Suicide Studies, 2022, 3(3), 1-40
ISSN 2771-3415

RESEARCH METHODS FOR SUICIDOLOGISTS

David Lester

Contents

Detailed Contents

CHAPTER 1: WHY WE CONDUCT RESEARCH

Many people do not see any usefulness in research. Some view it as a waste of time and effort while others are merely uninterested in it. In this first chapter, therefore, we want to suggest some reasons why it is useful, and perhaps important, to carry out research.

Let us begin with an example. I have frequently read and heard in the media that depression is especially common on the major national holidays, such as Christmas. As a suicidologist, this raised the question in my mind of whether suicide is more common on the major national holidays than at other times.

In order to examine this possibility. I conduct research. Research is a logical and systematic sequence of steps carried out in order to judge the probability that an assertion is correct. The words I used here were carefully chosen and each is important.

Because the steps are logical, it is more likely that my conclusion will be acceptedby others. Of course, if I do not care what others think, then research serves no purpose. But as long as we are interested in communicating with one another (and, after all, the fact that I wrote this book and that you are reading it indicates that both of us are interested in communicating), then it is useful if we conduct our research and report it ina logical manner.

The fact that the steps in research are both logical and systematic means that others can follow the procedures which we took and, if they choose to do so, repeat the piece of research we reported. Repeating a piece of research is called replication, and replication is important since sometimes the results of research conducted by one investigator cannot be replicated by others working independently. Only when research results have been replicated by others can we place confidence in their conclusions.

I said above that the research helps us to judge the probability that an assertion is correct. Because of the variability of human behavior and because of the logic behind the methods of research, we can never prove or disprove an assertion. All that we can do is say that it is likely to be correct or likely to be incorrect. We can usually specify precisely how likely the assertion is to be correct or incorrect. Later we will see that convention permits us to claim that assertions are correct when there is less than a one in twenty probability that we are in error in making this claim. To be put it more positively, if the results of my research permit me to claim that suicide rates are different on the major national holidays, then ordinarily there will be at least a 95 percent probability that I am correct in concluding this from my research.

As I noted at the beginning, some individuals are not concerned with verifying or disconfirming assertions. For example, some counselors and psychotherapists are not interested in showing to others that counseling helps clients. They claim to know that it does. In these cases we are dealing with beliefs. There is no problem in believing something. But you must not expect that others will share your belief or be converted by your belief. Your audience is likely to ask, "What evidence do you have for what you are saying?" or "Prove it!"

Decisions

Research requires us to make many decisions. Let us return to my initial example about whether suicide rates are different on the major national holidays. Let us assume that we want to see whether this is possible or not. Carrying out the research requires us to make many decisions.

(1) From where can we obtain a sample of suicides whose date of death is known? A sample is a selection of people from the total population of the world from whom to collect data. There are many ways to do this, some sounder than others, and we must be aware of these different ways.

(2) How many suicides should we study? The decision about the size of our sample is made easier if we have some knowledge of statistics. A statistic is a number or score of a person on some measure. Your age is a statistic, as is your sex, your religion, or your intelligence test score. Statistics, therefore, are a set of scores.

Statistics also refers to the methods for describing these scores to others and for drawing inferences from the scores. If I find that there were 531 suicides on New Year'sDay from 1972 to 1978, as compared to 453 on January 8th and 387 on December 25th (Lester, 1985), then I have described the scores to you, and you can obtain some idea ofthe degree of the difference. In order to conclude that suicide is less common on Christmas Day, I have to draw an inference from the scores. As I have noted above, thisconclusion will have an associated probability, that is, I may be able to conclude that I am 99 percent certain that suicide is less common on Christmas Day.

(3) How shall I obtain the dates of the suicidal act? Will I rely on published accounts? Published by whom - the government or newspapers? Will I examine medical examiner records so that I will be able to distinguish between the date of the suicidal act and the date of death which may not always coincide?

All of these questions concern the procedure followed by the investigator. Investigators do not try out every possible way of conducting the study. However, it is important that investigators report exactly how they carried out their research. Then others can choose to replicate their procedure exactly or choose to alter some feature of the procedure.

(4) What do you do with all of the scores? How do you summarize them, describe them and draw inferences from them? This requires some knowledge of statistics. These days, there are a variety of calculators and computers with associated software packages to help you do this. You can even pay a statistical consultant to assist you. But these aides still require that you have some knowledge of statistics, for you have to tell your computer or your consultant what it is you want done in order for them to do it.

(5) Finally, you have to write up a report of your investigation. There are standard ways of writing a report so that it contains all of the information that others require in order to judge the soundness of your research and the validity of your conclusions. Reports can remain unpublished and simply circulated to friends and colleagues, or they can be published. Publication requires that your report conforms to one of the many different styles that magazines (popular accounts of research) and journals (more formal scholarly publications) have adopted.

Judging the Research of Others

A knowledge of research methods also permits you to evaluate the research and conclusions of others. Again, some people are content to let experts tell them what is correct and what is incorrect. Others are skeptics. If you are a skeptic, and I certainly am one, then you rarely believe an assertion until you can examine the evidence that the expert has produced to convince you. My knowledge of research methods enables me to judge whether the evidence presented to me comes from well-designed research or poorly designed research.

Let me give you some examples. First, here is an example from statistics. Notice the next time your national television news (or newspaper or magazine) tells you that 65 percent of those polled supported the decision made by the President of the United States. They will often have a note on the screen (or in a footnote to the table in the newspaper) that they polled 1,000 people and the error is "plus or minus 3%." What does that mean? With some knowledge of statistics, you will know that this means that the pollsters reckon that the percentage of people in the nation supporting the President's decision is 95 percent likely to be somewhere between 62 percent and 68 percent. They cannot be sure exactly what the percentage is in the nation as whole, but it 95 percent likely to be a majority. This also means that there is a 5 percent probability that what they have just told you is incorrect.

Now here is an example of research with a methodological flaw. Several studies have found that people who kill themselves are more likely to have experienced the loss of a parent during their childhood. Indeed, I have published such a study (Lester, 1989). But there is a great deal of research which indicates that those who kill themselves are more depressed than those who do not commit suicide. Thus, the research on the childhood experience of loss must compare those who kill themselves with equally depressed individuals who have not killed themselves. Without this matching for depression, we can conclude only that loss of a parent in childhood leads to subsequent depression. However, if we match the subjects of our research for the intensity of their depression, then we can conclude that loss of a parent in childhood increases the risk of subsequent suicide.

Some assertions made by experts have important implications for us. Smoking cigarettes causes lung cancer. Vitamin C in large doses prevents colds. Intelligence is innate, determined by our genes. The world is flat.

To judge the soundness of these assertions we need to know what research was conducted, and we need to be able to evaluate this research. If we do not have access to the research or if we are unable to evaluate it, then we have to rely on the conclusions of others.

Conclusions

In this first chapter I have presented two rationales for being interested in research. First, perhaps you have a curiosity about the world, or at least certain featuresof the world, to the extent that you want to test the validity of particular assertions. This is certainly true for me. I am a psychologist and, often when I read some statement, I am motivated to see if I can test its validity through research.

Second, perhaps you are a skeptic with a tendency to doubt what others assert. Even if you do not take some time to conduct research into the topic, perhaps you want some knowledge which you can apply to judge the adequacy of the research that others have conducted?

Let me end this chapter with a brief report of a simple study on whether people who are depressed and suicidal show other self-defeating behaviors, such as choosing people and situations that lead to disappointment or failure or failing to accomplish tasks crucial to personal objectives despite demonstrated ability to do so. Read the following report and ask yourself:
is the evidence convincing?
is the report of the research clear enough?
how else might the study have been conducted?
can you come up with ideas for future studies on the topic?

Psychological Reports, 1992, 70, 1006.

SELF-DEFEATING BEHAVIOR, DEPRESSION, AND SUICIDAL PREOCCUPATION

DAVID LESTER AND SUZANNE HOFFMAN

Summary. - Scores on a measure of self-defeating personality were associated with depression and suicidal preoccupation in a sample of 58 college students.

To explore whether scores on a measure of self-defeating personality (Schill, 1990) were related to depression and suicidal preoccupation, a questionnaire was administered anonymously to 24 male and 34 female college students aged 18 to 23 (M_{age} = 20.1 yr., SD = 1.8) and enrolled in social science courses. The questionnaire contained Schill's measure, the Beck Depression Inventory (Beck, et a1., 1961), and a question asking whether the respondent had even thought about committing suicide. Mean scores were 19.1 (SD = 7.0) on the measure of self-defeating personality and 8.1 (SD — 7.1) on depression.

Self-defeating personality scores were correlated with depression scores (Pearson r = 0.55, one-tailed $R < .001$), current suicidal ideation (r = 0.58), and to ever having thought about suicide (r = 0.53). Using partial correlation coefficients to control for depression, sex, and age, self-defeating personality scores were still associated with current and prior suicidal ideation (rs — 0.43 and 0.33).

Self-defeating personality appears to be associated with both depression and suicidal preoccupation, a result consistent with an earlier study showing an association

between a measure of self-destructiveness (Kelley, et al., 1985) and depression and suicidal ideation (Lester and Gatto, 1989). Together these studies suggest that depression and suicidal ideation are associated with a general tendency to neglect the care of oneself.

REFERENCES

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. An inventory for measuring depression. *Archives of General Psychfatry*, 1961, 4, 561-571.

Kelley, K., Byrne, D., Przybyla, D., Eberly, C., Eberly, B., Greenlinger, V., Wan, C., & Gorsky, J. Chronic self-destructiveness. *Motivation and Emotion*, 1985, 9, 135-151.

Lester, D., & Gatto, J. L. Self-destructive tendencies and depression as predictors of suicidal ideation in teenagers. *Journal ofAdolescence*, 1989, 12, 221-223.

Schill, T. A measure of self-defeating personality. *Psychological Reports*, 1990, 66, 1343-1346.

I will return to comment on this study in the following chapters as I cover each step in the process of conducting research.

CHAPTER 2: THE HYPOTHESIS

Before beginning a research study, it is necessary to have a hypothesis. A hypothesis is a tentative assumption about the behavior you are planning to study. For example, in my brief research report presented in Chapter 1, the hypothesis was that people who are suicidal are also generally self-defeating. In this chapter, we will consider some of the sources for hypotheses and the best ways in which to formulate them.

Sources of Hypotheses

Informal Sources

Hypotheses can stem from a variety of sources. You might read something in a novel or a newspaper that suggests an hypothesis to you. You might have a personal experience which you think might be true for others in general. You may be having a conversation with others, and an idea is thrown out which intrigues you. Let me give some examples.

I like detective stories, but I am also interested in suicidal behavior. I came acrossthe following passage in a novel by John MacDonald (1968):

With the jumpers and drowners, McGee, you don't pick up a pattern. That's because a jumper damned near always makes it the first time, and a drowned is usually almost as successful.....So the patterns mostly come from the bleeders and the pill-takers and the shooters....Like the bleeders cut themselves again, and the pill-takers keep trying. It's always patterns. Never change. They pick the way that they want to go and keep after it until they make it. A pill-taker doesn't turn into a jumper, and a drowner won't shoot himself. (MacDonald, 1968,p. 102)

MacDonald, a novelist, is saying that people who make repeated attempts to kill themselves choose and stay with one method. This generates the hypothesis that people, on the whole, use the same method for suicide in successive attempts more often than they switch.

I was motivated to collect some data on this (Lester, et a1., 1989). Later at a conference on suicide, a colleague suggested that those who switch methods for suicide might be more intent on dying than those who stay with one method. I hope to pursue thisby comparing those who switch methods and those who do not in a future study.

To take another example, there is a widespread set of beliefs about the personalityof short people as compared to tall people. One of my students, a police officer, was convinced that short police officers behave differently from (and less professionally than)tall police officers. The preliminary study we did to explore this found no differences (Sheehan and Lester, 1980), but in a second study we did find that the belief that short police officers behave differently from tall police officers was widespread (Lester and Sheehan, 1980).

Formal Sources

The majority of hypotheses come from more formal sources. Most research derives from research conducted earlier by others. The simplest study based on earlier research is to simply replicate the study, exactly as the original investigators carried out the research, in order to explore whether their findings are replicable. However, their research may suggest alternative ways to test the hypothesis which guided the original study and may stimulate the development of new hypotheses for further research.

For example, in my study of the association between suicidal ideation and self-defeating behavior, I based my research on my previous study of the association between suicidal ideation and self-destructive behavior and on Menninger's (1938) theory of the relevance of the suicidal impulse to all kinds of self-defeating behavior.

Basing hypotheses on earlier research involves searching out this earlier research. Often it is possible to find reviews of the earlier research. For example, I published a review of the research on suicide (Lester, 1992), and this book can serve as a handy reference for those interested in research on suicide.

But how did I locate the research which I reviewed? There are a variety of reference sources which help us to locate research.

(1) Psychological Abstracts, which appears monthly now with an annual index, locates and publishes a brief abstract on most psychological research and is a good sourceof research on suicide. These abstracts are now available for search by computer, and most reference librarians can help you locate research articles.

(2) Other major reference sources for social and behavioral science research on suicide are Sociological Abstracts, Biological Abstracts, and Index Medicus, and there are now many highly specialized abstracts.

(3) There are journals and magazines which focus on specialized topics. For example, Suicide and Life-Threatening Behavior clearly could be a major source of hypotheses on suicide.

(4) More popular articles on suicide and other topics, primarily from magazines rather than journals, can be located through the Reader's Guide to the Periodical Literature.

A good book on how to utilize a reference library is by Beasley (1988).

The Advantage of Theory

Let us suppose that your mother said that you should never trust a person whose eyes are too close together. You plan and carry out a study in which you administer a psychological test of trustworthiness (which luckily another psychologist has already devised for you) to one hundred people, and you measure the distances between the centers of the pupils of their two eyes. You find no association at all.

The problem with this study is that other people may not be interested in your result. They may, in fact, ridicule you for bothering to test a hypothesis derived from your mother.

(Of course, had you confirmed the hypothesis, and others replicated your results, then you might become famous.)

The advantage of basing your hypothesis on a well-established theory is that even a failure to confirm the hypothesis is of interest. For example, Sigmund Freud's psychoanalytic theory predicts that trauma occurring in the first six years of life has moreof an impact on your subsequent life than trauma occurring later. This accounts for why my study on the experience of loss in suicides which I referred to earlier (in Chapter 1) ashaving a methodological flaw (a failure to take into account the severe depression of suicides) was published. Almost all of the suicides in my sample had lost a parent between the ages of six and fourteen, which was not expected on the basis of Freudian theory.

In suicide research, therefore, it is safer to base research on one of the more established theories or on published research for then the results of your research will be of more interest to others than if you base your hypotheses on your own idiosyncratic theory. (Of course, even the great theories were originally believed in by only one person, the one who proposed it.)

So to sum up, it is safer at first to base your hypotheses on the research of others and well-established theories. At least you can sharpen your research skills in this way before you branch out and test your own newly proposed and highly creative theories.

Having said this, of course, often researchers come across a topic or, even more likely, a source of data, for which it is not possible to formulate a good hypothesis. For example, I found out that the Society for American Baseball Research keeps incredibly accurate records of all of the major league baseball players who had at least one at-bat or threw one pitch. They also track down the causes of death of every player. I did a study once on whether the cause of death was related to major league performances. Did batterswho subsequently died from murder have different batting averages from batters who later committed suicide (Lester and Topp, 1989)?

There was no theory behind the study, and I had no expectation of what the result might be. It was simply that I located a source of data, and I was interested to explore the associations. It was what I call a "fishing-expedition." Fishing expeditions are fine, but you must not expect other researchers to perceive such research as good.

Specificity

It is important to make a formal statement of your hypothesis. Write it down. It is all right to make your first draft of it general, but eventually you must be very specific in stating it.

For example, we could start with an hypothesis that short people differ from tall people in their personality. Then we must consider how to define short and tall. Next, what personality traits are the critical ones? Extraversion or assertiveness? If we decide on assertiveness, how shall we define assertiveness? Our final, highly specific, hypothesis might, therefore, read something like this:

Men who are among the shortest five percent of the population will obtain higher scores on Smith's test of assertiveness than men who are among the tallest five percent of the population.

or

In general, the shorter the police officer, the more citizen complaints per year willbe made about him.

There must be no doubt here. To make your research acceptable to others, your hypothesis must eventually be stated with great specificity.

Conclusions

Research must be guided by hypotheses. Hypotheses can be stimulated by both informal and formal sources, but to be useful in guiding the design of your research studythey must be stated with great specificity. Hypotheses based on formal sources and, in particular, upon well-accepted theories, will generate research whose results will be of interest whether you confirm the hypothesis or not.

In general, the shorter the police officer, the more citizen complaints per year willbe made about him.

There must be no doubt here. To make your research acceptable to others, your hypothesis must eventually be stated with great specificity.

CHAPTER 3: VARIABLES AND THEIR MEASUREMENT

A variable is simply a concept that can be measured. Since hypotheses relate concepts to one another, hypotheses propose relationships between variables. In the example in Chapter 1 in which the relationship between suicidal ideation and self- defeating behavior was examined, there were two variables -- the score of the subjects on a test of self-defeating behavior and their answer to a question about prior suicidal ideation.

The critical question about variables which must be answered before we can conduct research into them is how to measure them. The operational definition of a variable is the sequence of steps which we take in order to measure the variable and assign it some score. There are several types of measurement (sometimes called levels of measurement) which are possible.

Types of Data

Nominal Measurement

The simplest types of measure is assigning the person to a category. For example, the scores for the sex of a person can be one of two categories: male or female. We say that the score of this particular subject for the variable of sex is female.

There may, of course, be several possible categories for our variable. For religion, the score may be Protestant, Roman Catholic, Greek Orthodox, Moslem, Hindu, Buddhist, Zoroastrian, Agnostic, Jewish, Shinto, etc. However, for most research there is often a limited number of categories in which we are interested. For example, in a study of religion in the Unted States, we might use three categories - Protestant, Roman Catholic and "other," lumping all of the other religions together in one residual category.

Some nominal measurements can be placed in order. For example, government data on the United States population often categorizes the age of the people by decade: 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70+. It makes sense to always list these variables in the text or in tables in the appropriate numerical order.

Ordinal Measurement

Consider a horse race. We often simply note who came first, who came second, and so, on down to who came last. These rankings (lst., 2nd., etc.) are ordinal measures.

Sometimes scores are not presented to us as rankings. For example, we may be told exactly how many votes each candidate in an election received. We then have to rank them first, second, third, etc.

Interval Measurement

Measurement on an interval scale means that one unit of the variable has the same significance or meaning as any other interval. For example, consider a simple ruler. An inch at the beginning of the ruler (say, on the left-hand side) is exactly the same length as an inch at

the end of the ruler, on the right-hand side. A dollar is the same amount of money whether you add it to a bank account of $100 or to a bank account of $1000.

Many psychological variables do not meet a strict criterion for being interval data. For example, consider intelligence test scores expressed as intelligence quotients (IQs). The average IQ of all people is defined to be 100. By convention, people scoring below 70 are considered to be retarded and those scoring above 130 are considered to be gifted. What does an increase of ten IQ points mean for a person? For the retarded, people scoring ten points higher than others may be able to tie their shoes or feed themselves. For the gifted, people scoring ten points higher than others may be able to beat a computer chess-playing program. Are these two intervals of ten IQ points equivalent? Probably not.

Thus, many psychological variables, though presented to us as apparently numerically precise scores, probably should be reduced to simple rankings. Despite this, statisticians and researchers using statistics often break this rule (and some others which we shall cover in subsequent pages).[1] However, the rule exists and makes logical sense, and inexperienced researchers should follow the rules until they have enough experience to judge which rules may be broken and under which circumstances.

Ratio Measurement

Ratio scores are interval scores with a true **zero.** A true zero is one in which there is none of the quantity which we are measuring. It makes sense to say that we have no money in our pocket. But consider temperature. Water freezes at 32^0 Fahrenheit. So 0° Fahrenheit is below freezing. But is there then no heat in the object? No. 0^0 Fahrenheit or 0° Centigrade are arbitrary zeros and not true zeros. Thus, temperature measured on these scales is an interval score but not a ratio score.[2]

Operational Definitions

As we noted above the operational definition of a variable describes the procedures by which we measure it. Often, we use shorthand to describe this in a research report. For example, we might simply state that we measured self-defeating behavior by using Schill's scale. However, sometimes the procedure should be spelled out more precisely.

How was the scale devised? How many items does the scale have, and what is the format for answering each item? What is the range of possible scores? What is the evidence that the scale is reliable and valid, concepts which I will define in the next section? What did we do about subjects who failed to answer one or two of the questions? How many subjects left the questionnaire so incomplete that we had to discard their questionnaire? In a longer report I might have answered these questions. The scores probably should be ranked, but most social science researchers would treat them as interval scores.

---

[1] The most common reason for breaking this rule is that the inferential statistics available for analyzing interval scores are far more varied and powerful than those available for ordinal and nominal scores. Thus, more comprehensive and interesting inferences can be made if we treat scores as interval measures.

[2] There is a ratio scale for measuring temperature called degrees Kelvin or degrees Absolute. The intervals are the same as those for the Centigrade scale, but water freezes at roughly 273.2^0 Absolute.

However, since the test is published and since I gave the reference for the test so that readers are able to locate it for themselves (by writing to the publishers or by consulting a psychologist who has a file of psychological tests), these details were omitted.

Research reports typically have a section entitled "Procedure" in which the operational definitions of the variables in the study are spelled out and in which the researcher tells the reader exactly how the research operations were carried out.

Reliability

An important question that must be answered about all measurement is whether the measuring instrument is reliable. A better phrase for this would be whether the measuring instrument is consistent, that is, regardless of what it is measuring, does it measure it consistently? There are several types of consistency.

Test-Retest Reliability

Let us say that I wish to measure the length of a table with a ruler. I measure the table and find it to be 32 inches long. What if I measure the same table with the same ruler tomorrow? Will it still be 32 inches long?

For the measurement of the lengths of tables to the nearest inch, we would expect great, even near perfect, reliability. But what about measuring your intelligence or your self-defeating behavior on Schill's scale? Will you get the same score tomorrow as you did today?

For an intelligence test, the fact that you saw the questions yesterday and had practice on them may increase your score today. You will be better prepared and less anxious about the type of questions and your ability to answer them, and you may have thought about the ones you could not answer and worked out solutions for them.

For the test of self-defeating behavior, these factors, though present, may be less important in determining your answer. But your psychological state may have changed. You might have had a migraine headache yesterday or a bad dream overnight, and these may affect your responses.

Scores may vary from test to retest, therefore, because the measuring instrument has poor reliability or because the variable being measured varies from day to day, or both.

Parallel-Forms

Another way of exploring whether a measuring instrument is reliable is to have at least two forms of the instrument. To measure our table, we could measure it with two rulers and see whether the two measurements are the same.

Psychological tests are sometimes created with two forms, each with different questions, so that one person may be given two versions of the test. A reliable instrument

would result in similar scores on the two instruments for the people we test. (Identical scores would be too much to hope for!)

Split-Half Reliability

If we have only one instrument but with many items, we can compare the scores on the odd-numbered items of the instrument with the scores obtained on the even- numbered items. This is almost as if we had created two parallel forms of the instrument but chose instead to combine them. There is some advantage to combining the two forms since longer psychological tests with more items tend to be more reliable on the test- retest criterion.

Item-Analysis

Reliability can also be explored by looking at whether the responses to the individual items are similar. For example, if we have forty items on our self-defeating behavior scale, people who answer, say, item number twenty-two indicating self- defeating behavior should, in general, answer item thirty-five in the same direction. Of course, individuals vary greatly in the pattern of their responding, but there should be a degree of consistency over all of our respondents.

Sometimes, investigation of a long scale reveals clusters of items for which people respond consistently within each cluster but not between clusters. Identification of these clusters enables us to derive sub-scales from the original scale.[3]

Validity

Once we are confident that the measuring instrument that we are using is reliable, we must next ask whether it is valid. The instrument is valid if it really does measure whatwe think it does. Does our ruler really measure length and not weight? Does our intelligence test really measure intelligence rather than test-taking anxiety? Again, there are several ways to explore the validity of a measuring instrument.

Concurrent Validity

The simplest way of exploring whether a measuring instrument is valid is to compare the scores it gives with the scores obtained frown another instrument which presumably to measure the same variable. For example, we can give our new improved intelligence test to a sample of people along with a second, more established, intelligence test. We hope that the two sets of scores will give similar rankings of our respondents. If they do, we will be more confident that our new intelligence test is valid.

Predictive Validity

[3] The complex statistical procedure for doing this is usually called factor analysis.

It may be that we expect scores on our measuring instruments to predict future behavior. If our measurements of the height of the tractor-trailer are valid it will not get stuck under the bridge one mile ahead on the highway.

Similarly, we may predict that scores on the SAT given to high school seniors will predict their grade-point averages at college. If the scores do so, then we may consider them valid.

Content Validity

In content or face validity, we look at the measuring instrument and ask ourselves whether it appears to be valid based on the content of the questions. Intelligence tests have face validity to the extent that the questions reflect what we think intelligence involves. For the measurement of intelligence, there is a group of psychologists who feel that intelligence tests are not face valid. Some believe that creative responses are penalized while others believe that the tests have a cultural bias in favor of the white middle-class to name just two objections to the validity of these tests.

Construct Validity

This is the most difficult type of validity to explain. Consider the scores obtained from intelligence tests. Do the scores follow the same patterns as we would expect if they really measured intelligence? For example, we might expect that intelligence increases with age during childhood. So, do older children obtain higher scores on our measuring instrument? We might expect that intelligence would decrease with advancing age in older people. Are the scores on our measuring instrument by octogenarians lower than those obtained by people in their sixties? Research which graves affirmative answers to these questions would provide construct validity for the measuring instrument.

In fact, every time we propose a hypothesis about some variable and confirm the hypothesis by our research, we both confirm the theory generating the hypothesis and we validate the measuring instrument.

Should You Devise Your Own Measuring Instruments?

In some courses on research methods, the instructors often have students practice constructing their own measuring instruments. It is a good way of teaching the concepts involved. However, for many major psychological traits, there are well-established measuring instruments which have been shown to be reasonably reliable and valid. For you to convince others that your new measuring instrument for the variable in which you are interested is reliable and valid would involve a great deal of research which is probably not relevant to your goals. Thus, it is usually preferable to use a published psychological test.

In my study on self-defeating behavior and suicidal ideation, I used Schill's inventory. (An inventory is simply a set of statements relevant to one or more

psychological traits.) The report of this instrument summarizes the results of the reliability and validity studies. Such reports may also give the distribution of scores which people who take the inventory obtain (which are called the norms for the inventory).

Locating Psychological Tests

There are several ways to locate what published tests exist for measuring psychological traits. For published tests which have to be purchased (and for which there may be requirements, such as being a licensed psychologist), Buros (1974) has produced Tests In Print and other related volumes.

There are some sources for tests, such as my fear of death scale, which are published in journals and which usually may be used free of charge by other researchers, for example, Robinson and Shaver (1978).

Failing these sources, researchers can use an abstracts service, such as Psychological Abstracts, and look up research on the psychological trait in which they are interested to see how other investigators have measured it.

The Reliability and Validity of Judgments

Often measures are simple judgments made by naive subjects or by experts. For example, Leenaars (1988) in his study of suicide note content, had judges decide whether a particular type of content was present or absent. One important question to ask is whether the judgments made are reliable and valid. A recent study of mine (Lester, 1991) exploring how well theories of suicide applied to cases of suicide failed to do this since only one judge was employed!

The reliability of judgments is easily examined. One way of doing this is to have one judge make two independent sets of judgments about the materials on two separate occasions. Do the judgments agree with each other? A second way is to have two judges make independent judgments about the material. Do the two judges agree? If the judges agree on the whole, they may meet to discuss the particular materials on which they disagree in order to arrive at a final score for each specimen.

The validity of the judgments will typically be assessed by the success of the research in confirming the hypothesis behind the research.

Conclusions

Research requires that we locate and use reliable and valid measuring instruments for the variables in the hypothesis guiding the research. For most psychological traits, there are many instruments already devised which have good reliability and validity. Finding these instruments is, however, not always easy. The reliability and validity of ratings made by judges must also be considered by researchers who use such judgments.

CHAPTER 4: CHOOSING SUBJECTS FOR THE RESEARCH

The people to whom you administer your procedures and psychological tests in your research are called subjects, and the group of subjects as a whole is called your sample. How do you choose the subjects for your research?

The requirements of statistical inference which is a necessary component of research demand that your sample be random. This does not mean that you, the researcher, wander around giving this person a questionnaire, and then that one, but not the next. It means that you first define your population in whom you are interested and from which you will draw your sample. Then you choose subjects such that every member of the population has an equal chance of being selected for your sample.

Let us say that I am interested in people. The costs and effort involved mean that Iam not going to consider all people in the world as my population. It will be too expensive to fly to Taiwan IG give a questionnaire to the person in my sample from that nation. Even America may be too big for my limited research project, though the polling companies who tell us about the preferences for Presidential candidates in the United States prior to general elections have to sample from the complete voting population of the country.

Let me limit myself to just the students in my college. To choose a random sample from these students, I would have to put the names of each student into a barrel and then choose, say, one hundred names. If the barrel is rolled until the names are well mixed, each student has the same probability of being selected for my research project.

It is important to realize that random samples may not be representative of the population. Some random samples of my college population, by chance, will on rare occasions consist of all men or all people over the age of thirty. That does not matter. Random samples do not have to be representative, only random.

In one classic mistake, the Literary Digest in 1936 polled a random sample of voters about their Presidential preferences. They predicted the election incorrectly because they sampled from telephone owners. In 1936, this procedure eliminated the poorer members of the electorate. The sample was a random sample of telephone owners but not representative of the electorate.

Nonrandom Samples

Stratified Random Samples

When polling companies are asked to find out the electorate's preferences for political candidates, random samples may give distorted estimates if they are not representative. Therefore, polling companies often decide upon a few characteristics of their sample which they consider to be essential. For example, they may want to make sure that each region of the USA contributes the correct proportion to the sample, and that the sample represents the electorate in its composition by sex. This means that they will choose, say, 65 men and 75 women living in New England, 145 men and 160 womenliving in the Pacific region, and

so on. Then, they will choose the 65 men in New England randomly, the 75 women in New England randomly, etcetera.

Systematic Sampling

For some researchers it is a nuisance to place the names of all of the population ina barrel and choose one hundred, or even to have a computer simulate this. It is easier to simply print out a list of all of the people in the population of interest and then choose every one hundredth name or five hundredth name. The spacing is determined by the size of the population and the size you have chosen for your sample. To choose a sample of 100 from a population of 5,000, one chooses every fiftieth person on the list.

Systematic sampling assumes that those at the top of the list will not differ in the variables you are measuring from those at the bottom of the list. This may be true most ofthe time, but it is possible for those whose names begin with A to differ in personality from those whose names begin with Z. Perhaps A's sit in the front of classes while Z's sit at the back. Perhaps A's (who usually get called first) have low frustration tolerance while Z's (who have always had to wait) have high frustration tolerance?

Despite the possible problems with systematic sampling, it is far easier (less cost and less effort) than truly random sampling.

Availability Sampling

As I mentioned in the previous chapter, some of the rules of research are occasionally broken. The choice of a sample often breaks the rules. Much psychological research, for example, is conducted on students in Introductory Psychology courses. Research participation OR writing an essay is often required of these students in major universities where a great deal of research is conducted. Students cannot be forced to participate in research. That would be unethical. But they can be required to put in four hours of work either by being a subject in research or by completing some other project.

Introductory Psychology students are used because they are <u>available</u>. They are not randomly chosen, and they may not be representative of students in general. But you will find that psychologists use inferential statistics regardless of this. On the plus side, the sample is clearly defined, and other investigators can easily choose similar subjects if they wish to replicate the research.

The subjects used in my study of suicidal ideation and self-defeating behavior reprinted in Chapter 1 came from an availability sample -- the students in my courses at my college.

Generalization and Replication

Little confidence can be placed in the results of any study until it has been replicated. Once the study has been replicated, we can generalize the findings. The first replication should

really be carried out by the original researcher. Some scholarly journals have suggested that no study should be published until the original investigator has replicated the study. This can be done most easily by dividing your sample into two equal parts and treating one part as the replication sample for the first part. The inferential statistics can usually incorporate this replication characteristic into the statistical analysis.

Research also needs to be replicated by other investigators working in other parts of this nation and by those in other nations of the world and, for example, in men and women, in different ethnic groups, and in those of different ages. Only after all of this canwe be confident about the results.

Biases Introduced by the Choice of Subjects

The choice of subjects for research can open introduces bias into the results of the research (Lester, 1969a). We have seen that much psychological research is restricted to college students, and this means that the results of the research might not hold true for thekinds of people who do not go to college.

Selection procedures may also introduce bias. In studies of psychiatric patients, it is easier to study the chronic patients who have settled down in the hospital than the acute patients who have been recently admitted. In older studies of homosexuals in the 1930s and 194Gs, the subjects were typically restricted to those in psychotherapy. In the 1990s, homosexuals can be selected from the larger community and include those who are not in psychotherapy.

Often people refuse to participate in research, and ethically they cannot of course be forced to participate. Other people agree to participate but do not show up for their appointment. Perhaps volunteers differ from those refusing to volunteer in important traits that might affect the results of your research?

Even those who volunteer fail to answer every item on the questionnaire you give them. Some omit many items. If too many items are omitted, or if important information is omitted, such as age, then the data from those subjects may have to be discarded.

Some subjects love to volunteer for research. Experienced subjects, therefore, may differ from inexperienced subjects. For example, they may be less anxious.

Researchers have also speculated and tried to investigate what the experience of being a subject in research is like. Do subjects try to guess what the researcher is measuring? Do those who guess correctly respond differently from those who guess incorrectly? When I was an undergraduate, my roommate, a physicist, volunteered to be a subject for the final exam in psychology. He was assigned to one student taking the examand, in order to try to help her, he faked all of his responses so that, when she plotted her results on a graph, she would get a straight line! She was probably the only student to obtain a straight line, and the examiners probably thought she had cheated.

Research designs often imply certain outcomes. These have been called the demand characteristics of the research. For example, if you put each subject in solitary confinement for several hours to explore the effects of solitary confinement on the mind, providing a panic button to press if they really want to be released early may suggest to the subjects that they may panic, and this many cause them to panic.

Related to the demand characteristics of the research are response sets (Berg, 1967). Subjects may have a particular set (or attitude) toward responding. A common response set is an acquiescence response set in which subjects tend to agree with statements. Thus, many psychological tests have items phrased both negatively and positively, for example, "I fear death a lot," and "I am not concerned by death." Those who say "Yes" to every item will not necessarily then come out with the highest score.

Conclusions

Choosing your sample often involves very difficult decisions, and the practicality *of* the choice may make your sample less than ideal for statistical purposes and open to criticism on methodological grounds. As a result, your research may be judged to be poorsolely on the basis of your sample.

Following the procedures for sample selection of other investigators who have worked on the topic increases the chances that your research will be considered acceptable. Replication of your results by yourself on new samples and by other investigators of their samples ensures that you can place some confidence in your results.

CHAPTER 5: CORRELATIONAL RESEARCH DESIGNS

In correlation research designs the variables studied are there in the subjects to be measured. The researcher does not manipulate or determine them. For example, let us assume that I want to see whether men use guns for suicide more than women do. I take one hundred consecutive suicides in Philadelphia and note the method used for their suicide and their gender. I then score each subject as male or female (a variable with nominal measurement) and whether they used a firearm for suicide or not (a second variable with nominal measurement).

I do not manipulate the gender of the subjects. They have their gender determined already when they arrived at the Medical Examiner's facility! Similarly, I do not manipulate their choice of method for suicide. They decided upon this themselves. I simply measured these two variables in the subjects.

This research design is very common and is called a correlational study. Correlation is simply the statistical term for association. A correlational study measures wh•ther two variables ar• associated (or correlated). *in the study above, th-• iiypotb.esis is that gender is associated with choosing a gun for suicide.

Correlational studies can be carried out on more than two variables. We can correlate hundreds of variables quite easily now that we have computers to perform the calculations. If we give a fifty-item psychological inventory to some subjects and wish to carry out an item-analysis in order to explore the reliability of the inventory, then we have to correlate the responses to each item with the responses to every other item. We have a total of fifty variables (each item in an item-analysis can be considered to be a separate variable), resulting in 1,225 correlation calculations.

Cause-and-Effect

Correlational research designs are useful, but they have one weakness. They do not permit cause-and-effect conclusions to be drawn. Let us say we have correlated two variables, A and B. It may be that variable A causes variable B. It may be that variable B causes variable A. Or it may be that some third variable C causes both A and B, and the association between A and B is simply incidental. Correlational studies do not permit us to distinguish between these three alternatives.

Let us say that we have found that depressed people as measured by a self-report questionnaire on mood obtain lower scores on a test of self-esteem. It might be that a depressed mood makes it more likely that a person will have low self-esteem. It might be that having low self-esteem makes it more likely that a person will be depressed. Or it may be that some third variable, such having a tendency to think irrationally, leads to both depression and low self-esteem in people. Our correlational study does not permit us to distinguish between these three alternatives.

Sometimes the nature of the variables enables us to eliminate some of the alternatives. For example, your choice of method for suicide cannot cause your gender, while your gender

could determine your choice of method for suicide. But in general, we cannot draw cause-and-effect conclusions from correlational studies.

Calculating Correlational Statistics

In this brief introduction to research methods I do not intend to cover the statistical analyses that accompany each research design. These may be found in a textbook on statistics, such as that by Witte (1993).

However, the particular correlational statistic you use will depend upon the level of measurement of each variable. Table 5.1 gives the names of some correlational statistics appropriate for each level of measurement.

Table 5.1: Correlational Statistics

Level of Measurement	Statistic
nominal	Phi coefficient
	Contingency coefficient
	Tretrachoric coefficient
ordinal data	Spearman's rho
	Kendall's tau
Interval/ratio	Pearson'r
one variable nominal one variable interval	Point-biserial coefficient

To make it easier to interpret research reports that you read, it is useful to know that the majority of correlation coefficients range from -1.0 to +1.0. The sign of the correlation coefficient (+ or -) indicates whether the association is positive (a high score on variable A is associated with a high score on variable B) or negative (a high score on variable A is associated with a low score on variable B). Subjects with higher scores on a test of depression typically have lower scores on a test of self-esteem (a negative association) but higher scores on a test of irrational thinking (a positive association).

The absolute size of the correlation coefficient (that is, regardless whether it is negative or positive) tells us how strong the association is. Zero means no association. Plus one or minus one means the strongest association possible.

A Case Study

It has long been suspected that men have a higher rate of fatal suicide whereas women have a higher rate of nonfatal suicide (Lester, 1969b). However, rarely does a researcher collect complete data on suicidal behavior in a community. However, Shneidman and Farberow (1961) tried to locate every fatal and nonfatal suicide in 1957 in Los Angeles.

Shneidman and Farberow located 768 fatal suicides -- 540 men and 228 women -- and 2652 nonfatal suicides -- 828 men and 1824 women. These data comprise two nominal variables: gender (male versus female) and outcome (death versus life). These numbers can be used to calculate a phi coefficient of correlation -- which comes to 0.33. This indicates only a moderate association but, because of the large sample size, we can be very certain that the association truly exists.

Conclusions

In this chapter we have described the correlational research design. Correlational research designs explore whether two or more variables are associated (related or correlated). Correlational research designs, however, do not allow us to draw cause-and- effect conclusions about the variables we are studying.

A Case Study

It has long been suspected that men have a higher rate of fatal suicide whereas women have a higher rate of nonfatal suicide (Lester, 1969b). However, rarely does a researcher collect complete data on suicidal behavior in a community. However, Shneidman and Farberow (1961) tried to locate every fatal and nonfatal suicide in 1957in Los Angeles.

Shneidman and Farberow located 768 fatal suicides -- 540 men and 228 women -- and 2652 nonfatal suicides -- 828 men and 1824 women. These data comprise two nominal variables: gender (male versus female) and outcome (death versus life). These numbers can be used to calculate a phi coefficient of correlation -- which comes to 0.33. This indicates only a moderate association but, because of the large sample size, we can be very certain that the association truly exists.

Conclusions

In this chapter we have described the correlational research design. Correlational research designs explore whether two or more variables are associated (related or correlated). Correlational research designs, however, do not allow us to draw cause-and-effect conclusions about the variables we are studying.

CHAPTER 6: EXPERIMENTAL RESEARCH DESIGNS

In the previous chapter we reviewed the features of correlational research designs. Experimental research designs are important because only these research designs enable us to draw cause-and-effect conclusions. Let us see why.

The Independent Variable

Experimental research designs are characterized by the fact that one of the variables is created (sometimes we say manipulated) by the researcher. The subject's gender, male versus female, cannot be affected by the researcher. Each subject comes with his or her gender, and the researcher can merely note it. Similarly, almost all psychological traits are part and parcel of the subject, and the researcher simply administers some psychological test to assess the strength of the trait.

In contrast, consider the following research study. Hood (1970) was interested in the fact that many scholars evaluate the psychological and psychiatric state of completed suicides prior to their suicidal actions knowing that the people killed themselves. He felt that this knowledge would distort the evaluations made by judges. Accordingly, he gave either the 32 genuine suicide notes or the 32 simulated suicide notes reprinted by Shneidman and Farberow (1957) to undergraduate psychology majors and told them that these were suicide notes written by men at an earlier period in their lives. He told the students that these men died much later; some deaths he identified as from suicide andsome deaths he identified as from natural causes. He asked the students to rate the suicidal intent of the note writers on a scale of 1 to 7.

Hood decided which note writers would be labeled as suicides and which wouldnot. He also determined which students would be given the genuine notes and which students would be given the simulated notes. Thus, this study is an experimental studysince Hood manipulated the two independent variables: genuine versus simulated notes and apparently dead from suicide versus natural causes. The dependent variable was the rating of intent made by the student judges.

One critical element in an experimental research design is that subjects must notbe allowed to choose which condition they experience (in this example, mode of eventual death). They must have no knowledge that there are different conditions, and they mustbe assigned randomly. Hood gave out the questionnaires at random to the students to ensure this, probably by shuffling the two different forms of the questionnaire (those with the genuine and those with the simulated notes) before handing them out. Hood shouldalso have have two forms of the questionnaires of each type, with the notes labeled as eventual suicides and eventual natural deaths in one form switched in the second form (in case the two sets of notes so labeled differed in perceived suicidal intent). He does not tell us that he did this.

Hood found that whether the students were given the genuine or the simulated notes had no impact on their judgments. However, students rated the eventual suicides as having more suicidal intent than the eventual natural deaths, as predicted. Since this was an experiment, Hood was able to conclude that knowledge of the person's eventual mode of death influenced the students' judgments of their suicidal intent in their suicide notes written at earlier age.

Mixed Designs

Research can also be designed with both correlational and independent variables. For example, we could explore the effects of the gender of the student judges on their evaluations of suicidal intent in Hood's experiment. We would simply have the questionnaires completed by men and by women and have them note their gender on the questionnaires. In these mixed designs, gender and eventual mode of death are sometimesboth called independent variables even though only one was manipulated by the researcher.

Pilot Studies

In all research, but especially in experimental research designs, it is often useful to select a few subjects to test the procedure. For this pilot study, or preliminary study, it is not necessary that the subjects be randomly chosen, but it is a good idea to make sure that they resemble the subjects to be used in their major characteristics (such as all college students or all senior citizens at a community center) so that you get a good idea as to how your subjects will react.

It is useful to question these pilot subjects after they have participated in the research to check what thoughts they had. For example, did they guess the hypothesis behind the study, and did they both understand and follow the instructions?

It is useful to do this even for correlational studies. Though correlational studies often employ standardized psychological tests which cannot be changed even if subjects misunderstand them (standardization means that all subjects must be administered the psychological test in the way prescribed by the designer or else the scores are not comparable to the norms for the test), you may have written some instructions yourself whose comprehensibility needs to be checked.

Statistical Analysis

As with correlational research designs, the present volume will not explain how todo the necessary statistical calculations so that inferences can be drawn from the scores. However, the most common statistical test carried out on scores from experiments is called analysis of variance (Witte, 1993).

If you have only one independent variable with three or more conditions, the statistical test is called a one-way analysis of variance. If you have two independent variables (each with two or more conditions), it is called a two-way analysis of variance; n independent variables requires an n-way analysis of variance.\

In the special case of one independent variable with two conditions (as if Hood's experiment had studied only the effect of knowledge of the eventual l\mode of death), it isalso possible to use a simpler statistical test called a t-test.

Conclusions

Experimental research designs are those in which the researcher creates (or manipulates) a variable. Groups of subjects may be given different sets of instructions or experience different conditions in the research. The advantage of experimental research designs is that researchers can conclude that the variable they manipulated caused the differences in the dependent variable.

CHAPTER 7: DRAWING CONCLUSIONS FROM RESEARCH

We have seen already that the type of conclusion we can draw from research depends upon the research design. Correlational research designs permit only inferences regarding the associations between variables, whereas experimental research designs permit us to draw cause-and-effect conclusions.

The conclusions we can draw from research are limited further by the probabilistic nature of statistical inference. This is not the place to go into detail into the logic of statistical inference, but some information is necessary so that you may report your results accurately and so that you will understand the reports of the research by others.

The Logic Of Statistics

There are several steps in the formal logic of statistical inference.

(1) First it is necessary that you state your formal hypothesis which is guiding your research. This is called the alternative hypothesis. Some formal hypotheses are directional (called two-tailed in statistics). This means that you predict the direction of the difference. For example, you may predict that women are more likely to engage in nonfatal suicidal behavior than are men, whereas men are more likely to engage in fatal suicidal behavior than are women.

In contrast non-directiona1 hypothèses (called one-tailed in statistics) predict only a difference. For example, you may predict that men and women will differ in the outcome (fatal versus nonfatal) of their suicidal actions, but you do not predict which group will commit more fatal acts.

Better theories permit us to specify the direction of the difference, as does the experience of previous research into the topic.

(2) It is also necessary to state the null hypothesis, that is, what will happen if your prediction is incorrect. The null hypothesis is one of no differences or no association. In the example given above, the null hypothesis would be that men and women do not differ in the outcome (fatal versus nonfatal) of their suicidal acts.

(3) You then design your study and collect your data.

(4) It is now time to apply statistical tests of inference to your data. In doing this, it is necessary to choose your level of statistical significance. This refers to the probability that you may draw an inference which is incorrect.

For example, in the example above, it may be that your sample of women does survive suicidal behavior more often than do men. However, if you could examine every man and woman in the world who engaged in suicidal behavior, you might find no differences between men and women in the outcome of their suicidal acts. Your sample was deviant. Thus, on the basis of your

sample you would draw a inference which was, infact, false. The probability of doing this is called the level of statistical significance. It is also called a Type I error. Look at the table below:

your hypothesis is actually	false	true
accept it the inference you draw about your hypothesis		
reject it		
	Type I error	correct decision
	correct decision	Type II error

Your hypothesis is not true for the total population but, if your sample permits you to accept it as true, then you have made a Type I error. On the other hand, if the hypothesis is true for the population, but your sample does not permit you to accept it, then you havemade a Type II error.

To take an analogy, if you don't have cancer, but the doctor decides you do have cancer, he has a made a Type I error (which in medicine is called a false positive). On theother hand, if you do have cancer, but the doctor say that you do not have cancer, he has made a Type II error (which in medicine called a false negative).

By tradition, Type I errors can have no more than a 5 percent probability of being made. Thus, every inference drawn from research has some probability of being incorrect. The probability of this may be as high as 5 percent or very very low (for example, 0.1 percent), but it can never be zero.

(5) Having seen the results of your statistical calculations, you now draw your inference. Again, by tradition, you accept or reject the null hypothesis, not your alternative hypothesis. This means that your conclusions should read formally "we rejected the null hypothesis" or "we accepted the null hypothesis."

Actually, you read such statements only in dissertations written for academic degrees. In research journals, researchers typically write more definitely. They will say something like "The results of the present study indicated that women survive their suicidal actions more often than do men."

(6) The statistical calculations and probability of Type I errors are typically incorporated into the text. For example, in the example printed in Chapter 1 of my study on self-defeating behavior and suicidal ideation, the text read:

Self-defeating scores were correlated with depression scores (Pearson r = 0.55, one-tailed R < .001), current suicidal ideation (¿ = 0.58), and to ever having thought about suicide (r = 0.53).

Notice that the result was stated in terms of my alternative hypothesis. Had I beenwriting a doctoral dissertation I would have said that I rejected the null hypothesis of no association between self-defeating scores and suicidal ideation.

Notice also that I reported the type of and the size of the correlation coefficient I used. I also reported the results of the test of statistical significance of the correlation coefficient and the number of subjects, which would have a meaning for other researchers who have used a Pearson correlation coefficient. The "p < .001)" is the probability statement. It means that the probability that what I have just said is incorrect in the total population is less than 0.001 or less than 0.1 percent. If the probability of being incorrect was less than 0.01 percent, I would have said that. So the fact that I did not say "p < .0001" means that I cannot be that certain. Since I stated that I was certain at the .1 percent level of statistical significance, this means that my result easily met the traditionally accepted criterion for accepting my hypothesis (or, to be absolutely precise, rejecting the null hypothesis).

Conclusions

We have mentioned two qualifications that may have to be made for all conclusions drawn from research. First, it may not always oe possioie io ôraw cause-anô- effect conclusions. Second, all inferences have some degree of probability of being incorrect.

It is only after taking a course in statistical inference and carrying out some research of your own that the logic of statistics and research will become clear. In this brief chapter, we wanted simply to alert you to the issues and to explain why there are various mathematical symbols embedded in the results sections of scholarly articles.

CHAPTER 8: ETHICAL GUIDELINES FOR RESEARCH

In the past, subjects have occasionally been harmed by research. Individuals have had experimental medical procedures tried out on them without being fully informed about the research or giving their consent. In the field of psychology, people have been subjected to stress, again without giving their informed concert.

In one of the classic cases of this research, Milgram (1963) carried out studies in which people had to guide another person through a learning task. Every time the leaner made a mistake, the teacher had to administer an electric shock to him. The teachers were led to believe that the learner, out of sight in another room, might have had a heart attack. The experimenter insisted they keep on with the study until the end, administering the electric shock for each error by the learner or non-response. Would they continue to obey the experimenter?

After the study was completed, the teachers were told that they had been deceived and that the learner had not really received any electric shock and certainly had not had a heart attack. However, a number of participants left the psychological laboratory having gone through an experience in which they thought they might have caused to someone's death and with the knowledge that they obeyed an experimenter to the extent of apparently causing someone's death.

Although the research seemed to meet the standards of the time, it was criticized for being unethical (Baumrind, 1964). Baumrind argued that the research involved loss of dignity, self-esteem and trust in rational authority without the people being informed about these consequences beforehand. By current standards, this research would today not meet the ethical standards of the profession.

All research involving human or animal subjects funded by the federal government must be approved by special committees set up in the institution where the research is to be conducted to ensure that ethical and humane guidelines are met. Institutions in which research takes place usually have such committees to investigate all proposed research regardless of whether there is federal funding or not. Some journals now require authors of research reports to certify that they have met the ethical standards for research, and various professional associations (such as the American Psychological Association) have established ethical guidelines for conducting research.

Let us look at some general rules for the ethical conduct of research.

Who is in Charge and Who Approves The Research?

There should be a clear Principal Investigator in charge of each research study who takes full responsibility for the study and who should carefully consider the ethical issues of the study.

There should be a review committee set up to check all research proposals to ensure that they meet acceptable standards and that subjects will be protected from harm when participating in the research. These committees can prove a nuisance for those wishing to cary out research since often the members of the committees from outside of the scientific profession are not fully conversant with the rules for conducting research, but these

committees force principal investigators to ensure that their subjects will not come to any harm.

Informed Consent

Subjects participating in research must give informed consent. In order that the consent is informed, the subject must be told about the essential features of the study which might influence his or her willingness to participate. On the other hand, some research involves varying the treatment given to different groups of subjects and may involve some deception.

In these cases, the information given to the subjects must be limited. However, after the subject has participated in the research, the nature of the research should be revealed more fully. The guidelines permit concealing certain procedures or intentions from the subjects. They do not permit concealment of risks and dangers.

We must also remember that subjects may consent to activities which are not in their best interests. Their consent does not justify research involving such activities. It is these situations that a "Human Subjects Committee" can be useful in deciding whether to let the research proceed. For example, several years ago, there was interest in whether a particular chromosomal defect (the XYY syndrome) was associated with later criminality in males. Research was proposed to identify such males at birth and to follow them up to see if they did indeed show a higher incidence of criminal behavior. However, it was felt that the identification and labeling of such men at birth might be a more important cause of later criminal than the chromosomal defect itself. So, although parents consented to participate, the project was abandoned.

Even if the subject consents to participate in the research, we might consider towhat has the subject consented?

(a) invasion of privacy,
(b) donations of resources such as time and effort,
(c) surrender of autonomy (as in hypnosis or sensory deprivation),
(d) exposure to temporary discomfort or stress, or
(e) exposure to procedures that may result in physical or mental harm.

There are certain types of subjects, such as children, those psychiatrically disturbed, and those mentally retarded for whom it is debatable whether informed consentis ever possible since we cannot be sure that they understand the nature of the research in which they may participate or the risks and dangers involved.

It is often advisable to have a printed explanation of the research for each subject to read and to have the subject sign this form to indicate his or her consent. However, subjects should remain free to decline to participate or to withdraw from the study at any time, even if they signed a consent form.

There are other guidelines to protect the subjects. They should remain anonymous, and the information they give should remain confidential at all times. Therefore, it is advisable not to store the data with their names attached. (This is especially important if the

data could be subpoenaed by a criminal justice agency.) It is also highly unethical for the researcher to discuss the data from a subject with someone else if that subject can be identified.

Psychological Tests

Psychological tests are often restricted to researchers who are members of the American Psychological Association or who have appropriate qualifications to administer psychological tests. These restrictions are in place both to prevent the content of the tests becoming known to those who might be given them later and to protect thosetested from abuse of the tests.

Tests can be abused when unskilled testers interpret the responses to those taking the tests. Many tests have implications for the psychiatric health of the respondent or their intelligence. Thus, inaccurate feedback or feedback which is not accompanied by appropriate counseling can harm the person.

After the Experimental Procedure

The researcher's ethical responsibilities do not end when the experiment is finished. Subjects must be debriefed, that is, informed as to the true nature of the researchwhich it may not have been possible to tell them before their participation in the research. It is now appropriate to tell them the hypothesis guiding the study, and the different conditions to which subjects were subjected.

In addition to simple debriefing, the researcher can now check whether subjects have any concerns about the research or their responses in it. For example, I frequently give college students a self-report paper-and-pencil measure of depression. Subjects may realize that they obtained a high score ana may want io discuss this.

If the experimental procedure was especially stressful, the researcher may want to keep in touch with subjects and contact them months later to check how they feel.

CHAPTER 9: WRITING A RESEARCH REPORT

Although there are variations from journal to journal, most research reports follow a standard form. After the title, the authors names and their place of employment, there is usually an Abstract. This summarizes in a few sentences the purpose and results of the study.

Introduction

The report proper begins with an Introduction. In this, the topic of the research is introduced and previous research into the topic mentioned. Individual studies may be mentioned, identified by the name of the authors followed by either the date of the report or the number of the reference in the Reference section later. In this book I have used the "Name (Date)" format.

The Introduction must next focus on the specific problem addressed by the present research and make explicit what the hypothesis is and what theory or research led to the hypothesis. There may also be a need to explain why the present research design is being used to test the hypothesis.

Method

The Method section describes in detail how the study was conducted. It describes the subjects chosen, the measuring instruments used, and the instructions given to each subject. This must be in sufficient detail so that a reader could replicate the study exactly as you did it.

Results

The Results section describes the results of the research with a sufficiently detailed presentation of the statistical calculations so that, again, the reader can follow exactly what you did.

Discussion

In the Discussion section, the results are summarized, and the hypothesis is examined for whether it was confirmed or supported. The discussion section is also where criticisms of the research you did can be placed, suggestions made for future studies, and the implications of your results explored.

References

The final section lists all of the articles and books referred to in the text, in sufficient detail so that others can easily locate them. The format for these varies from journal to journal. The references given at the end of the present volume give you one style for references.

An Example of a Research Report

The following is an example of a typical research report. It is the report of a brief unpublished study by myself.

Title

Is It Rational To Have Reasons For Not Committing Suicide?

Authors

David Lester and Barbara G. O'Neill

COMMENT:

Who is Barbara O'Neill? She was a student enrolled in a research tutorial with me for credit toward her BA degree. She collected the data and did the initial data analysis. Some scholars would not give a student assistant co-authorship but thank them instead in a footnote for their help. Although Barbara did the initial data analysis, I ***never*** publish student-collected data without first checking the scoring of the questionnaires, the recording of the data onto code-sheets, and the key-punching of the data into the computer.

Affiliation

Center for the Study of Suicide

Abstract

Scores on two of the reasons for not committing suicide on the inventory devised by Linehan, et al. were associated with scores on a measure of general irrational thinking in a sample of people who had thought about suicide during a crisis. Thus, some of the reasons for not committing suicide may not be rational.

Introduction

Cognitive therapy views dysfunctional emotions and behavior as a result of irrational thinking (Ellis, 1973), and Lester (1989) found that persons with higher levels of depression scored higher on a test of irrational thinking, as did persons who havel previously attempted suicide. Thus, the particular dysfunctional behaviors of suicidal preoccupation and depression do appear to be associated with a general tendency to think irrationally.

Linehan, et a1. (1983) devised an inventory to measure the strength of beliefs for various reasons for staying alive when a person is thinking of committing suicide. The present study was designed to explore whether belief in these reasons might also be associated with irrational thinking.

COMMENT:

This is a very brief introduction. In the first paragraph, it is noted that previous research indicates that depression and suicidal preoccupation may be more common in those who think irrationally

in general. Lester and O'Neill are trying to take a more positive view of suicidal preoccupation. Linehan has devised a questionnaire to assess people's reasons for not committing suicide. Lester and O'Neill wanted to see if agreement with these reasons for not committing suicide was also associated with a general tendency to think irrationally. If they were to find this to be the case, then suicidal people would not appear to be quite so pathological as hitherto. There has not been much research on irrational thinking in suicidal people, though Lester and O'Neill could have cited a few more studies had they so desired. However, there have been no previous studies on their particular hypothesis -- namely that those with strong reasons for not committing suicide also have a tendency to think irrationally. Notice also that the last sentence in their introduction states the hypothesis explicitly and that it is phrased as a one-tailed hypothesis -- belief in the reasons for not committing suicide is associated with irrational thinking (not rational thinking).

Method

A questionnaire was administered anonymously to 37 male and 104 female students enrolled in undergraduate social science courses. Their mean age was 23.6 years (standard deviation 5.8). They were administered a test of irrational thinking based on Ellis's list of irrational thoughts (Lee, et al., 1979), a measure of self-esteem (Rosenbaum and deCharms, 1960), the Reasons For Staying Alive inventory (Linehan, et a1. 1983) andone question about suicidal ideation. For answering the Linehan inventory, respondents were asked to "Consider a time in your life when you had the greatest personal difficulties. Look at the following arguments against suicide. For each statement circle how important was it at that time as a reason for you for not committing suicide." At the end of the inventory, the respondents were asked "At that time, did the thought of suicide cross your mind?"

COMMENT:

This Methods section is very clear. The number of subjects is described, and their sex and age specified. They were not a random sample, but rather an availability sample obtained from several of my courses. Incidentally, my students typically cooperate very well, and hardly anyone refuses to complete the questionnaires.

References for the questionnaires used are given so that the reader can locate them. Since I changed the instructions a little on Linehan's questionnaire, I reprinted the exact wording I used.

The one item of information omitted concerns the order the questionnaires were in. Did I vary the order or give everybody the questionnaires in the same order? Since this is a correlational study, my practice is always to give everyone the questionnaires in the same order. If I were to use, say, three different orders (ABC, BCA, and CAB), this variable of order of administration could be introduced into the statistical analysis to see if the order made a difference, but my practice is to avoid this.

Results

The Linehan inventory is scored for six reasons for staying alive: survival and coping beliefs, responsibility to family, child-related concerns, fear of suicide, fear of social disapproval, and moral objections. The scale for child-related concerns was not used in this research since so few of the respondents had children. Seventy-three of the respondents had thought of suicide at their time of crisis, and only data from these respondents were used in the analysis.

Partial correlation coefficients were calculated controlling for age and sex (see Table 1). Scores on the irrational thinking test were associated with scores on survival and coping beliefs (r — 0.21, $p < .05$) and fear of suicide ($r = 0.25$, $p < .05$). Thus, respondents who rated these reasons as more important obtained higher irrational thinking scores. Self-esteem scores were related to survival and coping beliefs ($r = 0.31$, $p < .01$) and responsibility to family ($r = 0.24$, $p < .05$). Respondents with higher self- esteem scored higher on these beliefs.

COMMENT:

Lester and O'Neill explain first that they omitted subjects who had never thought about suicide. This item could have been included in the Methods section rather than herein the Results section.

The results are presented in the text and with an accompanying table of results. The particular correlation coefficient used is noted -- partial correlation coefficients are for interval data and are based on the Pearson correlation coefficient. The size of the correlation coefficients is noted and their statistical significance. All those reported in thetext are statistically significant at the 5% level or better. The correlation coefficients which failed to reach statistical significance are noted in the table without an asterisk.

Discussion

The present study has shown that irrational thinking, in addition to being related to depression and suicidal preoccupation, is also related to belief in some reasons for staying alive. Thus, these beliefs may be as dysfunctional as the desire to commit suicide.However, since self-esteem was also associated with scores on two of the reasons for staying alive, these beliefs may also reflect facets of psychological health.

COMMENT:

This is a very brief discussion. Note that it states the conclusion in the logicallycorrect way, but in the way which is common. I did not say that I rejected the null hypothesis; rather I said that I confirmed the study's hypothesis in part, that is, for someof the reasons. I did not point to any problems with the research which could be changedif the study was repeated, and I did not discuss possibilities for future research or the implications for counseling.

This discussion reflects my attitude when reading articles by other scholars. I find introductions and discussions boring, and I prefer to focus on the Method and Results sections. However, comments from editors of journals and reviewers of my articles which have been

submitted to journals for publication indicate that I am deviantin this respect. Editors and reviewers prefer longer and more substantive introductions and discussions.

References

Ellis, A. Humanistic psychotherapy. New York: Julian, 1973.
Lee, D. Y., Hallberg, E. T., & Haase, R. F. Endorsement of Ellis's irrational beliefs as afunction of age. Journal of Clinical Psychology, 1979, 35, 754-756.
Lester, D. Depression, suicidal preoccupation, and irrational thinking. Journal of General Psychology, 1989, 116, 221-223.
Linehan, M., Goodstein, J., Nielsen, S., & Chiles, J. Reasons for staying alive when you're thinking of killing yourself. Journal of Consulting & Clinical Psychology, 1983, 51, 276-286.
Rosenbaum, M., & deCharms. Direct and vicarious reduction of hostility. Journal of Abnormal & Social Psychology, 1960, 60, 105-111.

Table 1: Correlations Between Reasons for Staying Alive, Irrational Thinking and Self-Esteem

	I	II	III	IV	V
Pearson r					
age	-0.16	-0.22*	-0.05	-0.15	0.18
sex	0.04	0.03	-0.05	0.22*	0.02
irrational thinking	0.22*	0.13	0.25*	0.15	-0.14
self-esteem	0.31*	0.19	-0.20*	-0.11	0.03
Partial r					
irrational thinking	0.21*	0.09	0.25*	0.11	-0.10
self-esteem	0.31*	0.24*	-0.19	-0.04	0.06

* two-tailed $p < .05$ or better

I survival and coping beliefs
II responsibility to family
III fear of suicide
IV fear of social disapprovalV moral objections

References

Baumrind, D. Some thoughts on the ethics of research. American Psychologist, 1964, 19,421-423.

Beasley, D. How to use a research library. New York: Oxford University Press, 1988.Berg, I. Response set in personality assessment. Chicago: Aldine, 1967.

Buros, O. K. Tests in Print II. Highland Park, NJ: Gryphon Press, 1974.

Hood, R. W. Effect of foreknowledge of manner of death in the assessment from genuineand simulated suicide notes of intent to die. Journal of General Psychology, 1970, 82, 215-221.

Leenaars, A. A. Suicide notes. New York: Human Sciences, 1988.

Lester, D. The subject as a source of bias in psychological research. Journal of General Psychology, 1969a, 81, 237-248.

Lester, D. Suicidal behavior in men and women. Mental Hygiene, 1969b, 53, 340-345.Lester, D. Suicide at Christmas. American Journal of Psychiatry, 1985, 142, 782.

Lester, D. Experience of personal loss and later suicide. Acta Psvchiatrica Scandinavica, 1989, 79, 450-452.

Lester, D. The Collett-Lester fear of death scale. Death Studies, 1990, 14, 451-468. Lester, D. The study of suicidal lives. Suicide & Life-Threatening Behavior, 1991, 21, 164-173.

Lester, D. Why people kill themselves. Springfield, IL: Charles Thomas, 1992.

Lester, D., Fong, C. A., & Angelo, A. A. Chronic suicide attempters who switch methodsand those who do not. Perceptual & Motor Skills, 1989, 69, 1390.

Lester, D., & Hoffman, S. Self-defeating behavior, depression, and suicidal preoccupation. Psychological Reports, 1992, 70, 1006.

Lester, D., & Sheehan, D. Attitudes of supervisors toward short police officers. Psychological Reports, 1980, 47, 462.

Lester, D., & Topp, R. Major league baseball performances of players who were latersuicides or homicide victims. Perceptual & Motor Skills, 1989, 69, 272.

MacDonald, J. The girl in the plain brown wrapper. Greenwich, CT: Fawcett, 1968.Menriinger, K. Man against himself. New York: Harcourt, Brace, 1938.

Milgram, S. Behavioral study of obedience. Journal of Abnormal & Social Psychology, 1963, 67, 371-378.

Robinson, J. P., & Shaver, P. R. Measures of social psvchological attitudes. Ann Arbor, MI: Institute for Social Research, 1978.

Sheehan, D., & Lester, D. Attitudes of short versus tall police officers. Perceptual & Motor Skills, 1980, 51, 578.

Shneidman, E. S., & Farberow, N. L. (Eds.) Clues to suicide. New York: McGraw-Hill, 1957.

Shneidman, E. S., & Farberow, N. L. Statistical comparisions betweeen attempted and committed suicides. In N. L. Farberow, & E. S. Shneidman (Eds.) The crv for help. New York: McGraw-Hill, 1961, 19-47.