ESSENTIALS OF

SURVEY RESEARCH AND ANALYSIS

by

Ronald Jay Polland, Ph.D.

TABLE OF CONTENTS

LESSON ONE

Basics of Survey Research

OBJECTIVES

- Describe what surveys are used to measure
- Give examples of how surveys can assess change
- Define “needs”, “assets”, “behavior”, “opinions”, “attitudes”, “beliefs”

Key Terms

Survey, questionnaire, descriptive research, causal explanation, prediction, evaluation, single sample, successive samples, panel study

What do surveys of people measure?

Ultimately, researchers wish to know how people act and think in many common situations. The best way to do that is to observe them repeatedly and directly over time. For example, if we want to know how often people go to food stores, we could follow them every time they leave their house. If we wanted to know what they eat for breakfast, we could sit at their table in the morning and record what they ate. However, as you may have guessed, we rarely have the chance to watch people and their behavior all the time. We cannot be like “flies on the wall” and record their every move. Even if we could do that, there are kinds of human behavior that are we cannot see directly —such as attitudes, beliefs and opinions. So, what are researchers to do?

One solution is to use surveys. Surveys (also called “questionnaires”) are a systematic way of asking people to volunteer information about their attitudes, behaviors, opinions and beliefs. The success of survey research rests on how closely the answers that people give to survey questions matches reality – that is, how people really think and act.

The first problem that a survey researcher has to tackle is how to design the survey so that it gets the right information. Is this survey necessary? Is the purpose of the survey to evaluate people or programs? Can the data be obtained by other means? What level of detail is required?

The second problem is how accurate does the survey have to be? Is this a one-time survey or can the researcher repeat the survey on different occasions and in different settings? How will the results be used? How easy is it to do the survey?

Uses of surveys

Survey uses range from highly practical public opinion polls and market research studies to highly theoretical analyses of social influence. Planners and administrators use surveys to get baseline information for policy decisions. Social scientists use surveys to measure voter behavior, psychological influences on the spending and saving behavior of consumers, attitudes, values and beliefs related to economic growth and the correlates of mental health and illness. Economists rely on regular consumer surveys for information on family financial conditions and surveys of business establishments to measure recent investment outlays.

The survey is an appropriate means of gathering information under three conditions: when the goals of the research call for quantitative and qualitative data, when the information sought is specific and familiar to the respondents and the researcher has prior knowledge of the responses likely to emerge.

Descriptive research

Surveys can be used to describe phenomena and summarize them. The goal of using surveys for descriptive research is to get a precise measurement of certain things such as political preference.

Causal explanation

Surveys often measure associations between things such as school grades and self-esteem. The data from surveys can provide a causal explanation to phenomena such as why teens become pregnant or why teens do drugs.

Evaluation

Surveys are useful for determining the degree to which a desired objective is attained as a result of a planned program.

Prediction

Survey data can be used to forecast future events.

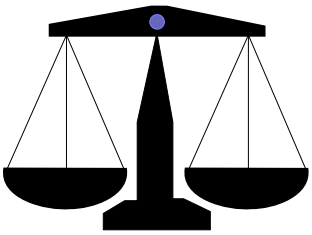
Types of survey design

The single cross section

This design involves collecting information at a single point in time from a sample selected to represent the total population.

Designs for assessing change

Successive samples and panel studies are methods for assessing change. In successive samples, the groups of people in each sample are different although their characteristics are similar . A panel study is one that collects information from the same group of people on two or more occasions. Panel studies permit better estimates of specific changes over time than do successive samples.



LESSON TWO

Measurement Levels and Types of Data

OBJECTIVES

- Define and give examples of three levels of measurement
- Define and give examples of eight types of scales
- Contrast the five types of measurement scales
- Compare and contrast interval versus ordinal scales

Key Terms

Categorical, ordinal, scale, attitudes, opinions, beliefs, behavior, attributes

Choosing the types of data to collect

The kinds of data you collect and how you collect them affects how you can use the data. That is why you have to decide beforehand what will be measured and how it will be measured. When something is measured, we assign to it a number or a label.

When numbers or words are used to group things, we call that type of data categorical data. Categorical data would be things like gender, race, religion, food group, or place of residence. When the numbers are used to order a list of things, that data is called ordinal. The ranking of football or basketball teams is done using ordinal numbers. A list of things to do would also be ordinal data. The third type of data is called scale data or interval data. Scale or interval data would be things like height, weight, age, class size, miles per hour and so on.

Types of data collected

There are six basic types of data that you might collect:

1. attitudes—what people say they want
2. opinions—what people think might be true
3. beliefs—what people know is true
4. behavior—what people actually do
5. attributes—what people are (demographic characteristics)
6. preferences—what people would choose

Each type of data may require a different type of question to get at that data. There are seven basic types of survey questions:

1. Open response
2. Partially open response
3. Closed response
4. Semantic differential scales
5. Agreement and rating scales
6. Ranking scales
7. Checklists

Open response questions are used when you need people to express themselves freely or when you cannot list all of the possible responses. (In your school, open response questions are also known as "essay questions"). Researchers use open response questions when they do not know how people will respond to a question. Open response questions are also used when no limit is placed on how much a person can say. Often, you will see surveys have additional lines for people to enter comments. This is a form of an open response question.

Open response questions that encourage free expression have some drawbacks. They tend to be very time-consuming for people to complete and for researchers to interpret. Often, the answers will be too short, confusing or irrelevant to be useful. Finally, the data from open response questions may be difficult to analyze for statistical purposes.

Unlike open response questions, Partially open response questions require specific, short answers that do not encourage free expression. Partially open questions are a compromise between closed response and open response forms. Partially open question provide an “Other” category where a person can provide additional information. An example of partially open questions are the blank spaces provided for the questions on racial background and persons living with you. Other examples of partially open questions are the ones following the question, "Have you ever had sex?" on page three of the TBOS.

You probably are most familiar with closed response questions from your work in school. These are the "multiple-choice" variety where a person has to choose among several possible answers.

There are two types of closed response questions: (1) those with *ordered* answer choices and (2) those with *unordered* answer choices. Ordered answer choices represent points along a continuum. For example, the second question on the TBOS (Teen Behavior & Opinion Survey) asks teens to indicate their highest grade completed. The choices are ordered from lowest (8th grade) to highest (12th grade).

Close response questions with unordered answer choices have each choice as an independent answer. The questions about racial background and religious background are examples of questions with unordered answer choices -- as well as the entire set of questions on who teens might talk to about their problems.

Semantic differential questions use a five to seven-point rating scale with each end of the scale having an adjective or phrase. These adjectives, called bipolar adjectives, are direct opposites. The questions at the bottom of Page 2 are semantic differential items. Semantic differential scales have three common factors: an Evaluative factor covering such dimensions as good-bad, pleasant-unpleasant and positive-negative; a Potency factor representing the dimensions of strong-weak, hard-soft and heavy-light, and an Activity factor with such scales as fast-slow, active-passive and excitable-calm.

Like the semantic differential, agreement scales have end points that are identified by adjectives or phrases—however, unlike the semantic differential, all of the steps may have an adjective or phrase associated with it. Typical of this type are ones that measures agreement using a five point scale with steps labeled “Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree.” The questions on Page 4 are of this type.

Related to agreement scales are rating scales. The questions at the top of Page 2 are rating scales that asks teens to rate how helpful each person would be if they were to talk to them about their personal problems. The questions at the top of Page 3 are also rating scales. These questions ask teens to rate how important to them are each of the goals listed.

Checklists are lists of items that people can check off. Usually they are questions about attributes or facts and people may check as many items as they feel applies to them. Checklists are like closed response questions except that for closed response questions, people are supposed to choose only one answer. The question at the top of Page 5 has a checklist below it.

Ranking scales require the respondent to “sort” items by assigning a numerical ranking (e.g., 1st, 2nd, 3rd...etc.) to each response item. The main difference between ranking scales and rating scales is that rating scales assign an absolute value to an item without regard to the other items; ranking scales assign a relative value to an item in relation to the items before and after it. Paired alternatives is a special type of ranking scale that compares every response item with each other. Paired alternative scales provide a more accurate ranking of items than can be done on a simple, rank-order scale.

LESSON THREE

Choosing a Questionnaire Format

OBJECTIVES

- Compare and contrast qualitative vs. quantitative data
- List advantages and disadvantages of interview, telephone and mailout questionnaires.
- Compare and contrast different types of questionnaire formats

Key Terms

Qualitative, quantitative, content analysis, longitudinal, panel study, social exchange.

Why people respond to surveys

The process of giving surveys to people, getting them to complete it in an honest manner and returning it can be viewed as a special case of "social exchange." The theory of social exchange says that people behave in certain ways when they know how others will react to their behavior. There are three things that must be done to increase the response to a survey: minimize the cost for responding, maximize the rewards for doing so and establish trust that those rewards will be delivered.

Rewards

If you have high regard for another person, that regard is rewarding to that person. For example, complementing a person is a way to express regard for them. Supporting their values can also be rewarding to people. If completing a survey or responding to an interviewer's questions can be seen as a rewarding act, then the survey process itself may provide the motivation to for a person to complete the survey or finish the interview.

Costs

A person acts on the basis of costs that may occur or that have already happened. For example, spending one's time completing a survey is a cost. Having to answer difficult questions is another cost. Even having to think of the proper answer can be a cost. The cost is high when a lot of physical or mental effort is required to answer questions or when the questions provoke embarrassment or anxiety. Disclosing personal information represents potential risks and questions of a very personal nature imply greater costs.

Trust

Building trust between the interviewer and the respondent is very important if the researcher wishes to get good reliable data. Sometimes, researchers will pay people for completing surveys -- as a symbol of trust. Paying respondents tends to encourage their belief that future promises will be kept. Trust is also obtained by assuring people that their responses will be kept confidential and that they will be given a copy of the results of the survey.

Why telephone and mail-out surveys are valuable

According to Donald Dillman, author of "Telephone and mail-out surveys - a total design approach," face-to-face interviews are not as successful as they once were and are becoming more expensive to do. As an alternative to in-person interviews, surveys by telephone and mail make it possible to reach hard-to-find respondents even though some may have moved. Also, doing telephone and mail-out surveys are cheaper than interviews and allow researchers to base sampling decisions on the nature of the research and not how many people they can afford to contact.

Response differences between mail-out and telephone surveys.

Typically, an initial contact by phone lasts only a few short minutes with the person who answers the phone having no prior warning. The person who answers the phone must decide whether to respond or not without knowing what the questions will be. In comparison, a person who receives a survey by mail has time to review the written materials. The decision to respond may be delayed until after the person thoroughly examines the survey. For these reasons, the effects of social exchange are much less for a telephone interview than for a mail-out survey.

LESSON FOUR

Elements of Item Construction

OBJECTIVES

- Identify seven types of bias in questionnaires
- Identify and state how to correct three common wording problems
- Define three types of response sets

Key Terms

Social desirability, acquiescence, bias, midpoint, operational definition, response set

Writing questions

Common Wording Problems

Writing questions for a particular survey means doing them for (1) a particular population, (2) a particular purpose and (3) for placement next to other questions in the survey. Words that are too difficult for some to understand may be perfectly acceptable for others. A question that is fairly vague may satisfy the objectives of one study but not the ones of another.

So that every respondent will understand a question, it is important to keep the reading level at or below the average reading level of the population. Complex words may be replaced by simpler ones or ones more easily understood. If you are giving a survey to a particular group, you would want to use words that are common to the group.

In addition to making questions more understandable, they should not lead a person to respond in any particular way. This is called “response bias.” Biased questions lead people to respond in a way that does not reflect how they really think and feel about a question.

Types of bias:

1. Objectionable questions
2. Unequal comparisons
3. Bias in more than one direction
4. Unbalanced categories
5. Emotionally charged words
6. Threats to self-esteem
7. Personalization

Objectionable questions are those that have words or phrases that produce negative feelings. One way to avoid objectionable questions is to create broad response categories (such as income or age categories). Another method of reducing objections to questions is to use a format that softens their impact. A third way to overcome objections is to use a series of questions rather than a single one.

Avoid double questions and double negatives. Do not assume that the person you are interviewing has a lot of prior knowledge. Make sure that the question you ask is accurate. Do not make questions too short - sometimes readability can be increased by adding words, not deleting them.

Forced choice vs. non-forced choice

Even number of alternatives creates forced choice; odd number offers a midpoint. The advantage of forced choice is to reduce the number of responses to the midpoint. However, research shows that midpoint draws equal number from both sides so that the overall distribution is not changed.

It is probably a good idea to include a midpoint like “Not applicable”, or “Unsure.” You want everyone to respond to every category. In some types of studies (like voter preference) , the number of “undecided” may be of value to know.

Look at the difference between these two questions:

In your opinion, how would you rate President Clinton's handling of the economy?

☐ EXCELLENT
☐ VERY GOOD
☐ GOOD
☐ FAIR
☐ POOR

President Clinton has done a good job of handling the economy

☐ STRONGLY AGREE
☐ AGREE
☐ NEITHER AGREE NOR DISAGREE
☐ DISAGREE
☐ STRONGLY DISAGREE

These two questions are not the same. In the first one, respondents are free to rate Clinton's performance on whatever criteria they choose. In the second question, the only option is whether one agrees that the statement is true. Someone who answered "Good" to the first one might answer "Agree" or "Strongly agree." Someone answering "Fair" to the first might answer "Agree" or "Disagree." In this example, it's unlikely anyone will answer "Neither agree nor disagree."

To agree with a statement, there are two possible judgments that may be made: (1) is the statement true? or (2) how does the person perceive what the statement refers to. The problem with questions like the second is that it can create a response bias by only having positive statements. "Good" is not clearly defined for all respondents. In the first question, "Good" is relative, not absolute as in the second question.

Look at this question:

Do you approve of the way President Clinton has handled the economy?

☐ STRONGLY APPROVE
☐ APPROVE
☐ NEITHER APPROVE NOR DISAPPROVE
☐ DISAPPROVE
☐ STRONGLY DISAPPROVE

In this form, there is no mention of whether his handling was “good” or “not good.” We are asking people if they approve of his actions or inactions. If doing nothing is what people perceive to be the best course of action, then they will approve of it. Responses to this question tell us nothing about what Clinton has or has not done. Approval ratings are popularity measures. Effective managers are often not very popular with their subordinates. So, too, is the case for presidents and the populace.

It is critical to state what the objective of a question is before the question is constructed. Then, after construction, the question should be compared back to the stated objective to determine its validity.

If you have a multiple response item, each one of those would be treated as a separate question. The following is a list of personality characteristics.

One way to avoid inconsistency is to define what each alternative means. Give sample questions answered and explain what the answer means.

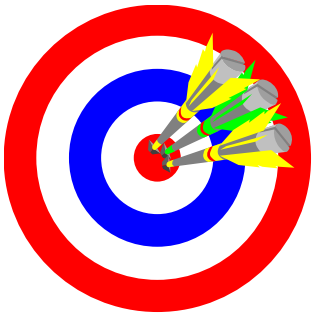
Response style and response sets

A response set happens when respondents choose too many of one type of answer such as “Yes” or “Agree.” To avoid response sets, the response options should vary their content. Social desirability is when people choose the response most favorable to their self-esteem or most in accord with their social norms. Again, response options should be given specific content. Additionally, response categories should be arranged so that their social desirability values are relatively close.

Positional biases occur when respondents mark the same area of a rating scale regardless of the question asked. Respondents will tend to respond toward the middle in a list of numbers, toward the extremes in a list of ideas and toward the second alternative in a list of two ideas. One solution is to develop two forms of the questionnaire in which the order of answers is reversed. Position bias may also be reduced by assigning more content to each of the rating categories.

Length of the questionnaire

Dillman’s research says that surveys should not exceed 11 pages or 125 items. Once people are on the telephone, however, the length of the interview does not appear to be a major problem.



LESSON FIVE

Assessing Reliability and Validity

OBJECTIVES

- Compare and contrast three types of reliability
- Compare and contrast six types of validity

Key Terms

Face validity, content validity, concurrent validity, discriminant validity, construct validity, test-retest reliability, internal consistency, stability.

Reliability and validity

In order for information to be useful, it has to be consistent, dependable, accurate and, most of all, true. Too often, we are presented with information that fails on one or more of these criteria. In research, these criteria are represented by the concepts of reliability and validity.

When we say that information is reliable, we mean that we can expect to obtain the same information time after time. The concept of reliability can be applied to sampling. If we repeatedly draw random samples of equal size from a population, we can expect to get the same sample values each time (plus or minus a certain amount due to sampling error).

When we say that information is valid, we mean that it is presented or used in the way for which it was intended. An IQ test is valid only if it is used to measure intelligence—it is not valid if it used to assign individuals to groups. A psychological test that is a valid measure of anxiety is not a valid measure of stress.

Types of validity.

Face validity is when the obtained information appears to be what was expected. A question that asked “Do you smoke?” would appear to have face validity as a measure of smoking behavior.

Content validity is when a question adequately reflects the underlying behavior or body of knowledge. Content validity is established by having experts evaluate the relevance of the test items.

Concurrent validity is when one question is comparable to another that validly measures the same content or construct. Concurrent validity is established by correlating one question with another that has previously been validated.

Discriminant validity is when a question can be used to differentiate one group from another. Surveys that are used to identify potential high school drop-outs would have discriminant validity if the students who graduate score higher on the test than students who leave before graduation.

Predictive validity is when a question can be used to predict behavior. One purpose of the SAT (Student Achievement Test) is to determine how successful students are likely to be in college. The SAT has “predictive validity” in its ability to predict a student’s Grade Point Average in college.

Construct validity is the last type of validity. A “construct” is a theoretical dimension like “self-esteem” that is measured by having several questions that all relate to how people view themselves. Self-esteem does not exist by itself but is represented by how people respond to these questions. In this example, construct validity measures the extent to which these responses can be called, “self-esteem.”

An example of construct validity on a survey of driving behavior would be if answers to the following questions are highly related. If they are, then one might assume that what is actually being measured is an underlying construct called “unsafe driving”:

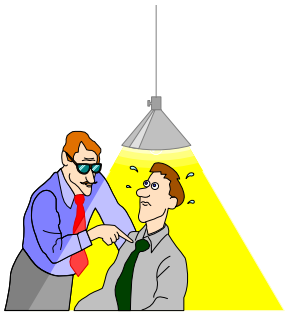
1. *I usually drive faster than the speed limit.* YES NO
2. *I try to “get even” with drivers who cut me off.* YES NO
3. *I usually don’t have time to signal lane changes*. YES NO

Types of reliability

Test-retest reliability is obtained by administering the same test on two or more successive occasions and then correlating the scores.

Internal consistency is obtained by correlating the scores on several questions that pertain to the same content to the sum total of the scores. The average item-total correlation is a measure of how consistently people respond to related items on a test. For example: if a math test had several sets of items that required people to multiply two and three digit numbers, you would expect persons who could not correctly multiply two digit numbers to be unable to multiply three digit numbers. Likewise, you would expect persons who can multiply three digit numbers to be able to multiply two-digit numbers. These expectations should be "consistent" from person to person and would be borne out by high item-total correlations of the items.

Stability refers to how much a person's score can be expected to change from one administration to the next. A perfectly stable measure will produce exactly the same scores time after time. This concept is similar to test-retest except that in test-retest situations there is no assumption that the absolute value of each person's test score will stay the same.



LESSON SIX

How to Identify the Research Problem and Define the Research Solution

OBJECTIVES

- Identify the characteristics of a research topic
- Define common goals of research
- Define what is meant by a "research problem."

Key Terms

phenomena, theoretical framework, research problem

Selecting a research problem

The first step in selecting a research problem is deciding what you really want to do. Therefore, the first goal of your research is to satisfy your needs for doing the research. The second goal of your research is to make a contribution to the body of knowledge—for your work to have utility for its audience. The third goal then is to satisfy the needs of the agency that funds your research.

Identifying the research problem

Perhaps, the hardest part of developing any research project is to narrow down the field of study and the "research problem." A distinction needs to be made between a "problem" and a "research problem." A "problem" is an observed discrepancy or "gap" between what is known and not known, between the ways things are and the way they should be or between two or more potentially related bodies of knowledge. The identification of the problem is an interpretation of the gap based on a set of observations. A "research problem" is a judgment drawn from the interpretation of the gap. The following are observations:

1. The proportion of black felons sentenced as “habitual offenders” by judges is twice that of white felons similarly sentenced.
2. Duval County has the highest rate of gonorrhea in the state.
3. Thirty-seven million Americans are without health insurance.
4. Men on low cholesterol diets are more likely to die violent deaths than men who eat normally.
5. Over 25,000 people died last year in alcohol-related traffic accidents.

While each of the above are observations about a specific area of concern, they are not research problems -- they do not define the gap. Here are the above observations transformed into research problems:

1. The race of convicted felons appears to influence the decision of judges during sentencing.
2. Too many Duval County residents engage in risky sexual behavior.
3. People without health insurance are denied access to medical care.
4. People on low cholesterol diets don’t live any longer than people who eat normally.
5. Too many alcohol-impaired drivers are still on the road.

Defining the research solution—an analogy

Taking the example of food shopping, here is how you would define the research problem:

⇒ *You need to prepare a covered dish to bring to a dinner party.*

Stemming from the research problem is the need for the study; i.e., what you will do to solve the problem or to help others understand the problem. Before you can decide on a solution, you need to consider what that solution will look like and what it will accomplish. These are the criteria for the problem solution:

1. Affordable - does not exceed your budget for food
2. Easy to prepare - can be completed with minimal additional expertise or training
3. Produces desired and expected effect - The finished product will what the recipe says it is. It will look and taste like it is supposed to.
4. Reliable - If the recipe calls for cooking 30 minutes in a 350 degree over, you expect no variance from 30 minutes or 350 degrees.

5. Conforms to requirements of hosts - the hosts don't want 30 dishes of baked beans or cole slaw nor do they want something like lobster newburg that only feeds one , so they will ensure that each participant bring something according to a set of specifications; i.e., salad, dessert, side dish, feeds 4 to 6 people.
6. Satisfies personal needs - you have a need to produce something of value either because its expected or required of you and/or because it will bring you recognition (assuming people like it).
7. Satisfies needs of hosts - the hosts do not have the resources to prepare a multitude of dishes to feed a large group: that is why they have others doing it.
8. People will benefit from it - Hopefully, it will satisfy their hunger and perhaps acquaint them with new ways of preparing food which they may use in the future.

The solution requirements imposed by a funded project are very similar: your solution must (1) fit within an allotted budget, (2) be doable given your resources and (3) reliably produce the desired outcome. After determining the solution criteria, the next step in the process is locating the sources of information you will use to support your solution to the problem. In this example, these would be cookbooks, newspapers, magazines, television shows and personal cooking experience.

A number of resources will be required to accomplish your solution - resources required might be a stove, an oven, a beater, mixing bowls, measuring devices, baking disk, utensils, time keeper and the ingredients for the meal. Some of these resources will already be in your possession; some may be borrowed and the rest will need to be purchased from outside sources. Even though the list of resources you use is typically presented later in the proposal, the decision process occurs at this point in time. The feasibility of your solution will greatly depend upon the quality of the resources you can assemble.

LESSON SEVEN

Sampling Theory and Practice

OBJECTIVES

- Identify five types of sampling methods giving examples of when they would be applicable.
- Identify common sampling errors and how to correct them.
- Describe what is meant by “statistical significance” and “margin of error”

Key Terms

sample, census, simple random sample, stratified sample, cluster sample, systematic sample, judgment sample, convenience sample, quota.

Sampling theory and sample selection

Probably the most asked question researchers can think of is, “How large a sample should I take?” The size of a sample is depends on the research design and how precise the answers need to be. What is of most value are the methods for determining the minimum sample size needed for a particular type of analysis and a particular degree of precision.

In survey research, the primary decision the researcher has to make is “How close to the true response do I want the obtained response to be?” In experimental research, the primary decision the researcher has to make is “How small is the phenomena I’m searching for and what level of risk am I willing to take to find it?” In both instances, the end result is determining how much confidence a researcher can have that the results he or she gets is valid and not due to chance.

If you flip a coin, the chances of it being heads or tails is one out of two (1/2) or 50 percent. Does that mean that every time you flip a coin, you will get a head followed by a tail (or vice versa)? Of course not. You're more likely to get several heads or tails in a row. For any number of coin flips, the expected number of heads will always be equal to the expected number of tails. However, in practice the actual number of heads obtained will be more or less than the number of tails due to random variation.

What is random variation? In the coin flip example, each coin flip is independent of every other coin flip. The likelihood of it coming up heads or tails on each successive coin flip has nothing to do with the outcome of the previous flip. This is what is meant by random: every coin flip is equally likely to produce a head or a tail and each head or tail flipped is independent of every other head or tail flip.

A sample "statistic"

One of the most useful and common statistics that we experience in our daily lives is the arithmetic average of a group of numbers—also known as the mean. If you are a baseball fan, you are well aware of batting averages. If you are a football fan, you can probably recite the average yards per game of your favorite team. When you flip a coin 100 times, the average number of heads you expect to get will be 50. The average serves two purposes: one, it describes the value that is most typical of a group of numbers and two, it is the most likely value you'd get if you were to randomly select a number from a group of numbers.

When you randomly select a number from a group of numbers, that value may not equal the average but will be more or less or equal to the average. This variation from the average is a function of sampling. The more numbers you sample, the closer will the average of those numbers be to the average of the main group of numbers. The average of the sample is used to estimate the average of the population from which its drawn. The larger the sample, the more alike will the numbers (scores) be to those of the population, the less will be the variation, and the sample mean will be closer to the population mean.

The other source of variation mentioned above was measurement error. If we use a coin with one side heavier than the other, the results of the coin flips will be biased toward that heavier end and the outcome will not be as expected. This is the variation due to the nature of the measurement. The main purpose of statistical analysis is to identify and separate the true value from measurement error and random variance.

Identifying the target population

From the target population will come the people to be studied or served by the project. Populations are groups of people assigned on the basis of common characteristics. The population is defined by its characteristics. Thus, all males in the Jacksonville area would be one population; all black males, single, 20 to 25, currently unemployed, would be another. Obviously, the size of the population decreases as the number of different characteristics increases.

The concept of population in research has often been a source of confusion. The reason is because in statistics, population also refers to a group of numbers such as IQ scores, age, sex, race, blood pressure, number of arrests, and so on. In quantitative data collection, researchers make judgments about the mathematical nature of the data. In qualitative data collection, researchers make judgments about the people who are the source of the data.

This is not to say that quantitative designs ignore populations of people for the sake of measurement. Numbers would be meaningless if they were not related back to the people who produced them. That is why quantitative designs pay particular attention to selecting samples of people that are large enough to be representative and typical of their population.

Census or sample?

A sample is, by definition, a part of a larger population. When the entire population is available for study, the result is a census. Since statistics describe the characteristics of populations based upon sample characteristics, there are no statistics to be collected from a census. This may seem a bit strange to the novice researcher who plans on doing a statistical analysis of the data.

When speaking of population characteristics, researchers refer to them as *parameters*. When speaking of sample characteristics that are used to estimate population parameters, researchers refer to them as *statistics*. A *measurement* or *observation* is a specific event that occurs at a specific point in time. That event represents one of thousands of other similar events that could occur at different points in time. As you will learn, true scientists never trust a single measurement made at a single point in time.

The basis for reliability theory is that the measurements we make are actually samples of the total number of possible measurements that could be made at any given time. When we say that a method yields a reliable measurement, we mean that the method will repeatedly produce a similar sample of measurements regardless of when or how it is made. Reliability of measurement is an important issue given that one of the goals of research is to produce reproducible results.

Obtaining an ample sample

When researchers are unable to obtain or analyze an entire population of people or numbers, they will need to take a sample of the population. Other reasons for taking samples are as follows:

1. to obtain a manageable collection of objects to study
2. to provide a qualitative representation of population characteristics
3. to provide quantitative estimates of population characteristics (parameters)
4. to control for factors that are extraneous to the design and focus of the research

Obtaining a manageable collection of objects to study

Ever wonder why political researchers sample about 1,000 people for their surveys? The main reason is that it is a lot easier (and cheaper) than surveying all 260 million people in this country! While this example is perhaps a bit extreme, it illustrates the problem faced by all researchers: “How large of a sample can I afford to take?” The counterpart to that question is “How small a sample do I need to take?”

Larger samples require more resources to manage them. As with all projects, there are limits to the amount of resources a researcher can dedicate to the sample being surveyed. Researchers on a budget must know how much they can afford to spend. On the flip side, the costs of taking a sample that is too small may be more serious. The dangers of taking too small of a sample are that (1) due to people becoming unavailable or unreachable, the study cannot be completed as designed, (2) the phenomena under study are too rare to be measured in a small sample, (3) the analytical procedures used will be invalid or unreliable and (4) the results of the project will not be generalizable to larger populations.

The first steps in the sample selection process are to identify the goals of the sampling, define the limitations and constraints of the project budget, sampling methodology, and target population and determine the minimum sample size required for a reliable, valid and relevant analysis.

Providing a qualitative representation of population characteristics

A sample can be thought of as a picture of a population. All that can be known about the population must be inferred from the sample. The goal therefore is to draw a sample that represents the population on all relevant and important characteristics. For characteristics to be relevant they must adequately define the target population and relate to the design and purpose of the study. To be important characteristics, they must be ones which influence the outcome of the study. An ideal sample is one that differs from its population only in size.

Providing quantitative estimates of population characteristics

Here, too, the goal is to draw a sample that represents the population on all relevant and important characteristics. The difference between qualitative and quantitative characteristics of a sample is that the qualitative characteristics of the population are known while the quantitative characteristics typically are not. When some of the quantitative characteristics are known, then, these can be used to assess how representative the sample is of the population in much the same way as was done with qualitative characteristics

Generally, the quantitative characteristics of a sample (statistics) are used to estimate the quantitative characteristics of a population (parameters). The larger the size of the sample, the more precise will be the population estimates. Statistical analysis is based on the mathematical relationships between sample statistics and population parameters. By observing the outcomes of sampling, statisticians have developed mathematical models that can be used to determine the likelihood of each of those outcomes. For example, using a model of probability called the binomial expansion, statisticians can determine the likelihood of getting 75 heads in 100 coin tosses (given that the expected number is 50).

What the likelihood estimate is really indicating is whether that particular sample was drawn from a population with known values. The “population with known parameters” may be either a real population or a theoretical population of numbers created mathematically by repeated sampling of numbers similar in nature to that of the sample (as in the case of the binomial population of coin flips). If the mean of a population is known (for the binomial population or distribution, the mean is .50), then, it is a simple mathematical procedure to determine the likelihood that the mean of the sample represents (is similar to) the mean of the population.

Controlling for extraneous factors

In research, one of the key goals is to have a well-controlled study. Control means reducing or eliminating the effects from sources other than what is being directly studied by the researcher. Recall that data consists of two sources of error: measurement error and random error. Since random error is strictly a function of sampling, proper random sampling methods will ensure that the error from sampling will not exceed a known maximum. Measurement error is another story since error can arise from many sources in the process of measurement. A typical problem in social research are population characteristics that are interrelated such as race and income. If a researcher is interested in looking at the effect of race on a particular outcome, the design of the research needs to account for (or control for) the possible effect that income may have on that outcome given its strong association with race.

A special form of sampling could be used to reduce the likelihood of income being a factor. The technique is called stratified sampling. In this method, the population is divided into groups or strata based upon differing levels of a population characteristic. In the case of income, the technique might be to let five income ranges represent the strata and then to draw an equal number of samples from each strata. If, on the other hand, you were interested in measuring the impact of income on an outcome, you would want to ensure that your sample has the same levels of income in the same relative amounts (proportions) as the population. To accomplish this would also require stratified sampling with one difference: the number of values sampled in each level or strata would be according to the same proportions as the population. This how a sample of 300 would be drawn:

Income Range	**Percent of Population**	**Number in sample**
less than 10,000	35	105
10,000 - 19,999	25	75
20,000 - 39,999	20	60
40,000 - 59,999	10	30
above 60,000	5	15

Here is a list of the types of sampling procedures commonly used in research:

1. simple random - assign numbers to objects and then use a table of random numbers to select objects
2. systematic - first object selected at random, then every other object selected at equal intervals.

3. stratified - sampling done within levels of a variable
4. cluster - sampling done within clusters of like objects
5. multiple/sequential - sampling done in stages according to progressively specific criteria
6. judgment/quota - set number needed for sample selected on basis of physical characteristics
7. convenience sample - you take whatever you can get

Intact groups

In social research, many groups receiving services, treatment, or analysis were already in existence before the decision to research them was made. Consequently, these groups were not created by any sampling procedure. A classroom of students or traffic violators would be such a group. When people are assigned to a group on the basis of some action they take to be in that group, then they are what is called, “a self-selected sample.” An example would be people who volunteer for a parenting class.

The problem with intact groups or self-selected groups is that there is no control over their representation: they may be totally unlike the population from where they originate. Additionally, in self-selected groups, the reason for the existence of the group depends upon factors that are common among group members but not common among population members. In other words, any analysis of this group would be biased towards whatever factors led to its creation.

One of the methods used to control for extraneous factors in intact groups is to find or assemble another group with similar characteristics. This research method is known as a cohort study or matched pairs design. The logic behind this approach is to strive for a situation where the only difference between two groups of people is group membership and not individual characteristics. The ideal research design then would be two groups of clones. While there are no such things as human clones, there have been studies of twins raised together versus twins raised apart that essentially accomplishes this degree of matching individuals on the basis of their physical and mental characteristics.

Methods for determining sample size

In statistics there are some general rules of thumb regarding what is a small and what is a large sample. The number that divides small from large is 30. This doesn't mean that you will only need 30 people or values in any situation; nor does it mean that you will need that many. What is means is that for most samples of thirty, the characteristics of the sample begin to approach those of the population. The kinds of analyses that you intend on doing will dictate the size of the sample you will need. In every case, though, the size will be a function of how precise you wish your measurements to be.

Sample size in survey research

Determining a sample size for estimating a single parameter (like the average number of heads in 100 coin tosses) is easy and straight-forward. Determining the appropriate sample size for a questionnaire that may ask 100 questions (and each question could be considered a parameter), is not so easy and requires some compromises. The compromise occurs in making an assumption about the data that's collected.

If each item on a questionnaire will be analyzed apart from every other item, the researcher will assume that each item was obtained independently of every other item. Alternately, the researcher will assume that each item was answered independently of every other item. Why is this important? Because the mathematics of random sampling require that the selection of a single response from a population of responses be independent of every other selection.

Survey researchers speak of "statistical confidence" and "error rate." For example, political surveys typically will quote the percentage of people responding "Yes" and "No" to a question along with a "margin of error" that is so many percentage points greater or less than the stated amount. For example, in a sample of 800 to 1,000 persons, the margin of error (error rate) will be plus or minus 3.5 percent. It is 3.5 percent regardless of the question asked since the error is a function of the size of the sample. A sample size of 400 would have an error rate of plus or minus 5 percent. The smaller the sample, the greater is the margin of error.

Since the questionnaire is only performed one time, researchers would like to know how confident they can be in the results they obtain. If they were to repeat the sampling 100 times, how many times would they get a sample that yields similar results. When political researchers quote error rates, they also quote their confidence in that error rate. The concept of confidence also relates to sample size.

If you have a very large sample, you need not repeat the sampling process many times in order to capture the true values of the population characteristics. However, with large populations and small samples, you would need to do a very large number of repeated samplings in order to be sure of capturing the true values. If you were to repeat the sampling process 100 times and 90 of the 100 samples taken would produce the same percentages and error rates, you would be confident that you've obtained the true value 90 out of 100 times. This is what is meant by a 90 percent level of confidence.

When you see or hear the results of a political poll, you can rest assured that the researchers are working under a 90 percent level of confidence. Can you have higher levels of confidence, say 100 percent? Yes, but, it will require a larger sample size for the same margin of error. If we think of error rate as relating to the precision of a sample, then the level of confidence would be its accuracy.

Sample size in hypothesis testing

The essence of hypothesis testing is to determine whether a sample is drawn from a population with known characteristics or from a population with unknown characteristics. If we know how all values in a population are distributed, then we can tell if a sample is from that distribution by looking at the value of its characteristics. The characteristics of samples that are most important are the measurements that describe the sample as a whole: its mean and standard deviation. In doing repeated sampling, the value of the sample mean will vary above and below an average value: the average being the mean of the sample means. How much it varies is its error rate. The larger the sample, the less it varies and the smaller the error rate. Using a similar procedure for determining sample size for categorical responses (such as "Yes" and "No") we can determine what sample is required given a particular error rate desired.

Since we can estimate all values of a distribution given its mean and standard deviation, we can determine the likelihood of any value being part of that distribution. That is what is done in hypothesis testing. If we have a distribution of sample means and the standard deviation of that distribution, we can determine if the sample we have was drawn from that population by comparing the mean of the sample to the distribution of sample means. If the sample mean value does not fall anywhere within the range of values in the distribution, then we can say that the sample was drawn from a "different" population.

Fortunately for researchers, there are published tables that indicate the required sample size given the degree of precision one desires and the type of analyses performed.

LESSON EIGHT

Determining the Coding System

OBJECTIVES

- Develop a coding scheme that reduces a data to quantifiable terms.
- Compare and contrast different types of coding schemes
- List coding requirements of various computer programs

Key Terms

coding, alphanumeric, missing values, fixed field, free field, recoding, reconversion

The fine art of coding

Coding is the process of converting information into a quantifiable format (usually numerical) so that a systematic analysis of the information can be done. The coding process is often made easier by precoding responses and enabling circled or checked numbers to be entered directly into a data set. The codebook is like a foreign language dictionary in that it translates English responses into numerical or categorical values.

Since nearly all analyses of large data sets will occur on a computer, the coding system used should model or be easily convertible to the format required by the computer's software. If the information is categorical in nature, then, the choice of whether to represent a data value as a number or a label depends upon the analyses that will be done. If you are only going to measure frequency counts and percentages, then either a number or label is acceptable.

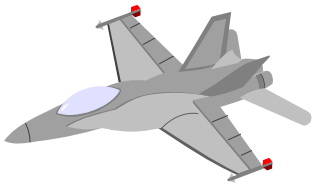
If you are going to do any statistical analyses, then numbers are mandatory. Labels cannot be manipulated mathematically. If you are collecting qualitative data and intend on doing content analysis, then your data elements might be entered as words and phrases instead of numbers.

When data can't be collected on one or more variables from one or more sources, then this creates the problem of missing values. Statistical computer programs deal with missing values in different ways. Generally, they give you the option of identifying what values are to be considered, "Missing," or allowing the computer to assume that zeroes or blank spaces represent missing values.

Should you assign a number to represent missing values? Yes, if you intend on either analyzing the missing values or replacing them with other values. However, you must inform the software what your missing values are before doing any analyses.

Data organization

When data is entered into a computer, it will generally be organized in a spreadsheet orientation where the rows represent different cases and the columns represent different variables. If you leave spaces blank instead of typing in a value, you may not know if the space represents a missing value or a missing data entry. By entering values in every space provided, you will know exactly what data has or has not been collected.



LESSON NINE

Formatting and Pilot-Testing the Survey

OBJECTIVES

- Describe the physical characteristics of the questionnaire
- List steps for grouping and ordering the questions
- Describe procedure for debriefing participants
- List types of information to be collected during pilot test

Key Terms

pilot test, debriefing, social usefulness

The format of the survey

Booklet format and printing procedures

Print the survey booklet on 81/2" x 11" paper. Place no questions on the front or back pages. The survey pages should be printed using a high quality laser printer on white or off-white paper.

Ordering the questions

Questions are ordered according to social usefulness or importance: those which people are most likely to see as useful come first and those least useful come last. Group questions that are similar in content. Establish a flow of responding from one question to the next. Questions in any topic area that are most likely to be objectionable to respondents should be positioned after the less objectionable ones. Demographic questions are usually placed at the beginning or at the end.

The first question is the most important. The first question should be clearly related to the survey topic and should be easy to answer. The first question should convey a sense of neutrality. The first question should be clearly applicable and interesting to everyone.

Formatting the pages

Use lower case letters for questions, upper case for answers. Identify answer categories on left with numbers—this allows precoding of responses. Establish a vertical flow. The purpose of vertical flow is to prevent inadvertent omissions, something that occurs often when respondents are required to move back and forth across a page with their answers. Vertical flow also prevents the common error of checking the space on the wrong side of the answers when answer categories are placed beside one another. Also, vertical flow enhances feelings of accomplishment.

The need to provide clear directions is extremely important. Use the same marking procedure throughout the survey. Directions for answering are always distinguished from the questions by putting them in parentheses.

Items in a series

Repeat the scale for each item. Ask one question at a time. The respondent should only be asked to do one thing at a time. The problem of asking two questions is that each request interferes with the other.

Use words for answer choices. Show a connection between items and answers. Use multiple column technique to conserve space. Show how to skip screening questions. Make questions fit each page. Use transitions for continuity -- for example, when a new line of questioning starts, when a new page starts or to break up the monotony of a long series of questions on a single topic. Transitions must also fit the situation. It is also useful to distinguish between major and minor transitions.

Designing the covers

The front cover receives the greatest attention and contains (1) a study title, (2) a graphic illustration, (3) any needed directions and (4) the name and address of the study sponsor. The title should sound interesting. Subtitles are often useful. Use graphic illustrations. The return address does not include the name of the researcher. The goal is to have the respondent view the researcher as an intermediary between the respondent and the accomplishment The back cover should consist of an invitation to make additional comments, a thank you and plenty of white space.

Why do a pilot-test?

The pilot-test is useful for demonstrating instrument reliability, the practicality of procedures, the availability of volunteers, the variability of observed events as a basis for power tests, participants' capabilities or the investigator's skills. The pilot test is a good way to determine the necessary sample size needed for experimental designs. From the findings of the pilot test, the researcher can estimate the expected group mean differences as well as the error variance. Even a modest pilot test conducted informally can reveal flaws in the research design or methodology beforehand.

Any surveys that have not been used in the past or have been modified in any way should always be pilot-tested. Any procedures that require complex instructions should be pilot-tested. Any methodology requiring time estimates should be pilot-tested.

Pilot-testing allows you to answer the following questions:

1. Is each of the questions valid?
2. Are all the words understood?
3. Are questions interpreted similarly by all respondents?
4. Does each close response question have an answer that applies to each respondent?
5. Does the questionnaire create a positive impression, one that motivates people to answer it?
6. Are questions answered correctly?
7. Does any aspect of the questionnaire suggest bias on the part of the researcher?

Selecting the pilot test sample

The sample for the pilot test should be as close as possible to the actual sample that will be drawn for the main project. When this is not possible, then you should try to get a sample with similar characteristics. Depending upon the availability of people, you may need to save as many participants for the main survey as you can—in which case, you don't want to include them in a pilot test.

Some researchers often will do a pilot test on a subset of their sample and then include them as part of the main sample. That is like mixing apples and oranges. If you make any change whatsoever to your study as a consequence of the pilot-test, then the participants in the pilot-test will have experienced something different from those in the main study. Additionally, one of the purposes of doing a pilot test is to debrief the participants after the study by asking questions about the methods, instruments, and procedures.

Information to be collected

The pilot test should be run exactly as if it were the actual study. The exception here is that you will be collecting data on how long procedures take, what actions facilitate or inhibit the operation of the study, whether instructions are understood and if the data you obtain is in the form expected.

It may be necessary to have more than one pilot test—especially in the situation where instructional materials or methods have been developed. In the case of instructional materials or methods, you would do a formative evaluation of the materials and methods. Unlike a pilot test where the researcher may not interact with participants, you would be asking questions of the participants as they read, watch, or listen to the instruction and when they are quizzed on what they have learned.

Participant debriefing

If your study involves questionnaires or interviews with people, you should have a debriefing session at the completion of the pilot test. Ask the participants if they understood all of the instructions, if they had any particular problem with any of the questions asked, if they understood the intent of the study, and if they had any recommendations how to improve the study.

LESSON TEN

Deciding How to Enter and Store the Data

OBJECTIVES

- Define the elements of a database
- Describe the steps in creating a database
- Describe how to use a form to enter data

Key Terms

database, table, records, fields

Creating the database structure

In a relational database, such as those created using a program such as Microsoft Access, you would store data from the surveys in *tables*. In Microsoft Access, a table is used to store data in records (rows) and fields (columns). The data from each survey is stored in its own row. Each column in the table represents the answers to each question on the survey. To bring the information together in a meaningful way, you would first decide what information you want to store in each table.

Generally, the first column of a data table will have an unique ID number that will identify the survey. This is to allow the researcher to go back and check the data on the survey with the data entered on the computer. The next columns of the table will usually contain the data from the survey in the order of the questions on the survey.

To make it easier to identify and manipulate the data, each column is given a unique short name that refers to the question number or subject on the survey. For example, a column labeled, “Grade” would refer to the TBOS question, “Highest Grade Completed,” on the survey. If the questions on a survey are numbered, then an easy way to label the columns would be “Q1, Q2, etc.”

When you create a Microsoft Access database, you create one file that contains the data and table designs as well as the forms, reports, and other things that make up the database. A data entry form is like an on-line survey: it has spaces or checkboxes specifically labeled that match the questions on the survey. The data entry form makes it easier to enter data because the data entry person is prompted by the database program as to what data to enter next.

If there are a lot of survey questions, then the data will take up many, many columns in the table. Thus, the data entry form will make the task of data entry a lot easier than having to enter the data a column at a time and a row at a time.

{Refer to the Appendix on how to use Microsoft Access to build a database}

Creating a data entry form

In Microsoft Access, you can create a form in two ways -- one, automatically by use of the *Form Wizard*, or two, manually. Before using either approach, it is best to sketch out on paper how you will want your form to look on the computer screen. Because you will be entering a lot of data and because the average computer screen shows about 23 lines of data, there will be several pages of the form displayed -- one at a time -- on the computer screen. You should decide how you want each page to look as well.

{Refer to the Appendix on how to use Microsoft Access to create a form}

LESSON ELEVEN

Data Entry, Data Checking and Quality Control

OBJECTIVES

- Define the variables of interest in the study
- Demonstrate procedures for developing a questionnaire
- Define the major methods of data collection in quantitative studies and qualitative studies
- Demonstrate how to determine if responses are random and representative

Key Terms

variable, coding scheme, tabulation, missing data, non-response bias

Data entry and quality control

Data entry is the process of transferring or transcribing the information from coding sheets, field notes, questionnaires and surveys onto the computer. Sometimes the process of data entry is facilitated by use of machine-readable sheets or by image scanners and optical character recognition software—in which case, manual data entry is not required. However, whether data entry is accomplished by person or by computer, there is a need to check the accuracy of the data (especially in the latter case). Since your analyses hinge on the accuracy of the data, good quality control procedures are needed for error checking and entry consistency.

There are computer programs that facilitate the quality control function of data entry directly by checking for acceptable values at the time they are entered. For example, if you recording ages of adults in a study, values lower than 18 would be errors. Manual checking of the data may or may not spot the error. What the data entry program would do is immediately reject any entered value that does not fall within the accepted range or is not of the proper form. If you don't have this program, then you might use the capability of your statistical analysis program to identify incorrect values and then to recode them.

Data coding problems

Missing data handling

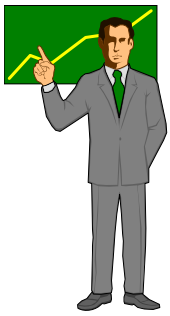
Prepare for nonresponse as treating it as one of the categories (example, a “No” response). Other situations, you might want to assign the average value to retain the rest of the data for analyses that require it. If you’re doing individual item analysis, then you’d drop the missing data—adding constants don’t change the overall results. For statistical purposes, you’d need to keep as much of the original sample size as possible.

Regression can be used to estimate missing data using answers to other responses. The problem may be avoided by having more than one question dealing with a topic. Not everything you ask will be of equal importance—some items you may afford to lose data. The main problem with nonresponse is that you don’t know if someone skipped a question or decided not to respond. Including a non-response category will ensure that people respond. Remind them both at the beginning of the questionnaire and at the end to answer ALL times. Don’t give people a way not to respond—give them the opportunity to respond. You can identify respondents who essentially are non-participants: people who mark the middle response for all items. Only slightly better than nonresponders. Do you keep them in the data bank? No, since their responses will skew the data.

Avoiding bias from non-respondents

It’s important to know whether those who did not respond differ greatly from those who responded. Relatively few refusals provide the theoretical potential for introducing considerable error into estimates of the sample characteristics. The extent of the differences between respondents and nonrespondents can seldom be determined. However, indirect methods may be used.

Mail questionnaires are also susceptible to other factors that affect response rate. Interest in the subject of the questionnaire may affect one’s answers and also determine if the questionnaire will be completed and returned. Another consideration is the ability of the respondent to provide written responses. People with less education are likely to be underrepresented, partly because of lower educational attainment, but also because of more difficulties with their seeing and writing capabilities.



LESSON TWELVE

Analyzing The Data

OBJECTIVES

- Discuss use of computer programs for data analysis
- Differentiate exploratory from confirmatory analysis
- Categorize the different levels of measurement and the tools used to analyze them
- Give examples of descriptive statistics
- Explain how graphical techniques can be used to analyze data.
- Given a data set, create a bar graph, pie chart and trend line.

Key Terms

content analysis, frequency distribution, mean, median, standard deviation, histogram, pie chart.

Descriptive analysis

Descriptive analysis involves describing the common underlying characteristics of data. In quantitative research, descriptive analysis involves arranging the data into a frequency distribution which groups each value into categories from low to high. If it is a normal distribution, then most of the values will fall towards the center of the distribution and decrease in frequency further out from the center. The two most important descriptive statistics of a normal distribution are the mean and the standard deviation. The mean is a measure of central tendency (in addition to the median and mode) and the standard deviation is a measure of dispersion (in addition to the range variance).

If you are analyzing nominal or categorical variables, then you would want to know how many and what percent of each value fell within each category. If you are analyzing ordinal variables, in addition to what you would show for categorical variables, you might want to know the average ranking of a variable. The average rank is calculated in the same manner as the mean except that ranks are added, not values. Ranking is a process of ordering values from low to high and then assigning a rank from 1 to n (where n is the number of values). In the Division 1-A College Football Poll, judges assign a rank from 1 to 25 to 25 of the 216 colleges in that division. The average of the judges' ranks then represents the actual ranking of a team in the poll.

An extension of constructing frequency distributions for one variable is to construct a combination of two related frequency distributions. The process is called cross-tabulation and the product is called a contingency table. The contingency table shows one variable's categories broken down by another. The contingency table is one if the ways that a researcher can assess the relationship between categorical variables.

To assess the strength and statistical significance of the relationship between variables, researchers have a number of procedures from which to choose. These procedures come under the heading of inferential statistics

LESSON THIRTEEN

Reporting and Presenting Results

OBJECTIVES

- Given a set of results, write conclusions that refer back to the questions raised in the study.
- Given a set of data and data analysis, choose the most appropriate presentation medium and format
- Describe the contents and format of the report.
- Given a set of conclusions and limitations, write a set of recommendations.

Key Terms

conclusions, interpretation, theoretical framework

The results and only the results

While some reports may combine the results and conclusions section, it's best to develop them separately. The results section should present the results of the analyses with minimal interpretation since the following section on conclusions will relate the findings to the research questions. Any charts, tables, graphs should be presented or referenced here. Charts, tables, and graphs should speak for themselves; i.e., the reader should be able to interpret them without textual explanation. A rule of thumb in discussing their contents is not to repeat in text what appears in the tables, charts and graphs. For example, you wouldn't say, "50 percent of the respondents answered, 'Yes'" if your table reports that. Also, you wouldn't say "the data shows an upward trend" if your line graph clearly indicates one.

The language of the results section should be short and sweet: say and show what happened and let the conclusions section take care of the interpretation and inference. The reasons for doing it this way is because no conclusions should be reached before all of the results have been digested—to do so would mean each outcome was evaluated in isolation of every other outcome. Your chance of understanding what all of the research can show would be minimized. Also, if there are inconsistencies among the results, you would be put in an untenable position to make a conclusion only to contradict it in a subsequent section.

Taking a step beyond—forming conclusions

When all the results are in, now is the time to make sense of it all. Each analysis should be related back to a corresponding research question and hypothesis. If your results supports your assumptions and answers your questions without any reservations, you should reaffirm that all went as planned. If the outcome was not as expected, then you'll need to explain any mitigating circumstances or give logical reasons why.

After relating your findings back to your original questions and research problem, you should then tie it into the theoretical framework presented in the related literature section. How does your research fit into what is already known? How does your findings expand upon or go beyond what is known? What will be the impact of your study?

Within the conclusions section (preferably at the end) would be a subsection focusing on the limitations of your research findings. Here is where you would discuss any deviation from the original proposed project and any unexpected outcomes. You also would remind the reader of the limitations mentioned earlier in the study

Making recommendations

After you've presented the results and conclusions, the logical question that would be asked by the reader (or the researcher) is, "What now?" In the recommendation section, you would discuss how to overcome the limitations of your research—how you (or others) would do it differently next time around. You would also discuss what sort of follow-up research, change in services or change in policy would be recommended in light of the results of your study.

The report format

In general, the final report should be written in the same style and format as was the proposal. Expanding your original format, you should include additional subheadings to represent the results, conclusions ad recommendations section. A general outline of a final research report might look like the following:

1. Cover sheet
2. Title page
3. Table of contents
4. List of Tables
5. List of Figures
6. Abstract
7. Introduction
8. Background
9. Purpose of the study
10. Research approach
11. Statement of the problem
12. Nature of the problem
13. Significance of the problem
14. Need for the study
15. Theoretical framework
16. Methodology
17. Research design
18. Target population
19. Sampling procedure
20. Discussion of variables
21. Description of data sources
22. Data collection methods
23. Instruments
24. Procedures
25. Data analysis techniques
26. Results
27. Conclusions
28. Recommendations
29. Summary
30. Appendices
31. Sample instruments
32. Tables

Attention to detail

Think of the appearance of your final report as a statement about you and your organization. Attractive, well-organized, good visuals—these are the qualities that your final report should have. The following are some guidelines for improving the appearance of the final report:

Always use standard, 8 ½ x 11 inch white paper

All pages, except the first page, should be numbered

All tables, charts and graphs should be consecutively numbered

Left margin should be at least 1 ¼ inches, top, bottom and right margins should be at least 1 inch

Use standard typestyles (fonts)

Right margin should be ragged, not justified, if typestyle is not a proportionally spaced font

Single-space the abstract (unless directed otherwise)

Do not use abbreviations. Place acronyms in parentheses following their full description

Tables should be interspersed with text only when they are essential to the understanding of the text