the prisoner group. The remaining subjects were placed on call in case they were needed. At the beginning of the experiment, there was no difference on any discernible factor between the individuals in the two groups. All subjects were required to sign an informed consent agreement, which told the subjects to expect some harassment, violation of privacy and other civil rights, and a minimally adequate diet during the course of the two-week experiment.

Without further notice to the subjects of the experiment, the local police—who were confederates in the study—went to the homes of those who had been assigned to the prisoner group and arrested them. Subjects were subjected to the full range of procedures associated with an arrest, including being charged, read their Miranda rights, spread-eagled on the police cruiser, searched, and handcuffed, often in front of their neighbors. The subjects were taken to the local police station, formally booked, read their rights again, and fingerprinted. Subjects were then left blindfolded in a holding cell. The prisoners (still blindfolded) were then transported to the "Stanford County Jail."

The jail simulation had been designed based on the inputs of several consultants, including one who had been incarcerated for nearly 17 years. A corridor in the basement of the Stanford University building was boarded up at each end and some of the doors to laboratory rooms were replaced with doors with steel bars and cell numbers. A closet approximately two feet square (large enough for a prisoner to stand but not to sit) was transformed into a cell for solitary confinement. No windows or clocks were included in the simulation environment that would allow prisoners to note the passage of time. Listening devices were concealed in the cells to record the prisoners' conversations and a hole in the wall at one end of the corridor allowed videotaping of activities in the hall.

Upon arrival at the "jail," the "prisoners" were lectured by the "warden" (a confederate of the experimenters) about the fact that they would be receiving different treatment as prisoners. They were then systematically searched and stripped naked and subjected to a spray for the purported purpose of delousing. These were some of the procedures to which the prisoners were subjected in order to humiliate them as is often done in real prison situations. Subjects then were given a smock with their prison identification number on the front and back as their sole piece of clothing. A chain was also padlocked around the foot of each prisoner. Many of these activities in the simulated prison (e.g., the chain around the foot) are not done in real prisons but were included in the simulation in order to increase the feeling of helplessness and oppression. Prisoners spent most of their time in the small cells that were barely large enough to hold three cots each.

Subjects randomly assigned to the "guards" group were not given any training on how to treat the prisoners. Within broad limits, they were allowed to do what they thought appropriate to maintain order and discipline in the "jail." Just as the prisoners were informed of the differences between being in jail and living in the outside world, the guards were warned of the dangerousness of their job. The guards made up their own rules for the running of the jail under the supervision of the warden. Guards were also given uniforms and wore mirrored sunglasses. They carried Billy clubs and whistles.

Following the arrests and indoctrination, subjects began the roles of prison guards and prisoners. At first, the subjects did not take the assigned roles very seriously: They were conscious of the fact that this was a simulation and that they were merely playing roles. At 2:30 a.m. the first morning, prisoners were awakened by blasting whistles and were required to gather in the hall to be counted. This was the first in a series of confrontations between guards and prisoners that resulted in increasingly polarized roles between the two groups. By the second day, the prisoners rebelled. The on-call guards were called in and the prisoners were forced back, stripped, and deprived of what little comfort they had. This began an increasing escalation of harassment and intimidation against the prisoners. After this incident, the guards developed a series of psychological tactics to maintain order and discipline among the prisoners. The guards' behavior was often arbitrary and humiliating to the prisoners. Within 36 hours of the start of the experiment, one of the prisoners began to "act crazy" and eventually was released from participation in the study. A little later in the experiment, another prisoner, who had been feeling ill and had refused to eat, broke down and became hysterical while talking to a priest (who had been brought in as a confederate) and the principal experimenter (in the role of prison "superintendent"). The researcher eventually removed the prisoner's chain and cap and reminded him that this was merely an experiment and not a real prison. The subject stopped crying and asked to go back to his cell to prove he was not a bad prisoner.

Eventually, the guards each developed one of three types of coping mechanisms. One group was tough but fair and followed the prison rules. The second group comprised "good guys" who did not punish the prisoners and would do small favors for them. The final group of guards was hostile, arbitrary, and very inventive in developing new ways to humiliate prisoners. Such behaviors, however, were not predicted by the initial personality tests that had been administered to each of the subjects.

Prisoners coped with their situation and concomitant feelings of powerless and frustration in a number of ways. At first, some of the prisoners attempt to rebel against the authority of the guards. Four of the prisoners broke down emotionally. One prisoner developed a psychosomatic rash over his entire body after learning that the "parole board" had denied his request for parole. Other subjects tried to be model prisoners and do everything required of them by the guards. However, by the end of the experiment, the prisoners had disintegrated both individually and as a group, and the guards were in total control of the prison.

Although the experiment was originally planned to last two weeks, the changes in behavior in the participants and the negative impact of the study made the researchers stop the study after only six days. By that time, the "prisoners" had become withdrawn and depressed and had started to behave in pathological ways. Some of the guards had started to behave sadistically, and none of the other guards attempted to intervene. The study was terminated after late-night videotapes showed the guards engaging in degrading and pornographic abuse of the prisoners when they thought that the experimenters were not watching. Another precipitating event was when a behavioral scientist brought in to interview the prisoners objected strongly to the behaviors she observed. However, she was the only one who ever questioned the morality of the experiment.

Conclusion

Experimental research is one of the foundations of science. In the simplest experimental design, a stimulus (independent variable) is presented to the research subjects and a response (dependent variable) is observed and recorded. Behavioral scientists also need to be concerned with the control of extraneous variables that can affect the outcome of the study without being directly related to the research question and intervening variables that occur after the manipulation of the independent variable but before the measurement of the dependent variable. As much as possible, these need to be controlled or eliminated through the experimental design. Laboratory experiments allow researchers great control over the variables in a study but have the drawback of not realistically emulating the real-world situation. Simulations and field experiments allow the design of an increasingly realistic experiment but also give the experimenter decreased control over the experimental situation. In addition, when working with human subjects, care must be taken in the design and conduct of experiments so that research subjects are not harmed.

Terms & Concepts

Confederate: A person who assists a researcher by pretending to be part of the experimental situation while actually only playing a rehearsed part meant to stimulate a response from the research subject.

Control Group: A subset of participants in an experiment that does not receive the experimental condition (i.e., does not experience the manipulation of the independent variable). Control groups help researchers determine whether the observed results of a research study were due to the manipulation of the independent variable or some other factor.

Data: (*sing.* datum). In statistics, data are quantifiable observations or measurements that are used as the basis of scientific research.

Ethics: In scientific research, a code of moral conduct regarding the treatment of research subjects that is subscribed to by the members of a professional community. Many professional groups have a specific written code of ethics that sets standards and principles for professional conduct and the treatment of research subjects.

Experiment: A situation under the control of a researcher in which and experimental condition (independent variable) is manipulated and the effect on the experimental subject (dependent variable) is measured. Most experiments are designed using the principles of the scientific method and are analyzed to determined whether or not the results are statistically significant.

Hypothesis: An empirically testable declaration that certain variables and their corresponding measure are related in a specific way proposed by a theory. A null hypothesis (H0) is a statement that the findings of the experiment will show no statistical difference between the current condition (control condition) and the experimental condition.

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data. Inferential statistics are used to make inferences such as drawing conclusions about a population from a sample and in decision making.

Operational Definition: A definition that is stated in terms that can be observed and measured.

Statistical Significance: The degree to which an observed outcome is unlikely to have occurred due to chance.

Subject: A participant in a research study or experiment whose responses are observed, recorded, and analyzed.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables (response). Extraneous variables are variables that affect the response but that are not related to the question under investigation in the study. Intervening variables occur between the manipulation of the independent variable and the measurement of the dependent variable.

Bibliography

Black, K. (2006). *Business statistics for contemporary decision making* (4th ed.). New York, NY: John Wiley & Sons.

Hickey, L. (2012). Little Albert and Stanford Prison Experiment. In *Psychology experiments and case studies*. Delhi, India: English Press. Retrieved November 5, 2013 from EBSCO online database eBook Collection (EBSCOhost). http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=406924&site=ehost-live

Mikėnė, S., Gaižauskaitė, I., & Valavičienė, N. (2013). Qualitative interviewing: Field-work realities. *Socialinis Darbas, 12*(1), 49–61. Retrieved November 5, 2013 from EBSCO online database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=88008107

The Stanford Prison Experiment . Retrieved 18 March 2002 from: www.prisonexp.org.

Wendt, O., & Miller, B. (2012). Quality appraisal of single-subject experimental designs: An overview and comparison of different appraisal tools. *Education & Treatment Of Children (West Virginia University Press), 35*(2), 235–268. Retrieved November 5, 2013 from EBSCO online database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=74231752

Witte, R. S. (1980). *Statistics* . New York, NY: Holt, Rinehart and Winston.

Suggested Reading

Calfee, R. C. (1975). *Human experimental psychology* . New York, NY: Holt, Rinehart and Winston.

L&ecirc ;, Q., & L&ecirc ;, T. (2013). *Conducting research in a changing and challenging world*. Hauppauge, NY: Nova Science Publishers. Retrieved November 5, 2013 from EBSCO online database eBook Collection (EBSCOhost). http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=607815&site=ehost-live

Weisburd, D., Morris, N. A., & Ready, J. (2008, Mar). Risk-focused policing at places: An experimental evaluation. *Justice Quarterly, 25*(1), 163-200. Retrieved April 2, 2008 from EBSCO Online Database SocINDEX with Full Text: http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=31334370&site=ehost-live

Zimbardo, P. G., Maslach, C., & Haney, C. (2000). Reflections on the Stanford Prison Experiment: Genesis, transformations, consequences. In T. Blass (Ed.), *Obedience to authority: Current perspectives on the Milgram paradigm* (pp. 193–237). Mahwah, NJ: Erlbaum.

Essay by Ruth A. Wienclaw, Ph.D.

Dr. Ruth A. Wienclaw holds a doctorate in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human/systems integration.

Reliability

Table of Contents

Abstract

To yield useable data, surveys, assessment tools, and other data collection instruments need to be both reliable and valid. Reliability is a measure of the degree to which such instruments consistently measure a characteristic or attribute. Statistically, reliability is a measure of the observed variability in obtained scores on an instrument. Variability can come both from true variance (such as differences in opinions, knowledge, or other characteristics of the individual) or from error variance. The total variability of a data collection or assessment instrument is the sum of the true variability and the variability due to error. Reliability can be estimated through the use of parallel forms of the instrument, repeated administration of the same form of the instrument, subdivision of the instrument into two parallel groups of items, and analysis of the covariance among the individual items.

Overview

In the case of data collection or assessment instruments, reliability—the degree to which a data collection or assessment instrument consistently measures a characteristic or attribute—is essential for the instrument to be valid. In other words, one must be confident that the instrument measures what it purports to measure. No matter how well-written a data collection or assessment instrument appears to be on its face, it cannot be valid unless it is reliable. If a measure is not reliable, it does not consistently measure the same thing. In other words, it is not measuring the construct that it was designed to measure, so the instrument is neither reliable nor valid. Therefore, both validity and reliability are essential when conducting survey research, so that the data collected in the study will actually give researchers the information they are trying to gather. Without both reliability and validity, the data collected are meaningless, and no conclusions can be drawn.

True Data Variance vs. Data Error

Even in the physical sciences, two sets of measures performed on the same individuals never exactly duplicate each other. To the extent that this is true, the measurement instrument is unreliable, whether it is a physical scale used to measure the weight of a chemical compound or a paper-and-pencil survey used to measure a person's attitude toward something. For example, on a scale of 1 to 10, what one person describes as a 10 another person may call a 9.5. This does not necessarily mean that their opinions are different, just that the two people are expressing them differently. Some of the total observed variance (the square of the standard deviation) in scores is due to true variance, or real differences in the way that people are responding to the question. The other part of the total variance is due to error.

Factors Affecting Reliability

Personal Factors

There are many reasons why a data collection instrument may not be reliable and thus may contribute to the error variance. In general, what social scientists try to measure are lasting and general characteristics of individuals related to the underlying construct that the assessment instrument is trying to measure. However, other types of characteristics that are not part of the underlying construct, such as the individual's test-taking techniques and general ability to comprehend instructions, may also be measured.

In addition to the permanent characteristics of individuals, there are also temporary characteristics that can affect their responses to questions on data collection instruments. These might include such factors as general health, fatigue, or emotional strain, all of which can affect the way that an individual responds to a question—a phenomenon familiar to anyone who has had to take a test in school when he or she was ill. Similarly, external conditions such as heat, light, ventilation, or even momentary distraction can impact one's responses in a way that does not reflect the underlying theoretical construct. Further, the subject's motivation can also impact the reliability of a data collection instrument. For example, for the most part teachers assume that their students are motivated to do well on any data collection instruments (e.g., a mid-term exam) given them. However, the same assumption cannot be made when asking a random sample of individuals to answer questions on the data collection or assessment instrument. For instance, it is often difficult to get shoppers to cooperate in opinion surveys because they are intent on accomplishing their errands so that they can go home. The motivation that may be offered to entice participation in the survey, such as a crisp new dollar bill or a carton of instant macaroni and cheese, is nothing compared to the motivation of students to do well in a course.

Difficulty in Understanding the Data Collection Instrument

Another source of variability in the way people respond to a data collection instrument may be individual differences in the way that people interpret the questions on the instrument. Care must always be taken in the development of a data collection instrument to write to a level that can be understood by all the people who will answer the questions. The questions need to be written unambiguously and with proper spelling, grammar, and punctuation to help reduce the possibility of low reliability because people do not understand what the questions are asking. For example, a child could easily take a question about a person who "lives near" him or her to mean the family members in the immediate household rather than a neighbor. Similarly, a question about how much one likes "sweet tea" means a different thing in the southern United States, where it refers to iced tea sweetened with simple syrup, than it does in Great Britain. The use of clear, concise language and operational definitions can help increase the reliability of the instrument.

Reliability problems may also stem from individual differences in the way that people interpret responses to a data collection instrument. Even in cases where the end points of the scale are operationally defined with clear examples, people who moderately dislike something could possibly vary their answers between 20 and 40 on a scale of 100, yet all mean the same thing. Similarly, some people never give a perfect score to anything on a rating scale because they believe that there is always room for improvement.

Inaccurate Measurements

Another potential cause of lack of reliability is when the data collection instrument is not valid and is actually measuring more than one thing. For example, a researcher might set up an experiment to determine whether men or women are more likely to stop and assist a stranger on the street who needs help. This could be done by having a confederate drop a sheaf of loose papers and counting how many times a man stops to help and how many times a woman stops to help. If more men stop than women, the researcher will probably conclude that men are more likely to help a stranger than are women. However, these results might not be replicable, particularly if the confederate is an attractive young female wearing a short skirt and halter top; although more men might stop to help, the experiment will more likely be measuring attraction rather than helpfulness. Similarly, great variability and concomitant low reliability can be found when data are collected through the use of a structured interview. Even when the questions are always asked with the same wording, differences between interviewers and how they are perceived by the people answering the questions can result in a situation where the same data collection instrument has widely disparate responses because of the interviewing styles of different interviewers.

Testing the Efficacy of Data Collection Instruments

In general terms, reliability is defined as the degree to which a data collection or assessment instrument consistently measures a characteristic or attribute. There are several ways that the reliability of such an instrument can be estimated.

The first of these methods involves the administration of two parallel forms of the instrument under specified conditions. The statistical correlation of the results of the two administrations is calculated to determine the degree of variance between the forms. However, it is important to note that it is typically difficult to develop two equivalent forms of the same assessment instrument that both have equal discriminability.

As a result of this difficulty, a second method of determining reliability, called test-retest reliability, is frequently used. In this approach, the same form of the data collection instrument is administered twice to the same sample of individuals, and the correlation between the two scores is calculated to determine the reliability.

A third approach to estimating reliability is to subdivide a single instrument into two presumably parallel groups of items. All of the items on the instrument are given to one sample of individuals at one time, and then the items are split out and treated as if they were two separate instruments. Each group is scored separately, and the resulting scores are correlated. This approach is called split-half reliability.

Finally, an analysis of covariance can be calculated among the individual items on the assessment instrument to determine the true score and error variances.

Applications

Case Study: Presenting Child Need

In an example of the application of reliability and validity assessment to a real-world problem, Forrester, Fairtlough, and Bennet (2007) examined the inter-rater reliability of methods to describe the needs of children to children's services in England and Wales. Although it is important to look at the unique characteristics of each case when determining what type of help a child needs, it is also important to be able to speak about "need" using a common language that will reliably discriminate between various classifications of need so that children can be given the help necessary for their well-being. Several typologies of need identification have been developed, but in order to maximize their usefulness, they must yield reliable results. Forrester, Fairtlough, and Bennet (2007) examined 200 consecutive file studies of closed referrals, first analyzing the files to classify the presenting needs or potential needs of the child in each case and then grouping them into clusters of variables for issues that occurred more than once. Fifty randomly selected cases were then tested to determine the relationship between the variables. The patterns were statistically analyzed using cross-tabulation and Spearman's rank correlation coefficient. Based on this analysis, the variables were reduced to a final list of ten, plus an "other" category.

The Results

The authors reached four main conclusions about the reliability of descriptions of children's need used by social services. First, it was found that descriptions of need that relied on a "main" need were not as reliable as other approaches, and that patterns of incidence could not be described adequately using the construct of "main" need. Although such an approach may simplify data presentation, it does not adequately describe the complexity of a child's situation, nor does it give any indication of the seriousness of the need. Second, although it was found that other approaches to describing need were more reliable than the "main" need approach, they, too, were not without their problems. In these typologies, classifications such as "dysfunctional family" or "unstable or otherwise detrimental family" were vague and had low levels of reliability. The authors urged that such terms be better defined in order to increase reliability. Third, the meaning of the legal definition of need had low levels of agreement between the raters. In part, this appeared to be a result of the fact that the legal definition emphasized seriousness of the need rather than presence of the need, as is the case in the other definitions and typologies.

The authors concluded that typologies needed to be developed for the full range of referrals to children's services. In addition, they cautioned that the concept of "main" need that was used in both research and government policy was unreliable and not a good indicator of a child's situation or problem. The use of this concept could lead to misclassification and inappropriate intervention for at-risk children. Third, some specific categories that were currently used, such as "dysfunctional family," required better definition to increase reliability of assessments. Finally, future typologies of need should be tested for inter-rater reliability before being implemented. It is only through a reliable instrument that at-risk children can be consistently identified and their needs appropriately assessed.

The Children's Physical Environment Rating Scale

In another example of a reliability study, Moore and Sugiyama (2007) examined the reliability and validity of a new scale to be used for assessing the physical environment of early childhood educational facilities. The literature links the physical environment of such facilities to cognitive and social development during early childhood. The Children's Physical Environment Rating Scale (CPERS) comprises 124 items clustered into 14 scales that focus on planning, overall architectural quality, indoor activity spaces, and outdoor play areas.

The reliability of the CPERS was tested for inter-rater reliability, test-retest reliability, and internal consistency. Inter-rater reliability was tested in 46 childhood development centers in Sydney, Australia. Each center was assessed by two of seven raters through several cycles of field testing. The resulting data were statistically analyzed to determine the degree of agreement between raters for each item and Cronbach's generalizability coefficient G for each subscale. These analyses showed a high degree of agreement and generalizability between the raters on the items on the CPERS. Based on this result, the authors concluded that the CPERS is a reliable instrument that can consistently be used to rate the physical environment of an early childhood facility, both for research purposes and in general.

In addition, the authors examined the degree to which scores on the CPERS were stable over time. Each of 11 early childhood development centers was assessed once and then reassessed again three to five weeks later. The results were analyzed using Cronbach's G. The results showed a high degree of test-retest reliability, indicating that scores on the CPERS are stable over time and are consistent measures.

Finally, internal consistency of the scales in the CPERS was assessed simultaneously but independently by two raters similar to center directors who might use the scale on a routine basis to assess 11 centers. The results of the ratings were analyzed using Cronbach's alpha to show the internal consistency of each subscale. In general, the results showed that the CPERS has very high internal consistency and is highly reliable for use in assessing early childhood centers.

Conclusion

In order for a data collection or assessment instrument to be valid and test what it purports to measure, it must be designed to be reliable and consistently measure a characteristic or attribute. If an instrument is not both reliable and valid, the resulting data are not of use to the researcher. There are many potential sources of variability in the results of data collection and assessment instruments. These include lasting and general characteristics of the individual, lasting but specific characteristics of the individual, temporary but general characteristics of the individual, temporary and specific characteristics of the individual, and systematic or chance factors affecting the administration of the instrument.

In addition, the data from every assessment instrument will also contain some degree of variability that is attributable to error. The total variability of a data collection or assessment instrument is the sum of the true variability and the variability due to error. Reliability can be estimated through the use of parallel forms of the instrument, repeated administration of the same form of the instrument, subdivision of the instrument into two presumably parallel groups of items, and analysis of the covariance among the individual items.

Terms & Concepts

Confederate: A person who assists a researcher by pretending to be part of the experimental situation while actually only playing a rehearsed part meant to stimulate a response from the research subject.

Correlation: The degree to which two events or variables are consistently related. Correlation may be positive (as the value of one variable increases, the value of the other variable increases), negative (as the value of one variable increases, the value of the other variable decreases), or zero (the values of the two variables are unrelated). Correlation does not imply causation.

Data: In statistics, quantifiable observations or measurements that are used as the basis of scientific research.

Operational Definition: A definition that is stated in terms that can be observed and measured.

Reliability: The degree to which a data collection or assessment instrument consistently measures a characteristic or attribute. An assessment instrument cannot be valid unless it is reliable.

Sample: A subset of a population. A random sample is a sample that is chosen at random from the larger population with the assumption that it will reflect the characteristics of the larger population.

Standard Deviation: A measure of variability that describes how far the typical score in a distribution is from the mean of the distribution.

Survey: (a) A data collection instrument used to acquire information on the opinions, attitudes, or reactions of people; (b) a research study in which members of a selected sample are asked questions concerning their opinions, attitudes, or reactions, and the responses are analyzed and used to extrapolate from the sample to the underlying population.

Survey Research: A type of research in which data about the opinions, attitudes, or reactions of the members of a sample are gathered using a survey instrument. The phases of survey research are goal setting, planning, implementation, evaluation, and feedback. Unlike experimental research, survey research does not allow for the manipulation of an independent variable.

Validity: The degree to which a survey or other data collection instrument measures what it purports to measure. A data collection instrument cannot be valid unless it is reliable. Content validity is a measure of how well assessment instrument items reflect the concepts that the instrument developer is trying to assess. Construct validity is a measure of how well an assessment instrument measures what it is intended to measure as defined by another assessment instrument. Face validity is when an assessment instrument appears to measure what it is trying to measure. Cross validity is the validation of an assessment instrument with a new sample to determine if the instrument is valid across situations. Predictive validity refers to how well an assessment instrument predicts future events.

Bibliography

Bowen, N. K. (2008). Cognitive testing and the validity of child-report data from the elementary school success profile. Social Work Research, 32(1), 18-28. Retrieved April 1, 2008, from EBSCO Online Database Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=31243646&site=ehost-live

Compton, D., Love, T. P., & Sell, J. (2012). Developing and assessing intercoder reliability in studies of group interaction. *Sociological Methodology, 42*(1), 348–364. Retrieved November 8, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=83576837&site=ehost-live

Forrester, D., Fairtlough, A., & Bennet, Y. (2007). Describing the needs of children presenting to children's services: Issues of reliability and validity. Journal of Children's Services, 2(2), 48-59. Retrieved April 9, 2008, from EBSCO Online Database SocINDEX with Full Text.

http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=27347409&site=ehost-live

Hendrick, T. M., Fischer, A. H., Tobi, H., & Frewer, L. J. (2013). Self-reported attitude scales: Current practice in adequate assessment of reliability, validity, and dimensionality. *Journal of Applied Social Psychology, 43*(7), 1538–1552. Retrieved November 8, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=88980290&site=ehost-live

Moore, G. T. & Sugiyama, T. (2007). The Children's Physical Environment Rating scale (CPERS): Reliability and validity for assessing the physical environment of early childhood educational facilities. Children, Youth and Environments, 17(4), 24-53. Retrieved April 9, 2008, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=31380663&site=ehost-live

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill Book Company.

Peterson, R. A., & Yeolib, K. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology, 98*(1), 194–198. Retrieved November 8, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=84917386&site=ehost-live

Teixeira de Melo, A., Alarcão, M., & Pimentel, I. (2012). Validity and reliability of three rating scales to assess practitioners' skills to conduct collaborative, strength-based, systemic work in family-based services. *American Journal of Family Therapy, 40*(5), 420–433. Retrieved November 8, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=80413978&site=ehost-live

Suggested Reading

Allen, M. J. & Yen, W. M. (1979). Introduction to measurement theory. Monterey, CA: Brooks/Cole Publishing Company.

Bulloch, S. (2013). Seeking construct validity in interpersonal trust research: A proposal on linking theory and survey measures. *Social Indicators Research, 113*(3), 1289–1310. Retrieved November 8, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=89480509&site=ehost-live

Conley, T. B. (2006). Court ordered multiple offender drunk drivers: Validity and reliability of rapid assessment. Journal of Social Work Practice in the Addictions, 6(3), 37-51. Retrieved April 9, 2008, from EBSCO Online Database Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=23270082&site=ehost-live

Dunn, T. W., Smith, T. B., & Montoya, J. A. (2006). Multicultural competency instrumentation: A review and analysis of reliability generalization. Journal of Counseling & Development, 84(4), 471-482. Retrieved April 9, 2008, from EBSCO Online Database Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=22177696&site=ehost-live

Gillaspy, J. A. Jr. & Campbell, T. C. (2007). Reliability and validity of scores from the Inventory of Drug Use Consequences. Journal of Addictions & Offender Counseling, 27(1), 17-27. Retrieved April 9, 2008, from EBSCO Online Database Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=23040456&site=ehost-live

Lemke, E. & Wiersma, W. (1976). Principles of psychological measurement. Chicago: Rand McNally College Publishing Company.

Lewis, C. A. & Cruise, S. M. (2006). Temporal stability of the Francis Scale of Attitude toward Christianity among 9- to 11-year-old English children: Test-retest data over six weeks. Social Behavior and Personality, 34(9), 1081-1086. Retrieved April 9, 2008, from EBSCO Online Database Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=23025211&site=ehost-live

Essay by Ruth A. Wienclaw, PhD

Dr. Ruth A. Wienclaw holds a PhD in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human-systems integration.

Qualitative Research Methods

Table of Contents

Abstract

Qualitative research encompasses a diverse range of theoretical and philosophical traditions that, in general, stem from an interpretivist view of the world. In contrast to quantitative research, which stems primarily from a positivist view of the world, and which is typically used to measure things that one can see, qualitative research seeks to explain or understand how a particular group of people experience and interpret their lives. Often, qualitative research is used to explore social phenomena and processes in their natural settings — such as how people cope with disease and illness, how the social category of gender is socially constructed through formal education, or how people learn to identify themselves as gay or lesbian. Qualitative research is used to make sense of such phenomena in terms of what they mean to people, or how people make sense of such experiences (Hallberg, 2006). Using methods such as detailed interviewing, participant observation, and focus groups, qualitative researchers try to gain insight into the actor's perspective and capture his or her point of view or lived experience.

Overview

Quantitative & Qualitative Research

Qualitative and quantitative research reflect different perspectives on the world and different sets of assumptions about what constitutes knowledge. At the heart of the distinction between qualitative and quantitative research lies a question about the nature of reality (Nicholls, 2009). **Positivism** is a philosophical tradition that emphasizes the importance of conducting social science research in a way that focuses strictly on causal relationships between behaviors and other social phenomena that can be measured or directly observed (Straus & Corbin, 2002). Positivism holds that objects, events, and social behaviors have objective properties that can be seen, measured, and predicted (Kuhn, 1970, p. 184–85). Thus, the positivist tradition emphasizes the importance of objectivity, which is an approach to creating knowledge that claims to be detached from the phenomenon under study (because the research process does not get involved in the lives of those being studied) and unbiased (because research methods do not reflect the personal values of the researcher).

This approach to research typically relies on quantitative methods — tests, questionnaires, standardized observation instruments, and formal records — to collect data, and analysis aims for precision by attaching numerical values to people's experience. Quantitative methods such as surveys, experiments, and intervention studies are used to measure things that one can see; in social science, what is typically measured are statistical relationships that stand for social behaviors (Shields & Twycross, 2003). These statistical relationships can provide insight into phenomena such as rates of illness and disease in populations or the impact of an education intervention on academic performance. **Transferability** , or the extent to which a finding is transferable or generalizable to other populations, is important in quantitative research (Silverman, 1993). It is also concerned with validity, of which there are two kinds. Internal validity refers to whether or not a study achieved its aims, and external validity has to do with the extent to which the results can be applied on other contexts.

Criticisms of Quantitative Research

However, quantitative research is generally unable to describe and explain social experience or explore what is taken for granted about a society and its cultural practices. Indeed, for some social groups, there has been something of a loss of faith in positivism (Carpenter & Suto, 2008) because of the tendency of researchers associated with this tradition to make generalizations about social experience. Yet, as Lincoln and Cannella (2004, p. 7) argue, quantitative research, especially when it relies on experimental design, is generally ill-suited to the complex and dynamic social world, which is characterized by differences in, at the very least, "gender, race, ethnicity, linguistic status, or class." Dissatisfaction with quantitative research methods was evident in early twentieth century social research such as that associated with the Chicago School, which in the 1920s and 1930s pioneered research methods that emphasized the importance of capturing the experience of people living in urban settings in their own words and through direct observation.

This kind of research relied on ethnography — fieldwork, case studies, and in-depth interviews — to generate richly detailed data (Hallberg, 2006). Classic qualitative studies include William Foote Whyte's study of young men in an Italian neighborhood of Boston (1993 [1943]), which is underpinned by the assumption that a social phenomenon needs to be studied in its entirety. This tradition relies on an interpretivist perspective on the world and is critical of the positivist emphasis on studying the parts of a phenomenon rather than the whole.

Specific critique of positivism in general and quantitative methods in particular by various protest groups, such as civil rights workers, Marxists, feminists, and disability advocates, followed in the 1960s and 1970s (Nettleton, 2006). These groups identified the ways in which quantitative research methods are based on values that have come to be associated with a male or masculine view of the world and can therefore be seen as androcentric, such as objectivity, detachment, and reason, and dismiss indigenous and vernacular knowledge (people's everyday language and customs) as irrational.

Feminist researchers have been especially critical of how social research using quantitative methods claims to be neutral and objective (Harding, 1991). Such an approach makes assumptions about women's lives and ignores aspects of life that are important to women. For instance, feminism has been critical of the idea that science can be detached and that the researcher is not involved in a relationship with the person who participates in the research. Qualitative methods involve an emotional closeness between the researcher and the research participant that can have the advantage of helping the researcher gain insight into sensitive and previously ignored areas of social life, such as life course changes (e.g., menopause), health care experiences, and sexuality.

Qualitative Research & Interpretivism

Qualitative research sets out to answer a different set of questions from quantitative methods. According to Hallberg,

> Qualitative researchers study phenomena and processes in their natural settings and intend to make sense of those matters in terms of the meanings people bring to them … Through detailed interviewing, participant observations, and rich descriptions of the social world, qualitative researchers hope to come close to the actor's perspective and try to capture his or her point of view or lived experience. (2006, p. 141)

The interpretivist tradition on which qualitative research is based tries to understand what it means to be human and asks questions about the meaning of social phenomena, how things work, and how people's perceptions influence their lives. Interpretivism is the basis of a diverse range of theoretical traditions, including ethnography, phenomenology, symbolic interactionism, and postmodernism. There are, however, some differences among approaches to qualitative research that reflect whether the researchers accept an objective reality that is subjectively lived (known as *subtle realism*) or believe that reality itself is best understood as multifaceted (*constructivism*).

These variations on how interpretivism is considered influence both data collection methods and analysis in qualitative research. For instance, research that is informed by subtle realism is more likely to use methods that try to represent reality (Mays & Pope, 2000), such as a case study or historical records. Analysis may involve triangulation between different kinds of data (for instance, reports and public records) to ensure that the findings will be seen as reliable and valid (Miles & Huberman, 1994). Moreover, in this view, qualitative research can be evaluated in the same way as quantitative research. That is, its validity can be assessed by reference to issues such as respondent validation, clear detailing of methods of data collection and analysis, and attention to negative cases (Mays & Pope, 2004, p. 51).

In contrast, constructivist accounts of social phenomena reject the importance of objectivity and reliability, emphasizing instead that social analysis (or knowledge) is provisional and context dependent. Constructivist research is more concerned with reflexivity than with reliability or even validity. Reflexivity refers to the way that a researcher's background and assumptions affect what questions are being asked, who the target population is for the study, how the questions are asked, and how the analysis is conducted. The qualitative researcher understands that he or she is not a neutral observer of social life (Haraway, 1991) and that what he or she sees is, in part, determined by his or her assumptions and background. Rather than viewing this insight as a limitation on the reliability or validity of research, reflexivity is viewed as a commitment to acknowledging and questioning the role of the researcher in collecting and analyzing data. Reflexivity does not ignore the potential for bias in the researcher's account of social life; it acknowledges the existence of any bias and makes that acknowledgement visible in his or her account of the research process.

Regardless of these differences, qualitative research generally seeks to gain access to subjectively lived experiences by talking with, listening to, and watching people in their everyday contexts. Moreover, although there are different theoretical or philosophical traditions that inform qualitative research, they share the common goal of trying to explain or understand how a particular group of people experience and interpret their circumstances. Broadly, qualitative research methods involve the systematic collection, organization, and interpretation of textual or visual material derived from talk or observation (Malterud, 2001). Unlike quantitative research, which collects data from many participants (typically hundreds, sometimes thousands) to explore a relatively small number of questions, qualitative research generates volumes of richly detailed data from a small number of cases.

Applications

Qualitative Research Methods

Broadly, qualitative research methods are used to gain insight into people's attitudes, behaviors, value systems, concerns, motivations, aspirations, culture, or lifestyles through the collecting of information that is typically unstructured. Unstructured information, or data, can include interview material, customer feedback forms, reports, or media clips. These data can be collected via in-depth interviews, analysis of textual or visual content (e.g., magazine articles or advertisements), participant observation (ethnography), or focus groups. Analysis involves a process of organizing data in order to interpret it, though the process varies according to the approach underlying the research. For instance, analysis can involve immersion (Miller & Crabtree, 1999), which entails looking closely at data and isolating what seem to most answer the research questions. Sometimes data can be isolated in order to examine it closely (decontextualization), and sometimes interpretations of data can be examined to see if they fit with broader patterns of data as a whole (recontextualization) (Silverman, 1993). A theory-based approach uses theory to organize data in order to shed light on a previously known phenomenon. Phenomenology looks at data in order to identify how people view and understand their experiences, or it may look for the stories that people tell about their experiences and what such stories reveal about social relationships and processes.

In-Depth Interviews

Qualitative researchers use interviews to gather information from people about their individual attitudes, beliefs, and feelings. In contrast to survey questions, which are typically closed (yes/no questions or questions with a selection of answers provided by the researcher), in-depth interviews typically are open-ended and encourage people to talk freely in their own words about topics of relevance to the research. In-depth interviews have the advantage of allowing participants to talk about issues that the researcher may not have considered but that are important to understanding the phenomenon under investigation. In-depth interviews are usually audio recorded, then transcribed, and analysis focuses on categorizing and exploring text. Analysis can focus on finding themes that recur throughout their interviews: how people organize their talk, the stories they tell about their experiences, or some combination of these approaches (Silverman, 1993). Research that uses the method of in-depth interview is often supported by an approach called grounded theory. In this approach, researchers rely mostly on the data to explain or predict an event, process, or set of experiences (Straus & Corbin, 1998).

Content Analysis

Content analysis is a method that focuses on both texts and images and is widely used in media research. Content analysis can be used quantitatively, to count, for instance, how many times a particular issue is discussed in the press. Such data can be used to gauge what the media view as important topics. Qualitative content analysis focuses on how topics are represented and uses analytic frameworks such as semiotics and psychoanalysis to explore the social construction of images and ideas and what they reveal about society. Content analysis involves researchers looking at what is in a text or image and coding (categorizing or grouping in themes) what they (think) they see (Rose, 2001). Identifying themes can help the research explain how ideas or images are put together and what they seem to be saying about social life.

Ethnography

Ethnography is a qualitative research technique that relies on direct observation of a particular social group or culture through talking to members of the group and looking at documents. In ethnographic research, the researcher is the research instrument and is trying to understand what is going on in a particular setting. Ethnography typically involves lengthy participation or immersion in the everyday life of a chosen or natural setting—what anthropologists have described as being "in the field" (Pope, 2005). This method requires patience and time, since it often depends on waiting for things to happen. For instance, many health researchers have used ethnographic methods to explore what happens in medical settings (e.g., conversational exchanges between doctors and patients in health clinics) and how this affects understandings and experiences of health. The data being collected (words, actions) are typically recorded in field notes, or it may be audio and/or videotaped. While all qualitative research needs to be grounded in the informed consent of research participants, ethnographic research methods raise particular issues around negotiating access to the research setting (Ames, Thompson & Thurston, 2001). Sometimes, researchers need to gain prior institutional approval, such as, for instance, when the research involves observation in a hospital or prison.

Focus Groups

Many qualitative researchers use focus groups to collect data. Focus groups are viewed as social events that are characterized by organized discussion (Kitzinger, 1994). A focus group is a form of group interviewing, but it differs from a group interview in the sense that interaction between group members is key (Kitzinger, 1995). The main purpose of using focus group research is to gain insight into the participants' attitudes, feelings, beliefs, experiences, and reactions as they arise in the context of group interaction. The primary assumptions in using the focus group method are that such attitudes are more likely to be revealed in the context of group interaction and that group processes can help people explore and clarify their views in new ways (Webb & Kevern, 2001).

Focus groups are led or facilitated by a moderator — usually someone on the research team — who is skilled at directing conversation, summarizing points, and keeping conversation on track. The moderator uses an interview protocol to guide discussion, which, as with in-depth interviews, reflects the broad themes that the researcher wishes to cover. However, in a focus group, the moderator must be able to pursue topics that participants raise in the interaction process (Litosseliti, 2003) and deal with emotions that may arise in the discussion, such as anger or sadness.

Focus groups have advantages over in-depth interviews or observation methods. They allow researchers to explore topics in a nonthreatening environment and gain insight into a range of views within a group of participants who share a similar background, such as doctors, teachers, or mothers (Krueger, 1994). For this reason, they are often used in health and marketing research (Nicholls, 2009) and can be used to find out what patients and practitioners think about health care or medical practice, identify gaps in practitioner learning, and determine whether practitioners have learned something after an educational intervention.

According to Kalvemark et al., "focus groups are particularly useful when there are power differences between the participants and decision-makers or professionals, when the everyday use of language and culture of particular groups is of interest, and when one wants to explore the degree of consensus on a given topic" (2004, p. 1075). However, there are also "difficulties associated with the focus group method, since it involves group dynamics" that can pose challenges for the moderator. It can be especially challenging to ensure that all group participants are invited to express their personal views, since groups are often characterized by power imbalances and internal tensions.

Software

Qualitative research generates an enormous volume of unstructured data that can be messy and time consuming to analyze (that is, to sort into themes and build into theories). Conventional tools for sorting and analyzing qualitative data include index cards or paper, scissors, and paste. These tools allow researchers to trawl through an interview transcript to pull out chunks of text that are especially meaningful and can be linked to chunks of text in another interview transcript. However, most qualitative researchers now use software packages to help them store, sort through, and code or index their data. Most programs allow researchers to store, sort, chart, model, and link videos, photographs, documents such as interview transcripts, and audio files. This allows the researcher to more systematically search the data and create an audit trail that makes the analysis process more transparent.

Conclusion

In contrast to quantitative research, which focuses on research questions that need to be answered by identifying and measuring phenomena, qualitative research is used when the research questions are focused on people's experiences or how they interpret the social world. There are several methods to support qualitative research, all of which have in common the desire to understand and explain experiences from the participants' perspectives. Consequently, qualitative research methods are typically used in contexts where the participants' perspective is crucial, such as in health care or in market research. Finally, because qualitative research methods generate a vast volume of data, researchers typically use software programs that can help store, code, and rapidly retrieve data more systematically.

Terms & Concepts

Androcentric: A view of the world based on values that have come to be associated with maleness or masculinity, such as objectivity, detachment, and reason.

Chicago School: A school of sociology at the University of Chicago in the 1920s and 1930s that focused on the experiences and perspectives of participants in applied research in an urban setting.

Constructivism: An interpretivist philosophy of learning that emphasizes that people construct their view of the world through their social experiences.

Ethnography: A qualitative research technique that relies on direct observation of a particular social group or culture through talking to members of the group and looking at documents. In ethnographic research, the researcher is the research instrument.

Interpretivism: A philosophical perspective that argues, first, that social phenomena need to be examined as a whole and, second, that there is not one single reality but rather multiple realities.

Positivism: A philosophical tradition that emphasizes the importance of conducting social science research in a way that focuses strictly on causal relationships between behaviors and other social phenomena that can be measured or directly observed.

Reflexivity: A commitment to acknowledging and questioning the role of the researcher in collecting and analyzing data.

Transferability: The extent to which a research finding is transferable or generalizable to other populations.

Triangulation: A way of enhancing research credibility by collecting data on the topic from different sources or using more than one researcher to collect the data.

Validity: Internal validity refers whether or not a study has achieved its aims. External validity is the extent to which the results can be applied in other contexts.

Bibliography

Ames, J.C., Thompson, S. & Thuston, M.N. (2001). Difficulties in negotiating research access. *Bulletin of Medical Ethics*. (166), 15-7.

Anyan, F. (2013). The influence of power shifts in data collection and analysis stages: A focus on qualitative research interview. Qualitative Report, 18(18), 1–9. Retrieved October 25, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=87514136&site=ehost-live

Carpenter C. & Suto, M. (2008). *Qualitative research for occupational and physical therapists*. Blackwell, London.

Davidson, J. (2012). The journal project: Qualitative computing and the technology/aesthetics divide in qualitative research. Forum: Qualitative Social Research, 13(2), 1–30. Retrieved October 25, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=86880597&site=ehost-live

Gringeri, C., Barusch, A., & Cambron, C. (2013). Epistemology in qualitative social work research: A review of published articles, 2008–2010. Social Work Research, 37(1), 55–63. Retrieved October 25, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=83402718&site=ehost-live

Hallberg, L.R-M. (2006). The "core category" of grounded theory: Making constant comparisons. *International Journal of Qualitative Studies on Health and Well-being, 1* (3), 141-148. Retrieved January 19, 2010 from EBSCO online database, Academic Search Complete http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=22107442&site=ehost-live.

Haraway, D. (1991). *Simians, cyborgs and women: The reinvention of nature*. London: Free Association Books.

Harding, S. (1991). *Whose science? Whose knowledge? Thinking from women's lives*. New York: Cornell University Press.

Kalvemark, S.K., Hoglund, A.T., Hansson, M.G., Westerholm, P., & Arnetz, B. (2004). Living with conflicts-ethical dilemmas and moral distress in the health care system. *Social Science & Medicine, 58* (6), 1075-1084.

Kitzinger, J. (1994). The methodology of focus groups: The importance of interactions between research participants. *Sociology of Health and Illness, 16* (1), 103-121. Retrieved January 19, 2010 from EBSCO online database, SocINDEX with Full Text http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=11347023&site=ehost-live

Kitzinger, J. (1995). Qualitative research: Introducing focus groups. *British Medical Journal, 311*, 299-302.

Krueger, R.A. (1994). *Focus groups: A practical guide for applied research*. Thousand Oaks, CA: Sage Publications.

Kuhn, T. (1970). *The structure of scientific revolutions*. 2nd Ed. Chicago: University of Chicago Press.

Lincoln, Y.S. & Cannella, G. (2004). Qualitative research, power, and the radical right. *Qualitative Inquiry, 10* (2), 175-201.

Litosseliti, L. (2003). *Using focus groups in research*. London and New York: Continuum International Publishing.

Malterud, K. (2001). Qualitative research: Standards, challenges, and guidelines. *Lancet, 358* (9280), 483-458. Retrieved January 19, 2010 from EBSCO online database, Academic Search Complete http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=4984151&site=ehost-live.

Mays, N. & Pope, C. (2000). Qualitative research in health care: Assessing quality in qualitative research. *British Medical Journal, 320* (7226), 50-52. Retrieved January 24, 2010 from EBSCO online database, MEDLINE with Full Text http://search.ebscohost.com/login.aspx?direct=true&db=mnh&AN=10617534&site=ehost-live.

Miles, M. & Huberman, A. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.

Miller, W.L. & Crabtree, B.F. (1999). Clinical research: A multimethod typology and qualitative roadmap. 3-30. In, B.F. Crabtree and W.L. Miller (Eds). *Doing qualitative research*. 2nd ed. Thousand Oaks: Sage Publications.

Nettleton, S. (2006). *The sociology of health and illness*. 2nd ed. London: Wiley.

Nicholls, D. (2009). Qualitative research: Part one — philosophies. *International Journal of Therapy and Rehabilitation,16* (10): 526-533. Retrieved January 19, 2010 from EBSCO online database, Academic Search Complete http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=44500594&site=ehost-live.

Pope, C. (2005). Conducting ethnography in medical settings. *Medical Education, 39*(12), 1180-1187. Retrieved January 19, 2010 from EBSCO online database, Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=18943121&site=ehost-live.

Rose, G. (2001). *Visual methodologies*. London: Sage.

Shields, L. & Twycross, A. (2003). The difference between quantitative and qualitative research. *Paediatric Nursing, 15*(9): 24. Retrieved January 19, 2010 from EBSCO online database, Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=11354676&site=ehost-live.

Silverman, D. (1993). *Interpreting qualitative data: Methods for analyzing talk, text and interaction*. London: Sage.

Strauss, A. & Corbin, J. (1998). *Basics of qualitative research techniques and procedures for developing grounded theory.* (2nd ed.). London: Sage Publications.

Webb, C., & Kevern, J. (2001). Focus groups as a research method: A critique of some aspects of their use in nursing research. *Journal of Advanced Nursing, 33*(6), 798-805. Retrieved January 19, 2010 from EBSCO online database, Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=5441440&site=ehost-live

Whyte, W.F. (1993 [1943]). *Street corner society: The social structure of an Italian slum*. Chicago: University of Chicago Press.

Suggested Reading

Atkinson, P., Coffey, A. & Delamont, S. (2003). *Key themes in qualitative research: Continuities and change*. Walnut Creek, CA: AltaMira Press.

Charmaz, K. (2004). Premises, principles, and practices in qualitative research: Revisiting the foundations. *Qualitative Health Research, 14*(7), 976-993.

Charmaz, K. (2006). *Constructing grounded theory. A practical guide through qualitative analysis.* Thousands Oaks, CA: Sage.

Denzin, N.K. & Lincoln, Y.S. (2005). Introduction: The discipline and practice of qualitative research. 1-32. In, N.K. Denzin and Y.S. Lincoln, (eds). *The Sage Handbook of Qualitative Research*. 3rd ed. Thousands Oaks, CA: Sage.

Gringeri, C., Barusch, A., & Cambron, C. (2013). Examining foundations of qualitative research: A review of social work dissertations, 2008–2010. Journal of Social Work Education, 49(4), 760–773. Retrieved October 25, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=90595347&site=ehost-live

Kumar, R. (2013). How are women's experiences of childbirth represented in the literature? A critical review of qualitative health research set in the global South. Women's Health & Urban Life, 12(1), 19–38. Retrieved October 25, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=87316553&site=ehost-live

Thorne, S. & Darbyshire, P. (2005). Land mines in the field: A modest proposal for improving the craft of qualitative health research. *Qualitative Health Research, 15*(8), 1105-13.

Essay by Alexandra Howson

Alexandra Howson, PhD, taught sociology for over a decade at several universities in the UK. She has published books and peer-reviewed articles on the sociology of the body, gender, and health and is an independent researcher, writer, and editor based in the Seattle area.

Quantitative & Qualitative Analysis

Table of Contents

Abstract

In the physical sciences, data are typically measurable and quantifiable so that they can be meaningfully analyzed using inferential statistics. In the social and behavioral sciences, on the other hand, not all data of interest can be reduced to numbers in this way. Therefore, it is important that social and behavioral scientists collect both quantitative and qualitative data to better understand behavior. Qualitative research paradigms (i.e., field observation, survey research, and secondary analysis) give researchers a depth and breadth of the understanding of human behavior that cannot be otherwise gained. However, it is through quantitative research paradigms (i.e., experiments, simulation, field experiments) that hypotheses can be tested and meaningful predictions made of real-world behavior. Both qualitative and quantitative data and their concomitant research paradigms are important parts of social and behavioral science research, and are essential to understanding behavior and advancing these sciences.

Overview

When one thinks of research methodologies and the scientific method, variables, experimental design, and inferential statistical tools usually comes to mind. Although the experimental paradigm is assuredly part of the social and behavioral science researcher's toolkit, the questions investigated by social and behavioral scientists tend to be complex, and researchers frequently need to test models that contain numerous independent and dependent variables, as well as extraneous variables that may affect behavior but cannot be easily tested using inferential statistics. Further, research designs for use with human subjects are often limited by practical and ethical problems. For example, it may be logistically impossible to control all the variables that may affect the way that one is perceived in the workplace. Similarly, although it may be possible to manipulate the independent variable of whether one's spouse lives or dies in order to determine its effect on depression, it would certainly be both unethical and illegal to do so. In addition, constructs such as attitudes, opinions, and beliefs often do not translate well into quantifiable data on a scale with a real zero and equal intervals between points. Fortunately, social and behavioral scientists are not limited to the use of experimental paradigms in their quest to better understand and predict human behavior. There are a number of research paradigms available that can help researchers in their tasks. The choice of which tool to use depends on the goals of the research study and any practical considerations that may limit the degree to which the researcher can control the variables in the study.

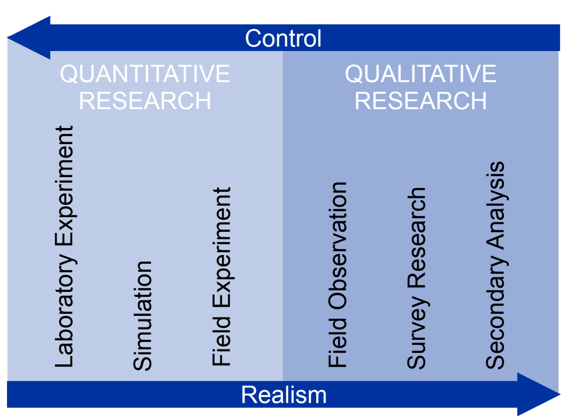
The continuum of research paradigms can be broken into two general categories. Quantitative research comprises research studies in which observations are measured and expressed in numerical form, such as in physical dimensions or on rating scales. The results of quantitative research studies are typically analyzed through the use of inferential statistics. Quantitative research paradigms also offer the researcher varying amounts of control over the research situation. However, even when multivariate statistical tools are used for the data analysis, quantitative research paradigms are restricted in scope.

Qualitative research paradigms, on the other hand, enable the researcher to look deeper into the data but generally allow less control over the research situation. Qualitative research comprises studies in which observations are not or cannot be quantified, that is, expressed in numerical form.

Quantitative Research Paradigms

There are three primary paradigms for quantitative research: laboratory experiments, simulations, and field experiments. These research paradigms offer scientists various degrees of control over the research situation and the extent to which the situation realistically reflects the complexity of the real world. The results of these research paradigms typically allow researchers to apply deductive logic and reason from a general principle to predict behavior in specific instances. As is illustrated in Figure 1, the more the research situation reflects the complexity of the real world, the less control the researcher has over the situation. Laboratory experiments allow researchers the most control, over both the level of the independent variable that is experienced by the subjects and the various extraneous variables that can affect the outcome of the study. For example, if one wanted to determine how different personality types behave in specific situations, one could set up a simple laboratory experiment in which subjects with different personality types were exposed to situations that contained what the researcher believed to be the important characteristics of the situations included in the hypothesis. Subjects might be exposed to two confederates of the researcher, one who acted neutrally and another who acted aggressively. Responses of the subjects could then be measured against some objective criterion and the results statistically analyzed. This approach to data collection gives the researcher a great deal of control over the experimental situation (e.g., whether the subject is exposed to the neutral or aggressive confederate). However, interactions with an experimental confederate tend to be far removed from interactions with people during the course of an actual day in the real world. Because of this fact, the results of the controlled experiment would not necessarily be very generalizable.

Figure 1: Research Paradigms Used in Quantitative & Qualitative Research



If the researcher were willing to give up some degree of control over the experimental situation, he or she could design an experimental condition with more realism. For example, if the researcher is trying to predict how people with different personality types react to various customer attitudes in the workplace, a simulation could be designed. The subject could play the role of the sales clerk, and the experimental confederate could play the role of the customer. The confederates would be assigned to play the role either in a neutral manner or in aggressive manner. The researcher could then measure the reaction of the subjects to the confederates and statistically analyze the results. This simulation design is more complex than the laboratory design. In addition to interacting with the experimental confederates, the subject also needs to perform the tasks of the sales clerk. This brings many other variables into the situation that are not directly related to the research question, such as stress levels resulting from the sales tasks and previous experience as a sales clerk. As a result, the researcher has less control over the situation than he or she does in the laboratory. However, the simulation offers the researcher data from a more realistic scenario, meaning that the results will likely be more generalizable than the results of the laboratory experiment.

If the researcher were willing to give up even more control of the experimental situation, he or she could conduct a field experiment instead of a laboratory experiment or a simulation. In this paradigm, the confederates might interact with the subjects in a real-world sales setting. For example, the subjects could be actual sales clerks. During the course of their daily activities at work, they would interact with the experimental confederates, who would be posing as customers. This is an even more realistic situation than either the simulation or the laboratory experiment. However, such a field experiment also introduces numerous other variables that are not present in the other two scenarios. For example, the experimenter has no control over what other types of interactions the subject has during that day or how other extraneous variables might impact the subject's reaction.

A real-world situation tends to be very complex. The reaction of a subject to an experimental confederate may depend not only on the confederate's behavior but also on numerous other variables, such as how the subject is feeling that day, whether or not the sales counter is very busy, whether the subject's last interaction with a customer was positive or negative, and what types of interactions the subject has had with other store personnel or supervisors that day, just to name a few. It is often virtually impossible for researchers to sufficiently articulate all the real-world variables that influence behavior in a way that allows a hypothesis to appear to be empirically tested using inferential statistics. In the real world, there are seemingly endless variables that need to be considered. As part of the scientific method, the researcher needs to observe real-world situations and gather data that can be used as part of the inductive reasoning process in order to develop theories about behavior in the real world. This is typically done through the paradigms of qualitative research.

Qualitative Research Paradigms

There are three primary paradigms for qualitative research: field observation, either participatory or non-participatory; survey research; and secondary analysis. Field observation and research allow the researcher no control over the experimental situation. However, data gathered through real-world observation allow the researcher to apply inductive reasoning to better understand the parameters of behavior and to generate testable hypotheses. Because of the complexity of human behavior in real-world situations, observing people in field settings often allows researchers to better understand the interaction of variables causing their behavior than observation in more controlled settings. In field observation and research, researchers directly observe behavior in natural, real-world situations. Field observations can be done with the researcher acting either as a participant or as a non-participant in the situation. For example, a researcher interested in what a suspect goes through during the period from arrest through arraignment could observe suspects interacting with law enforcement and criminal justice personnel and take notes on how they are treated and how they react. This is an example of non-participatory observation. Alternatively, the researcher could participate in these same interactions in order to better understand the feelings and actions of the suspects. In this participatory scenario, the researcher might pose as a suspect and allow himself or herself to be arrested, booked, and arraigned, taking notes on his or her reactions at each stage of the procedure.

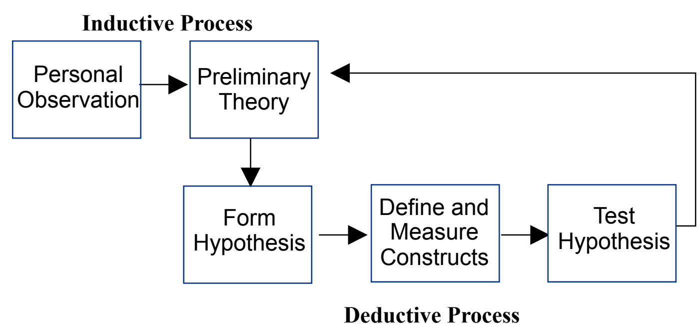
Another way to collect qualitative data is through survey research. In this paradigm, data about the opinions, attitudes, or reactions of the members of a sample are gathered using a survey instrument that is administered in a paper-and-pencil or electronic form or by an interviewer. Survey research has the advantage of allowing the researcher to collect information on non-tangible constructs, such as feelings, attitudes, and opinions, that are difficult to collect directly. In addition, surveys can be relatively inexpensive to administer, as they require no manipulation of variables and have relatively low costs associated with data collection. However, even well-written surveys cannot always provide the researcher with the answers to all questions of interest in a research study. What one says and what one does are often two completely different things. Research subjects may lie on a survey instrument in order to look good to the researcher, or even to themselves, or they may give responses that are not well thought out because they are not motivated to participate honestly in the survey.

Data for qualitative analysis can also be collected through existing records. In this paradigm, often referred to as secondary analysis, researchers review and analyze data found in actuarial records, political or judicial records, government records, the mass media, or any other hard copy or electronic document. The analysis of data found in archives does not require the researcher to interact with the subject. However, archival information is limited because the archival data necessary to investigate a research question is not always available to the researcher.

Applications

Both qualitative and quantitative methods have their place in social and behavioral science research. Paradigms resulting in qualitative data enable the researcher to look at a breadth and depth of behavior that cannot be accomplished in scenarios that generate quantitative data. On the other hand, the fact that quantitative data can be analyzed with inferential statistics makes them very useful for interpreting the results of a study and confirming or refining theories. As shown in Figure 2, the scientific method in general and the theory-building process specifically are aided by both qualitative and quantitative methods. The observations that result in qualitative data are invaluable for use in the inductive process, which culminates in the development of a testable research hypothesis, while the experimental paradigms that enable the collection of quantitative data that can be statistically analyzed contribute greatly to the deductive process, helping the researcher either confirm the hypothesis or determine where refinement is needed. Any necessary refinement is done through further observation and inductive reasoning, followed by additional hypothesis testing as the theory is refined to better reflect the real world.

Figure 2: The Theory Building Process



Because of the importance of both qualitative and quantitative data to the advancement of science, scientists in these fields are increasingly looking to methodologies that include both aspects in order to better understand and describe human behavior. Although hard, quantitative data are deemed superior to soft, qualitative data in some sectors, both are necessary if one is to understand behavior in the real world. Both qualitative and quantitative methods for social science research have advantages and disadvantages. In general, researchers employ quantitative methods to have greater emotional distance from their subjects. In some ways, this is good. It is necessary for researchers to maintain some measure of objectivity in order to not read their preconceived notions into the interpretation of the data. On the other hand, the emotional distance required in order to design and conduct a research experiment does not necessarily give one the depth and understanding of a real-world situation necessary to fairly interpret the results. Both qualitative and quantitative data are required in order to do this well.

The real world is a complex place, and human behavior in the real world can be impacted by seemingly endless variables. Increasingly, social and behavioral scientists are taking a more holistic approach to research. Qualitative data can be used to support and help interpret the results of quantitative studies, while quantitative data can be used to interpret and validate qualitative observations. Further, for many real-world situations, it is not possible to collect truly quantifiable data. Many inferential statistical tools make assumptions about the underlying shape of the distribution, the nature of the rating scale, and the existence of a meaningful zero point—expectations that simply cannot be met by many social and behavioral research studies. For example, although the attitudes, opinions, and feelings that are collected through survey research methods are often quantified and statistically analyzed, such analysis sometimes violates the assumptions of parametric statistics. Although a ten-point attitude scale can be used to collect data about subjects' opinions, without a meaningful zero for the scale, it is difficult to know what the various responses mean. Further, there is no way to determine whether or not the difference between a score of one and a score of two is the same as the difference between a score of seven and a score of eight on such a scale, or if every subject interprets the score values in the same way.

Conclusion

The data used in social and behavioral research can be either qualitative or quantitative. Although qualitative research paradigms such as field observation, survey research, and secondary analysis can enable the researcher to understand behavior in great depth and breadth, the results of such data collection methods cannot be analyzed using the tools of inferential statistics. Conversely, quantitative data can be statistically analyzed, but such data are often more limited in scope. The levels of control and analyzability that are gained by using quantitative data are balanced by the realism and depth of understanding that can be gained by using qualitative data. Ultimately, both qualitative and quantitative data are essential to understanding behavior and advancing social and behavioral sciences.

Terms & Concepts

Confederate: A person who assists a researcher by pretending to be part of the experimental situation, playing a rehearsed part meant to stimulate a response from the research subject.

Deductive Reasoning: A type of logical reasoning in which it is demonstrated that a conclusion must necessarily follow from a sequence of premises, the first of which is a self-evident truth or agreed-upon data point or condition. Deductive reasoning is the process by which predictions are drawn from general laws or theories.

Ethics: In scientific research, a code of moral conduct regarding the treatment of research subjects that is subscribed to by the members of a professional community. Many professional groups have a specific written code of ethics that sets standards and principles for professional conduct and the treatment of research subjects.

Experiment: A situation under the control of a researcher in which an experimental condition (independent variable) is manipulated and the effect on the experimental subject (dependent variable) is measured. Most experiments are designed using the principles of the scientific method, and the results are analyzed to determine whether or not they are statistically significant.

Hypothesis: An empirically testable declaration that certain variables and their corresponding measures are related in a specific way proposed by a theory.

Inductive Reasoning: A type of logical reasoning in which inferences and general principles are drawn from specific observations or cases. Inductive reasoning is a foundation of the scientific method and is the process by which testable hypotheses are formed from particular facts and observations.

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data, as well as in decision making.

Model: A representation of a situation, system, or subsystem. Conceptual models are mental images that describe the situation or system. Mathematical or computer models are mathematical representations of the system or situation being studied.

Multivariate Statistics: A branch of statistics that is used to summarize, represent, and analyze multiple quantitative measurements obtained from a number of individuals or objects. Examples of multivariate statistics include factor analysis, cluster analysis, and multivariate analysis of variance (MANOVA).

Qualitative Research: Scientific research in which observations cannot be or are not quantified, that is, expressed in numerical form.

Quantitative Research: Scientific research in which observations are measured and expressed in numerical form, such as in physical dimensions or on rating scales.

Scientific Method: General procedures, guidelines, assumptions, and attitudes required for the organized and systematic collection, analysis, interpretation, and verification of data that can be verified and reproduced. The goal of the scientific method is to articulate or modify the laws and principles of a science. Steps in the scientific method include problem definition based on observation and review of the literature, formulation of a testable hypothesis, selection of a research design, data collection and analysis, extrapolation of conclusions, and development of ideas for further research in the area.

Subject: A participant in a research study or experiment whose responses are observed, recorded, and analyzed.

Survey Research: A type of research in which data about the opinions, attitudes, or reactions of the members of a sample are gathered using a survey instrument. The phases of survey research are goal setting, planning, implementation, evaluation, and feedback. Unlike experimental research, survey research does not allow for the manipulation of an independent variable.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables, also known as the response. Extraneous variables are variables that affect the response but are not related to the question under investigation.

Bibliography

Allwood, C. (2012). The distinction between qualitative and quantitative research methods is problematic. *Quality & Quantity, 46*(5), 1417–1429. Retrieved November 6, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=77493397&site=ehost-live

Anderson, M. L. & Taylor, H. F. (2002). *Sociology: Understanding a diverse society* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Masue, O. S., Swai, I. L., & Anasel, M. G. (2013). The qualitative-quantitative "disparities" in social science research: What does qualitative comparative analysis (QCA) brings in to bridge the gap? *Asian Social Science, 9*(10), 211–221. Retrieved November 6, 2013, from EBSCO Online Database Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=89892948&site=ehost-live

Frels, R. K., & Onwuegbuzie, A. J. (2013). Administering quantitative instruments with qualitative interviews: A mixed research approach. *Journal of Counseling & Development, 91*(2), 184–194. Retrieved November 6, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=85872801&site=ehost-live

Žydžiumaite, V. (2007). Methodological considerations: Sequential linking of qualitative and quantitative research. *Social Sciences, 55* (1), 7-14. Retrieved 23 April 2008 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=25228925&site=ehost-live

Suggested Reading

Bischoping, K. (2005, Jan). Social research methods: Qualitative and quantitative approaches/the practice of social research. *Teaching Sociology, 33* (1), 95-97. Retrieved 23 April 2008 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=16053114&site=ehost-live

Coxon, A. P. M. (2005 May). Integrating qualitative and quantitative data: What does the user need? *Forum: Qualitative Social Research, 6* (2), 1-9. Retrieved 23 April 2008 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=17713334&site=ehost-live

Creswell, J. W. (2008). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed). Thousand Oaks, CA: Sage Publications.

Li, H. Z. (2003, Jun). Inter- and intra-cultural variations in self-other boundary: A qualitative-quantitative approach. *International Journal of Psychology, 38* (3), 138-149. Retrieved 23 April 2008 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=10875135&site=ehost-live

Morse, J. M. (2006). Insight, inference, evidence, and verification: Creating a legitimate discipline. *International Journal of Qualitative Methods, 5* (1), 1-7. Retrieved 23 April 2008 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=21331314&site=ehost-live

Rabinowitz, V. C. & Weseen, S. (1997, Win). Elu(ci)d(at)ing epistemological impasses: Re-viewing the qualitative/quantitative debates in psychology. *Journal of Social Issues, 53* (4), 605-630.

Tacq, J. (2011). Causality in qualitative and quantitative research. *Quality & Quantity, 45*(2), 263–291. Retrieved November 6, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=56588602&site=ehost-live

Essay by Ruth A. Wienclaw, PhD

Dr. Ruth A. Wienclaw holds a doctorate in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human-systems integration.

Sociology & Probability Theory

Table of Contents

Abstract

Probability theory is a branch of mathematics that deals with the estimation of the likelihood of an event occurring. An integral part of the scientific method and the principles of probability theory are applied through inferential statistical techniques in the analysis of data to determine the likelihood that a hypothesized relationship between variables exists. Applied probability theory, however, does not "prove" that a hypothesis is correct: it only expresses the confidence with which one can state that variables are related to each other in a way stated by a theory. Even when the results of a research study are statistically significant, the researcher still accepts the possibility of error by either rejecting the null hypothesis when it is, in fact, true (Type I error), or accepting the null hypothesis when it is, in fact, false (Type II error).

Overview

It is typically safe to assume that the majority of students studying sociology or other behavioral and social sciences are not doing so primarily because of their interest in pure mathematics. People interested in math and science are usually more immediately drawn to the "hard" sciences, where things can be weighed, measured, and counted and where the difference between a score of 12.00 and a score of 12.01 has a tangible meaning. People drawn to the social and behavioral sciences tend to be more interested in the wide variety and great unpredictability of human behavior. However, mathematical and scientific tools are essential to sociologists and other behavioral scientists as they seek to understand, interpret, and predict the behaviors that they see around them.

Unfortunately, too often the social sciences are considered by outsiders to be nothing more than the mere articulation of "common sense." Yet common sense is often not at all common, and examining one's own motivations and behaviors yields only limited insight into the motivations and behaviors of others. I may know why I act the way I do in certain situations, but it is not logically valid to generalize my knowledge about my opinions or behavior to apply to other people.

For example, I may have strong feelings about a certain political candidate and cast my vote accordingly. It might seem, therefore, that everyone confronted with the same evidence should vote the same way I do. However, neither horse races nor elections are that easy to predict. I cannot simply extrapolate my opinions and voting behavior to predict how others will vote. In order to do so, I must understand my deeper motivations, as well as the factors that impact others' decisions. For example, I may favor a certain candidate because of a speech he or she made, an article that I read in the newspaper, or the candidate's record in previous elected positions. However, it is unwise to base my vote on one speech, article, or even record. Speeches tend to be stylized and without deep content, newspaper columnists often have their own agendas and fail to be objective, and public records do not necessarily reflect how a candidate will perform if elected into a different office, or if he or she has changed his or her mind about a political issue. Someone who disagrees with my choice of political candidate may possess some of these missing pieces of information and, therefore, reach a different conclusion than mine. Alternatively, the person disagreeing with me may not have all the information that I possess. Social scientists try to unravel these and other complex issues by studying large populations of people.

When one talks about human behavior, one must consider other factors as well. Human beings cannot necessarily judge evidence objectively. One person may always vote for candidates in a certain political party because his or her parents did so. Another person may vote for a political party because he or she dislikes the general platform of the opposing party. Someone else may believe that a candidate belonging one social group could not possibly understand or fairly represent the needs of another social group and vote accordingly. Still others might vote for a candidate because of the way the candidate dresses or because he or she appears to be a "nice" person.

Because the human decision-making process is so complex, it is virtually impossible to generalize from the opinions and behaviors of one person, or even a small group of people, to make conclusions about how all human beings think or behave. For this reason, it is important to use the scientific method and probability theory to better understand why people act the way they do. Without an understanding of probability theory, one can easily fall into the trap of thinking that a single set of statistics "proves" a theory, or that the results of one study will be replicated in all future studies on the same topic. Because it is impossible to extrapolate from the behaviors and motivations of one individual to the behaviors and motivations of individuals in general, it is necessary to use scientific and mathematical tools to better interpret the world around us. Probability theory, and the inferential statistical tools that are based on it, helps scientists and researchers determine whether the results that they observe are due to chance or to some other underlying cause.

Applications

Virtually every semester, at least one of my students proudly announces that the statistical analysis he or she has performed on his or her research data "proves" that his or her hypothesis is right. As satisfying as this conclusion might be, in truth, statistics do not prove anything, nor is the scientific method a quest for proof. Statistics merely express confidence and describe probabilities concerning whether or not the null hypothesis is more likely to be true than the alternate hypothesis. This fact is frequently demonstrated in scientific literature when one group of scientists attempts to replicate the research of another group and finds that their research results lead to conclusions that are different from the original group's. A lack of understanding of the way that probability works can lead to poor experimental design and spurious results. The results of a statistical data analysis do not prove whether or not one's hypothesis is true, only whether or not there is a probability of the hypothesis being true at a given confidence level. So, for example, if a *t*-test or analysis of variance yields a value that is significant at the $p = .05$ level, this does not mean that the hypothesis is true; it means that the analyst runs the risk of being wrong 5 times out of 100.

Inferential Statistics

Without an understanding of probability theory and what statistics can and cannot accomplish, it may be tempting to look at the results of a research study or experiment, apply a few descriptive statistical techniques, and draw a conclusion about whether or not one's hypothesis is true. However, the objects of scientific study rarely yield black-and-white results. Even in the physical sciences, results can vary depending on the conditions under which a study was done. Therefore, inferential statistics are used to test hypotheses to determine if the results of a study have statistical significance, meaning that they occur at a rate that is unlikely to be due to chance, and to evaluate the probability of the null hypothesis (H0) being true. Inferential statistics allow the researcher to make inferences about the qualities or characteristics of the population that are based on observations of a sample. However, to understand what the results of statistical tests mean, one needs to understand the influence of probability on statistics and the abilities and limits of statistics.

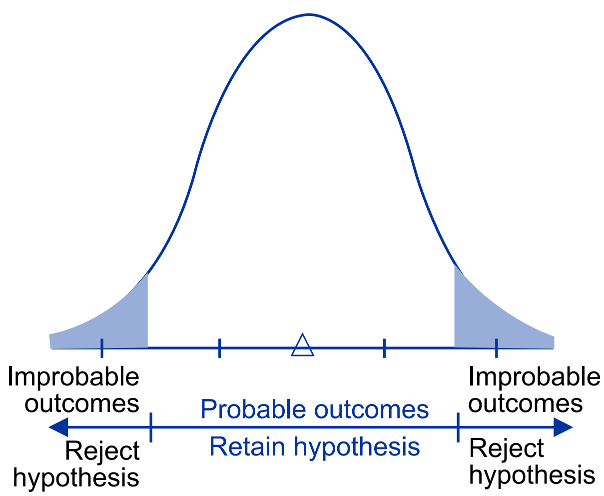
Hypothesis Testing

A hypothesis is an empirically verifiable declarative statement that the independent and dependent variables and their corresponding measures are related in a specific way as proposed by a theory. The independent variable is the variable that is manipulated by the researcher. For example, if a behavioral researcher wants to know if people are treated differently at work depending on how they dress, the independent variable in this hypothesis is the type of attire that people wear to work. The dependent variable—in this case, the way people are treated at work—is the variable whose value depends on the value of the independent variable.

For purposes of statistical testing, a hypothesis is stated in two ways. The null hypothesis (H0) is the statement that there is no statistical difference between the status quo and the observations of the researcher after the manipulation of the independent variable. In other words, it states that the variable being studied (e.g., the way that people dress at work) has no bearing on the end result (e.g., the way that people are treated at work). This means that if the null hypothesis is true, the manipulation of the independent variable in an experiment does not change the results. The alternative hypothesis (H1) states that there *is* a relationship between the two variables (e.g., people who dress in a business-like manner get better treatment in the workplace). In general, inferential statistical tests are used to test the probability of the null hypothesis (H0) being true.

When one accepts the null hypothesis, one is concluding that if the data in the underlying population are normally distributed, the results observed in a research study are more than likely due to chance. This is illustrated in Figure 1 as the unshaded portion of the distribution. Going back to the previous example, a researcher who accepts the null hypothesis is stating that it is likely that people who wear business suits and people who wear casual clothing are not treated any differently in the workplace. However, if the null hypothesis is rejected and the alternate hypothesis is accepted, the researcher is concluding that the results are unlikely to have occurred due to chance and are more likely attributable to some underlying factor. This conclusion assumes that there is a statistically significant difference between the way that the two groups are treated, due not to chance but rather to a real underlying difference in people's attitudes about how others dress in the workplace. Thus, since the results are statistically improbable, the differences observed in the study would lie within the shaded portions of the graph.

Figure 1: Hypothesized Sampling Distribution of the Mean Showing Areas of Acceptance and Rejection of the Null Hypothesis



(adapted from Witte, p. 118)

The results of statistical hypothesis testing are expressed in terms of probability, or the likelihood of an event occurring. Probability is the mathematical expression of the number of actual occurrences in relation to the number of possible occurrences of the event, expressed as a value between 0 and 1.0. A probability of 0 signifies that there is no chance that the event will occur, while 1.0 means that the event is certain to occur. When used within the paradigm of the scientific method, probability theory can be applied to real-world situations to estimate whether or not the observed results of the research are more likely to be due to an underlying cause or to mere chance.

Several steps are involved in this estimation. First, the researcher must define a population that has certain parameters. Continuing with the example of voting preferences used above, one might want to define the population as all voters within the United States, a two-part requirement: one part distinguishes between people who vote and people who do not vote, and the other part distinguishes between people who vote in the United States and people who vote in other countries that may have different political systems. Second, the researcher needs to draw a random sample from this defined population under the assumption that such a random draw will decrease the possibility of introducing bias into the sample. Third, the researcher needs to be able to reason probabilistically. In other words, he or she must be willing to run the risk of being wrong and must specify what an acceptable probability of being wrong is (e.g., 5 times out of 100). By specifying this probability, the researcher expresses the threshold of the probability of occurrence at which he or she would feel confident rejecting the null hypothesis. Finally, the researcher needs to set up a null hypothesis and an alternative hypothesis that cover the range of possibilities regarding the research question. The normal curve, as shown in Figure 1 above, represents all possible values for the results of the study. The shaded areas in the distribution represent the unlikely possibility that the results are not due to chance—typically a 1 or 5 percent probability—while the larger area represents the possibility that the observed results are due to chance.

Interpreting Statistical Results

As discussed above, statistics does not state with certainty whether or not the null hypothesis is correct; it only estimates the probability of it being correct. Therefore, because of the laws of probability, no matter whether one accepts or rejects the null hypothesis, there is always a possibility of error when interpreting statistics. When interpreting statistical results, two types of error are possible. Type I error, also referred to as α error, occurs when one incorrectly rejects the null hypothesis and accepts the alternate hypothesis. For example, a Type I error would have occurred if the researcher concluded that people who dress more formally in the workplace receive better treatment when, in fact, the way people dress has no effect on how they are treated. Type II error, also referred to as β error, occurs when one incorrectly accepts the null hypothesis. For example, a Type II error would have occurred if the researcher interpreted the statistical results to mean that people receive the same treatment in the workplace no matter how they dressed when, in fact, people who dress in a more businesslike manner tend to receive more advantages or rewards. As one decreases the likelihood of the alpha error, one concomitantly increases the likelihood of the beta error. However, this does not occur proportionately. The goal, therefore, is to determine the best balance between alpha and beta errors. For this reason, a p-value of .05 is usually taken as a minimum requirement for rejecting the null hypothesis. The conditions of Type I and Type II errors are shown in Table 1.

Table 1: Types of Error in Statistical Decision Making

	… and H_0 Is True	…and H_0 Is False
If reject H_0…	Type I error (α)	OK
If accept H_0…	OK	Type II error (ß)

Conclusion

In order to better understand, interpret, and predict the phenomena around us, researchers cannot rely solely on their own experience or motivations; they must apply probability theory in the form of inferential statistics to the data they collect. Probability theory, and the inferential statistical tools that are based on it, helps scientists and researchers determine whether the results that they observe are due to chance or to some other underlying cause. Without an understanding of probability theory and what statistics can and cannot accomplish, it can be tempting to look at the results, perform a few descriptive statistical techniques, and draw a conclusion about whether or not one's hypothesis is true. Such results tend to be specious, however. It is the application of the scientific method, including statistical analysis, that determines whether a body of knowledge is philosophy or science.

Terms & Concepts

Bias: The tendency for a given experimental design or implementation to unintentionally skew the results of the experiment due to a nonrandom selection of participants.

Data: In statistics, quantifiable observations or measurements that are used as the basis of scientific research.

Descriptive Statistics: A subset of mathematical statistics that describes and summarizes data.

Distribution: A set of numbers collected from data and their associated frequencies. A normal distribution is a continuous distribution that is symmetrical about its mean and asymptotic to the horizontal axis. The area under a normal distribution is 1. The normal distribution, also called the Gaussian distribution or the normal curve, describes many characteristics observable in the natural world.

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data, as well as in decision making.

Null Hypothesis (H0): The statement that the findings of an experiment will show no statistically significant difference between the control condition and the experimental condition.

Population: The entire group of subjects belonging to a certain category, such as all women between the ages of 18 and 27, all dry-cleaning businesses, or all college students.

Probability: A branch of mathematics that deals with estimating the likelihood of an event occurring. Probability is expressed as a value between 0 and 1.0, which is the mathematical expression of the number of actual occurrences compared to the number of possible occurrences of the event. A probability of 0 signifies that there is no chance that the event will occur, while 1.0 signifies that the event is certain to occur.

Sample: A subset of a population. A random sample is a sample that is chosen at random from the larger population with the assumption that it reflects the characteristics of the larger population.

Scientific Method: General procedures, guidelines, assumptions, and attitudes required for the organized and systematic collection, analysis, and interpretation of data that can then be verified and reproduced. The goal of the scientific method is to articulate or modify the laws and principles of a science. Steps in the scientific method include problem definition based on observation and review of the literature, formulation of a testable hypothesis, selection of a research design, data collection and analysis, extrapolation of conclusions, and development of ideas for further research in the area.

Statistics: A branch of mathematics that deals with the analysis and interpretation of data. Mathematical statistics provides the theoretical underpinnings for various applied statistical disciplines, including business statistics, in which data are analyzed to find answers to quantifiable questions. Applied statistics uses these techniques to solve real-world problems.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables, or response variables. Extraneous variables are variables that affect the dependent variables but are not related to the question under investigation.

Bibliography

Armore, S. J. (1966.) *Introduction to statistical analysis and inferences for psychology and education*. New York: John Wiley & Sons.

Courgeau, D. (2012). *Probability and social science: Methodological relationships between the two approaches.* Dordrecht: Springer. Retrieved November 6, 2013, from EBSCO Online Database eBook Collection. http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=523080&site=ehost-live

Hanson-Hart, Z. (n.d.). *Statistical reasoning*. Retrieved July, 24 2007, from http://www.math.temple.edu/~zachhh/ch5.pdf

Huff, D. (1954). *How to lie with statistics*. New York: W. W. Norton & Company.

Uprichard, E. (2013). Sampling: Bridging probability and non-probability designs. *International Journal of Social Research Methodology, 16*(1), 1–11. Retrieved November 6, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=85318568&site=ehost-live

Witte, R. S. (1980). *Statistics*. New York: Holt, Rinehart and Winston.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and Internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly, 75*(4), 709–747. Retrieved November 6, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=67133710&site=ehost-live

Suggested Reading

Eğecioğlu, Ö., & Giritligil, A. E. (2013). The impartial, anonymous, and neutral culture model: A probability model for sampling public preference structures. *Journal of Mathematical Sociology, 37*(4), 203–222. Retrieved November 6, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=90274270&site=ehost-live

Ghilagaber, G. (2005). Incompatibility between hazard- and logistic-regression in modeling survival data with multiple causes of failure. *Quality & Quantity, 39* (1), 37-44. Retrieved April, 7 2008 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=16525781&site=ehost-live

Sartori, R. (2006). The bell curve in psychological research and practice: Myth or reality? *Quality & Quantity, 40* (3*),* 407-418. Retrieved April, 7 2008 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=20907730&site=ehost-live

van den Berg, G. J., Lindeboom, M., & Doton, P. J. (2006). Survey non-response and the duration of unemployment. *Journal of the Royal Statistical Society, 169* (3), 585-604. Retrieved April 7, 2008 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=21097418&site=ehost-live

Essay by Ruth A. Wienclaw, PhD

Dr. Ruth A. Wienclaw holds a doctorate in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human-systems integration.

Variables in Sociological Research

Table of Contents

Abstract

One of the key factors in developing and testing scientific theories is the identification and operational definition of the variables that describe phenomena observed in the real world. Variables are measured in a research study, and they can have more than one value. There are several types of variables of interest to the researcher. Independent variables are stimuli that are manipulated in order to determine their effect on the value of dependent variables. Extraneous variables are variables that affect the value of the dependent variable but that are not related to the question under investigation in the study. Intervening variables are variables that occur between the manipulation of the independent variable and the measurement of the dependent variable and that contaminate the relationship between the two. In order to be of use in scientific research, variables need to be operationally defined so that they can be measured and their effects analyzed.

Overview

Sociologists attempt to make sense out the world by observing the behavior of people within society, developing theories to explain this behavior, translating their theories into working hypotheses that can be tested, and conducting empirical research to test whether or not their theories are supported. Based on the results of the research, they then either accept or revise their theories in a continuing attempt to explain the world around them. One of the key factors in this process is the identification and operational definition of variables — traits, characteristics, or other measurable factors that can have different values — that impact the phenomenon of interest.

One is primarily interested in two types of variables: independent variables and dependent variables. The independent variable is the variable that is being manipulated by the researcher. For example, Dr. Harvey has a theory that the way that people dress affects how they are treated by others in the workplace. He believes that if people dress as if they are successful professionals (e.g., well-groomed, business attire), they will be treated that way and receive a disproportionately high percentage of raises, promotions, high performance appraisals, and other recognition. The independent variable in this theory is whether or not people dress like successful professionals. This is the variable that Dr. Harvey will manipulate in his research study to determine how it affects the way that people are treated in the workplace. The second major variable of interest to researchers is the dependent variable. The dependent variable (so called because its value depends on which level of the independent variable the subject received) is the response to the independent variable. In Dr. Harvey's research study, the dependent variable is the way that people are treated in the workplace. Dr. Harvey's theory is that the value of this variable (i.e., whether or not people receive recognition in the workplace) is dependent on how they dress.

Concepts such as "dressing as if one is a successful professional" and "how one is treated in the workplace," however, are rather nebulous and open to different interpretations. To one person, "professional attire" may be a power suit with white shirt and tie while to another it may be a clean polo shirt with Bermuda shorts rather than cutoffs. Therefore, to be of use to researchers, variables need to be operationally defined in such a way that they can be tested and statistically analyzed. An operational definition is a definition that is stated in terms that can be observed and measured. To turn his question into a hypothesis, Dr. Harvey needs to operationally define both the independent and dependent variables. For example, he may decide that "dressing professionally" means that the person wears a dark suit with a white shirt and tie for men and a dark suit with white blouse and pearls for women. Of course, this is not the only definition of "professional dress" possible. Business casual, blazers and slacks, or any number of other possibilities is also possible. However, since it is typically impossible to consider the entire range of possibilities in one research study, Dr. Harvey will have to restrict his study to include only those values of the independent variable that are of most interest. Similarly, Dr. Harvey will have to operationally define what it means not to dress professionally in the workplace (e.g., jeans and a t-shirt). These definitions, of course, restrict Dr. Harvey's hypothesis. His results will not really answer the question about "professional" versus "not professional" attire, but only about the difference in treatment that people wearing power suits receive from those who wear casual attire.

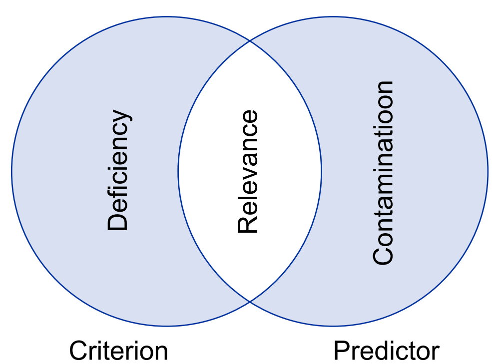
Based on this discussion, it would seem that Dr. Harvey would be well off to pick multiple operational definitions for the independent variable. Although in some ways this is true, operationally defining a variable can be a tricky proposition. The goal of operationally defining variables is not just so that they can be tested in research, but to adequately and accurately define them so that they completely represent the underlying concept as much as possible. For example, as discussed above, the concept of "dressing professionally" means different things to different people. These differences affect not only the persons who need to decide how to dress for success, but also the persons who judge them based on the clothing choices. For example, if one's boss is "old-fashioned" and dresses in a suit and tie, dressing in a suit and tie would be more likely to impress this person even if the standard for "business attire" for that company was jeans and a polo shirt.

In addition, Dr. Harvey will have to operationally define what he means by "how one is treated in the workplace." Operational definitions of this dependent variable could include the supervisor's performance appraisal ratings of the individual, the average time it takes before the person receives a raise or bonus, or whatever other factors Dr. Harvey thinks are indicative of success. Some statistical techniques allow researchers to design experiments where to test multiple conditions of both the independent and dependent variables (e.g., power suit, blazer and slacks, business casual, and casual clothing). However, given the infinite variety of human nature and behavior, it is unlikely that he will be able to include every possible condition in his operational definitions.

Operationally defining dependent variables in human research can be a complicated process. For example, the construct underlying the variable "success in the workplace" is a nebulous and complex concept. Unless one is willing to wait for the end of the subject's career and look back to determine the value of the ultimate criterion of how successful that person was in the end, one can only estimate the ultimate criterion of success in the workplace using one or more predictor measures that one can operationally define. The underlying criterion is a dependent or predicted measure that is used to judge the effectiveness of persons, organizations, treatments, or predictors. However, one does not truly know whether or not a person is successful until s/he retires and can look back on the entire career. Practically, however, this is typically not possible in social science research. Rather than choosing an ultimate criterion of success such as success at the point of retirement, it is typically necessary instead to pick an intermediate criterion of success such as how many promotions one receives within a given period of time, how many (or how large) the raises are that the person receives during that same time period, or the performance appraisal ratings the person receives from his or her supervisor.

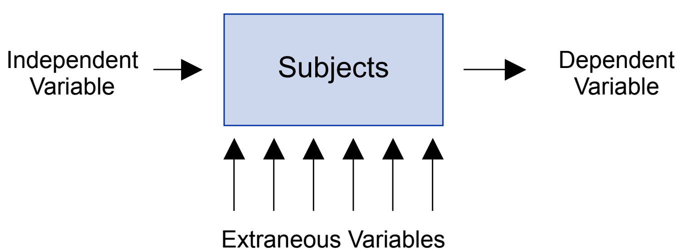
When operationally defining predictors to estimate the underlying criterion (in this case, success in the workplace), one strives to define measures that will collect data on as much of the underlying criterion as possible without measuring other extraneous variables that are not related to the criterion of "success." As shown in Figure 1, a condition known as criterion deficiency occurs when the predictor measures that are used as operational definitions of the criterion do not adequate define it. When this happens, the variable as operationally defined is not completely measuring the underlying criterion or hypothesis independent variable, and the research results will less than perfectly reflect the real relationship between the independent and dependent variables. Similarly, to the extent that the predictor measure is actually measuring something other than the criterion (i.e., is contaminated), the results will be imperfect reflections of the actual relationship between the variables. One way to help minimize the problem of criterion deficiency is to use multiple predictor measures, each of which measures a different aspect of the underlying criterion (e.g., use supervisor ratings, number of promotions, and amount of raises rather than just one of these measures). However, when doing this, one runs the risk not only of further contamination, but also of achieving spuriously high results because the predictors are related to each other (e.g., supervisor ratings are typically related to both promotions and raises).

Figure 1: Relationship between a Criterion & Predictor



Independent and dependent variables are not the only types of variables about which researchers need to be concerned, however. As shown in Figure 2, a third type of variable called extraneous variables can also unintentionally affect the outcome of a research study. These are variables that affect the outcome of the experiment (i.e., how a person is treated at work) that have nothing to do with the independent variable itself. For example, whether or not the person has visible tattoos or piercings may affect how s/he is perceived by others in the workplace no matter how s/he dresses. Similarly, the type of organization in which the experiment is conducted may affect the outcome of the study: a high-end consulting firm may have different expectations for "professional attire" than does a start-up software company. Other possible extraneous variables might include the expectations of the person giving the rating, how that person dresses, or any number of other factors that are not directly related to the relationship between the independent and dependent variable. As much as possible, of course, these extraneous variables need to be controlled. For example, one could restrict the hypothesis to only deal with consulting firms or business executives. However, the more a hypothesis is restricted, the less it reflects the real world. In addition, no matter how many extraneous variables are taken into account in an experimental design, it is virtually impossible to control literally every possible extraneous variable that may affect the outcome of a study. However, the more of these that are accounted for and controlled in the experimental design, the more meaningful the results will be.

Figure 2: Research Variables



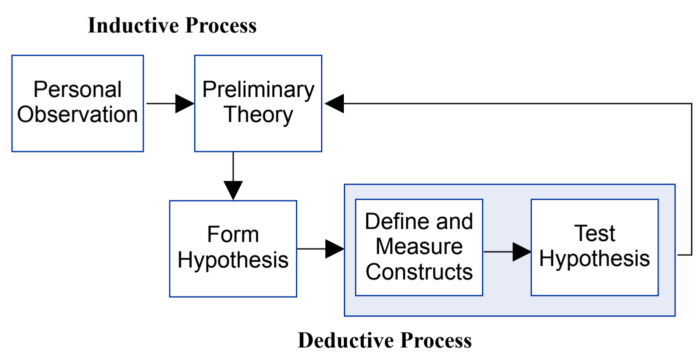
There is a fourth type of variable that may affect the outcome of the research. Intervening variables are variables that occur between the manipulation of the independent variable (e.g., how one dresses at work) and the measurement of the dependent variable (e.g., how long it takes to receive a promotion). For example, if during the time intervening between the change in the person's dressing habits and the time that the person's success is rated s/he receives further training, the resultant rating of the person's professionalism may be related to the training rather than to the way that the person dresses. Like extraneous variables, intervening variables need to be controlled as much as possible so that the effect of manipulation of the independent variable on the dependent variable can be determined.

Applications

The articulation and operational definition of variables is typically not done in isolation, but as part of the process of theory development and hypothesis testing through empirical research. Designing a good research study depends in part on two factors. First, one must try to control the research situation so that the variables measure only what they are supposed to measure. Second, one must try to include as many of the relevant factors as possible so that the research fairly emulates the real world experience. Both of these aspects require the development of good operational definitions for the variables in the study.

As shown in Figure 3, research design starts with a theory based on real world observation. For example, from personal experience Dr. Harvey may know that he is taken more seriously in professional situations when he dresses in a suit and tie. From this observation, he may develop a preliminary theory that if someone who wears "professional attire" is more likely to be perceived by others as being a competent professional than if one does not wear professional attire. Based on his observations and this preliminary theory, he next forms an empirically-testable hypothesis (e.g., "People who wear professional attire are more likely to be successful in the workplace"). To find out if this hypothesis is true, Dr. Harvey next operationally defines the various terms (i.e., constructs) in the hypothesis. As discussed above, he needs to determine how he is going to measure both professional attire and success in the workplace. To do this, he might conduct a study using research confederates who wear specific clothing that he has chosen for them. He might also define success using not only readily available measures such as raises, promotions, or annual performance evaluations, but also might develop a series of rating scales that measure the various components of success in the workplace. He would then run the experiment, using confederates dressed in different ways, collect the measures of the dependent variable, statistically analyze the resulting data using inferential statistics, and — based on the statistical significance of the answer — determine whether or not his hypothesis was correct.

Figure 3: The Theory Building Process



Despite the problems with operationally defining the various measures associated with the dependent and independent variables, however, conducting such an experiment in a laboratory setting is a relatively easy task. However, in behavioral research in general and in sociology research specifically, the phenomena are sometimes too big to be controlled in a laboratory setting or the mere fact that the research is conducted in a laboratory changes the results. Because of this fact, many sociology research studies are not conducted in laboratory settings. One approach to research that overcomes some of these limitations is the use of a simulation that approximates the real world setting. Simulations allow the researcher to bring in more real world variables but still control many of the extraneous variables. For example, a laboratory experiment about the relationship between business attire and perceptions of professionalism could be done by having people sit in an empty room and rate pictures of people who are dressed in different ways. Although this would yield some interesting data, it has little to do with the way the supervisors, customers, and employees interact and judge each other in the real world. A possible simulation would be to set up a workplace-like setting and have experimental subjects try to lead people (e.g., teach them a task) and rate how well they did. Another approach would be to conduct a field experiment in which real supervisors would rate the professionalism of real employees in the workplace who — at the behest of the researcher — wear specific types of attire. Although this approach has the advantage of being more realistic than laboratory research or simulations, it has the concomitant disadvantage of giving the researcher less control over extraneous variables that may taint the results of the study.

Sometimes, of course, the researcher does not even have enough control over the situation to manipulate the variables at all. For example, letting real employees in the workplace know that one is interested in the effects of different kinds of clothes on the perceptions of others might be enough to taint the experiment. A field study could be used in such a situation. This is an examination of how people behave in the real world. In a field study, the experimenter would just look at the way that people dress and collect data on the operationally defined dependent variables. Frequently, this approach is combined with another research technique called survey research in which subjects are interviewed by a member of the research team or asked to fill out a questionnaire regarding their preferences, reactions, habits, or other questions of interest to the researcher. This could be used to gather information about various extraneous or intervening variables that might taint the results. Unfortunately, although a very thorough interview or survey instrument can be written that would hypothetically gather all the data needed for the researcher to make decisions about the impact of work attire, such instruments are often more lengthy than the potential research subject's attention span. Further, as opposed to the other research techniques, surveys and interviews are not based on empirical data. Therefore, there is no way to know whether or not the subject is telling the truth.

Although the underlying theory is the same in all these research paradigms, the operational definitions of the independent and dependent variables may change depending on the degree of control that the researcher has over the situation. In addition, the possibility of contamination of the research results by extraneous variables becomes greater the less control the researcher has over the experimental situation.

Conclusion

Sociologists attempt to make sense out the world by applying the scientific method to their theories about the way that people act in society. One of the key factors in this process is the identification and operational definition of various variables that account for the observed phenomenon. The two primary variables of interest are the independent variable, which is manipulated, and the dependent variable whose value changes depending on the value of the independent variable. These concepts must be operationally defined in such a way that they can be tested and statistically analyzed. In addition, extraneous variables can unintentionally affect the outcome of a research study while having nothing to do with the independent variable itself.

Intervening variables are variables that occur between the manipulation of the independent variable and the measurement of the dependent variable. They can contaminate the relationship between the independent and depending variables. The articulation and operational definition of variables is part of the process of theory development and hypothesis testing through empirical research and is essential for the conduct of research that yields meaningful results.

Terms & Concepts

Confederate: A person who assists a researcher by pretending to be part of the experimental situation while actually only playing a rehearsed part meant to stimulate a response from the research subject.

Criterion: A dependent or predicted measure that is used to judge the effectiveness of persons, organizations, treatments, or predictors.

Data: (sing. datum) In statistics, data are quantifiable observations or measurements that are used as the basis of scientific research.

Dependent Variable: The outcome variable or resulting behavior that changes depending on whether the subject receives the control or experimental condition(e.g., how long it takes to receive a promotion) .

Empirical: Theories or evidence that are derived from or based on observation or experiment.

Extraneous Variable: A variable that affects the outcome of the experiment that has nothing to do with the independent variable itself.

Hypothesis: An empirically-testable declaration that certain variables and their corresponding measures are related in a specific way proposed by a theory.

Independent Variable: The variable in an experiment or research study that is intentionally manipulated in order to determine its effect on the dependent variable (e.g., how one dresses at work) .

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data. Inferential statistics are used to make inferences such as drawing conclusions about a population from a sample and in decision making.

Intervening Variable: A variable that occurs between the manipulation of the independent variable and the measurement of the dependent variable).

Operational Definition: A definition that is stated in terms that can be observed and measured.

Statistical Significance: The degree to which an observed outcome is unlikely to have occurred due to chance.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables (response). Extraneous variables are variables that affect the response but that are not related to the question under investigation in the study.

Bibliography

Black, K. (2006). *Business statistics for contemporary decision making* (4th ed.). New York: John Wiley & Sons.

Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology, 38*(1), 37-72. Retrieved November 6, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=77755951

Karlson, K., Holm, A., & Breen, R. (2012). Comparing regression coefficients between same-sample nested models using logit and probit: A new method. *Sociological Methodology, 42*(1), 286-313. Retrieved November 6, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=83576838

Magill, M. (2011). Moderators and mediators in social work research: Toward a more ecologically valid evidence base for practice. *Journal Of Social Work*, 11(4), 387-401. Retrieved November 6, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=66698630

Witte, R. S. (1980). *Statistics*. New York: Holt, Rinehart and Winston.

Suggested Reading

Anderson, M. L. & Taylor, H. F. (2002). *Sociology: Understanding a diverse society* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Kalmijn, M., & Liefbroer, A. C. (2011). Nonresponse of secondary respondents in multi-actor surveys: Determinants, consequences, and possible remedies. *Journal of Family Issues, 32*(6), 735-766. Retrieved November 6, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=60221243

Schaefer, R. T. (2002). *Sociology: A brief introduction* (4th ed.). Boston: McGraw -Hill.

Stockard, J. (2000). *Sociology: Discovering society* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Tilden, T., Hoffart, A., Sexton, H., Finset, A., & Gude, T. (2011). The role of specific and common process variables in residential couple therapy. *Journal Of Couple & Relationship Therapy, 10*(3), 262-278. Retrieved November 6, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=62823250

Essay by Ruth A. Wienclaw, PhD

Ruth A. Wienclaw holds a doctorate in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human/systems integration.

Confidence Intervals

Table of Contents

Abstract

A confidence interval is a statistical tool that estimates the range of values with a given probability of including the unknown, true value of a population parameter (e.g., mean, variance, correlation coefficient). Although hypothesis testing tends to be more frequently used in behavioral and social science research, in many ways, confidence intervals reveal more information about the underlying population. Confidence intervals approximate how much uncertainty is associated with the researcher's estimate of the underlying parameter. They also enable researchers to better understand how much confidence can be placed in the observed results of a quantitative research study.

Overview

According to the old adage, nothing in this life is sure except death and taxes. We see supporting evidence for this statement all around us. One can stare out the window at a dismal rain while listening to the weather report predicting sun all day. Although it may be sunny somewhere in the area, it certainly is not outside the window. Similarly, one can read the paper predicting the victory of one political candidate at the polls only to read the next day that the opposition candidate has won. The candidate may have won in some districts but lost overall. Such phenomena can also be found in behavioral and social science research. One researcher will triumphantly find support that a theory is correct. However, when another researcher tries to replicate the study, no such support is found. In research, such phenomena can be due to a number of reasons, including the complexity of human behavior, the inadequacy of the theory, and the nature of probability and inferential statistics.

Building a theory that realistically models the real world can be a difficult task. As human beings, we are constantly flooded with data from the world around us. Some of this is irrelevant to the task at hand: I do not care at this moment that the birds are singing outside my window or that there is a plane flying over head. Other of these data are important: I need to keep track of the words that my computer transcribes in order to make sure that the voice recognition software has correctly captured what I have said and that what I have said adequately expresses what I am trying to articulate. Some of these data are only important in the background, not needed now but potentially needed later: the heat of the halogen lamp at the back of my desk is unimportant unless a flammable piece of paper (or my hand) strays too near it. In order to be able to function, we need to prioritize the data. I shut out the sounds of the outside world, concentrate on the task before me, and remain just aware enough of my surroundings that I do not accidentally hurt myself.

The same is true for data concerning human behavior. Behavior is very complex, and it can be difficult to determine which pieces of information are important when building a theory and which are not. Every time we interact with someone, either in person or through communication media, we learn another piece of information or reinforce something we already know. We tend to try to move from a position of uncertainty to one of certainty. In general, knowing "truth" is not only comforting, allowing us to feel more in control of our surroundings, but it can also help us make decisions and plan for the future. However, life does not work that way. If we are open-minded, we find that there is another piece of data that challenges our assumptions and makes us rethink our theories. Similarly, statistics do not work that way, either. Statistically significant results in a research study do not "prove" anything. Rather, statistics point us with various degrees of confidence (or lack thereof) to the conclusion that one interpretation of the results is more likely than the other. Statistics do not yield black-and-white answers; they give best guesses or scientific estimates.

Null Hypothesis

In the behavioral and social sciences, quantitative research data are most frequently analyzed using inferential statistical tools. Most of the commonly used inferential statistical tools are used to test the probability of the null hypothesis (H0) being true. A null hypothesis is the statement that there is no statistical difference between the status quo and the experimental condition. If the null hypothesis is true, then the treatment or characteristic being studied makes no difference to the end result. For example, a null hypothesis might state that peer pressure has no effect on adolescents' decisions to use drugs. The alternative hypothesis (H1), on the other hand, would state that there *is* a correlation between peer pressure and adolescents' drug use decisions. If the researcher accepts the null hypothesis, he or she is saying that if the data in the population are normally distributed, the results of the experiment are more than likely due to chance. By accepting the null hypothesis, the researcher concludes that peer pressure has no impact on whether or not adolescents use drugs. In order for the null hypothesis to be rejected and the alternative hypothesis to be accepted, there must be a statistical significance that the difference observed between the drug-use behavior of adolescents who experienced peer pressure to use drugs and those who did not is probably due not to chance but to the real, underlying influence of peer pressure on their decision. The statistical significance is the degree to which an observed outcome is unlikely to have occurred due to chance rather than some underlying factor.

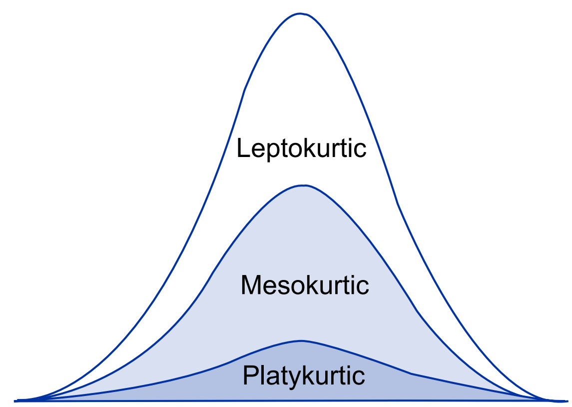
Although the ability to accept or reject a null hypothesis gives the researcher some information about the parameters of the underlying distribution, the amount of information gained is limited.

Confidence Interval

In addition to statistical tests for hypothesis testing, there is another approach to determining the statistical significance of one's research data. A confidence interval is an estimated range of values that has a given probability of including the unknown, true value of a given population parameter, such as the mean, the variance, or the correlation coefficient. This probability is called the confidence level and is expressed as a percentage, often 95 percent, meaning that if several samples are collected from the population, the unknown, true value being sought will fall within the confidence intervals of 95 percent of the samples.

The width of the confidence interval indicates the degree of uncertainty about the parameter. The narrower the interval, the more certain the researcher can be that the estimate is valid. A wide confidence interval often means that more data are needed before conclusions can be drawn about the parameter with any degree of certainty. The taller (i.e., more leptokurtic) a distribution is, the more data points are located within the confidence interval. Likewise, the wider and flatter (i.e., more platykurtic) a distribution is, the fewer data points are located within the confidence interval (see Figure 1). When the confidence interval is larger, any deviations in the research data are less likely to be significant. A narrow confidence interval means the researcher can have a high degree of confidence in the data's statistical significance.

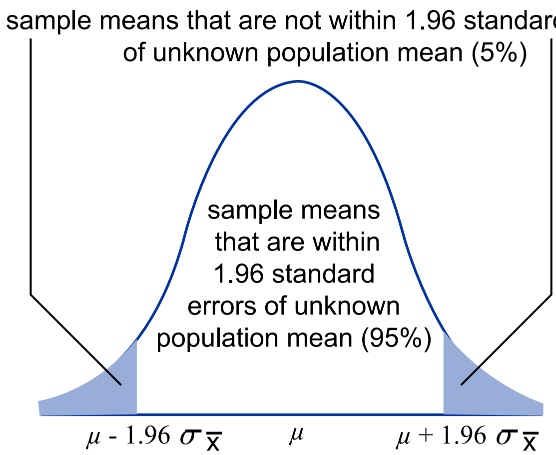
Figure 1: Comparison of Three Levels of Kurtosis ("Peakedness") of a Probability Distribution



There are three factors used in the calculation of a confidence interval. The first of these is the obtained value of the statistic (e.g., mean, variance, correlation coefficient) of the sample. In research, one assumes that this obtained value is a good estimate of the same underlying value for the wider population, called a parameter. Confidence intervals allow the researcher to better understand how much confidence can be placed in this assumption. The second element of a confidence interval is the standard error of the measure. This can be defined as the standard deviation of the distribution of the means of all the samples. The third element is the desired confidence level. Typically, confidence intervals are calculated so that the confidence level is 95 percent, but other confidence intervals can also be calculated. A confidence interval is attached to upper and lower boundaries (values) called confidence limits.

It is important to note that a 95 percent confidence interval is not the same as saying that there is a 95 percent probability that the interval contains the population parameter. The interval either contains the parameter or it does not. The 95 percent is a statement that if a large number of samples are collected from the same population, 95 percent of these samples will contain the true parameter within their confidence intervals. Figure 2 shows how a 95 percent confidence interval for sample means relates to a normal distribution.

Figure 2: Ninety-Five Percent Confidence Interval



(adapted from Witte, p. 180)

Although hypothesis testing tends to be more frequently used than confidence intervals in behavioral and social science research, in many ways, confidence intervals reveal more information about the underlying population. Rather than just indicating whether or not the null hypothesis should be rejected, confidence intervals indicate a range of plausible values for the population parameter. They are often a helpful statistic to calculate because of the nature of sampling. In sampling, a sample is selected from a larger population so that research can be done with a manageable group and extrapolated to the larger population. Typically, this is done randomly to help ensure that the sample is representative of the underlying population. However, sampling error—an error that occurs in statistical analysis when the sample does not represent the population—can occur, causing the results to not be truly representative of the underlying sample. Alternatively, the sample may not be representative simply due to the nature of probability and the luck of the draw when choosing the sample. This means that results from the sample will not be generalizable back to the population, and estimates of the various attributes of the population will vary from sample to sample. Confidence intervals can help researchers better understand a sampled data set by giving a lower and upper limit for the mean or other population parameter rather than a single estimate for the parameter. The confidence interval indicates the amount of uncertainty in the estimate of the true parameter for the population.

Applications

Distefan, Pierce, and Gilpin (2004) used confidence intervals to investigate the question of whether or not favorite movie stars influence adolescent smoking behavior. Previous observational research has suggested that the placement of smoking within movies might influence adolescents to start smoking. The portrayal of adolescents' favorite stars smoking in movies is a standard marketing gambit to advertise smoking products and promote smoking to viewers. In particular, advertising literature suggests that the placement of products within movies is effective if the viewer associates the brand image with the character and how the brand is used by that character. Further, the literature suggests that the optimal placement of products is in scenes in which the brand is used by the star of the film. This applies not only to smoking products but to a wide range of items, including candy, automobiles, and beverages. To investigate this phenomenon, the authors conducted a longitudinal study between 1996 and 1999. A representative sample of 3,104 California Adolescents, initially aged between 12 and 15 years, was used in the study. At the beginning of the study, non-smoker subjects were asked to choose two favorite male and female movie stars. The most popular movies featuring the stars from the previous three years were then reviewed to determine whether or not the star had smoked on-screen. In addition, data concerning the adolescents' perception of their parents' disapproval of smoking was also collected. The smoking status of the adolescents was reassessed three years later in follow-up interviews with the 2,084 subjects who could be located.

Among other analyses to provide population estimates of behaviors and attitudes, the researchers computed estimated variances and 95 percent confidence intervals for the data. Other statistical analyses were performed to evaluate demographic differences and identify independent predictors of smoking. The researchers also analyzed interactions between gender, receptivity to tobacco advertising and promotion, and the amount of smoking done by the favorite star. Also investigated were the interactions between receptivity and the amount of smoking by the favorite star as well as the interactions between independent variables and age and gender.

The confidence intervals performed indicated that susceptibility to smoking had a significant effect on future smoking. Further, confidence intervals indicated that those subjects whose favorite movie stars smoked on-screen were significantly more likely to have started smoking themselves during the three years between the initial interview and the follow-up interview. In addition, girls whose favorite movie stars smoked on-screen tended to be twice as likely to start smoking than others in the study.

Another example of the use of confidence intervals for analyzing social science research data is the study performed by Lipsky and Caetano (2007). The researchers examined the relationship between women who were victims of violence by an intimate partner and their use of emergency departments. Previous research found that violence perpetrated by intimate partners against women was associated with an overall increase in health-care utilization and with non-primary care services in particular. Statistics show that the most common cause of women's emergency-room visits is independent partner violence. The researchers "sought to discern whether race [and] ethnicity moderates this relationship and to explore these relationships in race [and] ethnic-specific models" (Lipsky & Caetano, 2007, par. 1).

A representative sample of 7,924 non-institutionalized civilians was used in the study. The analysis included non-Hispanic black, non-Hispanic white, and Hispanic married or cohabitating female respondents between 18 and 49 years of age. Subjects were asked how many times they had visited the emergency department for any reason in the previous 12 months. The exposure measure was defined as any intimate partner victimization during the previous 12 months. Married or cohabitating women were asked follow-up questions regarding the number of times their intimate partner had hit or threatened to hit them within the past 12 months. The researchers also collected information on substance abuse and sociodemographic factors. Specifically, measures of substance abuse included level of alcohol abuse as well as illicit drug use. Sociodemographic factors included race and ethnicity, age group, marital status, education level, employment status, household income, health insurance, number of children, and household density.

Bivariate and multivariate analyses were performed to determine the relationship between emergency department utilization and victimization by intimate partners. The study also computed 95 percent confidence intervals. Results of the analysis indicated that women who had been victimized by their intimate partners in the previous 12 months were twice as likely as non-victims to utilize an emergency department. Once sociodemographic and substance abuse factors were taken into account, women who reported intimate partner violence were one and a half times more likely than non-victims to utilize an emergency department. The results of the analysis also showed an interaction between intimate partner violence and race and ethnicity. Hispanics were found to use the emergency department three times as often as the black or white respondents. Acculturation was not found to be a significant or compounding factor in these relationships. Based on the findings of the study, the researchers suggested that Hispanic women would benefit from intimate partner violence screening in emergency departments. Further, they concluded that health-care providers and social-services personnel need culturally sensitive and specific responses for intimate partner violence. The researchers also suggested the need for further research to determine the antecedents of these findings.

Conclusion

Although quantitative research data are most frequently analyzed by hypothesis testing in the behavioral and social sciences, this approach to understanding the results of a research study does not necessarily tell the researcher how likely the observed results are to be representative of the underlying population. A confidence interval is an estimated range of values that has a given probability of including the unknown, true value of the parameter for the population. Confidence intervals approximate how much uncertainty is associated with the researcher's estimate of the underlying parameter. In many ways, confidence intervals reveal more than hypothesis testing does about the underlying population.

Terms & Concepts

Confidence Interval: An estimated range of values that has a given probability of including the unknown, true value of a particular parameter of the population.

Data: In statistics, quantifiable observations or measurements that are used as the basis of scientific research.

Distribution: A set of numbers collected from the data and their associated frequencies.

Experiment: A situation under the control of a researcher in which an experimental condition (independent variable) is manipulated and the effect on the experimental subject (dependent variable) is measured. Most experiments are designed using the principles of the scientific method and are statistically analyzed to determined whether or not the results are statistically significant.

Hypothesis: An empirically testable declaration that certain variables and their corresponding measures are related in a specific way proposed by a theory. A null hypothesis is the statement that the findings of the experiment will show no statistical difference between the control condition and the experimental condition.

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data, as well as in decision making.

Mean: An arithmetically derived measure of central tendency in which the sum of the values of all the data points is divided by the number of data points.

Normal Distribution: A continuous distribution that is symmetrical about its mean and asymptotic to the horizontal axis. The area under the normal distribution is 1. The normal distribution, also called a Gaussian distribution or a normal curve, is actually a family of curves and describes many characteristics observable in the natural world.

Population: The entire group of subjects belonging to a certain category (e.g., all women between the ages of 18 and 27, all dry-cleaning businesses, all college students).

Probability: A branch of mathematics that deals with estimating the likelihood of an event occurring. Probability is expressed as a value between 0 and 1.0, which is the mathematical expression of the number of actual occurrences to the number of possible occurrences of the event. A probability of 0 signifies that there is no chance that the event will occur, and 1.0 signifies that the event is certain to occur.

Sample: A subset of a population. A random sample is a sample that is chosen at random from the larger population with the assumption that such samples tend to reflect the characteristics of the larger population.

Sampling: A group of techniques that are used to select a sample from a larger population so that research can be done with a manageable group and extrapolated to the larger population.

Standard Deviation: A measure of variability that describes how far the typical score in a distribution is from the mean of the distribution. The larger the standard deviation, the farther away it is from the midpoint of the distribution.

Statistical Significance: The degree to which an observed outcome is unlikely to have occurred due to chance.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables. Extraneous variables are variables that affect the outcome but are not related to the question under investigation in the study.

Bibliography

Ching-hong Li, J., Ying, C., & Wai, C. (2013). Bootstrap confidence intervals for the mean correlation corrected for case IV range restriction: A more adequate procedure for meta-analysis. *Journal of Applied Psychology, 98*(1), 183–193. Retrieved November 5, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=84917385&site=ehost-live

Distefan, J. M., Pierce, J. P., & Gilpin, E. A. (2004, Jul). Do favorite movie stars influence adolescent smoking initiation? *American Journal of Public Health, 94* (7), 1239-1244. Retrieved April 28, 2008 from EBSCO Online Database Academic Search Complete: http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=13672929&site=ehost-live

Kuijpers, R. E., Van der Ark, L., & Croon, M. A. (2013). Standard errors and confidence intervals for scalability coefficients in Mokken scale analysis using marginal models. *Sociological Methodology, 43*(1), 42–69. Retrieved November 5, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=90132015&site=ehost-live

Lipsky, S. & Caetano, P. (2007, Dec). The role of race/ethnicity in the relationship between emergency department use and intimate partner violence: Findings from the 2002 National Survey on Drug Use and Health. *American Journal of Public Health, 97* (12), 2246-2252. Retrieved April 28, 2008 from EBSCO Online Database SocINDEX with Full Text: http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=28073644&site=ehost-live

Rueda, M. M., Arcos, A. A., Sánchez-Borrego, I. I., & Muñoz, J. J. (2012). An approximation method to derive confidence intervals for quantiles with some applications. *Quality & Quantity, 46*(4), 1197–1208.

Witte, R. S. (1980). *Statistics* . New York: Holt, Rinehart and Winston.

Suggested Reading

Cutrona, S. L., Woolhandler, S., Lasser, K. E., Bor, D. H., McCormick, D., & Himmelstein, D. U. (2008, Feb). Characteristics of recipients of free prescription drug samples: A nationally representative analysis. *American Journal of Public Health, 98* (2), 284-289. Retrieved April 28, 2008 from EBSCO Online Database SocINDEX with Full Text: http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=30103025&site=ehost-live

Karter, A. J., et al. (2008, Feb). Educational disparities in rates of smoking among diabetic adults: The Translating Research Into Action For Diabetes Study. *American Journal of Public Health, 98* (2), 265-370. Retrieved April 28, 2008 from EBSCO Online Database SocINDEX with Full Text: http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=30103037&site=ehost-live

Khalifeh, H., Hargreaves, J., Howard, L. M., & Birdthistle, I. (2013). Intimate partner violence and socioeconomic deprivation in England: Findings from a national cross-sectional survey. *American Journal of Public Health, 103*(3), 462–472. Retrieved November 5, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=85384460&site=ehost-live

Maddigan, S. L., Feeny, D. H., Majumdar, s. R., Farris, K. B., & Johnson, J. A. (2006, Sep). Understanding the determinants of health for people with Type 2 diabetes. *American Journal of Public Health, 96* (9), 1649-1655. Retrieved April 28, 2008 from EBSCO Online Database SocINDEX with Full Text: http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=22304268&site=ehost-live

Russell, M., Welte, J. W., & Barnes, G. M. (1991, Apr). Quantity-frequency measures of alcohol consumption: Beverage-specific vs. global questions. *British Journal of Addiction, 86* (4), 409-417. Retrieved April 28, 2008 from EBSCO Online Database SocINDEX with Full Text: http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=6625211&site=ehost-live

Strawbridge, W. J., Cohen, R. D., Shema, s. J., & Kaplan, G. A. (1997, Jun). Frequent attendance at religious services and mortality over 28 years. *American Journal of Public Health, 87* (6), 957-961. Retrieved April 28, 2008 from EBSCO Online Database SocINDEX with Full Text: http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=9708191858&site=ehost-live

Essay by Ruth A. Wienclaw, Ph.D.

Ruth A. Wienclaw holds a doctorate in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human-systems integration.

Correlation

Table of Contents

Abstract

In statistics, correlation is the degree to which two events or variables are consistently related. This measure indicates both the degree and direction of the relationship between variables. However, it yields no information concerning the cause of the relationship. Correlation techniques are available for both parametric and nonparametric data. The Pearson Product Moment Correlation is also used in other inferential statistical techniques such as regression analysis and factor analysis to help researchers and theorists build models that reflect the complex relationships observed in the real world.

Overview

Every day we make assumptions about the relationship of one event to another in both our personal and professional lives. "My alarm clock failed to go off this morning, so I will be late for work." "The cat ate an entire can of cat food so she must be feeling better." "I received a polite e-mail from Mr. Jones, so he must not be angry that my report was not submitted on time." Sociologists attempt to express the relationship between variables in the same way on a broader scale. "Advertisements induce previous purchasers to buy additional lottery tickets." "People tend to act more openly with strangers who outwardly appear to be similar to themselves." "Younger males tend to be less prejudiced towards women in the workplace."

From a statistical point of view, the mathematical expression of such relationships is called correlation. This is the degree to which two events or variables are consistently related. Correlation may be positive (i.e., as the value of one variable increases the value of the other variable increases), negative (i.e., as the value of one variable increases the value of the other variable decreases), or zero (i.e., the values of the two variables are unrelated). However, correlation does not give one any information about what caused the relationship between the two variables. Properly used, knowing the correlation between variables can give one useful information about behavior. For example, if I know that my cat gets sick when I feed her "Happy Kitty" brand cat food, I am unlikely to feed her "Happy Kitty" in the future. Of course, knowing that she gets sick after eating "Happy Kitty" does not explain why she gets sick. It may be that she is sensitive to one of the ingredients in "Happy Kitty" or it may be that "Happy Kitty" inadvertently released a batch of tainted food. However, my cat's digestive problems might not have anything to do with "Happy Kitty" at all. The neighborhood stray may eat all her "Happy Kitty" food, causing her to have eaten something else that causes her to get sick, or I changed her food to "Happy Kitty" at the same time she was sick from an unrelated cause. All I know is that when I feed her "Happy Kitty" she gets sick. Although I do not know why, this is still useful information to know. The same is true for the larger problems of sociology.

There are a number of ways to statistically determine the correlation between two variables. The most common of these is the technique referred to as the Pearson Product Moment Coefficient of Correlation, or Pearson r . This statistical technique allows researchers to determine whether the two variables are positively correlated (i.e., my cat gets sick when she eats "Happy Kitty"), negatively correlated (i.e., my cat is healthier when she eats "Happy Kitty"), or not correlated at all (i.e., there is no change in my cat's health when she eats "Happy Kitty").

Correlation vs. Causation

However, as mentioned above, knowing that two variables are correlated does not tell us whether one variable caused another or if both observations were caused by some other, unknown, third factor. As opposed to the various techniques of inferential statistics where we attempt to make inferences such as drawing conclusions about a population from a sample and in decision making by looking at the influence of an independent variable on a dependent variable, correlation does not imply causation. For example, if I have two clocks that keep perfect time in my house, I may observe that the alarm clock in my bedroom goes off every morning at seven o'clock just as the grandfather clock in the hallway chimes. This does not mean that the alarm clock caused the grandfather clock to chime or that the grandfather clock caused the alarm clock to go off. In fact, both of these events were caused by the same event: the passage of 24 hours since the last time they did this. Although it is easy to see in this simple example that a third factor must have caused both clocks to go off, the causative factor for two related variables is not always so easy to spot. To act on such unfounded assumptions about causation as inferred from correlation is part of the cycle of superstitious behavior. Many ancient peoples, for example, included some sort of sun god in their pantheon of deities. They noticed that when they made offerings to their sun god, the sun arose the next morning, bringing with it heat and light. So, they made offerings. From our modern perspective, however, we now know that the faithful practice of making offerings to a sun god was not the cause of the sun coming up the next morning. Rather, the apparent phenomenon of the rising sun is caused by the daily rotation of the earth on its access.

The classic example of showing the absurdity of inferring causation from correlation was published in the mid 20th century in a paper reporting the results of an analysis of fictional data. Neyman (1952) used an illustration of the correlation between the number of storks and the number of human births in various European countries. The result of the correlation analysis of the relationship between the sightings of storks and the number of births was both high and positive. Without understanding how to interpret the correlation coefficient, someone might conclude from this evidence that storks bring babies. The truth, however, was that the data were analyzed without respect of country size. Since larger northern European countries tend to have both more women and more storks, the observed correlation was due to country size. The correlation was incidental and not causal: correlation tells one nothing about causation. Although this example was originally meant to make people laugh, it was also meant as a warning: as absurd as these examples may sound, coefficients are frequently misinterpreted to imply causation.

Pearson Product Moment Correlation

The Pearson Product Moment Correlation is a parametric test that makes several assumptions concerning the data that are being analyzed. First, it assumes that the data have been randomly selected from a population that has a normal distribution. In addition, it assumes that the data are interval or ratio in nature. This means that not only do the rank orders of the data have meaning (e.g., a value of 6 is greater than a value of 5) but the intervals between the values also have meaning. For example, weight is a ratio scale. It is clear that the difference between 1 gram of a chemical compound and 2 grams of a chemical compound is the same as the difference between 100 grams of the compound and 101 grams of the compound. These measurements have meaning because the weight scale has a true zero (i.e., we know what it means to have 0 grams of the compound) and the intervals between values is equal. On the other hand, in attitude surveys and other data collection instruments used by sociologists, it may not be quite as clear that the difference between 0 and 1 on a 100 point rating scale of quality of a widget is the same as the difference between 50 and 51 or between 98 and 99. These are value judgments and the scale may not have a true zero. Even if the scale does start at 0, it may be difficult to define what this value means. It is difficult to know whether a score of 0 differs significantly from a score of 1 on an attitude scale. In both cases, the rater had a severe negative reaction to the item being discussed. Since ratings are subjective, even if numerical values are assigned to them, these do not necessarily meet the requirement of parametric statistics that the data be at the interval or ratio level.

Spearman Rank Correlation Coefficient

Fortunately, the Pearson product moment correlation is not the only method for determining the relationship between variable. For these situations, the Spearman Rank Correlation Coefficient can be used instead to determine the degree of relationship between two variables. The Spearman is a nonparametric statistical test that makes no assumptions about the underlying distribution of the data. Unlike the Pearson coefficient of correlation which requires interval or ratio level data, the Spearman can be used with ordinal level (i.e., ranked) data. In addition, the Spearman does not require interval data nor does it assume that there is a linear relationship between the variables. For example, the Spearman Rank Correlation could be used in a situation where one wanted to determine if the ratings of the violence level of television shows done by two different raters was close enough to be pooled (i.e., to determine whether or not both individuals were using the same subjective criteria when rating the shows).

Applications

Coefficients of Correlation & Their Use

Coefficients of correlation are used not only as stand alone statistics, but also as inputs into other statistical techniques including regression analysis and factor analysis. These techniques are used to develop multidimensional models that describe the complex nature of real world situations.

Factor Analysis

Factor analysis is a multivariate technique that is used to analyze the interrelationships between variables and attempts to articulate their common underlying factors. Factor analysis is used in situations where it is assumed that the nature of reality is not actually chaotic, but it is attributed to multiple underlying factors. Multidimensional mathematical techniques are applied to the data to examine how they cluster together into "factors." In many ways, factor analysis is more a logical procedure than a statistical one although it is based on the analysis of Pearson correlation coefficients between data. Factor analysis performs a causal analysis to determine the likelihood of the same underlying processes resulting in multiple observations. Although factor analysis can yield interesting information about the relationships between seemingly unrelated data, the determination of factors, in the end, is a qualitative decision requiring the insights of the researcher. Further, factor analysis does not determine "the" set of factors that underlie the data, but typically can reveal several likely sets of factors. This means that the researcher needs to give careful consideration to what is known about the situation in order to determine which potential set of factors is superior to the others. If such considerations are not available, however, the resulting factors will not be meaningful.

Factor Analysis & Pearson r

An example of a research study that uses the Pearson *r* as an input for model building was performed by Brennan, Molnar, and Earls (2007). The researchers used correlation and factor analysis to refine a measure of adolescents' exposure to violence in Chicago neighborhoods. Trained interviewers conducted separate interviews with each adolescent and his/her primary care giver about potentially harmful events that had occurred in the adolescent's life (both witnessed events and personal victimization). Subjects also answered 35 items concerning their anxiety/depression, aggression, and delinquency levels as well as a number of items used to collect demographic data. Among other methods, the researchers used correlation to identify items related to the measurement of exposure to violence that might not fit well with the construct of violence exposure. These were items that either correlated with other items on the scale and/or did not increase the reliability of the scale. Three scales were of particular interest: victimization, witnessing of violence, and learning of violence. To test whether or not these were truly three separate scales, the researchers performed a confirmatory factor analysis. Based on the results of the study, the researchers concluded that these were three different factors contributing to the exposure to violence in urban youth.

Regression Analysis & Pearson r

Another statistical technique that uses the Pearson *r* as an input is regression analysis. This is a family of statistical techniques used to develop a mathematical model for use in predicting one variable from the knowledge of another variable. Advanced regression techniques allow researchers to use both multiple independent and multiple dependent variables in developing models. The regression equation is a mathematical model of a real world situation that can be invaluable for forecasting and learning more about the interaction of variables in the real world. There are many types of multivariate regression including multiple linear regression, multivariate polynomial regression, and canonical correlation.

Subrahmanyam and Lin (2007) used correlation as an input into regression analysis to investigate the effects of Internet use on the well-being of adolescents. The researchers examined the effect of Internet usage on adolescents' feelings of loneliness and their perceptions of support from friends and family. Data were collected from 78 females and 78 males between the ages of 15 and 18.4 years of age. Each participant completed an Internet access questionnaire that asked questions about how they used the Internet (e.g., total time spent online, time spent using e-mail, place of access). The questionnaire also explored the subjects' knowledge of and familiarity with their online correspondents, as well as their relationships with these individuals. The loneliness level of the subjects was measured using the eight-item *Roberts Revision of the UCLA Loneliness Scale* and the availability of others to whom the participants could turn in times of need and how satisfied they were with that support was accessed using the 24-item *Social Support Scale for Children* . The data were analyzed using regression analysis. These results suggested that loneliness was not predicted by the length of time spent on the Internet. However, gender and the participants' perceptions of the participants on their online relationships did predict loneliness. Participants who felt that their online partners could be counted on in times of need also tended to be more lonely than those who did not. Finally, perceived support from significant others was not related to the amount of time spent online, time spent on e-mail, relationships with online partners, or perceptions about these relationships.

Conclusion

Coefficients of correlation mathematically express the degree of relationship between two events or variables on a scale of 0.0 (demonstrating no relationship between the two variables) to 1.0 (demonstrating a perfect relationship between the variables). In addition, coefficients of correlation may be positive (demonstrating as the value of one variable increases so does the value of the other variable) or negative (demonstrating that as the value of one variable increases the value of the other variable decreases). Two of the most common methods of determining correlation between variables are the Pearson Product Moment Coefficient of Correlation for use with parametric data and the Spearman Rank Order Coefficient of Correlation for use with nonparametric data. In addition, the Pearson statistic can be used as an input to other statistical techniques such as regression analysis and factor analysis in the building of models of complex real world behavior. Although correlation is an important statistical tool for sociologists, it is important that correlation by itself does not imply causation: Correlated variables may be both caused by a third, unknown factor or only speciously related.

Terms & Concepts

Correlation: The degree to which two events or variables are consistently related. Correlation may be positive (i.e., as the value of one variable increases the value of the other variable increases), negative (i.e., as the value of one variable increases the value of the other variable decreases), or zero (i.e., the values of the two variables are unrelated). Correlation does not imply causation.

Data: (*sing.* datum) In statistics, data are quantifiable observations or measurements that are used as the basis of scientific research.

Demographic Data: Statistical information about a given subset of the human population such as persons living in a particular area, shopping at an area mall, or subscribing to a local newspaper. Demographic data might include such information as age, gender, or income distribution, or growth trends.

Dependent Variable: The outcome variable or resulting behavior that changes depending on whether the subject receives the control or experimental condition (e.g., a consumer's reaction to a new cereal).

Distribution: A set of numbers collected from data and their associated frequencies.

Factor Analysis: A multivariate statistical technique that analyzes interrelationships between variables and attempts to articulate their common underlying factors.

Independent Variable: The variable in an experiment or research study that is intentionally manipulated in order to determine its effect on the dependent variable (e.g., the independent variable of type of cereal might affect the dependent variable of the consumer's reaction to it).

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data. Inferential statistics are used to make inferences such as drawing conclusions about a population from a sample and in decision making.

Model: A representation of a situation, system, or subsystem. Conceptual models are mental images that describe the situation or system. Mathematical or computer models are mathematical representations of the system or situation being studied.

Nonparametric Statistics: A class of statistical procedures that is used in situations where it is not possible to estimate or test the values of the parameters (e.g., mean, standard deviation) of the distribution or where the shape of the underlying distribution is unknown.

Normal Distribution: A continuous distribution that is symmetrical about its mean and asymptotic to the horizontal axis. The area under the normal distribution is 1. The normal distribution is actually a family of curves and describes many characteristics observable in the natural world. The normal distribution is also called the Gaussian distribution or the normal curve of errors.

Parametric Statistics: A class of statistical procedures that is used in situations where it is reasonable to make certain assumptions about the underlying distribution of the data and where the values to be analyzed are either interval- or ratio-level data.

Regression Analysis: A family of statistical techniques used to develop a mathematical model for use in predicting one variable from the knowledge of another variable.

Reliability: The degree to which a psychological test or assessment instrument consistently measures what it is intended to measure. An assessment instrument cannot be valid unless it is reliable.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables (response). Extraneous variables are variables that affect the response but that are not related to the question under investigation in the study.

Bibliography

Armore, S. J. (1966). *Introduction to statistical analysis and inferences for psychology and education* . New York: John Wiley & sons.

Brennan, R. T., Molnar, B. E., & Earls, F. (2007). Refining the measurement of exposure to

violence (ETV) in urban youth. *Journal of Community Psychology, 35* (5), 603-618. Retrieved March 28, 2008, from EBSCO Online Database SocINDEX with full Text. http://web.ebscohost.com/ehost/pdf?vid=5&hid=108&sid=e7761ab9-64c2-4b80-aea4-8ef8ea7ef230%40sessionmgr102

Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate data analysis*. New York: John Wiley and Sons. ..F.- Holgado–Tello, F., Chacón–Moscoso, S., Barbero–García, I., & Vila–Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, 44(1), 153–66. Retrieved October 25, 2013 from EBSCO Online Database

SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=47161252

Holosko, M. J. (2010). What Types of Designs are We Using in Social Work Research and Evaluation?. *Research On Social Work Practice*, 20(6), 665–73. Retrieved October 25, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=54489183

Hollander, M. & Wolfe, D. A. (1973). *Nonparametric statistical methods* . New York: John Wiley and Sons.

Huff, D. (1954). *How to lie with statistics*. New York: W. W. Norton & Company.

Neyman, J. (1952). *Lectures and Conferences on Mathematical Statistics and Probability* (2nd ed.). US Department of Agriculture: Washington DC.

Segal, E. A., Cimino, A. N., Gerdes, K. E., Harmon, J. K., & Wagaman, M. (2013). A Confirmatory Factor Analysis of the Interpersonal and Social Empathy Index. *Journal of The Society For Social Work & Research*, 4(3), 131–53. Retrieved October 25, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=90515573

Subrahmanyam, K. & Lin, G. (2007). Adolescents on the net: Internet use and well-being. *Adolescence, 42* (168), 659-677. Retrieved March 18, 2008, from EBSCO Online Database SocINDEX with Full Text. http://web.ebscohost.com/ehost/pdf?vid=4&hid=7&sid=0448787f-afa0-4373-819c-88135f67c7ab%40sessionmgr7

Thurstone, L. L. (1947). *Multiple-factor analysis* . Chicago: University of Chicago Press.

Witt, R. S. (1980). *Statistics*. New York: Holt, Rinehart and Winston.

Suggested Reading

Brady, H. E. & Seawright, J. (2004) Framing social inquiry: From models of causation to statistically based causal inference. Paper prepared for the American Political Science Association Annual Meeting. Retrieved March 24, 2008, from http://www.asu.edu/clas/polisci/cqrm/APSA2004/BradySeawright.pdf

Eldeleklioglu, J. (2007). The relationships between aggressiveness, peer pressure and parental attitudes among Turkish high school students. *Social Behavior & Personality: An International Journal, 35* (7), 975-986. Retrieved March 24, 2008, from EBSCO Online Database SocINDEX with Full Text. http://web.ebscohost.com/ehost/pdf?vid=8&hid=16&sid=f1afadd7-c398-46e6-9c2c-f70804ccc820%40sessionmgr9

Fricke, T. (2003). Culture and causality: An anthropological comment. *Population & Development Review, 29* (3), 470-479. Retrieved March 24, 2008, from EBSCO Online Database SocINDEX with Full Text. http://web.ebscohost.com/ehost/pdf?vid=5&hid=16&sid=f1afadd7-c398-46e6-9c2c-f70804ccc820%40sessionmgr9

Gangl, M. (2010). Causal Inference in Sociological Research. *Annual Review Of Sociology*, 36(1), 21–47. Retrieved October 25, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=53357050

Pearl, J. (2010). The Foundations of Causal Inference. *Sociological Methodology*, 40(1), 75–149. Retrieved October 25, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=55203480

Petrovec, D., Tompa, G., & Šugman, K. (2007). Poverty and reaction to crime - irresponsibility proven. *Sociologija: Mintis ir Veiksmas, 2007/2* (20), 32-42, Retrieved March 24, 2008, from EBSCO Online Database SocINDEX with Full Text. http://web.ebscohost.com/ehost/pdf?vid=6&hid=16&sid=f1afadd7-c398-46e6-9c2c-f70804ccc820%40sessionmgr9

Essay by Ruth A. Wienclaw, Ph.D.

Ruth A. Wienclaw holds a Ph.D. in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human/systems integration.

Analysis of Secondary Data

Table of Contents

Abstract

Due to various ethical and logistical considerations, it can be impossible in some settings to gather primary data for research analysis. However, many sources of secondary data are available that can be further analyzed by researchers seeking to answer other research questions. Secondary analysis may be qualitative or quantitative in nature and may be used by itself or combined with other research data to reach conclusions. Although the use of secondary data can be more cost-effective than the use of primary data, the fact that the researcher has no control over how the data were collected means that there are several disadvantages as well. However, a well-designed meta analysis or other study that incorporates secondary data can be very useful to the researcher in answering questions about social issues and significantly aid in the advancement of the social sciences.

Overview

Acquiring data for research in the social and behavioral sciences can be a difficult process necessitating the application of great creativity. Sometimes, ethical considerations mean that it is impossible to experimentally manipulate variables. For example, when studying the detrimental effects of length of unemployment, one cannot in good conscience randomly decide which subjects will lose their jobs and which ones will not or for how long they will be without income. In other cases, the mere fact that a researcher is observing the subjects changes the way that the subjects act. For example, the Hawthorne Effect refers to a well-known study of the effects of lighting levels on assembly line employees at the Hawthorne works of Western Electric outside Chicago. Researchers found that productivity increased not only when lighting levels were increased, but also when they were decreased because of the subjects' expectations that the experimental interventions would enable them to increase productivity. In still other cases, it is simply not possible to gather the data needed for a research study for practical or logistical reasons. For example, to test the effectiveness of a new training program for aircraft maintenance personnel, one could easily design a controlled study to see whether personnel performed better after training or without training. It would be relatively simple to operationally define dependent variables for the study including number of fatal crashes. However, it is highly unlikely that any airline would be willing to risk the lives of their employees or customers to collect such data.

Fortunately, researchers are not restricted to the use of primary data (i.e., data that are collected specifically for the research study). Many types of secondary data that have been collected and analyzed for other purposes are often available for re-analysis. In secondary analysis, further analysis of existing data (typically collected by a different researcher) is conducted. The intent of secondary analysis is to use existing data in order to develop conclusions or knowledge in addition to or different from those resulting from the original analysis of the data. Secondary analysis may be qualitative or quantitative in nature and may be used by itself or combined with other research data to reach conclusions.

Sources of Secondary Data

Secondary data are available from many sources. In some cases, one must contact the researchers of previous studies and gain access to their data. In other cases, it may be possible to use public access data.

- veteran's issues, and women's issues. The Census Bureau can be accessed at www.census.gov.
- University of Minnesota's Minnesota population Center is an integrated series of census microdata samples for US and international population studies. The data are intended for use by economists and social scientists. The data date back to the 1960s and includes 80 samples from 26 countries, with more scheduled for release in the future. The IPUMS data can be accessed at www.ipumns.umn.edu. The Bureau of Labor Statistics collects and maintains data on employment, earnings, living conditions, productivity, and other factors of interest to social scientists. The portal for the Bureaus of Labor Statistics data is found at http://stats.bls.gov.
- The Inter-University Consortium for Political and Social Research (ICPSR) maintains the world's largest archive of digital social science data. The goals of the consortium are to acquire and preserve social science data, provide open and equitable access to these data, and promote their effective use. The ICPSR web site is found at www.icpsr.umich.edu.

In addition to these sources, secondary data can be obtained for analysis through a wide variety of sources including newspaper and periodicals, organizational records and archives, videotapes of motion pictures and television programs, web pages, scientific records (e.g., patent applications), speeches of public figures, votes cast in elections or by legislators, as well as personal journals, diaries, e-mail, and correspondence. Many other sources of secondary data are available depending on the needs of the researcher.

Advantages to Using Secondary Data

There are a number of advantages to using secondary data for analysis. As discussed above, there are certain situations in which it is impossible for ethical, logistical, or other practical reasons to collect primary data. The analysis of secondary data allows researchers to examine data collected for other purposes to find the answers they seek to research questions. For example, the study of the effects of unemployment could include the reanalysis of questionnaires routinely collected by government or private employment agencies. The re-analysis of previously collected survey data could also be used in some cases to answer other questions about the effects of various levels of the independent variable on the dependent variable without the presence of the researcher or other observer changing the results. Similarly, a historical study of routinely collected data might be divided into groups for aircraft that had been worked on by technicians who had received the new training *vs* . those who had not. In addition, the collection of data for secondary analysis is typically much faster because the data have already been collected. Similarly, the researcher does not have to develop a new data collection instrument or run a new experiment, other factors that both reduce the time to gather the data as well as the costs associated with data collection. A major advantage of the analysis of secondary data is that the collection of such data is non-reactive. In other words, particularly for archival data, subjects will act naturally because they do not realize that their behavior is being observed and recorded. This advantage, of course, does not extend to data collected with surveys or direct observation where subjects know that their reactions are being observed.

Disadvantages of Using Secondary Data

On the other hand, the analysis of secondary data is not without its potential disadvantages as well. Unless one has collected the data oneself, it is virtually impossible to be completely confident in the quality of the data. Although the survey instruments associated with data sets may be available, one does not necessarily know what the inter-interviewer reliability is for surveys not under one's own control or whether or not interviewer bias or other interviewer effects may have tainted the data. Further, it is not always possible to find available data sets that contain the data that one needs to analyze. Another disadvantage in the use of secondary data arises from questions concerning the way subjects were selected. In most research studies, subjects are chosen from a representative sample so that results can be extrapolated to the general population. However, just as it is not always possible to know if interviewer affects were unintentionally introduced into data collection, it is similarly impossible in many cases to know whether or not a sample selected by someone else is truly random or if it was biased. When one uses primary data and research analysis, one can be confident about the way data were collected, samples were selected, and the relevance of survey items and other measurements to the research hypothesis. However, the same cannot always be said for analyses performed in secondary data. In samples where sampling error or bias occur, any conclusions drawn from the data cannot be extrapolated to the population at large.

Considerations

There are a number of issues that must be considered before embarking on a secondary analysis. First, if using secondary data collected using a survey instrument, it must be determined whether or not the wording of the question(s) of interest on the survey are a good fit for the data being used in the current analysis. If the wording is ambiguous or otherwise questionable for use in the current study, a better source of data needs to be found. When the results of a secondary analysis are reported, it is important to also consider the experimental conditions under which the data were originally collected. These conditions may impact the usefulness of the data for the current study. It can be tempting to "make do" with the data that are available; extrapolating data to relationships or conclusions that are not warranted. However, an ethical researcher will make certain that the data used in a research analysis are appropriate to the study whether performing a primary or secondary analysis.

Another type of analysis that uses secondary data is meta analysis. This is an analysis technique used to synthesize the results of multiple existing quantitative research studies of a single phenomenon into a single result. Statistically, meta analysis combines the effect size estimates of the individual studies into a single estimated effect size or a distribution of effect sizes. Due to the probabilistic nature of inferential statistics, the statistical significance of research results is only an estimate as to whether or not the hypothesis being tested is true. Therefore, even when the results of a study show statistical significance, the hypothesis still may not be true. This is one of the reasons that the results of research studies cannot always be replicated by other researchers. Through meta analysis, a body of research results performed by different researchers can be examined to help the researcher get a better picture of the overall pattern of the results of numerous studies conducted on the same phenomenon.

Applications

Use of Secondary Data in Social Science

Depression in Mothers

Secondary data analysis is frequently used in the social sciences. For example, Horwitz *et al* (2007) performed secondary analysis on data to understand the prevalence, correlates, and persistence of depression in the mothers of young children. The authors' review of the literature on depression found that only bad depression tends to occur more frequently in women, but also adds that “depression is expected to replace cancer as the second leading cause of morbidity within the next decade” (Horwitz, 2007). This study is an eight-stratified in gender-stratified random sample of children born in the all-New Haven Hospital between July 1995 and September 1997 and who lived in the New Haven Meriden Standard Metropolitan Statistical Area. Individuals in this group were excluded from the sample if they were born prematurely, had low birth weight, were likely to have developmental delays due to birth complications, or head chromosomal anomalies. Of an original pool of 7433 subjects, 1605 are found to be eligible under these criteria, of which 1278 participated in the initial data collection. Of these, 1095 participated in the one-year follow-up.

Several measures were collected for these individuals. Birth record information was obtained from birth records provided by the State of Connecticut Department of Public Health. These data included birth weight, gestational age, 1- and 5-minute Apgar scores, parental age, maternal education, and similar measures. Information about sociodemographic variables was collected using a short written survey instrument that was answered by the mothers. Mothers were also asked to report the difficulty in financial strain on a five-point scale ranging from "easy" to "difficult." In addition, the mothers were asked to rate their child's current physical health on a five-point scale. Mothers were also asked to respond to several standard questionnaires: The Beck Anxiety Inventory, the expressiveness and conflict scales from the Family Environment Scale, an adaptation of the of the Life Events Inventory, the social support items of the Medical Outcomes Study parent questionnaire, the short form Parenting Stress Index, and the Quality of Marriage Inventory. Maternal depression was measured using the Center for Epidemiologic Studies Depression Scale.

The variables were placed into groups as follows: Maternal social demographic characteristics, maternal mental and physical health characteristics, maternal support and stretched measures, child characteristics, and spouse/partner characteristics. Variables in each domain were statistically evaluated. Variables that did not have an effect on the outcome were investigated. Data analysis indicated that elevated self-reported symptoms of depression were related to various factors. These included “younger maternal age, lower maternal education, unemployment, minority race, maternal physical health status, single parenting, poverty, difficulty paying bills, high anxiety, high family conflict, family expressiveness, high parenting stress, social support, and high parental life events” (Horwitz, 2007). However, the analysis showed no significant relationship between elevated levels of depressive symptoms and birth of a child within the previous year.

The results of the study suggested that elevated depressive symptoms in mothers of 11- to 42-month-old children are prevalent. The analysis also a relationship between elevated symptoms of depression and associated characteristics for the other sample points (i.e., anxiety, high parent distress, poor physical health, financial strain, high life events, low social support, having younger children). The findings also suggested that women with co-occurring anxiety to live and conflict laden environments ever greater tendency to continue to report elevated symptoms of depression. This result supports the previously identified importance of anxiety as a predictor of non-repression of depression symptoms reported in the literature.

Self-Attribution in Victims

In another example of the use of secondary data in the social sciences, Littleton, Magee, and Axsom (2007) performed a meta analysis of self-attribution following three types of trauma: sexual victimization, illness, and injury. Self-attributions can be defined as the victim accepting responsibility or blame for the event of which s/he was a victim. Although the literature discusses theoretical models concerning the causes of self-attribution, the authors found little empirical research that investigated why self-attributions occur. Their investigation had four goals. First, they desired to determine the prevalence of self-attribution following a traumatic event. Second, they examined the effect of variable research methodology on reports of self-attribution. Third, they attempted to identify predictors (including individual differences and trauma variables) of self-attributions. Fourth, they desired to determine whether behavior and character are distinct.

To answer these research questions, the authors conducted a meta analysis of all existing studies of self-attributions which followed three types of trauma (i.e., sexual, illness, injury). The reason these factors were chosen for this study was because self-attributions are most frequently studied following these types of traumas. Therefore, self-attributions following these three types of trauma were the best candidates for meta-analysis. Further, studies concerning these types of traumas typically were found to include clear operational definitions of predictor variables. In addition, the researchers found few studies of self-attribution studies for other kinds of trauma, thereby making these three types of trauma the best candidates for meta-analysis.

Studies for inclusion in the meta-analysis were identified in several ways. The authors conducted literature searches in several professional databases using a number of keywords and phrases. In addition, researchers in the field were contacted directly and asked for data from any unpublished studies. Studies compiled in these two ways were discarded if they did not involve individuals reporting self-attribution following a trauma s/he actually experienced or if the data on effect size that is necessary to conduct of the meta-analysis could not be obtained from the study or its authors. This resulted in a total of 69 studies of self-attributions that were used in the meta-analysis. Thirty-four of the studies analyzed self-attributions following sexual victimization (i.e., rape, sexual abuse, incest, sexual harassment, attempted rape). Twenty-two studies analyze self-attributions following illness. The final 13 studies examine self-attribution following severe injury (i.e., spinal cord injury, severe burns, head injury). Effect sizes or self-attributions were calculated and reported level of self-attribution in each study was compared to the level expected due to chance. In addition, several potential predictors of self-attributions as well as degree of life thread (i.e. high or low) were coded. For each categorical variable examined, an analysis of variance was conducted.

The results of the meta analysis showed that self-attribution occurred at a lower than chance level for most of these traumas. The analysis of methodological differences showed that open-ended queries regarding attribution resulted in significantly lower levels of self attribution than did closed-end queries, although there was considerable heterogeneity in effect sizes among the studies. The results also showed that reporting self-attribution did not vary significantly among studies based on the measure used for this variable. The analysis also found that reported the level of self-attribution decreased as the lengths of the trauma in the studies increased. Similarly, there is a tendency for the level of self-attribution reported to decrease as the mean age of the subjects increased. Other potential predictor variables do not appear to predict self-attribution.

Conclusion

Obtaining primary data for social science research can be difficult for a number of ethical and logistical reasons. However, there are a number of public and private resources that offer the researcher secondary data that can be reanalyzed for further investigations. Although this makes data collection for the most part both easier and less expensive, the analysis of secondary data is not without its disadvantages as well. However, a well-designed study using secondary analysis or meta analysis can be of great advantage to the researcher who cannot obtain primary data from another source and can significantly aid in the advancement of the science.

Terms & Concepts

Experiment: A situation under the control of a researcher in which an experimental condition (independent variable) is manipulated and the effect on the experimental subject (dependent variable) is measured. Most experiments are designed using the principles of the scientific method and are statistically analyzed to determined whether or not the results are statistically significant.

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data. Inferential statistics are used to make inferences such as drawing conclusions about a population from a sample and in decision making.

Inter-Interviewer Reliability: The consistency with which different interviewers obtain similar responses from subjects using the same interview instrument. Interviewer bias and interviewer effects can lead to low inter-interviewer reliability.

Interviewer Bias: The expectations, beliefs, prejudices, or other attitudes that may affect the interview process and the subsequent interpretation of data collected through the interview process.

Interviewer Effects: The influence of the interviewer's behaviors and attributes on the subject's response in an interview situation. For example, the appearance, demeanor, training, age, gender, and ethnicity of an interviewer may all affect the way that a subject perceives the interview and/or responds to the questions on an interview. In some cases, the subject may try to please the interviewer by giving responses that s/he thinks the subject may want to hear or in other cases may give non-responsive answers in order to negatively impact the value of the data collected by an interviewer that s/he does not like.

Meta Analysis: A secondary analysis technique used to synthesize the results of multiple existing quantitative research studies of a single phenomenon into a single result. Statistically, meta analysis combines the effect size estimates of the individual studies into a single estimated effect size or a distribution of effect sizes.

Qualitative Research: Scientific research in which observations cannot be or are not quantified (i.e., expressed in numerical form).

Quantitative Research: Scientific research in which observations are measured and expressed in numerical form (e.g., physical dimensions, rating scales).

Secondary Analysis: A further analysis of existing data typically collected by a different researcher. The intent of secondary analysis is to use existing data in order to develop conclusions or knowledge in addition to or different from those resulting from the original analysis of the data. Secondary analysis may be qualitative or quantitative in nature and may be used by itself or combined with other research data to reach conclusions.

Subject: A participant in a research study or experiment whose responses are observed, recorded, and analyzed.

Survey: (a) A data collection instrument used to acquire information on the opinions, attitudes, or reactions of people; (b) a research study in which members of a selected sample are asked questions concerning their opinions, attitudes, or reactions are gathered using a survey instrument or questionnaire for purposes of scientific analysis; typically the results of this analysis are used to extrapolate the findings from the sample to the underlying population; (c) to conduct a survey on a sample.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables (response). Extraneous variables are variables that affect the response but that are not related to the question under investigation in the study.

Bibliography

Campbell, D. A. (2007). Secondary analysis. *Orthopaedic Nursing, 26* (4), 241-242. Retrieved April 23, 2008, from EBSCO Online Database Academic Search Premier. http://web.ebscohost.com/ehost/pdf?vid=13&hid=21&sid=62757d2f-5bcb-4125-bd0a-894bb563f1a8%40sessionmgr2

Horwitz, S. M., Briggs-Gowan, M. J., Storfer-Isser, A., & Carter, A. S. (2007). Prevalence, correlates, and persistence of maternal depression. *Journal of Women's Health, 16* (5), 678-691. Retrieved April 24, 2008, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=25705603&site=ehost-live

Littleton, H. L., Magee, K. T., & Axsom, D. (2007). A meta-analysis of self-attributions following three types of trauma: Sexual victimization, illness, and injury. *Journal of Applied Social Psychology, 37* (3), 515-538. Retrieved April 24, 2008, from EBSCO Online Database SocINDEX with Full Text. http://web.ebscohost.com/ehost/pdf?vid=12&hid=13&sid=044e731e-f467-4ff0-b74c-178c235c732d%40sessionmgr9

Schaefer, R. T. (2002). *Sociology: A brief introduction* (4th ed.). Boston: McGraw-Hill.

Stockard, J. (2000). *Sociology: Discovering society* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Vartanian, T.P. (2011). *Secondary data analysis*. New York: Oxford University Press.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences.* Chicago: Rand McNally College Publishing Company.

Young, R., & Johnson, D. (2013). Methods for handling missing secondary respondent data. *Journal of Marriage & Family, 75*(1), 221-234. Retrieved October 23, 2013 from EBSCO online database, SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=84935698&site=ehost-live

Suggested Reading

Andrews, L., Higgins, A., Andrews, M., & Lalor, J. G. (2012). Classic grounded theory to analyse secondary data: Reality and reflections. *Grounded Theory Review, 11*(1), 12–26. Retrieved October 23, 2013 from EBSCO online database, SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=77668319&site=ehost-live

Bennett, T., Holloway, K., & Farrington, D. (2006). Does neighborhood watch reduce crime? A systematic review and meta-analysis. *Journal of Experimental Criminology, 2* (4), 437-458. Retrieved April 24, 2008, from EBSCO Online Database SocINDEX with Full Text. http://web.ebscohost.com/ehost/pdf?vid=13&hid=13&sid=044e731e-f467-4ff0-b74c-178c235c732d%40sessionmgr9

Dixon, A., Khachatryan, A., & Yang, T. (2012). Socioeconomic differences in case finding among general practices in England: Analysis of secondary data. *Journal of Health Services Research & Policy, 17*(s2), 18-22. Retrieved October 23, 2013 from EBSCO online database, SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=78123891&site=ehost-live

Gibbs, D., Berkman, N., Weitzenkamp, D., & Dalberth, B. (2007). Federal support for adoption subsidies: State-level variations and the impact for adoptive families. *Journal of Public Child Welfare, 1* (2), 71-90. Retrieved

April 24, 2008, from EBSCO Online Database SocINDEX with Full Text. http://web.ebscohost.com/ehost/pdf?vid=9&hid=13&sid=044e731e-f467-4ff0-b74c-178c235c732d%40sessionmgr9

Hall, M., & Farkas, G. (2011). Adolescent cognitive skills, attitudinal/behavioral traits and career wages. *Social Forces, 89*(4), 1261–1285. Retrieved October 23, 2013 from EBSCO online database, SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=60914643&site=ehost-live

Petrosino, A. & Lavenberg, J. (2007). Systematic reviews and meta-analyses: Best evidence on "what works" for criminal justice decision makers. *Western Criminology Review, 8* (1), 1-15. Retrieved April 24, 2008, from EBSCO Online Database SocINDEX with Full Text. http://web.ebscohost.com/ehost/pdf?vid=11&hid=13&sid=044e731e-f467-4ff0-b74c-178c235c732d%40sessionmgr9

Trzesniewski, K.H., Donnellan, M.B., Lucas, R.E. (2011). *Secondary data analysis: An introduction for psychologists*. Washington, DC: American Psychological Association.

Essay by Ruth A. Wienclaw, Ph.D.

Ruth A. Wienclaw holds a doctorate in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human/systems integration.

Inferential Statistics

Table of Contents

Abstract

In order to better describe, explain, and predict the behavior of groups of people in society, sociologists make observations, develop hypotheses, and collect data with the intent of drawing conclusions from their findings. Inferential statistics is a family of tools that are used to support these efforts by allowing sociologists to draw conclusions from data and test whether or not the results of a study are due to chance or to some underlying phenomenon. A wide range of statistical methods are available for testing hypotheses. Each of these methods is appropriate to a different type of experimental design. Some of these tools include t-tests, the z statistic, analysis of variance (ANOVA), and regression analysis.

Overview

Using Statistics to Analyze Data

The overarching goal of sociological research is to describe, explain, and predict the behavior of people within society. To this end, sociologists observe behavior, develop hypotheses, collect data, and draw conclusions from their findings. Although it would be possible to perform these activities based on the input from a few people, human beings are infinitely diverse. Consequently, in most situations, it is virtually impossible to predict the behavior of a large group of people based on the actions of just one individual.

For example, despite our attempts at prognostication, it can be difficult to predict the outcome of a political election. Different people can look at the same data concerning opposing candidates and draw vastly different conclusions about the candidates' likelihood of being elected. Even the fact that a certain percentage of eligible voters cast their ballots for a given candidate in a primary election does not necessarily mean that they will do so again in the general election. An independent, for example, may vote in the primary of one party in order to help ensure that the candidate he or she prefers from that party is nominated in case the candidate of his or her choice from an opposing party is not elected. Another voter might try to ensure that an opposing party's least electable candidate is nominated so that the candidate from his or her preferred party has a better chance of winning the general election. Therefore, it is difficult to extrapolate from the fact that a candidate won a primary election that he or she will win the general election.

To be better able to meet the sociological goal of predicting behavior within society, it is important that data be collected from a wide range of individuals. In this way, patterns greater than the opinions or actions of a given individual or small group can emerge, and the sociologist can draw conclusions about the actions of a population that are based on a representative sample of the population.

One way to do this is by collecting data from a wide variety of people and determining what the average response to a situation is. The use of descriptive statistics, which measure the central tendency or "average" (mean, median, and mode) of a sample, may give us a better picture of the inclinations of the population. However, this is still a very restricted picture.

For example, if we give registered voters a questionnaire asking them to rate on a scale of 1 to 10 how much they like a certain candidate, the average answer might be 4.5. Based on this piece of information, we might conclude that the candidate is neither well liked nor strongly disliked. However, how the raw data falls on the scale is very important. If the raw data were clustered around the middle of the scale, this conclusion would probably be correct; if the raw data were evenly distributed across the scale, this conclusion would be less warranted, and we would need to conduct further investigations to determine how much the candidate is really liked. Similarly, if the data were polarized, with approximately half the people polled disliking the candidate extremely and the other half liking the candidate extremely, we would still have the same 4.5 "average" score, despite the fact that no one was ambivalent about the candidate.

Such problems with interpretation are not the only drawback to solely using descriptive statistics to draw conclusions from a sample. On the same 10 point scale, can we say with confidence that there is truly a difference between a score of seven and a score of eight? To overcome these and other limitations of descriptive statistics, sociologists and other scientists turn to inferential statistics in order to draw conclusions, or inferences, from their data.

What Is Inferential Statistics?

Inferential statistics is a subset of mathematical statistics that is used in the analysis and interpretation of data. In the examples above, inferential statistics could better help us understand the results of the primary data or meaningfully interpret the results of the polling data. Inferential statistics is used to test hypotheses to determine if the results of a study have statistical significance, meaning that they occur at a rate that is unlikely to be due to chance.

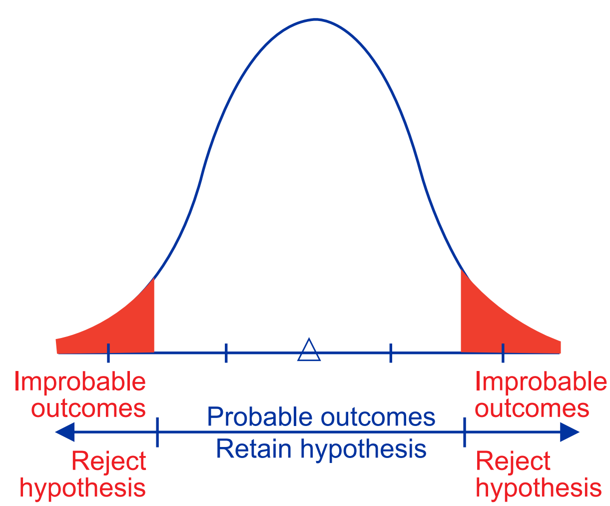
Testing Hypotheses

A hypothesis is an empirically verifiable declarative statement concerning the relationship between independent and dependent variables and their corresponding measures. An example of a hypothesis might be the assertion that people offer friendship more readily to those they feel are similar to themselves than they do to others. In this example, the independent variable, or the variable that is manipulated by the researcher, is the degree of similarity between the subject and the people around him or her. The dependent variable, or the subject's response to the independent variable, would be to whom the subject offers friendship.

Hypotheses are stated in two ways. A null hypothesis (H0) is a statement that denies that there is a statistical difference between the status quo and the experimental condition. In other words, it states that the independent variable being studied makes no difference to the end result. For example, a null hypothesis about people's preference for befriending others with whom they believe they have something in common might be, "There is no difference in the number of overtures of friendship made to strangers based on whether or not the strangers have something in common with the subject." This null hypothesis states that there is no relationship between the independent variable of perceiving that one has something in common with another person and the dependent variable of whether or not the person being studied makes an overture of friendship to that other person. The alternative hypothesis (H1) would be that there *is* a relationship between the two variables—for example, "People offer friendship more readily to those they feel are similar to themselves than they do to others."

Once the null hypothesis has been formulated, an experimental design is developed that allows the researcher to empirically test the hypothesis. Typically, the experimental design includes a control group that that does not receive the experimental conditions and an experimental group that does. In this case, the individuals in the control group would not be exposed to people who are markedly different from them, while the individuals in the experimental group would be. The researcher then collects data from people in the study to determine whether or not the experimental condition had any effect on the outcome. After the data have been collected, they are statistically analyzed to determine whether the null hypothesis should be accepted or rejected. Accepting the null hypothesis means that if the data in the population are normally distributed, the results are more than likely due to chance. This is illustrated in Figure 1 as the unshaded portion of the distribution. By accepting the null hypothesis, the researcher is concluding that it is likely that people do not react any differently to the people they perceive as different than they do to those whom they perceive as the same. For the null hypothesis to be rejected and the alternative hypothesis to be accepted, the results must lie within the shaded portion of the graph shown in Figure 1. When this occurs, there is a statistically significant chance that the difference observed between the two groups is due not to chance but rather to a real underlying difference in people's attitudes toward various types of strangers.

Figure 1: Hypothesized Sampling Distribution of the Mean Showing Areas of Acceptance and Rejection of the Null Hypothesis



(adapted from Witte, p. 118)

Analyzing Data with Inferential Statistics

Part of the process of designing an experiment is determining how the data will be analyzed. There are a number of different inferential statistical methods available for testing hypotheses. However, each method is appropriate to a different type of experimental design, and an experiment needs to be appropriately designed in order to use any method.

One class of frequently used statistical tests is the *t*-test. This type of statistical technique is used to analyze the mean of a population or compare the means of two different populations. In other situations where one wishes to compare the means of two populations, a *z* statistic may be used.

Another frequently used technique for analyzing data in applied settings is analysis of variance (ANOVA). This is a family of techniques that are used to analyze the joint and separate effects of multiple independent variables on a single dependent variable and determine the statistical significance of those effects. For example, ANOVA might be used if one wished to determine the types of reactions that subjects have to three different kinds of people (e.g., those who are similar, those who are of a different gender, and those who are of a different socioeconomic status). For more complex problems, an extension of ANOVA called multivariate analysis of variance (MANOVA) allows researchers to test hypotheses involving the simultaneous effects of multiple independent variables on multiple dependent variables.

Other types of applied statistics allow researchers to predict one variable based on knowledge of another variable. Correlation shows the statistical degree of relationship between two variables. For example, one might want to know the correlation between a child being raised in a single-parent home and academic success. Correlation coefficients allow researchers to determine whether the two variables are positively correlated (i.e., the more time a child spends in a one-parent home, the greater success he or she tends to have in school), negatively correlated (i.e., the more time a child spends in a one-parent home, the less success he or she tends to have in school), or not correlated at all. However, correlation does not tell us anything about causation. Even if we find that there is a high positive correlation between being in a single-parent home and success in school, we do not know if this success is due to the home environment, if the home environment is due to the academic success, or if both conditions are caused by a third, unknown condition.

Real-world problems do not always have easy answers that involve only two variables. How well a child does in school may depend on whether or not he or she lives with two parents, but other factors may affect academic performance as well. For example, success may also be affected by the parent's involvement in the child's life, the parents' prior academic success, the family's socioeconomic status, and various other factors. Regression analysis is a family of statistical techniques that allow researchers to predict the dependent variable score when given the scores of one or more independent variables. Multiple regression techniques analyze the effects of multiple predictors on behavior. Using these techniques gives the researcher a better understanding of these predictors' relative contributions to the behavior, as well as the factors that determine an individual's response to a situation.

Applications

Inferential statistics are used by sociologists in a wide variety of situations in which it is possible to gather quantifiable data. One such a study using several of these techniques was performed by Lee and Chang (2006) to investigate the social inequalities of lottery advertising on consumer welfare.

How Does Lottery Advertising Affect Consumer Welfare?

It has often been observed, both informally and empirically, that lotteries are most frequently played by those who can least afford to lose their money. Yet the promise of high return for a minimal investment is seductive, and lotteries continue to be an important source of income for many governments. In fact, lotteries have been referred to as a painless or voluntary tax, or even as a tax on stupidity.

If playing the lottery were truly a voluntary and uncoerced activity over which the sponsoring governments or agencies had no control, one might be tempted to agree with the pundits that lotteries are, in fact, a tax on stupidity. However, when these governments or agencies advertise their lotteries with full knowledge that the probability of winning is extremely low and those who play are disproportionately of lower socioeconomic statuses, lotteries become an interesting sociological and ethical question. The goal of lottery advertising is not only to gain new players but also to encourage existing players to play more frequently, a tactic that further increases lotteries' negative effects on those who play. Like most advertising, lottery ads tend to emphasize the rewards of participation and give short shrift to the potential pitfalls. Strong, positive slogans emphasizing the possibility of attaining sudden wealth typically overshadow the risks lotteries pose.

Lee and Chang performed an empirical study of the effect of lottery advertisements in Taiwan. They investigated "exposures to various ad channels, advertising recalls, cognitive and affective responses to lottery ads, perceived positive and negative consequences of lottery ads, and perceived social responsibility of lottery ads" (Lee & Chang, 2006). The authors used inferential statistics to evaluate "the statistical significance of socioeconomic, demographic, psychological, cognitive, and attitudinal variables" on the decision to purchase a lottery ticket (Lee & Chang, 2006). The study also examined the similarities and differences between the effects of the advertisements on varying socioeconomic and demographic groups. Data for the study were gathered using a four-page questionnaire, which was administered to 975 participants who were old enough to legally purchase lottery tickets at various locations.

Several inferential statistical tools were used to analyze the data, including a regression analysis to estimate the effects of the various demographic variables and perceptions of lottery advertisements on the probability that an individual would purchase a lottery ticket. The results of this analysis suggested that age, income, prior purchase of lottery tickets, exposure to lottery advertisements, and negative reaction to lottery advertisements were among the predictors of participation the lottery.

An ANOVA technique was used to analyze the data in order to investigate the interaction between effects of lottery advertising and income levels. The dependent variable in this study was the amount of money spent on the last lottery purchase. The researchers found that there was no relationship between awareness of lottery advertising and income level but that perceptions of lottery advertising did vary across income levels. Participants with greater incomes tended to have more negative perceptions of the lottery ads and the lottery's potential to bring them income than did participants with lower incomes. This finding supports the critics who maintain that lottery advertisements have a stronger effect on individuals of lower socioeconomic statuses who can ill afford to spend money on the lottery. Based on the results of the statistical analyses, the authors suggested that lottery ticket advertisements should place more emphasis on the negative aspects of playing the lottery, and they recommended further study to determine how this could best be done.

Conclusion

Inferential statistics is a subset of mathematical statistics that enables sociologists to analyze and interpret the data that they observe either in the field or in the laboratory. These statistical tools help sociologists make inferences from data so that they can better describe, explain, and predict behavior. Specifically, inferential statistics is used to test hypotheses to determine if the results observed in a study occurred at a rate that is unlikely to be due to chance. If this is the case, the results are said to have statistical significance. A wide range of inferential statistical tools can be applied to sociological problems to help researchers better understand the behavior of people in a society.

Terms & Concepts

Analysis of Variance (ANOVA): A family of statistical techniques that analyze the joint and separate effects of multiple independent variables on a single dependent variable and determine the statistical significance of the effect.

Data: In statistics, quantifiable observations or measurements that are used as the basis of scientific research.

Dependent Variable: The outcome variable, or the resulting behavior that changes depending on whether the subject receives the control or experimental condition.

Descriptive Statistics: A subset of mathematical statistics that describes and summarizes data.

Distribution: A set of numbers collected from data and their associated frequencies.

Hypothesis: An empirically verifiable declarative statement concerning the relationship between independent and dependent variables and their corresponding measures.

Independent Variable: The variable in an experiment or research study that is intentionally manipulated in order to determine its effect on the dependent variable.

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data. Inferential statistics is used to make inferences, such as drawing conclusions about a population from a sample. It can also be used in decision making.

Null Hypothesis (H0): The statement that the findings of an experiment will show no statistical difference between the control condition and the experimental condition.

Population: The entire group of subjects belonging to a certain category, such as all women between the ages of 18 and 27, all dry-cleaning businesses, or all college students.

Regression Analysis: A family of statistical techniques used to develop a mathematical model for predicting one variable from the knowledge of another variable.

Sample: A subset of a population. A random sample is a sample that is chosen at random from the larger population with the assumption that it will reflect the characteristics of the larger population.

Statistical Significance: The degree to which an observed outcome is unlikely to have occurred due to chance.

Statistics: A branch of mathematics that deals with the analysis and interpretation of data. Mathematical statistics provides the theoretical underpinnings for various applied statistical disciplines, including business statistics, in which data are analyzed to find answers to quantifiable questions. Applied statistics uses these techniques to solve real-world problems.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables, or response variables. Extraneous variables are variables that affect the dependent variables but are not related to the question under investigation in the study.

Bibliography

Anderson, M. L. & Taylor, H. F. (2002). *Sociology: Understanding a diverse society* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Calderwood, K. A. (2012). Teaching inferential statistics to social work students: A decision-making flow chart. *Journal of Teaching in Social Work, 32*(2), 133–147. Retrieved November 6, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=74639228&site=ehost-live

Chan, L. S. (2013). Minimal clinically important difference (MCID): Adding meaning to statistical inference. *American Journal of Public Health, 103*(11), e24-e25. Retrieved November 6, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=90627354&site=ehost-live

Gupta, S. K. (2012). The relevance of confidence interval and P-value in inferential statistics. *Indian Journal of Pharmacology, 44*(1), 143–144. Retrieved November 6, 2013, from EBSCO Online Database Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=70787218&site=ehost-live

Lee, Y.-K. & Chang, C.T. (2006). Social inequalities of lottery advertising on consumer welfare. In *Conference Papers - American Sociological Association, 2006 Annual Meeting, Montreal* (1-17). Washington, DC: American Sociological Association. Retrieved March 13, 2008 from EBSCO Online Database SocINDEX with Full Text. http://web.ebscohost.com/ehost/pdf?vid=22&hid=17&sid=a3878266-c212-4e78-95f9-556697cc9da2%40sessionmgr102

Witte, R. S. (1980). *Statistics* . New York: Holt, Rinehart and Winston.

Suggested Reading

Feld, S. L.. (1997, Mar). Mathematics in thinking about sociology. *Sociological Forum, 12* (1), 3-9. Retrieved March 13, 2008 from EBSCO Online Database Academic Search Complete. http://web.ebscohost.com/ehost/pdf?vid=11&hid=17&sid=a3878266-c212-4e78-95f9-556697cc9da2%40sessionmgr102

Gravetter, F. J. & Wallnau, L B. (2006). *Statistics for the behavioral sciences*. Belmont, CA: Wadsworth/Thomson Learning.

Moulton, L. H., Peña, J. B., Masyn, K. E., & Wang, Y. (2012). Show us the data. *American Journal of Public Health, 102*(10), e5–e6. Retrieved November 6, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=82057554&site=ehost-live

Young, R. K. & Veldman, D. J. (1977). *Introductory statistics for the behavioral sciences* (3rd ed.). New York: Holt, Rinehart and Winston.

Essay by Ruth A. Wienclaw, PhD

Dr. Ruth A. Wienclaw holds a PhD in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human-systems integration.

Descriptive Statistics

Table of Contents

Abstract

Descriptive statistics comprises a set of statistical tools that help sociologists, researchers, and other analysts better understand the masses of data with which they need to work. These tools include various types of charts and graphs to visually display the data so that they can be more easily understood, measures of central tendency that estimate the midpoint of a distribution, and measures of variability that summarize how widely dispersed the data are over the distribution. Each measure of central tendency and variability has particular strengths and weaknesses and should only be used under certain conditions. Descriptive statistics do not allow one to make inferences about the data or to determine whether or not the data values are statistically significant. Rather, they only describe data.

Overview

At its most basic, sociology is the study of humans within society. In order to better understand human behavior from this perspective, sociologists attempt to describe, explain, and predict the behavior of people in social contexts. At first glance, this task seems deceptively simple. After all, we usually know how and why we react the way we do in various situations. It should seem a simple step to extrapolate from our own attitudes and behavior to those of people in general. However, it is not valid to assume that everyone thinks or behaves in the same way. Human beings are infinitely diverse, and often two people can look at the same data or situation and arrive at two very different conclusions.

For example, although all voters have access to the same information during a presidential race, these races can be hotly contested, and voters can fiercely disagree over a candidate's merits. Even within the same party, voters can be divided over a candidate, with some giving credence to one piece of information about the candidate and others valuing another piece. It is a truism that people can look at the same situation and honestly disagree. For this reason, it is impossible to extrapolate from the attitudes or behavior of one individual to society at large. To truly describe, explain, and predict the behavior of people in social contexts, sociologists must acquire data on the attitudes and behaviors of more than one individual.

Just as data collected from only one individual is not of much use to sociologists, neither is data collected from a mere two or three people. Sociologists need to gather data from a large number of people in order to have any confidence that their findings can be extrapolated to people in general. The number of people used in sociological research studies routinely reaches in the hundreds for just this reason. Although hundreds or even thousands of inputs will give us a better picture of how people actually react or behave, this massive amount of data leads to another problem: How can we make sense of all the data and interpret them in a meaningful way? Fortunately, the field of mathematics offers us numerous statistical tools that can aid us in this task.

When thinking of statistics, most people think of inferential statistics, which is a subset of mathematical statistics used in the analysis and interpretation of data. Inferential statistics are used to make inferences from data, such as drawing conclusions about a population based on a sample. This branch of statistics comprises the seemingly arcane formulae and mathematical computations that so many students dread.

However, there is another class of statistical tools that is used to summarize data and develop inputs for use in inferential statistical computation. Although not a substitute for inferential statistics, descriptive statistics is very useful in helping sociologists better understand the masses of data with which they need to work. In general, descriptive statistics is a subset of mathematical statistics that describes and summaries data. Descriptive statistics are used to summarize and display data through various types of charts and graphs, such as histograms and pie charts. Using these tools, one can easily get a rough idea of the shape of the data; describe the "average" of the data through measures of central tendency, including the mean, median, and mode; and summarize the variability of the data through such measures as the standard deviation, the semi-interquartile deviation, and the range.

Applications

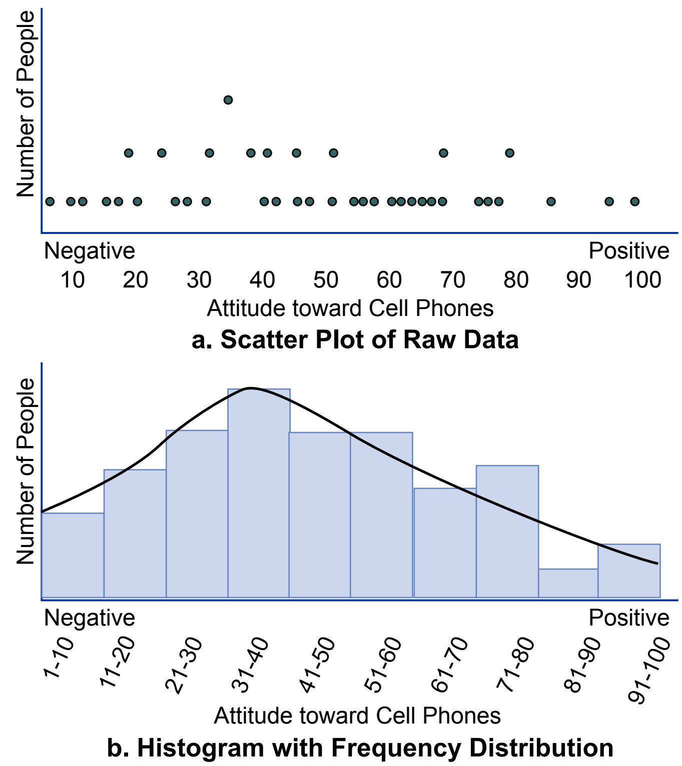
Graphing

One subset of descriptive statistics comprises various graphing techniques that help one organize and summarize data so that they are more easily comprehended. One of the most common and helpful methods for doing this is a frequency distribution. In this technique, data are divided into intervals of typically equal length using techniques such as a stem-and-leaf plot or a box-and-whiskers plot. Graphing data within intervals rather than as individual data points reduces the number of data points on the graph, making the graph—and the underlying data—easier to comprehend.

For example, one might seek to understand people's attitudes about the effects of cell phone use on driving behavior by asking 1,000 people to rate the effects on a scale of 1 to 100, with 1 being the most negative and 100 being the most positive. However, it would be difficult to display these results by graphing all 1,000 points. There would be several clusters of data points where a number of people gave the same response, as well as clusters of data points where people gave similar but not identical responses. Although displaying the data in this way certainly shows the full range of people's responses, it is difficult to interpret the data because of the large number of data points. In addition, one must question whether there is truly a meaningful difference between a rating of 22 on a 100-point scale and a rating of 23. Both of the people responding believed that cell phone usage had a negative effect on driving behavior, but can one really say that the person who responded with a 22 felt that much more negatively about the effects of cell phone usage than the person who responded with a 23? Probably not.

Therefore, it is reasonable to aggregate the data into ranges within the span of scores (e.g., 1–10, 11–20, etc.) before graphing them. As a result, the number of points on the graph is decreased and larger patterns can emerge. Figure 1 shows a comparison between a scatter plot of raw data and a histogram with a superimposed frequency distribution.

Figure 1: Scatter Plot & Histogram with Frequency Distribution for the Same Data Set



Measures of Central Tendency

Although graphing the data using this or other graphing techniques is helpful for better understanding the shape of the underlying distribution, other statistical tools, like measures of central tendency and measures of variability, can be used to understand the data even more thoroughly.

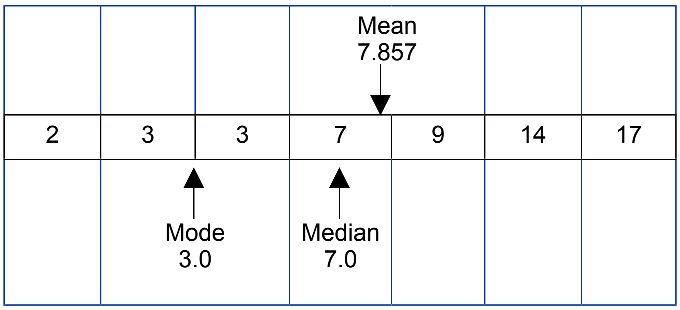
Measures of central tendency estimate the midpoint of a distribution. These measures include

- the median, or the number in the middle of the distribution when the data points are arranged in order;
- the mode, or the number that occurs most often in the distribution; and
- the mean, or the sum of all data values in the distribution divided by the total number of data points in the distribution.

These three methods frequently give different estimates of the midpoint of a distribution because they are all affected differently by the shape of the distribution and by any outlying points.

For example, as shown in Figure 2, for the data set 2, 3, 3, 7, 9, 14, 17, the mode is 3, as there are two 3s in the distribution, but only one of each of the other numbers; the median is 7, since, when the seven numbers in the distribution are arranged numerically, 7 is the number that occurs in the middle; and the mean (or arithmetic mean) is 7.857, since the sum of the seven numbers is 55 and 55 ÷ 7 = 7.857.

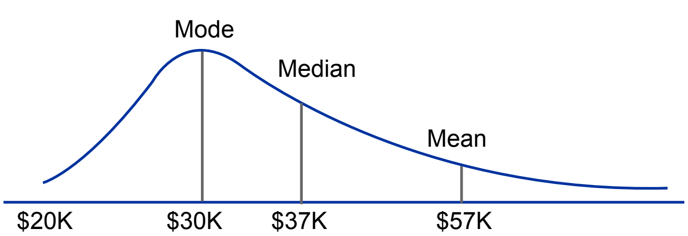
Figure 2: Measures of Central Tendency for a Simple Distribution



The three measures of central tendency all have different characteristics. It is important to remember that these differences are real and the measures are not interchangeable. For example, in a skewed distribution, where one end has extreme outliers but the data is otherwise normally distributed, the median may be pulled toward the skew (i.e., toward the end of with the outliers). Because of this, when the ends are not balanced and data are clustered toward one end of the distribution, the median may disproportionately reflect the outlying data points.

The mean is even more affected by extreme scores. If, for example, one wants to know the "average" salary of the salaries shown in Figure 3, one has to carefully consider just how the distribution affects the mean. In this case, it may be more accurate to report the average salary as the mode rather than the mean due to the small proportion of people who make a much higher salary than other people in the same field. As shown in the figure, this small proportion of people pulls the mean in the direction of the skew.

Figure 3: Placement of Different Measures of Central Tendency on a Skewed Distribution



(Adapted from Huff, p. 33)

Each measure of central tendency is best used under different circumstances. The mode has the obvious advantage of being quick and easy: one need only determine which number occurs most frequently in the distribution. However, this same characteristic means that the mode also has the disadvantage of lacking stability: a small change in the numbers can lead to a great change in the mode. Because the mode does not take actual score values into account, it is not really valuable for any purpose other than to state which number has the highest frequency.

The median is a more stable measure than the mode, and occasionally it may stand alone as a statistic. In fact, the median is preferred to both the mode and the mean for use in non-symmetrical distributions because it is less variable than the mode and less affected by extreme scores than the mean. However, like the mode, the median is a terminal statistic: it cannot be used to make statistical inferences about the data.

The mean has advantages in most situations over the other two measures of central tendency, and it is not a terminal statistic, meaning it can be used as an input for many inferential statistical techniques. The mean is highly stable, and its value does not vary greatly because of a change in a single score. Because of these traits, it is generally advisable to always use the mean as the measure of central tendency unless there is a compelling reason not to do so, such as a non-symmetrical distribution with extreme outliers.

Measures of Variability

It is important to note that although measures of central tendency give a quick measure of the "average" value in a distribution, this information by itself is insufficient to truly understand the distribution of the underlying data. For example, the data may be evenly spread across the distribution, cluster in the middle, or cluster at either end. Yet all of these distributions can yield the same value for the mean. Measures of central tendency are helpful for better understanding large amounts of data, but they are only one part of the puzzle. For example, without seeing the graph of the distribution, knowing that a sample of data has a mean of 10 does not give one much information about the data. One needs additional information in order to really understand what the data signify. The scope and signification of the data set can be better understood by knowing how far the data points are from each other, what the end points of the distribution are, and, in general, how the data are distributed. To better understand this aspect of a collection of data, one uses measures of variability. Measures of variability are descriptive statistics that summarize how widely dispersed the data are over the distribution. Specifically, these measures are the range, the semi-interquartile deviation, and the standard deviation, corresponding to the mode, the median, and the mean, respectively.

The range is a statement of the difference between the highest and lowest scores in the distribution. In conjunction with a measure of central tendency, this information helps one better understand the data. For example, if a class's mean score on a test was 60 out of a total possible score of 100, one would draw different conclusions about the class's abilities if the lowest and highest scores were 0 and 100 than if they were 50 and 70. Looking at the distribution within the first range, it would appear that more people got over half of the questions correct, because otherwise the mean would be less than 50. In the second case, it would appear that either no one understood the material well enough to get a large majority of the questions correct or a significant number of questions were badly worded, because no one earned a score of more than 70 out of 100. Some distributions, as in the first case, have outlying data, or stragglers at one or both ends of the distribution that are far removed from the rest of the data. The range, however, treats all values in the distribution alike and does not give consideration to whether or not they are outliers.

Like the median, the semi-interquartile deviation is a positional measure that eliminates the extreme scores on both ends of the distribution. To determine the semi-interquartile deviation, one divides the distribution into quarters. The first quartile (Q1) is determined by finding the median of the lower half of the distribution (i.e., the number with 25 percent of the numbers in the distribution below it). The third quartile (Q3) is determined by finding the median of the upper half of the distribution (i.e., the number with 25 percent of the numbers in the distribution above it). The semi-interquartile deviation (Q) is then calculated by subtracting the value of the first quartile from the value of the third quartile and dividing this number by two: $Q = (Q3 - Q1)/2$.

Just as the mean is a mathematically derived measure of central tendency, the standard deviation is a mathematical determination of the variability of a distribution. This statistic is an index of the degree to which scores differ from the mean of the distribution, making it a measure of variability that describes how far the typical score in a distribution is from the mean of the distribution. This statistic is obtained by subtracting the mean of the distribution from each score in order to determine the deviation of each score from the mean, squaring each resulting deviation, adding the squared deviations, and dividing this number by the total number of scores. The larger the standard deviation, the farther away the typical score is from the mean of the distribution.

Like measures of central tendency, each measure of variability has its own strengths and weaknesses. One of the uses of the range is to determine how many intervals should be used when developing a frequency distribution. The range is also the best method for determining variability if all one wants to do is look at the distribution. However, the range is highly unstable and easily affected be extreme scores. Further, it is a terminal statistic, not useful for much more than describing the distribution of the data. The semi-interquartile deviation has an advantage over the range in that it eliminates the extreme scores at both ends of the distribution, thereby making it more stable. In addition, the semi-interquartile deviation is a quick method for finding out whether or not a distribution is skewed. However, like the range, it is a terminal statistic. For most circumstances, particularly those in which one wants to do additional analysis of the data and make statistical inferences, the standard deviation is the best tool to use for describing the variability in a distribution. Like the mean, the standard deviation is used as the basis for inferential statistical techniques.

Conclusion

Descriptive statistics is a class of statistical tools that is very useful in helping sociologists, researchers, and other analysts better understand the masses of data with which they need to work. Descriptive statistics are used to summarize and display data in various types of charts and graphs, such as histograms and pie charts; mathematically describe what the "average" of the data is through measures of central tendency, including the mean, median, and mode; and summarize the variability of the data through such measures as the standard deviation, the semi-interquartile deviation, and the range. Each measure of central tendency and variability has different strengths and weaknesses, and the measures are not interchangeable.

It is important to remember that descriptive statistics do just that: describe the data. They do not allow one to make inferences about the data or determine whether or not the data values are statistically significant. This type of operation belongs to the realm of inferential statistics.

Terms & Concepts

Box-and-Whiskers Plot: A graphing technique that summarizes a data set by depicting the upper and lower quartiles, the median, and the two extreme values of a distribution. Also known as a box plot or a candlestick chart.

Data: In statistics, quantifiable observations or measurements that are used as the basis of scientific research.

Descriptive Statistics: A subset of mathematical statistics that describes and summarizes data.

Distribution: A set of numbers collected from data and their associated frequencies.

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data, as well as in decision making.

Mean: An arithmetically derived measure of central tendency in which the sum of the values of all the data points is divided by the total number of data points.

Measures of Central Tendency: Descriptive statistics that are used to estimate the midpoint of a distribution. Measures of central tendency include the median, the mode, and the mean.

Measures of Variability: Descriptive statistics that summarize how widely dispersed the data are over the distribution. The range describes the difference between the highest and lowest scores, the semi-interquartile deviation is a positional measure that eliminates the extreme scores on both ends of the distribution, and the standard deviation is a mathematically derived index of the degree to which scores differ from the mean of the distribution.

Median: The number in the middle of a distribution when all values are placed in order. A measure of central tendency.

Mode: The number that occurs most often within a distribution. A measure of central tendency.

Population: The entire group of subjects belonging to a certain category, such as all women between the ages of 18 and 27, all dry-cleaning businesses, or all college students.

Quartile: Any of three points that divide an ordered distribution into four equal parts, each of which contains one quarter of the data.

Sample: A subset of a population. A random sample is a sample that is chosen at random from the larger population with the assumption that it will reflect the characteristics of the larger population.

Skewed: A distribution that is not symmetrical around the mean, meaning that there are more data points on one side of the mean than on the other.

Statistics: A branch of mathematics that deals with the analysis and interpretation of data. Mathematical statistics provides the theoretical underpinnings for various applied statistical disciplines in which data are analyzed to find answers to quantifiable questions. Applied statistics uses these techniques to solve real-world problems.

Stem-and-Leaf Plot: A graphing technique in which individual data points are broken into the rightmost units ("leaves") and the leftmost units ("stems"). For example, the number 42 would have a stem of 4 and a leaf of 2; the number 47 would have a stem of 4 and a leaf of 7.

Bibliography

Cibois, P. (2012). The interpretation of statistics in sociology. *BMS: Bulletin De Methodologie Sociologique, 114*(1), 50–58. Retrieved November 5, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=89974832&site=ehost-live

Gringeri, C., Barusch, A., & Cambron, C. (2013). Examining foundations of qualitative research: A review of social work dissertations, 2008–2010. *Journal of Social Work Education, 49*(4), 760–773. Retrieved November 5, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=90595347&site=ehost-live

Guo, J., Li, W., Li, C., & Gao, S. (2012). Standardization of interval symbolic data based on the empirical descriptive statistics. *Computational Statistics & Data Analysis, 56*(3), 602–610. Retrieved November 5, 2013, from EBSCO Online Database Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=67136175&site=ehost-live

Huff, D. (1954). *How to lie with statistics* . New York: W. W. Norton & Company.

Witte, R. S. (1980). *Statistics* . New York: Holt, Rinehart and Winston.

Suggested Reading

Feld, S. L.. (1997). Mathematics in thinking about sociology. *Sociological Forum, 12* (1), 3-9. Retrieved March 13, 2008 from EBSCO Online Database Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=ioh&AN=1537075&site=ehost-live

Gravetter, F. J. & Wallnau, L. B. (2006). *Statistics for the behavioral sciences* . Belmont, CA: Wadsworth/Thomson Learning.

Iyengar, S. (2013). Artists by the numbers: Moving from descriptive statistics to impact analyses. *Work & Occupations, 40*(4), 496–505. Retrieved November 5, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=91553862&site=ehost-live

Young, R. K. & Veldman, D. J. (1977). *Introductory statistics for the behavioral sciences* (3rd ed.). New York: Holt, Rinehart and Winston.

Essay by Ruth A. Wienclaw, PhD

Dr. Ruth A. Wienclaw holds a doctorate in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human-systems integration.

The Misuse of Statistics

Table of Contents

Abstract

Without an understanding of the purpose and limitations of statistical tools, even the most well-intentioned person can easily misuse statistics to support a conclusion that is not valid. Both descriptive and inferential statistics are open to misuse if one is not careful. However, an understanding of what various statistical tools can and cannot do, what assumptions need to be met when using them, and how to appropriately interpret the results of statistical tests can enable one to learn what questions to ask when presented with statistical findings, become a better consumer of statistical information, and be less prone to succumb to the allure of misused statistics.

Overview

By definition, science requires the application of the scientific method, in which observations of the real world are turned into testable hypotheses, data are collected and analyzed, and conclusions are drawn based on these results. Hypothesis testing and the concomitant use of statistical tools is the way that any science is advanced and theories are validated or changed. However, the presentation of graphs, charts, or numbers derived from arcane formulae alone is not enough to "prove" whether a hypothesis is correct. Unless one understands the limitations of such statistical tools and how to interpret them, it can be easy for even the most well-intentioned person to misuse statistics to support a conclusion that is not valid. At best, statistics give estimates: scientific gambles, as it were, that one's interpretation of observed behavior approximates the actual underlying causes.

Unfortunately, for many people, the use of statistics seems to throw an aura of arcane acceptability over whatever conclusion they are attached to. We are much more likely to believe a conclusion supported by charts, graphs, or numbers than we are to believe the same conclusions if they are unsupported. "Our company has a combined experience of 112 years" sounds so much more venerable than "We have lots of experience," and "80% of students fear taking a statistics course" is more scientific than "Lots of students hate statistics." But the truth is, unless we know where these numbers come from, we do not know what they really mean. The 112 years of experience may actually be the combined ages of the president, vice president, and treasurer of the organization; the 80% of students may refer to a sample drawn from a group of art majors rather than math majors.

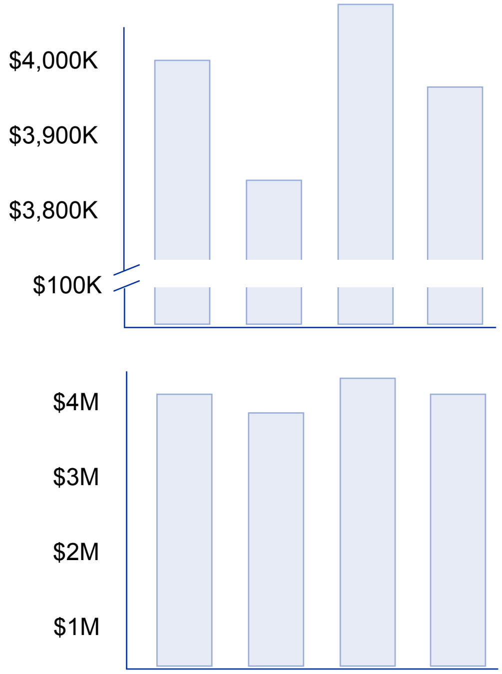
Admittedly, the proper use of inferential statistical tools requires training. However, even deceptively simple descriptive statistical techniques can be misused. In most cases, such situations arise due to a lack of understanding of the nature and limitations of the various statistical tools on the part of the person presenting the statistics. In a few cases, however, the person reporting the statistics may actually be trying to mislead the reader. Fortunately, even a little understanding about the nature of statistics can go a long way in helping one be a better informed reader of scientific reports, research studies, and even the daily newspaper. When armed with an understanding of what various statistical tools can and cannot do, what assumptions need to be met when using them, and how to appropriately interpret the results, one can learn what questions to ask when presented with statistical findings, become a better consumer of statistical information, and be less prone to succumb to the allure of misused statistics.

Applications

Misuse of Descriptive Statistics

Descriptive statistics can appear to be deceptively simple. Most people learn the basics of calculating a mean and preparing graphs and charts before they reach high school. Newspaper articles, television advertisements, and professional journals all present data summarized by descriptive statistics. However, descriptive statistics cannot be used to draw inferences about or make predictions from a sample of data. The purpose of descriptive statistical techniques is merely to organize and summarize data. Further, one must be careful about how data are displayed using graphical methods so that the data are not misrepresented. One type of misuse of statistics that is commonly seen is shown in the two graphs in Figure 1. Both graphs present the same data. However, the graph on top is designed so that it unfairly magnifies the differences in quarterly income for the four quarters, while the graph on the bottom is drawn to scale, showing that in actuality there is little difference between the quarterly earnings for the four quarters.

Figure 1: Sample Biased & Unbiased Histograms



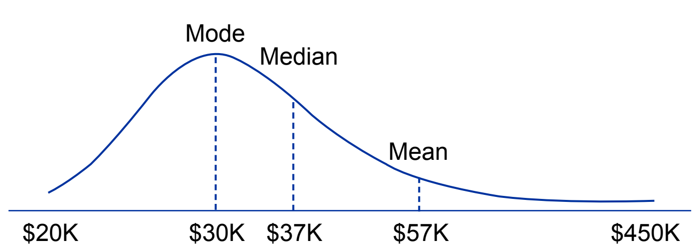
Advertisements and articles in news media and other publications frequently are illustrated with graphs and statistics from which conclusions are drawn. However, as illustrated above, these data can be misleading, due either to a poor understanding of descriptive statistics or to an intentional attempt to mislead. Therefore, one needs to take into account the type of descriptive statistics used and understand how the shape of a distribution can distort its meaning.

Descriptive Statistics

Descriptive statistics is a subset of mathematical statistics that describes and summarizes data. Included under this umbrella are various methods for summarizing and presenting data so that they are more easily understood, such as graphs, charts, distributions; measures of central tendency that estimate the midpoint of a distribution, such as mean, median, and mode; and measures of variability that summarize how widely dispersed data are in a distribution, such as range, semi-interquartile deviation, and standard deviation. These tools are deceptively simple; in truth, descriptive statistics are misused every day. For example, the three measures of central tendency are all ways to determine the "average" of a distribution of scores. It would be easy to assume that since they are all methods for finding the average, they must be interchangeable. This, however, is not true. Each different approach to determining central tendency has different characteristics from the others and is influenced by different things.

If the underlying distribution were a perfect normal distribution, these three techniques would all yield the same result. However, real-world data are messy, and underlying distributions are virtually never a perfect bell-shaped curve. Yet often only the "average" is reported, with no indication as to whether it is the mean, median, or mode, so that they reader has no idea how the measure may have been affected. For example, in a skewed distribution, where one end has extreme outliers but the data is otherwise normally distributed, the median may be pulled toward the skew (i.e., toward the end of with the outliers). Because of this, when the ends are not balanced and data are clustered toward one end of the distribution, the median may disproportionately reflect the outlying data points. On the other hand, if the extreme ends are balanced (i.e., not skewed), the median is not affected. The mean is also affected by extreme scores, and in a skewed distribution it tends to be pulled even more toward the skew than the median. These tendencies can make significant differences in the resulting values of central tendency. For example, if the mode were used to report the "average" salary for a given career and it was found that most of the people in that occupation only made $30,000 per year, it would give a different impression than if the statistic reported were the mean, which is pulled in the direction of the skew. As shown in Figure 2, the average salary for this hypothetical distribution is much closer to the mode than it is to the mean because of the small proportion of people who earn significantly more than the rest.

Figure 2: Placement of Different Measures of Central Tendency on a Skewed Distribution



(Adapted from Huff, p. 33)

These are not the only differences between these three measures of central tendency. Although the mode is quick and easy to calculate, it also has the disadvantages of lacking stability (i.e., a small change in the numbers can lead to a great change in the mode), not taking the score values into account, and not being valuable for any purpose other than to state which number has the highest frequency. The median is more stable and is the preferred measure of central tendency for use in non-symmetrical distributions because it is not as affected by extreme scores as the other two measures. The mean has advantages in most situations over the other two measures of central tendency: it is highly stable (its value does not vary greatly because of a change in one or a small handful of scores), and it is the basis of many inferential statistical techniques.

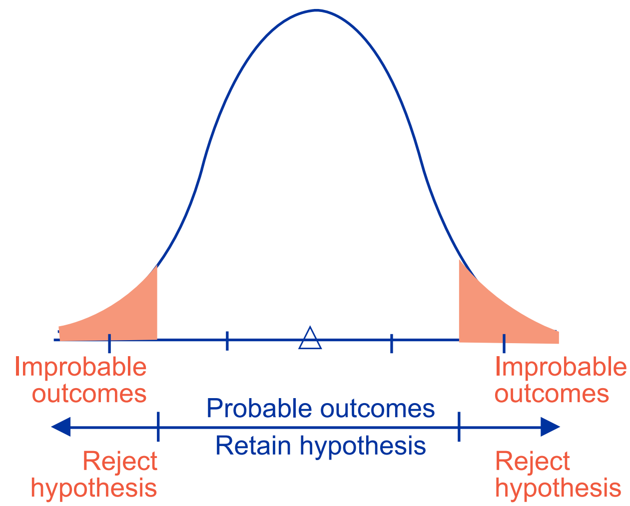
Misuse of Inferential Statistics

Inferential statistics are equally easy to misuse without a proper understanding of their limitations. Every semester, there is always at least one eager student in one of my classes who presents his or her research findings and proudly declares that the statistical results "prove" that the original hypothesis was correct. However, the truth is that statistics do not prove anything. Rather, they merely express probabilities and the degree of confidence with which one can say that the hypothesis being tested is more likely to be true than the alternative hypothesis. This fact is frequently shown in the literature, when one set of scientists attempts to replicate the research of other scientists and finds to their dismay (or, in some cases, delight) that the original results cannot be replicated.

To understand why this occurs, one needs to understand the influence of probability on statistics. In general, statistics are used to test the probability of the null hypothesis (H0) being true. The null hypothesis is the statement that there is no statistical difference between the status quo and the experimental condition. If the null hypothesis is true, then the treatment or characteristic being studied made no difference on the end result. For example, a null hypothesis might state that people are treated no differently in the workplace when they wear a business suit than they are when they wear casual clothing. The alternative hypothesis (H1) would state that the way people dress actually does have an effect the way they are treated in the workplace.

Accepting the null hypothesis means that if the data in the population are normally distributed, the results are more than likely due to chance. This is represented in Figure 3 as the unshaded portion of the distribution. By accepting the null hypothesis, the analyst concludes that it is likely that people do not react any differently to people wearing business suits than they do to those wearing casual clothing. For the null hypothesis to be rejected and the alternative hypothesis to be accepted, the results must lie in the shaded portion of the graph. This would mean that there is a statistically significant likelihood that the observed difference between the way the two groups are treated is probably due not to chance but to a real underlying difference in people's reactions to how others dress in the workplace. Statistical significance is the degree to which an observed outcome is unlikely to have occurred due merely to chance.

Figure 3: Hypothesized Sampling Distribution of the Mean Showing Areas of Acceptance & Rejection of the Null Hypothesis



(adapted from Witte, p. 118)

Another reason that statistics are sometimes misused is that not every statistical technique is appropriate for use in every situation. For example, some techniques assume that the samples that are being analyzed are not dependent, whereas other techniques do not make this assumption. A researcher needs to be careful to pick the technique that is most appropriate for the data being analyzed. In addition, sometimes researchers with less expertise in statistics prefer to use multiple simple statistical tests rather than a more comprehensive but complicated test. They may perform multiple t-tests rather than an analysis of variance, or multiple analyses of variance rather than one multivariate analysis of variance. However, one of the implications of the laws of probability is that the more tests that are run on a single set of data, the more probable it is that spuriously significant results will occur merely by chance. This approach is often referred to as "shotgunning." Conclusions drawn based on the results of such analyses are suspect at best, because this approach can compound the error inherent in the data and lead to false results.

In addition, no matter what type of inferential statistical technique is being used, one needs to look at the underlying assumptions of that technique to determine whether or not it is appropriate for what one is trying to do. Many of the inferential statistics that are commonly used (e.g., t-tests, analyses of variance, Pearson product moment correlation coefficients) are parametric and make certain assumptions about the parameters of the data being analyzed and the distribution of the underlying population from which a sample is drawn, including the assumption that the data have been randomly selected from a population with a normal distribution. Further, parametric statistics require data that are interval or ratio in nature. This means that the rank orders of the data have meaning (e.g., a value of 6 is greater than a value of 5), as do the intervals between the values. However, real-world data do not always meet these assumptions. For example, although one knows exactly what the difference is between 96 grams of a chemical compound and 95 grams of the same compound, it is less clear what the difference between a score of 96 and a score of 95 on an attitude survey means. To attempt to use parametric statistics in a nonparametric situation is to run the risk of producing misleading results.

Fortunately, in situations where data do not meet the assumptions of parametric statistics, one need not either rely on the misuse of parametric statistics or forgo statistical analysis completely. A number of nonparametric procedures are available that correspond to common tests used when the shape and parameters of a distribution are known. Nonparametric tests make no assumptions about the underlying distribution. Although they are not as powerful as standard parametric statistics, they do allow the analyst to derive meaningful information from a less-than-perfect data set.

Finally, statistics advances the state of science slowly. Not all answers can be found in one research study. For example, one statistic in particular that is frequently misinterpreted is the coefficient of correlation. The purpose of this inferential statistic is to determine the degree to which values of one variable are associated with values of another variable. For example, one could generally say with assurance that weight gain in the first year of life is positively correlated with age (i.e., the older the baby is, the more it is likely to weigh). However, this same correlation would not apply to most adults, as heavier adults are not necessarily older than lighter adults. Correlation only shows the relationship between the two variables; it does not explain why the relationship occurs or what caused it. Two events may be highly correlated but caused by a third factor. For example, two clocks that keep perfect time always chime at the same time. Neither causes the other to chime; rather, it is the movement of the mechanisms over time itself that causes the clocks to chime.

In a classic example of the misuse of correlation, Neyman once gave an illustration of the correlation between the number of storks and the number of human births in various European countries (1952). Someone not understanding how to interpret the correlation coefficient might conclude from this evidence that storks bring babies. The truth, however, was that the original calculation did not take into account the size of the countries in the data set. Larger countries tend to have both more women and more storks. The storks did not bring the babies, just as living in a larger country does not increase the probability of having a baby. The correlation was incidental, not causal.

Conclusion

The use of statistics is an important part of any science. A wide variety of techniques are available to those who desire to summarize large amounts of data or to make inferences and predictions about a larger underlying population based on observations of a sample. However, in order for the statistics to be meaningful, care must be taken to understand both their potential and their limitations. Both descriptive statistics and inferential statistics are open to abuse and misuse, with the result that the user may reach a conclusion unsupported by the data. By understanding what various statistical tools can and cannot do, what assumptions need to be met when using them, and how to appropriately interpret the results, one can learn what questions to ask when presented with statistical findings. Such knowledge helps both professionals and interested laypeople alike become better consumers of statistical information.

Terms & Concepts

Descriptive Statistics: A subset of mathematical statistics that describes and summarizes data.

Distribution: A set of numbers collected from data and their associated frequencies.

Inferential Statistics: A subset of mathematical statistics used in the analysis and interpretation of data. Inferential statistics are used to make inferences, such as drawing conclusions about a population from a sample, as well as in decision making.

Measures of Central Tendency: Descriptive statistics that are used to estimate the midpoint of a distribution. Measures of central tendency include the median (the number in the middle of the distribution), the mode (the number occurring most often in the distribution), and the mean (a mathematically derived measure in which the sum of all data in the distribution is divided by the number of data points).

Measures of Variability: Descriptive statistics that summarize how widely dispersed the data are over the distribution. Measures of variability include the range (the difference between the highest and lowest data points) and the standard deviation (a mathematically derived index of the degree to which data points differ from the mean of the distribution).

Nonparametric Statistics: A class of statistical procedures that are used when it is not possible to estimate or test the values of the parameters of the distribution or when the shape of the underlying distribution is unknown.

Normal Distribution: A continuous distribution that is symmetrical about its mean and asymptotic to the horizontal axis. The area under the normal distribution is 1. The normal distribution is also called the Gaussian distribution or the normal curve.

Null Hypothesis: The statement that the findings of the experiment will show no statistical difference between the control condition and the experimental condition.

Parametric Statistics: A class of statistical procedures that are used when it is reasonable to make certain assumptions about the underlying distribution of the data and the values to be analyzed are either interval- or ratio-level data.

Population: The entire group of subjects belonging to a certain category, such as all women between the ages of 18 and 27, all dry-cleaning businesses, or all college students.

Quartile: Any of three points that divide an ordered distribution into four equal parts, each of which contains one quarter of the data points.

Sample: A subset of a population. A random sample is a sample that is chosen at random from the larger population with the assumption that it will reflect the characteristics of the larger population.

Skewed Distribution: A distribution that is not symmetrical around the mean (i.e., there are more data points on one side of the mean than there are on the other).

Statistics: A branch of mathematics that deals with the analysis and interpretation of data. Mathematical statistics provides the theoretical underpinnings for various applied statistical disciplines, including business statistics, in which data are analyzed to find answers to quantifiable questions. Applied statistics uses these techniques to solve real-world problems.

Bibliography

Armore, S. J. (1966.) *Introduction to statistical analysis and inferences for psychology and education*. New York: John Wiley & Sons.

Hollander, M. & Wolfe, D. A. (1973). *Nonparametric statistical methods*. New York: John Wiley & Sons.

Huff, D. (1954). *How to lie with statistics*. New York: W. W. Norton & Company.

Mansell, W. (2013). Misleading the public understanding of assessment: Wilful or wrongful interpretation by government and media. *Oxford Review Of Education, 39*(1), 128–138. Retrieved November 8, 2013, from EBSCO Online Database Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=85751059&site=ehost-live

Neyman, J. (1952). Lectures and Conferences on Mathematical Statistics and Probability (2nd ed.). US Department of Agriculture: Washington DC.

Prewitt, K. (2012). When you have a hammer. . . . *Du Bois Review: Social Science Research on Race, 9*(2), 281–301. Retrieved November 8, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=84359609&site=ehost-live

Scheff, T. (2011). The catastrophe of scientism in social/behavioral science. *Contemporary Sociology, 40*(3), 264–268. Retrieved November 8, 2013, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=69671015&site=ehost-live

Witte, R. S. (1980). *Statistics*. New York: Holt, Rinehart and Winston.

Suggested Reading

Gardenier, J. (2012). Recommendations for describing statistical studies and results in general readership science and engineering journals. *Science & Engineering Ethics, 18*(4), 651–662. Retrieved November 8, 2013, from EBSCO Online Database Academic Search Complete. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=83839288&site=ehost-live

Gravetter, F. J. & Wallnau, L B. (2006). *Statistics for the behavioral sciences*. Belmont, CA: Wadsworth/Thomson Learning.

Hogben, L. (1957). *Statistical theory: The relationship of probability, credibility and error*. London: George Allen & Unwin.

Keller, D. K. (2006). The Tao of statistics: A path to understanding (with no math). Thousand Oaks, CA: Sage Publications.

Young, R. K. & Veldman, D. J. (1977). *Introductory statistics for the behavioral sciences* (3rd ed.). New York: Holt, Rinehart and Winston.

Essay by Ruth A. Wienclaw, PhD

Dr. Ruth A. Wienclaw holds a PhD in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human-systems integration.

Research Ethics in Sociology

Table of Contents

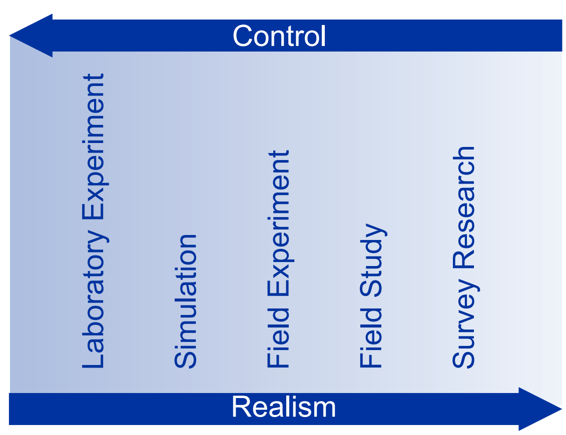
Abstract

The planning and conducting of research that will aid in the advancement of any behavioral or social science requires the choice of a research paradigm that appropriately balances the researcher's ability to control the research setting while maintaining an adequate and accurate representation of the complexity of the real world situation. When working with human subjects, the researcher is further required to take into account various ethical considerations to ensure that no physical or psychological harm occurs to the subjects as well as that the data are as valid as possible. For this purpose, professional codes of ethics have been developed to guide researchers in the ethical conduct of experiments and studies. In particular, researchers need to guarantee as much as possible the confidentiality of subjects' private information, acquire informed consent from research participants, and take all precautions possible during research planning, implementation, and dissemination to adequately and accurately present their research findings.

Overview

The conduct of the scientific research, by which sociology and other behavioral and social sciences advances, can be both challenging and rewarding. The antecedents of human behavior that we see all around us are complicated and interrelated, requiring the creative application of the scientific method in order to gather data to better understand and predict human behavior. There are a number of research tools available to social and behavioral scientists which supports this task. As shown in Figure 1, these research paradigms offer scientists various degrees of control over the research situation and the degree to which the research situation realistically reflects the complexity of the real world.

Figure 1: Research Paradigms for Primary Analysis



Laboratory Experiments

Laboratory experiments allow researchers the most control not only over the level of the independent variable that is experienced by the subjects, but also over the various extraneous variables that can erroneously affect the outcome of the study. For example, if one wanted to determine the relationship between how an individual dresses at work and how that person is treated by others, one might set up a simple laboratory experiment in which pictures of individuals in various types of attire (e.g., dark business suit, business casual, casual) are presented to subjects who are then asked to rate the professionalism of the person in the picture. This approach to collecting data to test the research hypothesis gives the researcher a great deal of control over the experimental situation (e.g., how the people in the picture are dressed). However, rating the "professionalism" of people portrayed in photographs is far removed from real world settings, so the results would not be widely applicable. If the researcher is willing to give up some control over the experimental variables, s/he would be able to design an experimental condition with more realism.

Simulations

A simulation could be set up in which subjects interact with experimental confederates who dress in various types of attire as specified by the experimenter. This research paradigm still offers the researcher a great deal of control over the experimental situation (e.g., she can specify exactly how the confederates will dress), but its increased realism concomitantly gives the researcher less control (e.g., extraneous variables such as the way the confederates talk, their attitude, and other variables not related to the research hypothesis can affect the response of the subjects). Further, although a simulation is more realistic than a laboratory experiment, it still only remotely emulates the real world situation.

Field Experiments

Giving up a little more control in favor of a higher degree of realism, the researcher could conduct a field experiment in which confederates interacted with the subjects in a real-world business setting. However, this situation would allow for the greater possibility of the influence of extraneous variables than the more controlled simulation and laboratory experimental paradigms. In some respects, this can be both an advantage and disadvantage. Although one's attire in the workplace has been shown to affect the way that one is treated, the way one is treated also depends on many other variables as well (e.g., behavior, grooming, attitudes of the other person, competence). The complexity of these variables can be better seen in field settings than in more controlled paradigms.

Real world situations tend to be very complex, however, particularly when one is trying to determine what variables affect human behavior. In many cases, it would be virtually impossible for a researcher to sufficiently articulate all the real world variables that influence behavior in a way that would allow a hypothesis to be empirically tested using inferential statistics. Statistical tools are available for modeling real-world behavior, but these typically require the collection of vast amounts of data from real-world observation. For such tasks or for the purposes of collecting individual observations for the application of inductive reasoning, more realistic research paradigms are needed. For example, although a researcher might be able to use a more controlled research paradigm to collect data on various levels of the dependent variable (e.g., business attire, business casual, and casual dress), in truth there are virtually infinite combinations of the ways that people can dress at work. Is a dark suit more impressive than light suit? If so, does the suit need to be black, or would dark gray or navy blue be just as impressive? Does the suit need to be plain or do pinstripes add to the professional aura? The list of permutations on just this one level of attire is seemingly endless. Similarly, how does one best define the way that a person is "treated at work"? Once again, in the real world there are seemingly endless ways in which this can be defined ranging from the politeness or friendliness with which they are treated by peers, supervisors, and customers to the hard data of number and frequency of promotions, amount and frequency of raises and bonuses, scores on performance appraisals, just to name a few measures.

Field Studies

Field studies are examinations of how people behave in the real world. For example, a researcher might either directly or unobtrusively observe how various people are treated in the workplace, recording his/her observations both on the treatment received as well as the person's attire. Another approach would be to employ the paradigm of survey research. Subjects could be interviewed by a member of the research team or asked to fill out a questionnaire regarding the way that they typically dress at work and how they perceive the treatment they receive from peers, supervisors, and customers. This could be combined with other information such as the amount and frequency of raises, bonuses, or promotions.

Survey Research

Theoretically, survey research allows researchers to gather the most information about the situation under investigation. However, although a very thorough interview or survey instrument can be written that would hypothetically gather all the data needed for the researcher to make decisions about the antecedents of treatment in the workplace, such instruments are often more lengthy than the potential research subject's attention span. Further, as opposed to the other research techniques, surveys and interviews are not based on observation. Therefore, there is no way to know whether or not the information being gathered from the subject is true. As a result, information gathered from research paradigms more to the right of the continuum in Figure 1 tend to be difficult to empirically test and determine the validity of the underlying hypothesis. So, the behavioral researcher is left with a dilemma.

The Middle Road

At first glance, it might seem that the best research paradigms lie somewhere in the middle of the continuum shown in Figure 1. Paradigms in the middle of the continuum still allow the researcher a good deal of control over the experimental situation while allowing for more widely applicable results due to the increased realism of the research situation. This tempting rule of thumb, however, is weakened by the fact that the ultimate goal of most researchers is to do research that adequately and accurately reflects the real world situation (i.e., is highly realistic) so that it can be extrapolated and used in predictions. It also allows a great deal of control over the variables so that the results can be statistically analyzed and the probabilities of the accuracy of the hypothesis can be estimated. However, the complexity of real world behavior places obvious limitations on the degree to which this can be accomplished. Further, when dealing with human subjects, there is another set of limitations that restrict the degree to which variables can be eliminated: Research ethics.

Research Ethics

In scientific research, the term *ethics* refers to a code of moral conduct regarding the treatment of research subjects that is subscribed to by the members of a professional community. Many professional groups had a specific written code of ethics that sets standards and principles for professional conduct and the treatment of research subjects. When dealing with human subjects, the researcher obviously wants first to do no harm, either physically or psychologically. However, the manipulation of variables often brings with it the potential for harming the subject. Further, research ethics are not relevant only to controlled experiments. Often, merely asking the question such as is done in survey research can bring up bad memories and unresolved psychological difficulties that can be harmful to the subject (e.g., questions about personal experiences in childhood or spousal abuse).

Applications

The American Sociological Association (ASA) has developed a code of ethics and policies and procedures for the professional and ethical conduct of research with humans. The general guiding principles stress "professional competence, integrity, professional and scientific responsibility, respect for people's rights, dignity, and diversity, and social responsibility." Three specific sections of the code of ethics deal with the specific ethical responsibilities of researchers performing sociological research: confidentiality, informed consent, and research planning, implementation, and dissemination.

Confidentiality

Frequently, the information collected by sociologists is sensitive in nature, particularly when that information can be traced back to the individual who provided it. Most people are loath to reveal personal information about themselves, their attitudes and feelings, or their family situation if they think there is a possibility of that information being made public and associated with them specifically. Virtually every person has personal information that s/he prefers to keep private for any number of reasons. Sometimes, this is a personal preference. Often, however, an individual's desire for confidentiality of personal information stems from a fear of harm should that information become public. Further, from a research point of view, it is important to maintain confidentiality in order to help ensure that the information obtained from research subjects is complete and accurate. Otherwise, it is quite likely that the data obtained will be tainted due to the very understandable desire on the part of the research subject to keep private information private or even to lie in order to appear better.

To this end, it is essential that sociologists and other behavioral and social science researchers take all reasonable precautions necessary in order to ensure the confidentiality of all subjects and organizations participating in their research studies. According to ASA guidelines, this confidentiality extends even after the death of the participant. This responsibility applies not only to the primary researcher, but to all members of the research team who have access to confidential information whether that information was collected as part of the current research study or involves the analysis of confidential information previously collected in other settings.

There are a number of ways that confidentiality of data can be ensured in many cases. One frequently used method for ensuring confidentiality is to remove any demographic data that specifically links confidential information to a particular individual from the database. For example, a researcher might not include the names of individuals or their specific addresses in the database. Instead, a unique identification number could be used to link all the data for a particular individual so that it could be analyzed without being traceable back to that individual.

However, in some situations, it is not possible to ensure confidentiality in this manner. In such situations, the researcher needs to discuss the relative limitations of confidentiality and foreseeable uses of the confidential information with the participants in the study. In such situations, it is also important for the researchers to take reasonable precautions to determine that they have obtained the informed consent of identifiable individuals before they transfer data to others or review the data collected by others. The twenty-first century has brought with it new technological means for storing and transferring data. From the perspective of confidentiality, this has both advantages and disadvantages. Researchers need to exercise extreme caution when transferring confidential data over public communication networks as well as maintaining the confidentiality of any audio or video recording by which a research subject may be identified.

Informed Consent

Another important ethical principle for researchers investigating human behavior is the concept of informed consent. Informed consent is an agreement between a participant and researcher or practitioner that discloses the nature of the procedure, its potential benefits, possible risks, and alternatives of the experimental or therapeutic procedure. An informed consent document is voluntarily signed by the participant before the conduct of the research or therapy. As a general principle, sociologists should never involve a human subject in research without his/her informed consent.

In some cases, researchers may not be ethically required to obtain informed consent (e.g., cases when there is minimal risk for the research participants or when the acquisition of informed consent would prevent the research from being carried out). Further, sociologists can "ethically conduct research in public places or use publicly available information" without first obtaining informed consent (ASA, 1999). If, however, the situation is questionable, the sociologist or researcher should consult with the research review board at his/her institution or another authoritative group that has expertise on research ethics. Obtaining informed consent is particularly important when working with vulnerable individuals such as children or juveniles, recent immigrants, the mentally disabled, or the mentally ill. When working with such populations, it is vital that the researcher take all reasonable precautions to ensure that participation in the research study is done consensually and is not coerced.

Informed consent documents need to be crafted using "language that is both understandable and respectful to the research subjects or the legal representatives" (ASA, 1999). Informed consent documents should reveal the nature of the research, the fact that participation or continued participation is voluntary, possible risks, benefits, while there are significant factors that may influence the subject's willingness to participate, as well as how confidential information will be handled. It is important that signed consent documents be kept on file.

Sometimes, however, the nature of the scientific inquiry requires the use of deception when dealing with the subjects (e.g., in situations where knowledge of the true intent of the research might affect the subject's response). Deception should not be used unless its use is justified by the potential value of the study and unless all equally valid alternative methods are available have been considered. Further, before using deceptive methods in research, the researchers should submit the research assigned for approval to the appropriate institutional research review board or other authoritative research group. Researchers should not deceive potential subjects with regard to factors that might negatively impact their willingness to participate in the study (e.g., physical risks, discomfort, negative emotional experiences). When it is necessary to use deception as part of the experimental design, the researchers should debrief the subjects as soon as possible after the end of the study and correct any misconceptions on the part of the subjects at that time.

Research Planning, Implementation, & Dissemination

The purpose of sociological research is to advance the state of the art in understanding and predicting human behavior in social groups. In order for research to accomplish this task, researchers need to rigorously apply the principles and procedures of the scientific method in the design, conduct, analysis, and interpretation of the study and its results. During the planning phase, this means the researchers design the study so that the results will be unambiguous and not misleading. When dealing with special populations (e.g., children, individuals with developmental disabilities), researchers should consult with experts in other areas of expertise as necessary to better understand how to ethically deal with the potential subjects. Although from time to time it may be appropriate to offer a financial token or other inducement to potential research subjects, this should only be done where necessary and only to the extent that will encourage participation, but not introduce the possibility of changing the results (e.g., inducing false results to "please" the researcher who paid the subject).

When reporting the results of their research studies, it is essential that researchers only report the actual results of the study and not use or report fabricated data or falsified results. Further, it is important that all results of a research study be reported, including any information or results that may contradict the expected outcomes of the hypothesis or other findings in the study. To better help others interpret the results of their study, researchers should disseminate relevant caveats on the results and conclusions of their research. Extraneous variables and potential factors that may have influenced the results in either the current research situation or that might affect the ability to replicate the results in other studies should be fully discussed. Any assumptions, theories, methods, measures, and research designs should be fully disclosed. To aid in the advancement of the state-of-the-art, researchers should also permit other professionals to examine their data and analysis as long as the confidentiality of the subjects is maintained. Finally, if the researcher later finds significant errors in their presentation, analysis, or interpretation of the research data, s/he should take all reasonable steps to correct these errors such as publishing a correction, retraction, errata, or through other appropriate communication (ASA, 1999).

Conclusion

It is through the application of the scientific method and the conduct of scientific research that the state-of-the-art in sociology or any other behavioral or social science advances. When working with human subjects, the researcher needs to find a research paradigm with the appropriate balance between allowing control of the research variables and one that adequately and accurately reflects the reality of the real world and all its myriad details. This complex decision is further complicated by the fact that researchers using human subjects in their studies must also adhere to professional codes of ethics for the conduct of human research. If this is not done, not only can the data become tainted, but research subjects may be physically or psychologically harmed as a result of the experimental intervention. In particular, the American Sociological Association has articulated guidelines for the ethical conduct of research through confidentiality, informed consent, and precautions to be taken during research planning, implementation, and dissemination. By adhering to this code of ethics and by taking all reasonable steps to ensure the ethical treatment of research subjects, sociologists can help ensure that they not only do no harm to their subjects ,but also that they collect data that will aid in the advancement of their science.

Terms & Concepts

Confederate: A person who assists a researcher by pretending to be part of the experimental situation while actually only playing a rehearsed part meant to stimulate a response from the research subject.

Demographic Data: Statistical information about a given subset of the human population such as persons living in a particular area, shopping at an area mall, or subscribing to a local newspaper. Demographic data might include such information as age, gender, or income distribution.

Ethics: In scientific research, a code of moral conduct regarding the treatment of research subjects that is subscribed to by the members of a professional community. Many professional groups had a specific written code of ethics that sets standards and principles for professional conduct and the treatment of research subjects.

Experiment: A situation under the control of a researcher in which an experimental condition (independent variable) is manipulated and the effect on the experimental subject (dependent variable) is measured. Most experiments are designed using the principles of the scientific method and are statistically analyzed to determine whether or not the results are statistically significant.

Hypothesis: An empirically testable declaration that certain variables and their corresponding measure are related in a specific way proposed by a theory.

Inductive Reasoning: A type of logical reasoning in which inferences and general principles are drawn from specific observations or cases. Inductive reasoning is a foundation of the scientific method and enables the development of testable hypotheses from particular facts and observations.

Informed Consent: An agreement between a participant and researcher or practitioner that discloses the nature of the procedure, its potential benefits, possible risks, and alternatives of the experimental or therapeutic procedure. The informed consent document is voluntarily signed by the participant before the conduct of the research or therapy.

Model: A representation of a situation, system, or subsystem. Conceptual models are mental images that describe the situation or system. Mathematical or computer models are mathematical representations of the system or situation being studied.

Scientific Method: General procedures, guidelines, assumptions, and attitudes required for the organized and systematic collection, analysis, interpretation, and verification of data that can be verified and reproduced. The goal of the scientific method is to articulate or modify the laws and principles of a science. Steps in the scientific method include problem definition based on observation and review of the literature, formulation of a testable hypothesis, selection of a research design, data collection and analysis, extrapolation of conclusions, and development of ideas for further research in the area.

Subject: A participant in a research study or experiment whose responses are observed, recorded, and analyzed.

Survey: (a) A data collection instrument used to acquire information on the opinions, attitudes, or reactions of people; (b) a research study in which members of a selected sample are asked questions concerning their opinions, attitudes, or reactions, gathered using a survey instrument or questionnaire for purposes of scientific analysis; typically the results of this analysis are used to extrapolate the findings from the sample to the underlying population; (c) to conduct a survey on a sample.

Survey Research: A type of research in which data about the opinions, attitudes, or reactions of the members of a sample are gathered using a survey instrument. The phases of survey research are goal setting, planning, implementation, evaluation, and feedback. As opposed to experimental research, survey research does not allow for the manipulation of an independent variable.

Variable: An object in a research study that can have more than one value. Independent variables are stimuli that are manipulated in order to determine their effect on the dependent variables (response). Extraneous variables are variables that affect the response, but that are not related to the question under investigation in the study.

Bibliography

American Sociological Association. (1999). Code of ethics and policies and procedures of the ASA Committee on Professional Ethics. Washington, DC: American Sociological Association.

Shaw, R. (2011). The ethical risks of curtailing emotion in social science research: The case of organ transfer. *Health Sociology Review, 20*(1), 58-69. Retrieved November 6, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=60728650 Spicker, P. (2011). Ethical covert research. *Sociology, 45*(1), 118-133. Retrieved November 6, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=58014070

Swauger, M. (2011). Afterword: The ethics of risk, power, and representation. *Qualitative Sociology, 34*(3), 497-502. Retrieved November 6, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=63995198

Suggested Reading

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. Washington, DC: American Psychological Association. Retrieved April 15, 2008, from http://www.apa.org/ethics/code2002.html

Bridges, D. (2007). Research ethics, academic virtue and the practice of higher education. Social Sciences, 56(2), 7-13. Retrieved April 15, 2008, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=25975088&site=ehost-live

Cordner, A., & Brown, P. (2013). Moments of uncertainty: Ethical considerations and emerging contaminants. *Sociological Forum, 28*(3), 469-494. Retrieved November 6, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=89888318

Rivera, R., Borasky, D., Rice, R., Carayon, F., & Wong, E. (2007). Informed consent: An international researchers' perspective. American Journal of Public Health, 97(1), 25-30. Retrieved April 15, 2008, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=23660822&site=ehost-live

Spicker, P. (2007). Research without consent. Social Research Update, (51), 1-4. Retrieved April 15, 2008, from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=28139181&site=ehost-live

Wilkinson, S., & Kitzinger, C. (2013). Representing our own experience: Issues in “insider” research. *Psychology Of Women Quarterly, 37*(2), 251-255. Retrieved November 6, 2013 from EBSCO Online Database SocINDEX with Full Text. http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=87656771

Essay by Ruth A. Wienclaw, PhD

Dr. Ruth A. Wienclaw holds a doctorate in industrial/organizational psychology with a specialization in organization development from the University of Memphis. She is the owner of a small business that works with organizations in both the public and private sectors, consulting on matters of strategic planning, training, and human/systems integration.